

ARTIFICIAL INTELLIGENCE & THE *MACHINE*-ATION
OF THE RULE OF LAW

Katie Szilagyi

Thesis submitted to the University of Ottawa
in partial Fulfillment of the requirements for a
doctoral degree in Law

Faculty of Law
Common Law Section
University of Ottawa

Dedication

*And what is good, Phaedrus,
And what is not good—
Need we ask anyone to tell us these things?
~ Robert Pirsig¹*

For Ian,
who helped me find my path.

&

For Andrew,
who held my hand as I walked along it.

¹ Robert Pirsig, *Zen and the Art of Motorcycle Maintenance: An Inquiry into Values* (New York, NY: HarperCollins, 1974) at x.

Acknowledgements

What if some day or night a demon were to steal into your loneliest loneliness and say to you: 'This life as you now live and have lived it you will have to live once again and innumerable times again; and there will be nothing new in it, but every pain and every joy and every thought and sigh and everything unspeakably small or great in your life must return to you, all in the same succession and sequence—even this spider and this moonlight between the trees and I myself. The eternal hourglass of existence is turned upside down again and again, and you with it, speck of dust!'
~ Friedrich Nietzsche²

In his thought experiment of eternal recurrence, Nietzsche asks us to think about whether we'd be willing to live our lives over and over, wishing nothing were different—not forward, not backward, not in all eternity. Early on this journey, I recall a conversation with Woody Hartzog about the potential decadence of the PhD experience: for the next few years, he emphasized, your only job is to read, think, and write! While I will remain forever grateful to have had the opportunity to indulge in such decadence, these years have contained more than their fair share of hardship, on both personal and global scales. I cannot have imagined I'd be completing this dissertation without the guidance of my beloved mentor and supervisor, nor during the throes of a global pandemic. All things considered, I just might choose to skip over these events the next time the cycle recurs.

But, as Nietzsche knew all too well, to skip over the hardship would be to miss out on all the joys. And the joys of this process were many. I am indebted to so many impactful individuals who contributed their intellect, passion, and ability to support my academic journey. I am blessed with wonderful friends, family, and colleagues—too numerous to name—who stepped in and stood up for me. I'll try my best to acknowledge them all here. (If I miss you, I'll catch you next time.)

Many of the ideas explored in this thesis first became part of my worldview in our SuperSixPack Robotnik Reading Group during my JD at the University of Ottawa, through my regular attendance at the wonderful We Robot conference, and during my LLM studies at Tel Aviv University. Special mentions to Kristen Thomasen and Sinzi Gutiu, for their steadfast friendship, and my LLM "Scholar's Workshop," which introduced me to Robert Cover's works. Thanks also to Ryan Calo, who first directed me to James Boyd White's *Heracles' Bow*, which became a cornerstone of my understanding of law as narrative.

My mentors and colleagues at the University of Ottawa supported my professional aspirations and encouraged me on this academic path. I'm especially grateful to Jeremy de Beer, Marina Pavlovic, and Michael Geist for their encouragement. Students in my seminar classes "Law, Technology, and the Future" and "Privacy Law" (at both uOttawa and later the University of Manitoba), were unsuspecting guinea pigs for many of my big ideas. Each of them contributed to creating vibrant intellectual environments where technology law ideas were circulated and shared. A similar environment was generated by our Algorithmic Justice Reading Group, comprised of Robin McLachlan, Ryan Mosoff, Anna Milner, Hong Nguyen, and Jon Khan, who always brought their

² Bernard Williams, ed, *Friedrich Nietzsche, The Gay Science* (Cambridge, UK: Cambridge University Press, 2001) at 194. Originally published in 1882, this quotation is often referred to as "Aphorism 341: The heaviest weight."

best selves to discussions of technology law. My years at uOttawa were also marked by regular accountability check-ins with Suzie Dunn, which helped keep my progress on track.

During these uOttawa years, I was grateful to be a recipient of the Social Sciences and Humanities Research Council's Joseph-Armand Bombardier Canadian Graduate Scholarship, as well as entrance awards and research bursaries from uOttawa, and the Canada Research Chair in Ethics, Law, and Technology. I benefitted from summer research in Brazil with Rio's Institute of Technology and Society, supported by uOttawa's Centre for Law, Technology and Society. Special mention to Florian Martin-Bariteau for his efforts. Thanks also to the members of the Canadian Robotics and AI Ethical Design Lab (CRAiEDL) for many fruitful interdisciplinary discussions.

My final year of this PhD was spent as an Assistant Professor at the University of Manitoba, where I am again lucky to be surrounded by supportive colleagues and friends. In particular, our Writing Club comprised of Kjell Anderson, Martine Dennie, Gerard Kennedy, and Amar Khoday was instrumental in my completion of this dissertation. Thanks also to the Pomodoro Method, Tycho, and Big Red Machine for creating the right writing ambiance.

In addition to my committee members, my draft documents benefitted from feedback generously contributed by Christina Fawcett, Gerard Kennedy, Jon Khan, Amar Khoday, Teresa Scassa, and James Young. Special thanks are owed to Christina for her Herculean efforts in assisting with proofreading and grammatical copyediting. (Eventually, I'll get the hang of those blank pronouns!)

My doctoral advisory committee of Peter Oliver, Jane Bailey, and Jason Millar were intellectually rigorous, unfailingly thorough, and deeply kind in executing their duties. I am appreciative of their considerable efforts in shepherding me through this grief-tinged process with empathy, compassion, and attention to detail. Peter, in particular, gave generously of his time with extensive written feedback and thought-provoking Zoom conversations. Many thanks also to Carys Craig for agreeing to serve as the external member of my examining committee.

I am endlessly grateful to my family, whose unfailing love, support, and encouragement helped me get over the finish line. I survived Summer 2022 thanks to my mom's cooking and pep talks; my dad's 21-second hugs; and Andrew's relentless positivity, razor focus on my schedule, and trips to Starbucks for my favourite London Fogs. I will never fully express my love or gratitude. Thanks to Krista Zeman for being my Ottawa family, for all the snack innovations, popcorn creations, and constant conversations. Special mention to my baby pingüino, also forthcoming 2022, who gave me the deadline I so desperately needed, while also making this process unimaginably more difficult. (I'm going to love you so much, baby, but you're the worst co-author ever.)

Above all, this work would not exist without the incredible guidance, mentorship, and inspiration provided by Ian Kerr. His influence on my life's journey has been unparalleled; I hope that my invocations of his work in this thesis will both do his *oeuvre* justice and continue some the stories he had only just started telling. I certainly missed having him cast his thoughtful eye over these pages, but not nearly as much as I miss his quick wit, his taste in music, his infectious laugh, and his genuine friendship. Ian: I cannot fathom how differently my life would have turned out but for our conversation at law school orientation, all those years ago. I'll keep working on becoming who I am. Thank you for believing in me. Once more and innumerable times more: *amor fati*.

Abstract

In this dissertation, I argue that the Rule of Law is made vulnerable by technological innovations in artificial intelligence (AI) and machine learning (ML) that take power previously delegated to legal decision-makers and put it in the hands of machines. I assert that we need to interrogate the potential impacts of AI and ML in law: without careful scrutiny, AI and ML's wide-ranging impacts might erode certain fundamental ideals. Our constitutional democratic framework is dependent upon the Rule of Law: upon a contiguous narrative thread linking past legal decisions to our future lives. Yet, incursions by AI and ML into legal process—including algorithms and automation; profiling and prediction—threaten longstanding legal precepts in state law and constraints against abuses of power by private actors.

The spectre of AI over the Rule of Law is most apparent in proposals for “self-driving laws,” or the idea that we might someday soon regulate society entirely by machine. Some academics have posited an approaching “legal singularity,” in which the entire corpus of legal knowledge would be viewed as a complete data set, thereby rendering uncertainty obsolete. Such “regulation by machine” advocates would then employ ML approaches on this legal data set to refine and improve the law. In my view, such proposals miss an important point by assuming machines can necessarily outperform humans, without first questioning what such performance entails and whether machines can be meaningfully said to participate in the normative and narrative activities of interpreting and applying the law. Combining insights from three distinct areas of inquiry—legal theory, law as narrative scholarship, and technology law—I develop a taxonomy for analysing Rule of Law problems. This taxonomy is then applied to three different technological approaches powered by AI/ML systems: sentencing software, facial recognition technology, and natural language processing. Ultimately, I seek the first steps towards developing a robust normative framework to prevent a dangerous disruption to the Rule of Law.

Table of Contents

Dedication	ii
-------------------	-----------

Acknowledgements	iii
-------------------------	------------

Abstract	v
-----------------	----------

Table of Contents	vi
--------------------------	-----------

Introduction	1
---------------------	----------

1. Thesis Overview	3
a. Part I: Coding the Law	3
b. Part II: Complicating the Law	4
c. Part III: Applying the Law	4
2. Theory & Methodology	6
a. What methods are used?	6
b. What theoretical framework is used?	7
c. What does this thesis demonstrate?	10
3. A Pre-Emptive Response to Critics	10

Part I: Coding the Law	16
-------------------------------	-----------

1 Introduction to Artificial Intelligence & Machine Learning	17
---	-----------

1. How Artificial Intelligence Works	20
a. How Machine Learning Works	23
b. How Data is Collected & Interpreted	26
2. AI and Governance	29
a. AI and Legal Regulation	30
b. AI and Corporate Power	32
3. AI and the Legal Futurists	33

2 What Cyberlaw & STS Might Teach	39
--	-----------

1. Lessons from Cyberlaw	41
a. The Saga Begins	41
b. Complications & Elaborations	43
c. Applying Cyberlaw's Lessons to Robotics & AI	44
2. Lessons From STS	47

a.	Technology as Prime Mover	47
b.	Technology as Neutral	50
i.	Technological Neutrality in STS Literature	50
ii.	Technological Neutrality in the Law	53
c.	Technology as Solutionism	55
3.	Marching Orders for the Future of Law	57
a.	Conceptualizing Quantification	58
b.	Understanding Affordances	59
c.	Generating Normativity	61
4.	The Cyberlaw & STS Curriculum	64
3	Delegating Decision-Making	67
1.	Introducing AI to Legal Spaces	68
a.	Framing Example: Red-Light Camera Deployed as “Automated” Police Officer	69
b.	Datafying the Administrative State	70
c.	Automated Analogues for Human Decision-Makers	74
d.	Proposed Regulatory Frameworks	76
2.	Complications for the Data Set	80
a.	Desperately Seeking Fairness	80
b.	What the Judge Had for Breakfast	82
c.	Problems of Induction and The Denial of Ignorance	84
d.	Traffic Bot in the Loop: What Does This Example Show?	87
3.	Mechanical Jurisprudence	90
Part II: Complicating the Law		95
4	The Rule of Law	96
1.	What is the law supposed to do?	98
a.	Theoretical Takes on Law & the Rule of Law	101
b.	Ronald Dworkin’s Law as Integrity	107
2.	Can the law offer right answers?	109
a.	Problem of hard cases	111
b.	Law as a Chain Novel	113
c.	Dworkinian Takeaways	115
3.	Rule of Law Conclusions: A Call for Storytelling?	117
5	Law as Narrative	121
1.	Law as Narrative Scholarship	122
2.	Lessons from Critical Theory	127
a.	Feminist Narratives	127

b.	Indigenous Narratives	129
c.	Critical Race Narratives	132
3.	Law as Storytelling	137
Part III: Applying the Law		142
<hr/>		
6	Sentencing Software	143
<hr/>		
1.	Risk Assessment Via COMPAS	145
a.	COMPAS Design Documentation	146
b.	COMPAS Critiques	147
2.	Rule of Law Problems	151
a.	Easy Answers to Hard Cases: What Process Dominates?	151
b.	Acknowledging Contested Concepts: Which Values Dominate?	155
c.	Constructing Truth: Whose Story Dominates?	156
3.	Storytelling Solutions	158
<hr/>		
7	Facial Recognition Technology	164
<hr/>		
1.	Facial Recognition Technology	166
a.	FRT Example: Clearview AI	169
b.	FRT Use in Predictive Policing	172
c.	Equality-Seeking Critiques	174
2.	Privacy Law's Legacy & Technological Pinch Points	177
a.	Criminal Law Dimensions	179
i.	<i>R v Jarvis</i>	180
b.	Tort Law Dimensions	181
i.	Appropriation of Personality	181
ii.	False Light	182
iii.	Public Disclosure of Private Facts	182
c.	Statutory Dimensions	183
3.	Rule of Law Problems	185
a.	Easy Answers to Hard Cases: What Process Dominates?	186
b.	Acknowledging Contested Concepts: Which Values Dominate?	189
i.	Privacy as Contextual Integrity	190
ii.	Privacy as Consent Management	190
iii.	Privacy as Trust/Loyalty	191
iv.	Applying Theory to FRT	191
c.	Constructing Truth: Whose Story Dominates?	193
4.	Storytelling Solutions	194

8	Natural Language Processing	199
1.	Natural Language Processing	200
a.	Equality-Seeking Critiques	204
2.	Rule of Law Problems	206
a.	Easy Answers to Hard Cases: What Process Dominates?	207
b.	Acknowledging Contested Concepts: Which Values Dominate?	210
c.	Constructing Truth: Whose Story Dominates?	212
3.	Storytelling Solutions	214
Conclusion		218
1.	Argument Summary	218
2.	Future Directions	219
3.	Concluding Thoughts	220
Bibliography		222

Introduction

I begin with a quote from Joan Didion: “we tell ourselves stories in order to live.”³ Our lives and the law are inextricably woven: elaborate stories assigning value, amplifying perspectives, and creating meaning. As critical and Indigenous traditions recognize, law and culture are inseparable; therefore, law and stories are inseparable. As we strive to assign value, amplify perspectives, and create meaning through the law, we acknowledge that we weave cultural threads of human endeavour into the tapestry.

Those cultural threads of human endeavour might also be described as technology. Our technological advancements govern social, political, and even legal norms. While some argue for technology’s neutrality, those assigned values, amplified perspectives, and created meanings can influence technology’s trajectory and application. And, the inverse is also true, as technology pushes back to offer its own stories. Today’s dominant technologies, like ubiquitous computing, sensor-enhanced “smart” applications, and networked devices, allow data-driven approaches to proliferate. Artificially intelligent or, more accurately, machine learning models are powered by the wealth of data such technologies provide. Business practices, government initiatives, and scholarly efforts increasingly resort to this deep well of data to define the next best move. It is perhaps inevitable that this same datafication is beginning to extend to the law; but the introduction of artificially intelligent law invites significant questions for the stories law tells.

Emerging literature on the use of artificial intelligence (AI) to govern, administer, or maintain law reveals an incoherence in what people mean by the term “law”, and more specifically, when they claim the maintenance of the “Rule of Law.” This thesis asks how delegating decision-making authority to AI impacts both our conceptions of law and the Rule of Law, and what a richer conception of the Rule of Law implies regarding AI. I assert that a better understanding of the legal context in which AI is being applied will help us to better understand whether AI applications promote or detract from the Rule of Law. What we take to be the values collectively referred to as the Rule of Law are rooted in different views on legal theory. Thus, the law’s normative aspect should not be presumed; it must be interrogated.

Writers help us make sense of the world that we inhabit. They offer acute observations about our sensibilities, our shortcomings, our collective delusions, our unintended conceits. Good writing can transform the way we see the world. In the hands of writers, mere lives become stories; mere conversations become narrative. And what is the law, other than an especially elaborate story that so many authors—lawyers, legislators, judges, juries, everyday citizens—continually co-write together: creating a story that grows and changes over time. Various rules and principles constrain how we tell that story, but one is traditionally considered paramount: the Rule of Law.

Writers need tools, and technological innovation has transformed the reach of writers. And far be it for me to decry the march of progress, but I started to feel concern about the authenticity of the voices telling these stories around the time that my phone started offering suggestions of what I should type next. The true challenge artificial intelligence presents for the information age is not

³ Joan Didion, *The White Album* (New York, NY: Simon & Schuster, 1979) at 1.

a hostile takeover by the fully embodied robots envisioned by Asimov,⁴ but the subtle ways that machine learning manipulates data in cyberspace. Today's online environment is unforgivingly shaped by enormous corporate conglomerates,⁵ and is replete with corresponding behaviour modifying technologies.⁶ The ways in which we forge relationships, communicate information, engage in commerce, and create culture have all been subtly shaped and reordered by algorithms. All of these machinations—or *machine*-ations,⁷ left unchecked, threaten to destabilize our core institutions, impoverish our existing models for technology regulation, and lead to an unintended reordering of our normative framework for the Rule of Law.

This project seems especially urgent in an era when disinformation makes its rounds on social media, radicalizing unsuspecting Facebook users and preventing any vestigial veneer of social cohesion. With proprietary algorithms held by private companies, there is no oversight to ensure the data being collected and deployed is representative of society, or free from bias. Literature on the disproportionate impact of algorithmic decision-making on racialized, minority, and other equality-seeking groups also gave me pause.⁸ If decisions were to be made in the justice system, what data would it harness? How would it be collected? Would it be in the interests of justice? And what stories might that data tell about the people it proposed to represent?

The discovery of Robert Cover's influential piece, "Nomos and Narrative," first bridged the gap I observed between theoretical accounts of rule-following and modern algorithmic dispensations of rules.⁹ Cover situates the legal universe within a normative structure designed to bring the stories we tell about the law to the forefront.¹⁰ Understood in the context of corresponding narratives, epics, and scriptures, he writes: "law becomes not merely a system of rules to be observed, but a world in which we live."¹¹ His account of law suggests a participatory social exercise, wherein the structure designed by legal rules offers society meaning.¹² This explanation reminded me of Dworkin's observations about the law as a chain novel being contiguously generated by an ongoing succession of interchangeable authors.¹³ There, I observed, was perhaps some kind of roadmap for applying Cover's method of narrative as an interpretative tool—one that is engaged by imposing normative force on the world around us.¹⁴ Though Dworkin's proposed chain novel is decidedly acontextual—and does not anticipate stories imbued with critical perspectives and lived

⁴ In 1950, science fiction writer extraordinaire Isaac Asimov published his collection of short stories, "I, Robot", envisioning a world in which bots worked for humans and operationalized with three paramount laws. See generally, Isaac Asimov, *I, Robot* (New York, NY: Bantam Books, 1950).

⁵ Five major businesses control the preponderance of the cyberspace marketshare: Amazon, Apple, Facebook (now aka Meta), Microsoft, and Alphabet (the parent company of Google). For discussion on this point, see: Alex Wilhelm, "Tech's 5 Biggest Players Now Worth 3 Trillion," (19 July 2017), online: *TechCrunch* <techcrunch.com/2017/07/19/techs-5-biggest-players-now-worth-3-trillion/> [perma.cc/G8FK-5E7H].

⁶ Paul Lewis, "'Our minds can be hijacked': the tech insiders who fear a smartphone dystopia" (6 October 2017), online: *The Guardian* <www.theguardian.com/technology/2017/oct/05/smartphone-addiction-silicon-valley-dystopia> [perma.cc/63FB-JHFF].

⁷ My term *machine*-ations refers to the computational processes performed in the background by artificial intelligence systems, the first step towards a possible future where many societal systems, including the legal system, are regulated by machine.

⁸ See, e.g., Safiya Noble; Ruha Benjamin; Cathy O'Neil; Meredith Broussard; Virginia Eubanks, all discussed herein.

⁹ Robert Cover, "Nomos and Narrative" (1983), 97 *Harvard L Rev* 4 at 5 [Cover, "Nomos"].

¹⁰ *Ibid* at 8.

¹¹ *Ibid* at 5.

¹² *Ibid* at 8.

¹³ Ronald Dworkin, "Law as Interpretation" (1982) 9 *Critical Inquiry* 179.

¹⁴ Cover, "Nomos," *supra* note 9 at 10.

experiences—I nonetheless found the metaphor apt. It offered a convenient bridge between critical scholarship and the Rule of Law, one I hoped would ultimately inform this thesis.

1. Thesis Overview

a. Part I: Coding the Law

This thesis begins with a thorough literature review. In Part I, I lay the groundwork for further discussion by explaining how AI and its primary enabling technology, machine learning (ML), use the ongoing datafication of social and economic life to encode various processes. The conversion of human activities to data points, enmeshed in a new surveillance capitalism, has permitted the dominant AI industries to develop highly marketable technological tools. Yet, it remains unclear whether these tools are beneficial to society. Since the advent of Internet-enabled activities, code has been prescribing human actions to follow particular scripts. In the era of ML, code has now become capable of manipulation; it is now able to transform, nudge, reformulate, and transgress existing boundaries. These predictive capabilities might someday pre-empt meaningful human choices, particularly if decision-making structures are built to be data-driven. The first three chapters of this thesis set the stage for the forthcoming discussion.

In **Chapter 1**, I lay out the scientific bases for AI and ML, explaining the complexities of these technologies in straightforward language with clear examples. While the potential of AI and ML are often written about with breathless enthusiasm, I keep this introductory chapter grounded in reality, offering real-world examples of how technologies are currently being deployed. I summarize major scholars' views on how AI and ML might disrupt governance structures and instances of power, including their impacts on legal regulation and market economics. I conclude by explaining the enthusiastic worldviews of the legal futurists, who would gladly allow law's empire to be regulated by machine, and note some preliminary critiques of this way of thinking.

In **Chapter 2**, I map the scholarly landscape of technology law, with particular focus on the regulatory emphasis of cyberlaw scholarship and the sociotechnical insights of science and technology studies (STS). A new tradition of legal academia grew up around commenting on the transformative impacts of the Internet, widely considered to be an exceptional force that challenged some traditional legal ideas. I summarize the key reflections of this vein of scholarship, situating my upcoming AI/ML analyses within a framework of lessons gleaned from cyberlaw and STS. I bring focus to two key concepts: technological affordances and normativity, which both offer additional nuance to considerations of law within a technological context. I push back against assumptions of technological neutrality in either law or design, demonstrating that this approach neglects technological realities and embedded expressions of power.

In **Chapter 3**, I review the current uses of AI and ML within the legal system, using the administrative state as a laboratory to think through ideas about human judgement, bias, and delegated decision-making. This chapter shows the limitations of existing AI and ML techniques, while still seriously considering the possibility that judicial decision-making is limited by human frailty. Noting insights from legal realism and behavioural psychology, I review the decision-making challenges that tend to generate calls for legal regulation by machine methods. To begin

making the transition to doctrinal legal scholarship, I conclude by engaging with theoretical accounts of robotic rule-following and mechanical jurisprudence.

b. Part II: Complicating the Law

Having established the relevant AI- and ML-informed vocabulary, I move on to developing a robust theoretical foundation, using analysis and discussion of two schools of thought within legal academia: analytical jurisprudence, specifically contested accounts of the Rule of Law; and law as narrative. In Part II, my goal is to remind readers that our vision of law is complicated. I explore the nuances of various contested concepts and the competing explanations they offer for certain legal phenomena. Through this conceptualization of law, I demonstrate that law is something greater than merely right answers, thereby making its administration by machine difficult, if not impossible. Going beyond the assumptions of the legal futurists, I analyse the problems presented by hard cases, thinking through law's role in resolving issues that are unclear or unprecedented.

In **Chapter 4**, I engage with Rule of Law scholarship. I begin with some of the conventional scholars writing in this space, including HLA Hart, Ronald Dworkin, Joseph Raz, Lon Fuller, and others. I set out a fairly conventional framework for thinking through Rule of Law issues, while highlighting some of the longstanding tensions from doctrinal legal scholarship. Through this analysis, I demonstrate the fractious attempts to define fundamental legal concepts, thereby frustrating any assumption that law could be easily transitioned to regulation by machine. While I adopt parts of Dworkin's interpretative theory of legal analysis, I also lay out its shortcomings, thereby setting the stage for a reparative, critical approach in the next chapter.

In **Chapter 5**, I discuss the thread of legal scholarship known as law as narrative. This is where I unpack Cover's "Nomos and Narrative," coupled with insights from James Boyd White's *Heracles' Bow*.¹⁵ I connect these ideas with the work of scholars from feminist, Indigenous, and critical race traditions, all of whom have emphasized the importance of centering excluded narratives and articulated the need for storytelling within law. I also emphasize critical accounts from technology law scholarship, where the objectivity proclaimed by the legal futurists is frustrated by data-driven approaches that replicate and extend social bias. I argue that incorporating narrative into a broad, relational concept of law allows context-sensitive storytelling to appropriately complicate our perspectives on what law is, casting further doubt on the potential for effective legal automation.

c. Part III: Applying the Law

After setting out the scientific and theoretical foundations in Parts I and II, and identifying key themes, terms, and approaches—both technological and legal—I apply these lessons to three different case studies. Each of these is afforded its own chapter in Part III. The three situations that I describe each operationalize the Rule of Law in situations where AI/ML technology is directed towards streamlining the administration of justice. I will refer to this overall process as "regulation by machine" and to the specific processes themselves as "self-driving laws." The use of "self-driving laws" as a plural concept refers to the fact that only parts of the legal system use automated technologies. Reflective of a view of the law as an ongoing, dynamic thing, only certain

¹⁵ James Boyd White, *Heracles' Bow: Essays on the Rhetoric and Poetics of Law* (Madison, WI: The University of Wisconsin Press, 1985) [White, *Heracles' Bow*].

aspects of the law are being impacted through automation. Because I am interested in showing both the impacts of AI/ML technology on the Rule of Law, and the insights that a Rule of Law lens might offer for emerging technology, I target three specific types of technology to underscore those impacts. To reinforce this bidirectional analysis, I select AI/ML technologies that engage different legal arenas, to increase the breadth of the conversation.

In **Chapter 6**, I consider sentencing software. This area of algorithmic justice has received considerable academic attention, making it fertile ground for significant descriptive review and analysis. Since judicial sentencing is necessarily discretionary, and given that the social sciences have often charged sentencing procedures as being fraught with bias, sentencing software offers a good starting point for my methodology. If sentencing is already biased, using a data scientific approach to ameliorate the problem may seem especially logical. In this chapter, I examine the much-maligned sentencing software program COMPAS to articulate the Rule of Law problems that aggregate in the criminal law context. Through these examples, I draw the constitutional dimension of self-driving laws into focus.

In **Chapter 7**, I assess facial recognition technology and its privacy law dimensions. This discussion builds on the analysis in Chapter 6, as criminal and constitutional concerns still arise with facial recognition technology. Yet, by taking a broader privacy law lens, other areas of law are also implicated. Privacy law is especially interesting because of its uneasy presence in the common law landscape generally. It is part statute, part tort, part criminal law with a quasi-constitutional dimension, part contractual obligation, part nebulous theoretical construct. Throughout legal history, privacy norms are typically adjudicated at moments when new technologies increase human capabilities in unanticipated ways. Privacy law's malleability to sociotechnical forces makes it an area of frequent commentary for technology law academics, as well as a space where law-related technology might serve to transform the law itself. Using the example of facial recognition technology, and a broad lens on the types of legal arenas that are implicated by privacy concerns, I identify the significant scope of Rule of Law problems generated using AI/ML technology.

In **Chapter 8**, I analyse natural language processing techniques, thinking ahead to a future where law might be administered using this tool. Going beyond the previous two chapters, this discussion takes the analysis squarely to the realm of the legal futurists, imagining what might happen if ML methods were permitted to autonomously draft laws or legal judgments. I use the online natural language processing tool GPT-3 to respond to some of this thesis' key inquiries, critiquing its responses to the same questions that have guided my considerations.

In each of my three case studies, I use the insights gathered in Part II to categorize and evaluate Rule of Law problems, using a taxonomy developed from the discussions contained therein. Then, I deploy law as narrative as a recuperative relational technique, attempting to repair the emergent Rule of Law issues arising in the self-driving laws scenarios. I argue that self-driving laws do not contend with the normative exercise that law requires, or the degree to which the Rule of Law has additional prospective meaning. It is here, I assert, that we see the true failing of self-driving laws: by using the past to predict the future, we are unable to imagine any possible future. We are instead resigned to a meagre version of Dworkin's Hercules, writing a predictable, plodding version of legal process, never executing social change nor reimagining what the future we might want should look like. We are confronted with the risk that future technology bidirectionally moulds the law,

even if we think we are maintaining the status quo or following a predictable evolutionary path. To preserve the integrity of our legal universe, we will likely want to reject self-driving laws as an overall governing mechanism. Overall, we must tread carefully to ensure that the normative frames smuggled in through AI/ML technologies' affordances do not wreak havoc on the Rule of Law.

2. Theory & Methodology

a. What methods are used?

In this thesis, I employ a combination of methods. First, I inform readers about how AI works, canvassing material that will be familiar to those working in the area, but that is perhaps unfamiliar to scholars focused on other areas of legal scholarship. I remind readers about what existing literature offers in three key areas: technology law scholarship; Rule of Law scholarship; and characterizations of the law as narrative. By combining these broad categories, I aim to generate insight into the nature of law itself, querying what might be lost if decision-making authority is delegated to algorithmic structures or if, alternatively, the implications of AI/ML are naively ignored. By fusing jurisprudential inquiry with studies of emergent technology, I seek to widen the analytical scope of how technology law is typically analysed. If law is culture, and culture is language, then society must conduct important normative inquiries before transposing that language into code.

Each of these three scholarship areas—technology law; Rule of Law; and law as narrative—offer unique insights to the broader theme of AI governance. I conduct qualitative analysis, confronting the research of key theorists in each field. In the technology law space, I map the insights offered from the fields of cyberlaw and STS. Some scholars have begun confronting questions of methodology to better guide law and technology scholars. Notably, Ryan Calo¹⁶ and Jack Balkin¹⁷ have set out contrasting methodological approaches, differing on the question of whether to treat certain technologies as exceptional. Rebecca Crootof and BJ Ard have proposed an innovative methodology for structuring technology law: they propose eschewing the framings of technological exceptionalism, or its inverse, in favour of identifying and resolving the legal uncertainties that arise when a new technology is deployed.¹⁸ I purport to do something similar. I offer comments on these methodological techniques, aiming to enrich my own analysis by clearly situating it within the broader technology law conversation. I do not intend for my methods to apply to the study of technology law generally; rather, I propose to merge these disparate views into a new lens on the Rule of Law.

The rich subject areas of cyberlaw and technology law have traditionally had little nexus with the fields of analytical jurisprudence or doctrinal study of the Rule of Law; by walking carefully through this landscape, I establish new insights that go beyond mere description. By comparing AI/ML to the last allegedly transformative technology of our time, the Internet, I offer a meaningful comparator for legal analysis and contrasting situations against which to measure the utility of

¹⁶ Ryan Calo, "Robotics and the Lessons of Cyberlaw" (2015) 103 Calif L Rev 513 at 514 (arguing that the law must treat as an exceptional technology, with certain "essential qualities" that create problems for the law) [Calo, "Lessons of Cyberlaw"].

¹⁷ Jack Balkin, "The Path of Robotics Law" (2015) 6 Calif L Rev Cir 45 at 46 (arguing that Calo's approach wrongly emphasizes exceptionalism and distinguishing what is essential about the technology from what features from human social life the technology makes "newly salient").

¹⁸ Rebecca Crootof & BJ Ard, "Structuring Techlaw" (2021) 34 Harv JL & Tech 347.

various explanatory theories and metaphors. Through these fields of study, I point out the similarities and differences between AI and previous technologies, applying lessons learned where possible and defining new, different issues that AI presents for the first time. Where controversies exist, I attack them head on. In the Rule of Law analysis, I grapple with the questions of what the law is supposed to do and whether it can provide objectively correct answers to all manner of legal questions. Finally, I present law as narrative as a theoretical innovation for contextualizing the challenges of algorithmic decision-making. In this section, I take a broad approach to narration, incorporating critical perspectives of law from feminist, Indigenous, and critical scholars. In combination, these three schools of thought present a novel method for analysing the law itself and for mapping the impacts of AI/ML technologies in the legal realm.

These lessons are then applied to the three case studies, each of which engages a different ML-enabled technological innovation: sentencing software, facial recognition technology, and natural language processing. The predictive capabilities of ML coincide with the Rule of Law when scholars argue in favour of using such techniques to streamline the administration of justice. Through these case studies, I aim to ask constitutive questions about the law for the first time. In which situations might it be appropriate for a machine to replace a person? How will justice be served? In which computer language might the law be written?

Ultimately, through this thesis, I seek to provide what is lacking in previous incarnations of statements of the Rule of Law. To do so, I develop a taxonomy to test how the implementation of different AI applications conforms to the Rule of Law. This pursuit, I argue, is more practical than opining on whether the law accords with particular conceptions of justice, which are unhelpfully abstract, highly varied, and deeply contested. However, it is difficult to approach questions about the Rule of Law without implicating literature that aims to differentiate between various concepts of law. This method points to the need to identify a theoretical framework, which in turn presents additional methodological questions.

b. What theoretical framework is used?

This thesis theorises and models the use of AI applications using a broad umbrella of what constitutes law. To obtain the necessary specificity, I engage with traditional analytical jurisprudence to eventually set out my own version of what law is. While this thesis is not intended as a theoretical treatise, I nonetheless set out certain parameters of what I mean when I write about the law, and what law is trying to do, in order to diagnose the shortcomings inherent with unduly delegating decision-making authority to AI/ML. This thesis is also not intended to set out a new concept of law, which would require more robust theoretical work. Instead, I propose a system of law not to advocate relentlessly for that system, but instead to ensure that readers are clear on what I refer to when I write about “law.”

I use a concept of law heavily influenced by a Dworkinian perspective. I choose this framework for two reasons: first, because I believe it to have substantial traction in the dominant contemporary legal worldview; and second, because I believe it to be a reasonable account—with some shortcomings, which I propose to discuss. Because Dworkin aims to respond to both undue

formalism and free-wheeling legal realism,¹⁹ his conceptualization is well-suited to some of the charges I wish to make against an algorithmic conception of justice. Furthermore, using Dworkin's theory of law as interpretation allows me to harness one of his chief theoretical innovations: the idea of law as a chain novel. Dworkin proposed a theory of law as integrity, asserting that the arc of the law contains fundamental fidelity to principle, which must be adhered to by legal officials.²⁰ He analogized the development of the law to a chain novel, in which a series of anonymous authors pass on responsibility for authoring the law to subsequent anonymous authors.²¹ To ensure integrity, each author must write in a manner consistent with the authors who came before.²² The arc of the development of the law must therefore be compatible with a narrative storytelling structure in order for it to be fully intelligible by those it proposes to govern.

From this theoretical standpoint, the subsequent step to law as narrative is clear.²³ Where I depart from Dworkin, however, is in his articulation of hard cases and the fundamental nature of legal inquiry that is revealed in those situations. In Dworkin's worldview, hard cases arise when judges have no obvious precedent or statute to guide their determinations.²⁴ Here, Dworkin asserts that judges have the duty to discover what the rights of the parties are, which he distinguishes from "invent[ing] new rights retrospectively."²⁵ Importantly, Dworkin acknowledges that reasonable lawyers and judges may sometimes disagree about legal rights, thus signalling the intrinsic complexity in some legal inquiries.²⁶

Until this point, I agree with Dworkin. His perspective underscores the difficulty inherent in transposing law to a coded, digital format: the right answer might simply not be written down anywhere—yet. However, despite this nod to complexity and ambiguity, Dworkin's framework nonetheless seems to suggest the availability of a singular legal answer.²⁷ A key component of Dworkin's theoretical framework is administration by the superjudge Hercules, a figure capable of making consistently correct decisions. Hercules has unlimited time and unlimited resources, differentiating him from actual judges, who suffer from real-world constraints in adjudication efforts. Hercules is pivotal to Dworkin's framework: he fully understands the purpose of the law and all preceding cases. Dworkin thereby demonstrates that there is only one way to decide each case, *i.e.*, with integrity. I take issue with Dworkin's fundamental belief that resolving legal problems with integrity necessarily leads to one right answer. A particular difficulty with this approach is that Hercules resigns himself to considering only past cases to generate his conclusions.

¹⁹ Setting out his vision for a theory of law as integrity, Dworkin pre-emptively responds to critics: "Law as integrity deplores the mechanism of the older 'law is law' view as well as the cynicism of the newer 'realism.' It sees both views as rooted in the same false dichotomy of finding and inventing law." Ronald Dworkin, *Law's Empire* (Cambridge, MA: Harvard University Press, 1986) at 228 [Dworkin, *Empire*].

²⁰ Dworkin, *Empire*, *supra* note 19 at 225-228.

²¹ *Ibid* at 228.

²² *Ibid*.

²³ In introducing the topic, Dworkin even goes so far as to say, "We can usefully compare the judge deciding what the law is on some issue not only with the citizens of courtesy deciding what that tradition requires, but with the literary critic teasing out the various dimensions of value in a complex play or poem." *Ibid*.

²⁴ Ronald Dworkin, *Taking Rights Seriously* (Cambridge, MA: Harvard University Press, 1977) at 83. Specifically, Dworkin defines hard cases as those in which "... no settled rule dictates a decision either way ..." [Dworkin, *TRS*].

²⁵ *Ibid* at 81.

²⁶ *Ibid*.

²⁷ Dworkin, for what it's worth, thinks that there are right answers, but acknowledges his thinking does not eliminate the presence of other possible, though less good, answers. He writes: "[t]his chapter describes the questions that judges and lawyers must put to themselves, but it does not guarantee that they will all give these questions the same answer." *Ibid*.

This approach puts undue pressure on legal metaphors to operate as blueprints for all possible legal futures. Such pressure is especially limiting, for example, as the Canadian judicial system grapples with integrating Indigenous legal orders to better represent the reality of Indigenous litigants. Even an infinitely wise Hercules, able to put himself in the place of various litigants, might sometimes be called upon to expand his conception of justice.

An expansive conception of justice demands a broader view of the Rule of Law as one encompassing a narrative arc from past to future, threaded through the legal realities of the present. Hercules, imbued with ultimate wisdom and limitless time, might benefit from a process of reflective equilibrium in his decision-making, engaging in an ongoing conversation with culture and context. In my view, cues can be taken from traditions of critical scholarship: critical feminist, Indigenous, and race scholars, who imbue the legal world with normative claims, complexity, and lived experiences. Their perspectives, in turn, can impact the world-building normative function of law. To bridge any remaining gap, I harness the conception of law as narrative described by Robert Cover: law's central tension operates in the space between law as it is and law as it ought to be.²⁸ Any vision of the law must be articulated through language and be anchored in a world of social forces and constraints.²⁹ The law is more than just one thing: it is also all of the things that it isn't. Delegation of legal decision-making to AI resolves a tension between the law as it has been stated and the law as it ought to be: a tension that needs to be persistently engaged by human endeavour for law to retain its governing power. Law cannot operate without this tension. In fact, if law were to be fully automated, law would be alienated from the human collective and condition whose movements give its existence meaning. There is vulnerability in the law to be parsed and perfected; assuming there is no vulnerability, there is no way to create meaning through law.

In related work, Dutch legal philosopher Mireille Hildebrandt offers pragmatic accounts of the Rule of Law that appeared to fit within both Cover's legal *nomos* and a technology-driven analysis.³⁰ Her explanation of the Rule of Law hinges on a demand that citizens be given an opportunity to contest the state's claim to be acting in their best interests.³¹ Using this lens reinforces the Rule of Law's prospective element: the contestability at the heart of the Rule of Law coincides with its procedural core, meaning debates and disagreements can be used to generate future decisions.³² This characterization of contestability reminds me of James Boyd White's description of the case method as a form of legal imagination: vital training for new lawyers on the arc of law's possibility.³³ Yet, this procedural aspect of the Rule of Law is not the only one that demands attention. Rule of Law's substantive aspect offers additional texture, as well as additional opportunities for regulation by machine to disrupt law's authority. Furthermore, these concepts are rooted in a liberal humanist worldview that presupposes the individual as a core consideration

²⁸ Cover, "Nomos," *supra* note 9.

²⁹ Julen Etxabe, "The Legal Universe After Robert Cover" (2010) 115 *Law and Humanities*, online: *Academia* <www.academia.edu/368582/The_Legal_Universe_After_Robert_Cover> [perma.cc/G9CS-6DS9].

³⁰ Hildebrandt defines legal norms in constitutional democracies usually consist of three key components: [1] they are democratically established by the governing legislative body; [2] it is possible for them to be violated; and [3] violations are evaluated (and penalized, if appropriate) by a court of law. Mireille Hildebrandt, *Smart Technologies and the End(s) of Law: Novel Entanglements of Law and Technology* (Cheltenham: Edward Elgar Publishing, 2015) at 10 [Hildebrandt, *End(s) of Law*].

³¹ *Ibid* at 85.

³² Mireille Hildebrandt, "Algorithmic Regulation & the Rule of Law" (2018) *Phil Trans R Soc A* 376: 20170355, online: *The Royal Society* <[dx.doi.org/10.1098/rsta.2017.0355](https://doi.org/10.1098/rsta.2017.0355)> [perma.cc/W6F9-EZAX] [Hildebrandt, "Algorithmic Regulation"] at 8.

³³ James Boyd White, "The Study of Law as an Intellectual Activity: A Talk to Entering Students" in White, *Heracles' Bow*, *supra* note 15 at 56.

for all theoretical constructs. By problematizing these assumptions, I make space for a more relational worldview, which coincides with the lessons from critical scholars.

c. What does this thesis demonstrate?

The leap from Dworkin's infinitely able superjudge Hercules to infinitely able supercomputer AI is a reasonably short one in some minds. Having identified what I view as both representative and more workable versions of a theoretical framework, I offer my preliminary conclusions about how AI exacerbates certain Rule of Law problems and how AI, done wrongly, might exploit certain weaknesses in the overall fabric of law. AI technology has developed in a way that underscores Rule of Law vulnerabilities that might not have otherwise been highlighted. AI thereby operates as a lens, revealing both problems and opportunities to repair the Rule of Law. There may well be situations in which supervised AI could resolve long-standing issues for the law, but these situations should not be green lit until the associated normativities are roundly interrogated.

Through the lens of AI, I analyse questions about legal decision-making and interpretation that have long preoccupied doctrinal jurisprudence scholars. By combining the insights of cyberlaw, STS, and law as narrative, I offer a unique framework for contextualizing the Rule of Law. Richer conceptions of sources for law, and corresponding storytelling approaches from various critical fields, provide opportunities for fine-tuning the interpretative exercises of law that introduce relational logics to legal understanding.

3. A Pre-Emptive Response to Critics

One can imagine a series of questions in response to my selected treatment of artificial intelligence, machine learning, and the law. Why the Rule of Law? Why approach the Rule of Law through the concepts of law put forward by legal philosophers? Why focus on the Rule of Law rather than just the concept of law? How do AI and ML implicate the Rule of Law's *machine*-ations?³⁴

Approaching these questions thematically, I here offer a pre-emptive response to critics.

Legal futurist scholars romanticize the potential of a legal system free from human involvement. Like much of technology law, themes seeded in law and economics undergird this objective.³⁵ Efficiency is sought as the primary goal, without attention paid to other important normative concerns. Automated justice is proposed as being more efficient, ameliorating longstanding complaints about backlogged courts and overloaded judges. Here, I am thinking specifically of legal futurist accounts that suggest we will be able to wholly delegate the business of both law and the courts to machines. I am thinking of legal futurist scholars who write about the future of law without acknowledging law's constitutive role in defining cultural and social norms. Scholars such Benjamin Alarie have written extensively on what he terms the "legal singularity:" the idea that the digitization of law will promulgate a "completion" of the legal data set, allowing law to propagate autonomously.³⁶ Anthony Casey and Anthony Niblett dream of the forthcoming death of rules

³⁴ Yes, I made up the world *machine*-ations. Ian Kerr's mother used to chide him for making up at least one word per academic paper. As it served him well, I've followed suit. ;)

³⁵ Lawrence Lessig took inspiration from the Chicago school's approaches to law & economics to conceptualize his much-lauded concept "code is law." See further discussion in Chapter 2.

³⁶ Benjamin Alarie, "The Path of the Law: Toward Legal Singularity" (2016) 66:4 UTLJ 443 at 445 [Alarie, "Legal Singularity"].

and standards, replacing generalized law with personalized microdirectives.³⁷ Daniel Goldsworthy claims machine learning technologies will generate superintelligence in the manner contemplated by Ray Kurzweil and be applied to legal decision-making to realize Dworkin's supposedly impossible dream of right answers.³⁸ Each of these legal futurist accounts idealizes a far-off future when law's operation is finally free of human intervention, while simultaneously asserting that this future is right around the corner. Even though the necessary technologies remain undeveloped—and perhaps even impossible—the literature still presents automated justice as a serious, imminent, near-future outcome. Law is presented as a problem to be solved, and technology as the silver bullet capable of solving it. This approach, often referred to as technological solutionism,³⁹ is further elaborated upon in Chapter 2. The efficiency and consistency that will be made possible through law's digitization is presented as laudable and, consequently, as synonymous with justice.

The legal futurists write enthusiastically about the capabilities of technologies that are not yet here, and may never be here, to justify their plans for law's eventual administration. These accounts are presented both for the legal system itself (*i.e.*, self-driving laws in the sense of automated court decisions or administrative tribunals) and participation of lawyers within the legal system (*i.e.*, the use of advanced software to facilitate research, predict case outcomes, and automate rote tasks). While the latter set of examples has become plausible in limited circumstances, any true delegation of judging activities to algorithms remains beyond current technological capability. The example of AlphaGo is often given to emphasize how far computers have come in their capacity to mimic human learning.⁴⁰ A computer's ability to play and succeed at Go, widely acknowledged to be one of the world's most conceptually difficult games due to its cognitive requirements, is certainly impressive.⁴¹ However, it does not neatly transpose onto the ability to make complex legal determinations. As further explained in Chapter 1, computers are excellent at activities involving math, or that are neatly reducible to math. For truly sophisticated exercises of judgment, most systems still require a human “in the loop.”⁴² Thus, in the near future, any delegation of legal decision-making from persons to machines will simply put the human involvement at a different place in the chain. While this reconfigured process may masquerade as digital decision-making, it does not guarantee an improvement over current practices, especially in light of the important roles that law plays in establishing and maintaining normative force.

The belief that law can universally offer right answers is, at best, a fringe view within the jurisprudence literature. Yet, this assumption seems to animate the chief contentions of legal futurists. The act of learning law in law schools offers a powerful metaphor here: some students simply want to learn to know what is going to be on the final exam or what rule to extract from a case, while their professors want them to think harder about hard cases. Empirical projects for law

³⁷ Anthony J Casey & Anthony Niblett, “A Framework for the New Personalization of Law” (2019) 86 U Chi L Rev 333 at 338.

³⁸ Daniel Goldsworthy, “Dworkin's dream: Towards a singularity of law” (2019) 44 *Alternative Law Journal* 286.

³⁹ Evgeny Morozov, *To Save Everything, Click Here: The Folly of Technological Solutionism* (New York, NY: PublicAffairs, 2013).

⁴⁰ Brian Christian, *The Alignment Problem: Machine Learning and Human Values* (New York, NY: WW Norton & Company, 2020).

⁴¹ In 1997, Dr. Piet Hut of the Institute for Advanced Study, claimed that it would be at least 100 years before a computer could beat a human at Go—and maybe longer. In 2016, the AI system Alpha Go defeated Go grandmaster Lee Sodol. See Benjamin Alarie, Anthony Niblett & Albert H Yoon, “Law in the Future” (2016) 66:4 *UTLJ* 423 [Alarie et al, “Future”].

⁴² Ian Kerr & Katie Szilagyi, “Asleep at the Switch: How Lethal Autonomous Robots Become a Force Multiplier of Military Necessity” in Ryan Calo, A Michael Fromkin & Ian Kerr, eds, *Robot Law* (Cheltenham, UK: Edward Elgar Publishing, 2016) 333 at 339.

are preoccupied with prediction in the sense of asking what the court will do: whether leave to appeal will be granted; whether a particular outcome will be reached; whether the “right answer” will be achieved. Alarie anticipates that computing power will eventually be able to handle law’s normative dimensions, writing hopefully that his legal singularity will “bridg[e] the gap between normative values and politics and lived factual consequences.”⁴³ Niblett and Casey acknowledge their proposal might have normative concerns, but explicitly state they are leaving such work for other scholars.⁴⁴ Goldsworthy appreciates the normative aspects of philosophical investigation, but his enthusiasm presents this task as a small piece of the puzzle in a legal system that has used technological solutions to collapse all other considerations.⁴⁵

In my view, talking about automating law’s processes through machine learning algorithms requires a careful look at rules, and their fellow traveler, principles. As elaborated in Chapter 1, machine learning consists of rules: computers are told which rules to follow and how to organize data based on adherence to those rules. And, when thinking about rules and principles in relation to the law and how it operates, one is paramount: the Rule of Law. While many can disagree about law’s contents, application, or extent, it is easy to commit to the Rule of Law as a mythologized ideal. In this sense, the Rule of Law operates above the changeable whims of lawmakers to secure the legitimacy and authority of the whole system. In my view, however, legal automation without careful calibration of normative interests risks a conversion of the Rule *of* Law into a Rule *by* Law, where law’s interpretative functions collapse into mere administration. Obscuring law’s machinations through inscrutable, data-driven methods that are difficult to parse threatens the contestability considered critical to legal process. In Chapter 3, I develop these ideas in conversation with the precepts of administrative law, focusing on procedural fairness.

A few other legal scholars have tackled questions about legal automation, artificial intelligence, and machine learning through a Rule of Law lens. Of course, the fundamentally contested nature of the Rule of Law means different scholars characterize the problem differently. Roger Brownsword approaches the question in light of a technological imperative to rework the Rule of Law, which he characterizes as “implicitly assum[ing] a rule-guided form of order” and notes its subtle regulatory impacts to constrain possible behaviours.⁴⁶ Advocating for “a rule of persons, not machines,” Frank Pasquale emphasizes the importance of human judgment in exercising discretion within a properly functioning administrative state.⁴⁷ In other work, Pasquale is joined by computer scientist Glyn Cashwell to catalogue challenges with ML approaches that are “devoid of

⁴³ Alarie, “Legal Singularity,” *supra* note 36 at 453.

⁴⁴ Casey & Niblett, *supra* note 37 at 1405.

⁴⁵ Goldsworthy’s analysis carves out a small piece of philosophical investigation to remain the purview of lawyers and judges. To me, however, what he presents as a small carve out is essentially the entirety of law’s empire. He writes: “It may well become the role of lawyers, judges and academics to deconstruct and reconstruct theories of law and justice where artificial intelligences of the type described allow for more than one cogent, logical outcome. In this way, synthesising incomprehensible volumes of judgments and theoretical analysis no longer remains an impossibility, but rather a technical problem to be solved. And once rendered possible, it is then left to argue why one coherent and logically sound outcome should be preferred over another. This is a normative consideration, a philosophical question of value and ethic. And that, at least for now, remains the purview of human beings.” Goldsworthy, *supra* note 38 at 289.

⁴⁶ Roger Brownsword, “Artificial Intelligence and Legal Singularity: The Thin End of the Wedge, the Thick End of the Wedge, and the Rule of Law” in Simon Deakin & Christopher Markou, eds, *Is Law Computable?: Critical perspectives on law and artificial intelligence* (Oxford, UK: Hart Publishing, 2020) at 13.

⁴⁷ Frank Pasquale, “A Rule of Persons, Not Machines: The Limits of Legal Automation” (2019) 87 *Geo Wash L Rev* 1 [Pasquale, “Rule of Persons”].

apprehension of meaning in the legal context.”⁴⁸ As introduced above, Hildebrandt offers robust Rule of Law analysis in her text *Smart Technologies and the End(s) of Law*,⁴⁹ but her approach draws heavily on a civil law tradition of rule-making that might accuse anything happening outside law’s explicit framework as being detrimental to democracy.⁵⁰ A continental European scholar, Hildebrandt’s work is particularly apt in light of recent movements around the *General Data Protection Regulation (GDPR)* and corresponding calls to ensure that citizens have a right to an explanation when algorithms make determinations affecting their lives. Similarly, the UK scholar Paul Gowder addresses the Rule of Law as requiring regularity, publicity, and generality: key criteria that “constitute a morally important kind of legal equality” when in operation.⁵¹ In a wide-ranging article on automating government decision processes, Australian scholars Monika Zalnieriute, Lyria Bennett Moses, and George Williams characterize the Rule of Law as a “political work in progress,” focusing on its oft-cited ideals.⁵² Through this lens, the focus is on the Rule of Law’s procedural guarantees, rather than any substantive philosophical ideologies.

What connects these scholars’ accounts is an awareness of the Rule of Law’s centrality in the administration of justice and a deep concern that justice might not be able to achieve its lofty objectives if its role becomes uncalibrated. Pasquale and Cashwell remind academics tackling such issues that merely predicting the future is not an “unalloyed good,” especially since social science research on reflexivity confirms how predictive analysis “exacerbates, rather than resolves, the problems it reveals.”⁵³ Jennifer Cobbe tackles the question differently in an article critiquing the legal futurists in light of their proposals’ implications for law’s reflexivity requirement.⁵⁴ Her explorations highlight law’s role as a reflexive, societal institution that has normative consequences, and that has “entrenched the power of capital, strengthened the position of the wealthy, reinforced inequalities, and protected established interests from outside challenges.”⁵⁵

Relatedly, legal futurists present automated justice as the mechanism by which we finally eliminate the judicial bias preventing justice from truly being achieved, at last remedying the caricatural legal realist’s account of law determined by what a judge had for breakfast rather than the law and facts before the court. This topic is a major theme in Chapter 3. Through this lens, human fallibility is emphasized and machine competence appears as the necessary means to solve it. Others have

⁴⁸ Frank Pasquale & Glyn Cashwell “Prediction, Persuasion, and the Jurisprudence of Behaviourism” (2018) 68 UTLJ (Supplement 1) 63 at 74.

⁴⁹ Hildebrandt, *End(s) of Law*, *supra* note 30 at 85.

⁵⁰ Joshua Fairfield offers a useful explanation of the distinction between common and civil law philosophies that mirrors my own observations. Rulemaking in common law systems is chaotic and occurs in a series of dyads, wherein the rules arise from judicial consideration of pairs of opposing interests. But the common lawyer’s interest in asking what happens next, or what narrative might flow from it, might just make up for the chaos. By contrast, a civil lawyer’s commitment to the Rule of Law is anchored in a belief that anything happening outside law’s framework is likely corrosive or problematic for democracy. This type of civil lawyer reasoning seems to be animating Hildebrandt’s central contentions. See Joshua Fairfield, *Runaway Technology* (Cambridge, MA: Cambridge University Press, 2021) at 86 [Fairfield, *Runaway*].

⁵¹ Paul Gowder, “Transformative Legal Technology and The Rule of Law” 68 UTLJ (Supplement 1) 82 at 86.

⁵² Saying that they don’t wish to provide yet another account of the Rule of Law, the authors condense the Rule of Law’s various aims into several key ambitions: transparency and accountability; predictability and consistency; and equality before the law. Monika Zalnieriute, Lyria Bennett Moses & George Williams, “The Rule of Law and Automation of Government Decision-Making” (2019) 82:3 *Modern Law Review* 1 at 4-5.

⁵³ Pasquale & Cashwell, *supra* note 48 at 81.

⁵⁴ Jennifer Cobbe, “Legal Singularity and the Reflexivity of Law” in Deakin & Markou, *supra* note 46.

⁵⁵ *Ibid* at 22.

written extensively on whether machines can legitimately make judicial determinations,⁵⁶ whether human reasoning has some unique characteristics that evade digitization,⁵⁷ and whether we should just replace all the judges in the court system with some highly advanced robots.⁵⁸ I do not wish to retrace this already well-trod ground. Instead, my inquiry here is slightly different: what would the normative consequences be for the law if the legal futurists achieve their goal? What is the law supposed to do? Would it still achieve its aims if its functionality were automated?

In my view, this analysis must be bidirectional. I am concerned both with what law's operations indicate for the future of legal technology and what our responses to those proposed technologies might tell us about the nature of law. To explore these ideas, I use three different AI/ML powered technologies, each the subject of a case study in Part III. Here, I take inspiration from Evgeny Morozov's work, *To Save Everything Click Here: The Folly of Technological Solutionism*, which approaches the technological solutionism themes sought by Silicon Valley as impoverishing our self-understanding as political objects. By quantifying everything, he argues, the narrative imagination drops out: an undue reliance on data science prevents us from understanding our role in the system, instead discretizing answers to zero-in on singular data points.⁵⁹ The intersections of law and technological design invite deep considerations of what law is supposed to do.

Part III's case studies each focus on the Rule of Law problems generated by the use of the respective AI/ML technologies. I do this because the act of using technology to govern law's systems promotes the efficacy by which it rules. I observe that the Rule of Law is harnessed by people spanning various political ideologies and, while possessing a contested character, is usually presented as being legitimate and synonymous with the good. Frank Pasquale notes that "legal futurists tend to present the reduction of legal obligations to computer code as a positive evolutionary step toward the realization of the Rule of Law."⁶⁰ By converting the operation of law to a machine-driven process, the hope is that law will be cured of its deficiencies, including bias, discrimination, arbitrariness, or other injustice.

One of the Rule of Law's most oft-cited objectives is to establish a rule of laws, not men. This idea is put forward in the context of limiting authoritarian rule: to ensure that states are not subject to the fickle wishes of their leaders, the Rule of Law offers a way of governing that can be consistent, non-arbitrary, and reliable. As Frank Pasquale points out, "one literal way of achieving the oft-quoted ideal 'a rule of law, not of men' is to dispense altogether with persons implementing or interpreting law."⁶¹ Far from curing deficiencies, such delegation merely reframes the playing field. Instead of asking whether the people making judicial decisions are pursuing justice, the question becomes whether the people setting up the machines are capable of generating a just system.

⁵⁶ Cass Sunstein, "Of Artificial Intelligence and Legal Reasoning" (2001) 8 University of Chicago Law School Roundtable 29 at 34.

⁵⁷ Dan M Kahan, "Laws of cognition and the cognition of law" (2015) 135 Cognition 56 at 59 (noting that the study of cognition confirms that professional judgment is special).

⁵⁸ Ian Kerr & Carissima Mathen, "Chief Justice John Roberts is a Robot" (2014) We Robot Conference paper, online (pdf): [SSRN <papers.ssrn.com/sol3/papers.cfm?abstract_id=3395885>](https://papers.ssrn.com/sol3/papers.cfm?abstract_id=3395885) [perma.cc/28YG-UFMY].

⁵⁹ Morozov, *supra* note 39.

⁶⁰ Pasquale, "Rule of Persons," *supra* note 47 at 4.

⁶¹ *Ibid.*

Related to this inquiry are additional questions: is the technology likely to get us there? What are the normative consequences of generating a growing collection of scholarly papers dependent on idealized machine learning systems that do not yet exist? What wisdom is wrought from assuming that all legal outcomes can be generated by parsing the library of existing decisions? If systems are built to produce singular correct legal answers, what becomes of the grand tradition of dissent? What happens to the ideas whose time has not yet come? What happens when the present context is suddenly very different from the context in which the existing decisions were handed down? What kinds of stories will the robots write for us? These sorts of questions animate my inquiries in this thesis.

I now turn to answering these questions.

Part I: Coding the Law

1

Introduction to Artificial Intelligence & Machine Learning

*An essentially mechanical world would be
an essentially meaningless world.*
~ Friedrich Nietzsche⁶²

Once, at the outset of this PhD, I moved into a rental house in Ottawa with a longtime friend. While the old red brick house in Ottawa's Byward Market had character in spades, my housemate and I had some challenges with our landlady. In a story familiar to renters everywhere, she was somewhat disengaged, prone to seek shortcuts wherever possible. My former housemate is both meticulous and accustomed to the banality of bureaucracy; so, shortly after we moved into the new place, she proactively located the necessary form to change the electricity account to our names. The form had to be completed by the rental authority, meaning it fell within the landlady's purview. Instead of completing the form, however, the landlady simply emailed the electricity company directly, with us in the cc: line, asking them to switch the account. To this, we all received an automated message that the electricity company would get back to us in 10-14 working days.

Four weeks later, we received at last our reply from the electricity company: attached is a form, could the landlady please fill it out and promptly remit the form to the agency? This form was, of course, the very same form my housemate had sent the landlady weeks before. Frustrated, my housemate forwarded the company's response to me, with the added sardonic commentary: "Amazing. This is the form I sent to her in the first place."

I received the email on my mobile device, using Google's Gmail application. The application provided me with three proposed responses to the text of the email: [1] "Looks great!"; [2] "Nice!"; and [3] "Looks good to me." For all their so-called wisdom, Google's robots had not yet sorted out my housemate's sarcasm. More troublingly, all the proposed responses were positive. Did this circumscribed range of possibilities perhaps flow from the fact that the word "amazing" was used in the original message? Did the predictive algorithm not anticipate any possibility of disagreement with the pronouncement of amazingness? Was this degree of positivity indicative of some larger corporate effort to placate the communications of the masses, or was it mere oversight?

Yes, this technology is still new, and perhaps it will learn to better predict responses over time. Perhaps it will learn to identify sarcasm, the typical tenor of conversations between my former housemate and me, or my usual word choices for such interactions. Such improvements, in my view, would do nothing to ameliorate the problem. Instead, they might well make it worse: if the algorithm improves its natural language capabilities and becomes better able to draft an appropriate response to the messages I receive, I will become less inclined to mock that response or use it as a

⁶² Morozov, *supra* note 39 at 244.

small allegory to introduce my doctoral dissertation. Eventually, I would be more comfortable assuming the language, the phrasing, and the voice that the corporation has assigned me.⁶³

And, if the technology further advances, perhaps my response will become altogether unnecessary. Perhaps the algorithm will prescribe my response irrespective of my input and autonomously respond to the email. Perhaps my former housemate's robot will follow a similar script. Perhaps our entire conversation can take place without us. We can read it when it's over, just to see what we have said.

In such a world, we must ask: what kinds of stories will the robots write for us?

If robots begin to author our stories, we must pay careful attention to the bids on meaning claimed through those stories. After all, meaning is difficult to generate through an artificial intermediary that is only capable of generating the specific outputs it has been taught. As artificial intelligence inculcates different areas of our culture and society, it has begun to tiptoe into the legal realm, usually under the guise of improved efficiency through advanced pattern-matching or statistical techniques. As this thesis will explore, some optimistic scholars predict legal futures powered by a "legal singularity," as envisioned by Benjamin Alarie, where law's empire would be overseen by sophisticated algorithmic processes.⁶⁴ Alarie's proposed regulation by machine would aggregate all existing legal knowledge, generating the correct response for each new situation based on an algorithmic output of the necessary data.⁶⁵ This approach, I argue, threatens to impoverish the narrative aspects of law, which have long served a cooperative world-building function: a collective story. If law were to someday be fully administered by machine, these machinations could threaten law's crucial interpretative function. Especially as artificial intelligence capabilities grow, as in the sphere of sentencing decisions for criminal law, there are substantial repercussions for the interpretative exercises that law traditionally connotes. Joshua Fairfield agrees, noting law's role as a cooperative fiction, or a "piece of social technology."⁶⁶ If AI develops its own bids on meaning, through already common technologies such as natural language processing, the potential transformations of human language might lead to corresponding transformations of law.

Ian Kerr sounded early alarm bells on the legal consequences of automated systems with his work on the nature of predictive analytics and how they impact legal frameworks.⁶⁷ He worried that incorporating artificial intelligence in law, termed law's "computational turn," might subvert law's traditional evidence and interpretation-based approach to predicting future legal outcomes, replacing it instead with wholesale pre-emption.⁶⁸ Harnessing Oliver Wendell Holmes' phrase, "the path of law," Kerr recalls that Holmes' predictive approach was intended to illuminate the

⁶³ Rich Firth-Godbehere, "Silicon Valley thinks everyone feels the same six emotions" (17 September 2018), online: *Quartz* <qz.com/1392130/silicon-valley-thinks-everyone-feels-the-same-six-emotions/> [perma.cc/8D9S-CGPE] (arguing that most contemporary emotion-detecting technology is based on the findings of Paul Ekman, whose research was largely inspired by his own experiences as a grieving teenage boy. The six emotions we purportedly share are: happiness, anger, sadness, disgust, surprise, and fear).

⁶⁴ Alarie, "Legal Singularity," *supra* note 36 at 445.

⁶⁵ *Ibid.*

⁶⁶ Fairfield, *Runaway*, *supra* note 50 at 49.

⁶⁷ Ian Kerr, "Prediction, pre-emption, presumption: The path of law after the computational turn" in Mireille Hildebrandt & Katja de Vries, eds, *Privacy and Due Process After the Computational Turn* (London, UK: Routledge, 2013) 91 [Kerr, "Prediction, pre-emption"].

⁶⁸ *Ibid* at 93.

nature of law itself by shifting focus to the perspective of those who might be most impacted by it.⁶⁹ By contrast, new pre-emptive approaches would do precisely the opposite: obfuscating the citizen's ability to perceive law's machinations. As he explained:

I argue that today's predictive technologies threaten due process by enabling a dangerous new philosophy of pre-emption. My concern is that the *perception* of increased efficiency and reliability in the use of predictive technologies might be seen as the justification for a fundamental jurisprudential shift from our current *ex post facto* systems of penalties and punishments to *ex ante* preventative measures that are increasingly being adopted across various sectors of society. This shift could fundamentally alter the path of law, significantly undermining core presumptions built into the fabric of today's retributive and restorative models of social justice, many of which would be pre-empted by tomorrow's actuarial justice.⁷⁰

Kerr articulated his concern in the context of due process, jokingly called the “prime directive” of the legal universe.⁷¹ But he hints at larger problems: at the possibility that predictive algorithms might “rewrite the code of the legal universe by reprogramming some of its core normative presumptions.”⁷² He expressed similar concerns by mapping the ongoing shift from automatic machines to autonomous machines: as emergent machine learning processes are designed to “supersede their initial programming,” they are actually designed to do the unexpected.⁷³ “With so-called autonomous machines, the possibility of generating unintended or unanticipated consequences is not a malfunction. It is a feature, not a bug.”⁷⁴ Invoking Stanley Milgram's idea of an “agentic shift,” Kerr explains the moral strain that accompanies delegated—or, perhaps, abdicated—responsibility for decision-making. Once this type of agentic shift takes place, a moral decision-maker no longer feels responsible or capable of making key decisions.⁷⁵ Building on Kerr's analysis, I argue that delegating decision-making to artificial cognizers threatens law on a doctrinal level. If law is to be mediated by artificial intelligence and machine learning, the potential consequences for the legal universe are enormous. In my view, such shifts in human language, morality, and interpretation threaten a key guiding principle: the Rule of Law.

Before delving too deeply into legal doctrine and jurisprudence, however, foundations must be laid. This chapter provides an overview of artificial intelligence (AI) and machine learning (ML), offering an introduction to the key technologies, vocabulary, and ongoing debates in this academic sphere. In the following sections, I survey the proposals for AI-enabled legal applications, provide common sense explanations on how the technology works, and begin to map its intersections with

⁶⁹ Kerr, “Prediction, pre-emption,” *supra* note 67 at 93

⁷⁰ *Ibid* at 92.

⁷¹ The “prime directive” refers to the central tenet of the Star Trek universe, which prohibited the members of the Star Fleet from interfering in the normal development of any society in the universe. Franz Joseph, *Star Trek Star Fleet Technical Manual* (New York, NY: Ballantine Books, 1975) cited in *ibid* at 107.

⁷² *Ibid*.

⁷³ This piece on medical decision-making was written as a guest blog post on Jack Balkin's website. It begins with an apt epigram from Ludwig Wittgenstein: “Why, would it be unthinkable that I should stay in the saddle however much the facts bucked?” Ian Kerr, “AIs as Substitute Decision Makers” (1 November 2018), online: *Balkinization* <balkin.blogspot.com/2018/11/ais-as-substitute-decision-makers.html> [perma.cc/5V85-QXU9] [Kerr, “Balkinization”].

⁷⁴ *Ibid*.

⁷⁵ *Ibid*.

exercises of both corporate and legal authority. I offer an entry point into the technological affordances of AI and ML that power the further chapters of this thesis.

1. How Artificial Intelligence Works

AI refers to the ability of computers developing the ability to “think” on their own. The culmination of decades of advances in mathematics, statistics, computer science, neurobiology, psychology, information science, and other related fields, AI is a suitcase word, in that it contains a host of meanings that need unpacking.⁷⁶ AI is typically used to describe technologies capable of executing complicated tasks in environments of uncertainty. Such tasks might include voice recognition, image identification, logic problems, pattern matching, and other types of optimization problems.⁷⁷ While AI often brings to mind fantastical science fiction films,⁷⁸ or sensational images of anthropomorphized humanoid robots waging war on humanity’s behalf,⁷⁹ most of today’s AI applications tend to be more routine.⁸⁰ Data mining, knowledge database discovery, predictive analytics: these ordinary office functions are increasingly powered by machine learning.⁸¹ Definitions in this space are important: when discussing AI ideas, one should take care to use specific terminology and avoid the “android fallacy.”⁸² Ultimately, the vision of machine intelligence promised by science fiction remains unrealized. We have not yet arrived at the singularity⁸³ and humanoid bots equipped to pass Turing tests⁸⁴ are not yet ready to march off the assembly line. We have yet to build artificial general intelligence (AGI). AGI would mirror human

⁷⁶ This term was coined by one of the so-called “fathers” of artificial intelligence, Marvin Minsky, in his book *The Emotion Machine*. Recently, Minsky was implicated in the Jeffrey Epstein papers as participating in his exploitative sexual trafficking activities. See Kate Crawford, *Atlas of AI: Power, Politics, and the Planetary Costs of Artificial Intelligence* (New Haven, CT: Yale University Press, 2021) at 245 [Crawford, *Atlas*].

⁷⁷ Christian, *supra* note 40.

⁷⁸ *Terminator, 2001: A Space Odyssey, Blade Runner, Her, The Matrix*. The choices appear endless. If you’re Nick Bostrom, perhaps it’s the Paperclip Armageddon. See Nick Bostrom, “Ethical Issues in Advanced Artificial Intelligence” (2003), online: nickbostrom.com/ethics/ai.html [perma.cc/FT77-BYA4]. In late 2017, an intrepid programmer picked up on this paperclip folk tale to create an end-of-the-world simulation you can play online. See Adam Rogers, “The Way the World Ends: Not With a Bang But With a Paperclip” (21 October 2017), online: *WIRED* <www.wired.com/story/the-way-the-world-ends-not-with-a-bang-but-a-paperclip/> [perma.cc/3MD2-4XF4].

⁷⁹ Kerr & Szilagyi, *supra* note 42 at 333.

⁸⁰ In a slightly sardonic report on AI’s current prowess, Miles Brundage writes: “Notable recent achievements of AI include surpassing human performance in the game of Go and achieving superhuman performance on a range of image processing tasks. AI technologies are widely distributed in modern life, with commonly used applications including search engines, voice recognition on phones, and online machine translation.” See Miles Brundage, “Scaling Up Humanity: The Case for Conditional Optimism about Artificial Intelligence” in Peter J Bentley, Miles Brundage, et al, “Should We Fear Artificial Intelligence?” at 13 (March 2018), online: *European Parliament* <[www.europarl.europa.eu/RegData/etudes/IDAN/2018/614547/EPRS_IDA\(2018\)614547_EN.pdf](http://www.europarl.europa.eu/RegData/etudes/IDAN/2018/614547/EPRS_IDA(2018)614547_EN.pdf)> [perma.cc/7CLG-SKUH].

⁸¹ For an industry-facing report on the state of the AI and ML in modern business practices, see McKinsey & Company, “The state of AI in 2020” (17 November 2020), online: *McKinsey* <www.mckinsey.com/business-functions/mckinsey-analytics/our-insights/global-survey-the-state-of-ai-in-2020> [perma.cc/9BX9-Y3PT].

⁸² Bill Smart and Neil Richards argue that when discussing AI, ML, and the implications of robots, we need to guard against the “android fallacy,” which they define as improperly conceptualizing of robots in a manner consistent with science fiction. They posit that an important element of discussing legal regulation for AI is sketching what robots can do, what they cannot do yet, and what they may never be able to do. Because “never” is quite a long time, they propose that “long after we are dead” is a workable standard. See Neil M Richards & William D Smart, “How Should the Law Think About Robots” in Calo et al, *Robot Law*, *supra* note 42 at 3.

⁸³ So envisioned by notable futurist Ray Kurzweil as the exponential leap in computing abilities when the exponential pace of machine ability outstrips human ability, fundamentally transforming the world and the role of humans within it. See Ray Kurzweil, *The Singularity is Near: When Humans Transcend Biology* (London, UK: Duckworth Overlook, 2005).

⁸⁴ Posited by computer pioneer Alan Turing, the Turing test is the baseline for assessing whether an artificial intelligence is capable of tricking a human that it too is human. For more discussion of Turing-test-enabled technologies, see Ian Kerr, “Bots, Babes and the Californication of Commerce” (2004) 1 *uOttawa L & Tech J* 285.

capacity for generalized skill development and drawing lateral connections between seemingly unrelated things.⁸⁵ Existing AI, more precisely known as artificial narrow intelligence (ANI), has a narrower skill set: it requires specific instructions and is coded to solve explicit problems.⁸⁶ Nonetheless, such pattern-based problem solvers are powerful computational tools that enable many of today's automated processes.⁸⁷ In subsequent chapters, I provide more thorough explanations of available AI systems: Chapter 6 explains sentencing software; Chapter 7 explains facial recognition technology; and Chapter 8 explains natural language processing. Each of these systems are enabled by similar types of ANI processes.

Chess aficionados might recall the development of IBM's Deep Blue, the chess supercomputer that defeated famed grandmaster Garry Kasparov in 1997.⁸⁸ When he narrowly edged the machine in 1996, Kasparov famously remarked that he "could feel—[he] could smell—a new kind of intelligence across the table."⁸⁹ A mere year later, small adjustments made Deep Blue's incredible processing power capable of evaluating hundreds of millions of potential future chess moves every second.⁹⁰ According to Deep Blue's project manager, what made the machine victorious was that although Kasparov was prepared to play against a computer, "we programmed it to play like a Grandmaster."⁹¹ Although this strategy proved victorious against Kasparov, imitating human decision-making ultimately stalled the development of such systems.⁹²

Similar initial advances were made in the development of AlphaGo, a computer program specifically designed to play the popular abstract strategy game, Go.⁹³ Like Deep Blue before it, AlphaGo trained to predict the moves that expert human players would make, through a database comprised of more than 30 million moves.⁹⁴ While AlphaGo successfully defeated significant human competitors, Brian Christian notes that although the computer had "transcended human play," it remained "an imitator at heart. It was not learning to play the *best* moves. It was learning to play the *human* moves."⁹⁵ For both Deep Blue and AlphaGo, the incredible popularities of the games they were training to play (chess and Go, respectively) created significant advantages for

⁸⁵ AGI is defined as AI that would have the ability to solve a variety of complex problems in a variety of different contexts. See Michael Guilhot et al., "Nudging Robots: Innovative Solutions to Regulate Artificial Intelligence" (2017) 20 Vand J Ent & Tech L 385 at 395. Note that AI is sometimes also divided into the categories of "strong" and "weak," wherein "strong" connotes AGI: computers would "think" at level that either meets or surpasses the cognitive capabilities of humans, with the ability to engage in abstract reasoning and thinking. In this dichotomy, "weak" AI is current, narrower AI applications. Ryan Calo, "Artificial Intelligence Policy: A Primer and Roadmap" (2017) 51 UC Davis L Rev 399.

⁸⁶ *Ibid.*

⁸⁷ Harry Surden, "Artificial Intelligence and Law: An Overview" (2019) 35 Georgia State University Law Review 1305 at 1309 [Surden, "AI & Law"]

⁸⁸ Christian, *supra* note 40 at 242.

⁸⁹ Kashyap Vyas, "Deep Blue vs Kasparov: The Historic Contest That Sparked the AI Revolution", (8 August 2019), online: *IE* <interestingengineering.com/deep-blue-vs-kasparov-the-historic-contest-that-sparked-the-ai-revolution> [perma.cc/4BC5-RT85].

⁹⁰ *Ibid.*

⁹¹ *Ibid.*

⁹² Christian, *supra* note 40 at 242.

⁹³ Go is especially popular in East Asia. Over 20 million of the world's 46 million Go players reside there, with the greatest concentrations in China, South Korea, and Japan. The International Go Federation, "Go Population Survey" (February 2016), online: *InterGo* <www.intergofed.org/wp-content/uploads/2016/06/2016_Go_population_report.pdf> [perma.cc/EK7A-EAJA].

⁹⁴ Christian, *supra* note 40 at 243.

⁹⁵ *Ibid.*

developing a database for the computer systems to learn from.⁹⁶ Their capacity to transcend human abilities were empowered by the large repositories of relevant human information that existed.

This distinction is subtle, but important: machine learning does not actually connote learning in traditional terms. In most instances, the ML method simply derives the best possible solution for a very narrow problem, constructed based on human understandings and drawn from the dataset that records those understandings. As Meredith Broussard explains: “when a machine ‘learns,’ it doesn’t mean that the machine has a brain made out of metal. It means the machine has become more accurate at performing a single, specific task according to a specific metric that a person has defined.”⁹⁷ A similar point is made by MIT Professor Luis Perez-Breva, who describes the process of image recognition by a computer as follows:

[I]f you’re trying to get a computer to recognize an image just feed it enough data and with the magic of math, statistics and neural nets that weight different connections more or less over time, you’ll get the results you would expect. But what you’re really doing is using the *human’s* understanding of what the image is to create a large data set that can then be mathematically matched against inputs to verify what the human understands.⁹⁸

Improvements in AI have been made possible in recent years due to leaps in ML, which have been enabled through newly rich databases of human behaviour derived from various tracking techniques. As I will discuss further, most people participate freely in these information-gathering activities, enabled by online platforms and social media, contributing personal data for quantification.⁹⁹ Despite the broad proliferation of AI and ML methods, straightforward explanations of what the technology entails can be difficult to find. The novelty of ML means its definition remains contested, with different leading sources proposing different options.¹⁰⁰ The collaborative nature of coding work often means that the latest solutions are usually posted on online hubs and forums frequented by developers, instead of being published in peer-reviewed academic journals.¹⁰¹ Open-source solutions to problems are frequently shared for other programmers to adapt to their own needs.¹⁰² While ML research is still sometimes presented at

⁹⁶ A few years later, a successor technology, AlphaGo Zero, outstripped AlphaGo’s abilities, even though it had never been shown any human moves at all. Rather than learning from the same large repository as its predecessor, it used a neural network to learn to imitate itself, removing humanity’s role as intermediary. *Ibid* at 244-246.

⁹⁷ Meredith Broussard, *Artificial Un-intelligence: How Computers Misunderstand the World* (Cambridge, MA: The MIT Press, 2019) at 92.

⁹⁸ Ron Schmelzer, “Is Machine Learning Really AI?” (21 November 2019), online: *FORBES* <www.forbes.com/sites/cognitiveworld/2019/11/21/is-machine-learning-really-ai/#69ab44a52621> [perma.cc/QM4N-SN74].

⁹⁹ McKinsey, *supra* note 81.

¹⁰⁰ Broussard, *supra* note 97 at 91-92. Broussard notes there is little consensus across texts, unlike for definitions of settled terms like “dog,” which tend to be consistent.

¹⁰¹ Examples of these include GitHub, for open-source code that has been shared with the community, and blogs, where developers often post problems they are solving and their proposed solutions. Some such blogs include the Amazon Web Services (AWS) Machine Learning Blog, the Berkeley Artificial Intelligence Research (BAIR), and the Cambridge Innovation Institute (CII). See AWS, “AWS Machine Learning Blog” (last accessed 30 July 2022), online: *Amazon* <aws.amazon.com/blogs/machine-learning/> [perma.cc/Y4WY-BXZM]; BAIR, online: *Berkeley* <bair.berkeley.edu/blog/> [perma.cc/E68N-WCQS]; CII, online: *CII* <www.cambridgeinnovationinstitute.com/> [perma.cc/2R2H-3LLF].

¹⁰² For example, GitHub’s page for the ML environment TensorFlow shows over 87,000 forks, meaning that the solution has been copied and used over 87,000 times. Users can also make their own solutions open source, sharing the results further, or can protect it, if they have adapted it to something proprietary. GitHub, “tensorflow/tensorflow”, online: <github.com/tensorflow/tensorflow> [perma.cc/77QY-T2JM].

major conferences or recorded in short papers,¹⁰³ the field's constant changes means that traditional knowledge sources, like journals or textbooks, have less of an authoritative footing in the field.

a. How Machine Learning Works

AI is often erroneously equated with ML, which is a computational technique that allows computer programs to “learn” from experience and gradually improve their performance at specific tasks over time.¹⁰⁴ This technique is operationalized through algorithms that execute advanced pattern-matching techniques.¹⁰⁵ The term algorithm, in its simplest form, refers to the conversion of an input to an output. If an algorithm is expressed in binary, this simply means that a switch is turned on or off: a bit is changed from a 0 to a 1, or from a 1 to a 0. Normally, this process means that data are the input for computational processes: the algorithm operates on the data and produces a corresponding output. What AIs, and particularly ML systems, do is turn this input/output mechanism around. Within a given training dataset, the input (the data) and the output (the desired result) are both already known. On this basis, the computational process derives its own program to convert one to the other. In essence, the computer drives its own program: the programmer—after setting up the problem-solving environment—takes a backseat.¹⁰⁶

ML algorithms can be divided into three main types: supervised learners, unsupervised learners, and reinforcement learners.¹⁰⁷ All three require some kind of training data, where all of the relevant variables are already established, and the models can be developed, iterated, and fine-tuned.¹⁰⁸ While similar in that each algorithm approaches problems in its own way, the three approaches differ at the outset, when the problem-solving environment is created.¹⁰⁹

Supervised learners are broadly characterized as either classification or regression systems. In classification systems, an algorithm is given both a training data set and the specific labels that

¹⁰³ These are often posted on a website called arXiv, which is maintained by Cornell University, and provides an open-access archive for scholarly materials. The dominant fields include physics, mathematics, computer science, quantitative biology, quantitative finance, statistics, electrical engineering and systems science, and economics. The papers are not peer reviewed by arXiv prior to posting. See arXiv, online: <arxiv.org> [perma.cc/2WTQ-72YL].

¹⁰⁴ Some commentators describe ML as a type of AI. This view flattens considerable nuance, but also introduces a debate about what it means for a computer to become intelligent and when the leap is made from mere mathematical regression models to something that we would describe like a human cognizing function. For our purposes, it is sufficient to describe ML, and particularly deep learning, as one of AI's enabling technologies. They are sometimes discussed synonymously, and other times presented as dichotomous. Professor Perez-Breva notes that most of what is branded as AI is not really AI at all: it consists of “complicated training and data-intensive learning systems.” While these are “most definitely ML capabilities, that does not make them AI capabilities.” Schmelzer, *supra* note 98. See also Broussard, *supra* note 97 at 92.

¹⁰⁵ Harry Surden, “Machine Learning and Law” (2014) 89 Washington L Rev 87 at 89 [Surden, “ML & Law”].

¹⁰⁶ In fact, two ML-enabled Facebook chatbots developed not only their own program, but also their own language that was unintelligible to anyone else. It was later determined that the bots had developed a *de facto* shorthand, still using English, but it nonetheless freaked everyone out and was shut down shortly thereafter. See Robby Berman, “Facebook Catches Two Chatbots Speaking Their Own Language” (6 August 2017), online: *BigThink* <bigthink.com/robbey-berman/facebook-catches-2-chatbots-speaking-their-own-language> [perma.cc/5N9F-YSBL].

¹⁰⁷ Broussard, *supra* note 97 at 93.

¹⁰⁸ *Ibid.*

¹⁰⁹ There are some other ways to train algorithms. For example, McGill researchers have attempted to measure the semantic distance between online concepts by having people play “Wikispeedia,” a variant of the popular online game “Wikiracing,” in which participants attempt to navigate from one Wikipedia page to another, seemingly unrelated page, using the least number of steps. See Robert West et al., “Wikispeedia: An Online Game for Inferring Semantic Distances between Concepts”, Proceedings of the Twenty-First International Joint Conference on Artificial Intelligence (IJCAI-09), online: *Stanford* <infolab.stanford.edu/~west1/pubs/West-Pineau-Precup_IJCAI-09.pdf> [perma.cc/XE8F-RXYY].

must be applied to appropriately categorize the data set.¹¹⁰ A long list of input variables are inputted, each of which is connected to a “correct” label. Combining an input variable with its correct label produces a value known as the “ground truth.” Over time, with practice, the algorithm learns how to predict the ground truth value based on the input alone, without access to its corresponding label. In regression systems, an algorithm is given a pattern and then applies the pattern to a data set. The algorithm then learns related patterns, becoming more accurate at predicting outcomes—as in classification systems. Over time, this “learning process” increases the likelihood of a positive matched result. In both types of supervised systems, user feedback is continuously solicited to continue to hone the algorithm. As users indicate how closely the machine identified the relevant ground truth criterion, the algorithm receives ongoing training.¹¹¹

Commonly used supervised learning methods include Naïve Bayes, nearest neighbour, and decision-trees. Naïve Bayes can be expressed as a simple equation but can be taught to do almost anything.¹¹² Based on Bayes’ theorem, a mathematical technique for calculating the probability that a particular belief is true, these algorithms can diagnose medical conditions, or identify fraudulent credit card transactions, or segregate spam email from your email inbox.¹¹³ Nearest-neighbour algorithms are empowered to search for the closest possible alternative, which can be used for diverse applications, ranging from detecting handwriting, to controlling robot hands, to recommending books or movies you might like.¹¹⁴ Decision tree learners, by contrast, follow a logic-fueled course of action to answer questions such as whether your credit card should be accepted or what should be the next move in a game of chess.¹¹⁵ Decision trees can be used to solve either regression or classification tasks, following along from the initial root nodes through various decision nodes, until eventually reaching the solution, or leaf node.¹¹⁶

By contrast, unsupervised learners are not given the labels for categorizing the data and do not function with respect to a ground truth.¹¹⁷ Instead, as the algorithm processes the training data, it spontaneously determines the categories independently by grouping similar objects together. This process is called clustering, and it is similar to what babies do as they learn to make sense of the world around them.¹¹⁸ In an unsupervised scenario, the training data might more accurately be called verification data: the algorithm is not receiving any actual training, but the only way that the human developer can confirm that the algorithm is developing a meaningful model is by deploying it over previously verified data, thereby ensuring that it is creating useful representations. Finally,

¹¹⁰ Broussard, *supra* note 97 at 93.

¹¹¹ For an everyday example, think about the small puzzles regularly encountered online to prove you are “human,” things like “select all the pictures of bicycles” in an image of a park. If you were asked to identify six images of bicycles, five of them would already be tagged, while one would be requiring human input to help train the algorithm. Rugare Maruzani, “Are You Unwittingly Helping to Train Google’s AI Models?” (26 January 2021), online: *Towards Data Science* (towardsdatascience.com/are-you-unwittingly-helping-to-train-googles-ai-models-f318dea53aee) [perma.cc/RS38-WQ5A].

¹¹² Emmanuel Ileberi, Yanxia Sun & Zenghui Wang, “A machine learning based credit card fraud detection using the GA algorithm for feature selection” (2022) 9 *Journal of Big Data* at 1.

¹¹³ Broussard, *supra* note 97 at 103.

¹¹⁴ *Ibid.*

¹¹⁵ Ileberi et al, *supra* note 112 at 3.

¹¹⁶ *Ibid.*

¹¹⁷ Broussard, *supra* note 97 at 93.

¹¹⁸ This is similar to how babies learn to make sense of the world around them: over time, a baby comes to understand that Mommy is not replaced by a smaller version of Mommy as she walks away. Clustering describes the ability to learn that both the large object in the foreground, and the small object in the background, are the same person. Pedro Domingos, *The Master Algorithm* (New York, NY: Basic Books, 2015) at 205 [Domingos, *Master*].

reinforcement learners take a hybrid approach: they are given freedom to make choices and mistakes but are given feedback after each effort as to their performance metrics and how to improve in future outings.¹¹⁹ Irrespective of which algorithmic approach is employed, detecting patterns is at the core of machine learning: it permits many commercial applications, including Internet searches, facial recognition software, credit card fraud detection systems, and data mining.¹²⁰ If the ML algorithm performs well, one might even say it appears “intelligent.”¹²¹ This technology powers predictive analytics and is, increasingly, viewed as a proxy mechanism for intelligence—despite the semantic concerns catalogued above.

To broadly construe the operation of ML systems, consider an example: an email spam filter.¹²² Email spam filters offer a clear opportunity for automation: receiving spam is undesirable, and the process of opening every received email to identify whether it is spam is laborious and time-consuming. Advances in ML have greatly diminished the quantity of spam email that makes it to our respective inboxes, as automated processes run in the backgrounds of most of today’s email clients.¹²³ The technology is powered by a classification algorithm: one designed to classify an input variable into a specific class. In this type of situation, the input is a vector comprised of unique feature values, while the output would be a classifier label.¹²⁴ An arbitrary condition is set, and each of the unique feature values from the input vector is checked against it. The algorithm’s output is 1 if the condition is met (*i.e.*, the email is spam) and 0 if the condition is not met (*i.e.*, the email is not spam).

For its training, the algorithm is shown a few baseline spam emails. Showing examples helps set the relevant arbitrary condition, like whether the message contains some word or phrase indicative of spam email, like “Earn Easy Money!” or common misspellings of lucrative pharmaceuticals.¹²⁵ In this process of supervised learning, the algorithm trains by reviewing a set of verified examples of the spam email phenomenon. When it first begins training, the algorithm will be somewhat anemic: having only seen a few examples of the phenomena it is meant to be identifying, its internal rule set will be underdeveloped. Only after seeing hundreds, or perhaps thousands, of examples does the algorithm “learn” to produce a highly effective model.¹²⁶ With continuous user feedback, the algorithm eventually becomes effective: users tagging unidentified spam emails as spam provides ongoing, on-the-job training.

Like chess or video games—other tasks that ML excels at—email spam filters lend themselves well to automated processes because they have right-or-wrong answers.¹²⁷ Right-or-wrong answers

¹¹⁹ Broussard, *supra* note 97 at 93.

¹²⁰ Surden, “ML & Law,” *supra* note 105 at 90.

¹²¹ *Ibid.*

¹²² Incidentally, Harry Surden uses this same example to explain ML. *Ibid.* at 1312.

¹²³ Litmus, “How does each spam filter score emails?” (16 December 2021) online: *Litmus* <help.litmus.com/article/205-how-does-each-spam-filter-score-emails> [perma.cc/5KV3-Z7TG].

¹²⁴ To be more specific, the input would be a Boolean vector, in the form $x = (x_1, \dots, x_j, \dots, x_d)$, where $x_j = 1$ if j th arbitrary condition is met, and $x_j = 0$ otherwise. The learner works by operating over a training set (x_i, y_i) , where $x = (x_{i,1}, \dots, x_{i,d})$ are the observed inputs, and $y_i =$ the corresponding output for each case, a classifier. The algorithm works if the classifier will produce the correct output y_i for future cases of x_i . For further explanations of the relevant vector algebra, see Pedro Domingos, “A Few Useful Things To Know About Machine Learning” (2012) 10 *Communications of the ACM* 78 at 79 [Domingos, “Few Things”].

¹²⁵ Domingos, *Master*, *supra* note 118 at 6.

¹²⁶ *Ibid.* at 92.

¹²⁷ Surden, “AI & Law,” *supra* note 87 at 1323.

work well in binary expression. Either an email is spam, in which case the algorithm will output a 1, relegating the message to a specific spam folder; or the email is not spam, in which case the algorithm will output a 0, allowing the message to end up in your inbox. Before calculating these binary results, however, each individual email is assigned a “spam score,” usually between 0-10, which represents how likely it is to be spam.¹²⁸ In the common email client Microsoft Outlook, a spam filter set to “high sensitivity” will rate an email with a spam score of 6.0 or higher as a failure.¹²⁹ These are the emails that get sent to the spam folder. While such systems are reasonably successful, sometimes errors are made and emails that are actually spam will end up in your inbox. The error rate underscores the limitations of ML systems: even though there are millions of emails identified as spam every year, mischaracterizations remain possible, notwithstanding the extremely well-developed data set. Many organizations still require automated headers added to external emails because spam is such a prevalent issue.¹³⁰ And, despite these challenges, spam remains a relatively straightforward use case. By contrast, most of the world’s problems do not lend themselves to binary, right-or-wrong answers. Even though ML technology struggles to deal with such complexity, automated systems continue to be proposed to handle many new scenarios. Such suggestions arise due to the incredible wealth of data that has been produced through modern telecommunications, social media, and Internet usage.

b. How Data is Collected & Interpreted

In the past decade, data aggregation became possible on a newly massive scale as increasing aspects of human life have moved online. Correspondingly, all major technology companies are turning their attention to developing AI applications. The most recent industry data reports are that total annual investment in AI in 2019 was around \$70 billion USD, with over \$37 billion being directed specifically to AI startups.¹³¹ Addition of the phrase “AI” to any technological application has become commonplace, and seems to generate newfound interest in pre-existing technologies.¹³² Somewhat ironically, however, the algorithms that power machine learning are not recent discoveries, but pre-existing technology rendered newly feasible through the transformative impact of recent Internet communications behaviours on the accumulation of data.¹³³ Deep learning, for example, became briefly popular in the 1970s, but then faded out of view for a time.¹³⁴ The framework on back propagation that enables many of today’s ML methods, such as self-driving cars, was invented by Yann LeCun in the late 1980s.¹³⁵

In today’s information economy, the business models of most smartphone-enabled Internet communications methods are part of this framework. The potential uses of ML have exponentially skyrocketed in response to the magnitude of available data. Correspondingly, corporations are becoming more interested in making sure their decisions are “data-driven.”¹³⁶ Shoshanna Zuboff

¹²⁸ Litmus, *supra* note 123.

¹²⁹ *Ibid.*

¹³⁰ Think of the bold red “Attention: courriel externe | external email” that adorns all emails that uOttawa email account holders receive from outside the University of Ottawa, designed to prevent phishing scams.

¹³¹ Raymond Perrault et al, “The AI Index 2019 Report”, AI Index Steering Committee, Human-Centered AI Institute, Stanford University, December 2019, online: *Stanford* <hai.stanford.edu/ai-index-2019 > [perma.cc/2ZQC-LZVT].

¹³² AI, after all, is simply statistics with better marketing.

¹³³ Christian, *supra* note 40 at 22-24.

¹³⁴ *Ibid.*

¹³⁵ *Ibid* at 234.

¹³⁶ McKinsey, *supra* note 81.

characterizes such activities as surveillance capitalism, noting that modern corporate revenues flow from “data assets appropriated through ubiquitous automated operations,” transforming our system of capital into the “default business model for most online companies and startups, where valuations routinely depend on ‘eyeballs’ rather than revenue as a predictor of remunerative surveillance assets.”¹³⁷ Enabled by a “logic of accumulation,” the datafication of modern business models commodifies and monetizes different aspects of human behaviour.¹³⁸ Julie Cohen refers to the oversaturation of data as “infoglut—of ‘an unimaginably unmanageable flow of mediated information.’”¹³⁹ Rather than information scarcity, political and epistemological problems flow from an overabundance of information, creating “the need for new ways of cutting through the clutter.”¹⁴⁰ Safiya Noble points out that classification systems necessarily entail the imposition of some kind of hierarchy, which can reinforce existing hegemonies if not thoughtfully interrogated.¹⁴¹

Thinking critically about big data reveals some issues with the “logic of accumulation” on an epistemological level, as the assumption that data represents an objective account of reality is quite easily displaced. Consumers can be nudged to pursue particular ends, with subtle behaviour modifications encouraged through corporate practices.¹⁴² Adjacent here are issues of online misinformation or disinformation, which can sully the acquisition of data and pursuit of knowledge,¹⁴³ or the reward-focused aspects of human psychology, which can nudge users to behave in certain ways on online platforms, in pursuit of a highly pleasurable dopamine boost.¹⁴⁴ Separately, one can also critique the quality, provenance, and organization of a compiled dataset. Researchers danah boyd and Kate Crawford note a series of critical questions for big data, problematizing the assumption that bigger is necessarily better.¹⁴⁵ Big data sets might fail to accurately represent marginalized social subgroups, lose meaning once decontextualized in a massive dataset, or skirt the established norms for ethically procuring research data.¹⁴⁶ Many ML researchers often rely on publicly available datasets to train and test algorithms, which are drawn from sources such as movies watched on the streaming service, Netflix; or emails sent between coworkers at the failed energy company, Enron.¹⁴⁷ While these datasets are easily accessed, they do not necessarily count for all eventualities and are not generated using robust social science

¹³⁷ Shoshanna Zuboff, “Big other: surveillance capitalism and the prospects of an information civilization” (2015) 30 *Journal of Information Technology* 75 at 81.

¹³⁸ *Ibid* at 85.

¹³⁹ Julie Cohen, “Chapter 3: The Information Laboratory” in *Between Truth and Power: The Legal Constructions of Information Capitalism* (2019) at 1; online: <juliecohen.com/between-truth-and-power/> [perma.cc/QWY6-L3ZW] [Cohen, *Truth & Power*].

¹⁴⁰ *Ibid*.

¹⁴¹ Safiya Umoja Noble, *Algorithms of Oppression: How Search Engines Reinforce Racism* (New York, NY: New York University Press, 2018) at 148.

¹⁴² Zuboff notes that because “big data analysis yields only correlational patterns,” big companies like Google or Facebook continuously experiments on their users, changing settings to gently nudge them towards certain behaviours, to “tease out issues of causality.” Eventually, insights gleaned from these experiments are monetized. Zuboff, “Big other,” *supra* note 137 at 84.

¹⁴³ Academic literature increasingly distinguishes between misinformation, which is incorrect but does not necessarily connote intention to mislead, from disinformation, which is deliberately misleading or biased. See generally, UW Bothwell & Cascadia College, Library Guides, “News: Fake News, Misinformation & Disinformation” (17 March 2022), online: *UW Guides* <guides.lib.uw.edu/c.php?g=345925&p=7772376> [perma.cc/WN3Y-ABZM].

¹⁴⁴ Christian, *supra* note 40 at 133-137.

¹⁴⁵ danah boyd & Kate Crawford, “Critical Questions for Big Data: Provocations for a cultural, technological, and scholarly phenomenon” (2012) 15 *Information, Communication & Society* 662.

¹⁴⁶ boyd & Crawford, *supra* note 145 at 671-672.

¹⁴⁷ Broussard, *supra* note 97 at 94.

research methods. Other behaviours produce no data, reversing the inference that might be drawn from a dataset, since the associated activity does not accumulate any data exhaust. This absence of information is called dark data, referring to circumstances where accumulated data does not accurately represent the full picture.¹⁴⁸ Meanwhile, some events are altogether unprecedented, offering no real warning. These are black swan events, which cannot be predicted based on past behaviour, because they have never happened before.¹⁴⁹ And some knowledge is tacit, absorbed from the world around us without ever being explicitly taught or discussed.¹⁵⁰ Tacit knowledge was described by chemist Michael Polanyi, who grew frustrated with how scientists treated knowledge, and wanted to address the degree to which certain “beliefs are anchored in ourselves.”¹⁵¹

While data debates continue amongst critical scholars, the developers charged with building ML tools are rarely involved.¹⁵² They tend to rely on the available datasets and the common mathematical or computer science understandings of terms, which can sometimes vary from how terms are used in other areas of inquiry. For example, one of the key problems arising in ML programming is overfitting, which refers to a tendency for the learner to align itself too specifically to the training data.¹⁵³ Domingos explains overfitting as a learner that correctly classifies 100% of the training data, but only 50% of the test data—when it could have scored 75% on both.¹⁵⁴ This tendency to develop an algorithm that is too closely aligned with the needs of the training data can lead to biased results when the algorithm, which functioned so well in training, is deployed on actual data. Although the developers use the term “bias,” they do so with a very different meaning than it carries in law or social science. In programming, bias refers to ML’s tendency to learn the wrong thing, time and time again.¹⁵⁵ Correcting bias, from the perspective of the developer, will result in the system responding consistently to the data. The inverse of bias is variance: the tendency of ML to learn random things, despite the actual signal.¹⁵⁶ Variance might result from a system trying too many different hypotheses, thereby leading to the counterintuitive conclusion that a simpler learner might be better than a more complicated one.¹⁵⁷ As a general rule, Domingos posits, “more data beats a cleverer algorithm.”¹⁵⁸

¹⁴⁸ David J Hand, “Dark Data: Why What You Don’t Know Matters” (Princeton, NJ: Princeton University Press, 2020) at 4-5. Hand analogies dark data to dark matter in physics: even though it cannot be seen or recorded, it still “can have a major effect on our conclusions, decisions, and actions.”

¹⁴⁹ Nassim Nicholas Taleb, *The Black Swan: the impact of the highly improbable* (London, UK: Penguin Books, 2010). Taleb notes that although black swan events are usually complete surprises, they also tend to be rationalized after the fact, justifying them as though they might have been expected.

¹⁵⁰ Kenneth A Grant, “Tacit Knowledge Revisited—We Can Still Learn from Polanyi” (2007) 6 *Electronic Journal of Knowledge Management* 173 at 176. Sometimes this knowledge is altogether ineffable; Grant gives the example: “Ask Tiger Woods how he hits his golf shots.”

¹⁵¹ *Ibid* at 175.

¹⁵² Admittedly, this has started to change in recent years, with a line of inquiry in computer science known as fairness, accountability, and transparency in machine learning (or FAT/ML). See generally Oxford Internet Institute, “Fairness, Accountability, and Transparency in Machine Learning” (last accessed 31 July 2022), online: *OII* <www.oii.ox.ac.uk/study/courses/introduction-to-fairness-accountability-and-transparency-in-machine-learning/> [perma.cc/9BWW-RWHZ]. This remains somewhat of a minority view, as evidenced by the recent controversy where Google fired well-known AI ethics researcher Timnit Gebru for participating in a draft paper that criticized the equality dimensions of natural language processing systems. See Chapter 8 for a full discussion.

¹⁵³ Domingos, “Few Things,” *supra* note 124 at 81.

¹⁵⁴ *Ibid*.

¹⁵⁵ *Ibid*.

¹⁵⁶ *Ibid*.

¹⁵⁷ *Ibid*.

¹⁵⁸ *Ibid* at 84.

The way models are derived is similarly difficult to understand, plagued by the “curse of dimensionality,” or the acknowledgement that “intuition fails in high dimensions.”¹⁵⁹ Generalizing a problem correctly becomes more difficult with the number of features that the problem exhibits.¹⁶⁰ While we might possess the mental faculties to express a problem with two or three dimensions, a model with 100 dimensions would defy intuition. And the problem grows exponentially: something represented in 100 dimensions with 100 variables would offer 100^{100} , or 1 trillion possible examples.¹⁶¹ Relatedly, scalability becomes an issue: even if a better model might assist, simpler models are usually selected because the more complicated ones are harder and more time-consuming to learn.¹⁶² Again, harnessing a massive data set will often yield better results than designing a better algorithm, removing the incentive to improve the tool.

Domingos struggles with this idea in *The Master Algorithm*; while he exhibits a technologist’s enthusiasm, he also injects a dash of scepticism. He writes:

Our digital future begins with a realization: every time you interact with a computer—whether it’s your smart phone or a server thousands of miles away—you do so on two levels. The first one is getting what you want there and then: an answer to a question, a product you want to buy, a new credit card. The second level, and in the long run the most important one, is teaching the computer about you. The more you teach it, the better it can serve you—or manipulate you. Life is a game between you and the learners that surround you. You can refuse to play, but then you’ll have to live a twentieth-century life in the twenty-first. Or you can play to win.¹⁶³

Playing to win in the age of surveillance capitalism seems to require a data-driven approach. Yet, even within this buzzword-y phrase, some ambiguity persists. What exactly are data-driven approaches driven toward? If we adopt the optimization frame of an ML developer, it follows that some value is being maximized or minimized. Do we seek to maximize something desirable, like justice or equality? How about minimizing something undesirable, like oppression or injustice? Or, in a more cynical capitalist frame, do the corporate interests involved seek to maximize profit, and perhaps someday, minimize human involvement all together? Even without clarity on these sorts of issues, AI is increasingly proposed as a governance mechanism across societal sectors.

2. AI and Governance

Despite the challenges with aggregated datasets and logics of accumulation, technologically enabled predictability has key business advantages. Humans tend to be uncomfortable with uncertainty. We like predicting things and we like to do it accurately. We used to rely on astrology, shamans, or soothsayers for predictions, but it turns out data science is more reliable. Our increased use of our technological devices is making that data ever more reliable: we can now train our

¹⁵⁹ Domingos, “Few Things,” *supra* note 124 at 82.

¹⁶⁰ Jenna Burrell, “How the machine ‘thinks’: Understanding opacity in machine learning algorithms” (2016) *Big Data & Society* 1 at 5.

¹⁶¹ Lillian Edwards & Michael Veale, “Slave to the Algorithm? Why a ‘Right to an Explanation’ Is Probably Not the Remedy You Are Looking For” (2017) 16 *Duke Law & Technology Review* 18 at 56.

¹⁶² Domingos, “Few Things,” *supra* note 124 at 85.

¹⁶³ Domingos, *Master*, *supra* note 118 at 264.

computers to predict our keystrokes, our purchasing habits, our commuting patterns, and so much more. In this vein, Mireille Hildebrandt envisions a not-so-distant future world where our houses will awaken us as required by our daily schedules, feed us in accordance with our calculated nutritional needs, and send us off to work perfectly calibrated to achieve quotidian success.¹⁶⁴

Hildebrandt distinguishes between different technological forms of regulation that shape society. Thinking through an oft-used example—machine-facilitated diagnoses in medicine—she observes an increase in “heteronomy” and a corresponding decrease in “autonomy.”¹⁶⁵ By virtue of using a mechanical solution, there is an increased likelihood of achieving consistent results across cases, and a corresponding decrease in how the individual practitioner thinks about, or engages with, the problem’s diagnosis. In some such scenarios, the mechanical diagnoses might end up being altogether invisible, thereby rendering them all-but incontestable.¹⁶⁶ As mechanical diagnoses improve over time, the eventual outcome of their use might well be ML-generated diagnoses that cannot be easily understood by medical professionals.¹⁶⁷ Moreover, automated solutions that rely on sophisticated pattern-matching assume that whatever patterns are in the dataset are also in the world around them—which isn’t necessarily the case.¹⁶⁸ Without empirical testing of the dataset, such patterns are impossible to confirm.¹⁶⁹

Some approaches to automation envision a hybrid system where humans and machines work collaboratively, each specializing in their respective areas of expertise. Such systems are typically referred to as having a “human in the loop.”¹⁷⁰ The machine functions autonomously insofar as possible, but defers to a human operator for certain decisions, such as “to make a difficult judgment or an assessment that remains outside the system’s capability or for which a computer decision is deemed societally inappropriate.”¹⁷¹ This modality is common for semi-autonomous systems deployed in the warfare context, acknowledging the legal and moral aspects of the laws of war that impede technology from exerting inappropriate authority on the battlefield.¹⁷² As automated systems are deployed in more diverse scenarios, many worry about their use in the legal realm.

a. AI and Legal Regulation

Hildebrandt worries about the consequences of applying algorithmic decision-making processes to legal regulatory arenas, including legislation, administration, and judicial adjudication.¹⁷³ Displacing human regulation with artificial analogues is fraught with challenges, she argues, especially given the degree to which legal authority is situated within texts, and how it is significant for “the force of law.”¹⁷⁴ She ponders the widespread consequences:

¹⁶⁴ Hildebrandt, *End(s) of Law*, *supra* note 30 at 4.

¹⁶⁵ *Ibid* at 68.

¹⁶⁶ *Ibid* at 69.

¹⁶⁷ Kerr, “Balkinization,” *supra* note 73.

¹⁶⁸ Hildebrandt, *End(s) of Law*, *supra* note 30 at 69.

¹⁶⁹ *Ibid*.

¹⁷⁰ Kerr & Szilagyi, *supra* note 42 at 339.

¹⁷¹ Surden, “AI & Law,” *supra* note 87 at 1320.

¹⁷² Kerr & Szilagyi, *supra* note 42 at 336.

¹⁷³ Hildebrandt, “Algorithmic Regulation,” *supra* note 32.

¹⁷⁴ *Ibid* at 2.

Are we confronting a conflation of legislation and administration, insofar as legislation becomes self-executing? Could legal judgment at some point be conflated with its prediction? Will automated systems run on proprietary software or be too complex to explain? If these systems are not testable, can they be contestable? Do algorithmic decision systems require new types of interpretability, on the nexus of machine learning and law? Does this require a new legal hermeneutics?¹⁷⁵

Hildebrandt then narrows her focus, describing two types of regulation: code-driven regulation and data-driven regulation.¹⁷⁶ Code-driven regulation depends on “if this then that”-styled regulation, or approaches that flow through a decision tree.¹⁷⁷ Inherently deterministic, such approaches require a formalization of legal norms to be expressed instead in computer code.¹⁷⁸ If one wants to appeal the outcome of a code-driven decision, the obvious complaint is that the legal norm has been improperly translated into code.¹⁷⁹ This approach contrasts data-driven regulation, where code evolves with its training data set, changing into something ever more powerful.¹⁸⁰ Data-driven approaches are not deterministic; they are powered by ML algorithms to form artificial legal intelligence.¹⁸¹ One’s ability to appeal the outcome of a data-driven decision is hampered by the inscrutability of the solution: depending on the algorithm’s design choices, it might not be possible to achieve insight into how its decision was made.¹⁸²

Sheila Jasanoff offers insight into the relationship between law and technology, writing in 1996: “[t]he law today not only interprets the social impacts of science and technology but also constructs the very environment in which science and technology come to have meaning, utility, and force.”¹⁸³ Citing cases that struggle with epistemological issues over “the authority of knowledge,” Jasanoff investigates when, and by what criteria, knowledge should in fact count as valid science.¹⁸⁴ While Jasanoff’s inquiries are focused on how science is utilized within the legal system, her observations offer prescient analogues for the cognitive and social aspects of self-driving laws. Traditionally, courts have exercised a gatekeeping function with respect to the construction of knowledge, certifying experts before purporting to take their credentials seriously.¹⁸⁵ The incursions of ML methods into law threaten such long-established principles, especially without insight into which datasets are used for training and verification, or what types of design choices are being made.

These outcomes create trouble for legal regulation, which Hildebrandt contends is focused “not merely on the modification of behaviour, but on coordinating, prohibiting and enabling action in a way that addresses individuals that are subject to law as capable of giving reasons for their actions

¹⁷⁵ Hildebrandt, “Algorithmic Regulation,” *supra* note 32 at 1-2.

¹⁷⁶ *Ibid* at 2.

¹⁷⁷ *Ibid*.

¹⁷⁸ *Ibid* at 3.

¹⁷⁹ *Ibid*.

¹⁸⁰ *Ibid*.

¹⁸¹ *Ibid*.

¹⁸² *Ibid*.

¹⁸³ Sheila Jasanoff, *Science at the Bar: Law, Science, and Technology in America* (Cambridge, MA: Harvard University Press, 1997) at 16.

¹⁸⁴ *Ibid* at 19.

¹⁸⁵ *Ibid* at 58.

and in a way that respects their autonomy.”¹⁸⁶ The unearned certainty at the heart of today’s ML methods thereby engages the Rule of Law, given the degree to which contestability is a critical consideration for legal procedure.¹⁸⁷ As I discuss further below, the obfuscation of our data, the power that it possesses in aggregate, the conflation of government and business interests: all of these reduce citizens’ abilities to contest legal decisions in the way that the Rule of Law anticipates.

b. AI and Corporate Power

In *New Laws of Robotics: Defending Human Expertise in the Age of AI*, Frank Pasquale begins a discussion of democracy through the lens of workplace economics.¹⁸⁸ Introducing robots to the workplace, he points out, is often a question of efficiency, and ensuring humans retain control over the workplace will necessitate careful calibration of capitalist tendencies. A similar phenomenon is described by Astra Taylor as “fauxtimation,” where the work accomplished through human effort is falsely described as automated.¹⁸⁹ Taylor reminds readers that in these situations, it is not the robots themselves who are gunning for your job, but rather the privileging of business owners’ quests for efficiency that impacts employment outcomes. This explanation demonstrates how AI might encroach on various aspects of private life in a way that engages the public interest. For this proposition, Pasquale quotes mid-century legal academic Robert Lee Hale: “there is government whenever one person or group can tell others what they must do and when those others have to obey or suffer a penalty.”¹⁹⁰

Harnessing Hildebrandt’s lessons of “code-driven regulation” as a “pre-set logic that remains constant over time,” Jennifer Raso notes that some instances of computational decision-making are still fairly low-tech and describing them as automated “obscures the degree to which these tools rely on the labour of many humans who enter data, maintain database infrastructures and communicate results to impacted individuals.”¹⁹¹ This “fauxtimation” phenomena has also been referred to as “ghost work” and can appear to amplify the success of AI or ML tools, when human decision-making still powers the key functionality of the technology.¹⁹² Workers participating in this gig economy continue to develop their capitalist selves, as hidden units of production that enable the success of AI-driven solutions.

Ultimately, the transformative impact of the AI industries adheres through digital platforms. Through the lens of surveillance capitalism, changes to various industries enabled by digitization are opportunities to be mined, permitting digital mediation of social and economic interactions, allowing interactions to be recorded, reported, and searched. Also known as platform economics or platform capitalism, the direct connection between consumers and products are often carefully

¹⁸⁶ Hildebrandt, “Algorithmic Regulation,” *supra* note 32 at 5.

¹⁸⁷ *Ibid* at 8.

¹⁸⁸ Frank Pasquale, *New Laws of Robotics: Defending Human Expertise in the Age of AI* (Cambridge, MA: Harvard University Press, 2020) at 176 [Pasquale, *New Laws*].

¹⁸⁹ Kerr, “Balkinization,” *supra* note 73.

¹⁹⁰ Pasquale, *New Laws*, *supra* note 188 at 176. Pasquale cites Robert Lee Hale, *Freedom through Law* (New York, NY: Columbia University Press, 1952), but doesn’t offer a page reference.

¹⁹¹ Jennifer Raso, “AI and Administrative Law” in Florian Martin-Bariteau & Teresa Scassa, eds, *The Law of Artificial Intelligence in Canada* (Ottawa: LexisNexis 2021) 181 at 186.

¹⁹² Raso, *supra* note 191 at 186. Another example is Amazon’s “Mechanical Turk” platform, which hires remote workers to do simple computational actions. These are often kinds of things AI is thought to be doing: Mechanical Turk’s tagline is “artificial artificial intelligence.” See Christian, *supra* note 40 at 22.

designed to avoid existing regulatory schemes.¹⁹³ This economic form, Cohen argues, has enabled a transition from market power to platform power, and dominant platforms “have thrived by developing ways to offer both individual and business users a wide variety of information services while controlling both advertising markets and the harvesting and processing of user data.”¹⁹⁴ The corresponding logics of extraction bring to mind colonial worldviews, as data is frequently analogized to oil as the latest world-changing resource.¹⁹⁵

3. AI and the Legal Futurists

Translating ML techniques from capitalist practices to enforcing legality amplifies the risks of pre-emption, as delineated by Kerr.¹⁹⁶ Very different consequences are at play for which advertisements a smartphone user engages with *versus* which legal outcomes are prescribed for a given court case, but the technology proposed to govern them is largely the same. Consisting of statistical methodologies, these ML techniques aim to predict the future based on past behaviour and other determining factors emerging as patterns from the dataset. While the legal system has historically been slow to change, the implementation of new technologies has been proposed to alleviate some of its exigent pressures. The predictive capabilities of ML coincide with legal spaces through the automation of certain tasks. In an overburdened justice system, there are demands to clear backlog and reduce the workload imposed on human decisionmakers. Furthermore, human decision-making remains fundamentally fallible, and thus subject to appeals. The technological gloss offered by introducing new algorithmic methods is appealing: the veneer of objectivity offered by technological solutions presents an alternative to the intrinsic messiness of human judgement. But is deviating from this system perhaps taking one form of messiness and replacing it with another? Why do we assume technological solutions, built by humans, will be free from the same fallibility we aim to correct?

As stated at the outset, in this thesis, I’ll refer broadly to the impacts of these AI-enabled legal technologies on the legal system as self-driving laws. Self-driving laws take many forms, popping up in a variety of different subject areas. In criminal law, sentencing algorithms have been used to predict the likelihood of recidivism.¹⁹⁷ Predictive policing programs now identify high risk areas to screen for criminal activity.¹⁹⁸ But despite the limited scope of implementation for such tools, equality concerns have already been identified. Criminal sentencing based on algorithms designed to predict likelihood of recidivism have been accused of racial discrimination. Predictive policing programs designed to identify high risk areas for criminal activity might mistake heightened levels of enforcement for higher crime rates, thereby further entrenching socioeconomic divides. Other areas of law also increasingly offer automated aspects, particularly in the administrative state, and in the everyday business of lawyering within law firms.¹⁹⁹

Certain optimistic scholars have contributed highly aspirational accounts to the literature. Law professor and legal entrepreneur Benjamin Alarie offers a particularly bold vision for the future of

¹⁹³ Cohen, “Chapter 6: The Regulatory State in the Information Age,” in Cohen, *Truth & Power*, *supra* note 139 at 25.

¹⁹⁴ *Ibid* at 5.

¹⁹⁵ Crawford, *Atlas*, *supra* note 76 at 113.

¹⁹⁶ Kerr, “Prediction, pre-emption,” *supra* note 67.

¹⁹⁷ This is the subject of Chapter 6.

¹⁹⁸ This is the subject of Chapter 7.

¹⁹⁹ These examples are considered in detail in Chapter 3.

self-driving laws: wholesale regulation by machine. In his TEDx lecture, he shared a personal narrative of his frustrations as a first-year law student frantically engaged in a “paper chase” as a library learning exercise.²⁰⁰ Days later, he realized that same information was digitally searchable in online databases. This anecdote did not lead him to the conclusion that the exercise was beneficial for understanding the systemization of legal information; instead, he declared books had reached the end of their era of cultural dominance. The book had a good run, he argued, and while helpful for defining the contours of the legal universe, it was now obsolete. Instead of seeking precedents in dusty library books, ML algorithms can now assist legal truth-seekers by combing a “complete” dataset and accessing ground truth.

Alarie’s proposed regulation by machine would aggregate all existing legal knowledge, generating the correct response for each new situation based on an algorithmic output of the necessary data.²⁰¹ Joined by Anthony Niblett and Albert Yoon, who are also the co-founders of his legal AI start-up, Blue J Legal, the three co-authors argue against the sceptics who wrongly claim “our industry is different.”²⁰² In their assertion, the coming tide of AI will disrupt all aspects of human industry, including law.²⁰³ In other work, Alarie calls this eventuality a “legal singularity.”²⁰⁴ Like Kerr, he begins his essay by quoting Holmes’ “The Path of Law,” building on Holmes’ concept of law as prediction.²⁰⁵ While Holmes used the frame of the bad man to better predict what courts might do from the view of person impacted, Alarie’s view of law as prediction is much less rhetorical. He explains his idea of the legal singularity as follows:

The legal singularity contemplates the elimination of legal uncertainty and the emergence of a seamless legal order, universally accessible in real-time. In the legal singularity, disputes over the legal significance of agreed facts will be rare. They may be disputes over facts, but the once found, the facts will map on to clear legal consequences. The law will be functionally complete.²⁰⁶

To be fair to Alarie, he does acknowledge that moral dimensions of law are currently beyond the purview of machines, admitting that “excelling at law ostensibly requires a sound heart as well as an able mind.”²⁰⁷ Yet, he is undeterred, steadfast in his optimism that the naysayers will “inevitably [be] demonstrated empirically to be incorrect.”²⁰⁸ Anchoring his analysis in the tax law space, he predicts “the coming decades will witness three gradual transitions” as we approach his legal singularity.²⁰⁹ First, we will experience an improved dispute resolution exercise, as we transition away from using standards to “query-able systems of complex rules.”²¹⁰ Next, we will transition to

²⁰⁰ Benjamin Alarie, “Transcending Books: What’s Next for Law and Society?” TEDxBeaconStreet (August 2018), online: *TED* <www.ted.com/talks/benjamin_alarie_transcending_books_what_s_next_for_law_and_society> [perma.cc/96UB-BFRN].

²⁰¹ Alarie, “Legal Singularity,” *supra* note 36 at 445.

²⁰² Alarie et al, “Future,” *supra* note 41 at 425.

²⁰³ *Ibid.* Attempting to dispel the sceptics, they write: “In the face of enormous technological disruption, however, the one thing that has not changed is the capacity of humans to underestimate the importance and impact of technological change.”

²⁰⁴ Alarie, “Legal Singularity,” *supra* note 36 at 445.

²⁰⁵ Alarie uses the exact same quotation as Kerr did to begin his piece (although, after using the quote in the abstract, he never invokes Holmes again): “For the rational study of the law the blackletter man may be the man of the present, but the man of the future is the man of statistics and the master of economics.” *Ibid* at 443. See Kerr, “Prediction, pre-emption,” *supra* note 67 at 91.

²⁰⁶ Alarie, “Legal Singularity,” *supra* note 36 at 446.

²⁰⁷ *Ibid.*

²⁰⁸ *Ibid.*

²⁰⁹ *Ibid* at 447.

²¹⁰ *Ibid.*

more sophisticated tax systems that better “distribute benefits and burdens.”²¹¹ Finally, we will arrive at a “complete specification” of tax law, by virtue of the arrival of the singularity, which “will thenceforth remain (more or less) in positive and normative equilibrium.”²¹² While Alarie begins with tax law, he nonetheless predicts this same “complete specification” will eventually arrive in all areas of law, which he suggests will fulfill John Rawls’ vision of “reflective equilibrium.”²¹³ Alarie’s analysis—peppered with language evocative of AI evangelists Ray Kurzweil and Nick Bostrom—commits the android fallacy that Smart and Richards caution against. Despite the hopeful promises for AI futures, its abilities to regulate law remain circumscribed. By positioning his analysis in a future where the technology does indeed work as he might hope, Alarie attempts to forestall critiques of the AI’s specific *machine*-ations. Yet, even before the technology is fully realized, the meagre abilities of today’s AI/ML systems produce sufficient red flags to temper his optimism.

Meanwhile, Anthony Casey and Anthony Niblett approach legal futurism from another angle. Acknowledging law’s longstanding efforts to tailor itself to the relevant context, and the exigent “frictions” that sometimes prevent all contextual factors from being taken into account, they propose a new personalization of law enabled by big data and predictive analytics.²¹⁴ This new framework will allow for “microdirectives” to be precisely responsive to an offender’s particular circumstances, considering factors that cannot have been known at the time of legislating.²¹⁵ Another enthusiast, Daniel Goldsworthy, seems to have independently conceptualized a “legal singularity,” inspired by Ronald Dworkin’s idealized Justice Hercules.²¹⁶ Excited by the successes of AI-enabled technologies described in this chapter, like Deep Blue, AlphaGo, and deep learning, Goldsworthy anticipates the day when neural networks will allow for constructive interpretation in legal texts, just as Dworkin dreamed Hercules could do.²¹⁷

The legal futurists’ proposals present vexing questions for law’s efficacy and overall scope. They also raise questions about day-to-day execution and enforcement of the law. Most importantly, regulation by machine seems to anticipate the existence of right answers. This “right answer thesis” has been a central feature of analytical jurisprudence, where its adherents remain in the minority. Most legal theorists understand law as requiring space for interpretation. Thinking through these problems engages theoretical concerns about the concept of law. Rethinking them with the added considerations of legal regulation, surveillance capitalism, and economic power also engages the Rule of Law. Hildebrandt explains this nexus to the Rule of Law as follows:

One can see the political as a separate realm and politics as a dirty game played by cynical men. Instead, the idea of the Rule of Law refers to a political order that is constrained in such a way that dirty play is ruled out or mitigated as dirty players are called to account; a political order where one power is counterbalanced by another, making sure that we are not ruled by men but by

²¹¹ Alarie, “Legal Singularity,” *supra* note 36 at 447.

²¹² *Ibid.*

²¹³ Alarie et al, “Future,” *supra* note 41 at 427.

²¹⁴ Casey & Niblett, *supra* note 37 at 333-335.

²¹⁵ *Ibid* at 347.

²¹⁶ Goldsworthy, *supra* note 38 at 286.

²¹⁷ *Ibid* at 288.

a law that can be transformed by those subject to its rule. Rule of Law is not rule *by* law by men; it is also not rule *by* law by machines.²¹⁸

For technology law scholarship, the Rule of Law is a somewhat unusual path. Scholars sometimes comment on it from a civil liberties lens, particularly in the context of technology enabling a surveillance society,²¹⁹ or removing barriers to accessing justice through streamlined technological processes.²²⁰ But this inquiry tends to be adjacent, operationalized in the context of writing about constitutional rights like due process or search and seizure. Apart from Hildebrandt's writing, only a few scholars have approached the topic using theoretical lenses. Roger Brownsword writes about "technological wedges" being driven into our understanding of law, which might require a "radical rebooting of our legal thinking," potentially "reshaping the Rule of Law and our conception of coherence in the law."²²¹ Helpfully, Brownsword acknowledges the substantive aspects of the Rule of Law in terms of the responsibilities regulators owe to the global commons, and the application of law to those regulators whether they happen to be acting in a public or private capacity.²²² While he aims to critique Alarie's approach through its impacts on Rule of Law values, he does not immediately set out what he believes the relevant values to be.²²³ Later, he summarizes the theoretical debate succinctly, writing "[w]hile we can spend a long time debating the details of the Rule of Law, the spirit surely is that, on the one side, the exercise of arbitrary power is to be constrained and that, on the other, the exercise of non-arbitrary power is to be respected."²²⁴

Paul Gowder, who developed his own substantive concept of the Rule of Law in his text *The Rule of Law in the Real World*, writes about the coercive aspects of power enabled through technological means.²²⁵ Gowder's insights help bridge the gap between the Rule of Law as a public law concept and the instances of arbitrary power realized through AI/ML methods deployed by private companies. Given private law's large role governing modern societies powered by surveillance capitalism, it may well be that thinking about Rule of Law actors should involve a broader spectrum of those who exert power over others. After all, as Gowder points out:

The inegalitarian dangers of arbitrary power are, it now seems to me, just as severe in the hands of a landlord or an employer or a distant corporation as in the hands of a police officer or a bureaucrat. The rules of Walmart can be just as Kafkaesque as those of the local welfare office. In either case, the individual is faced with a potentially vast disparity of power which, if it goes unchecked, can be used to

²¹⁸ Mireille Hildebrandt, "Law as an Affordance: The Devil Is in the Vanishing Point(s)" (2017) 4 *Critical Analysis of Law* 116 at 128 [Hildebrandt, "Affordance"].

²¹⁹ Kerr, "Prediction, pre-emption," *supra* note 67.

²²⁰ Julie Cohen, "From Lex Informatica to the Control Revolution" (forthcoming 2022) 36 *Berkeley Tech LJ* __; available online: SSRN <papers.ssrn.com/sol3/papers.cfm?abstract_id=3854219> [perma.cc/KN86-YAAT].

²²¹ Brownsword, *supra* note 46 at 2.

²²² *Ibid* at 3.

²²³ *Ibid* at 5.

²²⁴ *Ibid* at 13.

²²⁵ Gowder writes: "Understanding legal technology from a rule-of-law perspective suggests that the potential mass availability of technologically-mediated legal cognition can serve as an instrument of novel collective and social action by and on behalf of the powerless." Gowder, *supra* note 51 at 85.

deprive the one over whom power is exercised of basic interests and establish unacceptable relationships of hierarchical status.²²⁶

Monika Zalnieriute, Lyria Bennett Moses, and George Williams investigate how Rule of Law values are implicated by automated systems—both the code-driven and data-driven varieties.²²⁷ They define Rule of Law values similar to administrative law scholars, focusing on “transparency and accountability,” “predictability and consistency,” and “equality before the law.”²²⁸ Applying these values to different case studies of automation, their analysis focuses on the need for better transparency and accountability in automated systems, as proposed by Danielle Citron and Frank Pasquale.²²⁹ They also argue for better design and implementation, noting disastrous consequences for the Rule of Law when technologies are deployed inappropriately.²³⁰ Meanwhile, Pasquale offers additional insights into human judgment’s appropriate role in the legal system, emphasizing that the role of someone trained in the law goes beyond mere administration or delivery of expertise.²³¹ He explains:

The rule of law entails a system of social relationships and legitimate governance, not simply the transfer and evaluation of information about behavior. There is necessarily some degree of self-governance among professionals, which gives them an occupational identity distinct from other workers. Their primary fiduciary duty is to clients, not managers or shareholders. The main reason they enjoy this autonomy is because they must handle intractable conflicts of values that repeatedly require thoughtful discretion and negotiation. A robust and ethical legal profession respects that discretion, founded on the flexibility and subtlety of legal language, as a prerequisite for a just and accountable social order. It ensures a rule of persons, not machines.²³²

From this view, Pasquale emphasizes the need for circumscribed uses of legal automation.²³³ Later, joined by Cashwell, the two extend Kerr’s fear of prediction, referring to the use of predictive analytics in judicial decision-making as a “jurisprudence of behaviourism.”²³⁴ Using such prediction models in the legal system, they argue, may endanger Rule of Law values.²³⁵ Meanwhile, Jennifer Cobbe criticizes Alarie’s vision of the legal singularity in Rule of Law language.²³⁶ She notes that uses of AI in the legal system tend to “deprecate[] normative values in favour of market-oriented metrics of efficiency, speed, cost, and consistency.”²³⁷ Her analysis suggests concerns about logics

²²⁶ Gowder, *supra* note 51 at 88.

²²⁷ Zalnieriute et al, *supra* note 52 at 3. The authors distinguish between “human-authored pre-programmed rules (such as expert systems)” and “tools that derive rules from historic data to make inferences or predictions (often using machine learning).” The use of Hildebrandt’s nomenclature for these concepts is mine.

²²⁸ *Ibid* at 5-6.

²²⁹ *Ibid* at 17. They refer here to Citron’s idea “technological due process,” which I discuss in Chapter 3.

²³⁰ *Ibid* at 25-26.

²³¹ Pasquale, “Rule of Persons,” *supra* note 47 at 60.

²³² *Ibid*.

²³³ If an area is a complete mess, he concedes, algorithms would be advantageous to find patterns that remain beyond human comprehension. If, however, an area of law is perfectly ordered, then “an expert system can reduce it to a series of rules to be applied.” *Ibid* at 57.

²³⁴ Pasquale & Cashwell, *supra* note 48 at 63.

²³⁵ *Ibid* at 64.

²³⁶ Cobbe, *supra* note 54.

²³⁷ *Ibid* at 18.

of accumulation and economic power, juxtaposed against Alarie’s vision of the eventual “complete specification” of all forms of law, to be realized at the advent of the legal singularity.²³⁸ She takes issue with how Alarie has characterized algorithmic systems as capable of “interact[ing] neutrally as merely technical phenomena. The reflexivity of algorithms, their contextual, contingent, and normative nature, suggests otherwise.”²³⁹ Her analysis accounts for law’s role as a reflexive societal institution, which is continually reconstituted by social norms and economic transformations. Inherent in this discussion is the interplay between technological affordances and the legal norms that technology disrupts: a key theme of Chapter 2.

None of these Rule of Law approaches, however, engage with the narrative aspects of law. The closest available analogue is that proposed by Joshua Fairfield in his book, *Runaway Technology*, which takes inspiration from some of the same sources that I used in this dissertation.²⁴⁰ Discussing narrative’s power to assist in seeing things at an appropriate level, Fairfield acknowledges the Rule of Law’s origins as a “co-operative fiction.”²⁴¹ Law is composed of a series of narratives, offering a vocabulary for talking about what might happen when new issues arise.²⁴² Absent a specific purpose, the narrative nature of law will allow unnecessary aspects to eventually fade away.²⁴³ Science, too, is a collection of narratives, but scientists often refuse to venture too deeply into the “unscientific realm” that is story and narrative.²⁴⁴ In my view, this denial of narrative in science calls into question the objectivity of numbers being represented as “just data,” as though the data were not drawn from a highly narrative world. As the further chapters of this thesis will explore, operationalizing data without a context-sensitive understanding of the story it might tell yields significant consequences for the power dynamics of the Rule of Law.

Before heading in this direction, I pause to consider the lessons that might be gleaned from existing technology law scholarship, particularly the wealth of regulatory framing generated by the advent of the Internet. This field, known as cyberlaw, offers lessons for thinking through legal issues pertinent to AI and ML. Additionally, the interdisciplinary field of science and technology studies (STS) adds texture, deepening the social insights realized through the construction of technological systems. I turn now to these lessons in Chapter 2.

²³⁸ Alarie, “Singularity,” *supra* note 36 at 447.

²³⁹ Cobbe, *supra* note 54 at 19.

²⁴⁰ Fairfield’s text was published in 2021, more than four years after I had drafted my thesis proposal. See Fairfield, *Runaway*, *supra* note 50.

²⁴¹ *Ibid* at 113.

²⁴² *Ibid* at 62.

²⁴³ Fairfield refers to Richard Rorty’s insights on the history of language, noting that old metaphors are constantly dying off into literalness, and then serving as a platform and foil for new metaphors. *Ibid* at 64.

²⁴⁴ *Ibid* at 117.

2

What Cyberlaw & STS Might Teach

Another flaw in the human character is that everybody wants to build and nobody wants to do maintenance.
~ Kurt Vonnegut²⁴⁵

Once, on a brief stopover in New York City, I attended a law and technology panel discussion at Fordham University School of Law. One of the panelists, a criminal law professor with some interest in technology law, appeared to champion the availability of computational methods for determining sentences. Given all the troubles with achieving proper insight into human decision-making, he suggested, perhaps having an algorithm make the decision would be superior. He scoffed at some of the usual critiques about algorithmic decision-making being inscrutable or opaque. After all, he emphasized, we don't really know how humans arrive at their decisions anyway. While some criticize machines for being a "black box,"²⁴⁶ what goes on inside the human brain is equally opaque. He put a hypothetical question to the assembled audience:

What is it that makes us uncomfortable about a three-pound computer sitting on a desk making a legal decision, as opposed to the two-pound computer that we carry around on our shoulders?

No one successfully answered the question, and while the panelists soon moved on to discuss other topics, his inquiry stayed with me.²⁴⁷ There was a difference, I thought, but it was difficult to quantify and even more difficult to articulate.

Throughout history, people have often been distrustful of technology, especially new technology. Their treatment of it was bound up in unhelpful analogues from the past, inoperative metonyms that failed to meaningfully describe the nature of the new tool. Technological progress has long been accompanied by sweeping declarations that the world will be truly transformed, or led to ruin, by some new technology. People have long worried about how technology might be used to remake the world we know, and what familiar aspects of human society might be lost along the way. Such concerns were raised about the written word, the printing press, and even books—by the early 1600s, many expressed fears that there was now too much content to read and innovation would be stymied by scientists unable to stay abreast of new developments by reading all the latest papers.²⁴⁸ Words, and then words in print, were the algorithms of another generation. Arguments levied about the advent of the Internet were no different.

²⁴⁵ Kurt Vonnegut, *Hocus Pocus* (New York, NY: Penguin Books, 1997) at 167.

²⁴⁶ Frank Pasquale, *The Black Box Society: The Secret Algorithms That Control Money and Information* (Cambridge, MA: Harvard University Press: 2015) [Pasquale, Black Box].

²⁴⁷ Pasquale and Cashwell offer another version of this anecdote from Wisconsin assistant attorney general Christine Remington: "We don't know what's going on in a judge's head; it's a black box, too." Pasquale & Cashwell, *supra* note 48 at 64.

²⁴⁸ Nicholas Carr, *The Shallows: What the Internet is Doing to Our Brains* (New York, NY: WW Norton, 2011) at 71. Carr quotes from Spanish dramatist Lope de Vega's 1612 play "All Citizens Are Soldiers," where he wrote: "So many books—so much confusion! All around us an ocean of print; And most of it covered in froth."

When scholars first began critiquing the Internet, a tenor of moral panic sometimes pervaded the conversation. Now, as more scholars sound alarm bells about AI's capabilities, some accounts can be unduly hyperbolic or anticipate dystopias resembling science fiction.²⁴⁹ What's more: many recent critiques of AI and predictive analytics in popular culture have portrayed this moment as exceptional.²⁵⁰ Without a clear articulation of method within technology law scholarship, it can sometimes be difficult to pinpoint what exactly the legal threat of new technologies really is.²⁵¹ Indeed, students in technology law seminars typically begin the semester perplexed that what they are studying is not really law, at all. Eventually, their conception of what constitutes law and regulation broadens significantly, leading them to develop a more pluralistic view of both the emergent problems and their potential solutions: I aim to provide an account of this journey in this chapter.

I begin this chapter by mapping technology law scholarship through changing technological and legal contexts. First, I present the lessons from the cyberlaw literature of the past few decades. The Internet is a key exemplar here, including the Declaration of Independence of Cyberspace,²⁵² Lessig's model for explaining the transformational impacts of digital space through understanding of code as law,²⁵³ and early Internet regulatory models.²⁵⁴ By focusing on Internet—another recent technology presented as truly transformative—I heed the lessons from methodological shortcomings in recent technology law contributions. Second, I review lessons from science and technology studies and its interdisciplinary insights into how technological force co-constitute society. Employing this lens, I discuss academic responses to three key ideas: the idea that technology is the prime mover driving history forward; the idea that technology is a neutral vessel that is neither good nor bad absent application; and the techno-solutionism that prefers technological solutions over all others. These lessons, I argue, will be necessary for better understanding the potentially transformative impact of AI and ML on social institutions; these principles will later be used to critique and analyse AI and ML in legal applications. Finally, I draw together some key threads from the foregoing analysis to present marching orders for the future of law. Most importantly, I discuss the ideas of technological affordances and normativity. Here, I present law's normativity in conversation with its affordances. By thinking critically about what

²⁴⁹ STS Professor Lee Vinsel cites historian David C. Brock's idea of "wishful worries" and discusses media coverage of emerging technology issues, including AI, blockchain, social media, and surveillance capitalism. "Wishful worries" denote analyses of problems that it would be nice to have, rather than the ones we actually have. Lee Vinsel, "You're Doing It Wrong: Notes on Criticism and Technology Hype" (1 February 2021), online: *Medium* <sts-news.medium.com/youre-doing-it-wrong-notes-on-criticism-and-technology-hype-18b08b4307e5> [perma.cc/WN6G-VA5N].

²⁵⁰ In an amusing example, well-known crusader Tristan Harris (Silicon Valley programmer turned anti-AI evangelist) claims in his Netflix special "The Social Dilemma" that no one got upset at the advent of the bicycle. Yet, research shows that moral panic did accompany the introduction of bicycles: people were afraid for the rapidity of movement, for changing relationships between pedestrians and public space, and for women developing the condition "bicycle face," as a result of exerting effort to remain upright. *Ibid.* See also Joseph Stromberg, "Bicycle face: a 19th-century health problem made up to scare women away from biking" (24 March 2015), online: *Vox* <www.vox.com/2014/7/8/5880931/the-19th-century-health-scare-that-told-women-to-worry-about-bicycle> [perma.cc/Q5T7-9TNU].

²⁵¹ Some scholars that have begun to develop technology law method, including Rebecca Crotoof, BJ Ard, Ryan Calo, Stephanie Ballard, Jack Balkin, and others. I engage with their accounts only so far as they point out problems with techlaw scholarship generally—but then chart a different course to repair those problems. See, e.g., Crotoof & Ard, *supra* note 18; Stephanie Ballard & Ryan Calo, "Taking Legal Futures Seriously" (2019) We Robot Conference Paper, online: *UMiami* <robots.law.miami.edu/2019/wp-content/uploads/2019/03/Calo_Taking-Futures-Seriously.pdf> [perma.cc/PH2J-LL96].

²⁵² John Perry Barlow, "The Declaration of Independence of Cyberspace" (1996), online: *EFF* <www.eff.org/cyberspace-independence> [perma.cc/7GTC-A9QN].

²⁵³ Lawrence Lessig, *Code and Other Laws of Cyberspace* (New York, NY: Basic Books, 1999).

²⁵⁴ James Grimmelmann, "Regulation by Software" (2005) 114 *Yale LJ* 1719 [Grimmelmann, "Regulation"].

the law affords, we can start to define a locus for how technology might interfere with law's affordances. This necessarily implicates the lofty ideals of the Rule of Law, which is addressed specifically in Chapter 4.

1. Lessons from Cyberlaw

I begin by offering some preliminary comments on the project of technology law scholarship generally, in order to both situate this contribution in the broader conversation and apply these scholarly approaches to AI and ML. By setting out this work, I demonstrate both how my approach is inspired by—and where it departs from—today's leading scholars in the technology law space. Building on this foundation, I seek unique methodological approaches to confront the problems posed by introducing ML to legal spaces.

a. The Saga Begins

When the Internet arrived, so did the breathless promise of how it was poised to change the world. One of the most hopeful expressions of this sentiment was penned by John Perry Barlow, erstwhile Grateful Dead lyricist, in his 1996 “A Declaration of the Independence of Cyberspace.”²⁵⁵ In his eloquent treatise, society was on the precipice of something new and unimaginable. Access to the Internet was poised to democratize society in bold new ways, bolstering a new generation of digital pioneers to take matters into their own hands. Barlow proposed that cyberspace was an “act of nature” that would “grow itself through our collective actions.”²⁵⁶ He asserted that the governments of the industrial world, those “weary giants of flesh and steel,”²⁵⁷ would have no place in the egalitarian, free-flowing new marketplace, emerging from courageous and creative electronic conversations.

Many of these ideas grew out of antiestablishment movements. The democratic potential of the new world made possible by the Internet was heralded by science fiction writers. Modern social media empires and attempts to relocate modernity to the so-called digital world are anticipated by speculative fiction and cyberpunk writers, including William Gibson's coining of “Cyberspace”²⁵⁸ and Neal Stephenson's imagined “Metaverse.”²⁵⁹ While these literary efforts contain political and dystopian aspects, they also engage meaningfully with creation of a new world order and technology's potential role in such a creation narrative. Crypto-anarchists, including Timothy C. May, wrote about the potential unmaking of the global financial system achievable through the new liberties presented by cyberspace.²⁶⁰ Feminist scholars also attached to the promise offered by the Internet: the opportunity to remake the world offered an opportunity to subvert the traditional patriarchal structures in favour of a more inclusive system, with different normative underpinnings. In her iconic “A Cyborg Manifesto,” Donna Haraway writes about the merger of humans and machines through a critical feminist lens, imagining a world that would erode boundaries between

²⁵⁵ Barlow, *supra* note 252.

²⁵⁶ *Ibid.*

²⁵⁷ *Ibid.*

²⁵⁸ William Gibson, *Neuromancer* (New York, NY: Ace Books, 1984).

²⁵⁹ Neal Stephenson, *Snow Crash* (New York, NY: Penguin Random House, 1992).

²⁶⁰ Timothy C May, “Libertaria in Cyberspace,” (1 September 1992), online: *Nakamoto Institute* <nakamotoinstitute.org/libertaria-in-cyberspace/> [perma.cc/T6TB-NV8U]; Timothy C May, “The Crypto-Anarchist Manifesto,” (1988), online: *Nakamoto Institute* <nakamotoinstitute.org/crypto-anarchist-manifesto/> [perma.cc/X5TV-8AUT].

not only men and women, but also humans and machines.²⁶¹ Sadie Plant described a new digital technoculture of women, finally able to reassess traditional distinctions between women and men, perpetuating a sexual revolution through information technology.²⁶² Sandy Stone engages similar ideas through the lens of art, culture, and transgender politics, using the transformation of her own body as an analogue for potential social changes empowered through virtual systems.²⁶³ Stone's ideas, in particular, made space for people at society's margins: she writes about disabled people using technology to express ideas and fully participate in spaces from which they had historically been excluded.²⁶⁴

Yet, even at this hopeful time, some prescient scholars began to articulate how the legal system might soon be transformed by shifting balances of information and power. Technology law scholarship is indebted to the foundational work of scholars such as Joel Reidenberg and Lawrence Lessig, who offered two foundationally different frameworks for conceptualizing the impact of a newly coded reality for legal regulation.²⁶⁵ In Reidenberg's watershed paper, "Lex Informatica: The Formulation of Information Policy Rules Through Technology," he begins the project of understanding a new form of law emerging from the heritage of *lex mercatoria*.²⁶⁶ Just as itinerant merchants in medieval Europe had once developed their own governance norms—outside of the authority of church, sovereign, or nobility—so too were customs and practices necessary to offer stability to participants in the new digital information economy. His account emphasized the impact of public governance as policy intervention and the importance of regulatory mechanisms outside the traditional law-making channels. Lessig varied Reidenberg's project in his seminal work, *Code*, in which he sketches a system where four simultaneous pressures are exerted on a central dot: law, code, markets, and norms.²⁶⁷ We, the "pathetic" dot at the epicentre of these concomitant forces, experience their collective impact as regulation. Regulation in this sense is not merely the law defining appropriate behaviours, but the manner through which imposed forces constrain individual behaviour. While the subtle impacts of norms and markets are sometimes discussed, Lessig's chief observation is that the coded structures of cyberspace often collapse into *de facto* regulation by defining the limits of possibility. Simply put, code is law. Through structural enforcement, our behaviours are encoded to follow particular paths.

These regulatory-focused analyses set the stage for the development of the field of cyberlaw. Scholars observed the design choices in Internet architectures as transformative for human behaviour and the corresponding potential for meaningful regulation. Since then, a significant proportion of technology scholarship has followed Lessig's lead, discussing structural aspects of

²⁶¹ Haraway's "A Cyborg Manifesto" was originally published in 1985, in the *Socialist Review*. Donna Haraway, "A Cyborg Manifesto: Science, Technology, and Socialist-Feminism in the late Twentieth Century" in Donna Haraway & Cary Wolfe, eds, *Manifestly Haraway* (Minneapolis, MN: University of Minnesota Press, 2016) 3 at 16.

²⁶² Sadie Plant, *Zeros and Ones: Digital Women and the New Technoculture* (London, UK: Doubleday, 1995).

²⁶³ Allucquère Rosanne Stone, "Will the Real Body Please Stand Up?" in Michael Benedikt, ed, *Cyberspace: First Steps* (Cambridge, MA: MIT Press, 1991) 81.

²⁶⁴ *Ibid* at 82.

²⁶⁵ Other early entrants to this space include David Post, David Johnson, Cass Sunstein, and Jack Balkin, although a full account of the history of cyberlaw is outside the scope of this dissertation.

²⁶⁶ Joel Reidenberg, "Lex Informatica: The Formulation of Information Policy Rules Through Technology" (1997) 76 *Tex L Rev* 553 at 554.

²⁶⁷ This model is drawn from Lessig's earlier work on legal regulation, wherein he describes four forces on humans at the centre of regulatory schemes: law, architectures, norms, and markets. In his updated model, architectures replaced by code—the architecture of cyberspace. Lawrence Lessig, "The New Chicago School" (1998) 27 *J Legal Studies* 661 at 662.

regulation. Many technology law papers offer prescriptive approaches or recommendations for law reform through statutory enactments. The subtlety in the alternative method proposed by Reidenberg, *i.e.*, that good information policy governance necessitates policy intervention at a foundational level, is sometimes underemphasized. Nonetheless, a strong body of scholarly work began to grapple with technological design, unintended consequences, and best practices for regulating emerging technology.

b. Complications & Elaborations

Technology is imbricated within systems of power, making certain outcomes almost inevitable when deploying technological analogues for human systems. Recent scholarly efforts in the technology law space have picked up on some of the earlier antiestablishment themes to point out how innovation can be used to replicate and extend power. These approaches, often proffered by feminist and/or critical race scholars, demonstrate how innovation tends to perpetuate the *status quo* through technological design choices. A recent notable work in this space is Ruha Benjamin's *Race After Technology*, which deals with how structural decisions can exacerbate social inequality.²⁶⁸ To elaborate on how technological design can encode certain values and guide behaviour, she offers the example of public benches where the seating area is interrupted with arm rests, making it impossible to lay down.²⁶⁹ This familiar example resonates, as many can sympathize with the experience of being desperate for a nap at the airport, only to have any chance of comfortably stretching out impeded by those pesky arm rests. The design subtly enforces a set of behaviours about what it is appropriate to use the technology to do. Benjamin calls this "hostile architecture," and points out that it can range from overtly hostile to less obvious, like a strangely shaped, overly artistic bench that is uncomfortable to sit on, thereby preventing you from staying on the bench for very long.²⁷⁰ Benjamin then goes on to analogize such technological codes to the codes enforced by Jim Crow laws: a system of laws that mandated racial segregation between white and black citizens in the United States, realized by diverse means like legal, social, and building codes.²⁷¹ Only some of these enactments were laws in the traditional sense; others simply imposed constraints on society through encouraging the development of certain norms or behaviours.

Safiya Umoja Noble's book, *Algorithms of Oppression*, also emphasizes the coded aspects that contribute to ongoing racial discrimination in society. Her analysis began with the observation that searching for issues related to young African American girls using the prominent search website Google generated inappropriate and alarming results.²⁷² Searching for topics relevant to her young stepdaughter and nieces, Noble found that Google's first result for the keyword search "black girls" was pornographic.²⁷³ This observation began a deeper evaluation of how Google's search results contribute to broader social narratives about marginalized subgroups, especially in light of how innocuous search terms generated overtly sexualized results, calling into question the majority's treatment of people existing outside the mainstream.²⁷⁴ In her text, Noble demonstrates that

²⁶⁸ Ruha Benjamin, *Race After Technology: Abolitionist Tools for the New Jim Code* (Cambridge, UK: Polity Press, 2019).

²⁶⁹ *Ibid* at 90.

²⁷⁰ *Ibid* at 90-91.

²⁷¹ *Ibid*.

²⁷² Noble, *supra* note 141.

²⁷³ *Ibid* at 3.

²⁷⁴ *Ibid* at 71.

algorithmic determinations, often thought of as being benign or neutral, can amplify discriminatory histories and encode inequality as automated systems begin to control social and political decisions.²⁷⁵ These equality-seeking approaches add additional texture and complexity to the lessons that cyberlaw might offer to emergent robotics and AI.

c. Applying Cyberlaw's Lessons to Robotics & AI

In the 25 years since Barlow's proclamation, governments and private companies alike have found countless ways to constrain the boundless freedom the Internet once promised. And, more problematically, information shared in that promising new landscape has been commodified in a way few might have anticipated. Informational privacy has been relinquished in the name of social cohesion.²⁷⁶ The most intimate details of our lives are contained in online profiles and traded in an online marketplace for the sake of simplified sign-ins and targeted advertising. Privacy, some have argued, is essentially over.²⁷⁷ All of these data points have equipped ML algorithms, described in Chapter 1, with predictive capabilities beyond 1996's wildest imagination. We have become—as Frank Pasquale describes—a black box society: one that inputs data into opaque online algorithms that control money and power, producing outputs by unknown means, obscuring the machinations that occur for A to become B.²⁷⁸ This datafication of society has embodied many decision-making capabilities in technological artifacts: both physical robots and background algorithms, each of which obscure decision-making processes in different ways.

As this transition occurred, scholars began to identify that the lessons offered by cyberlaw might need to be adapted from Internet-focused analyses to encompass robotics and AI. Ryan Calo's important contribution, "Robotics and Lessons from Cyberlaw," began to map how legal scholars interested in regulation of robotics might learn from academic treatment of the early Internet.²⁷⁹ Through his analysis, Calo encourages us to identify the essential features of a technology to inform how it should interact with law and policy. Calo argues robots have three distinctive qualities: social valence, which permits anthropomorphization; emergence, which is characterized by adaptive behaviour; and embodiment, which refers to how robots act physically in the world. These essential features of robotics insert disruptive forces into traditional legal spheres. In a concluding section, Calo reflects on what cyberlaw might teach for future regulatory efforts.²⁸⁰ Here, he notes that transformative technologies like the Internet and robotics matter because they "... [change] the range of human experiences in ways that undermine the balance the law hopes to strike."²⁸¹ In Calo's example of social valence, he describes the challenges of distinguishing a robot's liminal status as either agent or object.²⁸² If contemporary psychology struggles to parse the distinction, it

²⁷⁵ Noble, *supra* note 141 at 71.

²⁷⁶ In 2010, Facebook founder Mark Zuckerberg famously said that privacy was no longer a social norm, because people had become more comfortable with sharing things online. See Bobbie Johnson, "Privacy is no longer a social norm," (11 January 2010), online: *The Guardian* <theguardian.com/technology/2010/jan/11/facebook-privacy> [perma.cc/U36X-92YV].

²⁷⁷ In 1999, Sun Microsystems founder Scott McNealy famously said about privacy: "You have zero privacy anyway. Get over it." And that was long before Facebook, Instagram, SnapChat, and its oh-so-many contemporaries. See Polly Sprenger, "Sun on Privacy: 'Get Over It'" (26 January 1999), online: *WIRED* <archive.wired.com/politics/law/news/1999/01/17538> [perma.cc/Y5HG-ANF3].

²⁷⁸ Pasquale, *Black Box*, *supra* note 246.

²⁷⁹ Calo, "Lessons of Cyberlaw," *supra* note 16.

²⁸⁰ *Ibid* at 558.

²⁸¹ *Ibid*.

²⁸² *Ibid* at 546.

is likely that similar struggles could appear in tort, criminal, or other areas of law.²⁸³ As power adheres in technological artifacts, the potential scope of corresponding abuses of power is broadened, which, as I will later argue, necessarily implicates the Rule of Law. And, as AI and ML enable a host of new behaviours, both technological and human, such new unpredictability might be at odds with the long-established structures of the common law. In an example furnished from tort law, Calo describes hospital protocols of ensuring meetings between doctors and patients before surgery, designed to eliminate standard of care issues arising from the nature and timing of doctor-patient interactions.²⁸⁴ Neither patients nor hospital administrators are concerned with meeting physical tools used to perform the surgery: no one asks for a meet-and-greet with the scalpel; yet, this might change as surgeons increasingly delegate responsibility for performing surgery to robotic surgery assistants.²⁸⁵

Calo's analysis is seeded with references to technology as being fundamentally disruptive. He acknowledges the Internet's potential for transformative change as being generated by the scope and scale of the communication it engenders, thereby creating a qualitative distinction in the world's connectivity.²⁸⁶ He discusses the datafication of technological artifacts as embodying information in a way that disrupts a longstanding distinction between informing and acting.²⁸⁷ This account of disruption asks us to look for the spheres where legal normativity might be transformed by the different, or emergent, behaviours that the technology now makes possible. Lyria Bennett Moses observes something similar, pointing out that it is technology's transformative capabilities that require asking the question of how to treat the activity under existing legal rules, or "whether it ought to be permitted, prohibited, required, or encouraged."²⁸⁸ She describes four types of legal problems that arise because of technological change: the need for new rules to regulate new forms of conduct; the need to clarify rules that have become uncertain; the need for the scope of rules to be modified; and a justification for those rules that persist.²⁸⁹

Yet, the lessons that should be drawn from cyberlaw's development to best inoculate against the potential threats of robots and AI remain up for debate. Many have likened robotics and AI to the Internet in terms of its disruptive potential: all share the common ability to upend traditional norms by virtue of the new possibilities generated. Kieran Tranter argues against the "disruption" framing, noting that it "does not presuppose a past; nor does it really see the future. It is anxiously stuck in the present; concerned with what the future holds yet ignorant of its past."²⁹⁰ Using the helpful example of autonomous vehicles, or self-driving cars, Tranter argues that characterizing innovation in terms of disruption negates law's ongoing constitutive nature, wherein laws pertaining to cars are always in some state of flux.²⁹¹ He writes:

²⁸³ Calo, "Lessons of Cyberlaw," *supra* note 16 at 546.

²⁸⁴ *Ibid* at 547.

²⁸⁵ *Ibid*.

²⁸⁶ *Ibid* at 520.

²⁸⁷ *Ibid* at 536.

²⁸⁸ Lyria Bennett Moses, "Why Have a Theory of Law and Technological Change?" (2007) 8 *Minn JL Sci & Tech* 589 at 594-595. She goes on to note: "[t]hese questions arise (with increasing urgency) when a new technological development is conceived, developed, and diffused into society."

²⁸⁹ *Ibid* at 595.

²⁹⁰ Kieran Tranter, "Disrupting Technology Disrupting Law" (2017) *Law, Culture & the Humanities* at 8 [Tranter, "Disrupting"].

²⁹¹ *Ibid* at 7.

The presentism of technology disrupting law leads directly to an essential contradiction; an emphasis on the need for legal change yet a continual affirmation of law as it has been known; a desire to manage disruption using the forms and tools of modern law.²⁹²

This contradiction points to an issue with Calo's exceptionalism approach: as Tim Wu points out, "[exceptionalism] depends on *what* you think it is an exception to."²⁹³ Meg Leta Jones argues that true technological exceptionalism would create analogous "linear patterns" across societies when new "exceptional" technologies are introduced—something that can be easily demonstrated to be false.²⁹⁴ No linear pattern can be identified for the introduction of the same technologies in different cultures or time periods; instead, the defining characteristics of analyzing technological adoption are "variation and messiness."²⁹⁵ However, this argument suffers from the cultural distinctions, including differences in legal systems, that characterize different societies, eliminating the possibility of a meaningful control group. To resolve this conundrum, Jack Balkin asks us to look instead to the features of social life that technology makes newly salient.²⁹⁶ Here, Balkin acknowledges the social relations at work in human interactions with technology. Power dynamics are at the heart of such observations, as the implementation of new tools threatens to recalibrate the allocation of power within our social systems. Even our use of the term "technology" is usually shorthand for complex interrelationships—for first order human interactions with new inventions and second order interactions with other people using those inventions.²⁹⁷ Acknowledging these tensions, Balkin writes:

We might say, then, that the most important lesson of cyberlaw for robotics is the need to attend to the relationships between affordance and imagination, between tools and relations of power, between technological substrate and social use. The characteristics of robotics that currently draw our attention, that disturb our legal complacency, that create puzzles for law and opportunities for social innovation, will not be the last puzzles and opportunities we face.²⁹⁸

Balkin suggests that the emergent behaviour of robotics and AI systems might lead to ML systems that behave unpredictably or subvert human expectations. Riffing on Lessig's famous adage, he contends: "[c]ode is lawless."²⁹⁹ Still, Balkin's analysis remains focused on how to appropriately assign liability for legal harms arising from emergent behaviour. Missing, in my view, is a full account of the emergent difficulties for the project of regulating such lawless code—and the downstream consequences for the entire legal system if AI and ML systems begin to subvert human expectations about the operation of the Rule of Law itself. Automated decision-making threatens to bake in undesirable bias and discrimination, allow corporate influence to seep into

²⁹² Tranter, "Disrupting," *supra* note 290 at 8.

²⁹³ This view, that technology's "essential qualities" is what drives corresponding conversations about the law and policy around them, is also shared by Lawrence Lessig, David Post, and others. Meg Leta Jones, "Does Technology Drive Law? The Dilemma of Technological Exceptionalism in Cyberlaw" (2018) 2 *Journal of Law, Technology & Policy* 102 at 107 [Jones, "Dilemma"].

²⁹⁴ *Ibid* at 105.

²⁹⁵ *Ibid*.

²⁹⁶ Balkin, *supra* note 17 at 46.

²⁹⁷ *Ibid* at 49.

²⁹⁸ *Ibid* at 48.

²⁹⁹ *Ibid* at 52.

public decision-making forums, and supersede human authority by virtue of the dominance generated through technological design. Similarly, Tranter's discomfort with the "disruption frame" might be managed through better acknowledgement of law's ongoing narrative structure, aiming to constitute reality as much as it aims to regulate it. Scholars in science and technology studies have also attempted to grapple with these issues, generating meaningful insights for legal scholars.

2. Lessons From STS

Insights from science and technology studies (STS) offer another framework for understanding the intersection between technology and law. Through an STS lens, interdisciplinary insights further complicate the regulatory modes of legal thinking that dominate most cyberlaw scholarship. In this section, I present three common academic understandings of technology that impede thoughtful consideration. We can understand technology as deterministic: generating a strong force capable of mandating how we will behave in any technologically mediated environment. We can understand technology as neutral: as simply offering up modes of behaviour that are inherently neither bad nor good. Or, finally, we can understand technology as presenting a solution: solving pesky problems that sometimes arise in the business of being human.

a. Technology as Prime Mover

Does technology drive history? Cultural anthropologists, sociologists, and philosophers of science have all offered stories that describe technology as society's prime governing force.³⁰⁰ This view, often called technological determinism, is at the heart of accounts that posit an inextricable link between social evolution and the technologies that bring about said evolution.³⁰¹ Technology is therefore an agent of change: shaping human behaviour and bringing corresponding changes to social institutions. While determinism in other contexts might connote predetermination, or a causal relationship between two phenomena, technological determinism offers a more nuanced explanation: the outcome may not be predetermined due to the complexity of the factors at play.³⁰² Different theorists have advocated for both harder and softer versions of determinism, but most agree that social context is an important consideration.³⁰³ Technological determinism is often contrasted against social constructivism, a theory which proposes that technology does not shape human action, but instead, that human action shapes technology.³⁰⁴ This competing approach, the social construction of technology (SCOT), focuses on how technology is embedded within social

³⁰⁰ Merritt Roe Smith, "Technological Determinism in American Culture" in Merritt Roe Smith & Leo Marx, *Does Technology Drive History? The Dilemma of Technological Determinism* (Cambridge, MA: The MIT Press, 1994) 1 at 2.

³⁰¹ Theofanis Papageorgiou & Panayotis G Michaelides, "Schumpeter and Veblen on technological determinism, individualism and institutions" (2016) 23 *The European Journal of the History of Economic Thought* 1 at 8.

³⁰² For example, a straight causal relationship between technology and outcome may be further influenced by path dependence, or multiple equilibria embedding some type of determinism. *Ibid.*

³⁰³ Allan Dafoe, "On Technological Determinism: A Typology, Scope Conditions, and a Mechanism" (2015) 40 *Science, Technology, & Human Values* 1047 at 1050.

³⁰⁴ Thomas P Hughes, "The Evolution of Large Technological Systems" in Wiebe E Bijker, Thomas P Hughes & Trevor Pinch, eds, *The Social Construction of Technological Systems: New Directions in the Sociology and History of Technology* (Cambridge, MA: The MIT Press, 1987) 51 at 51-54.

context.³⁰⁵ In this account, it is the ongoing social construction of society that determines which technologies will be propagated, and not the other way around.³⁰⁶

The term “technological determinism” is usually attributed to American economist Thorstein Veblen in the 1920s,³⁰⁷ although other theorists expressed similar ideologies. For example, Karl Marx is often associated with technological determinism, with examples like his famous statement from *The Philosophy of Poverty*: “[t]he hand-mill gives you society with the feudal lord; the steam-mill society with the industrial capitalist.”³⁰⁸ Marx’s emphasis on forces of production introduces a unifying concept to describe how the tools, instruments, and machines available to humans—coupled with their expenditures of energy and labour—are intrinsically “involved in the active effort of [how they] change material reality to suit their needs”.³⁰⁹ Marx’s description promotes a materialist understanding of history where social change is bound up in forces of production: making the changes that occur in any particular time period necessarily tied to that era’s chief industry or methods of exchange. Importantly, in Marx’s conception, the nature of emerging technologies themselves provide new structure to human needs. Similarly, in a classical account of technological determinism, Robert Heilbroner investigates the impact of technology on “determining the nature of the socioeconomic order.”³¹⁰ Taking an empirical approach, he attempts to understand first why technology evolves in its particular evolutionary order, and second, how specific forces of production “affect the superstructure of social relationships.”³¹¹ Technological determinism, he concludes, is particularly “a problem of high capitalism and low socialism”—when the influence of technology is pervasively felt, but regulatory efforts are meagre at best.³¹²

This description of the gap between technology’s influence and the corresponding efforts to regulate that technology will sound especially familiar to those concerned with Internet regulation. It leads to what some have described as a pacing problem: technological innovation is introduced quickly, outpacing regulation or legislation.³¹³ This overly linear perspective portrays law as forever playing catchup with technological innovation. From this view, an assembly line runs from Silicon Valley directly to the consumer, with regulators being asked to walk innovation back long after it has already transformed daily life. This approach is inherently deterministic: technology defines the path we take. Jones takes issue with the pacing problem: if we accept that legal regulators, as well as technology law scholars, are forever chasing the latest technologies with attempted regulatory efforts, we also accept they have a degree of irrelevance.³¹⁴ A better approach, she

³⁰⁵ Jones, “Dilemma,” *supra* note 293 at 110.

³⁰⁶ Dafoe, *supra* note 303 at 1060.

³⁰⁷ Papageorgiou & Michaelides, *supra* note 301 at 4.

³⁰⁸ Karl Marx, *The Poverty of Philosophy*, cited in Robert L Heilbroner, “Do Machines Make History?” (1967) 8 *Technology and Culture* 335 at 335. The excerpted passage reads in full, “In acquiring new productive forces men change their mode of production; and in changing their mode of production, in changing the way of earning their living, they change all their social relations. The hand-mill gives you society with the feudal lord; the steam-mill, society with the industrial capitalist.”

³⁰⁹ Langdon Winner, *Autonomous Technology: Technics-out-of-Control as a Theme in Political Thought* (Cambridge, MA: The MIT Press, 1977) at 78.

³¹⁰ Heilbroner, *supra* note 308 at 335.

³¹¹ *Ibid* at 336.

³¹² *Ibid* at 345.

³¹³ Jones, “Dilemma,” *supra* note 293 at 105.

³¹⁴ *Ibid* at 108.

suggests, asks for a legal construction of technology, which borrows from SCOT to understand regulation of technology within a web of human relations. She describes as follows:

The legal construction of technology focuses on law as cultural corner of societies with its own customs and rituals, players and roles, institutions and relationships, and rules and power—and how this cultural corner makes sense of a technology, technological system, or technological concept.³¹⁵

Such an approach is preferred by many STS scholars, who see SCOT as better able to explain more nuanced, co-constructive role between humans and technology in the process of shaping society.³¹⁶ Taking a slightly different tack, Langdon Winner refines the concept as a more subtle version of determinism: technological somnambulism, which permits us to “so willingly sleepwalk through the process of reconstituting the conditions of human existence.”³¹⁷ From this view, it is not so much that the “invisible hand of technology magically removes all other social outcomes or possibilities, or that ‘it’ somehow punishes those who do not respond accordingly with heavy costs.”³¹⁸ Instead, our social uses of technology seem to reconstitute our perceptions of the world, making it so that we will have difficulty looking to other possibilities.³¹⁹ The availability of AI and ML solutions for legal problems has already transformed the process of legal research: most would not choose to manually sort through physical case reporters in lieu of high-powered search algorithms on popular case aggregation services like Westlaw or QuickLaw. Yet, some reported decisions never find their way online, languishing only in the fine print of a physical case reporter, deep in the stacks of books in a law professor’s office. Similarly, the availability of electronic discovery methods allows keyword searches to organize huge volumes of evidence produced for litigation to be sorted into tranches, with the supervising lawyer simply doing spot checks to confirm that the produced materials fit the required description and seem relevant. This is considered adequate to comply with a lawyer’s professional responsibility to produce all relevant documents in a discovery process.³²⁰ Yet, what is the likelihood that some smoking gun piece of evidence remains hidden deep in the morass of documentation, unencumbered by one of the keywords identified? As these types of technologies reconstitute our legal reality, we become more likely to sleepwalk towards a world where we trust technology’s outcomes, setting aside other possibilities that may have remained at the margins.

Some regulators see the breakneck pace set by the pacing problem as a problem that can be solved by drafting legal protections as technologically neutral, thereby enabling law to stay one step ahead of technological change. As I discuss in the next section, by conceptualizing technology as a neutral tool, understood through its impacts and regulated at a higher, value-driven level, perhaps the issues of outdated regulations can be avoided.

³¹⁵ Jones, “Dilemma,” *supra* note 293 at 133.

³¹⁶ *Ibid.* For example, see Bruno Latour’s work on actor-network theory, further elaborated in the next section.

³¹⁷ Langdon Winner, “Technologies as Forms of Life” in David Kaplan, ed, *Philosophy of Technology* (2004) 103 at 107.

³¹⁸ Kerr & Szilagyi, *supra* note 42 at 359.

³¹⁹ *Ibid* at 360.

³²⁰ Amy Salazyn, “A Taxonomy for Lawyer Technological Competence” (18 December 2020), online: *SLAW* <www.slaw.ca/2020/12/18/a-taxonomy-for-lawyer-technological-competence/> [perma.cc/PH8P-DFXF].

b. Technology as Neutral

Technology's presentation as a neutral tool is often emphasized by technologists and designers, many of whom tend to eschew responsibility for ethical frames or social uses of technology. Guns don't kill people, or so goes the old adage: people kill people. However, guns remain a technological innovation designed for the purpose of killing people as efficiently as possible. Such understanding of the technology's purpose is often discounted in efforts to portray any technological advancement as neutral. Technological neutrality is a common theme in both STS literature and law, as STS scholars attempt to add nuance to the notion, and legal regulators aim to create sufficiently broad regulations to operate despite changing technological contexts.

i. Technological Neutrality in STS Literature

Technology's neutrality is often approached through the nuclear weapons problem. This concept is emphasized by Langdon Winner's discussion of inherently political technologies.³²¹ Winner insists that creating and implementing technological innovations is inextricably linked to expressions of power and authority.³²² In a series of illustrative examples, he demonstrates that deploying seemingly neutral new tools can yield distinctly political outcomes: low-hanging overpasses on Long Island serve the racist, classist purposes of their creator, preventing those who travel by bus from reaching upper-class hotspot Jones Beach.³²³ Poorly-conceived pneumatic moulding machines, which produced poorer quality product but required fewer workers, were used to bust unions during a dispute between a manufacturing tycoon and an iron moulding union.³²⁴ Mechanical tomato harvesting machines (combines) not only reconfigured tomato farmers as larger corporate entities able to afford them, but also transformed the tomato itself to make it easier to harvest.³²⁵

Bruno Latour takes this concept one step further. Although one could ignore what the technology is saying, design mandates a type of "script"—an item's technological nature can suggest that it be used in a particular way to better accord with its purpose.³²⁶ Through design choices, some actions are encouraged, while others are discouraged. This nuanced theory, the actor-network approach, bridges determinism and constructivism.³²⁷ Because human action is mediated by those

³²¹ Langdon Winner, "Do Artifacts Have Politics?" in Langdon Winner, ed, *The Whale and the Reactor: A Search for Limits in an Age of High Technology* (Chicago: University of Chicago Press, 1986) 19 at 22 [Winner, "Artifacts"].

³²² *Ibid.*

³²³ These bridges are alleged to reflect the racist and classist predispositions of their architect Robert Moses. By preventing public transit from being able to travel on the parkway, he was able to *de facto* prevent anyone of a lower socioeconomic class from reaching his acclaimed public park, Jones Beach. *Ibid* at 23. Some contest the accuracy of this story, arguing that the low bridges were due to project cost constraints, and that those without cars could still get to Jones Beach using trains or buses. Others, including aides to Moses, affirm the account, noting that he indeed held racist views, and that the Moses parkway had substantially lower clearances than other parkways in the area constructed around the same time. See Glenn Kessler, "Fact Checker: Robert Moses and the saga of the racist parkway bridges" (21 November 2021), online: *Washington Post* <www.washingtonpost.com/politics/2021/11/10/robert-moses-saga-racist-parkway-bridges/> [perma.cc/VV3E-QNSU].

³²⁴ Winner, "Artifacts," *supra* note 321 at 24.

³²⁵ To withstand the agitation and rough motion imposed by the combine, tomatoes were harvested earlier in the season, when they were smaller, less ripe, less juicy, and less delicious.³²⁵ *Ibid* at 25.

³²⁶ Peter-Paul Verbeek, "Materializing Morality: Design Ethics and Technological Mediation" (2006) 31 *Science, Technology, & Human Values* 361 at 362 [Verbeek, "Materializing Morality"].

³²⁷ Bruno Latour, "Where Are the Missing Masses? The Sociology of a Few Mundane Artifacts" in Deborah G Johnson & Jameson M Wetmore, eds, *Technology and Society: Building Our Sociotechnical Future* (Cambridge, MA: The MIT Press, 2009)

technologies they use, human behaviour is shaped by materials selected to perform tasks. The material world pushes back against human action, owing to the design of the objects themselves.³²⁸ When faced with decisions about replacing people with technological analogues, Latour's playful investigation poses an "age-old Mumfordian choice."³²⁹ Do you accept the unreliability of people inherent in human nature, or do you substitute a technological solution for the unreliable people?

Latour's example of the Berliner key brings this argument home. In multi-family dwellings in Berlin, residents needed to use their keys to both lock and unlock the front door of their buildings. Contrasted with the simplicity of doors that slam shut and lock behind you, nothing requires a forgetful resident to lock the door as she leaves. Although she has to unlock the door to access to her own home, the process does not necessarily require that she then lock the door to ensure that no one else will be able to inappropriately enter the dwelling. Latour explains, "[h]ow can you force [her] to lock it? A normal key endows you with the *competence* of opening the door—it proves you are *persona grata*—but nothing in it entails the *performance* of actually using the key again once you have opened the door and closed it behind you."³³⁰ Non-technological means of solving this problem are complicated and fraught with their own challenges: you could put up a sign, but that might not influence anyone; you could position a police officer at each dwelling, but that seems draconian—and unlikely. No, instead, a technological solution is achieved, which Latour defines as a "program of action."³³¹ The Berliner key is designed such that, once inserted into the lock to unlock the door, it cannot be removed in the usual fashion. Instead, the key needs to be pushed through the lock to the other side, and then used to relock the door, before removal is possible. The key itself says what the sign might: you need to lock the door behind you, because this use is proscribed by the functioning of the key itself. Latour points out that you could give up and leave the key behind, but you would then be sacrificing the competence afforded you as resident: you would never again be able to get in or out on your own accord.³³² Of course, there is a cost to transitioning away from reliance on either morality or signage: instead, you have to replace most of the locks in Berlin.

Verbeek offers corresponding observations about technological behaviour mediation on an individual scale, invoking the microwave.³³³ The ease of reheating an individual meal means families gather at mealtime less often than before.³³⁴ However, this observation has a caveat: technology can mediate behavior without determining it.³³⁵ Families can still choose to eat together, notwithstanding the appeal of the microwave's electromagnetic hum. The distinction between these experiences depends on the influence the product has on the individual: these can

225 at 229 [Latour, "Missing Masses"]. See also Bruno Latour, "On actor-network theory: A few clarifications" (1996) 47 *Soziale Welt* 369 at 379.

³²⁸ Verbeek, *supra* note 326 at 366.

³²⁹ Latour, "Missing Masses," *supra* note 327 at 230.

³³⁰ *Ibid* at 252.

³³¹ *Ibid*.

³³² *Ibid*.

³³³ Nynke Tromp, Paul Hekkert & Peter-Paul Verbeek, "Design for Socially Responsible Behaviour: A Classification of Influence Based on Intended User Experience" (2011) 27 *Design Issues* 3 at 5. See also Peter-Paul Verbeek, *What Things Do: Philosophical Reflections on Technology, Agency, and Design* (University Park, PA: The Pennsylvania State University Press, 2005).

³³⁴ *Ibid* at 5.

³³⁵ *Ibid*.

be categorized as coercive, persuasive, seductive, or decisive.³³⁶ The families who stop eating together following the introduction of the microwave found it to be persuasive, while those who remain uninfluenced merely find it seductive.³³⁷ Understanding physical artifacts as entering into a dialogue with their users, by virtue of incorporated design choices, aids in making sense of technology within a cultural, human framework.

Complex perspectives like those of Latour or Verbeek are often relegated to the STS realm, while those participating in actual engineering or technical design processes eschew nuance in favour of a more traditional take on technological innovation as a neutral vessel. Technologically neutral perspectives can sometimes hamper nuanced design or transdisciplinary efforts to integrate engineering efforts with ethics and law. Seeing technology only as a tool to be used to solve a problem, irrespective of the larger social impacts, might have downstream consequences in design choices on a larger scale. On Twitter, for example, controversy erupted when it was suggested that the social media giant's image cropping algorithm was "racist," since it always seemed to prioritize the image of a light-skinned person over a darker-skinned person.³³⁸ In fact, such images were being cropped according to the contrast within the image. If an image was too large to fit within the typical square thumbnail images that populate Twitter's main feed, it was cropped in a way that prioritized the lighter aspects of the image over the darker aspects. This might work well for highlighting a sunny afternoon in the park as opposed to the shady spot under a tree. It does not, however, work well when a light-skinned person is pictured alongside a darker-skinned person. The algorithm consistently cropped such images so that the light-skinned person was featured in the main Twitter thumbnail; only when you click on the image to expand it will you discover that a darker-skinned person was also originally present in the image.³³⁹

Some Twitter users jumped to the algorithm's defence, explaining that it was not racism, just math.³⁴⁰ As one such Twitter user explained:

Guys, the algorithm just prioritizes contrast. All you have to do is lower the contrast of the white woman and increase the contrast of the black man to see this. This isn't racism, he's a black man standing in front of a dark backdrop causing him to blend in more.³⁴¹

Thinking of the mathematical aspects of the cropping algorithm as neutral, or merely prioritizing the high-contrast images, negates the complexity of the social relations in the world in which such an ML algorithm is deployed. The "it's just math" apologists neglect the broader narrative of its

³³⁶ Tromp et al, *supra* note 333 at 12.

³³⁷ *Ibid.*

³³⁸ Andrew Paul, "Remember Twitter's racist photo-cropping tool? It's still a thing." (9 December 2020), online: *InputMag* <www.inputmag.com/culture/remember-twitters-racist-photo-cropping-tool-its-still-a-thing> [perma.cc/JPW4-8KUA].

³³⁹ This was on full display during the run-off election for the U.S. Senate in the state of Georgia. After the debate between Democratic candidate Raphael Warnock, who is black, and Republican candidate Kelly Loeffler, who is white, Twitter users began noticing that a photo of the two of them side by side (too large to fit in the typical Twitter image square), always prioritized Ms. Loeffler—even if the image were flipped or reversed. *Ibid.*

³⁴⁰ @IllustriousVar, "Guys, the algorithm just prioritizes contrast. All you have to do is lower the contrast of the white woman and increase the contrast of the black man to see this. This isn't racism, he's a black man standing in front of a dark backdrop causing him to blend in more." (7 December 2021 at 12:38), online: *Twitter* <twitter.com/IllustriousVar/status/133583594944877730>.

³⁴¹ *Ibid.*

impact: the prioritization of the image of a white person in a society already plagued by inequality. One might imagine such apologists making the same type of argument about a hammer; it is just a tool—one which might be used to build something, or be used to attack someone, as though it is a weapon.³⁴² While this is true on its face, legal frameworks have been put in place to circumscribe the appropriate uses of hammers. It is not permitted to use a hammer as a deadly weapon: society has already done the work of thinking through the types of limitations you might want to impose on a hammer's use. For emerging technologies, like ML-powered Twitter image-cropping tools, this work has not yet been completed.³⁴³

In STS accounts, moral frameworks are mediated by technology: designing for ethics is not about getting rid of power; instead, it entails entering into meaningful dialogue with power. And if ethics asks questions about how to act, technology might give us physical answers.³⁴⁴ Inquiries into ethics are epistemologically close to inquiries into law. Taken holistically, this co-creation conception leaves some space to reconcile the narrative aspect of making technology with the narrative aspect of creating law. Yet, technologically neutral approaches in legal statutes or regulation tend to avoid specificity in order to maximize their longevity and scope of application.

ii. Technological Neutrality in the Law

Modern legal interpretative principles have espoused technological neutrality as fundamental to producing statutory schemes that are able to withstand the rapidity of technological change. Using a doctrinal legal example, consider the postal acceptance rule from contract law. The traditional position established legal norms for delivery of contracts via mail: contracts were considered signed at the moment when the letter was placed into the mailbox, relinquished into the care of the postal system.³⁴⁵ This was easier than trying to assess when the recipient had received the signed document. Instead, the postal service was understood to be the agent for both parties: by delivering the acceptance to the post office, it was understood to be complete and binding. This convention also prevented instances of fraud or efforts to interfere with the delivery of the mail. Yet, the availability of new methods of communication transformed this longstanding rule: technological means enabled instantaneous communication, meaning the relevant documents were not in the custody of the postal system for a variable period of time.³⁴⁶ To understand what type of situation

³⁴² One need not imagine too hard: this exact scenario happened to me when I presented an early version of my doctoral research to an interdisciplinary research group. After I presented the same Twitter example, an engineering master's student raised his hand, saying: "Well, this is interesting, but at the end of the day, it's just a tool. It's like a hammer! You can use a hammer to build something, but you can also use it as a weapon."

³⁴³ Note that as of July 2022, Twitter has attempted to solve this problem by simply prioritizing the centre of an image when cropping. So, if you were to put a black person at one end of a wide landscape image, and a white person at the other end, separated by white space, the algorithm will now prioritize the white space. This can be easily seen in some early articles about Twitter's image-cropping algorithm, which simply linked to the relevant Twitter posts rather than taking screenshots. As of July 2022, the once-impugned images now simply show white squares. See Alex Hern, "Twitter apologises for 'racist' image-cropping algorithm" (21 September 2020), online: *The Guardian* <www.theguardian.com/technology/2020/sep/21/twitter-apologises-for-racist-image-cropping-algorithm> [perma.cc/JPW4-8KUA].

³⁴⁴ Verbeek, "Materializing Morality," *supra* note 326 at 369.

³⁴⁵ *The Household Fire & Carriage Accident Insurance Company v Grant*, [1879] LR 4 Ex D 216.

³⁴⁶ *Brinkibon Ltd v Stahag Stahl und Stahlwarenhandelsgesellschaft mbH* [1983] 2 AC 34. In this case, the acceptance of an offer was sent by telex. The inquiry before the court was whether communicating via telex should be characterized as instantaneous communication (where the contract is formed at the moment it reached the ear of the offeror) or non-instantaneous communication (where the contract is formed at the moment it entered the custody of the delivery service, *i.e.*, the postal acceptance rule). In *Brinkibon*, the court likened telex to instantaneous communication: a telex machine requires a recipient to turn it on and be present to receive transmission, affirming the earlier decision of *Entores v Miles Far East Co* [1955] 2 QB 327.

was created by the technology required asking whether the new technology enabled instantaneous or non-instantaneous communication. Statutory schemes have since stepped in to specifically indicate what happens when contracts are communicated by electronic means.³⁴⁷

The Supreme Court of Canada (SCC) regularly appeals to technological neutrality, reasoning by analogy in situations when technology has transformed the operation or applicability of some legal rule. For example, in a 2012 copyright case about receipt of royalties for musical works incorporated into video games, the SCC discussed whether there was merit to distinguishing between situations where the game was distributed over the Internet *versus* situations where the game was sold in a store.³⁴⁸ Counsel for the appellant, the Entertainment Software Association, argued that downloading a video game on the Internet did not fit into the statutory provision set out in section 3(1)(f) of the *Copyright Act*, which required a game be communicated to the public by telecommunication to trigger the availability of royalties. Conversely, the respondent, the Society of Composers, Authors and Music Publishers of Canada (SOCAN), argued that downloading a game was a form of “communication to the public by telecommunication,” which meant SOCAN’s members were entitled to compensation through the requisite tariff.³⁴⁹ Relying on the principle of technological neutrality to equate Internet downloading with other sales, the SCC opined:

The principle of technological neutrality is reflected in s. 3(1) of the Act, which describes a right to produce or reproduce a work “in any material form whatever.” In our view, there is no practical difference between buying a durable copy of the work in the store, receiving a copy in the mail, or downloading an identical copy using the Internet. The Internet is simply a technological taxi that delivers a durable copy of the same work to the end user.³⁵⁰

Failing to design legal regulations as technologically neutral can lead to calls for legislators to regularly update them as technology changes. For example, Canada’s private sector statutory privacy scheme is accompanied by the *Regulations Concerning Publicly Available Information*, which carve out certain exceptions from requiring explicit consumer consent before personal information is collected, used, or disclosed.³⁵¹ The *Regulations* permit exception for five specific categories of information: (a) telephone directories; (b) business directories; (c) registries managed by public authorities (d) public records of judicial/quasi-judicial bodies; and (e) publications of magazines, books, and newsletters.³⁵² In recent years, these *Regulations* have attracted criticism for being outdated, given their explicit references to phone books or business records, and silence on emerging topics like social media. Commentators seeking modernization note that references to specific categories in (a) through (e) of the *Regulations* reveals their age and creates a technologically specific scheme, especially in light of the vast array of information that has now

³⁴⁷ In Ontario, this is governed by the *Electronic Commerce Act, 2000*, SO 2000, c 17, s 22(1), which states: “Electronic information or an electronic document is sent when it enters an information system outside the sender’s control or, if the sender and the addressee use the same information system, when it becomes capable of being retrieved and processed by the addressee.” Additional provisions deal with presumptions, time of receipt, place of business, and habitual residence.

³⁴⁸ *Entertainment Software Association v Society of Composers, Authors and Music Publishers of Canada*, 2012 SCC 34.

³⁴⁹ *Ibid* at para 127.

³⁵⁰ *Ibid* at para 5.

³⁵¹ *Regulations Concerning Publicly Available Information*, SOR/2001-7.

³⁵² *Ibid*.

become accessible in the digital age.³⁵³ Still, cries for updating the *Regulations* seem to confuse technological neutrality with modernity, since they seem to demand inclusion of new types of media enabled through technology. Teresa Scassa notes that exempting further categories, like social media, upsets the original legislative purpose of the *Regulations*, which are carefully constructed as narrow exceptions, never intended to create “a free-for-all open season on the personal information of Canadians.”³⁵⁴ The perils of publicly available information in the context of privacy violations and new AI/ML-powered technologies are taken up in Chapter 7.

Translating old common law concepts to keep pace with new technological innovations can require careful selection of metaphors to power the translation. Joshua Fairfield describes the transition of the old property tort trespass to chattels as it moves “from sheep-bothering to cybertrespass.”³⁵⁵ As we think about AI bots that are webcrawling across websites, we can employ reasoning from longstanding common law principles to furnish our thinking about how to best describe this new phenomenon in legal terms. Fairfield notes that the law of trespass offers us good intuitions about “balancing one person’s ownership with another’s casual interaction.”³⁵⁶ Similar observations about metaphor and analogical reasoning have long peppered technology law literature. Writing on cryptography, Michael Froomkin describes the function of metaphor as going beyond merely putting a particular spin on the explanation: they lead to fundamentally different definitions of the thing being characterized.³⁵⁷ While metaphors can be used as a sort of “cognitive hook” to better explicate something unfamiliar in familiar terms, uncritical use of metaphors can impoverish explanatory efforts.³⁵⁸ Ian Kerr invokes lessons from Lon Fuller to remind us that “the danger in using legal fictions varies inversely with the degree of our awareness that the words used are not literal.”³⁵⁹ Kerr relies on Neil Postman, arguing that “it is the ideology of the computer that gives rise to this ‘dominant metaphor of our age.’”³⁶⁰ Writing in 1999, he offers as example the “now commonplace computer virus.”³⁶¹ In my view, this human-machine metaphor is only further operationalized in our current context of “artificial intelligence” and “machine learning.”

c. Technology as Solutionism

Technological solutionism assumes that every problem needs a technological solution, often one that you can spend money on. Evgeny Morozov describes technological solutionism as the will to improve almost everything, which seems to involve “recasting all complex social situations as either as neatly defined problems with definite, computable solutions or as transparent and self-evident

³⁵³ ETHI, “Towards Privacy by Design: Review of the Personal Information Protection and Electronic Documents Act” (42nd Parliament, 1st Session), online: *House of Commons* <www.ourcommons.ca/DocumentViewer/en/42-1/ETHI/report-12> [perma.cc/Y5CH-PGEE].

³⁵⁴ Teresa Scassa, “Open Season on Social Media Data: ETHI’s Report on PIPEDA Reform – Part II” (22 March 2018), online: *Teresa Scassa* <www.teresascassa.ca/index.php?option=com_k2&view=item&id=273:open-season-on-social-media-data-ethis-report-on-pipeda-reform-part-ii&Itemid=80> [perma.cc/L5ZB-5GG6].

³⁵⁵ Fairfield, *Runaway*, *supra* note 50 at 56.

³⁵⁶ *Ibid.*

³⁵⁷ A Michael Froomkin, “The Metaphor is the Key: Cryptography, The Clipper Chip, and the Constitution” (1995) 143 *University of Pennsylvania Law Review* 709 at 862.

³⁵⁸ Ian Kerr, “Mind Your Metaphors: An Examination of the Inefficacy Argument as a Reason Against Regulating On-line Conduct” in Lester Pourciau, ed, *Ethics and Electronic Information in the Twenty-First Century* (West Lafayette, IN: Purdue University Press, 1999) 231 at 234 [Kerr, “Metaphors”].

³⁵⁹ *Ibid* at 235.

³⁶⁰ *Ibid.*

³⁶¹ *Ibid.*

processes that can be easily optimized—if only the right algorithms are in place!”³⁶² This “unabashedly pejorative” term is borrowed from “the world of architecture and urban planning, where it has come to refer to an unhealthy preoccupation with sexy, monumental, and narrow-minded solutions.”³⁶³ In some accounts, technological solutionism is tied closely to technological determinism, tying “technological progress tightly to social progress and ... quickly identify[ing] technological solutions to social problems.”³⁶⁴ This approach requires that the latest technological solutions be implemented in order to keep a particular space up-to-date. Law is especially vulnerable to this phenomenon, even as some argue that a lawyer’s duty of competence includes a specific duty of technological competence to use the latest tools. Kerr described the challenges of living in what Postman termed a “technopoly,” wherein value is attached to using the latest technology.³⁶⁵ This shift in values transforms how technology is viewed: rather than as a tool, it appears as an end in itself.³⁶⁶

Morozov also describes the reality of a quantified self, where individuals use electronic sensors and other tracking devices to record their own behaviours. According to one of the movement’s founders, technology journalist Gary Wolf, self-tracking has been enabled by the advent of four factors: smaller, more powerful electronic sensors; their ubiquity enabled by smartphone use; their normalization of sharing through social media; and cloud computing’s ability to merge and aggregate user data to yield better results.³⁶⁷ Quoting directly from Wolf:

We tolerate the pathologies of quantification—a dry, abstract, mechanical type of knowledge—because the results are so powerful. [...] Numbering things allows tests, comparisons, experiments. Numbers make problems less resonant emotionally but more tractable intellectually.³⁶⁸

Acknowledging this willingness for humans to quantify aspects of their lives might spell disaster for the concept of narrative imagination. As Martha Nussbaum explains, humans require something more than mere facts to be able to relate to the complex world around them.³⁶⁹ Thus, narrative steps in: enabling people to be self-reflexive or be able to imagine themselves in circumstances broader than their own. Morozov proposes that narrative imagination also entails “one’s interaction with complex sociotechnological and political systems and the ability to see one’s own role in them.”³⁷⁰ Contrasting narrative imagination against some “oxymoronic numeric imagination,” Morozov worries that merely seeking out quantifiable metrics or linear relationships between cause and effect will impoverish deeper understandings of our complicated world.³⁷¹ This same idea percolates in proposals for AI legal governance, as will be further elaborated in Chapter

³⁶² Morozov, *supra* note 39 at 5.

³⁶³ Morozov goes on to say that these shallow solutions are “the kind of stuff that wows audiences at TED Conferences,” giving overly simplistic responses to “problems that are extremely complex, fluid, and contentious.” *Ibid* at 6.

³⁶⁴ Jones, “Dilemma,” *supra* note 293 at 108.

³⁶⁵ Kerr, “Metaphors,” *supra* note 358 at 233.

³⁶⁶ *Ibid*.

³⁶⁷ Morozov, *supra* note 39 at 230.

³⁶⁸ *Ibid* at 231.

³⁶⁹ *Ibid* at 260.

³⁷⁰ *Ibid*.

³⁷¹ *Ibid*.

3. Putting the same idea another way, data scientist Cathy O’Neil writes that moral imagination is something only humans can provide.³⁷²

In Chapter 1, I discussed proposals for the personalization of law, where Casey and Niblett suggest that personalized approaches could carefully cater legal responses to the unique circumstances of each offender.³⁷³ With adequate data, such as the aspirational dataset of the quantified self aficionados, Casey and Niblett’s vision might someday be a reality. Fearing that the Luddites might stand in the way of the realization of their vision, however, they caution:

It will not be technological infeasibility or lack of data standing in the way of personalized law. It will be the inability of humans to agree on what law is designed to do.³⁷⁴

Thinking about personalized law seems to preview some of the upcoming concerns described in the subsequent chapters. What goal, specifically, might personalized law serve? Would it appropriately discharge the legal system’s obligations to take the unique circumstances of each individual offender into account during sentencing? Or, might the consequences prove disastrous, as Morozov fears, through the loss of narrative imagination. Morozov invokes Friedrich Nietzsche’s observations on the quantification phenomenon, noting Nietzsche’s critiques of an undue “information reductionism” that limited the scope of human experience to only that which might be quantified.³⁷⁵ He writes:

Most perceptively, Nietzsche understood quantifiable information might be nothing but low-hanging fruit that is easy to pick but often thwarts more ambitious, more sustained efforts at understanding. ‘Would it not be rather probable that, conversely, precisely the most superficial and external aspect of existence—what is most apparent, its skin and sensualization—would be grasped first—and might even be the only thing that allowed itself to be grasped?’ he wondered.³⁷⁶

Ultimately, if decisions are unduly data-driven, we risk becoming embedded in a never-ending feedback loop, likely to reproduce the same biases and discriminations as all those who came before us. We also risk segmenting society, alienating decision-makers from the human collective whose movements give law meaning. The Rule of Law, emphasized earlier but described in more detail in Chapter 4, might then collapse into rule *by* law. Law, then, would resemble an overtly utilitarian system, prioritizing efficiency over the pursuits of other normative ends.

3. Marching Orders for the Future of Law

Living in a technopoly also contributes to a sense that we understand advanced technologies to be authoritative. From the ongoing quantification movement flows what I term a “technological

³⁷² Cathy O’Neil, *Weapons of Math Destruction: How Big Data Increases Inequality and Threatens Democracy* (New York, NY: Broadway Books, 2016) at 204.

³⁷³ Casey & Niblett, *supra* note 37 at 336.

³⁷⁴ *Ibid* at 358.

³⁷⁵ Morozov, *supra* note 39 at 243.

³⁷⁶ *Ibid* at 244.

gloss” on all types of human problems. The potential insights that might be gleaned through data change our interactions with human information, our understanding of what behaviour a particular technology affords, and our operative normative frameworks. I begin by conceptualizing the notion of quantification, by observing certain psychological phenomena that animate concern in the human-automation interaction space. I then discuss a major theme from the cyberlaw and STS literatures, technological affordances, defining its intersections with law. Finally, I engage with the notion of normativity, particularly how it operates in law and how its operation might be destabilized by certain technological affordances.

a. Conceptualizing Quantification

Big data enables a new quantification of human life that extends the purview of technology into additional dimensions. Through big data, as unpacked in Chapter 1, ML algorithms have become able to generate significant insights into human behaviour on a population level, which leads to more advanced computing techniques that can increasingly mimic human intelligence. When automated systems operate as though they possess “intelligence,” their interactions with humans can induce certain psychological states.³⁷⁷ In the human-automation interaction literature, two main terms are used to describe the issues that arise when humans interact with automated systems. These are automation bias and automation complacency.³⁷⁸

Automation bias occurs when users rely on imperfect automated decision-making processes, which can lead to both omission errors—things that are missed—and commission errors—things that are done improperly—due to inappropriate reliance on the automated system.³⁷⁹ A related, complementary phenomenon is automation complacency, which occurs when people interacting with an automated system are forced to perform multiple tasks simultaneously, increasingly relying on the automated system instead of their own instincts.³⁸⁰ In concert, automation bias and automation complacency present overlapping challenges: they directly interface with how humans allocate their attentional resources, and cannot be easily overcome.³⁸¹ These mental states can also lead to issues of over-trust, as humans delegate increasing amounts of responsibility to automated systems, stymied in system administration by innate psychological phenomena.³⁸²

For humans interacting with algorithmic decision-makers, automation bias can amplify pre-existing social bias, as reliance increases on the technological solution. If data analytics have baked in longstanding misogyny or racism, undue trust in the system will allow those undesirable

³⁷⁷ In other work, I explore the significance of automation bias and automation complacency in the context of a highly regularized task: driving. Much of the driving task is located in procedural memory, meaning its dominant behaviours are reflexive and unconscious. When interspersed with conscious activities like managing alerts from an automated system, new increases in cognitive load can create issues at both the attention allocation and information processing levels. See Katie Szilagyi, Jason Millar, AJung Moon, Shalaleh Rismani, “Driving into the Loop: Mapping Automation Bias & Liability Issues for Advanced Driver Assistance Systems” Digital Society (forthcoming 2022).

³⁷⁸ See also Raja Parasuraman & Dietrich Manzey, “Complacency and Bias in Human Use of Automation: An Attentional Integration (2010) 53 The Journal of the Human Factors and Ergonomics Society 381.

³⁷⁹ *Ibid* at 381.

³⁸⁰ *Ibid*.

³⁸¹ *Ibid*.

³⁸² Szilagyi et al, *supra* note 377 at 6. In the context of automated driver assistance systems, we observe how manufacturers and regulators ask human operators to serve in a managerial role over the system, even though doing so contradicts intrinsic psychological traits. If accidents happen, the human is then deemed to have “misused” the system, even though they have been placed into a position that robustly induces a particular state.

characteristics to pass through unabated. And, as will be explored in Part III, technologies that are originally deployed simply to “assist” human decision-makers tend to occupy an inappropriately large space in the human field of vision, as the technological gloss offered by such machinations exerts authority. By allowing omission errors and commission errors to remain unremedied, subtle incursions are made into human dominion over decision-making, likely to increase over time with further reliance on such systems. If multiple such systems are deployed simultaneously, they can begin to obscure their origins, resulting in automation complacency as technological authority is further concentrated inside the so-called black box. To better understand such conclusions, I turn now to analyzing what different technological choices afford.

b. Understanding Affordances

Another theme can be drawn from cyberlaw literature: the idea of technological affordances and their impact on legal regulation. Some argue that the law is rooted in the technology of the written word: that the book transformed legal thinking by requiring logical structures and meaningful engagement with language.³⁸³ Scholars use the language of affordances to underscore the importance of a dialogic understanding between what technology is and the consequences flowing from its application. Mireille Hildebrandt argues that modern law—and especially the Rule of Law—is anchored in a particular understanding of words and language.³⁸⁴ Through this lens, the inclusion of legal technology in the administration of law is troubling because the operation of law itself is meaningfully tied up with the affordances of the written word in physical texts. Ambiguities of language have created a system of interpretative exercises that would be altogether abandoned through technological processing. In Hildebrandt’s estimation, the overall project of self-driving laws necessarily imposes new constraints on the operation of law itself, as particular language choices—coded in ML form—mandate specific courses of action. Such constraints will be particularly problematic in unstable legal contexts, *i.e.*, edge cases or hard cases, where the appropriate legal answer is not immediately clear.

Hildebrandt’s expansive text *Smart Technologies and the End(s) of Law* offers a comprehensive analysis of our shifting technological landscape, combining insights from a variety of academic traditions. She offers significant insight into the origin of modern legal systems, but her central contention that Internet and communications technologies destabilize the Rule of Law remains underexplained. Her core argument focuses on the proliferation of text-driven law made possible by Gutenberg’s printing press: widespread availability of the governing text brought more and more people under the ambit of law’s legitimacy, with judicial exercises of power and interpretation following in turn. What is it, then, that Hildebrandt seeks to preserve? The open publication of laws, seemingly under threat by proprietary algorithms used for more and more rote decisions? The interpretative exercise of parsing words and meanings, applications of judicial maxims, that seems to rest at the heart of judicial decision-making? Or is it something more subtle: the possibility that one could be wrong? Do we need to preserve that possibility because excising all of the interpretation from the practice of law means mere administration—rote applications of rules that may themselves be wrong? In response to reviewers, Hildebrandt offers a potential answer: that the affordances of modern law depend on the affordances of human language and their

³⁸³ Carr, *supra* note 248 at 8; Fairfield, *supra* note 50 at 157.

³⁸⁴ Hildebrandt, *End(s) of Law*, *supra* note 30 at 173-175.

corresponding proliferation.³⁸⁵ Law has traditionally mandated divisions of labour between those who interpret and apply the law and those who enact it, creating a nuanced system of checks and balances fundamental for the Rule of Law.³⁸⁶ The Rule of Law's ability to offer authority might be tied up in fundamentally human exercises of discretion and interpretation. But such commentary requires more doctrinal heavy lifting to set out a meaningful concept of law and the corresponding normativity that the Rule of Law imposes—which I undertake in Chapter 4.

Hildebrandt's perspective uses a civilian understanding of law as driven by text and codes. Though similar global principles can be found at the heart of both civilian and common law traditions, the common law is quicker to acknowledge the malleability of language in the face of interpretative exercises. Some of this is by design: civil law's heritage as a formal codex and its continuing codification often presumes the existence of a right answer to legal quandaries. Technology even had a role in early civil law's codification: the transition from scroll to codex precipitated the early codification efforts of Roman law.³⁸⁷ Hildebrandt's arguments about the force of law being amplified by Gutenberg's printing press accord with a code-based civilian worldview. While similar arguments might be levied at the common law, its oral traditions are more firmly entrenched. Even when printed material became more readily available, common law doctrine remained an oral patchwork, with statutes playing only a supporting role in conflict resolution.³⁸⁸ Written, reasoned decisions became common only after the 1750s, when the judiciary felt pressured by the emerging supremacy of Parliament and began presenting "judicial opinions in the same semi-official guise as legislation," offering a rival track of legal authority.³⁸⁹

Since this shift happened so long ago, one might argue that the statutory supremacy that emerged in this period is analogous to the alleged completeness of civilian codices. Modern law's entanglement with text might therefore start to look similar from either common or civil law's vantage points. In both systems, modern law's complexity is increasingly enmeshed with the development of digital technology and new mechanisms for presenting the written word. Scholars such as Ethan Katsh have emphasized that the shift to electronic media has already transformed modern legal paradigms, with doctrinal changes stemming from modifications to how information is moved, stored, and processed.³⁹⁰ François Ost proposed a word processing model for better conceptualizing the role of law in a transitional, networked society.³⁹¹ Florian Martin-Bariteau contends that Ost's framework might even be better updated for the twenty-first century as a wiki, acknowledging collaborative effort within a networked, distributed society.³⁹² Each of these accounts further complicates the legal paradigm in the face of emergent technology: suggesting law has long departed from the path set by Gutenberg. Concurrently, understanding the history of judicial decisions as an exercise to attempt to maintain common law's superiority in the face of developing parliamentary power underscores the narrative thread that has long been woven into

³⁸⁵ Hildebrandt, "Affordance," *supra* note 218 at 119.

³⁸⁶ *Ibid.*

³⁸⁷ Tranter, "Disrupting," *supra* note 290 at 12.

³⁸⁸ William D Popkin, *Evolution of the Judicial Opinion: Institutional and Individual Styles* (New York, NY: New York University Press, 2007) at 5. I am indebted to Jon Khan for sharing this notion with me.

³⁸⁹ *Ibid.* at 15.

³⁹⁰ M Ethan Katsh, *The Electronic Media and the Transformation of Law* (New York, NY: Oxford University Press, 1989) at 15.

³⁹¹ François Ost, "Le temps virtuel des lois postmodernes ou comment le droit se traite dans la société de l'information" in Jean Clam & Gilles Martin, eds, *Les transformations de la régulation juridique* (Paris, FR: LGDJ, 1998) at 24.

³⁹² Florian Martin-Bariteau, "The Matrix of Law: From Paper, to Word Processing, to Wiki" (2014) *Lex Electronica* 23.

legal reasoning. Interpretative exercises seek to tell stories in an explanatory, cohesive manner: a process that has been embedded in the “thinking like a lawyer” adages usually offered as the goal of legal education.³⁹³ The role of narrative, therefore, is coupled with the interpretative task.

One of law’s purposes is to offer some structure to society. While it has never had an exclusive monopoly on this function, it has also never been faced with automation as a component of how it executes its force on society. Robert Cover once remarked that legal interpretation takes place on a field of pain and death.³⁹⁴ He was writing about the violence embedded in the criminal law: the inherent complexity of sentencing that brings the blunt force of the state to bear against an accused individual. Incorporating AI and ML tools in the exercise of legal power means a possible extension of the ambit of such violence. This exercise of power, once constrained in the thrust of the lawmaker, might now be shared by for-profit institutions, doled out by inscrutable algorithms, and informed by biased, discriminatory datasets. Power’s ability to shape spaces prompts another consideration: the normativity generated by the world-building function of law and the sense-making generated by asking law to serve as an organizing force for society.

c. Generating Normativity

Like AI, the term “normativity” invites careful consideration as to its underlying meaning. This term can be invoked for different reasons in different disciplines. When Lessig refers to norms as a key component of legal regulation, he observes the world-building function that can flow from tacit agreement on how things should be. This incorporates cultural aspects of knowledge: even without explicit coding, norms can be rigorously enforced within social settings. They can also be malleable, capable of shifting depending on various factors. Consider the example of space sharing on public transportation. If you climb aboard a city bus, you typically begin by looking for an empty seat. If the seats were arranged in twos—one by the window, one by the aisle—you would typically take the window seat first so that, as the bus fills up, someone could easily take the seat beside you. If the bus had only a few people already aboard, you would not usually go sit in the aisle seat beside one of the few occupied window seats. And if you did, you could expect a surprised glance from the person in the window seat, wondering why you had chosen to crowd them despite the availability of other empty spots. You would first ensure that all the sets of two had at least one occupant before you began the process of doubling up, unless you encountered a friend or colleague on the bus: then, it would be appropriate to join them and chat together for the duration of your shared journey. If, perhaps, they were engrossed in a book, wearing headphones, or attempting to sleep, you might not. And so on.

Such public transportation etiquette offers a normative frame for typical bus behaviour. Of course, exceptions might arise. You would likely not look askance if a small child sat down beside you, unaware of usual social mores. Someone from a different culture, accustomed to different traditions around use of space, might behave differently. Some buses might have special seats reserved for elderly folks, those with mobility issues, or pregnant people, and while you might give a quick glare to a fit, healthy youngster taking such a seat from someone more deserving, you might ignore an able-bodied person in such a seat if they were in the role of caring for a group of others or

³⁹³ White, *Heracles’ Bow*, *supra* note 15 at 49.

³⁹⁴ Robert Cover, “Violence and the Word” (1986) 95 Yale LJ 1601 [Cover, “Violence”].

attempting to move something large and bulky. You might observe the difficulty and offer them grace. Despite the usual norms, you naturally acknowledge situations where an exception is appropriate. Other behaviours might be locally determined, too. You might observe changes to these norms when traveling to other cities or countries. In another part of the world, you might notice that the norm is the precise opposite of what you are used to: people might occupy the aisle seats first, prioritizing their own easy departures, leaving the window seats open—and leaving you to scuttle over their legs in order to gain access to a seat.

No explicit rule governs these sorts of behaviours. Yet, members of a community are able to discern agreed behaviours on the use of space, governing themselves and making exceptions when appropriate. This type of normativity differs from instances of rules made visible and explicit—like the articulation of formal legal codes.³⁹⁵ Yet, law is often described as generating normative force on the world, by imposing some particular state of affairs on society. And the categories can become surprisingly muddled from time to time, with scholars grappling with instances where legal norms are made visible from social facts. In this way, law’s normativity can appear somewhat elastic, capable of being stretched to meet different definitions, depending on who has set them out. When we think about the normativity of law, fundamentally, we are asking how the law functions—or might function—in its efforts to guide human behaviour. Law professor Leslie Green notes the absurdity of engaging in a conversation about the meaning of normativity in a legal sense. He writes, in a series of provocative questions:

If [...] laws are norms, then to ask how we might explain their normativity is to ask how to explain the normativity of norms. Is that a sensible question? If someone asked us to explain the triangularity of triangles, it is not clear what we could do except say what a triangle is and what properties it has. Of course, no one ever asks for an explanation of the triangularity of triangles, because by the time they are old enough to frame questions like that they already know what triangles are.³⁹⁶

Green’s subsequent analysis goes to the heart of accounts of legal rule-following, asking whether law’s norms are mandatory, and whether only that which is backed by the pain of sanction should be called law.³⁹⁷ Of course, most common-sense accounts acknowledge that some laws, like the formalities of contract, cannot meaningfully be called mandatory. They merely offer suggestions for the best way to do business—and how to couch that business in the safety of legal recourse and remedies. Teasing out this distinction between when law obligates *versus* when it merely suggests is a separate project: one that is further complicated by social facts of how people really behave. Many legal philosophers have struggled with this question of where to allocate social facts within a coherent articulation of law’s norms. HLA Hart proposes that every legal system is made up of

³⁹⁵ Some conceptions of law broaden this perspective considerably. For example, Anishinaabe teachings would view relinquishing one’s seat to an elder to be a matter of law, not simply a matter of etiquette or normativity. These teachings are usually referred to as the seven grandfather/grandmother teachings: treating one another with wisdom, respect, love, bravery, truth, humility, and honesty. See, e.g., John Borrows, “Seven Gifts: Revitalizing Living Laws Through Indigenous Legal Practice” (2016) 2 Lakehead Law Journal 1 at 3, 11.

³⁹⁶ Leslie Green, “The Normativity of Law: What is the Problem?” (last accessed 26 July 2022), online (pdf): *UVic* <www.uvic.ca/victoria-colloquium/assets/docs/Green_Normativity.pdf> [perma.cc/2HLD-SV9V] at 23.

³⁹⁷ *Ibid* at 5.

the social practices of the officials of that system.³⁹⁸ Scott Shapiro rejects this classification problem on metaphysical grounds, arguing that legal rules and behavioural practices are different beasts entirely: rules are “abstract objects” while practices are “concrete events.”³⁹⁹

While these questions of legal theory might seem like impractical abstractions, they matter a great deal for the translation of legal authority from the judicial system and the administrative state to AI and ML. Translating these guiding concepts asks us to agree on what kind of society we are building and what are the rules that demand articulation into code. Shapiro’s “metaphysical” distinction between rules and practices becomes somewhat meaningless when both categories are obfuscated by algorithmic processes. Furthermore, the same social facts that have long bothered legal philosophers now form the basis of enormous data sets of human behaviour.

Lessig’s observations on the impact of digital society might appear especially true here. His model explicitly acknowledges the impact of norms on society, exerting their influence on the central pathetic dot, concurrently with markets, law, and code. When he cautioned that code ultimately collapsed into law, he acknowledged that digital architectures would mandate particular pathways of behaviour. Early Internet designers would enable or constrain particular online activities, effectively reconfiguring concepts like obligation or rule-following. What Internet users were permitted to do was cordoned into components by code. We therefore needed to be especially careful with code, to ensure that Internet design promoted its early utopian ideals of democracy and free-flowing information.⁴⁰⁰ We needed to write code in ways that aligned with the laws we wanted legislated, the morals we wanted reflected in widespread behavioural norms, and the business models we wanted to define market values.

What we didn’t consider was what might happen if the code started writing itself.

Despite Lessig’s insights into norms as constraints, his model does not assist with understanding how the coherence of the legal order might be disrupted by subtle algorithmic manipulations of behaviours. Predictive analytics enables a reordering of behaviour not anticipated by Lessig’s modalities: past practices become data points to suggest future practices. Imagine, when you climb aboard that city bus, being ushered into your seat by a small digital nudge offered by a personalized advertisement—right above the same seat you sat in last time. Or, if subtle targeted advertisements sent your way throughout the day precluded you from choosing the bus at all, leading you instead to purchasing a new electric bicycle or vehicle. The ML techniques used to generate these conclusions can confound intuition and prove surprisingly complicated.⁴⁰¹ That’s why conversations about today’s AI need to transcend previous analogies to Internet regulation and engage more deeply with the social facts that are lapped up through social uses of technology—and aggregated into large data sets. If code was once the regulator, it has become the manipulator.

³⁹⁸ Green, *supra* note 396 at 7.

³⁹⁹ Scott Shapiro, *Legality* (Cambridge, MA: Harvard University Press, 2013) at 102.

⁴⁰⁰ This was especially true since state law could be constitutionally reviewed by the courts, but any law enacted or enforced by code would inject private technology companies’ interests into the generation of normativity. Lessig, *supra* note 253 at 317-318.

⁴⁰¹ Following one’s intuition can be fraught with confirmation bias, while understanding what is technologically possible can be wracked with counterintuitive turns. Selbst & Barocas note that intuition can be dangerous, because it can lead us to blithely accept things we wrongly believe to be true, believing false discoveries since they accord with our pre-existing beliefs. See Andrew Selbst & Solon Barocas, “The Intuitive Appeal of Explainable Machines” (2018) 87 *Fordham L Rev* 1085 at 1129.

Law's normativity has an especially important role to play here, particularly for nebulous, contested concepts like the Rule of Law. It's one thing for a digital nudge to encourage us to choose a particular seat on the bus; it's another matter entirely for the same type of algorithm to calculate a prison sentence, identify a suspect via facial recognition, or autonomously draft a judicial opinion using the same types of data sets as the ones that surreptitiously predicted our usual bus behaviour. Normativity matters for explicitly setting up systems where algorithms will learn by example, governing the question of how to get AI to do precisely what we want, rather than something close to what we want (but not quite right).⁴⁰² Code as law was already problematic with respect to the invisibility of digital systems; it is almost cataclysmic with respect to the inscrutability of ML ones.

Normativity also matters for how it might be incorporated into technologies themselves. Imagine a digital nudge that always wanted us to sit on the left side of the bus first, even if there were plenty of available seats on the right. Such a system would ignore the long-established space sharing norms in favour of its own preferences. Of course, acknowledging any shift also means observing how the original system design created particular norms to begin with: the space sharing behaviours would never have arisen in the same way if window seats were arranged individually or in groups of three. In this way, technology can invite us to follow particular scripts and the incorporated normative frames of AI technologies can smuggle their embedded values into different scenarios.

4. The Cyberlaw & STS Curriculum

In concert with our STS insights, lessons about technological affordances and normativity punctuate our treatment of new technological systems. We can understand technology as deterministic: determining our behaviour for how we will sit on the bus forevermore. We can understand technology as neutral: simply offering a certain mode of bus behaviour that is neither bad nor good, but that is perhaps possible of being weaponized by those with poor intentions. We can understand technology as solutionism: finally solving the pesky problem of having to decide where the best seat on the bus might be. In each instance, acknowledgements of normative frames or values in design are tied to expressions of power. At the same time, we see the regulatory conundrums resulting from the direct pipeline from Silicon Valley to user experience, with efforts to constrain certain uses arising only if problems are identified in practice. The lessons of cyberlaw show the frustrations with curtailing certain online behaviours: issues that are only exacerbated in the age of big data, AI systems, and ML algorithms.

Additionally, this chapter contributes to a broader conversation about the merits of a technology-law-specific method. Harnessing interdisciplinary insights from diverse scholars, it is easy to see how overly focusing on a technology's exceptional qualities can extract it from the overall social context. For the analysis to be undertaken in the coming chapters, I aim to avoid this shortcoming. By the same token, some technology law scholarship can be overly prescriptive, focusing on law reform and statutory changes; as Tranter puts it, a hope that technology's disruption of law can be

⁴⁰² One of the struggles of translating between code and law is the counterintuitive nature of what machines are capable of doing and the methodology behind giving them instructions about what to do next. Christian describes this as part of the alignment problem: the challenges in translating between human and machine instructions, to have a machine behave as intended. Christian, *supra* note 40 at 13.

remedied with more law.⁴⁰³ Indeed, some of the seminal texts in the technology law space, like Hildebrandt's *Smart Technologies and the End(s) of Law*, and Julie Cohen's *Configuring the Network Self*,⁴⁰⁴ suffer from this shortcoming: they present deep, thought-provoking analysis of society's interaction with technology, coupled with somewhat unconvincing solutions. Calo makes this point in his review of *End(s) of Law*: he describes Hildebrandt's prescriptive solutions, like Cohen's before hers, as being "notably less sophisticated than her development of the problem."⁴⁰⁵

My goal in this dissertation is to undertake a slightly different type of analysis: one that parses the problems from a broad, jurisprudential standpoint and presents simply alternate modes for thinking about solutions—rather than proposing specific solutions outright. Given the nature of the critique within this sphere, this approach makes sense to further explicate ongoing social challenges. Any prescription that is developed in the later chapters, particularly in Part III, is designed as more of a meditation on the Rule of Law and corresponding technological futures, especially as they connect with storytelling and narrative. By incorporating lessons from cyberlaw, STS, and legal scholarship generally, I seek to offer interdisciplinary insights on how we delegate power within our democracy's legal system. While this system was intentionally designed to have formal constraints on power, the affordances of new automated technologies might displace such constraints, allowing innovation to drive legal change.

The interrelationship between malleable human action and scripted technological design creates a zone where particular values can be espoused and political ends can be achieved. For AI systems and ML algorithms, the zone where values like efficiency and repeatability are pursued is laden with political potholes: bias, discrimination, inscrutability, privacy violations, entrenching undesirable societal characteristics in code, to name only a few. Law itself also engages with these types of malleable constructs; as Fairfield puts it, law is language *about* language.⁴⁰⁶ When law is intended to pursue ideals like fairness, justice, and equality, these pursuits have a recursive, self-referential aspect, since the pursuit of the value implicates the value itself.⁴⁰⁷ Efforts to characterize law as being somewhat neutral, especially ones predicated on a singular liberal conception of the self, are frustrated by established psychological phenomena at the intersection of automation bias, automation complacency, and over-trust. The narrative imagination, so critical for imagining lives outside our own experience, is threatened by overly quantified approaches. I keep these insights in mind in my exploration of the impacts of automation on law, especially for what they might mean to the articulation of law's stories.

The linkages between story and critical scholarship are long established. Informed by those traditions, Benjamin calls for the articulation of counternarratives, emphasizing the importance of transforming structures to enhance the complexity of how technological progress is viewed. Taking cues from Paulo Freire's *Pedagogy of the Oppressed*, she notes the need to change structures if they do not permit dialogue.⁴⁰⁸ She identifies that how we catalogue the connections between ideas

⁴⁰³ Tranter, "Disrupting," *supra* note 293 at 8.

⁴⁰⁴ Julie Cohen, *Configuring the Network Self: Law, Code, and the Play of Everyday Practice* (New Haven, CT: Yale University Press, 2012).

⁴⁰⁵ Ryan Calo, "Technology, Law, and Affordance: A Review of *Smart Technologies and the End(s) of Law*" (2017) 4 *Critical Analysis of Law* 72 at 76.

⁴⁰⁶ Fairfield, *Runaway*, *supra* note 50 at 50.

⁴⁰⁷ *Ibid.*

⁴⁰⁸ Benjamin, *supra* note 268 at 177.

is important, especially for establishing how power is created and sustained.⁴⁰⁹ While acknowledging the degree to which power is implicated in the state's authority over law, her view also takes into account concentrations of power in corporate settings, which have meaningful results on the technological affordances flowing from their products. With these lessons in mind, I continue to Chapter 3, in which I analyse the role that legal decision-making plays in society, the existing technological incursions into administrative law, and the corresponding implications for jurisprudence.

⁴⁰⁹ As she eloquently puts it: "Genealogies reflect and reproduce power relations." Benjamin, *supra* note 268 at 177.

We have to abandon the arrogant belief that the world is merely a puzzle to be solved, a machine with instructions for use waiting to be discovered, a body of information to be fed into a computer in the hope that, sooner or later, it will spit out a universal solution.
~ Václav Havel⁴¹⁰

Once, at a critical moment in our legal past—in the late 19th century—contentious debates about the content of legal education raged within the American academic community. During this period, the then-Dean of Harvard Law School, Christopher Columbus Langdell, was the vanguard of a movement to formalize legal education. Dean Langdell is widely remembered for a number of reasons: moulding law school into the three-year degree programme taught by professional faculty that we still use today; his initial efforts in conceptualizing modern contract doctrine; and his views on legal formalism.⁴¹¹ Langdell, it seems, would have appreciated the arguments of the legal futurists discussed in the previous chapters: “the only resource that one needs to create a science of law,” he reasoned, “is a law library.” He insisted upon teaching the law as deductive reasoning via syllogism.⁴¹² Arguments proceeded strictly: rule, then facts, then outcome. This conceptualization not only made law reliable and straightforward, but also generalizable—easy to define for all analogous circumstances.

Oliver Wendell Holmes’ seminal “The Path of Law” was an early attempt to break from this classic version of legal formalism.⁴¹³ This paradigmatic piece of legal writing, often touted as the best legal article of all time,⁴¹⁴ began as a speech that Holmes delivered at the opening of Boston College’s new lecture hall in the late 1890s. Boston College, perhaps obviously, was not Harvard Law School, where Holmes taught law under Dean Langdell. It was the region’s “other” law school—and, perhaps for that very reason, it presented Holmes with an opportunity to set out his path-breaking point of view to the assembled titans of the legal community. Through this lens, “The Path of Law” was not only a speech, but a ceremonial address, given on a significant occasion, which elevated and venerated the profession of law.⁴¹⁵ He begins, emphatically, that the study of law is “not studying a mystery, but a well-known profession.”⁴¹⁶ Law is not logical, he argued: law is not syllogism and reason. Instead, law is a business: the business of prediction. When people study the law, they are studying something that is well-documented and inherently practical. Both

⁴¹⁰ Daniel Kahneman, Oliver Sibony, Cass R Sunstein, *Noise: A Flaw in Human Judgement* (New York, NY: Little Brown, Spark, 2021) at 332 [Kahneman et al, *Noise*].

⁴¹¹ Catherine Pierce Wells, “Langdell and the Invention of Legal Doctrine” (2010) *Buffalo L Rev* 551.

⁴¹² In formal logic, syllogisms are logical results flowing from two assumed propositions (premises). For example, Premise 1: All animals are mortal. Premise 2: All humans are animals. Conclusion: All humans are mortal.

⁴¹³ Oliver Wendell Holmes Jr, “The Path of Law” (1897) 10 *Harv L Rev* 457.

⁴¹⁴ Judge Posner calls it “the best article-length work on law ever written.” See Richard A Posner, ed, “Introduction” in *The Essential Holmes: Selections from the Letters, Speeches, Judicial Opinions, and Other Writings of Oliver Wendell Holmes Jr.* (Chicago, IL: University of Chicago Press, 1992) ix at x.

⁴¹⁵ Robert Gordon, “Law as a Vocation: Holmes and the lawyer’s path” in SJ Burton, ed, *The Path of Law and its Influence: The legacy of Oliver Wendell Holmes Jr.* (New York, NY: Cambridge University Press, 2000) 7.

⁴¹⁶ Holmes, *supra* note 413.

overly simplistic representations and overly theoretical abstractions detracted from the ability to characterize law's real adherence in everyday life.

We might be forgiven for thinking that the issues generated by artificial intelligence create new, thorny problems for the articulation of law and legal practice. The high technology associated with AI and ML suggests things are diverging from the well-trodden path of books and words to something sleek, digital, and ephemeral. Yet, these problems highlight the same debates that have been ongoing for centuries: what is law; how does it work; what are its constituent components; and when is the answer something more than "it depends"?

In this chapter, I map the existing legal technology in decision-making contexts. I present the justifications for why such approaches have already been implemented in certain aspects of the legal system. To begin casting doubt on such processes, I refer to sociological research about how people make decisions and what processes people believe to be legitimate. I conclude this section by beginning to cast doubt on the quality of data that can be collected, identifying certain epistemological problems along the way.

This section showcases AI's potential to ameliorate legal problems. This balanced approach should avoid any (misplaced) critiques according to which I appear categorically opposed to any innovation in the law. Accordingly, this section highlights some of the potential benefits of AI enhanced systems, rather than focussing merely on the detractors. It is perhaps possible that AI can have certain circumscribed responsibilities within the legal sphere, though this stops well short of wholesale regulation by machine. This section lays a foundation for those possibilities for AI and ML that are appropriate, while also articulating problem areas that might arise going forward.

1. Introducing AI to Legal Spaces

AI's association with innovation has led to the former becoming the buzzword *du jour* in many industries. Law and legal industry are no exception. Law firms emphasize their expertise in AI through highlighting innovative uses of technology.⁴¹⁷ Some even argue that a modern lawyer's duty of competence is only fulfilled if it embraces a need to modernize legal practice by understanding recent technological innovations and implementing new techniques.⁴¹⁸ Conferences are organized focused on legal technology, where industry players are dedicated to selling the latest tech to the big law firms.⁴¹⁹ Both new and established law schools highlight opportunities for students to learn coding, develop skills in the latest technology, and improve upon the traditional

⁴¹⁷ Representative technologies in this space include Legal Beagle, a smartphone application that connects lawyers with prospective clients and offers electronic document signing, and Kira, a machine learning powered document review software that electronically highlights and produces extracts of documents. See Legal Beagle, "Attorneys" (last accessed 29 July 2022), online: *Legal Beagle* <legalbeagle.net/attorneys/> [perma.cc/WD4W-559C]; Kira, "Deepen Client Relationships. Grow Revenue. Deliver Total Diligence." (last accessed 29 July 2022), online: *Kira Systems* <kirasystems.com/> [perma.cc/YRW7-M5BS].

⁴¹⁸ In 2019, the Federation of Law Societies of Canada updated its Model Code of Professional Competence, which states that lawyers have a duty of technological competence. Amy Salyzyn, "It's Finally (Sort Of) Here!: A Duty of Technological Competence for Lawyers" (26 November 2019) online: *SLAW* <www.slw.ca/2019/11/26/its-finally-sort-of-here-a-duty-of-technological-competence-for-canadian-lawyers/> [perma.cc/6PU2-CFN6]. Since the COVID-19 pandemic began in March 2020, many court proceedings have moved online, underscoring the need for both lawyers and judges to understand modern technological capabilities.

⁴¹⁹ The overlap here between start-up culture and legal industry is significant. See, e.g., University of Alberta Faculty of Law, "2022 Legal Innovation Conference: Introducing & Embracing Tech & Automation in the Legal Industry" (28 January 2022), online: *UAlberta* online: <www.ualberta.ca/law/about/legal-innovation.html> [perma.cc/5SKA-VE87].

avenues of learning the law.⁴²⁰ In all of these cases, innovation is proffered as an opportunity to modernize and streamline law.

The justification for employing innovative AI/ML tools is immediately clear: lawyering is a complicated field, wherein lawyers hold professional responsibilities to their clients to do the best they possibly can. If technology is going to make something better or easier, it follows that legal practitioners, bureaucrats, and other guardians of the judicial system might need to adopt the newest tools. Moreover, today's judicial system is rife with real concerns about overcrowded courtrooms, significant wait times creating barriers for access to justice, and the presence of bias in decision-making generally. Legal realists assert empirical evidence against fairness in judicial decision-making, arguing that sentences are more often impacted by a judge's state of mind or hunger level than the particulars of the instant case.⁴²¹ A growing movement calls for increased data science in legal practice to better support decisions about whether to pursue litigation *versus* settlement.⁴²² Following the example of medicine in moving from an art to an evidence-based field, there are those who call for increased legal empiricism and more quantitative analysis of data in law.⁴²³ As seen through this lens, there are certain fundamental ideas about reasonableness in the law that ask for a decision-making process to be represented through a clear decision tree, outlining how the decision-maker arrives at their reasonable conclusion. To map the transformation from human decision-maker to replacement bot, consider the implementation of a relatively straightforward automated process as a framing example: a red-light camera installed in place of a human police officer to monitor red light violations.

a. Framing Example: Red-Light Camera Deployed as “Automated” Police Officer

Red lights seem to impose a strict, non-negotiable rule: cars cannot cross an intersection if the stoplight is red. Typically, violating this rule would attract the attention of a police officer, responsible for ensuring that all drivers operate in accordance with this rule. Failure to do so will result in a ticket under the relevant local ordinance that mandates compliance with traffic rules. In theory, a system designer could interview an experienced, dedicated, and well-meaning police officer for the list of steps typically undertaken when monitoring red light violations, aiming to implement those same factors in the technological replacement. One might imagine that it would be relatively easy to map the human's decision-making process to a computer-enabled equivalent using a decision tree. And, if one is exclusively process-oriented, the red-light camera will easily achieve its goal: it will determine whether someone has broken the law by driving through a red light. This is a binary criterion: either it has happened, or it has not. Mireille Hildebrandt describes

⁴²⁰ See, e.g., the Legal Innovation Zone at Toronto Metropolitan University (TMU, formerly Ryerson University). They boast a legal technology incubator that offers support to start-ups “building technology based solutions for legal consumers.” TMU, “Legal Innovation Zone” (last accessed 29 July 2022), online: <www.torontomu.ca/zone-learning/legal-innovation-zone/about/> [perma.cc/XFX6-M7LM].

⁴²¹ Shai Danziger, Jonathan Levav & Liora Avnaim-Pesso, “Extraneous Factors in Judicial Decisions” (2011) 108 *Proceedings of the National Academy of Sciences* 6889.

⁴²² See, e.g., Paul-Erik Veel, “Getting Leave to Appeal to the Supreme Court: Empirical Insights from Tax Cases” (20 January 2021), online: *Lenczner Slaughter Blog* <litigate.com/OnTheDocket#/getting-leave-to-appeal-to-the-supreme-court-empirical-insights-from-tax-cases> [perma.cc/F6ES-PWQ5].

⁴²³ *Ibid.* See also Jena McGill & Amy Salazyn, “Judging by the Numbers: Judicial Analytics, the Justice System and its Stakeholders” (2021) 44 *Dal LJ* 249.

this phenomenon as “if this, then that regulation,” referencing to the Boolean logic that fuels many automated decision processes.⁴²⁴

Given that red-light violation is a simple, binary adjudication, the potential advantages of using AI are clear. Our automated bot will treat all situations equally, unimpeded by contextual circumstances. Any concerns about latent, unconscious, or conscious biases will be eliminated, as the bot will not harbour antiquated ideas about the inadequacy of women drivers or exhibit racist tendencies. The bot is not impacted by hunger, fatigue, retribution, bad weather, or other sensory experiences. It will not become distracted by its mobile phone, sneeze at an inopportune moment, doze off unexpectedly, or take an unauthorized coffee break in the middle of a shift. It will not fudge the outcomes of a few sneaky drivers making dangerous choices by entering the intersection on a waning yellow light to make up its ticket quota at the end of the month. The bot will create an instance of the justice system that is fair, consistent, and evenly applied across gender, racial, and socioeconomic lines—among other distinguishing factors.

Despite these perceived advantages, we must ask ourselves: is this really so? Would the automated red-light camera bot provide a better way to dispense justice than a human police officer? What about in other incarnations of the judicial system? Would the same be true in criminal law, tax law, or administrative law?

Through the lens of the bot’s installation as an automated police officer, certain themes emerge. Some advantages for automated processes are immediately apparent, like fairness, consistency, accuracy, and affordability. Administrative law offers an interesting laboratory to evaluate some of these concepts, since they tend to be emphasized in this legal sphere. In addition, quasi-administrative or quasi-judicial processes where human decision-makers make important decisions impacting individuals’ lives are also vulnerable to being supplanted by automated processes.

b. Datafying the Administrative State

In Canadian law, guarantees of natural justice and procedural fairness are deeply entrenched in the operation of the administrative state. Even after recent modifications to the operative legal framework in administrative contexts,⁴²⁵ certain central guarantees remain fundamental to the operation of law. These natural justice and procedural fairness obligations include the right to a fair hearing, impartiality, and independence, all enshrined in an administrative system that allows discretionary decision-making while upholding central tenets of morality.⁴²⁶ Since at least the *Nicholson v Halimand-Norfolk (Regional Municipality) Commissioners of Police* decision in 1978, the SCC has adopted a relatively flexible way of interpreting administrative decision-makers’

⁴²⁴ Hildebrandt, “Algorithmic Regulation,” *supra* note 32 at 2.

⁴²⁵ The Canadian administrative framework underwent significant changes over the past few decades, with significant upheaval in the 2008 decision *Dunsmuir v New Brunswick*, which jettisoned some standards of review, and culminating with the SCC’s administrative law trilogy in 2019. The cornerstone case of that trilogy, *Canada (Minister of Citizenship and Immigration) v Vavilov*, 2019 SCC 65, reinforced that the standard of review was presumed to be reasonableness unless there is a basis to rebut that presumption.

⁴²⁶ Former Justice Louis LeBel, one of the architects of the SCC’s ongoing attempts to renew administrative law during his tenure on the court, has spoken at length about the role of procedural fairness in Canadian administrative tribunals. He posits that the three core values of procedural fairness are access to a fair hearing (*audi alteram partem*), impartiality, and independence. See The Hon. Louis LeBel, “Notes for an Address: Reflections on Natural Justice and Procedural Fairness in Canadian Administrative Law” (2013) 26 Can J Admin L & Prac 51 at 53-57.

obligations of procedural fairness.⁴²⁷ Even in the face of somewhat rigid procedural rules, administrative decisionmakers depend on discretion for the ability to “take individual circumstances into account and allow decision makers to adapt their approaches accordingly.”⁴²⁸ In *Baker v Canada (Minister of Citizenship and Immigration)*, the SCC confirmed the contents of the duty of procedural fairness, opining on its flexibility and context-dependency, and prescribed a framework for its operation.⁴²⁹ In *Blencoe v British Columbia (Human Rights Commission)*, the SCC explains the co-operative aspects of principles of natural justice and the duty of fairness to generate fair procedure for litigants.⁴³⁰ Later, writing extrajudicially, Justice Louis LeBel explained the centrality of administrative law’s operations as warranting careful protection, given that “Canadians will deal with administrative action and justice more often than with the civil or criminal courts in their daily life.”⁴³¹ To achieve the goals of the administrative state, a certain flexibility is required to ensure that the precepts apply meaningfully to individual litigants’ life circumstances. As eloquently expressed by the majority in *Dunsmuir v New Brunswick*:

Procedural fairness is a cornerstone of Canadian administrative law. Public decision makers are required to act fairly in coming to decisions that affect the rights, privileges or interests of an individual. Thus stated, the principle is easy to grasp. It is not, however, always easy to apply. As noted many times, “the concept of procedural fairness is eminently variable and its content is to be decided in the specific context of each case.”⁴³²

Administrative boards and tribunals are typically tasked with frontline adjudication responsibilities that can really impact people’s daily lives. Since World War II, increasing delegation of decision-making authority to administrative agencies has dramatically increased the scope and size of the administrative state, broadening the scope of application for duties of procedural fairness.⁴³³ The entire legal mechanism of administrative law is organized around an assumption of expertise by those frontline adjudicators, who have both longstanding experience in the relevant specialty field and the benefit of the evidentiary record before them when making decisions. Yet, these types of decisions are frequently maligned in an administrative law context, where principles of deference in reasonableness analysis can lead to incoherent case law. Indeed, standard of review in Canadian administrative cases has been in a state of flux since *Dunsmuir v New Brunswick*,⁴³⁴ and the SCC’s recent pronouncements in *Canada v Vavilov* attract similar criticisms.⁴³⁵ While there is not

⁴²⁷ *Nicholson v Halimand-Norfolk (Regional Municipality) Commissioners of Police*, [1979] 1 SCR 311 (SCC).

⁴²⁸ LeBel, *supra* note 426 at 57.

⁴²⁹ *Baker v Canada (Minister of Citizenship and Immigration)* [1999], 2 SCR 817.

⁴³⁰ *Blencoe v British Columbia (Human Rights Commission)*, 2000 SCC 44 (characterizing the duty of procedural fairness owed to a complainant and whether a five-month delay violated that duty).

⁴³¹ LeBel, *supra* note 426 at 57.

⁴³² *Dunsmuir v New Brunswick*, 2008 SCC 9 at para 79. For the source of their quotation, three previous cases are listed: *Knight v Indian Head School Division No 19*, [1990] 1 SCR 653 at p. 682; *Baker*, *supra* note 429 at para 21; and *Moreau Bérubé v New Brunswick (Judicial Council)*, 2002 SCC 11 at paras 74-75.

⁴³³ Some observers seek to differentiate between the content of *procedural* duties of fairness, which have expanded considerably since WWII with the growth of the administrative state (developed doctrinally by the SCC in *Baker*), and the content of *substantive* duties of fairness, which have a longer history in the common law (and tend to connote deference to frontline decisionmakers, provided the decision is reasonable). See Paul Daly, “Canada’s Bipolar Administrative Law: Time for Fusion” (2014) 40 Queen’s LJ 213 (arguing that the firm distinction between procedural and substantive aspects of review should not hold post-*Baker*).

⁴³⁴ *Dunsmuir*, *supra* note 432.

⁴³⁵ *Canada (Minister of Citizenship and Immigration) v Vavilov*, 2019 SCC 65.

particular need in this thesis to wade into administrative law principles, this discussion is nonetheless important for contextualizing the manner through which automated functions can be deployed in daily life. The same is true in quasi-judicial or other administrative settings that expand the scope of law to include law-like decisions impacting a person's livelihood.

At first glance, it may seem that a quantified approach to legal decision-making would satisfy the process-oriented concerns highlighted in non-instrumental accounts of procedural justice. Emphasizing consistency, non-arbitrariness, and the guarantees embodied by procedural fairness, automated decision-making (ADM) appears as the antidote to the legal realists' concerns about inconsistencies in human judgement. ADM has been deployed for decisions impacting education, credit scoring, financial futures, citizenship and immigration determinations, and others. As Jennifer Raso points out, in direct contradiction of visions of "paper pushers and red tape," administrative agencies tend to be early adopters of new technological processes for automating decision-making.⁴³⁶ Such technologies often promise glorious outcomes like "streamlin[ing] data management" and "speed[ing] up routine decision-making," which Raso argues are connected to Canadian administrative law's normative commitments of "fair process; responsive and responsible decision-makers; and substantive justice."⁴³⁷ Yet, because administrative legal developments tend to be rooted in the doctrinal considerations brought before courts, there are few significant commentaries from legal decisions on the use of automated tools. Even though such technologies have been in use for years, there is limited discussion of their merits. Existing decisions tend to focus on whether data should be disclosed, often in response to freedom of information requests.⁴³⁸

Even before the proliferation of algorithmic processes and machine learning methodologies, legal commentators observed how automation threatens to supplant human decision-making in administrative agencies. In 2008, Danielle Citron was already describing how automated processes were increasingly making primary decisions in diverse administrative settings, ranging from terminating social benefits, banning individuals from air travel, quantifying child support payments, purging voters from the official register of voters, and determining eligibility for federal contracts.⁴³⁹ Since at least 2014, Canada's immigration and refugee process has experimented with ADM of varying capabilities, including technological solutions for classifying cases, scoring probability assessments, flagging cases for review, providing recommendations, and even rendering decisions.⁴⁴⁰ Canada is not alone in using immigration bots; other countries are also experimenting with similar technologies.⁴⁴¹ Other areas of the Canadian government have implemented similar processes. The Canadian Revenue Agency (CRA) uses predictive analytics to rapidly investigate non-compliance and better understand taxpayer decisions regarding tax debt.⁴⁴² Economic and Social Development Canada (ESDC) uses a risk-scoring algorithm for its investigations into

⁴³⁶ Raso, *supra* note 191 at 165.

⁴³⁷ *Ibid* at 181.

⁴³⁸ Raso lists a number of example cases that are limited in this way, usually resulting from freedom of information requests for procurement documents. Because of this, she worries, judicial reviews will be unlikely, and contracts often bar disclosure of specific innovative algorithmic processes. *Ibid* at 198.

⁴³⁹ Danielle Keats Citron, "Technological Due Process" (2008) 85 Wash UL Rev 1249 at 1252.

⁴⁴⁰ Petra Molnar & Lex Gill, "Bots at the Gate: A Human-Rights Analysis of Automated Decision-Making in Canada's Immigration and Refugee System" (2018), online (pdf): *Citizen Lab* <citizenlab.ca/wp-content/uploads/2018/09/IHRP-Automated-Systems-Report-Web-V2.pdf> [perma.cc/3DZT-9XP2].

⁴⁴¹ *Ibid* at 14.

⁴⁴² Diagnostic Report, Steering Committee on Big Data, ATIP request by Lex Gill, online: *Google Drive* <drive.google.com/file/d/1HpvGzMdF7SPH319iNA2wkkQkDwDax7F7/view> [perma.cc/L5KM-AGER] at 11.

overpayment.⁴⁴³ The Department of Justice (DOJ) is developing predictive models for legal risk management and resource allocation of legal service delivery.⁴⁴⁴ National Resources Canada (NRCan) uses consumer data to map trends and identify targets for fuel consumption on personal vehicles.⁴⁴⁵ Both NRCan and Industry Canada (IC) use social media data to assess consumer attitudes and better tailor messaging for receptive audiences.⁴⁴⁶ Taken together, these types of statistical and technological machinations show a governmental preoccupation with data-driven decision-making, greasing the wheels for more advanced ADM techniques in the future.

Certain advantages to automation are immediately clear: reduction in human-power can mean cost-saving in salaries and offices; automated processes generate consistent results; and people interacting with the system are saved the hassle of physically attending an agency office.⁴⁴⁷ Such streamlining, Citron points out, “risks dismantling critical procedural safeguards at the foundation of administrative law” as ADM combines rule-making and adjudicative processes into a single, seamless task.⁴⁴⁸ Her analysis is supported by James Grimmelmann’s complementary observations on how processes are, in effect, regulated by software: automation, immediacy, and plasticity each limit possible human behaviours when interacting with software systems.⁴⁴⁹ Software is thereby relegated to the extreme end of the continuum for the debate on rules *versus* standards, with software as a peak example of the rule-bound end of the continuum.⁴⁵⁰ Automation, once set into motion by a programmer, autonomously makes its decisions.⁴⁵¹ Its effects are immediate, which means any forbidden behaviour never occurs.⁴⁵² Its profound plasticity means that almost any system can be implemented and precisely described, leading to flexibility and granularity that far exceeds physical architectures.⁴⁵³ This combination of nearly costless replication and extremely specific application deployment makes software’s ability to regulate particularly robust.

Harnessing Hildebrandt’s distinction of data-driven *versus* code-driven regulation, catalogued in Chapter 1, Raso observes that the relative sophistication of the ADM tools is less important than how they impact decision-making processes in practice.⁴⁵⁴ Her primary concern is anchored in questions of substantive fairness, which triggers reasonableness review. Noting courts’ different treatment of decisions requiring correctness review *versus* reasonableness review, she observes that ADM would not pose serious challenges for correctness review (associated with procedural fairness) since a reviewing court would look only to whether the answer is correct.⁴⁵⁵ Conversely, challenges might arise in reasonableness review, where scrutinizing the administrative decision-

⁴⁴³ *Ibid* at 11.

⁴⁴⁴ *Ibid*.

⁴⁴⁵ *Ibid* at 12.

⁴⁴⁶ *Ibid*.

⁴⁴⁷ Citron, *supra* note 439 at 1252-1253.

⁴⁴⁸ *Ibid* at 1253.

⁴⁴⁹ Grimmelmann, “Regulation,” *supra* note 254 at 1723. Grimmelmann intends this observation to add further specificity to Lessig’s famous modality: code is law. See Chapter 2 for discussion of Lessig’s model.

⁴⁵⁰ Grimmelmann points out that absent discretion, increasing complexity does not turn rules into standards. He writes: “Software rules can become almost unimaginably complex without their hard edges blurring.” *Ibid* at 1733.

⁴⁵¹ And, once it has been written, the cost to apply that same process to another decision set can be “vanishingly small.” *Ibid* at 1729.

⁴⁵² *Ibid*.

⁴⁵³ Grimmelmann describes this using the Church-Turing thesis, which posits that all moderately sophisticated programming languages are equally powerful, since they can all be used to express the same ideas. *Ibid* at 1730-1731.

⁴⁵⁴ Raso, *supra* note 191 at 187.

⁴⁵⁵ *Ibid* at 197.

maker's reasons form a significant aspect of the process.⁴⁵⁶ For simple code-driven ADM analogues, like the red-light traffic bot, how the bot arrived at its answer might prove straightforward if it adhered to a determinative decision tree. Yet, this would not prove true for more complex data-driven ML algorithms, capable of shifting and adapting through progressive, iterative functions. Such situations might obscure the relevant aspects of decision-making processes in a manner incongruent with robust reasonableness review. Adjacently, Raso notes that ADM-generated reasons might deploy standard-form or boilerplate phrases, which have been recognized by courts as insufficient reasons.⁴⁵⁷ Given current ADM capabilities, Raso is right to express concerns about the validity of their integration into administrative contexts. Similar concerns exist in diverse quasi-judicial settings, which historically insisted upon human judgment.

c. Automated Analogues for Human Decision-Makers

As technology has continued to advance, the corresponding concerns have amplified. Uses of automation in administrative settings have led to numerous recent controversies. Consider, for example, the use of automation by British schools to assign automatic grades to students when COVID-19 pandemic restrictions prevented them from being able to sit their A-level examinations.⁴⁵⁸ Students received grades commensurate with historical scores for their school, instead of the estimated grades their teachers submitted to the Orwellian-sounding qualifications regulator, "Ofqual." For students from socioeconomically disadvantaged areas, which had historically obtained lower test scores, this might mean that their grades were downgraded, sometimes by as much as two letter grades from their teachers' submitted estimates (*e.g.*, from an A to a C).⁴⁵⁹ Not only did this impact students' future plans, it also negated their hard work in revising and preparing for their examinations. As one student explained, normally exam results mean that students only have themselves to blame if they do badly, but this system removed their personal accountability from the equation.⁴⁶⁰ When students took to the streets protesting Ofqual's assigned grades, their selected cheers included "Fuck the Algorithm!" or "The algorithm stole my future!"⁴⁶¹ Different from previous outcries emanating from digital data collection or egregious privacy breaches, these protests "weren't focused on how their data might be used in the future, but how data had been actively used to *change* their futures."⁴⁶² At the heart of this discontent was concern about how algorithmic decision-making negates individual agency, extracting humanity from decisions about how to rule human lives.

⁴⁵⁶ Raso, *supra* note 191 at 197.

⁴⁵⁷ *Ibid.*

⁴⁵⁸ Louise Moore, "Why 'Ditch the algorithm' is the future of political protest," (19 August 2020), online: *The Guardian* <www.theguardian.com/commentisfree/2020/aug/19/ditch-the-algorithm-generation-students-a-levels-politics> [perma.cc/5ZHN-7YDS].

⁴⁵⁹ Over 40% of students received downgraded grades. Although students were promised that their grades would be downgraded by no more than one letter grade, this was quickly revealed to be false, especially from students from less affluent areas. Commentators observed the possibility that these results would serve to widen the educational and economic gap between students from affluent and less affluent areas. See Emma Gill, "'I am expected to just live with these unfair grades' – Student's open letter to the government as she slams A-level results system" *Manchester Evening News* (13 August 2020), online: *Manchester Evening News* <www.manchestereveningnews.co.uk/news/greater-manchester-news/a-level-results-unfair-downgraded-18764743> [perma.cc/9PXS-5KKJ].

⁴⁶⁰ *Ibid.*

⁴⁶¹ Moore, *supra* note 458.

⁴⁶² *Ibid.*

This kind of dystopian outcome often appears when one size-fits-all solutions are substituted for nuanced, individual decision-making. Cathy O’Neil writes compellingly about this phenomenon in her book, *Weapons of Math Destruction*, in which she illustrates how automated processes can wreak havoc on human dignity in a variety of fora.⁴⁶³ Statistical or mathematical processes, while appearing impartial and fair, can lead to fundamentally unjust results. Notably, she recounts the challenges teachers face when their employment outcomes are linked to pupils’ performances on standardized tests;⁴⁶⁴ recidivism algorithms and their connection to stop-and-frisk policing practices in poorer neighbourhoods;⁴⁶⁵ credit scoring as it relates to insurance premiums and car insurance;⁴⁶⁶ and the actuarial tables that link race and poverty with poor life expectancy, thereby quantifying the social determinants of health.⁴⁶⁷ In each of these scenarios, O’Neil’s careful analysis quantifies how blunt weaponization of mathematical models leads to unjust results, even when the ideas are presented under the guise of equality. Far from treating people equally, these sorts of algorithms punish those in already disadvantaged circumstances, further exacerbating social limitations evoked by traditionally limiting biases.

For example, Latanya Sweeney reports targeted online advertising suggestive of arrest records as appearing more often for black-identifying names than white-identifying names.⁴⁶⁸ Closer to home, a Toronto-area school board made headlines by automating selection of online classes alphabetically by last name.⁴⁶⁹ While this seemed both efficient and neutral, it resulted in entire classrooms of students with last names common to certain ethnic groups; for example, an entire 30-person group all had the last name “Wong.” Like the Twitter-cropping algorithm discussed in Chapter 2, this example raises the spectre of racial segregation, demonstrating how even seemingly neutral policies can have discriminatory results.

Discrimination can also adhere in sectors where automation is used to identify defining characteristics of individuals. The information garnered can then be used to power various ADM processes or to produce targeted online advertisements. In an especially poignant example, algorithmic processes can be used to classify a person’s gender, combining algorithmic techniques of facial recognition and body recognition in audio, images, or video to determine gender.⁴⁷⁰ Yet, as researchers point out, gender is a complex concept, increasingly revealed to have “multifaceted internal aspects, which exhibit much more diversity and fluidity than thought before.”⁴⁷¹ Looking at gendered analyses through the lens of transgender individuals reveals the damage that such technological processes can impose on those living outside of binary conceptions of gender.⁴⁷² Taken a step further, the example reminds us that complexity, fluidity, and diversity are typically

⁴⁶³ O’Neil, *supra* note 372.

⁴⁶⁴ *Ibid* at 135-140.

⁴⁶⁵ *Ibid* at 101-103.

⁴⁶⁶ *Ibid* at 142-150; 164-166.

⁴⁶⁷ *Ibid* at 161-164.

⁴⁶⁸ Latanya Sweeney, “Discrimination in Online Ad Delivery” (2013) 56 *Communications of the ACM* 44.

⁴⁶⁹ Caroline Alphonso, “Toronto-area school board sorts classes alphabetically, raising concerns of racial segregation” (24 Sept 2020), online: *The Globe & Mail* <www.theglobeandmail.com/canada/article-toronto-area-school-board-sorts-online-classes-alphabetically-raising/> [perma.cc/4LB8-MXX8].

⁴⁷⁰ Foad Hamidi, Morgan Klaus Scheuerman, Stacy M Branham, “Gender Recognition or Gender Reductionism? The Social Implications of Automatic Gender Recognition Systems” (2018) *CHI 2018* (April 21-26, Montreal, QC, Canada).

⁴⁷¹ *Ibid* at 1.

⁴⁷² Hamidi et al, *supra* note 470 at 3. In this study, transgender participants described the harms experienced through automated misgendering processes through interviews.

flattened by automated processes, which can result in disproportionate impacts felt by marginalized individuals living outside of society's majority.

Virginia Eubanks' *Automating Inequality* offers more compelling examples for how automated processes can create disproportionate impacts in disadvantaged socioeconomic communities.⁴⁷³ One of her key observations is that ADM is often tested in "low-rights environments" where the impacted individuals are already marginalized in other ways.⁴⁷⁴ This is by design: the people the programming impacts are already less likely to push back against the implementation of such policies, due to lack of time, skill, transparency, or political power.⁴⁷⁵ And, the specific areas in which such technologies are initially deployed are those that dominant social spheres tend to ignore. Still, vigilance is necessary. This digital poorhouse, Eubanks cautions, can begin in specific sectors of society, but is often being primed for transference to the wider community.⁴⁷⁶

... and while the most sweeping digital decision-making tools are tested in what could be called 'low rights environments' where there are few expectations of political accountability and transparency, systems first designed for the poor will eventually be used on everyone.⁴⁷⁷

With these challenges in mind, it is not surprising that some governments have attempted to impose regulations that guide the implementation of ADM. Next, I consider some of those efforts, as well as accompanying critiques of further delegating decision-making power in spaces already characterized by the delegation of authority from legislature to administrative agency.

d. Proposed Regulatory Frameworks

Canada's Directive on Automated Decision-Making (DADM) offers guidance for systems that employ AI to either make or assist with administrative decisions.⁴⁷⁸ Hopeful that AI might enhance service delivery, the DADM purports to prioritize key administrative law principles, such as procedural fairness, accountability, legality, and transparency.⁴⁷⁹ These goals offer a normative frame for the DADM's operation. Section 6 of the DADM delineates requirements for anyone using automated decision-making substitutes. It mandates an Algorithmic Impact Assessment (AIA) before implementing the technology.⁴⁸⁰ It creates transparency requirements, offering notice of an ADM process before the decision is rendered and providing meaningful explanations after the decision itself.⁴⁸¹ It also sets out quality assurance criteria focused on testing, peer review, legal compliance, and the possibility of human intervention, among other factors.⁴⁸² Taken together,

⁴⁷³ I offer more commentary on this text in Chapter 5. See generally Virginia Eubanks, *Automating Inequality: How High-Tech Tools Profile, Police, and Punish the Poor* (New York, NY: Picador, 2019).

⁴⁷⁴ *Ibid* at 12.

⁴⁷⁵ *Ibid* at 16.

⁴⁷⁶ *Ibid* at 12.

⁴⁷⁷ *Ibid*.

⁴⁷⁸ Government of Canada, "Directive on Automated Decision-Making" (1 April 2021), online: *TBS Canada* <www.tbs-sct.gc.ca/pol/doc-eng.aspx?id=32592> [perma.cc/NRX3-W887] [DADM].

⁴⁷⁹ *Ibid*.

⁴⁸⁰ *Ibid*, s 6.1.

⁴⁸¹ *Ibid*, s 6.2.

⁴⁸² *Ibid*, s 6.3.

these requirements aim to enshrine principled application of administrative law ideals within automated applications.

Other administrative law standards are explicitly incorporated into the DADM. To comply with administrative law's precepts of providing notice whenever a decision will impact an individual's rights,⁴⁸³ the DADM mandates user notification that a given process is being undertaken by an ADM.⁴⁸⁴ Specifically, the DADM identifies that some situations will mandate human intervention, where the AIA returns a high or very high impact decision.⁴⁸⁵ Where administrative law might traditionally allocate the right to an oral hearing if some threshold level of rights-engagement is surpassed, Teresa Scassa observes that the procedural equivalent in an ADM-context seems to be guaranteeing in-person involvement.⁴⁸⁶ The level of human involvement required, Scassa points out, seems to correlate to "the potential impact of the decision on the affected party's rights or interests."⁴⁸⁷ While the DADM remains silent on the availability of oral hearing rights, this right should be differentiated from the right to an explanation for a given decision. Relatedly, a petitioner's right to be provided with reasons for a given administrative decision raises the spectre of arbitrariness and transparency.⁴⁸⁸ Absent some scrutiny of the process, fear that the decision was made for arbitrary reasons might persist in a petitioner's mind.

The shift to ADM sublimates administrative law's theoretical policy objectives into definitive, coded procedures. The scope of the DADM's recommendations is motivated by the same types of objectives that procedural fairness typically entails. Scassa notes that these organizing concerns regarding the appropriate roles for procedural fairness and nature justice are long-contested issues within government decision-making processes.⁴⁸⁹ Such issues are not unique to the advent of AI and ML, since defining how to best represent these ideals within government procedures is a longstanding challenge. Still, Scassa observes how ADM might signal, "at least in part, a return to legal formalism," where decision-making flows from identification of legal rules, explicit encoding of those rules, and the lessons learned from patterns of past decision-making.⁴⁹⁰ She maps the distinctions that arise as administrative law principles are translated to ADM. One of her chief questions is what differences emerge in the translation process—and what those differences might tell us about the ADM's fundamental nature.⁴⁹¹

In another vein, Scassa wonders whether administrative law's right to a hearing can ever be satisfied through delegation to an ADM.⁴⁹² Is there something fundamental about speaking your case aloud

⁴⁸³ *Canada (Attorney General) v Canada (Commission of Inquiry on the Blood System)* [1997], 3 SCR 440 (opining on the reasonable notice provision of the *Inquiries Act* and confirming the issuance of notices as an aspect of procedural fairness).

⁴⁸⁴ DADM, *supra* note 478, s 6.2.1.

⁴⁸⁵ *Ibid*, s 6.3.9, Appendix C.

⁴⁸⁶ Teresa Scassa, "Administrative Law and the Governance of Automated Decision-Making: A Critical Look at Canada's Directive on Automated Decision-Making" (2021) 54 UBC L Rev at 22 [Scassa, "ADM"].

⁴⁸⁷ *Ibid* at 22.

⁴⁸⁸ The right to be provided with reasons in Canadian administrative law is located in *Baker*, *supra* note 429. Per the reasons of Cory J at para 43: "[I]n certain circumstances, the duty of procedural fairness will require the provision of a written explanation for a decision. The strong arguments demonstrating the advantages of written reasons suggest that, in cases such as this where the decision has important significance for the individual, when there is a statutory right of appeal, or in other circumstances, some form of reasons should be required."

⁴⁸⁹ Scassa, "ADM," *supra* note 486 at 1.

⁴⁹⁰ *Ibid* at 9.

⁴⁹¹ *Ibid* at 2.

⁴⁹² *Ibid* at 23.

and being heard by a human decision-maker that cannot be replicated by a technological process? And if so, what is the innate characteristic from this process that demands preservation? Some critics approach this question through the lens of transparency, while others have considered values including equality, privacy, and fairness.⁴⁹³ In particular, the importance of explainable administrative processes have been emphasized by writers like Frank Pasquale, who characterizes many of today's exercises of power in business, finance, and governance as occurring in an inscrutable black box.⁴⁹⁴ Regardless, all of these concerns have the potential to present as Rule of Law issues when considered in a public law context.⁴⁹⁵

Since the introduction of Europe's *General Data Protection Regulation (GDPR)*, commentators have discussed the availability of a so-called "right to an explanation" for any decision generated by an automated process. Although the *GDPR* contains several provisions that reflect this underlying rationale, no explicit protections exist within the regulation itself. People seeking legal justifications most commonly point to Articles 13 through 15, which discuss the meaningfulness of explanations about the logic of decisions.⁴⁹⁶ Relatedly, Article 22(1) guarantees that "a data subject shall have the right not to be subject to a decision based solely on automated processing, including profiling, which produces legal effects concerning him or her, or similarly significantly affects him or her."⁴⁹⁷ Yet more explicit language around the right to an explanation exists only in the recitals,⁴⁹⁸ which are not legally binding. While the interpretation of these provisions remains untested until initial legal challenges are brought, more and more commentators are suggesting that the *GDPR* may not be the panacea previously anticipated.⁴⁹⁹

More recently, the European Commission (EC) has proposed new AI regulations, which would categorize different uses of AI with corresponding risk levels: minimal/no risk, limited risk, high risk, and unacceptable risk.⁵⁰⁰ Under the EC's framework, minimal/no risk technologies would be permitted with no restrictions, while those identified as exhibiting limited risk (described as "AI with specific transparency obligations") would be permitted, but made to comply with certain

⁴⁹³ Scassa, "ADM," *supra* note 486 at 4.

⁴⁹⁴ Pasquale, *Black Box*, *supra* note 246.

⁴⁹⁵ Scassa, "ADM," *supra* note 486 at 4.

⁴⁹⁶ GDPR Articles 13-15 discuss personal data and right of access by the data subject. EU, *Regulation (EU) 2016/2017 of the European Parliament and of the Council of 27 April 2016 on the protection of natural persons with regard to the processing of personal data and on the free movement of such data, and repealing Directive 95/46/EC (General Data Protection Regulation)*, [2016] OJ L 119/1, online: *EU* <gdpr-info.eu/> [perma.cc/2ETN-9CDA] [*GDPR*].

⁴⁹⁷ GDPR Article 22(1) reads: "The data subject shall have the right not to be subject to a decision based solely on automated processing, including profiling, which produces legal effects concerning him or her or similarly significantly affects him or her." Article 22(3) provides the possibility of a similar mechanism to protect data subjects' rights in the case of certain enumerated government and consent-based carveouts in 22(2). *Ibid.*

⁴⁹⁸ See especially Recital 71, which states in part: The data subject should have the right not to be subject to a decision, which may include a measure, evaluating personal aspects relating to him or her which is based solely on automated processing and which produces legal effects concerning him or her or similarly significantly affects him or her, such as automatic refusal of an online application or e-recruiting practices without any human intervention. [...] In any case, such processing should be subject to suitable safeguards, which should include specific information to the data subject and the right to obtain human intervention, to express his or her point of view, to obtain an explanation of the decision reached after such assessment and to challenge the decision. For the full text, see *ibid.*

⁴⁹⁹ See, e.g., Andrew Burt, "Is there a 'right to an explanation' for machine learning in the GDPR?" (1 June 2017), online: *IAPP* <iapp.org/news/a/is-there-a-right-to-explanation-for-machine-learning-in-the-gdpr/> [perma.cc/Z9UG-FFVP].

⁵⁰⁰ European Commission, "Proposal for a Regulation laying down harmonised rules on artificial intelligence" (April 21, 2021), online: *European Commission* <digital-strategy.ec.europa.eu/en/library/proposal-regulation-laying-down-harmonised-rules-artificial-intelligence> [perma.cc/9RA5-69D2].

informational or transparency obligations.⁵⁰¹ Meanwhile, high risk technologies, such as recruitment tools or medical devices, would only be permitted if subjected to strict compliance requirements and an *ex ante* conformity assessment.⁵⁰² AI technologies deemed to exhibit an unacceptable level of risk, like social scoring, exploitation of children, or remote biometric identification for law enforcement, are strictly prohibited under this framework.⁵⁰³ Still, many analysts have characterized the AI regulations as a missed opportunity, particularly in the area of privacy and data protection.⁵⁰⁴ Further, the overall framework has a definite marketing angle, seeking to set “enabling conditions” for the development and deployment of AI-based solutions, encourage investment attraction, and more.⁵⁰⁵

Citron asks targeted questions about substantive aspects of the expertise of administrative agencies.⁵⁰⁶ Reaffirming the “expertise commitment” of specialized agencies is an ongoing task, one which requires constant reinterrogation to confirm that the substantive demands of a specified expertise are being met. The process of constantly reinterrogating principles and priorities requires flexibility in addition to expertise: agencies’ ability to affirm their own expertise offers a guarantee of competence in an ever-changing social and technological landscape.⁵⁰⁷ In more recent work, Citron joins Ryan Calo to map the current usages of ADM in the United States’ administrative context.⁵⁰⁸ Their analysis is preoccupied with a sense of various U.S.-based agencies having rushed into the use of ADM solutions, which then undermines agencies’ unique claims to be “better situated than the legislature to govern daily life.”⁵⁰⁹ Already, over half of federal administrative agencies in the United States are either using or investigating the possibility of using AI.⁵¹⁰ Scholars critical of this transition have tended to focus on the human values supplanted by introducing automation, discussing how procedural fairness guarantees like transparency, accountability, or due process can disappear when a machine takes over a previously human-held task.⁵¹¹ Calo and Citron are critical of this framing, seeking to instead reframe the debate to focus on whether agencies’ use of ADM threatens the constitutionally-blessed delegation of authority that permits the existence of administrative agencies.⁵¹² They remind readers that “[t]he famously ‘functionalist’ rationale for delegation rests on the affordances of bureaucracies, particularly their ability to accrue expertise and the prospect of flexible and nimble responses to complex problems.”⁵¹³ If agencies adopt

⁵⁰¹ Lucilla Sioli, “Shaping Europe’s Digital Future: CEPS Webinar – European approach to the regulation of artificial intelligence” (April 23, 2021), online: *Centre for European Policy Studies* <www.ceps.eu/wp-content/uploads/2021/04/AI-Presentation-CEPS-Webinar-L.-Sioli-23.4.21.pdf> [perma.cc/28KZ-NC37].

⁵⁰² *Ibid.* The categories are not necessarily mutually exclusive, with one technology able to appear in more than one of the AI-risk category.

⁵⁰³ *Ibid.* Such technologies are described as contradicting EU values (see Title II, Article 5).

⁵⁰⁴ See, e.g., David Ruiz, “Artificial Intelligence ban slammed for failing to address ‘vast abuse potential’” (April 23, 2021), online: *MalwareBytes Labs* <blog.malwarebytes.com/artificial-intelligence/2021/04/artificial-intelligence-ban-slammed-for-failing-to-address-vast-abuse-potential/> [perma.cc/S9MB-BPFD].

⁵⁰⁵ Sioli, *supra* note 501.

⁵⁰⁶ Citron, *supra* note 439 at 1249.

⁵⁰⁷ *Ibid.*

⁵⁰⁸ Ryan Calo & Danielle Keats Citron, “The Automated Administrative State: A Crisis of Legitimacy” (2021) 70 *Emory LJ* 797.

⁵⁰⁹ *Ibid.* at 845.

⁵¹⁰ *Ibid.* at 801.

⁵¹¹ *Ibid.* at 802.

⁵¹² *Ibid.* at 803-804.

⁵¹³ *Ibid.*

technological solutions that they do not fully understand, their claim to specialized expertise and nimble response becomes fundamentally compromised.⁵¹⁴

Calo and Citron employ a historical lens, reminding readers that technological progress in the law has always been somewhat uneasy. Through this lens, technologies fascinate us “precisely for their ability to substitute for humans.”⁵¹⁵ Their observations dovetail with the full accounting of cyberlaw history undertaken in Chapter 2, reminding us about Lessig’s observations of how technology reveals law’s “latent ambiguities.”⁵¹⁶ As technological change transforms human behaviours and makes available a broader spectrum of possible choices, corresponding legal puzzles are revealed. For Calo and Citron, undue focus on supplanting specific values precludes an evaluation of law’s first principles.⁵¹⁷ In their view, this results in scholars simply advocating for a return to the status quo, instead of interrogating problems with the existing system that technological approaches might well solve.⁵¹⁸ Scassa offers a similar warning, arguing against allowing concerns about preserving procedural fairness to overshadow other critical systemic issues.⁵¹⁹ Heeding their advice, I turn now to thinking through the complicated nature of the data set ADM is asked to administer, the key characteristics it exhibits, and the corresponding complications that arise when approaching its compilation from a critical perspective.

2. Complications for the Data Set

Technological incursions into the administrative state have created vast stores of data. As the preceding section demonstrates, however, results have been mixed on related objectives: achieving equality through automated analogues or generating regulations that meaningfully adhere to longstanding legal precepts. Yet, many advocates for self-driving laws still extol the virtues of the consistency that it would offer over flawed humans, providing improvements over the fickleness of human decision-making. In this section, I further analyze these types of concerns, thinking through the critiques of legal realists that charge the existing legal system with bias or inaccuracy. Would introducing algorithms truly offer substantial improvements over the system currently in place? Then, I apply these insights to the framing example of the automated red-light camera bot, which provides a demonstration of how decisions might be delegated to a technological substitute.

a. Desperately Seeking Fairness

Concerns about receiving fair treatment animate many inquiries into law’s legitimacy: the ability to attach social meaning to decisions and reasons anchored in consistent procedure; the public

⁵¹⁴ Calo & Citron, *supra* note 508 at 804. They further elaborate: “The automated administrative state is less and less the imperfect compromise between the text of the Constitution and the realities of contemporary governance. At some point, the trend toward throwing away expertise, discretion, and flexibility with both hands strains the very rationale for creating and maintaining an administrative state. This is especially true where, as often, the very same processes of automation also frustrate the guardrails put in place by Congress and the courts to ensure agency accountability.”

⁵¹⁵ *Ibid* at 810. The example Calo & Citron offer is that legal rules assume humans will drive cars; now robots do.

⁵¹⁶ *Ibid*. See also Lessig, *supra* note 253 at 22.

⁵¹⁷ Calo & Citron, *supra* note 508 at 810.

⁵¹⁸ *Ibid*.

⁵¹⁹ Scassa, “ADM,” *supra* note 486 at 7. Citing Montreal Declaration’s preamble, Scassa notes that AI can “restrict the choices of individuals and groups, lower living standards, disrupt the organization of labor and the job market, influence politics, clash with fundamental rights, exacerbate social and economic inequalities, and affect ecosystems, the climate and the environment.” See *Montreal Declaration For a Responsible Development of Artificial Intelligence* (Montreal: Université de Montréal, 2018), Preamble, online: *Université de Montréal* <www.montrealdeclaration-responsibletai.com/> [perma.cc/VC4F-LEVU].

availability of those reasons once given; the faithful observances of due process—these features offer the stability and predictability needed to “feel at home within the framework of the law.”⁵²⁰ Notably, Tom Tyler’s studies into comfort with judicial decision-making underscore the importance of process as a signal of equality and legitimacy.⁵²¹ Tyler maps extensive academic research into the psychology of procedural justice, contrasting instrumental and normative views on process control.⁵²² The instrumental perspective focuses on the degree to which individuals can exert control over the outcomes of third party decisions.⁵²³ By contrast, normative views focus on more aspects of an individual’s overall procedural experience, instead of only the outcomes realized.⁵²⁴ As with our earlier discussion on law’s normativity catalogued in Chapter 2, expressions of particular values are tied to defining the appropriateness of a certain way of doing things.

Tyler defines normative aspects of procedural justice as including “neutrality, lack of bias, honesty, efforts to be fair, politeness, and respect for citizens’ rights.”⁵²⁵ While both instrumental and normative perspectives emphasize the importance that people attach to stating their own opinions, instrumental views consider such opportunities to be “valuable only to the extent that those stating their opinions think they are influencing outcomes.”⁵²⁶ Despite this viewpoint, studies suggest that people value the opportunity to state their individual case to the powers that be, regardless of its potential impact on the eventual outcome.⁵²⁷ People feel more fairly treated when they have the opportunity to participate in the process, even if their participation will ultimately have “little or nothing to do with what the authorities decide.”⁵²⁸ Tyler describes non-instrumental priorities as being “value-expressive,” which means they are not associated with attaining particular outcomes. The preconditions for creating value-expressive outcomes are not entirely clear, but potential hypotheses include impartiality, good faith, and consideration of one’s views.⁵²⁹ Tyler’s research suggests that people feel most secure that a legal proceeding is just when it seems to be impartial, when its authority is exercised with good faith, and when their viewpoints are given the opportunity to be aired.

Another study analysed secondary research on Leventhal’s six criteria of procedural justice to evaluate the relative fairness of a procedure: representativeness, consistency, suppression of bias, accuracy, correctability, and ethicality.⁵³⁰ Overall, the results converged around a deep preoccupation with consistency.⁵³¹ In this sense, consistency refers to similarity of both treatment and outcomes.⁵³² Consistent procedure requires “equal treatment for all affected parties” and means

⁵²⁰ Joseph Raz, “The Law’s Own Virtue” (November 2018), online: *SSRN* <papers.ssrn.com/sol3/papers.cfm?abstract_id=3262030> [perma.cc/U6EC-ZCWX] [Raz, “Law’s Own Virtue”] at 2.

⁵²¹ Tom R. Tyler, “Part Four: The Meaning of Procedural Justice” in *Why People Obey The Law* (Princeton: Princeton University Press, 1990).

⁵²² Tyler notes that most efforts to understand fair procedure flow from the “control theory of procedural justice” developed by Thibaut & Walker (1975), which is centred on “the nature of the control exercised by disputants over the forums used to resolve their disputes ...” and distinguishes between decision control and process control. *Ibid* at 115.

⁵²³ *Ibid* at 6.

⁵²⁴ *Ibid* at 7.

⁵²⁵ *Ibid*.

⁵²⁶ *Ibid* at 116.

⁵²⁷ *Ibid*.

⁵²⁸ *Ibid* at 117.

⁵²⁹ *Ibid*.

⁵³⁰ *Ibid* at 118-120.

⁵³¹ *Ibid* at 120.

⁵³² *Ibid* at 118.

that, over time, a procedure should “follow the same rules and be enacted in the same way each time it is used.”⁵³³ Using sports analogies, Tyler explains that consistency requires baseball umpires to define the same strike zone for all players and that consistency ensures the amount of physical contact that a basketball referee considers a foul should remain the same over the entirety of the game.⁵³⁴ The expectation is that people should be treated the same, with the rules applied consistently to their unique situation. For many reasons, however, such uniform applicability can be questioned in terms of law’s operation.

b. What the Judge Had for Breakfast

Charges from contemporary legal realism can interfere with the procedural fairness guarantees that law proposes to offer. The legal realist view is often represented with the caricatured statement that “justice is what a judge had for breakfast.”⁵³⁵ Legal realism accuses legal decision-making of being influenced by politics, ideology, bias, racism, sexism, other prejudices, or even just the vagaries of human intuition.⁵³⁶ Such accusations appear in many facets of human life: sunshine is well-documented as improving mood, which also means sunshine can improve stock market performance.⁵³⁷ But, defendants seeking lenient sentences must hope the sun does not shine overly brightly: a study of 207,000 immigration decisions found that hot weather outside made it less likely that asylum-seekers would be granted political asylum.⁵³⁸

The chaos, unintended consequences, and subtle pressures induced by quotidian life might have significant consequences in a judicial context, where the implications of subtle biases can impact an accused person’s freedom or life circumstances. Similar critiques are levied across the diverse fields of legal analysis. In a well-known study, authors Shai Danziger, Jonathan Levav, and Liora Avnaim-Pesso analysed Israeli parole board decisions for the impact of judges’ daily food breaks on decision-making.⁵³⁹ Based on their analysis of the decisions judges made in three distinct “decision sessions,” demarcated by two separate food breaks that split the day into sections, they concluded that judges were more likely to produce favourable rulings either at the very beginning of the work day or immediately after a food break.⁵⁴⁰ Even more alarmingly, the percentage of favourable decisions dropped from approximately 65 per cent at the outset of each decision session to nearly zero by its conclusion—but returns “abruptly” to approximately 65 per cent following a food break.⁵⁴¹ The authors interpreted their findings “through the lens of mental depletion,” arguing that their findings supported a view of law’s indeterminacy by demonstrating how “legally irrelevant situational determinants” regularly find their way into legal reasoning.⁵⁴² This research

⁵³³ Tyler, *supra* note 521 at 118-119.

⁵³⁴ *Ibid* at 119.

⁵³⁵ Andrea Glöckner, “The irrational hungry judge effect revisited: Simulations reveal that the magnitude of the effect is overestimated” (2016) 11 *Judgment and Decision-Making* 601.

⁵³⁶ *Ibid*.

⁵³⁷ Ozkan Eren & Naci Mocan, “Emotional Judges and Unlucky Juveniles” (2018) 10 *Applied Economics* 171.

⁵³⁸ Kahneman et al, *supra* note 410 at 17.

⁵³⁹ Danziger et al, *supra* note 421 at 6889.

⁵⁴⁰ *Ibid* at 6890.

⁵⁴¹ *Ibid* at 6889.

⁵⁴² *Ibid* at 6892. In response to being told about this study, a student once asked me: “Do judges need to remind themselves to eat so that they don’t give someone a life sentence?”

suggests a phenomenon of decision fatigue, where the laborious work of deliberation reduces a judge's mental resources to the point that they are less able to make good decisions.⁵⁴³

Acknowledging that psychological consequences of anger, sadness, or sense of well-being can negatively impact decision-making means acknowledging the human frailty that can be reflected in the business of judging. Similarly, a study by Ozkan Eren and Naci Mocan investigates the impact of negative emotions arising from unexpected losses by a Louisiana State University's (LSU) football team.⁵⁴⁴ They hypothesized that unexpected negative outcomes—upset losses—would generate emotional shocks, triggering feelings like anger, frustration and sadness, which might impact judicial decisions.⁵⁴⁵ Key to their study was the fact that the LSU football team had a storied history and a loyal fan base.⁵⁴⁶ Using the Las Vegas pregame point spread as a baseline for fans' rational expectations about the likely outcome of the game, they found that upset losses tended to increase the sentence length imposed on juvenile defendants.⁵⁴⁷ Judges who were LSU alumni were particularly vulnerable to being influenced by this phenomenon.

In *Noise: A Flaw in Human Judgment*, authors Daniel Kahneman, Olivier Sibony, and Cass Sunstein describe the incursions into reliable human decision-making. First, they paint an overly rosy picture of the traditional position of criminal sentencing: that judicial discretion over sentencing enables a just, humane approach to criminality that permits “individualized tailoring” and an approach that transcends the bluntness that would surely result if judges were unduly constrained by rigid rules.⁵⁴⁸ Due process requires a high degree of judicial discretion, to ensure that individuals are treated as such, permitting their specific sentences to be tailored to their individual circumstances. Yet, they argue, this rosy picture of due process is impugned by noise—the observation that different results are highly scattered, making them inaccurate. Looking at criminal sentencing through this lens, Kahneman et al present numerous studies identifying inconsistencies in the types of sentences awarded, revealing substantial disparities.⁵⁴⁹ For example, they report on a 1977 survey of judges who, when tasked with responding to the same five cases detailing low-level offences, produced highly disparate results.⁵⁵⁰ Some judges assigned prison time for simple possession of marijuana, while others merely recommended probation.⁵⁵¹ A burglary

⁵⁴³ The idea of “decision fatigue” or “ego depletion” suggests that willpower is a finite resource that can be expended through the making of many small decisions throughout the day. Proponents of this idea suggest that eliminating small decisions can preserve executive function for more important tasks. This is credited as the reason why executives like Steve Jobs, Mark Zuckerberg, and Barack Obama tended to wear the same outfit as “uniform” day in and day out—it prevented them from using finite decision-making capacity to decide on the day's attire. The research leading to this conclusion was first published by R.F. Baumeister in 1998 and has since been cited over 2763 times. See generally, RF Baumeister, “Limited Resources for Self-Regulation: A Current Overview of the Strength Model” in Edward R Hirt et al, *Self-Regulation and Ego Control* (London, UK: Elsevier, 2016) 185. Of course, follow-on research has complicated the findings, with some researchers arguing the papers resulted from publication bias and *p*-hacking: see generally, Malte Friese et al, “Is Ego Depletion Real? An Analysis of Arguments” (2019) 23 *Personality and Social Psychology Review* 107.

⁵⁴⁴ Eren & Mocan, *supra* note 537.

⁵⁴⁵ *Ibid* at 172.

⁵⁴⁶ *Ibid*

⁵⁴⁷ *Ibid* at 172-173.

⁵⁴⁸ Kahneman et al, *Noise*, *supra* note 410 at 13-14.

⁵⁴⁹ *Ibid* at 16.

⁵⁵⁰ This study was performed by William Austin and Thomas Williams, who gave 47 different judges the same 5 cases to review. Each judge obtained “summaries of the information used by judges in actual sentencing, such as the charge, the testimony, the previous criminal record (if any), social background, and evidence relating to character.” The study revealed “substantial disparity.” *Ibid*.

⁵⁵¹ *Ibid*.

case ran the gamut of possible sentences: ranging from five years in prison to thirty days—plus a \$100 fine.⁵⁵²

Other, similar studies undergird their chief contention: human decision-making is highly noisy, meaning that humans given the same information might arrive at disparate results. Importantly, for our purposes, this means judges given the same fact situations will arrive at very different results, calling into question the important belief that everyone who faces the judicial system is being treated equally under the law. Kahneman et al distinguish noise from bias: noise generates scattered results, while bias generates consistent results that are nonetheless off target in some systematic way.⁵⁵³ Using the metaphor of target shooting, they show the distinct appearance of these two types of human error, with bias appearing as “systematic deviation” and noise appearing as “random scatter.”⁵⁵⁴ For a legal system preoccupied with treating litigants equally and assigning just sentences, this human shortcoming is a clear problem. In 1954, psychology professor Paul Meehl tested the limits of this problem by reviewing studies of clinical judgment compared against a modelled, mechanical prediction, concluding that deploying rote mechanical rules reliably outperformed human judgment.⁵⁵⁵ Overall, his findings stressed that “clinicians and other professionals are distressingly weak in what they often see as their unique strength: the ability to integrate information.”⁵⁵⁶

Faced with these types of conclusions, the draw towards automated forms of decision-making is strong. Through meticulous, data-driven assessments, one might argue, the legally irrelevant situational determinants can be extricated from the process, producing purer, more determinate articulations of law. As Meehl’s conclusions posit, deploying mechanical processes will enable the integration of information—a process that will only improve in accuracy and consistency with the new power offered by ML models. However, while models do offer some improvement over human decision-making, the relative advantage is meagre at best. In a median review of mechanical formulae *versus* clinicians, the formulae boasted a success rate of 73% of the time—only 5% better than the clinicians’ score of 68% success.⁵⁵⁷ For mechanical prediction *versus* human judges of employee performance, the separation was similar: 59% compared against 65%. AI might offer even better success rates. Yet, predicting the future remains a thorny task, further complicated by problems of induction and the denial of ignorance.

c. Problems of Induction and The Denial of Ignorance

Philosopher Bertrand Russell explained the difficulty of predicting the future based on the past with his parable of the inductivist turkey.⁵⁵⁸ He describes a turkey on a farm, who is fed each day at nine o’clock in the morning. Although he begins to see a pattern exhibited, our turkey is a good inductivist: he does not jump to conclusions without sufficient evidence and withholds judgment.

⁵⁵² Kahneman et al, *Noise*, *supra* note 410 at 16.

⁵⁵³ *Ibid* at 4.

⁵⁵⁴ *Ibid*.

⁵⁵⁵ *Ibid* at 114.

⁵⁵⁶ *Ibid*.

⁵⁵⁷ *Ibid* at 143.

⁵⁵⁸ This story has attained folk-legend status, and many versions of it can be found online and in philosophy texts. Its foundational citation, however, does not mention a turkey at all: the parable is about a chicken! Bertrand Russell, *The Problems of Philosophy* (New York, NY: Henry Holt and Company, 1912) at 93.

Day after day, month after month, our turkey is fed at nine o'clock. Eventually, after months of careful data collection, across different days of the week and weather conditions, our turkey draws his conclusion: he will be fed breakfast at nine o'clock every day until the end of time. Of course, our turkey reaches this conclusion on the morning of December 24—Christmas Eve, when the farmer has a different idea for what our turkey's fate should hold.

Law also suffers from this problem of induction. Things are always done a certain way, binding precedents are always followed, until one day, when there's a sudden break from precedent. Perhaps the Supreme Court's composition changed to judges with different visions for law's trajectory. Perhaps social context has shifted such that a once-radical idea has become commonplace—its time has finally come. To suppose that legal knowledge could attain completion is to commit the same fallacy as Russell's inductivist turkey. Sometimes, things happen that have never happened before. These occurrences are simply not in the dataset.

On one hand, the observation that predicting the future is difficult may not be an innovative contribution.⁵⁵⁹ Kahneman et al make a similar observation in their text, writing “[a]dmittedly, asserting that the future is unpredictable is hardly a conceptual breakthrough.”⁵⁶⁰ The co-authors are quick to note that despite this fact's sheer obviousness, it is nonetheless regularly ignored, as studies repeatedly demonstrate “predictive overconfidence.”⁵⁶¹ While models may consistently produce better results than people, their overall improvement is somewhat marginal.⁵⁶² Moreover, people who believe themselves capable of accurately predicting the future tend to overestimate their own predictive capabilities. It seems likely that judges would fall into this category. Not only are they overconfident in their own abilities, or prone to deny the possibility of noise or bias in their results, but as Kahneman et al caution, they also believe themselves able to predict the truly unpredictable.⁵⁶³ The authors term this phenomenon the “denial of ignorance.”⁵⁶⁴

For Meehl, who remained dumbfounded that his observations on the advantages of modeling were largely unheeded in his lifetime, the denial of ignorance offers a plausible explanation. When decision-makers rely on their own intuition, they are rewarded emotionally, by heeding an internal signal and listening to a confident voice capable of “knowing without knowing why.”⁵⁶⁵ Quantifying the actual predictive metric will not provide a numerical basis for that level of confidence, making it hard to relinquish the emotional reward, especially in scenarios of uncertainty.⁵⁶⁶ When the dataset does not contain the necessary data to delineate a clear path

⁵⁵⁹ As Yogi Berra famously said, “It's tough to make predictions, especially about the future.” The provenance of this famous quotation is often contested, with different sources attributing it to Danish physicist Niels Bohr, Danish artist Robert Storm Peterson, and American writer Mark Twain—among others. See Quote Investigator, “It's Difficult to Make Predictions, Especially About the Future” (20 October 2013), online: <quoteinvestigator.com/2013/10/20/no-predict/> [perma.cc/3P6Z-9WJN].

⁵⁶⁰ Kahneman et al, *Noise*, *supra* note 410 at 144.

⁵⁶¹ *Ibid.*

⁵⁶² *Ibid* at 143.

⁵⁶³ *Ibid* at 145.

⁵⁶⁴ *Ibid.*

⁵⁶⁵ *Ibid.*

⁵⁶⁶ *Ibid.*

forward, people are more likely to turn to their intuition to provide “the sense of understanding and confidence they crave.”⁵⁶⁷

People are prepared to trust an algorithm that achieves a very high level of accuracy because it gives them a sense of certainty that matches or exceeds that provided by the internal signal. But giving up the emotional reward of the internal signal of a high price to pay when the alternative is some sort of mechanical process that does not even claim high validity.⁵⁶⁸

For those who continue to extol the virtues of the traditional legal system over self-driving laws, these observations about the emotional reward of heeding one’s intuition ought to be studied carefully. Advocates for human judgment might point to the marginal benefits of mechanical or automated processes, emphasizing the close percentages in accuracy. If the difference is so slight, human advocates might say, why bother shifting to the slightly better process? Kahneman et al respond to this “why bother?” attitude by pointing out that no one would deliberately select a lottery ticket with a 59% chance of winning over one with a 65% chance.⁵⁶⁹

Proposals for self-driving laws anticipate the law’s ability to offer right answers that can be easily reduced to datapoints to train ML algorithms to replicate those right answers. This type of formalist conception of law presumes that the law never runs out: there are no real hard choices because law’s dominion always offers a right answer; those right answers can be drawn from a comprehensive database of all legal knowledge. And information science is on track to deliver some version of this reality. With the advent of surveillance capitalism, corporations have amassed large datasets of user behaviour that produce increasingly uncanny predictions of what users might want to see.⁵⁷⁰ The enormous financial benefits enable advertisers to showcase products their algorithms predict users will want.⁵⁷¹ Predictive analytics has transformed corporate business models, but its impacts also implicate individual autonomy in nefarious ways. When nudged towards a particular advertisement, how certain can a user be that they would have purchased that item without direction? Prediction can collapse into pre-emption, as real individual choices are precluded.⁵⁷²

Presuming the availability of right answers in system design discounts important aspects of legal reasoning: it refuses acknowledgment of ambiguity; it fails to engage in interpretation of words or phrases with multivalent meanings; it refuses to accept that new developments in society present new, often unanticipated questions; it risks entrenching particular realities for future generations. The common law’s inherent flexibility anticipates re-imaginings and reversals to remain relevant and capable of maintaining social force in changing society. Transposing legal rules to automated administration risks a “set it and forget it” mentality, a phenomenon that Rebecca Crootof calls “technological-legal lock-in.”⁵⁷³ And, this mentality might just lock in an unequal society before it can make headway on the significant work of being reimagined.

⁵⁶⁷ Kahneman et al, *Noise*, *supra* note 410 at 145.

⁵⁶⁸ *Ibid* at 146.

⁵⁶⁹ *Ibid*.

⁵⁷⁰ Shoshanna Zuboff, *The Age of Surveillance Capitalism* (New York, NY: PublicAffairs, 2019) at 16.

⁵⁷¹ *Ibid* at 67.

⁵⁷² Ian Kerr, “Prediction, pre-emption,” *supra* note 67.

⁵⁷³ Rebecca Crootof, “‘Cyborg Justice’ and the Risk of Technological-Legal Lock-In” (2019) 119 *Columbia L Rev Forum* 233.

d. Traffic Bot in the Loop: What Does This Example Show?

Fundamentally, Tyler's lessons ask us to think about legitimacy in a substantive way. One of Tyler's lessons is that people are preoccupied with consistency, meaning a focus on this value might promote a normative understanding of law as legitimate. Concurrently, however, Tyler's lessons also teach that people need to feel heard to make sense of their legal situation, even if being heard doesn't ultimately influence the overall outcome. While not diametrically opposed, these values can sometimes be difficult to reconcile. The experience of being heard seems to connote some degree of personalization, being responsive to the specific situation at hand. In fact, the development of modern case law has tended away from treating all offenders the same, as prescribed by strict penal codes. One might characterize this desire as a recognition of the need for context-dependency. Kahneman et al tackle context-dependency head-on in their text, citing the US Supreme Court case *Woodson v North Carolina*, which found mandatory death sentencing to be unconstitutional because it was comprised of general rules meant to apply across the board.⁵⁷⁴ The key problem with such policies, the Supreme Court expressed, is that:

It treats all persons convicted of a designated offense not as uniquely individual human beings, but as members of a faceless, undifferentiated mass to be subjected to the blind infliction of the penalty of death.⁵⁷⁵

While the death penalty confers “especially high stakes,” Kahneman et al note that its essence is easily extrapolated to other scenarios: the application of “overly rigid, noise-reducing rules” always threatens the holistic evaluation of an individual.⁵⁷⁶ Coupled with the observations that reliance on big data can create biased results that exacerbate inequality, concerns arise about the legitimacy of widespread deployment of ADM.⁵⁷⁷ While judges may indeed be fickle decision-makers, it seems the available algorithms are only marginally better—and might result in other losses that further prejudice the system. Making predictions remains difficult, even for people who have spent years cultivating a specific area of expertise. Compiling additional data, it seems, does not necessarily solve this problem. And, research into human cognition reveals that the narrative schema and prototypes that people use to understand the world might ensure that those predisposed to view judges' decisions as biased will continue to do so, even if research confirms the opposite.⁵⁷⁸

With these lessons in mind, we return to the framing example of the automated red-light camera bot. Even deploying this simple binary automated replacement raises at least two possible objections. First, there is the usual objection: there may be scenarios where it is appropriate to let

⁵⁷⁴ Kahneman et al, *Noise*, *supra* note 410 at 333.

⁵⁷⁵ *Woodson v North Carolina*, 428 U.S. 280 (1976) at 304.

⁵⁷⁶ This is true in many diverse scenarios, many of which don't involve law at all: teachers and students; doctors and patients; employers and employees; coaches and athletes; etc. Kahneman et al, *Noise*, *supra* note 410 at 333.

⁵⁷⁷ And for this reason, as discussed in Chapter 1, Casey and Niblett recommend using AI to develop personalized law: Casey & Niblett, *supra* note 37.

⁵⁷⁸ Kahan draws this conclusion from his work into cognition in the context of judicial decision-making. He writes: “Observers often perceive judges' decisions to reflect the influence of ideological values or other extraneous influences. But the pervasiveness of this impression is obviously not evidence that judges' decisionmaking is infected by the dynamics of untamed Bayesianism or any other mechanism hostile to neutral decisionmaking. Indeed, mock juror studies that investigate how heterogenous narrative scheme and prototypes interact with identity-protective cognition imply that lay persons of opposing cultural predispositions could be expected to *perceive* judges' decisions to be biased even if profession judgment rendered judges wholly immune to such dynamics.” Kahan, *supra* note 57 at 60.

an infraction slide because rote application of the rule is at odds with basic notions of justice. Second, there is a more nuanced one: that even though the outcome is presented as binary, the decision-making steps taken to get there are complex, variable, and easily obfuscated—and their true operation will not be meaningfully represented by an automated process.

Beginning with the second objection: if we carefully interview the experienced, dedicated, and well-meaning police officer for the list of steps typically undertaken when monitoring red light violations, we can expect to obtain a reasonable preliminary list of operational criteria. On the basis of this interview, designers could then define a set of criteria that would need to be strictly followed in order to produce an automated process to achieve the same results. At first glance, it might appear as though this is a simple machine to generate. However, the individualized steps discretized in the officer interview would not necessarily be reflected in the red-light camera's methodology. The steps might be slightly different, depending on which officer was selected for interview. And the complexity of human thought processes poses an additional constraint: some knowledge is tacit, and not easily recognizable or explainable.

As technology advances, this automated process might be transformed from a simple decision-tree powered by “if this, then that” architecture to a more advanced ML approach. As the ML algorithm learns, it continues to develop more efficient, more advanced versions of itself. The ML approach would necessarily rely on information collected from powerful sensors to augment, streamline, or “personalize” each individual outcome. Subtleties in information collection from powerful sensors mean that the ML approach may derive alternative means for achieving its goal that are not represented on the decision tree nor explicitly coded into the system. As Grimmelmann explains, software systems can frequently regulate without any meaningful transparency into their operative processes, fueled by software's plasticity and automated operation, which “bedevil attempts to have software explain itself.”⁵⁷⁹ Even experts, Grimmelmann notes, may sometimes not be able to explain why an automated program acts the way that it does.⁵⁸⁰

Moreover, the first objection remains: context-dependency. Indeed, flexibility is extolled as a virtue of the administrative state in direct recognition of such scenarios. The rote application of rules may generate disproportionate results, where litigants experience the blunt force of law, if justice is not assessed on a case-by-case basis. Such is the ongoing debate in situations involving mandatory minimums, where criminal codes prescribe set sentences for given crimes, irrespective of the unique life circumstances of the accused, or other mitigating factors that might traditionally be considered in sentencing.⁵⁸¹ Our red-light traffic bot is designed to behave uniformly, always issuing a ticket if someone were to proceed through the intersection while the light is red. However, we can imagine scenarios where a police officer deployed in the same role would offer leniency. Imagine, for example, a middle-of-the-night rush to the hospital with a pregnant passenger in the throes of labour. While the traffic bot would still issue the ticket, for it to arrive in the mail a few weeks later in the sleepless haze of early parenthood, it is hard to imagine a police officer doing the same. A police officer would likely, at the very least, allow the prospective parents

⁵⁷⁹ Grimmelmann, “Regulation,” *supra* note 254 at 1723.

⁵⁸⁰ *Ibid.*

⁵⁸¹ See, for example, debates in the United States about the validity of mandatory minima for crack cocaine drug offences. ACLU, “Cracks in the System: Twenty Years of Unjust Federal Crack Cocaine Law” (October 2006) online: *ACLU* <www.aclu.org/other/cracks-system-20-years-unjust-federal-crack-cocaine-law> [perma.cc/RR6Q-J22G].

to pass unabated—if not offer a police escort to the hospital. Even though a consistent approach to red light violation management would require issuance of a ticket, asking new parents to accept both a conviction and the corresponding fine seems unduly harsh in the circumstances.

The nuances of rule-following are a central consideration in analytical jurisprudence. HLA Hart conceptualizes law as a union of primary and secondary rules. In his characterization, primary rules force a particular obligation onto an individual, while secondary rules allow for the ability to change or transform the primary rules.⁵⁸² Secondary rules are what Hart calls “power-conferring rules.”⁵⁸³ Meanwhile, primary rules arise in places like criminal codes or specific regulations: they are “duty-imposing” in that they tell individuals how they ought to behave.⁵⁸⁴ Hart distinguishes between the ability to be concerned with rules as either “an observer who does not accept himself them” or as “a member of the group which accepts them and uses them to guide conduct.”⁵⁸⁵ These distinctions are more commonly known as an external *versus* an internal point of view. From an external point of view, it is not possible to understand how the rules function in the lives of those members of society who have adopted them. Yet, for social and legal rules to persist, they must also have an internal aspect, which explain group members’ “regularities of behaviour” by virtue of a “critical reflective attitude.”⁵⁸⁶ As Scott Shapiro explains, “[m]embers of the group act *as a rule* because they accept *there is a rule*.”⁵⁸⁷

Hart also uses the example of a traffic light to convey his analysis of rule-following. He further explains the distinction between internal and external rules:

If, however, the observer really keeps austerely to this extreme external point of view and does not give any account of the manner in which member of the group who accept the rules view their own regular behaviour, his description of their life cannot be in terms of rules at all, and so not in the terms of the rule-dependent notion of obligation or duty. Instead, it will be in terms of observable regularities of conduct, predictions, probabilities, and signs. For such an observer, deviations by a member of the group from normal conduct will be a sign that hostile reaction is likely to follow, and nothing more. His view will be like the view of one who, having observed the working of a traffic signal for some time, limits himself to saying that when the light turns red there is a high probability that the traffic will stop. He treats the light merely as a natural *sign that* people will behave in certain ways, as cloud as a *sign that* rain will come. In so doing he will miss out a whole dimension of the social life of those whom he is watching, since for them the red light is not merely a sign that others will stop: they look upon it as a *signal for* them to stop, as so a reason for stopping in conformity to rules which make stopping when the light is red a standard of behaviour and an obligation. To mention this is to bring into the account the way in which the group regards its

⁵⁸² HLA Hart, *The Concept of Law* (Oxford: Oxford University Press, 1961) at 82.

⁵⁸³ *Ibid* at 96.

⁵⁸⁴ *Ibid* at 85-86.

⁵⁸⁵ *Ibid* at 89.

⁵⁸⁶ Scott Shapiro, “What Is the Internal Point of View?” (2006) *Fordham LJ* 75 at 1164-1165.

⁵⁸⁷ *Ibid* at 1165.

own behaviour. It is to refer to the internal aspect of rules seen from their internal point of view.⁵⁸⁸

On this account, it would be difficult to characterize any ADM, no matter the degree of sophistication of its AI or ML, as capable of making a rule-based decision that would satisfy Hart. In some ways, ADM seems like the ultimate outsider, destined to “participate” in law only computationally, *i.e.*, “in terms of observable regularities, predictions, probabilities and signs.”⁵⁸⁹ It would not possess the tacit knowledge gleaned from community participation, the appreciation of bias levied by equality-seeking critiques, or the context-sensitivity needed to know when it is appropriate to let a particular rule slide. For understanding a Rule of Law powered by self-driving laws, this outsider status presents theoretical difficulties, especially in the face of such a contested concept. I return to analytical jurisprudence in Chapter 4. First, however, I consider the idea of rule-following through the lens of Roscoe Pound’s “mechanical jurisprudence.”

3. Mechanical Jurisprudence

While the automated traffic bot offers a simplistic entry point into some of the important thematic concerns of this thesis, my primary analytical framework is powered by a more advanced version of AI or ML: the “self-driving laws” described in Chapter 1. Self-driving laws would automate the key functions of legal research and analysis, creating machine-generated outcomes for problems of rule-application and more. Recent technology law scholarship, canvassed in Chapter 2, has attempted to describe such innovative algorithmic techniques and their intersections with the administration of justice. Yet, some of the most profound comparators for how legal theory intersects with rote, mechanical denunciations of legal ideas come from scholars whose work significantly pre-dates the availability of AI/ML methods. Notably, Ronald Dworkin’s idealized judge Hercules, to be discussed at length in Chapter 4, exhibits many of the same features as an algorithmic process, with immediate access to all knowledge and unlimited wisdom powering the calculation of the perfect solution. Dworkin’s ideas about right answers and interpretative theory, primarily written in the 1970s and 1980s, came long before the personal computing revolution or widespread availability of significant database services. Writing in 2001, Cass Sunstein explains that AI will be able to meaningfully participate in analogical legal reasoning when it becomes capable of making good evaluative judgments.⁵⁹⁰ Sunstein suggests this contention is illuminated by Dworkin’s insights on legal reasoning: it is Dworkin who points out that analogical thinking must be powered by theory or principle to unify (or divide) the case at hand from those that have come before.⁵⁹¹

Even more prescient are the observations of Roscoe Pound, who followed in Oliver Wendell Holmes’ footsteps by attempting to repair unduly formalist approaches to law. In a 1908 paper, Pound coined the phrase “mechanical jurisprudence” to describe what later generations would refer to as robotic applications of rules to legal scenarios.⁵⁹² Pound’s chief contention is that rote

⁵⁸⁸ Hart, *supra* note 582 at 89-90.

⁵⁸⁹ Kerr and Mathen make this point with respect to their idealized robot decision-maker, JR-R. See Kerr & Mathen, *supra* note 58 at 29.

⁵⁹⁰ Cass Sunstein, “Of Artificial Intelligence and Legal Reasoning” (2001) 8 University of Chicago Law School Roundtable 29 [Sunstein, “Of AI & LR”].

⁵⁹¹ *Ibid* at 32.

⁵⁹² Roscoe Pound, “Mechanical Jurisprudence” (1908) 8 Columbia L Rev 605.

applications of rules impoverish the common law method, especially considering how procedure operates to generate rigid, codified practices.⁵⁹³ Such severe abstractions of procedure can deny justice, as “[e]very time a party goes out of court on a mere point of practice, substantive law suffers an injury.”⁵⁹⁴ Pound worries that legislative law-making is “in the grip of the imperative theory and its arbitrary results” while judicial decision-making is “in the grip of a jurisprudence of conceptions and its equally arbitrary results.”⁵⁹⁵ Over 110 years ago, he identified a similar desire to produce consistent, scientific legal results at odds with the pragmatic, sociological impetus behind much of law and policy-making. As remedy, he called for legal scholars to “test the conceptions worked out in the common law by the requirements of the new juristic theory,” grappling between those jurists “who think it is a great display of learning and of judicial independence to render ... ‘strong decisions,’ and those who fix their gaze upon the raw equities of a cause and forage in the books for cases to sustain the desired result.”⁵⁹⁶

In a 2014 conference paper, Ian Kerr and Carissima Mathen provide an intriguing thought experiment by asking readers to assume that the Chief Justice of the United States Supreme Court, John Roberts, is found to have been an advanced social robot, unaware of his own mechanical nature during his distinguished legal career.⁵⁹⁷ The goal of this experiment was to test the integration of a social robot into society without assistance or recognition.⁵⁹⁸ Their conceptualization, the authors posit, is meaningful for a coming era of mechanical jurisprudence, which they describe as “an era where the production of at least some legal knowledge and decision-making is delegated to machines and algorithms, not people.”⁵⁹⁹ Importantly, they note that Justice Roberts, Robot (JR-R)’s accidental accension to his position is a critical aspect of the thought experiment; they are not interested in analysing claims of illegitimacy that might arise from human power being inappropriately delegated to machine analogues.⁶⁰⁰

Kerr and Mathen acknowledge that their thought experiment offers a specific lens on the mechanical jurisprudence question. The central conceit of the paper requires that JR-R happened into his authority accidentally, with no living humans left to oversee the project, thereby obviating the need for discussions of legitimate delegation. Early in the paper, however, they lay out some complementary inquiries, asking whether JR-R may ethically or legally delegate his decision-making authority to a machine, and what kind of transparency is required if he did so.⁶⁰¹ They do not propose to deal with these inquiries, but they nonetheless lay out a template for the animating concerns of this thesis. They point out that judges legitimately rely on law clerks to do research and drafting; those clerks might already use technological means of producing their work that the assigning judge neither knows about nor understands.⁶⁰² Would similar delegation to an artificial cognizer be appropriate? What kinds of information would need to be known by democratic official about the artificial cognizer’s operations and potential biases before one could sign off on

⁵⁹³ Pound, *supra* note 592 at 617.

⁵⁹⁴ *Ibid* at 619.

⁵⁹⁵ *Ibid* at 621.

⁵⁹⁶ *Ibid* at 622.

⁵⁹⁷ Kerr & Mathen, *supra* note 58.

⁵⁹⁸ *Ibid* at 2.

⁵⁹⁹ *Ibid* at 4.

⁶⁰⁰ *Ibid* at 5.

⁶⁰¹ *Ibid* at 6.

⁶⁰² *Ibid*.

the delegated authority? What sorts of limitations ought to be imposed to ensure that judicial power is retained by those who legitimately hold office?

As delegation increases, automated decision-making analogues may grow ever closer to the “self-driving laws” theorized by Alarie and others. While it is likely that any adoption of AI solutions for legal industry will be incremental, with lawyers maintaining supervisory authority through the “human-in-the-loop” mentality,⁶⁰³ this outcome is by no means guaranteed. And, even if things start out incrementally, cautious philosophies might later be overridden by calls for innovation, efficiency, or other desirable values. Supplanting human knowledge with automated, mechanical jurisprudence does sometimes appear to be an inevitability as technology progresses.⁶⁰⁴ Kerr and Mathen express their fear about the eventual culmination of this process, writing:

The professional may trust his or her own intuitions. But evidence-based reasoning will suggest that the machine’s decision ought to be followed. The more often professionals delegate decision making to machines, the more they will relinquish control. The more they relinquish control, the more they become dependent on machines. The more dependent they become on the machine, the more they relinquish professional expertise. And, so on.⁶⁰⁵

Relinquishing professional expertise may spell disaster for the cultivation of ethical decision-making as a uniquely human skillset.⁶⁰⁶ The act of judging necessarily requires the “ability to meaningfully follow rules and adopt a particular point of view of a legal system.”⁶⁰⁷ The professional expertise judges bring to their roles as experts can be difficult to artificially replicate.⁶⁰⁸ While an AI/ML-enabled bot is certainly equipped to follow rules, it offers little beyond their rote application. Kerr and Mathen point out that meaningful judicial participation has a community-focused social aspect, where judges are expected to be “a member of the community, understanding its history, its moral convictions, having a point of view about its current character and having a stake in its future.”⁶⁰⁹ Pound’s response to legal formalism encapsulated many of the same ideas.⁶¹⁰ He observed the task of the judge being to “make a principle living,” which should be achieved not by merely deducing it from rules, but instead by “achieving thoroughly the less ambitious but more useful labor of giving a fresh illustration of the intelligent application of the principle to concrete

⁶⁰³ Recall the discussion on “human-in-the-loop” decision-making in automation, discussed in Chapter 1.

⁶⁰⁴ Cass Sunstein, pessimistic about the abilities of “strong” AI in the early 2000s, was careful to leave open the possibility that computers would one day be sufficiently powerful to complete analogical reasoning. The undertone of his analysis suggests the coming potential of AGI: not quite yet, but one day soon. Sunstein, “Of AI & LR,” *supra* note 590 at 34.

⁶⁰⁵ Kerr & Mathen, *supra* note 58 at 8.

⁶⁰⁶ In other work, Kerr cautions against automation that prescribes certain courses of human behaviour, noting that the trial-and-error of typical human activities allows for the cultivation of virtue. Following Aristotle’s *Nicomachean Ethics*, Kerr argues that allowing automation of a certain “correct” course of action will supplant the very human activity of making a mistake and learning from it, with disastrous consequences. Ian Kerr, “Digital Locks and the Automation of Virtue” in Michael Geist, ed, *From Balanced Copyright to Radical Extremism* (Toronto: Irwin Law, 2010) 247 [Kerr, “Digital Locks”].

⁶⁰⁷ Kerr & Mathen, *supra* note 58 at 39.

⁶⁰⁸ Kahan, *supra* note 57 at 59.

⁶⁰⁹ Kerr & Mathen, *supra* note 58 at 39-40.

⁶¹⁰ Kerr and Mathen describe Pound’s position taken in this paper as follows: “Pound believed that the logic of previous precedents alone would not solve jurisprudential problems and decried the ossification of legal concepts into self-evident truths.” *Ibid* at 9.

cause, producing a workable and a just result.”⁶¹¹ Pound describes this as the genius of our common law: distinct from the deeply codified civil codes.⁶¹²

Yet, for self-driving laws, the ability to respond dynamically to emerging social problems will be limited by the imposed systems. While the legal futurists dream of truly emergent legal singularity-styled technology that can proactively respond to new challenges as they arise, the current state of the technology is at odds with such dreams. Meanwhile, locking in a particular version of reality through binary code would serve to undermine sociological responses to jurisprudence that Pound believed should be used to better describe legal matters.

* * *

Many technology law scholars’ discussions of regulation for algorithmic decision-making are rooted in an explanatory function of the law. We know inherently about the value of an explanation: it is much more satisfying to be told why you cannot do something, as opposed to be simply told no and expected to abstain accordingly. People care deeply about both process and the stories we tell about the law. The entire administrative mechanism rests on the idea that justice can only be served if the appropriate steps are carried out. As Tyler’s research reminds us, people are so attached to process that they report being more satisfied with a proceeding if their procedural justice needs are satisfied than if they received their preferred outcomes.⁶¹³

Central to this approach is an expectation that legal decision-making requires a detailed account of how a decision was arrived at, which would allow the subject of the decision or a neutral third-party to follow the thread of the decision-maker’s reasoning from beginning to end. This expectation is powered by the uniquely human context described by Kerr and Mathen. Arguing in parallel, Selbst and Barocas suggest that the legal system is designed to be “inherently contrastive.”⁶¹⁴ If you’re cross-examining a bank robber, asking him the question “Why do you rob banks?” is necessarily context-specific. Posing the question “Why do you rob banks?” is associated with a silent juxtaposition of: “... as opposed to obey the law?” The second half of this query is understood, albeit unspoken. You are not inquiring, “Why do you rob banks, as opposed to something else?” Everyone understands, from context, the purpose of the question.⁶¹⁵

Such contrast requires some realization of context to achieve the necessary balance. If we believe Hart on the necessity of adopting the internal point of view, then there is something tricky about appreciation of context and how it fits into an account of legal decision-making. Certainly, there might be simple situations that could be more easily delegated to ADM. Imagine a similar argument in a robot-assisted surgery, where technological approaches would typically improve on accuracy and medical outcomes. Still, there might be some sort of line where it is necessary that human decision-making come back into the loop. Perhaps it would be realized during hand surgery

⁶¹¹ Pound, *supra* note 592 at 622.

⁶¹² *Ibid* at 623. His final lines read almost as a benediction: “Let the principles be formulated by whom or derived from whence you will. The Common Law will look to courts to develop and expound them, the Civil Law to doctrinal treatises. It is only a lip service to our common law that would condemn it to a perpetuity of mechanical jurisprudence through distrust of legislation.”

⁶¹³ Tyler, *supra* note 521.

⁶¹⁴ Selbst & Barocas, *supra* note 401 at 8, citing W Bradley Wendel, “Explanation in Legal Scholarship: The Inferential Structure of Doctrinal Legal Analysis” (2011) 96 Cornell L Rev 1035 at 1059.

⁶¹⁵ I borrow this excellent example from Selbst & Barocas, *ibid*.

on a concert pianist, where the stakes of a mistake seem elevated. Abstracting back from that example, equality or consistency ought to require the same level of care irrespective of the patient on the table.

It is not novel to be identifying some issues *vis-à-vis* the implementation of technology. To simply decry the loss of human values in the face of technological process would be an overly simplistic critique of what is happening when AI systems displace human decision-making. Recall the Ofqual results, where students were given algorithmically-determined grades irrespective of their actual performance on their A-level exams. This piques a sense of basic (in)justice, where students worked so hard only to be assigned random scores. These scores obviated their sense of individuality before a system that was working to consider their future. Even if the scores were statistically accurate, the outcome nonetheless raises a sense of human impotence in the face of the power of AI and ML. And even if it was human decision-makers at a higher, more abstract level, who made the choice to delegate authority to the AI, the twin phenomena of automation bias and automation complacency add a layer of technological gloss to this administrative decision. When such power is connected to legal decision-making, an area where the rationale for how decisions are appropriately made continues to be contested through a jurisprudential lens, the issue becomes all the more complicated.

The issues canvassed in this Part, and particularly Chapter 3, demonstrate both practically and theoretically that law retains issues of self-understanding that it struggles to deal with. Taken together, these critiques signal a need to make some deeper decisions about what law really is before its regulation may be delegated to an external entity. Before robots can be assigned law's power, we need a better idea of what law is. We take up these issues in the next Part.

Part II: Complicating the Law

*The life of the law has not been logic:
it has been experience.*
~ Oliver Wendell Holmes⁶¹⁶

Once, at a robot law workshop, I was listening to a panel of well-known technology law academics discussing the difficult-to-define contours and limitations of technology law *versus* policy. At the heart of this discussion were the types of concerns discussed in Chapters 2 and 3: the embedded values articulated by technology, the scripts they propose for decision-makers to follow, the pacing problem generated by innovation moving straight to consumers without public understanding or legislative oversight, and so on. The conversation turned to the separate spheres occupied by the scientists and engineers who build the technology and the lawyers, regulators, and policymakers who attempt to control its use. One eminent scholar expressed his frustration with laypeople who claim lawyers, as contrasted with scientists and engineers, don't "build" anything. He boldly declared that lawyers actually do build something: "we build the Rule of Law, and it's incredibly important, thank you very much!"

But I found myself wondering: is this really true? Do the actions of lawyers contribute to the creation of some critically important and univocal Rule of Law? Is the overall operation of the justice system aimed towards this objective? If so, does it typically succeed? And would its success continue if its operations were delegated to self-driving laws? Or would the interpretative functions of law collapse into mere administration, separating human judgment from the Rule of Law?

As a concept, the Rule of Law is often asked to do a lot of heavy lifting. It is caricatured in law, in politics, and at their intersections, usually under the guise of providing legal certainty. At the extreme authoritarian end of the spectrum, those who seek to justify instances of state enforcement, like policing, militarization, or a general preoccupation with law and order, often invoke the Rule of Law as the laudable ideal being pursued. These refrains become particularly plaintive in moments of civil disobedience, perceived inconvenience, or protest. At the liberal democratic end of the spectrum, the Rule of Law is championed as a central guarantee of Western democracies, often due to the inclusion of more robust substantive content.⁶¹⁷ It is also readily invoked in the variable middle, across a seemingly contrasting range of societies, cultures, and worldviews.⁶¹⁸ Two recent examples in contemporary Canadian society typifying this tenor of conversation are the

⁶¹⁶ Oliver Wendell Holmes Jr, *The Common Law* at 1, online (eBook): *Project Gutenberg* <www.gutenberg.org/files/2449/2449-h/2449-h.htm> [perma.cc/YZ6L-SW3H].

⁶¹⁷ Yet, as positivists charge, pouring additional substantive content into the Rule of Law can cast doubt on the certainty of legal outcomes. Raz explains: "If the rule of law is the rule of the good law then to explain its nature is to propound a complete social philosophy. But if so the term lacks any useful function. We have no need to be converted to the rule of law just in order to discover that to believe in it is to believe that the good should triumph." Joseph Raz, "The Rule of Law and its Virtue" in *The Authority of Law: Essays on Law and Morality* (Oxford, UK: Oxford University Press, 1979) 210 at 211 [Raz, "RoL & Virtue"].

⁶¹⁸ Brian Tamanaha discusses the appeal of the Rule of Law as a concept in a broad range of cultures and styles of government, including Putin's Russia, Mugabe's Zimbabwe, and Khatami's Iran, among others. Brian Tamanaha, *On the Rule of Law: History, Politics, Theory* (Cambridge, UK: Cambridge University Press, 2004) at 2-3 [Tamanaha, *Rule of Law*].

SNC Lavalin scandal⁶¹⁹ and the Wet’suwet’en rail blockades.⁶²⁰ In such moments, the Rule of Law seems to possess an especially conservative character, or, perhaps more egregiously, is made to seem like a straightforward concept that cannot be reasonably questioned or discussed. In the media, appeals to adhere to the Rule of Law often consist of politely obeying officials or court orders, sometimes in lieu of the right to peaceful protest.⁶²¹ It is invoked as a singular entity, synonymous with justice. Brian Tamanaha characterizes the Rule of Law’s “rampant divergence of understandings” as being “analogous to the notion of the ‘good,’ in the sense that everyone is for it, but have contrasting convictions about what it is.”⁶²²

As canvassed in Chapter 2, understanding technology in dialogue with social norms is often described using the language of affordances.⁶²³ This invites legal commentators to look critically at the types of behaviours that are enabled by a given environment.⁶²⁴ Much like the sociotechnical insights of STS, affordances assist in making sense of which activities are made possible, promoted, or perhaps even mandated, by surrounding ecosystems. When considered in the legal context, one might characterize modern law as affording an environment hospitable to particular values; possibilities include human dignity, freedom of expression, and reliable commercial frameworks, among others.⁶²⁵ Traditionally, common law systems have imposed Rule of Law limitations on the delegation of authority, including repeatability, non-arbitrariness, and universal application.⁶²⁶ From this view, these oft-cited characteristics are the Rule of Law’s affordances: they afford the type of environment where decisions can be made, contracts concluded, and commerce carried out without retributive justice, cronyism, or needing to duck Damocles’ unexpected sword. In this way, normativity flows from affordances—the Rule of Law’s power to shape an ecosystem for our behaviour generates a corresponding normativity. To the extent that this normativity is successful in mandating particular ways of being, we concede to follow the law because law’s normative power makes a successful claim for our obedience. Or, at the very least, the state’s exclusive right to use violence to enforce law leaves us feeling like we have little choice but to obey.

Combining the Rule of Law with algorithmic technologies to generate self-driving laws infuses additional affordances into the mix. Recognizing that the technological affordances of self-driving laws also make claims for our obedience is critical to understanding technology and law as instances of power. In combination with the existing normativities justified by the Rule of Law, technology

⁶¹⁹ See, e.g., Patricia Hughes, “The SNC-Lavalin Controversy in IV? Acts: A Play About the Rule of Law” (5 March 2019), online: *SLAW* <www.slaw.ca/2019/03/05/the-snc-lavalin-controversy-in-iv-acts-a-play-about-the-rule-of-law/> [perma.cc/X89C-ETVP].

⁶²⁰ See, e.g., Corey Shefman, “Wet’suwet’en blockades: Rule of law is a convenient weapon” (28 February 2020), online: *The Hamilton Spectator* <www.thespec.com/opinion/contributors/2020/02/28/wet-suwet-en-blockades-rule-of-law-is-a-convenient-weapon.html> [perma.cc/9RCN-7BED].

⁶²¹ Describing so-called threats to the rule of law by blockading protestors defying injunctions, Corey Shefman writes: “At best, these voices treat the rule of law as if it is a neutral, amorphous thing, swooping in like a superhero to ‘do justice’ when the law demands.” Shefman, *supra* note 620.

⁶²² Tamanaha, *Rule of Law*, *supra* note 618 at 3. This observation stems from debates in moral philosophy.

⁶²³ Ryan Calo describes affordance theory as being “famous everywhere but law”, citing James J. Gibson’s insights that the “same environment or artifact holds different possibilities and dangers for different organisms.” Ryan Calo, “Can Americans Resist Surveillance?” (2016) *University of Chicago Law Review* 23 at 25.

⁶²⁴ Hildebrandt concedes that affordances are not a common sense concept but are used in ecological psychology to describe the sorts of behaviours that a particular environment makes possible for a given organism. Mireille Hildebrandt, “Affordance,” *supra* note 218 at 117.

⁶²⁵ Hildebrandt cites Tommaso Bertolotti & Lorenzo Magnani’s term “cognitive niche.” *Ibid* at 122.

⁶²⁶ Raz, “RoL & Virtue,” *supra* note 617 at 214-218.

might offer an amplifying, or at the very least, a confounding effect. Hildebrandt notes “as a lawyer” that characterizing the law in terms of an affordance is “pivotal in times when (1) *law as an affordance* as well as (2) *the affordances of the law* are on the verge of radical change.”⁶²⁷ Such radical change is imminent through the delegation of decision-making authority to algorithmic structures.⁶²⁸ By making law more efficient, proponents of self-driving laws might suggest, we have done something inherently good. But is efficiency necessarily an end to be pursued for its own sake? In defining any strong normative framework, system values must be clearly articulated. Otherwise, an ends-oriented utilitarianism might creep in, emphasizing maximal efficiency ahead of all other values. Furthermore, and perhaps most critically, despite the Rule of Law’s ubiquity, its most usual definition is as a contested concept.⁶²⁹ Despite the Rule of Law’s normative force on law’s relations, other normativities interact with its world-building function.

To dig into these normativities, I begin this chapter with theoretical accounts of the Rule of Law. I discuss approaches to the Rule of Law theorized and delineated by various scholars. I map the contributions of scholars like HLA Hart, Lon Fuller, Joseph Raz, and Ronald Dworkin, while still showcasing the Rule of Law as an essentially contested concept. From the insights of these established scholars, I aim to identify a relatively uncontroversial version of the Rule of Law, rooted in mainstream doctrine. I am mindful that using these scholars reinforces a particular hegemonic view of the law: one that is traditional, white, and male.⁶³⁰ This is by design. Through these legal theorists’ key contributions, I begin telling the story that eager advocates for self-driving laws might expect (and prefer) to hear: that automation guarantees an adherence to—or even a preservation of—the Rule of Law. By starting with this procedural perspective, I demonstrate that even stalwart legal positivists acknowledge the existence of situations where the law might come up short.⁶³¹ These happen to be the situations that occupy most discussion time in legal education and most court time on appeal. The question of what to do in these hard cases has long troubled legal philosophers. Clearly, careful application of the Rule of Law as a concept demands attention to not only the *rule* aspect, but also to the *of law* aspect. This differentiation is sometimes described as formal *versus* substantive conceptions of the Rule of Law.⁶³² To flesh out the substantive components, I draw the conclusion that Dworkin’s interpretative theory of law as integrity offers a meaningful path forward for conceptualizing the AI/ML conversation, even if it still contains some shortcomings. I summarize those shortcomings in Dworkin’s theory, setting the stage for a reparative, critical legal approach in Chapter 5.

1. What is the law supposed to do?

Few concepts are more central to contemporary constitutional democracy than the Rule of Law: society is subtly organized around certain ideas designed to preserve the Rule of Law.⁶³³

⁶²⁷ Hildebrandt, “Affordance,” *supra* note 218 at 121.

⁶²⁸ Scassa, “ADM,” *supra* note 486 at 19.

⁶²⁹ Jeremy Waldron, “The Concept and the Rule of Law” (2008) 43 *Georgia Law Review* 1 at 52.

⁶³⁰ Catharine MacKinnon observes the Rule of Law and the rule of men as one indivisible thing. She references Hart’s rule of recognition and Dworkin’s values as concomitant ways men make law law-like. Catharine A MacKinnon, *Towards a Feminist Theory of the State* (Cambridge, MA: Harvard University Press, 1989) at 170.

⁶³¹ Hart, *supra* note 582 at 154.

⁶³² Zalnieriute et al, *supra* note 52 at 3.

⁶³³ In this section, my treatment of key jurisprudence theorists overlaps some with the explication of the Rule of Law. This, as I explain further below, is an inevitable consequence of the Rule of Law’s positioning as a contested concept, the various claims the theorists advance, and the variable concepts of law that travel alongside its ability to rule. Waldron, *supra* note 629 at 10.

Originating as an alternative to mob rule—a rule of laws, not humans—this structure is typically presented in legal theory as non-negotiable, as it aims to constrain political might from the seductive allure of the moment.⁶³⁴ This precept has its roots in Aristotle’s *Politics*, as a concept designed to lift law above politics and guard against tyranny.⁶³⁵ Yet, despite the Rule of Law’s ideological centrality, it is sometimes difficult to characterize beyond its function as a constraint on the arbitrary exercise of power.⁶³⁶ Like eyeglasses that we use to see the world, we seldom take them off to examine them for themselves.⁶³⁷

In Canadian legal scholarship, the 1959 constitutional case *Roncarelli v Duplessis* often provides an entry point to the topic of the Rule of Law.⁶³⁸ The case centered on a political actor, Quebec Premier Maurice Duplessis, who ordered the provincial liquor commission to revoke the liquor license of Montreal restaurant owner Frank Roncarelli.⁶³⁹ Roncarelli was a Jehovah’s Witness and vocal critic of Quebec’s Catholic church who, after many Jehovah’s Witnesses were jailed for violating a city ordinance prohibiting flyer distribution, provided them with bail.⁶⁴⁰ For such behaviour, Roncarelli was known as a thorn in the side of the Catholic-Church supporting Duplessis government: he alleged that his liquor license had been revoked in retaliation for his religious and political activities, and brought a civil action, ultimately succeeding against the Premier of the province. Using this case, the Rule of Law is presented as a concept that enables universal, non-arbitrary, and prospective application of law. Yet, the majority offered several concurring opinions explaining contrasting rationales for the decision. Famously, the opinion of Justice Rand emphasized the limitations on “absolute and untrammelled ‘discretion’” and the corresponding duty of good faith in discharging public authority that such discretion necessarily implies.⁶⁴¹ Despite such lofty sentiments, some judges were content to dispose of the matter using a notice provision in Quebec’s Civil Code.⁶⁴² This seminal case, so often held up as a stalwart example of a central organizing principle in Canadian law, reveals internal discord at the highest judicial level about how the Rule of Law operates, what credence its force lends to society, and how to characterize its power appropriately.

Rule of Law accounts are sometimes described as being either formal or substantive. Formal accounts focus on the rule aspect: the systematization of legal rules, the places those rules derive their authority, and how those rules make a legitimate claim to obedience.⁶⁴³ Substantive accounts focus on the character of the rules themselves: what the rules actually are, what themes exist within the rules, and which theories of morality might animate their precepts or justify their operation.⁶⁴⁴ This dichotomy, however, does little to grapple with more nuanced perspectives of the concept of

⁶³⁴ Lawrence B Solum, “Equity and the Rule of Law” (1994) 36 *Nomos* 120 at 121.

⁶³⁵ Aristotle, *Politics* 1282b, cited in The Stanford Encyclopedia of Philosophy, “The Rule of Law” (22 June 2016), online: *The Metaphysics Research Lab* <<https://plato.stanford.edu/entries/rule-of-law/>> [perma.cc/5445-UY3C] [SEP, “Rule of Law”].

⁶³⁶ Paul Burgess, “The Rule of Law: Beyond Contestedness” (2017) 8 *Jurisprudence* 480 at 482, online: <<https://doi.org/10.1080/20403313.2017.1341203>> [perma.cc/2NQG-B3H2].

⁶³⁷ Richard Delgado, “Storytelling for Oppositionists and Others: A Plea for Narrative” (1989) 87 *Michigan L Rev* 2411 at 2413 [Delgado, “Storytelling”].

⁶³⁸ *Roncarelli v Duplessis*, [1959] SCR 121.

⁶³⁹ *Ibid.*

⁶⁴⁰ *Ibid.*

⁶⁴¹ *Ibid* at 140.

⁶⁴² Genviève Cartier, “The Legacy of *Roncarelli v. Duplessis* 1959-2009” (2010) 55 *McGill LJ* 375 at 385-388.

⁶⁴³ Tamanaha, *Rule of Law*, *supra* note 618 at 92.

⁶⁴⁴ *Ibid.*

law and what it might offer for society. For my analysis, it would be all too easy to construct a strawman version of legal formalists, *i.e.*, those who are so eager for law to offer a complete code for social organization that they believe law has the capacity to answer all possible questions. While this type of reasoning does sometimes appear in self-driving laws literature, it is not a serious claim advanced in analytical jurisprudence.⁶⁴⁵ Each of the theorists below acknowledge the fuzzy edges of law's empire; however, they advocate for different ways of dealing with it.

As I work through these theories, there is some conflation between the theorists' overall *concepts* of law and their corresponding insights into the Rule of Law. This, I believe, is necessary as we attempt to tease out what impact self-driving laws might have. If the power of the Rule of Law adheres despite its fundamental character as a contested concept, this might pose difficulties for articulating a concept of law to travel alongside it.⁶⁴⁶ Jeremy Waldron is particularly instructive on this point: he argues that we cannot understand the concept of law separately from the Rule of Law.⁶⁴⁷ Even if grammar seems to demand definition of a constituent component of the phrase before ascribing meaning to the phrase itself, Waldron remains skeptical. He argues we need to understand the terms as a package, instead of seeing law as a "separable component" of the Rule of Law—or *vice versa*.⁶⁴⁸ By contrast, Raz would demand enunciation of a system of law first, before delving into the specific content of the Rule of Law.⁶⁴⁹ In this way, he argues, the Rule of Law is a negative value—it protects against the danger of arbitrary power made possible by the existence of law itself.⁶⁵⁰ While the Rule of Law does not offer legal systems a precise list of idealized principles they must necessarily conform to, it comes close to being a condition precedent required before any legal system can be considered worthy of our obedience.⁶⁵¹

Varied perspectives are common when attempting to enunciate central concepts like the Rule of Law. By canvassing these viewpoints, I show the challenges that might accompany a transition to self-driving laws, where systems require the designation of "ground truth" as a baseline comparator. Even in attempting to chart a way forward, I acknowledge the contested nature of the terrain, the multiple plausible perspectives, and the long-neglected minority views that might need to be incorporated before law could truly function equally for all. This framing might offer the opportunity to understand all that law is—or can be.⁶⁵² The battles of legal interpretability that have captivated theorists⁶⁵³ become an advantageous feature: because legal norms are expressed in human language, they are necessarily ambiguous.⁶⁵⁴ Contestability, then, might just serve as a key

⁶⁴⁵ The strictest versions of formal legality see law as simply that which is done by the state or the sovereign, which is an appealing option for authoritarian governments. Tamanaha, *Rule of Law*, *supra* note 618 at 92.

⁶⁴⁶ And doing so would require developing an entirely new coherent theory of law, which is outside the scope of this project.

⁶⁴⁷ Waldron, *supra* note 629 at 10.

⁶⁴⁸ *Ibid.*

⁶⁴⁹ *Ibid* at 11.

⁶⁵⁰ Joseph Raz, *The Authority of Law: Essays on Law and Morality* (Oxford: Oxford University Press, 2009) at 224 [Raz, *Authority*]

⁶⁵¹ Raz, "Law's Own Virtue," *supra* note 520 at 17.

⁶⁵² Writing from an Indigenous perspective, Larry Chartrand argues that "law is much richer and deeper than Western legal thought." Larry Chartrand, "Indigenizing the Legal Academy from a Decolonizing Perspective," (2015) University of Ottawa WP 2015-22 at 14. Broadening the conception of law to accept simultaneous operation of multiple systems and visions of law might be important for goals of reconciliation.

⁶⁵³ Dworkin, *Empire*, *supra* note 19 at 255-259 (arguing that judges deciding hard cases strive to find coherent principles, while acknowledging that even strict principles can sometimes permit multiple interpretations).

⁶⁵⁴ Hildebrandt, "Algorithmic Regulation," *supra* note 32 at 3.

feature of legal process, offering a prophylactic against overzealous or underinclusive perspectives on law's extent.⁶⁵⁵

a. Theoretical Takes on Law & the Rule of Law

HLA Hart's seminal *The Concept of Law* sets out a vision for modern legal positivism. One of the most important works of twentieth century jurisprudence, Hart offered his response to John Austin's separation thesis, attempting to repair Austin's positivism by extricating it from command theory.⁶⁵⁶ Importantly, Hart's description of law attempts to bridge the divide between positivism and normativity. His careful attempts to identify the different types of rules and their corresponding sources of power demonstrate exceptional sophistication for a concept of law.⁶⁵⁷ Jennifer Nadler describes Hart's middle ground between extremes as being "normative in the sense that it tries to offer an account of legal authority and obligation, but positivist in the sense that it tries to explain law's normativity in terms of something other than its substantive morality."⁶⁵⁸ By beginning with Hart, I here illustrate the overlap between articulating a concept of law with the operation of the Rule of Law. For the project of the Rule of Law's normativity delineated above, this means Hart is interested in creating a system that guides citizens' behaviour by asking what it is that makes people believe certain actions have become obligatory. When does a general concept of law lead to obligations best characterized as the Rule of Law?

Addressing the application of rules, Hart concedes that any "honest description" of the judicial project requires balancing some opposing forces.⁶⁵⁹ Explicit rules to be extracted from cases are usually clear, except in some rare circumstances. The vast majority of the time, Hart asserts, the rule that derives from any given case will have a core of clear meaning.⁶⁶⁰ Such rules are taken to be as determinate as any statutory rule; indeed, judges will often comment in their opinions that they find themselves to be bound by a common law rule they consider contrary to the interests of justice and implore the legislator to change it. Hart acknowledges that "at the margin of rules and in the fields left open by the theory of precedents, the courts perform a rule-producing function which administrative bodies perform centrally in the elaboration of variable standards."⁶⁶¹ Rules with a core of clear meaning might still be accompanied by a penumbra of uncertainty, which can complicate the judicial project. The way rules are used in social life means they do not necessarily offer predictions: because Hart acknowledges rules can have an open texture, he correspondingly notes that "at the points where the texture is open, individuals can only predict how courts will decide and adjust their behaviour accordingly."⁶⁶² Prediction in this sense is not the absolute data scientific answer proposed by the futurists, or what is sometimes offered through empirical methods or predictive analytics. Rather, it is the uncertainty that arises at law's edges, where the definite answers the law sometimes seems to offer are simply absent. Hart's concept of law is

⁶⁵⁵ Hildebrandt, "Algorithmic Regulation," *supra* note 32 at 3.

⁶⁵⁶ The "separation thesis" refers to a system of law that separates it from morality. Hart, *supra* note 582 at 18. See also Jennifer Nadler, "Hart, Fuller and the Connection Between Law and Justice" (2007) 27 *Law and Philosophy* 1 at 3.

⁶⁵⁷ Hart, *supra* note 582 at 28-36.

⁶⁵⁸ Nadler, *supra* note 656 at 1.

⁶⁵⁹ Hart, *supra* note 582 at 134.

⁶⁶⁰ Here, Hart acknowledges that in defining the *ratio decidendi*, or the rule descending, from a given common law case, "the head-note is usually correct enough." This philosophy is oddly reminiscent of advocates for self-driving laws or empirical analytics, where headnotes are common fodder for machine analysis. *Ibid.*

⁶⁶¹ *Ibid* at 135.

⁶⁶² *Ibid* at 138.

formal, but also pragmatic: there may be some areas where law simply cannot reach—an idea further developed below.

Lon Fuller's concept of "legality" also influenced the modern formulation of the Rule of Law. Fuller's project conceived of law as "the enterprise of subjecting human conduct to the governance of rules."⁶⁶³ Unlike Hart, Fuller's approach is anti-positivist: he contends that law requires some reciprocity between the lawmaker and those governed by the law in order to secure law's internal morality.⁶⁶⁴ Fuller described eight formal "desiderata" for the laws of a legal system to live up to the Rule of Law ideal: general, open, prospective, clear, consistent, stable, capable of being obeyed, and upheld by officials.⁶⁶⁵ Although Fuller's desiderata are also linked to law's internal procedural morality, from a strictly Rule of Law standpoint, they propose a series of conditions that must be met before the activities of the state can legitimately be considered lawmaking.⁶⁶⁶ In this way, they can be seen as conditions precedent that must be achieved before state law meets the threshold required by the Rule of Law. As Colleen Murphy explains:

When lawmakers respect the eight principles of the rule of law, their laws can influence the practical reasoning of citizens. Citizens can take legal requirements and prohibitions into consideration when deliberating about how to act. They can predict how judges will interpret and apply rules, enabling them to form reliable expectations of the treatment different actions are likely to provoke. When the rule of law is realized, their expectations of congruence will not be disappointed. Taken together with the reasonable expectation that fellow citizens will also obey the law, these expectations justify the belief that the law gives citizens reasons to act or refrain from acting in certain ways.⁶⁶⁷

In Fuller's view, imposing such requirements gives society the structure it needs for individuals to live out their lives against the "firm base lines" needed for sustained moral interaction.⁶⁶⁸ Reliable legal precepts are a precondition for living a good life, since it is hard to understand the implications of one's actions if the goalposts are continuously moved. Fuller presents this view through the reasons of Foster J. in his famous Speluncan Explorers piece, writing that "our positive law is predicated on the possibility of men's coexistence in society."⁶⁶⁹ The natural lawyer John Finnis similarly engages this idea of the good in defining what constitutes the Rule of Law, asserting that the availability of Rule of Law precepts signify a legal system "in good shape."⁶⁷⁰

⁶⁶³ Lon L Fuller, *The Morality of Law* (New Haven, CT: Yale University Press, 1969) at 106 [Fuller, *Morality*].

⁶⁶⁴ Fuller is frustrated by the positivists' refusal to acknowledge the role legal rules play in effectuating moral behaviour in people. Absent some kind of baseline for society, he proposes, principles are less effective, as they attempt to govern everything simultaneously, "in a social vacuum or in a war of all against all." *Ibid* at 203-205. The idea of reciprocity is expanded below.

⁶⁶⁵ These eight principles comprise an inner morality of law. *Ibid* at 47.

⁶⁶⁶ Note that Fuller's desiderata flow from an internal procedural conception of morality, where morality is achieved through following the proper process of "law-making and law-administering," rather than a substantive morality, which would engage with the content and character of the laws themselves. *Ibid* at 200-201.

⁶⁶⁷ Colleen Murphy, "Lon Fuller and the Moral Value of the Rule of Law" (2004) 24 *Law and Philosophy* 239 at 241.

⁶⁶⁸ Fuller, *Morality*, *supra* note 663 at 205.

⁶⁶⁹ Lon Fuller, "The Case of the Speluncan Explorers" (1949) 62 *Harvard L Rev* 616 at 620. Fuller continues: "When a situation arises in which the coexistence of men becomes impossible, then a condition that underlies all of our precedents and statutes has ceased to exist. When that condition disappears, then it is my opinion that the force of our positive law disappears with it."

⁶⁷⁰ Finnis cites Fuller, amongst others, and indeed: Finnis' requirements for the Rule of Law are similar to Fuller's desiderata. As Hilary Nye explains: "His account of what the Rule of Law entails includes prospectivity, promulgation, clarity, coherence,

Finnis' work distinguishes between kinds of legal validity: law in the legal sense (which is positive law, properly enacted) *versus* law in the focal sense (which is aspirational, focused on law's ideal).⁶⁷¹ Judges must enforce laws in a legal sense, which are properly supported by the Rule of Law, but citizens might disregard laws that they consider contrary to the interests of justice.⁶⁷² By separating these two senses of legal validity, Finnis avoids leading law's grand questions to theories of judicial interpretation.⁶⁷³ Even so, his view promotes a pursuit of the common good: laws in a legal sense are supported by a robust Rule of Law and laws in a focal sense are supported by the citizens' idealized pursuit of justice.

Joseph Raz conceptualizes the Rule of Law as playing a key role in protecting human dignity.⁶⁷⁴ Law is something that people must obey and be ruled by; it might also successfully guide them in their endeavours.⁶⁷⁵ As a matter of "analytic clarity," Raz aims to distinguish the Rule of Law from other values in liberal political morality, like human rights, social justice, or democracy.⁶⁷⁶ Similar to Fuller and Finnis, he offers a series of five characteristics necessary for the law to evince quality *qua* law: reasonably clear; reasonably stable; publicly available; general rules and standards; and applied prospectively, not retroactively.⁶⁷⁷ Differing from Fuller and Finnis, however, Raz's philosophy is positivist, in that he makes no moral claims in his delineation of law's procedures or content; he focuses on the aspects of law that permit stability and certainty, *i.e.*, its authority. He differentiates between a person's behaviour conforming with the law, which could happen accidentally, and a person explicitly obeying the law, which requires that the person know the law and it be capable of guiding their behaviour.⁶⁷⁸ For the Rule of Law to function as a meaningful precept, people need to know what the law is and be able to act on it.⁶⁷⁹ Raz's famous pronouncement that the Rule of Law is to law as sharpness is to a knife showcases the priority of this concept to law's power and efficiency.⁶⁸⁰

These diverse perspectives offer several unifying themes. Taken together, the Rule of Law protects us against arbitrary exercises of power, retribution, or secrecy. It not only governs behaviour, but also ensures a consistent application of legal rules and outcomes; it generates a corresponding trust in public institutions. Clear, open rules help individuals make informed choices about how to exercise their free will. While law is something that people must obey and, indeed, be ruled by, it also offers the possibility that law will guide them in their endeavours.⁶⁸¹ This engenders the stability necessary to build a productive society. This permits some social consensus, embedded in

possibility of compliance, stability, limitation on specific decrees, and requirements with respect to the application and administration of law, as well as institutional or procedural aspects such as the independence of the judiciary, judicial review, access to courts, and so on." John Finnis, *Natural Law and Natural Rights* (Oxford, UK: Oxford University Press, 1980) at 270. See also Hilary Nye, "Finnis's divided view of law: problems for adjudicative theory" (2020) 11 *Jurisprudence* 503 at 524.

⁶⁷¹ Finnis, *supra* note 670 at 18.

⁶⁷² *Ibid* at 360.

⁶⁷³ Nye notes that Finnis' theory is different from Ronald Dworkin's, which relies "heavily on interpretation in order to show the law in its best light. Law in the focal sense, for Finnis, is good positive law – law that has in fact been posited by the officials and also promotes the common good." Nye, *supra* note 670 at 521.

⁶⁷⁴ Raz, *Authority*, *supra* note 650.

⁶⁷⁵ *Ibid* at 212.

⁶⁷⁶ SEP, "Rule of Law," *supra* note 635.

⁶⁷⁷ Raz, "Law's Own Virtue," *supra* note 520.

⁶⁷⁸ Raz, *Authority*, *supra* note 650 at 213-214.

⁶⁷⁹ *Ibid* at 214.

⁶⁸⁰ Kristen Rundle, *Forms Liberate: Reclaiming the Jurisprudence of Lon L Fuller* (Portland, OR: Hart Publishing, 2013) at 149.

⁶⁸¹ Raz, *Authority*, *supra* note 650 at 212.

widely held social expectations of law-abiding behaviour.⁶⁸² In such an environment, society can develop dependable and explainable legal norms, and individuals can realize their human dignity by exercising free choices against a consistent legal backdrop.⁶⁸³ For supporters of self-driving laws, this might seem like a perfect marriage: the AI can be coded with system specifics, offering a rule-oriented output for each situation. This presupposes the view that law is what the courts say, *i.e.*, when the court has given its interpretation of previous case law and statute for a given set of facts.

Law being merely what the courts say is vulnerable to various criticisms. Brian Bix points out that such pronouncements suffer from a “time-frame problem,” where the “settledness” of any legal issue is really only valid relative to a particular point in time.⁶⁸⁴ At best, one could assert that something is legally correct—for now. Correspondingly, specificity presents its own issues: trying to derive general principles from specific fact patterns will necessarily create legal propositions at higher levels of generality than those rules operated at in cases from which they were derived.⁶⁸⁵ Such abstraction starts to evoke the rhetorical meanings of words and concepts, perhaps even playing language games.⁶⁸⁶ And moving in the other direction, converting abstract legal values and principles into a series of concrete, reproducible legal rules requires certain engagement with the abstraction. John Finnis cautions that translating principles of practical reasoning into actual rules needs an act of implementation that exceeds mere deduction.⁶⁸⁷ This means that judges seeking to interpret meaningful statutes from abstract values and principles have some guidance, but are ultimately exercising their judgement, somewhat arbitrarily choosing from a range of reasonable alternatives.⁶⁸⁸ All of these cautionary points suggest some degree of incommensurability in legal decision-making that mandates interpretative exercises to make sense of law’s extent.

Meanwhile, managerial conceptions of law—where law proceeds in a unidirectional manner, originating with an authority flowing from the top of a chain of command—invite criticism from the world-building sense of law. To fully articulate why people obey the law, any legitimate legal system needs to demonstrate the source of, or the justification for, its authority. Fuller points out that a managerial conception, as employed by positivists like Hart, is unhelpful for making sense of why people obey the law.⁶⁸⁹ He characterizes the inner life of the law as requiring a fundamental reciprocity between the governor and the governed: only through a fully reciprocal relationship between the legislator and the citizens can a legal order function properly.⁶⁹⁰ This reciprocity

⁶⁸² Brian Z Tamanaha, “Always imperfectly achieved rule of law: Comments on Jeremy Waldron” (2021) 10 *Global Constitutionalism* 106 at 113.

⁶⁸³ Raz, *Authority*, *supra* note 650 at 212.

⁶⁸⁴ Brian Bix offers this argument in a critique of a passage from AD Woozley’s “No Right Answer” piece, where Woozley proposes legal questions obtain a settled answer once a number of courts have answered that question, and the answer endures. Bix thinks the best way to conceptualize this argument is by limiting the applicability of rules to a certain point in time. Brian Bix, *Law, Language & Legal Determinacy* (Oxford, UK: Oxford University Press, 1993) at 83.

⁶⁸⁵ AD Woozley, “No Right Answer” (1979) 29 *The Philosophical Quarterly* 25 at 27. See also *ibid*: while Bix presents himself as disagreeing with Woozley on the truth-value capable of being extracted from general statements of legal opinions, both scholars do an excellent job of articulating the issues with language.

⁶⁸⁶ Wittgenstein describes language games as the processes by which children learn to speak their native languages, where teachers point to objects to direct children’s attention to them, while simultaneously uttering the corresponding word. He goes on to think through the expansion and ostensive teaching of languages through such language games. See Ludwig Wittgenstein, *Philosophical Investigations*, trans by GEM Anscombe (Oxford, UK: Basil Blackwell, 1958) at 6.

⁶⁸⁷ Finnis uses *determinatio* to describe this “act of implementation ... that goes beyond deduction.” Bix, *supra* note 684 at 89

⁶⁸⁸ *Ibid*.

⁶⁸⁹ Fuller, *Morality*, *supra* note 663 at 207.

⁶⁹⁰ *Ibid* at 209.

between the person giving the orders and the person receiving them is needed, lest law's claim to authority fall apart.⁶⁹¹ Kristen Rundle reflects on Fuller's frustration that so much of legal education focuses on the moments when things fall apart and legal remedies are sought to put things back together.⁶⁹² Conversely, as Fuller was known to argue, a more interesting aspect of law's normativity is how it creates and sustains productive social relationships.⁶⁹³ This observation is especially salient in the face of proposals for self-driving laws: not only might their outcomes be incorrect, but their claim to obedience as a social force might be implausible. Will people truly respect the authority of self-driving laws? Will they feel that their actions are compelled by authoritarian technology?

Harnessing control by fiat seems associated with authoritarian regimes rather than participatory democracy. While judicial restraint is lauded as a key Rule of Law feature, predictability is not the only value encapsulated within it. The Rule of Law, anchored in some concept of the good, might sometimes require updating—like when social norms have evolved from the law on the books, or when technology has rendered a particular behaviour obsolete or impossible. In such cases, some deeper commitment to equality might intercede, asking for a departure from past practice. Hilary Nye offers observations on the Rule of Law from this very vantage point, agreeing that predictability is not the only value promoted by the Rule of Law, and arguing that appropriate judicial intervention can sometimes be at odds with Rule of Law-generated norms of judicial restraint.⁶⁹⁴ She proposes that “clarity, prospectivity, and the rest are valuable because they permit a person to have some say in the shape and structure of her own life.”⁶⁹⁵ Nye goes on to highlight the value in creating an autonomous life, supported by the Rule of Law in both the predictability of operative legal frameworks and the substantive conditions of the corresponding environment.⁶⁹⁶ Invoking substantive conditions requires a broader understanding of the Rule of Law; it does not serve Rule of Law values to altogether neglect changing context or “newly current values,” which might well pursue the ultimate aim of the common good.⁶⁹⁷ Nye's analysis is aimed at the ability to thrive in a community free from unjust persecution.⁶⁹⁸ However, her framing suggests another key criterion of an autonomous life: one that a citizen should be free to develop without the subtle nudges and interventions of technological systems. To me, the necessary corollary for such autonomous self-creation is the potential messiness of novel situations that will take creative legal thought to resolve, as opposed to being instantly resolvable through algorithm. Such creative legal thought might necessitate a measure of lived human experience, as discussed in Chapter 3.⁶⁹⁹ The conversion of the Rule of Law into a Rule by Law offers little to no recourse for human autonomy.

⁶⁹¹ Fuller, *Morality*, *supra* note 663 at 209.

⁶⁹² A posthumous tribute from Fuller's Harvard colleague Albert Sacks focused on how he often recommended changes to the law curriculum to better educate students on law's role in forming and guiding social relationships, rather than focusing solely on law's intervention when those same social relationships break apart. Albert M Sacks, “Lon Luvois Fuller” (1978) 92 *Harvard L Rev* 349 at 350. See also Rundle, *supra* note 680 at 124.

⁶⁹³ *Ibid* at 3.

⁶⁹⁴ Nye, *supra* note 670 at 527.

⁶⁹⁵ *Ibid*.

⁶⁹⁶ *Ibid* at 528.

⁶⁹⁷ *Ibid* at 526-527.

⁶⁹⁸ Like for one's religious beliefs, she proposes. *Ibid* at 528.

⁶⁹⁹ This was one of the chief contentions argued by Kerr and Mathen with respect to JR-R's ability to meaningfully serve as judge. Refer to Chapter 3 for the full discussion.

Influence on human autonomy is also exerted by private actors in law-like ways, where exercises of power flow from the corporate sphere. In the self-driving laws context, this conversion of Rule of Law to Rule by Law also implicates the private actors who control development of AI and ML tools, many of which are large technology companies with considerable financial and social clout.⁷⁰⁰ While the more doctrinal approaches to the Rule of Law tend not to engage with this instantiation of power, focusing solely on state power, some scholars do point out the connection. British judge Lord Bingham defined the Rule of Law very broadly, suggesting that “all persons and authorities within the state, whether public or private, should be bound by and entitled to the benefit of laws publicly and prospectively promulgated and publicly administered in the courts.”⁷⁰¹ As introduced in Chapter 1, Paul Gowder’s real-world articulation of the Rule of Law emphasizes the need to constrain coercive power that employers hold over employees, tenants, and others in positions of vulnerability, noting that private actors’ actions are often backed up by the state.⁷⁰² Robin West agrees, noting that “the stigma, punishment, harm, and so on that threaten the enjoyment of the lives of many people, ... at least on occasion come not from states but from powerful nonstate entities. Part of the point of law is to do something about that.”⁷⁰³ Taking this broad view of the ambit of the Rule of Law will become important for thinking through the consequences of AI/ML applications, which I return to in Part III.

Imagining a transition from Rule of Law to Rule by Law also engages the longstanding question of what happens when law runs out. Even legal positivists like Hart and Raz acknowledge that some hard cases offer no clear answers, requiring either a non-legal exercise of judgment or an appeal to something more.⁷⁰⁴ Answering the question of what we should do when the situation calls for that pesky something more has long confounded legal scholars. It might be tempting here to wade into the debates on law’s morality.⁷⁰⁵ Hart and Raz would prefer not to enter such debates, since it would require resorting to something outside law, injecting political philosophy into the mix, and generating consequent losses for values prized by positivists: certainty and authority.⁷⁰⁶ By offering somewhat neutral rules for generating systems of law, both scholars leave any morally-tinged specifics outside their frameworks, relegating them to the world of politics. This type of approach is attractive for the philosophical minefields of analytical jurisprudence but it is altogether

⁷⁰⁰ Recall that five companies control the preponderance of Silicon Valley’s capital and marketshare: Google (Alphabet), Facebook (Meta), Microsoft, Apple, and Amazon. See Wilhelm, *supra* note 5.

⁷⁰¹ Lord Bingham, “The Rule of Law” (2007) 66:1 *The Cambridge Law Journal* 67 at 69. I note that I was first directed to this quotation by Jeff King, through a presentation given at the University of Ottawa’s Public Law Group in January 2020.

⁷⁰² Gowder, *supra* note 51 at 87. This is especially true if legal provisions fail to protect workers, tenants, or others. Gowder explains: “Such a dictatorship is backed up, albeit indirectly, by the coercive force of the state, insofar as the state enforces the property and contract regimes that enable some employers to take dominant labour market positions over many people, structure the terms of that relationship, and, indeed, acquire the concentrated control over resources that forces people into the employment market to meet their basic needs in the first place.”

⁷⁰³ Robin West, “The Limits of Process” in James Fleming, ed, *Getting to the Rule of Law: Nomos L* (New York, NY: New York University Press, 2011) 32 at 48.

⁷⁰⁴ Hart’s approach to hard cases is discussed in his *Concept of Law*, *supra* note 582 at 135; see also Raz, *Authority*, *supra* note 650 at 49-50. Peter Oliver also offers a concise summary of the hard case approach; see Peter C Oliver, “A Constitution Similar in Principle to That of the United Kingdom?: The Preamble, Constitutional Principles, and a Sustainable Jurisprudence” (2019) 65:2 *McGill LJ* 207 at 258.

⁷⁰⁵ At this point, law just might collapse into politics. Hart, *supra* note 582 at 166 (discussing the limited criticisms of law that are valid through the lens of morality).

⁷⁰⁶ Raz specifically spells out this caveat to operation of the Rule of Law: mere adherence to it will not guarantee other ideals like justice, democracy, or respect for human rights. Law will therefore need additional virtues in order “to avoid injustices and other moral blemishes.” Raz, “Law’s Own Virtue,” *supra* note 520 at 15.

unhelpful for the practical exercise of generating an algorithmic version of justice, given that all cases—hard or not—must be accounted for in the algorithm.

I discuss the specific problem of hard cases in further detail below. But first, I offer a brief introduction to Dworkin’s theory of law as integrity. He proposes another answer to reconcile issues of morality: through his one-system approach, such questions are properly dealt with as a matter of legal interpretation.⁷⁰⁷

b. Ronald Dworkin’s Law as Integrity

In Dworkin’s theory of law as integrity, he asserts that the arc of the law requires fidelity to certain principles, to which legal officials must always adhere.⁷⁰⁸ But Dworkin’s account also acknowledges the role discretion must play alongside rules. These two concepts must be understood in tandem to define the scope of discretionary power.⁷⁰⁹ Articulating the broad outlines of his theory, Dworkin insists that “law as integrity denies that statements of law are either the backward-looking factual reports of conventionalism or the forward-looking elements of legal pragmatism.”⁷¹⁰ Instead of engaging in the old debate of whether judges merely apply or create new law, Dworkin rejects the debate in favour of his own framing to better understand the complexity of legal reasoning. He calls this framing “relentlessly interpretative,”⁷¹¹ asking judges to conceptualize legal principles as horizontal rather than vertical, casting a wide net across community standards to encompass as many as possible.⁷¹²

Beginning in the present, and pursuing the past only to the extent that it makes sense to characterize current standards, law as integrity consists of both past decisions and the associated principles needed to justify them.⁷¹³ This is the material accessed by Dworkin’s imagined superjudge Hercules, who possesses the unparalleled capability to carefully and rapidly parse the entire corpus of the common law to seek the legal result that best promotes the integrity of the system. Unlike real-world judges, whose judging experiences are necessarily constrained by overloaded dockets and limited resources, Hercules’ enhanced ability enables him to consider all possible outcomes to arrive at the answer that best accords with integrity. Through the model of Hercules, Dworkin’s own acknowledgements of ambiguity, nuance, or different reasonable outcomes are cast aside in favour of certainty; Hercules’ depth of understanding proposes the availability of a singular correct legal answer in hard cases, elaborated upon further below.⁷¹⁴

Infinitely able Hercules, with unlimited time and unlimited resources, resembles some form of automated justice, meted out through powerful computation, reminiscent of Alarie’s legal

⁷⁰⁷ Through this lens, Dworkin claims that viewing law as interpretivism denies conceptualizing law and morality as separate systems. Ronald Dworkin, *Justice for Hedgehogs* (Cambridge, MA: Harvard University Press, 2011) at 402-404 [Dworkin, *Hedgehogs*].

⁷⁰⁸ Dworkin, *Empire*, *supra* note 19.

⁷⁰⁹ Hildebrandt, “Algorithmic Regulation,” *supra* note 32.

⁷¹⁰ Dworkin, *Empire*, *supra* note 19 at 225.

⁷¹¹ *Ibid* at 226.

⁷¹² *Ibid* at 227.

⁷¹³ *Ibid*.

⁷¹⁴ While Dworkin retreated from the right answer thesis advanced earlier in his career, his theory of law as integrity nonetheless assumes that lawyers argue as if there is one right answer.

singularity. In *Law's Empire*, Dworkin all but anticipates the coming threat of self-driving laws. Nearing the end of his tome, he writes in reflection:

Have I said what law is? The best reply is put to a point. I have not devised an algorithm for the courtroom. No electronic magician could design from my arguments a computer program that would supply a verdict everyone would accept once the facts of the case and the text of all past statutes and judicial decisions were put at the computer's disposal.⁷¹⁵

Despite the robustness of interpretative theory, Dworkin acknowledges the possibility of multiple equally plausible interpretations authored by equally well-meaning judges. This, he wants us to conclude, is not a shortcoming of law as integrity; instead, it is a feature. To illustrate his point, Dworkin offers six methodological approaches that Hercules might take in order to offer a “coherent theory about legal rights to compensation for emotional injury” in the *McLoughlin* case, all while knowing that both sides have cited meaningful precedents, and both sides would count a decision in their favour as a continuing narrative in the ongoing chain novel.⁷¹⁶ Hercules is described as being very careful, creating a short list of six potential precedents, each of which is relatively concrete; even allowing some space for complexity, these accounts will invariably contradict one another.⁷¹⁷ His different methods require different resolutions on the boundaries of moral rights to compensation, the importance of economic calculations, the foreseeability of emotional injuries, or the balancing of foreseeability *versus* financial risk.⁷¹⁸ In the pages that follow, Dworkin goes through a painstaking analysis of all the potential interpretations Hercules might take. Noting that law as integrity asks judges to “make the law coherent as a whole,” he acknowledges that this process can require ignoring the silos that different areas of law are routinely slotted into.⁷¹⁹

Translating Dworkin’s interpretative approach from theory to practice presents a significant challenge for self-driving laws. It places intense pressure on legal metaphors to operate as blueprints for all possible legal futures. Such legal metaphors experience habitual strain from being squeezed into application for technological innovations never contemplated at the time of their creation.⁷²⁰ The likelihood that the existing dataset contains the right answer for any given legal situation is only possible if the ambit of potential legal futures remains correspondingly narrow

⁷¹⁵ Dworkin, *Empire*, *supra* note 19 at 412.

⁷¹⁶ Dworkin, *TRS*, *supra* note 24 at 240.

⁷¹⁷ *Ibid* at 240-241.

⁷¹⁸ Hercules’ six different methods are as follows: (1) No one has a moral right to compensation except for physical injury. (2) People have a moral right to compensation for emotional injury suffered at the scene of an accident, but not suffered later. (3) People should recover for emotional injury if it would diminish the overall cost of accidents or make the community richer long term. (4) People have a moral right to compensation for any injuries, emotional or physical, that is a direct consequence of careless conduct, no matter its unlikelihood or unforeseeability. (5) People have a moral right to compensation for emotional or physical injury only if the injury was reasonably foreseeable. (6) People have a moral right to compensation for reasonably foreseeable injuries, but this right should not be imposed to give massive/destructive financial burdens. *Ibid*.

⁷¹⁹ Dworkin notes that judges often begin their work by assigning a particular case into some compartment, which means the precedents and statutes considered are derived from this subject area. *Ibid* at 251. The classification at work here brings to mind the operation of self-driving laws as categorization systems; the degree to which Hercules might be sometimes asked to ignore the boundaries between categories is not consistent with the datafication of law (at least in its current, non-singularity incarnation).

⁷²⁰ Kerr, “Metaphors,” *supra* note 358.

and consistent with past practice.⁷²¹ Certain situations call for an expansion of what happened before, to best respond to unprecedented black swan events or ameliorate historical wrongs; for example, reconciliation with Indigenous peoples through the legal system will be impossible if strict precedents, rooted in colonial legacies, are resolutely maintained.⁷²² In such situations, Hercules might be called upon to renovate his conception of justice.

2. Can the law offer right answers?

Some of Dworkin's early work advanced the right answer thesis in law: the idea that each case has a single right legal answer, buried deep in some logical, principled morass of common law precedent, evidentiary record, and legal argument, just itching to be found.⁷²³ In his book *Taking Rights Seriously*, Dworkin proposed that proper legal decision-making flows from the wider net he has cast, including rules and principles.⁷²⁴ Judges have an obligation to justify their choices as emanating from a set of possible options, thereby suggesting that the act of eventually selecting one answer from a multiplicity of options meant that it was the correct option. This positioning was Dworkin's response to Hart's open texture thesis, briefly elaborated below in part (a).

Brian Bix argues that Dworkin's right answer thesis might be divided into two subclaims: first, that there exist unique correct answers to every possible legal question and, second, that these answers are delivered through legal considerations only (meaning that no extra-legal consideration is required).⁷²⁵ One might concede that there are correct answers to every possible question without conceding that law offers those answers: morality would come into the equation to explain away some types of decisions. But the breadth of Dworkin's theoretical approach requires that correct answers be both generally available and specifically available within the systematization of law. In essence, Dworkin insists that legal positivism is simply inadequate to explain the nature of what happens when law runs out. His theory, he proposes, better elucidates what happens on those occasions when judges make decisions despite the absence of a clear answer.

Later, Dworkin retreated somewhat from his positioning on right answers: when introducing the character of Hercules in *Law's Empire*, he cautioned that his readers should not misinterpret Hercules' answers to legal problems as offering a complete theory of law. Rather, Hercules' framing is offered as an approach for finding legal meaning, consisting of "questions rather than answers."⁷²⁶ As readers follow Hercules' career, observing how he deals with particular legal issues, they are given a window into how law as integrity might work in practice. Dworkin wants readers to conclude that the idealized judicial decisionmaker thoroughly weighs a series of reasonable options

⁷²¹ Oliver, *supra* note 704 at 255-256.

⁷²² John Borrows, *Canada's Indigenous Constitution* (Toronto: University of Toronto Press, 2010) at 6 [Borrows, *Indigenous Constitution*]. See also TRC Call to Action #28: Truth & Reconciliation Commission of Canada, "Calls to Action" (2015), online (pdf): <www2.gov.bc.ca/assets/gov/british-columbians-our-governments/indigenous-people/aboriginal-peoples-documents/calls_to_action_english2.pdf>.

⁷²³ Dworkin, *TRS*, *supra* note 24: see generally "Chapter 4: Model of Rules I."

⁷²⁴ *Ibid.*

⁷²⁵ Bix characterizes Dworkin as a maximalist on when to describe considerations as being "legal considerations." Dworkin seems willing to run the gamut of possible inclusions. Bix contrasts this with Raz, who is a minimalist on legal considerations: he wouldn't even include moral standards that seem to be embedded in statutes. Bix, *supra* note 684 at 82.

⁷²⁶ Dworkin, *Empire*, *supra* note 19 at 239.

before arriving at the ideal judicial outcome. Intriguingly, Dworkin's method invites his critics into Hercules' approach, thereby inoculating himself against extensive criticism:

But I offer this caution in advance. We must not suppose that his answers to the various questions he encounters define law as integrity as a general conception of law. They are the answers I now think best. But law as integrity consists in an approach, in questions rather than answers, and other lawyers and judges who accept it would give different answers from his to the questions it asks. You might think other answers would be better. (So might I, after further thought.) You might, for example, reject Hercules' views about how far people's legal rights depend on the reasons past judges offered for their decisions enforcing these rights, or you might not share his respect for what I shall call "local priority" in common-law decisions. If you reject these discrete views because you think them poor constructive interpretations of legal practice, however, you have not rejected law as integrity but rather have joined its enterprise.⁷²⁷

Despite Dworkin's clever equivocation here, his framework nonetheless anticipates that lawyers argue as if there is one correct answer. As a practical matter, law's empire demands some initial agreement about what constitutes the law to ensure an even playing field. Indeed, Dworkin describes the phenomenon as crucial for law as an interpretive exercise: that lawyers are essentially arguing about "the best interpretation of roughly the same data."⁷²⁸ As judicial decision-making is necessarily informed by the act of lawyering, the presumption of right answers remains important for the overall architecture of legal reasoning. Court proceedings are conducted with opposing lawyers battling for a particular interpretation to be seen as the right answer. Judges pen their reasons knowing that they must both offer a comprehensive account of their decisions that will survive challenge on appeal, and also persuade the losing party that their perspective was thoroughly taken into account. While one might assert that either the lawyers' positions or the judges' reasons consist of right answers, doing so would present an impoverished vision of law, discounting the essence of contested concepts and the interpretative practices at the heart of meaningful legal inquiry. Dworkin makes this point by emphasizing a concept of law that remains neutral as between "competing explanation of the connection between a community's reigning opinions and its legal commitments."⁷²⁹ Especially given the flexibility of language and linguists' observations on the malleability of meaning, it is difficult to presume that even other humans will agree about the meaning of a selected word, let alone when a machine becomes interlocutor of claims for linguistic meaning. Correspondingly, any implicit assumption that law has the ability to offer right answers might present problems for the adjudication and the systematization of law.

This assumption, in turn, will pose conceptual problems for deploying legal solutions involving AI/ML. Take, for example, the various "equally correct" interpretations that Hercules considered

⁷²⁷ Dworkin, *Empire*, *supra* note 19 at 239-240.

⁷²⁸ *Ibid* at 91. Dworkin elaborates: "I do not mean that all lawyers everywhere and always must agree on exactly which practices should count as practices of law, but only that the lawyers of any culture where the interpretative attitude succeeds must largely agree at any one time. We all enter the history of an interpretative practice at a particular point; the necessary preinterpretive agreement is in that way contingent and local.

⁷²⁹ *Ibid* at 97.

in response to the *McLoughlin* case. These equally correct answers are simply various approaches that different lawyers, motivated by different aspects of law's empire, and the different needs of their respective clients, might emphasize in constructing their legal argument. We might, as Cass Sunstein suggests, need to modify Dworkin's interpretative approach to acknowledge that preferring any one answer requires intrinsic value judgements about the efficacy of courts and the fallibility of the operative political system.⁷³⁰ Otherwise, to assume that the argument that carried the day is "correct" in the sense of ground truth is to discount the various political and philosophical machinations operating in the background. Self-driving laws are thereby faced with a credibility problem, even if they are deeply rooted in Dworkin's sophisticated legal theory. For AI/ML applications, the presumptive existence of a singular "correct" answer tends to be a key feature of system design. Data analysis, particularly in instances of classification or regression, works to provide an answer that the system designer believes to be correct. Acknowledging the possibility of multiple plausible answers adds a layer of complexity that might not be available per the given system design. Even in this datafied framework, it seems inevitable that system designers attempting to answer legal questions will be doing so as matters of (at least some) judgment, rather than as matters of pure calculation.⁷³¹

In the world of data science approaches to problems, one could construe this issue as one of insufficient data: by acquiring additional information about particular problems, one might eventually be able to see all possible angles and arrive at a singular "correct" answer. This particular view of the law is appealing to authoritarians and libertarians: that the law might finally become clear if you just invest enough effort into fully understanding it. In some ways, Dworkin's theory of law as integrity seems to demand this type of conclusion. His legal universe claims a whole-system approach: a system that can offer all possible legal answers. However, even though Dworkin's theory requires lawyers to argue *as if* there is a right answer, nowhere does he provide anything more than his own (or, Hercules' own) best argument for a given "right" answer, all the while explicitly acknowledging that other lawyers, critics, and judges will be equally convinced of their own views, and, in turn, Hercules' best answer.

With this introduction in place, I turn now to discuss Dworkin's answers to two particular issues: (a) the problem of hard cases; and (b) the concept of law as a chain novel.

a. Problem of hard cases

If hard cases really do exist, that is, if there are truly some cases that cannot be resolved with an appeal to either statute or precedent, then there needs to be a way to answer them. This irresolvability requires inquiries into the indeterminacy of law. Famously, HLA Hart presented

⁷³⁰ Sunstein points out that we cannot really understand an interpretative calculus without acknowledging the possibilities of errors or problems. He writes: "The argument for any particular approach must depend, in large part, on a set of judgments about institutional capacities—above all, about the strengths and weaknesses of legislatures and courts. We cannot assess decision costs and errors costs without making those judgments. If judges are excellent and error-free, their excellence bears on the choice of a theory of interpretation. If judges are likely to blunder, their fallibility bears on the choice of a theory of interpretation." Cass Sunstein, "There is Nothing that Interpretation Just Is" (2015) Constitutional Commentary 193 at 209, online: *University of Minnesota* <<https://scholarship.law.umn.edu/concomm/120>> [perma.cc/5N8E-4LND].

⁷³¹ Oliver notes the different usages of "judgement" and "judgment," with the former acknowledging the expansive nature of the task undertaken by judges as involving genuinely hard choices, empathy, wisdom, and other characteristics, but the latter being used in most scenarios involving the judiciary or court proceedings. Oliver, *supra* note 704 at 212, and in particular: footnote 12.

this issue as a “penumbral problem” concerning the open texture of law.⁷³² If a rule exists, then there will be some factual situations where circumstances dictate that the rule clearly applies. Yet, in other factual circumstances, it might be equally appropriate either to assert or to deny that the very same rule applies. Hart acknowledges that this “fringe of vagueness or ‘open texture’” of law might affect our ability to recognize the “ultimate criteria” used to identify what constitutes law just as well as a particular statute.⁷³³ In other words, law is hard, facts are fickle, and there are some situations where the best legal answer is: no one really knows, but we’ll take our best guess or exercise our most considered discretion, and call it law.

Hart was adamant that such problems arose only occasionally—at the very fringes of more fundamental rules—and in contrast with the “unquestionably rule-governed operations over the vast, central areas of law.”⁷³⁴ Dworkin, on the other hand, introduces hard cases as those where “no settled rule dictates a decision either way.”⁷³⁵ These, by definition, may happen more regularly than Hart expected; they might also require settling the debates about judicial activism, political interventions, and inclusion of principles apart from doctrinal rules. Judges may strive to produce coherent results,⁷³⁶ or they may invent law from the ether—or a range of options in between. The question, then, is not only about what individual judges choose to do when hard cases arise, but also what the legal system is prepared to use in order to resolve them. In essence, we must ask: how are gaps in the law filled?

This issue animated the famed Hart-Dworkin debates, where the two legal philosophers elaborated upon competing theories about how to navigate the interplay between rules and principles in the legal system.⁷³⁷ In a helpful overview, Shapiro characterizes the Hart-Dworkin debate as consisting of three main theses, which he draws from Dworkin’s essay “The Model of Rules I.”⁷³⁸ The first is the Pedigree Thesis: how a community differentiates between what is law from what is not law based on its “pedigree” or source.⁷³⁹ The second is the Discretion Thesis: what to do in situations where the existing legal rule does not cover the field.⁷⁴⁰ The third is the Obligation Thesis: whether someone feels bound to act in accordance with a particular legal obligation is generated by the existence of legal rules.⁷⁴¹ I do not wish to wade too deeply into the treacherous, deep waters of this longstanding debate, other than to observe that many scholars have dedicated considerable attention to the critical questions of where law runs out and how judicial discretion works when that happens, without ever reaching any kind of meaningful

⁷³² Hart, *supra* note 582 at 123; see also Raz, *Authority*, *supra* note 650 at 49-50; 53-77.

⁷³³ *Ibid.*

⁷³⁴ Hart, *supra* note 582 at 154. While Hart wants to relegate the rule-sceptic to fringe cases, he also acknowledges that “in the vast majority of cases that trouble the courts, neither statutes nor precedents in which the rules are allegedly contained allow of only one result. In most important cases there is always a choice.” *Ibid.* at 12. Even in situations where ambiguities operate, Hart suggests judges frame their responses “as to give the impression that their decisions are the necessary consequences of predetermined rules whose meaning is fixed and clear.” *Ibid.* This nod to certainty helps promote confidence in judges’ decisions.

⁷³⁵ Dworkin, *TRS*, *supra* note 24 at 83.

⁷³⁶ Dworkin, *Empire*, *supra* note 19 at 255-259 (arguing that judges deciding hard cases strive to find coherent principles, while acknowledging that even strict principles can sometimes permit multiple interpretations).

⁷³⁷ Scott J Shapiro, “The Hart-Dworkin Debate: A Short Guide for the Perplexed” (2007), online: *SSRN* <papers.ssrn.com/sol3/papers.cfm?abstract_id=968657> [perma.cc/Q8GU-9AKT] [Shapiro, “Perplexed”].

⁷³⁸ *Ibid.* at 6.

⁷³⁹ *Ibid.* at 7.

⁷⁴⁰ *Ibid.* at 8.

⁷⁴¹ *Ibid.*

consensus. Fundamentally, to adopt Dworkin's position is to reject legal positivism as an animating theory for law and to allow for strong judicial discretion in hard cases.

Dworkin's emphasis on interpretation takes into account the broad influences of legal institutions, their authority, and their connections with morality. Shapiro helpfully explains that:

Dworkin believes that no rule could be fashioned that accurately reflects the verdicts of all these political principles, presumably because the possibilities that would have to be considered and codified are infinite in number. Moreover, these principles and their weights fluctuate over time, based on their own degree of institutional support, and hence any resulting master rule would fail to be stable.⁷⁴²

Dworkin asserts that judges have the duty to discover what the rights of the parties are, which he distinguishes from “invent[ing] new rights retrospectively.”⁷⁴³ Importantly, as noted earlier, Dworkin acknowledges that reasonable lawyers and judges may sometimes disagree about legal rights, thus signalling the complexity that is inherent in some legal inquiries.⁷⁴⁴ When Dworkin invites his critics into his method, arguing that those who disagree with law as integrity have in fact joined its enterprise, he sets the stage for incorporating additional viewpoints into the fabric of law. He even hints at this when introducing his concept of law as narrative, writing: “[w]e can usefully compare the judge deciding what the law is on some issue ... with the literary critic teasing out the various dimensions of value in a complex play or poem.”⁷⁴⁵ I doubt he would agree, but in my view, such literary allusions are the perfect justification for law as narrative.

b. Law as a Chain Novel

Dworkin analogized the development of the law to a chain novel, wherein a series of authors pass on responsibility for authoring the law to subsequent authors.⁷⁴⁶ To ensure integrity, each author must write consistently with all prior authors.⁷⁴⁷ This asks judges to adopt a particular role in authoring their legal decisions: the arc of the development of the law must remain compatible with a storytelling structure for it to be fully intelligible by those it proposes to govern. Judges might usefully be compared to literary critics, since their interpretative efforts requiring the extraction of some creative meaning from “text or data” resemble literary engagement of some “complex play or poem.”⁷⁴⁸ This interpretative feat imposes normative force. By creating this space for narrative within the arc of the law, the chain novel anticipates a storytelling structure as a key component of how law generates its normative force and achieves a world-building function.

Admittedly, the stories in Dworkin's chain novel do not consist of social scientific stories. While this conception of a novel is different from the storytelling tradition for which critical legal scholars have advocated, there are obvious similarities in the establishment of a clear narrative arc that unites past precedent with future legal application. In *Law's Empire*, Dworkin takes aim at critical

⁷⁴² Shapiro, “Perplexed,” *supra* note 737 at 14.

⁷⁴³ Dworkin, *TRS*, *supra* note 24 at 81.

⁷⁴⁴ *Ibid.*

⁷⁴⁵ Dworkin, *Empire*, *supra* note 19 at 228.

⁷⁴⁶ *Ibid.*

⁷⁴⁷ *Ibid.*

⁷⁴⁸ *Ibid.*

legal scholarship through the lens of internal scepticism.⁷⁴⁹ He seems to shrug at the various analyses that critical scholars might propose, asserting that their emphasis on peoples' lived experiences are too simple to make sense. He writes: "nothing is easier or more pointless than demonstrating that a flawed and contradictory account fits as well as a smoother and more attractive one. The internal skeptic must show that the flawed and contradictory account is the only one available." In so doing, he is essentially arguing the inverse of Occam's razor: the simplest answer is not, in fact, the correct one. I agree with Dworkin on this point, except that we would differently characterize what the simpler account is. It is not simpler or more attractive to incorporate the confusing aspects offered by critical accounts. It is actually a grand challenge—one that I will chart a course for in Chapter 5, further complicating our conception of law.

Robert Cover notes astutely in "Violence and the Word" that while the law might appear to speak with a single voice, its voice is not derived from "a single coherent and consistent" interpretation of legal materials.⁷⁵⁰ Indeed, attempts to extract a singular reasonableness from the complexity of lived human experiences tend to subvert realities that do not reflect the majority back to itself. Standards like the "reasonable person" as a benchmark for comparison in many subject areas of law promote a myopic understanding of the human experience.⁷⁵¹ Mayo Moran writes about how the idea of an objective, reasonable standard is a concept deeply embedded in "common sense ideas about what is normal or ordinary."⁷⁵² She discerningly observes that standards of reasonableness seeded in concepts of what constitutes "normal behaviour" lead to "troubling or discriminatory social understandings of particular groups," which ultimately give legal legitimacy to entrenched systemic inequalities.⁷⁵³ Social constructions of power have long reinforced a certain prototype required before a subject can be believed. The fact finding that courts do in order to develop a full recounting of the events giving rise to a legal dispute can sometimes be cloaked in expectations about what proper truth-telling looks like. Sarah Banet-Weiser synthesizes many critiques on how claims to truth-telling must be understood through a power lens, "including racial, economic, gendered, and colonialist power."⁷⁵⁴ Invoking Foucault's lessons on the nature of truth-telling and confessions, she exceeds his observations on confession logics by emphasizing the merits of cascading accounts.⁷⁵⁵ Narrative assists with this mission by complicating the types of information that are presented into a judge's frame of reference.

Through this narrative lens, the interpretative exercise required by law's epistemological processes is transformed from fact-finding into storytelling. Facts are often at the core of judicial or administrative processes: significant procedural attention is paid to fact-finding, ranging from cross-examination at trial to deference to lower tribunals. If we understand facts as mere datapoints, divorced from the social contexts in which they take place, self-driving laws might make the same errors as Bertrand Russell's inductivist turkey, poised to look to the future based only upon the past. Understanding facts to be bound up in the stories we tell in order to make

⁷⁴⁹ Dworkin, *Empire*, *supra* note 19 at 239.

⁷⁵⁰ Cover, "Violence," *supra* note 394 at 1625.

⁷⁵¹ Mayo Moran, *Rethinking the Reasonable Person: An Egalitarian Reconstruction of the Objective Standard* (New York, NY: Oxford University Press: 2003).

⁷⁵² *Ibid* at 164.

⁷⁵³ *Ibid*.

⁷⁵⁴ Sarah Banet-Weiser, "Ruined lives: Mediated white male victimhood" (2021) *European Journal of Cultural Studies* 60 at 63.

⁷⁵⁵ *Ibid* at 76.

sense of the world inoculates them against misleadingly neutral presentations. Things cannot be presented as truly neutral or descriptive when understood to be the constituent components of stories—*i.e.*, constitutive rhetoric, or narrative.⁷⁵⁶

c. Dworkinian Takeaways

Dworkin's theory offers a helpful starting point for my Rule of Law inquiries. His recognition of law's interpretative aspects demonstrates the degree to which law's epistemic inquiries are linguistic processes: the construction of legal knowledge depends on the development of certain cooperative fictions.⁷⁵⁷ To maintain their explanatory power, legal explanations offer social cooperation through the marshaling of abstract concepts into cohesive stories. The issues that Dworkin had with the positivists' legal rules offer coherent explanations for why the self-driving laws project misunderstands how legality operates: Dworkin's interpretative approach treats the doctrinal aspects of law as fundamentally interpretative. While the positivists would generate tests for whether and to what extent lawyers and legal officials have adhered to doctrinal propositions of law,⁷⁵⁸ Dworkin's interpretivism mandates a co-creation element, where the various actors of the legal system are working together to create the stuff and structure of law. This approach is helpful for understanding both the concept and the Rule of Law as being participatory. In my view, to merely disagree with the positivists to demonstrate dissatisfaction with self-driving laws fails to fully elucidate the specific challenges I see as being most troubling in such proposals. By contrast, Dworkin's understanding of the concept of law as an interpretative concept begins to cast doubt on the viability of the self-driving laws project, even without further complicating the field.

However, simply adopting Dworkin's worldview as a theoretical framework is not the way forward either. Dworkin's insistence on cohesiveness presents challenges for how to negotiate meaning at the edges of what law might be or what experiences in society might look like. Even if most statistically significant behaviour operates at the centre of a bell curve, a complete account of reality will still need to deal with the outliers: the bell curve's tails. In Dworkin's appeals to single, correct answers—to ground truth—he mistakes consistency with uniformity. His adamance against the inclusion of social scientific data or critical narratives in his chain novel analogy supposes the simplistic brush with which he would prefer to paint reality.⁷⁵⁹ His later work, while more nuanced, still allows for the same types of criticisms to be levied against an overly-rosy one-system approach. Seeing the challenges of attacking the two-system dichotomy of positivism *versus* interpretivism that he wrestled with his whole career, he finally attempted to unite them into a one-system unifying theory. He summarizes his updated worldview in *Justice for Hedgehogs*.

When more than forty years ago I first tried to defend interpretivism, I defended it within this orthodox two-systems picture. I assumed that law and morals are different systems of norms and that the crucial question is how they interact. So I said what I have just said: that the law includes not just enacted rules, or rules with pedigree, but justifying principles as well. I soon came to think, however, that the

⁷⁵⁶ White, *Heracles' Bow*, *supra* note 15 at 39.

⁷⁵⁷ Fairfield, *Runaway*, *supra* note 50 at 187.

⁷⁵⁸ Or at least, this is how Dworkin characterizes the doctrinal approach of the positivists. See Dworkin, *Hedgehogs*, *supra* note 707 at 402.

⁷⁵⁹ Dworkin, *TRS*, *supra* note 24 at 252-253.

two-systems picture of the problem was itself flawed, and I began to approach the issue through a very different picture.⁷⁶⁰

Dworkin then goes on to pull together his various threads of exploration, including politics and morality, to form a singular account. The details of his unified approach are unimportant here; what matters, in my view, is the readiness with which he generates a simple account of a highly complicated system. Yet valuing equality, as Dworkin purports to do, might require acknowledging the deleterious impacts of law's uniform application across sectors and social groups, which generates bias. As Moran writes, “[t]he very centrality of some idea of equality across cases ... suggests that we must attend to the possibility that there will come a point at which inequalities in the application of the law do raise a rule of law concern.”⁷⁶¹ Even though the precept that “like cases must be treated alike” is usually presented as an uncontroversial aspect of the Rule of Law,⁷⁶² its consequences might not be straightforward, especially for how we understand rule-following.

This critique presents its own problems for the advancement of legal theory. However, when applied to the unique challenges presented by self-driving laws, the theoretical approach merges with calls for empirical evidence and quantification. Overemphasizing the “correct” answer that law might offer misapprehends the overall project of law to negotiate coherent fictions that can operate at the societal level. Unduly concentrating on data-driven analyses forces a solutionist, empirical, and pragmatic approach to law that discounts its normative dimensions. Law's method requires a narrative imagination. In this way, the construction of rationality is a social, not a deductive, exercise.⁷⁶³ Dworkin's theories elegantly present law as interpretation to make sense of how judges think and behave, but broad sweeping explanations only go so deep. Joshua Fairfield assists in this indictment of abstract reasoning with his example of the “illusion of explanatory depth.”⁷⁶⁴ He observes that we think we understand how complicated things like zippers or home plumbing work, until the moment when we are asked to really explain them.⁷⁶⁵ Applying this concept to law, it seems that Dworkin's theory might offer a cohesive account of law's empire, but plumbing the explanatory depths of this approach reveals that it only adheres at the surface level.

If we accept Dworkin's interpretative approach as a theoretical framework, then the unification of all interests lends credence to the availability of “correct” legal answers. The potential application to self-driving laws presents concerning plausible futures. In the far-off vision of the legal futurists, if we succumb to the android fallacy, law might adopt a continuously updating, generative mode, capable of adaptation or even personalization. Yet, this future could very well impose the political morality and embedded systemic biases of the corporate interests that have produced its operative datasets and algorithms, consolidating power in worrying ways. Nuanced aspects of rule-following and the contested nature of the Rule of Law would give way to specific visions of those ideals. However, in a more accurate representation of technology's current capabilities, the datafication of law's precedents will set them in stone in a way never seen before. Assuming that legal knowledge is complete will collapse any sense of narrative imagination in legal reasoning. This

⁷⁶⁰ Dworkin, *Hedgehogs*, *supra* note 707 at 402.

⁷⁶¹ Moran, *supra* note 751 at 169.

⁷⁶² Moran discusses this formulation in reference to John Rawls' political philosophy. *Ibid* at 171.

⁷⁶³ Fairfield, *Runaway*, *supra* note 50 at 99.

⁷⁶⁴ *Ibid* at 96.

⁷⁶⁵ *Ibid*.

centralization will remove the chain novel feature altogether, imposing one singular author who presumes to anticipate all eventualities. Hercules will be writing all laws from here on in, and even though Dworkin believes he is a careful judge, he will set the tone going forward with a single moral compass and a singular voice.

With Dworkin's theory—and its problems identified thus far—as a jumping-off point, I further elaborate upon the need to generate cohesive legal fictions through the approach put forward by law as narrative. The narrative lens puts the collaborative work of law at the centre of the interpretative exercise. Rather than relegate uncomfortable truths to the edges, as Dworkin's law as integrity might do, narrative approaches re-centre the lived experiences of those who might otherwise stay anchored at law's margins. Throughout Chapter 5, I complicate Dworkin's theoretical approach to law to better elucidate the conceptual problems faced by the project of self-driving laws. Before going there, however, I offer few final thoughts on the Rule of Law as a benchmark and its intersections with law as narrative.

3. Rule of Law Conclusions: A Call for Storytelling?

If there are competing conceptions of the Rule of Law, as would seem to be the case, the corresponding question is obvious: which one should we choose? Perhaps choosing one particular singular voice to tell the story is the wrong approach.⁷⁶⁶ The Rule of Law, and the concept of law itself, remain—at their core—contested concepts.⁷⁶⁷ Waldron describes the Rule of Law as “multi-faceted.”⁷⁶⁸ The many efforts that have been undertaken to describe it, to map its contours and determine its extent, to invoke it as justification for myriad factors and experiences, to offer a cacophony of explanations: these should be seen as a feature rather than a deficiency. We could no sooner choose one particular conception of the Rule of Law than we could accept one citizen's experience as a determinative, universal claim on the experience of being governed.

Because legal theories aim to smooth out wayward edges to produce a singular, universally applicable approach, they tend to avoid making claims that are overtly political or normative.⁷⁶⁹ Despite Dworkin's efforts to account for politics, his commentary on critical legal studies suggests his commitment to a neutral, objective account of legality—at any cost.⁷⁷⁰ And, despite the appeals to universality advocated by some theorists, not enough heavy lifting explains the relationship between the Rule of Law as a theoretical construct worthy of protection and the individual experiences most citizens might have *vis-à-vis* the force of law. Many citizens will go their entire lives without being arrested, sued, or even appearing on a jury. Others will be in constant contact. Still other citizens' experience of the law is the patchwork combination of rules and regulations imposed by the administrative state and its many tribunals. This complexity and variation become crucially important when faced with technological architectures comprised of default settings and

⁷⁶⁶ This is the type of critique Stanley Fish offered Ronald Dworkin, arguing that all texts are amenable to multiple interpretations and meanings, with no clear authority. Stanley Fish, “Working on the Chain Gang: Interpretation in Law and Literature” (1982) 60 Tex L Rev 551.

⁷⁶⁷ Even Raz acknowledges at the outset of his updated work that a lack of agreement; instead of privacy, he uses the example of freedom of speech. Raz, “Law's Own Virtue,” *supra* note 520 at 1.

⁷⁶⁸ Waldron, *supra* note 629 at 6.

⁷⁶⁹ Rundle, *supra* note 680 at 11-12.

⁷⁷⁰ Dworkin accuses critical legal scholars of “self-conscious leftist postur[ing]” and, as excerpted above, notes that “nothing is easier or more pointless than demonstrating a flawed account fits as well as a more attractive one.” It is clear, in his estimation, that his theory is the more attractive one. Dworkin, *Empire*, *supra* note 19 at 271-274.

limited options, which then constrain mechanisms for citizen participation in building, resisting, and agitating legal norms.⁷⁷¹ Through this lens, technological mediation not only disrupts the Rule of Law but changes the ambit of what might be called law. Such a system is best understood with a broad, expansive worldview, one that not only recognizes the simultaneous function of concomitant legal orders and the contexts in which they apply, but also technological enforcement that resembles law in its ordering and impacts. In this way, technology expands law's empire, as the nature of technological systems require people to conform and obey in much the way laws do. If this expansion of law's empire is to be achieved through self-driving laws, the potential normative consequences are endless.

In the face of increasing automation in the legal field, some have proposed identifying certain areas where decision-making is simply too important to be delegated to algorithms. Using human judgment to insulate certain areas of decision-making from algorithmic intervention, discussed in Chapter 3, might still seem like an attractive remedy. However, the creation of two tiers of fact-finding or judicial reasoning presents problems on an epistemic level. The judicial system is preoccupied with finding or propagating truth. Fundamental components of the system, like discovery, undertakings, and cross-examination, presume the possibility of the attendant lawyers or, at the very least, the court itself, identifying what really happened in the events leading up to the legal dispute. As detailed above, this approach can be criticized from an equality-seeking perspective, where the veneer of objectivity required by truth-seeking processes replicates the world as it is, further entrenching the status quo, rather than envisioning the world that could be.

The task of constructing reality is a Herculean one. Perhaps this is why Dworkin gave this name to his all-powerful judge. Dworkin's deification of Hercules is at odds with his desire to construct a legal framework that does not deal in morality. By operating as an ideal for what judging could be, and imposing a uniform structure on law's empire, Hercules functions as omnipotent—almost like a kind of God. Even though Dworkin wants the reader to conclude that Hercules is above playing politics, he nonetheless acknowledges that politics arise from people exerting power on other people, which has to happen outside of his Herculean ideal. The elegance proffered by Dworkin's Hercules falls short of a complete vision of reality: in its efforts to offer a whole-system approach, it smooths out the many imperfections and varied blemishes that are common in human life. Such messiness offers a more convincing representation of reality. Instead, Dworkin's efforts to construct an overly shiny and polished version of our messy reality coincide with programmatic attempts to model reality in the AI/ML space. As computers process vast quantities of data to produce the models they claim represent reality, they similarly engage in a smoothing process to create something far slicker than the unvarnished truth. As Kate Crawford explains:

Yet now it's common practice for the first steps of creating a computer vision system to scrape thousands—or even millions—of images from the internet, create and order them into a series of classifications, and use this as a

⁷⁷¹ Ian Kerr, "The Devil Is in the Defaults" (2017) 4:1 *Critical Analysis of Law* 91 at 94. See also this discussion on NPR's "Fresh Air" between host Terry Gross and technology journalist Farhad Manjoo about overwhelming control exerted by the "Frightful Five"—Apple, Amazon, Google, Facebook, and Microsoft—the most valuable companies in the world. NPR, "How 5 Tech Giants Have Become More Like Governments Than Companies (26 October 2017), online: *NPR* <www.npr.org/2017/10/26/560136311/how-5-tech-giants-have-become-more-like-governments-than-companies> [perma.cc/QQ4Q-YXKW].

foundation for how the system will perceive observable reality. These vast collections are called training datasets, and they constitute what AI developers often refer to as ‘ground truth.’ Truth, then, is less about a factual representation or an agreed-upon reality and more commonly about a jumble of images scraped from whatever various online sources were available.⁷⁷²

As I read Crawford’s insightful passage on the construction of truth by a machine learning developer, I was struck by how the world-building processes from which an AI developer might construct ground truth resembles that of how legal inquiry arrives at a similar conclusion. For example, consider the different standards of proof at play in criminal law proceedings (beyond a reasonable doubt) *versus* civil law proceedings (balance of probabilities). In these two spheres, the expectations for the amount of information needed to draw an appropriate legal conclusion are related to the severity of the impact on the accused person. In a criminal law case, the possibility of an accused person losing their freedom demands the higher evidentiary standard of beyond a reasonable doubt. The character of the epistemological inquiry justifies the potential impact on the accused person’s liberty interests. Truth, then, becomes less about what really happened, and more commonly about what can be credibly constructed from whatever evidentiary sources were available.

Crawford’s meditation on the availability of truth evokes similar insights from Nietzsche:

What then is truth? A movable host of metaphors, metonymies, and anthropomorphisms: in short, a sum of human relations which have been poetically and rhetorically intensified, transferred, and embellished, and which, after long usage, seem to a people to be fixed, canonical, and binding. Truths are illusions which we have forgotten are illusions—they are metaphors that have become worn out and have been drained of sensuous force, coins which have lost their embossing and are now considered as metal and no longer as coins.⁷⁷³

Nietzsche offers another lens on the construction of truth. His perspective reminds us of the cooperative fictions that are built and maintained to harness the operational power of law.⁷⁷⁴ So much of law’s power adheres in narratives, including courtroom dramas, journal articles, and judicial decisions.⁷⁷⁵ Technological encroachment into legal spaces has long required the application of reasoning by analogy and use of metaphor to attach novel innovations to past practices. But, as noted above, metaphors can sometimes impoverish legal reasoning, as they strain

⁷⁷² Crawford, *Atlas*, *supra* note 76 at 96.

⁷⁷³ Friedrich Nietzsche, *On Truth and Lies in a Nonmoral Sense* (1873), online (pdf): *AustinCC* <www.austincc.edu/adechene/Nietzsche%20on%20truth%20and%20lies.pdf> [perma.cc/2HF2-KVRA].

⁷⁷⁴ A good example of this is the “beyond a reasonable doubt” standard. Today, it is extolled as a virtuous maxim that emphasizes the criminal law’s caution to never imprison anyone unless their guilt seemed nearly certain. When I was in law school, it was promoted as a beautiful concept in honour of justice, *i.e.*, that our vaunted justice system would permit 1000 guilty men to walk free before ever allowing a single innocent one to languish in prison. In reality, the maxim originated at a time when the citizens on juries were highly concerned with the religious consequences of wrongfully imprisoning someone. They worried their own souls would hang in the balance if they erred, since wrongly convicting someone was a mortal sin, punishable by damnation. To improve the experience of those serving on juries, the “beyond a reasonable doubt” standard was introduced to assuage the jurors’ consciences. See James Q Whitman, *The Origins of Reasonable Doubt: Theological Roots of the Criminal Trial* (New Haven, CT: Yale University Press, 2008) at 4.

⁷⁷⁵ Fairfield, *Runaway*, *supra* note 50 at 162.

to encompass novel applications far beyond their original scope; as Justice Benjamin Cardozo once observed, “[m]etaphors in law are to be narrowly watched, for starting as devices to liberate thought, they end often by enslaving it.”⁷⁷⁶ Recognizing metonymy in the construction of law’s narratives further complicates the concept of “ground truth” as proposed by the legal futurists. Cognitive linguist George Lakoff describes metonymy as “one of the basic characteristics of cognition.”⁷⁷⁷ He offers the example of a housewife-mother stereotype, showing how it works as a metonymic model to represent a category as a whole.⁷⁷⁸ He shows how such social stereotypes are both inaccurate and, simultaneously, capable of defining cultural expectations.⁷⁷⁹

Such observations underscore law’s role in constructing meaning, through its inevitable interventions in both language and culture. As Fairfield astutely observes, some of modern law’s crowning achievements are “highly fictive stories about how things ought to be.”⁷⁸⁰ For contested concepts like the Rule of Law, the complexities of legal meaning *versus* everyday usage are exacerbated by classification methods and datasets that entrench biased societies, the types of challenges elucidated in Chapter 3. Assuming a whole system approach can both absorb longstanding biases and still generate a just result is somewhat unbelievable. For Dworkin’s interpretative method, this might mean stepping away from whole system approaches to explain complicated ideas. Perhaps the best possible version of interpretation offers another way of knowing, one where truth is negotiated instead of simply found.⁷⁸¹ Instead of anemic metonyms, we need to seek a more robust version of truth.

Dworkin’s storytelling approach—the chain novel—is acontextual: it presumes that the context on the ground was already lapped up by previous cases. What if we are presented with a totally new situation? This past-focused storytelling approach might not be enough.⁷⁸² Automating decision points presumes that there is a regularized approach, one that can be taken out of context and generalized for other situations: a list of steps than can be followed, much like a recipe. In other contexts of automation, like robot-assisted surgery or self-driving cars, a human-in-the-loop often remains necessary, in case something goes wrong. Given that hard cases can defy even the most robust legal reasoning, perhaps there is a similar need for human-in-the-loop in legal decision making. And, perhaps, this function can be achieved through insisting upon storytelling, in the manner hypothesized by critical theorists. We take up this hypothesis in Chapter 5.

⁷⁷⁶ Kerr, “Metaphors,” *supra* note 358 at 232.

⁷⁷⁷ George Lakoff, *Women, Fire and Dangerous Things: What Categories Reveal about the Mind* (Chicago, IL: University of Chicago Press, 1987) at 77.

⁷⁷⁸ *Ibid* at 78-86.

⁷⁷⁹ *Ibid* at 85.

⁷⁸⁰ Fairfield, *Runaway*, *supra* note 50 at 161.

⁷⁸¹ *Ibid* at 187-188.

⁷⁸² Oliver similarly notes that Dworkin’s chain novel analogy struggles with how to adapt social scientific facts for fidelity to interpretation. To regulate real people in real societies in an ongoing and efficacious way might require a more robust version of reflexive equilibrium that offers interplay between principles and real-world context. This contextual approach allows well-meaning principles to achieve coherence in the real-world based on exigent circumstances. Oliver, *supra* note 704 at 253, 257.

*Like law, rhetoric invents, and like law it invents
not out of nothing but out of something.*
~ James Boyd White⁷⁸³

Once, in a break between conference sessions, I was engaged in a bizarre conversation with two colleagues about which of the three movies in the Matrix trilogy is the best. While two of us were firmly of the view that the first movie was the clear winner, our third was putting in a surprisingly compelling case for the second sequel when a fourth conference-goer casually joined our merry group. We welcomed her into our discussion although, as she immediately confessed, she had never seen any of the Matrix movies. Quickly, following a dizzyingly deft sequence of lateral leaps I can neither recount nor replicate, our conversation veered from the Matrix trilogy to science fiction books espousing similar themes, to blockbuster movies generally, to high-action staple Top Gun, to its star Tom Cruise, before finally arriving at Cruise's religion, Scientology. Our new conversation partner laughed: "I came in at the Matrix and now I'm talking about Scientology; I'm out!" As suddenly as she'd joined us, she wandered away, leaving the three of us wondering how we had so quickly traversed this expansive intellectual terrain without first taking the red pill.

Later, I caught up with her. We laughed about how completely—and quickly—she had derailed our desperately nerdy discussion. Why, I asked, had she joined a fervent conversation between three science fiction aficionados earnestly debating movies she had never seen? She told me about a book called *How to Talk About Books You Haven't Read* by French literature professor Pierre Bayard.⁷⁸⁴ He argues that the relations between ideas are more important than the ideas themselves. This meant that, in order to talk meaningfully about something, one really only needs a cursory idea of what any particular book is about in order to place it in the collective library. She asked me whether I had ever read Vladimir Nabokov's *Lolita*. Resisting that imposter syndrome urge to lie to appear more well-read than I really was, I admitted: no, I had not. Still, she pointed out, "I imagine you could explain its central themes." I could, and easily: an older, lecherous man; a younger, possibly teenage girl; an inappropriate relationship between them. Contextually, and from how the phrase *Lolita-esque* had been used in various settings over the course of my life, I knew what the book was about. Similarly, she argued, it was possible to participate in a conversation about movies she had never seen, having absorbed some of their key themes from the zeitgeist, and able to connect the broader ideas to other related topics of conversation.

Of course, the kicker was: she had never read Bayard's book, either! But she had read a book review detailing it and knew enough about its central argument to place it in the collective library.

As we spoke, I started thinking about the implications for a ML-enabled, data-driven legal library, which knows only what it has been taught. How would it make those lateral leaps between seemingly incommensurate ideas? Would our collective library be improved or impoverished by

⁷⁸³ White, *Heracles' Bow*, *supra* note 15 at 39.

⁷⁸⁴ Pierre Bayard, *How to Talk About Books You Haven't Read*, trans by Jeffrey Mehlman (New York, NY: Bloomsbury, 2007).

computational methods attempting to map semantic distance between concepts? Would natural language processing be able to trace subtle, nuanced references to art, history, or literature? What about all the tacit knowledge that we have never really been taught, but glean from merely existing in the world? Would it be successfully replicated in the deep learning models running over repositories of human behaviour reduced to data? Or would our stories be punctuated differently? And what would this mean for the cataloguing of our stories in the collective legal library?

This chapter grapples with the question: whose stories are told? If we understand the law as ripe for adjudication and administration through data-driven approaches, then we must also understand that such data might present us with a particular worldview. Without a critical lens, I argue, there is no reason to believe that a data-driven approach will do anything to ameliorate the very bias it intends to target. To understand law as narrative, however, one must first concede the possibility of some ambiguity in the sources comprising law. If true, this creates a corresponding need for interpretation: law is something more than merely that which is written. Recognizing this ambiguity in the system casts immediate doubt on the project of automated law, which is poorly equipped to handle the messiness inherent in human life. Law already struggles to make space for minority lived experiences; asking for an automated approach necessitates establishing baselines that are bound to struggle with inclusivity and representativeness. If only the dominant stories are told, we will be faced with further entrenchment of privileged worldviews that discredit the inherent variety of human experiences. Automation bias, flowing from the unearned certainty of machine-based solutions, has the potential to transform nuanced traditions of finding truth through storytelling, replacing subtlety with whatever data best fits the operative model. Despite Dworkin's protestations catalogued in Chapter 4, the accounts of critical scholars create meaningful opportunities for stories to travel alongside the development of law.

Law as narrative theorists have infused their scholarship with compelling accounts of history, literature, and personal experience. I begin this chapter by summarizing several influential descriptions of law as a storytelling process—one that recognizes how humans participate in the creation of reality through the very act of describing it.⁷⁸⁵ Acknowledging, however, additional room for equality-seeking perspectives, I then move to a discussion of critical theory, in which I employ the traditions of marginalized and intersectional scholarship to demonstrate why legal thinking benefits from a plurality of perspectives. These traditions, including feminist, Indigenous, and critical race perspectives, showcase the importance of storytelling as method for expanding the potential parameters of truth. Storytelling, I argue, might just mitigate the problems of regulating by algorithm.

1. Law as Narrative Scholarship

Law as narrative begins from a premise that the articulation of stories—the act of forming a cohesive, meaningful narrative—offers a different lens through which to view the creation, structure, and development of the law. In this way, the arc of the development of the law must be compatible with a narrative storytelling structure, in order for it to be fully intelligible to those it proposes to govern. Taking a similar approach, James Boyd White uses this idea to describe law as a system of “constitutive rhetoric,” wherein law functions as “a system of resources for claiming,

⁷⁸⁵ Delgado, “Storytelling,” *supra* note 637 at 2416.

resisting, and declaring significance.”⁷⁸⁶ In his text *Heracles’ Bow*, White presents a series of literary essays that offer a compelling demonstration of the potential expansiveness of law. Through this text, law is revealed as interdisciplinary: the constituent essays bounce from emphasizing the importance of telling stories, to incorporating the semiotics of Wittgenstein, to thinking critically about why we use the case method to teach law in the classroom, to comparing reading judicial opinions with reading poetry, to reimagining a conversation between Socrates and modern lawyers in the style of Plato’s *Gorgias*. This breadth and exploration of rhetoric showcases the uses of writing about law from untraditional vantage points and in accessible styles.⁷⁸⁷

White offers us a powerful explanation for the idea of constitutive rhetoric, likening it to our conceptualization of what the word “fishing” means. If we want to explain that a bear is fishing for salmon in a river, we could, without complication, describe that activity as “fishing.”⁷⁸⁸ Yet, if a man were at the same river engaged in the very same activity, saying that he was “fishing” could potentially connote so much more than it might for the bear.⁷⁸⁹ People can engage in all kinds of complicated behaviours, with varying motivations and complexities rationalizing their actions. Is the fishing person engaging in a “sentimental escape to the wilderness, [...] sportily clad in [an] L.L. Bean outfit, demonstrating his place in a certain social class”?⁷⁹⁰ And how would that person’s “fishing” be contrasted with the same activity undertaken by an Indigenous person, who might have a traditional link to the activity, a spiritual connection with the land, and a relationship with the fish while fishing? These kinds of background factors complicate the expression of what seems like a simple fact: the man was fishing. Attaching a story to the fact also frustrates the representation of that fact as a mere data point.

It is easy to imagine additional motivations and complexities for the fishing person. Perhaps it is a woman, overcoming prejudice by participating in what in many societies has been a traditionally male-dominated activity. Perhaps it is a third-generation fisherperson in a small Newfoundland town, trying to make a living in the same manner as their forebears, despite declining cod stocks. Perhaps it is a large commercial trawler, decimating the local ecosystem for profit. White’s example brings home the distinctions that can arise when different people, despite being engaged in an activity that might be referred to using the same umbrella term, assign different meanings to that term. As White emphasizes, understanding whether there is “tension or disharmony” in the various uses of a shared term is fundamental to establishing shared meaning—to “mak[e] the kind of community that enables people to say ‘we’ about what they do and to claim consistent meanings for it.”⁷⁹¹ The creation of shared meaning becomes a key feature for law, especially through its magic solving words,⁷⁹² which imbue specific meanings. Approaching meaning through rhetorical

⁷⁸⁶ White, *Heracles’ Bow*, *supra* note 15 at 39.

⁷⁸⁷ White aims to confront the question of what we are building, in law, when we say what we mean. This is much like Gerald Wetlauffer’s approach to the denial of rhetoric in law, when he writes: “But if law is, at its core, the practice of rhetoric, the *particular* rhetoric that law embraces is the rhetoric of foundations and logical deductions. And that particular rhetoric is one that relies, above all else, upon the denial that it is rhetoric that is being done.” Gerald B Wetlauffer, “Rhetoric & its Denial in Legal Discourse” (1990) 76 Va L Rev 1545 at 1555.

⁷⁸⁸ White, *Heracles’ Bow*, *supra* note 15 at 37.

⁷⁸⁹ No disrespect to the bear, who might have all kinds of motivations that remain opaque to human thought processes.

⁷⁹⁰ White, *Heracles’ Bow*, *supra* note 15 at 37.

⁷⁹¹ *Ibid.*

⁷⁹² Felix Cohen uses the phrase “magic solving words” to refer to phrases that have special operating power within the different spheres of jurisprudence. Such phrases are often somewhat meaningless, in that they cannot be described from experience, and

analysis, he argues, enables a manner of addressing collective knowledge in a “tacitly creative” way.⁷⁹³ Taking cues from artistic pursuits, bids on meaning achieved rhetorically offer creative solutions that are organized and consistent, without being rule-bound.⁷⁹⁴ This expanded sense of rhetoric “directs our attention to the most significant questions of shared existence, which are wholly outside the self-determined bounds of science.”⁷⁹⁵ As we make sense of the legal universe through rhetorical processes, we are granted questions to ask that assist in generating cohesion of meaning, in light of cultural forces.

White’s erudite explanation of rhetorical function is enriched by literary allusions to Homer’s *The Iliad* and Milton’s *Paradise Lost*.⁷⁹⁶ In particular, White draws a comparison between the fallen angels of *Paradise Lost*, unable to create community out of their language of selfishness and hatred, with the challenges faced by the participants of the first American Constitutional Convention in 1787.⁷⁹⁷ The same types of questions are faced by both the angels and the would-be legislators: “[w]hat kind of community shall it be? How will it work? In what language shall it be formed?”⁷⁹⁸ These, White argues, are questions of rhetorical analysis, asking us to grapple with great questions of justice and ethics. Is this not the core purpose of law, one might ask after reading his assessment?

The case method of learning common law in law schools seems to resonate with this narrative form.⁷⁹⁹ In learning to cultivate legal imagination, law students learn a body of knowledge tantamount to being introduced to new cultural practices. The similarities to an epic story or narrative arc are significant, as exposure to law’s body of work expressed through cases resembles learning a culture’s foundational norms expressed through parables. Robert Cover explains:

No set of legal institutions or prescriptions exists apart from the narratives that locate it and give it meaning. For every constitution there is an epic, for each decalogue a scripture. Once understood in the context of the narratives that give it meaning, law becomes not merely a system of rules to be observed, but a world in which we live.⁸⁰⁰

Cover asks us to understand the law as built by narratives. These narratives impose insistent demands on our comprehension of the structure of the social milieu. Generating a great legal tradition requires going beyond function or structure—or what Cover terms “technical virtuosity.”⁸⁰¹ This disambiguation separates some deeper undercurrent of law from its more perfunctory regulatory aspect. Instead, legal excellence is achieved through establishing a rich, situated *nomos*. Cover’s use of the term *nomos* is interesting, as it is from the ancient Greek meaning law, but it connotes an especially expansive version of law that also embraces the norms and customs governing human behaviour. The term *nomos* is often associated with Peter L.

require that meanings be appropriately defined. Felix Cohen, “Transcendental Nonsense” (1944) 2 ETC A Review of Legal Semantics 82 at 91.

⁷⁹³ White, *Heracles’ Bow*, *supra* note 15 at 44.

⁷⁹⁴ *Ibid.*

⁷⁹⁵ *Ibid.*

⁷⁹⁶ *Ibid* at 38-39.

⁷⁹⁷ *Ibid* at 39.

⁷⁹⁸ *Ibid.*

⁷⁹⁹ *Ibid* at 49-59.

⁸⁰⁰ Cover, “Nomos,” *supra* note 9 at 4-5.

⁸⁰¹ *Ibid* at 6.

Berger's concept of "a meaningful order imposed on individual's discrete experiences" in the sociology of religion.⁸⁰²

While Cover emphasizes the interpretative practices of law that hold the normative world together, he takes care to distinguish this from more quotidian distinctions between theory and practice. He is looking specifically at human behaviours in relation to norms, which are not only demanded by law or regulation, but also offer methods of communicating with one another within our social milieu. For example, he distinguishes between snacking and desecrating Judaism's sacred Yom Kippur fast; he differentiates between depositing funds in your bank account and explicitly refusing to pay your taxes.⁸⁰³ On both sides of his examples, the same actions—eating and keeping money in the bank—would be performed. Yet, understanding each of these actions relative to the existence of specific norms imbues them with different meanings. Through these examples he elucidates "the capacity of law to imbue action with significance."⁸⁰⁴ He writes:

Law is a resource in signification that enables us to submit, rejoice, struggle, pervert, mock, disgrace, humiliate, or dignify. The sense that we make of our normative world, then, is not exhausted when we specify the patterns of demands on us, even with each explicated by Hercules to constitute an internally consistent and justified package. We construct meaning in our normative world by using the irony of jurisdiction, the comedy of manners that is *malum prohibitum*, the surreal epistemology of due process.⁸⁰⁵

Cover's phrase, the "surreal epistemology of due process," asks us to consider the knowledge that is built through the operation of the legal system. Here, we might recall the emphasis on process built by the administrative state, as described in Chapter 3. Building legal knowledge through a prescribed course of action allows us to make careful, successive updates to the project of social control. Yet, in my view, the automation of those same functions—their *machine*-ations—has the potential to alienate us from the formative decisions that bind society. Already, the impartial ideal of justice is impugned by racism, classism, misogyny, and other structural deficiencies that have disenfranchised individuals who are part of minoritized communities. When automation is proposed, it is usually suggested as offering increased consistency or reliability to legal decisions.⁸⁰⁶ Encapsulated in this goal is an objective of increased efficiency—an efficiency relentlessly pursued irrespective of the normative goals sought by law's reflexivity.⁸⁰⁷ Cover's explanation offers further insight into the important role that narratives serve:

The precepts we call law are marked off by social control over their provenance, their mode of articulation, and their effects. But the narratives that create and

⁸⁰² Peter L Berger, *The Sacred Canopy: Elements of a Sociological Theory of Religion* (New York, NY: Doubleday, 1967) at 19. Cover references Berger's scholarship in numerous introductory footnotes.

⁸⁰³ Cover, "Nomos," *supra* note 9 at 8.

⁸⁰⁴ *Ibid.*

⁸⁰⁵ *Ibid* at 8-9.

⁸⁰⁶ Sometimes, ML is presented as the antidote to unconscious bias, systemic discrimination, or inconsistency that human decision-making inevitably creates. Proponents adopt a decidedly technological solutionist worldview, asserting that datafied approaches allow insight into otherwise opaque decisioning processes, ensuring repeatability and the removal of irrelevant factors. See Jon Kleinberg et al, "Human Decisions and Machine Predictions" (2017) 133 *Quarterly Journal of Economics* 237.

⁸⁰⁷ Willem DeVries, "Wilfrid Sellars" *Stanford Encyclopedia of Philosophy*, online: *SEP* <plato.stanford.edu/entries/sellars/> [perma.cc/BQM9-M8H6].

reveal the patterns of commitment, resistance, and understanding—patterns that constitute the dynamic between precept and material universe—are radically uncontrolled. They are subject to no formal hierarchical ordering, no centralized, authoritative provenance, no necessary pattern of acquiescence. Such is the radical message of the first amendment: an interdependent system of obligation may be enforced, but the very patterns of meaning that give rise to effective or ineffective social control are to be left to the domain of Babel.⁸⁰⁸

Describing the first amendment (the United States' free speech law) operating as an interdependent system controlled through emergent patterns of meaning offers insight into Cover's views on narrative's power. Not only do narratives offer supplementary explanations for why legal orders adhere within society, but the adhesion is propagated by the chaotic contents of the narratives themselves. These narratives may, at first blush, seem contradictory, but there is significance revealed through the contrasts. Within the contingent system of legal obligations, social power is made effective or ineffective through the deployment of either more or less compelling narratives, arising organically from "the domain of Babel." Cover offers thorough explanation for his use of the term "Babel," writing in a footnote:

I use the term Babel advisedly. It suggests not incoherence but a multiplicity of coherent systems and a problem of intelligibility among communities. If law is given meaning through mythos, and if the domain of mythos is characteristically narrower than that of precept, we are indeed in Babel. Dworkin's concerns converge in some ways with those expressed here. In his later work, Dworkin concedes the open character of the materials to which the "Herculean" judge appeals in reaching the "right answer." This openness is tantamount to the preconditions for the "Babel" I posit in the text. Dworkin's chain novel analogy suggests the intelligibility, through retrospective harmonization, of any single interpretative effort even though it be interpersonal in character. But like the "Herculean" dimension of Dworkin's jurisprudence, the chain novel concept ignores the problem of interpenetrability or comprehensibility between interpretive efforts or traditions, each of which is independently defensible or even "right."⁸⁰⁹

Cover's exploration brings us back to Dworkin. At their core, Dworkin's critiques about legal positivism rely on the exacting nature of the rules that judges claim to rely on when they form their legal opinions. As Cover points out, Dworkin's critiques do not deny that social control may contribute to the articulation of those rules, or the special control that flows from judicial decision-making.⁸¹⁰ However, Dworkin's efforts to achieve coherence through the chain novel analogy discount the unique claims on truth that may be put forward from different perspectives or traditions. This disregard is unsurprising, given Dworkin's motivation to find a robust explanation for legal interpretation without the availability of discretion for hard cases.

⁸⁰⁸ Cover, "Nomos," *supra* note 9 at 17.

⁸⁰⁹ *Ibid.*, n 45.

⁸¹⁰ *Ibid.* (In n 44, Cover describes Hart's view of judging as one particular mode of privileged precept articulation; while Dworkin disagrees with Hart's characterization, this does not negate other mechanisms of social control).

Cover's observation that law consists of uncontrolled, non-hierarchical patterns sits uncomfortably alongside proposals for self-driving laws, which presume the ability to predict legal futures based solely on the past. His account of law suggests a participatory social exercise, wherein the structure designed by legal rules offers society meaning.⁸¹¹ The realm of Babel that creates our dominant mythos is inseparable from our cultural milieu. He writes: "law becomes not merely a system of rules to be observed, but a world in which we live."⁸¹² Law's normativity therefore requires the incorporation of narratives, especially those that might not be recognized by majority society. Narrative, then, is an interpretative tool that we engage when imposing normative force on the world around us. Self-driving laws' affordances would not easily carve out space for the insertion of such narratives—especially critical ones.

2. Lessons from Critical Theory

a. Feminist Narratives

Many feminist scholars believe that majority society has quieted, or altogether failed to acknowledge, women's voices. Majority society proceeds from a male viewpoint: women's perspectives have not been incorporated into a robust understanding of the human experience. The prototypical denizen of the so-called liberal state approaches life's problems from an individualized, inherently male perspective. Such assumptions animate mainstream legal scholarship and liberal humanism. In her seminal *Towards a Feminist Theory of the State*, Catharine MacKinnon asserts that taboo topics are hierarchy in disguise.⁸¹³ In successive chapters, she attacks important issues for the lives of women—abortion, rape, pornography, and sex equality—and analyzes women's legal treatment in each area. By approaching these topics in this way, she emphasizes the unique aspects of how women move through the world and the male hegemony that discounts lived female experiences, even in these areas of concern that are far more urgent and relevant for women's lives.

One of MacKinnon's most powerful observations is her description of the "point-of-viewlessness" at the centre of liberal accounts of rationality or objectivity.⁸¹⁴ Efforts to reckon with the legitimacy of law are hamstrung by a reasonableness analysis that counts anything as reasonable if it corresponds to the way that things already are.⁸¹⁵ Law's normativity, and the state of affairs posited by the Rule of Law, evoke a similar point-of-viewlessness. With this in mind, I propose that a relational account, such as the framework developed by Jennifer Nedelsky, helps us to better understand the situatedness of human experience that reveals itself through data. Nedelsky reminds us that once people get into the habit of relational thinking, it is easier to see that the very personal relationships and choices that are epitomized as being "individual" are the product of much larger systems.⁸¹⁶ To think of our choices as truly personal is to remain unaware of the influence enforced by the broader construct of society.

⁸¹¹ Cover, "Nomos," *supra* note 9 at 17.

⁸¹² *Ibid* at 5.

⁸¹³ MacKinnon, *supra* note 630 at 200.

⁸¹⁴ *Ibid* at 162.

⁸¹⁵ *Ibid*.

⁸¹⁶ Jennifer Nedelsky, *Law's Relations: A Relational Theory of Self, Autonomy, and Law* (Oxford, UK: Oxford University Press, 2011).

Compellingly, Nedelsky uses relational framing to indict society's treatment of violence against women.⁸¹⁷ As she reflects on the failures of the liberal state to protect women and children from violence, she spurs a reconceptualization of state power. She, too, finds parallels with the work of Robert Cover and the violence that enabled policies of "racial subordination in the American South."⁸¹⁸ In a pre-emptive response to critics, she notes that most accounts of liberalism would likely criticize her perceived expansion of the confines of the state and would "see taking on the kind of transformation I have in mind as dangerously enlarging the appropriate scope of the state with vague, open-ended and inevitably contested objectives, thus inviting both intrusion and expanded state power to which no clear limits could be drawn."⁸¹⁹ Rushing towards state expansion offers its own challenges, given the threat that many folks, already otherwise marginalized by the state, feel in the shadow of state power.

Writing about the criminalization of sexual violence, Sara Deer and Abigail Barefoot note that using state power to confer safety unduly extends carceral logics by strengthening the state's punitive abilities and increasing the possibility of violence against marginalized populations.⁸²⁰ Scholars like Beth Richie specify that the efforts of the feminist anti-violence movement to extend policing and arrests as tactics for ensuring women's safety has disproportionately poor outcomes in communities of colour.⁸²¹ Such critiques underscore the importance of truly intersectional policy-making to achieve stability and cohesiveness; communities that have traditionally suffered at the hands of the state are not eager to see it expanded to then deal with their long-standing challenges. The oppressor, or so the argument goes, cannot properly liberate the oppressed with its uninterrogated oppressive policies.

Yet such critiques, in my view, do not negate the utility of relational thinking as a centerpiece of liberal feminist thought—and its potential utility to other areas of inquiry. To help put Nedelsky's framework into a technological context, consider privacy as a legal concept. Most modern privacy approaches require consent from the impacted individual in order to collect, use, or disclose information about them. This is a key criterion in Canadian privacy legislation, inspired by the OECD's ten principles on sound data collection, and used all over the world.⁸²² Daniel Solove calls such activities privacy self-management, where there is a focus on individual behaviour, such as obtaining consent, rather than larger constructs.⁸²³ Yet, simply arguing that individuals have the capacity to "opt out" negates the larger systemic pressures at play and the default settings of most digital services provided by the major technology companies; this is true for many data-collecting activities in modern life. Ian Kerr wrote about this phenomenon in an effort to situate privacy as a relational concept.⁸²⁴ This reasoning does much to elucidate the shortcomings of the liberal state for systemic concepts: to better understand which privacy interests are implicated, we need to understand how the information might aggregate, what insights are being gleaned from that

⁸¹⁷ Nedelsky, *supra* note 816 at 201.

⁸¹⁸ *Ibid.*

⁸¹⁹ *Ibid.*

⁸²⁰ Sara Deer and Abigail Barefoot, "The Limits of the State: Feminist Perspectives on Carceral Logic, Restorative Justice and Sexual Violence" (2019) 28 Kan JL & Pub 505 at 510.

⁸²¹ *Ibid.*

⁸²² *PIPEDA*, *infra* note 1037, Schedule 1.

⁸²³ Daniel J Solove, "Privacy Self-Management and the Consent Dilemma" (2013) 126:7 Harv L Rev 1880 at 1882.

⁸²⁴ Ian Kerr, "Schrödinger's Robot: Privacy in Uncertain States" (2019) 20 Theoretical Inquiries L 123 at 151 [Kerr, "Schrödinger"].

aggregation, and how those insights might be used. The possibility for advanced pattern-matching techniques to extract population-level insights from personal data means insulating consent as an individual choice impacting the data subject's insights alone is a naïve way of proceeding. Instead, privacy as a relational construct, where the information is understood as standing in relation to other people's data and being interpreted *en masse* supports a clearer perspective on what the legal right looks like in practice. I return to this analysis in Chapter 7, to better conceptualize the impacts of facial recognition technologies.

Cynthia Khoo employs an analogous tactic, applying the lens of relationality to tort law's foreseeability requirement.⁸²⁵ Inspired by Kerr's application of relationality to privacy law, and Carys Craig's application of relationality to copyright law, she proposes a critical feminist lens to disrupt the traditional point-of-viewlessness that tends to animate these legal doctrines. Through this lens, tort law's foreseeability requirement is expanded to consider the social relationships and cultural realities of dominant actors in the technology law space, which are presented neutrally while relying on outdated tropes of genius.⁸²⁶ Drawing from Frank Pasquale's work on legal automation, Khoo presents the need for relationality to animate the Rule of Law itself, preferring an expansive legal process conception to legal technology's incomplete attempts at defining normativity.⁸²⁷ Taken together, these relational approaches offer a mode of application for some of the finest contributions of feminist legal thought: expansiveness, recognition of complex sociopolitical interrelationships, and legal treatment of interconnected issues that cannot be compartmentalized.

b. Indigenous Narratives

Indigenous legal traditions have long recognized the value of storytelling as a feature of describing and enforcing legal tradition. These traditions offer understandings of law rooted in narratives, which may be instructive for conceptualizing how law propagates through stories. My intention in incorporating Indigenous legal narratives here is to emphasize the diversity of missing or ignored narratives in contemporary Canadian common law. In so doing, I point to a broader conception of what constitutes law, one that might be expanded to better reflect the lived realities of those the law proposes to govern. While Canadian law has made some initial steps to think through legal treatment of Indigenous peoples, most efforts would properly be designated as Aboriginal law—*i.e.*, Canadian law in relation to Aboriginal peoples—not Indigenous law. Constitutionalizing the treatment of Aboriginal peoples, which are specifically defined as First Nations, Métis, and Inuit peoples in section 35 of the *Constitution Act, 1982*, has created a body of law directed at the Canadian state's interactions with people meeting this definition.⁸²⁸ Yet, this constitutional approach does not adequately reflect Indigenous narratives or law. Such tension is consistently engaged in efforts to merge traditional Indigenous legal orders with colonial law. Notwithstanding the SCC's clear recognition of Indigenous legal traditions as primarily oral, Aimée Craft notes that

⁸²⁵ Cynthia Khoo, "Missing the Unintended Forest despite the Deliberately Planted Trees: Reasonable Foreseeability and Legal Recognition of Platform Algorithm-Facilitated Emergent Systemic Harm to Marginalized Communities" (2020) We Robot Conference Paper, online (pdf): <drive.google.com/file/d/1ZVV9yD8KH4KUYLMxE6mlORiCcFKNXIJU/view> [perma.cc/5GSW-RA96] at 11.

⁸²⁶ *Ibid* at 68.

⁸²⁷ *Ibid* at 67. See also Pasquale, "Rule of Persons," *supra* note 47 at 59.

⁸²⁸ *Canadian Charter of Rights and Freedoms*, Part I of the *Constitution Act, 1982*, being Schedule B to the *Canada Act 1982* (UK), 1982, c 11 [*Charter*].

the common law's preference for written legal precepts poses continuous barriers for inclusion.⁸²⁹ Laws of evidence discount oral narratives due to longstanding common law evidentiary standards for necessity and reliability; while adversarial methods allow expert evidence to be tested, questioning truth or reliability through cross-examination creates serious culture clash and potential for disrespect of Elders.⁸³⁰ This discussion reveals a central problem: common law often treats Indigenous stories as matters of fact or evidence for disposition of a matter under common law, rather than true articulations of law themselves.

Different Indigenous groups and scholars describe diverse modes of articulating Indigenous laws. Darcy Lindberg writes about the traditions of his home nation, the nêhiyawak (or Plains Cree) peoples, as expansive and located in institutions other than central enactments or proclamations, including “narratives, songs, artistic renderings, ceremonies, spiritual and place names, kinship models, bundles, and language.”⁸³¹ Craft describes Anishinaabe understandings of law as oral histories; if written at all, they were traditionally written “on the land,” through pictograms, wampum belts, birch bark scrolls, or petroforms.⁸³² Val Napoleon writes about Gitxsan legal personhood being both legally and politically defined through access to certain oral histories.⁸³³ The Gitxsan collective oral history, the *adaawk*, links houses (matrilineal kinship groups) to their land; establishes ownership of the land and its resources; describes covenants made with the land; and explicates cultural artifacts like songs, crests, and names that flow from a spiritual connection with the land.⁸³⁴ Danielle Lussier describes the activity of Métis beadwork as being a visual language that conveys Indigenous knowledge in assorted forms, including as mnemonic device and visual language to communicate stories and songs.⁸³⁵ Lussier articulates how law lives and moves through beading practices and argues for its incorporation in decolonizing legal education.⁸³⁶

These traditional methods for conveying law differ considerably from Eurocentric legal orders. Sarah Morales notes Indigenous law cannot be easily located in a law library, because “law is not only a matter of ideas but rather a process—an activity.”⁸³⁷ Part of the reason why colonizers did not recognize the value of the existing Indigenous legal orders pre-dating European contact was that Indigenous systems of governance did not resemble European law. Without codified statutes

⁸²⁹ Aimée Craft, “Reading Beyond the Lines: Oral Understandings and Aboriginal Litigation” (2013) Canadian Institute for the Administration of Justice Conference at 7.

⁸³⁰ *Ibid* at 10.

⁸³¹ Darcy Lindberg, “Drawing Upon Indigenous Laws in the Yukon” (2020) 50 *The Northern Review* 179, online (pdf): *The Northern Review* <doi.org/10.22584/nr50.2020.007> [perma.cc/E9CR-7XDT].

⁸³² Craft, *supra* note 829 at 5.

⁸³³ Napoleon’s approach to Gitxsan law is informed by Kristen Rundle’s treatment of Lon Fuller’s legal theory. Napoleon observes the relational demands that are constitutive of Fuller’s legalities, especially how governing relationships are framed by law, as analogous to how Gitxsan institutions and processes express legal ways of knowing. See Val Napoleon, “Gitxsan Legal Personhood: Gendered” in Austin Sarat, ed, *Interrupting the Legal Person (Studies in Law, Politics, and Society)*, Volume 87A (Bingley, UK: Emerald Publishing Limited, 2022) 19 at 22.

⁸³⁴ *Ibid* at 4.

⁸³⁵ Danielle Lussier, “Law with Heart and Beadwork: Decolonizing Legal Education, Developing Indigenous Legal Pedagogy, and Healing Community” PhD Dissertation (2021), online (pdf): *UOttawa* <ruor.uottawa.ca/handle/10393/42012> [perma.cc/9Z8P-LRY3] at 276.

⁸³⁶ *Ibid* at 275. She also notes the rich history of beading as part of the Haudenosaunee tradition of wampum belts, which embody treaties and are traditionally read by highly skilled knowledge-keepers but is cautious to situate her key arguments in Métis traditions—her home nation. Such caution and respect are instructive for anyone seeking to engage respectfully with Indigenous legal traditions and pedagogy. *Ibid* at 270-273.

⁸³⁷ Sarah Morales, “Snuw’uyulh: Fostering an Understanding of the Hul’qumi’num Legal Tradition” PhD Dissertation (2014), online (pdf): *University of Victoria* <dspace.library.uvic.ca/handle/1828/6106?show=full> [perma.cc/5CS9-EM5C] at 4.

like the Napoleonic code, or familiar judicial precedents, Indigenous legal systems were unfairly discredited. A critical Indigenous perspective, such as that of scholar John Borrows, proposes that Canadian state law is itself incomplete without the incorporation of Indigenous perspectives.⁸³⁸

Borrows offers a pragmatic approach: while critical of colonial legal hegemony, he proposes a reorganization to include Indigenous perspectives alongside common law and civil law in a trijural Canadian system.⁸³⁹ His call for the incorporation of narrative in law reflects a lengthy tradition of storytelling in Indigenous culture.⁸⁴⁰ In his account of Indigenous legal orders, storytelling is not merely an additive feature to inject lived experiences into legal realities, but a method for propagating and generating law itself. Instead of statutes, there are stories: oral histories, allegories, folktales—generating the apotheosis of untold generations of legal thought and culture. Like MacKinnon, he observes that the purported “point-of-viewlessness” at the heart of liberal society discounts alternative perspectives (and ignores the reality that a privileged white male viewpoint is in fact imprinted on common law). Viewed through a colonial lens, Indigenous traditions were considered to be less tangible than imported common or civil law because they were unwritten, ceremonial, and different from colonial expectations.⁸⁴¹ Yet, it is comparison between systems that renders one superior to another, and the very act of comparison is rigged against diversity: if the European system is defined as reasonable and right, anything that does not mirror the dominant form will necessarily be othered.

While I do not claim expertise in the space of what indigenizing colonial law or the legal academy might look like, I do wish to acknowledge the failures of the status quo in accounting for the lived realities of Indigenous peoples within Canada and the many atrocities perpetuated through the legal system. This subject area offers a provocation for thinking about the creation of legal futures and governance structures that do not seriously reckon with the legal future of Indigenous peoples. If this is true, then the necessary corollary is that it is impossible to simply input existing legal decisions to form a dataset on which to base future predictions. No justice will be derived from this exercise, no matter the empirical exactness, until Indigenous understandings comprise part of the dataset. Yet simultaneously, other scholars disagree that any colonial Canadian law could legitimately include Indigenous legal orders.⁸⁴² Some advocates argue for wholesale Indigenous self-governance, where the most appropriate method is to prevent any inclusion of Indigenous precepts in colonial legal structures whatsoever.

Pursuing justice might yet result in either full inclusion of Indigenous ideals within the civil/common Canadian law structure—or something else that obviates its jurisdiction altogether. Regardless, it is important not to equate these perspectives on Indigenous storytelling with the advocacy of other marginalized groups, who do not contest the jurisdiction of common law and

⁸³⁸ Borrows, *Indigenous Constitution*, *supra* note 722 at 6.

⁸³⁹ The need to create space for Indigenous content within Canadian legal culture for Indigenous legal precepts suffuses Borrows’ text. He calls for a renegotiation of constitutional principles to fully embrace Indigenous law. *Ibid* at 6, 23-24.

⁸⁴⁰ Borrows begins his text with a compelling anecdote of the trickster, Nanabush, following lifecycles of both the natural world and the law on his laptop screen. Nanabush observes the trickiness of the words of power in the law—he both hates them and loves them. *Ibid* at 4.

⁸⁴¹ *Ibid* at 119.

⁸⁴² Some caution that the Truth & Reconciliation Commission’s goals carry significant European colonizer baggage and must avoid framing Indigenous traditions as exact translations for colonial legal concepts. See Kirsten Anker, “Reconciliation in Translation: Indigenous Legal Traditions and Canada’s Truth and Reconciliation Commission” (2016) 33 Windsor YBAJ 15.

civil law itself. Many minoritized and marginalized populations orient their arguments towards making state legal narratives wholly inclusive, as opposed to rejecting its applicability outright. Such arguments, especially in the critical race tradition, are focused on law's more elitist or discriminatory aspects; it is to these arguments that we now turn.

c. Critical Race Narratives

Intersectional scholars are increasingly writing about coded bias in the technology law space, but critical race scholars have levied similar critiques against traditional legal thought for years. To bridge these complementary concepts, I begin with the conception of law as narrative used by Richard Delgado, who centers the diversity of voices telling stories that might constitute law and offers a vision for how storytelling might disrupt prevailing social power dynamics.⁸⁴³ Similarly, Lucie White's analysis of structural discrimination and lower-value language showcases the classification processes at work in coding, and assigning legal value to, different types of speech.⁸⁴⁴ I connect this work to that of critical technology scholars, like Safiya Noble,⁸⁴⁵ Virginia Eubanks,⁸⁴⁶ and Ruha Benjamin,⁸⁴⁷ who write about the continued oppression that datafied approaches create for minoritized communities. Through the work of these scholars, I emphasize the discrepancy between the aspirations of a technologized solution and the lived reality of the impacted population.

Delgado's conception of law as narrative is centered on the diversity of the voices telling the stories that might constitute law. In "Storytelling for Oppositionists & Others,"⁸⁴⁸ he sets out a vision for how storytelling might disrupt prevailing social power dynamics. His astute observations highlight the distinctions in the lived experiences of those who experience inequality in white supremacist culture.⁸⁴⁹ Majority individuals might believe inequality persists due to "cultural lag" or an "inadequate enforcement of currently existing beneficial laws," either of which might be easily remedied.⁸⁵⁰ Conversely, minority individuals observe a majority mindset that allows those in power to continue to casually hold their power without questioning how or why they came to accrue it, and who else might suffer due to its grip on society. He argues that the incorporations of "stories, parables, chronicles, and narratives" serve to destroy that problematic mindset.⁸⁵¹

Methodologically, his piece functions as both narrative and meta-narrative on the importance of adopting a diversity of perspectives. The example that he uses to showcase his point—the job application process of a racialized prospective legal academic—is also instructive: it shows the gatekeeping function of the legal academy for the propagation of ideas and the withholding of status linked to privilege. As he walks through five narratives explicating a single event from a

⁸⁴³ Delgado, "Storytelling," *supra* note 637.

⁸⁴⁴ Lucie E White, "Subordination, Rhetorical Survival Skills, and Sunday Shoes: Notes on the Hearing of Mrs. G." (1990) 38 *Buffalo Law Review* 1 at 5 [White, "Sunday Shoes"].

⁸⁴⁵ Noble, *supra* note 141.

⁸⁴⁶ Eubanks, *supra* note 473.

⁸⁴⁷ Benjamin, *supra* note 268.

⁸⁴⁸ Delgado, "Storytelling," *supra* note 637.

⁸⁴⁹ Other works by Delgado use conversations between composite characters to typify the concerns and lived experiences of racialized individuals interacting with the law and legal academy. See, e.g., Richard Delgado, "Rodrigo's Ninth Chronicle: Race, Instrumentalism, and the Rule of Law" (1994) 143 *U Pa L Rev* 379.

⁸⁵⁰ Delgado, "Storytelling," *supra* note 637 at 2413.

⁸⁵¹ *Ibid.*

variety of perspectives, he easily demonstrates the significance of incorporating competing narratives to generate a composite image. In so doing, he provides a compelling case for his stated premise: social reality is mostly constructed.⁸⁵² Almost by stealth, he demonstrates the importance of the majority listening more closely to the narratives of outgroups, in order to form a better, more cohesive social fabric. Stories, then, provide a wellspring of perspective—an easy mechanism for creating cohesion between majority and minority groups—and maybe even, one day, allowing for a wholesale disruption of such distinctions.

White addresses intersectional issues of race, class, and gender in the legal system through the lens of language. In her piece, “Subordination, Rhetorical Survival Skills, and Sunday Shoes: Notes on the Hearing of Mrs. G,” she provides a compelling argument that claimants who speak lower-value language face structural discrimination. The piece is grounded in linguistic theory: illuminating how marginalized groups inadvertently devalue their own speech. White employs the story of one of her legal aid clients, Mrs. G., to exemplify how a young, poor, black woman is specifically disadvantaged. Mrs. G. faced an inquiry for overpayment of state benefits and possible welfare fraud; among other things, at issue was whether her purchase of new shoes for her daughters could be characterized as “life necessities.”⁸⁵³ After initially agreeing to White’s expertise on legal strategy, Mrs. G. disappointed in the hearing: she first refused to speak, and then abandoned her carefully curated script about life necessities in favour of a defiant statement that she had bought her daughters “Sunday shoes.”⁸⁵⁴ Three key facets of the incivility afforded her by the system are named and expounded upon: intimidation, humiliation, and objectification.⁸⁵⁵ By naming and describing this “terrain” thoroughly, White does the important work of breaking through smokescreens of race, class, and gender to specifically elucidate the mechanisms of Mrs. G.’s oppression. While clearly drawing from law and society, critical race, and feminist methodologies as well, I selected this piece as an exemplar of narratives of client’s experiences within the legal system itself and due to its power as a vehicle for storytelling within the law.

The text describes two potential storytelling strategies: relational and rule-oriented. The socially powerless tend to adopt relational methods of conveying their truth, which is problematic because the system prefers “rule-breach-injury” frameworks of conceptualizing legal claims.⁸⁵⁶ Specifically, White notes that jurors find speakers who use language typical of socially powerless groups to be less credible, competent, or trustworthy than those who use “typically male” speech patterns.⁸⁵⁷ Methodologically, it is here that the narratives of people moving through the legal system intersect with sociological observations about how doctrinal structures might be used to oppress those people. Mrs. G.’s participation in the courtroom is a particularly bold example of what Jane Baron describes as narrative law and literature: an evidentiary account of how the law actually functions in real-world settings.⁸⁵⁸ By putting Mrs. G.’s narrative front and centre, White highlights the ways that the law has traditionally undervalued female voices, young voices, black voices, low socioeconomic voices. Even in this critical framework, however, the recognition of certain language

⁸⁵² Delgado, “Storytelling”, *supra* note 637 at 2416.

⁸⁵³ White, “Sunday Shoes”, *supra* note 844 at 30.

⁸⁵⁴ *Ibid* at 31.

⁸⁵⁵ *Ibid* at 33-39.

⁸⁵⁶ *Ibid* at 17.

⁸⁵⁷ *Ibid*.

⁸⁵⁸ Jane B Baron, “Law, Literature, and the Problems of Interdisciplinarity” (1999) 108 Yale LJ 1059 at 1071.

as being less powerful might encourage adoption of the voices typically favoured by the system itself: male voices, older voices, white voices, wealthy voices. For White, her pragmatism as Mrs. G's lawyer wins out: her desire for Mrs. G to achieve a favourable outcome means she will encourage Mrs. G to abandon her sense of self to move more successfully within a system that disenfranchises such selves.

One of the most compelling elements of this piece as narrative scholarship is the dual effect of White describing Mrs. G.'s experience in the legal system in tandem with her own account of being a new legal aid lawyer, back in North Carolina after ten years away, struggling to fully understand and assist the community.⁸⁵⁹ This positioning bolsters her storytelling claim about inequality in the system, as intended in Farber and Sherry's recognition of storytelling as a rejection of formalism in favour of pragmatism.⁸⁶⁰ Ultimately, White celebrates Mrs. G. for finding her voice within a system designed to leave her voiceless. She concludes with bold claims that post-bureaucratic institutions can arise from localized institution-building activities,⁸⁶¹ but her narrative accounts lend anecdotal power and relationality to law's transformative potential.

Both Delgado and White's approaches signal the importance of storytelling for centering lived realities. In the technology law space, intersectional scholars are increasingly writing about coded bias.⁸⁶² This conversation centres so-called edge cases, bringing impacts that are felt by groups that have been historically marginalized or otherwise minoritized.⁸⁶³ As technological design attempts to treat users in a uniform fashion, these scholars highlight the uncomfortable consequences of such treatment. In *Algorithms of Oppression*, Noble presents critical analysis through the lens of structural aspects of discrimination presented by the Google search algorithm and its troubling results for young black women.⁸⁶⁴ Her study signals the pitfalls of minoritized folks traversing a terrain built by those who do not understand their life circumstances or challenges. Similar observations have been recorded by other researchers, including the landmark "Gender Shades" project by Joy Buolamwini and Timnit Gebru, who catalogued the racialized and gendered aspects of facial recognition technology.⁸⁶⁵ Their study showed that facial recognition more reliably recognized lighter faces than darker faces and male faces rather than female faces—with darker skinned women being recognized least often. Such observations emphasize various structural

⁸⁵⁹ White, "Sunday Shoes," *supra* note 844 at 22.

⁸⁶⁰ Daniel Farber & Suzanna Sherry, "Telling Stories Out of School: An Essay on Legal Narratives (1993) 45 Stan L Rev 807 at 820.

⁸⁶¹ White, "Sunday Shoes," *supra* note 844 at 57-58.

⁸⁶² See, e.g., Benjamin, *supra* note 268; see also Joy Buolamwini & Timnit Gebru, "Gender Shades: Intersectional Accuracy Disparities in Commercial Gender Classification" (2018) 81 Proceedings of Machine Learning Research 1-15.

⁸⁶³ I note here the ongoing debate about how to best refer to groups that have experienced a history of minoritization or marginalization. Many now prefer this framing (as opposed to the more segregatory term "minorities") to refer to such groups, since the passive word choice reflects a power dynamic that has been imposed upon the groups in question. Sometimes, the term "historical" is used an additional modifier, to show this treatment has been longstanding, but its usage can also be seen as improperly relegating this treatment to the past, when the issue is ongoing. I use the words minoritized and marginalized to specifically acknowledge the existence of power dynamic, making such descriptions the result of social processes instead of meaningful descriptors of inherent qualities about the group itself. The passive language also emphasizes that minoritized individuals are not necessarily numerical minorities, but instead those that have been put into a subjugated position by the oppressor. See generally Sylk Sotto-Santiago, "Time to Reconsider the Word Minority in Academic Medicine" (2019) 12 J Best Pract Health Prof Divers 72. Of course, others disagree see K Ward Cummings, "'Minoritized': a violent word" (27 March 2019), online: *Baltimore Sun* <www.baltimoresun.com/opinion/op-ed/bs-ed-op-0328-minoritized-word-20190320-story.html> [perma.cc/ZZP8-QJX].

⁸⁶⁴ Noble, *supra* note 141. I take up this example in greater detail in Chapter 8.

⁸⁶⁵ I take up this example in greater detail in Chapter 7.

aspects of technology industries in North America, where women—and especially Black women—remain underrepresented. Design choices that disproportionately impact these populations might therefore remain unnoticed or simply not be revealed in testing.

For self-driving laws, the consequences are significant. The enduring legacy of how minoritized groups have been treated by the legal system means judicial decisions can bear the worst scars of society's failures. Transposing these decisions into digital enforcement risks embedding, or even exacerbating, differential treatment by cloaking it in data to offer a veneer of neutrality. Reliance on scientific "objectivity" or mathematical "neutrality" creates an obscuring influence known as "math-washing," where the use of mathematical methods papers over subjective reality.⁸⁶⁶ As discussed in Chapter 2, Benjamin calls technological design systems that impose racialized consequences on citizens the "New Jim Code," a nod to how datafication and discriminatory design can entrench structural inequalities.⁸⁶⁷ Even when technology is presented as either "magnifying or obliterating" racial divisions, the persistent ideologies themselves do not seem to shape technological design.⁸⁶⁸ Benjamin's work signals the colonial aspects of technological design, which are made clear when they offer a comprehensive theory of everything.⁸⁶⁹ She observes:

These technological advances are sold as morally superior because they purport to rise above human bias, even though they could not exist without data produced through histories and discrimination.⁸⁷⁰

Noble presents a helpful example about the inherent biases that persist in library classification systems, which employ "information shorthand" to categorize data.⁸⁷¹ The translation of such systems from information sciences to digital information technologies yields a particular reality that appears true due to its definitive presence on one's computer screen in black and white. As background machinations that generate reality are increasingly obscured by algorithmic processes, identifying the sources or validity of information becomes correspondingly difficult.⁸⁷² Over time, the twin phenomena of automation bias and automation complacency pervade the human response to such information. There is a sense that something presented in digital format—the news, a Twitter timeline—has been verified.⁸⁷³ The same outcome seems probable for self-driving laws.

Categories require constant negotiation. They are not set, and undue focus on tradition will not generate the unlearning necessary to create knowledge better oriented towards justice.⁸⁷⁴ As MacKinnon signals when she discusses engaging at the epistemological level, work remains at the

⁸⁶⁶ The term "math-washing" was coined by Kickstarter data scientist Fred Benenson. Kate Robertson, Cynthia Khoo, & Yolanda Song, "To Surveil and Protect: A Human Rights Analysis of Algorithmic Policing in Canada" (2020), online (pdf): *Citizen Lab* <citizenlab.ca/wp-content/uploads/2020/09/To-Surveil-and-Predict.pdf> [perma.cc/A7C5-R8WK] at 121.

⁸⁶⁷ Benjamin, *supra* note 268.

⁸⁶⁸ *Ibid* at 43.

⁸⁶⁹ *Ibid* at 176.

⁸⁷⁰ *Ibid* at 10.

⁸⁷¹ Noble, *supra* note 141 at 137.

⁸⁷² This can lead to what Mireille Hildebrandt calls an "unwarranted aura of objectivity" or "magical expectations." Mireille Hildebrandt, "Algorithmic Regulation," *supra* note 32 at 3.

⁸⁷³ Rachel Botsman, *Who Can You Trust? How Technology Brought Us Together and Why It Might Drive Us Apart* (New York, NY: Hachette Book Group, 2017).

⁸⁷⁴ Borrows acknowledges the potentially destructive force of traditions if they become static, frozen in time, overly romanticized, or unduly inflexible. Borrows, *supra* note 722 at 8.

information curation stage to create more just and equal outcomes.⁸⁷⁵ Noble cautions that knowledge management systems tend to reproduce society's prejudices, since people responsible for assembling information hold their own biases.⁸⁷⁶ The ordering of society absent critical perspectives on the collection and curation of data will not generate the future we want to create. Algorithm design choices can have significant impacts on marginalized communities, as designers are often alienated from the lived realities and actual consequences of their programs on the communities in question.⁸⁷⁷ For Alarie and his collaborators, this means that the book should not be found obsolete just yet, in light of the number of stories that remain unwritten, undiscovered, or uncatalogued.

Eubanks' book *Automated Inequality: How High-Tech Tools Profile, Police, and Punish the Poor* discusses the perceived infallibility of many machine-driven methods that are directed towards assisting or streamlining processes.⁸⁷⁸ Eubanks' method went to the heart of communities impacted by automated decision-makers, employing first-person interviews and information collection from assorted examples of social services. She notes the philosophical challenges flowing from conceptualizing human decision-making processes as necessarily opaque—and contrasted against an alleged transparency made available through machines.⁸⁷⁹ Writing in an American context, she emphasizes that the poorhouse preceded the Constitution as an institution by over a century.⁸⁸⁰ It is a fantasy, she cautions, to expect statistical models or predictive algorithms will suddenly replace poverty policies, and the corresponding culture, built over centuries.⁸⁸¹ Technology and oppression have a troubled history of traveling together, ranging from the Nazi's Hollerith punch card systems to better identify Jews, and the passbook system that controlled movements of over 25 million Black South Africans in apartheid South Africa.⁸⁸² Efforts to maximize efficiency have an unfortunate tradition of targeting social outgroups, especially in the absence of adequate human rights protections. Storytelling, she notes, is central to the work of activists to encourage an integration of the experiences of outgroup members in ethical system design.⁸⁸³

Throughout Eubanks' studies, preferential treatment of automated decision-making tools, which are presumed to be infallible, defeat longstanding procedures designed to guarantee procedural fairness.⁸⁸⁴ Her work makes clear that complicated aspects of the human experience might be simply absent from the data set due to nuances in how data is collected. For example, in her studies of social services, she observes gaps in the data for professional middle class families' use of assistive services like nannies, babysitters, rehabilitation centres, and counselling, since wealthier families

⁸⁷⁵ A recent guide on amplifying the voices of feminist scholars of colour in the technology space emphasizes the gatekeeping function of citation. As a way of identifying lineage and claiming importance, we can "get stuck on thinkers who others have already designated as important: (mostly) dead white men." Rigoberto Lara Guzmán & Sareeta Amrute, "How to Cite Like a Badass Tech Feminist Scholar of Color" (22 August 2019), online: *Data & Society* <points.datasociety.net/how-to-cite-like-a-badass-tech-feminist-scholar-of-color-ebc839a3619c> [perma.cc/X5LL-SFG2].

⁸⁷⁶ Noble, *supra* note 141 at 141.

⁸⁷⁷ Robertson et al, *supra* note 866 at 120.

⁸⁷⁸ Eubanks, *supra* note 473 at 167.

⁸⁷⁹ *Ibid* at 168.

⁸⁸⁰ *Ibid* at 178.

⁸⁸¹ *Ibid*.

⁸⁸² *Ibid* at 196.

⁸⁸³ *Ibid* at 206, 212.

⁸⁸⁴ *Ibid* at 179.

tend to have private insurance or the ability to pay fees out of pocket.⁸⁸⁵ Meanwhile, those in poorer socioeconomic circumstances use the free services offered by the state, signalling different kinds of participation in the larger social sphere through their mode of accessing social services. Reporting statistics from these institutions recalls the phenomenon of dark data: this information would not accurately reflect the actual assistance that wealthier families receive in their lives, creating challenges for quantifying need, modeling patterns, and ultimately, predicting the future. Because the next step for predictive models tends to be machine learning, such data gaps will impoverish the possible results if not remedied. Through such outcomes, Eubanks notes, data mining generates “statistical social groupings,” thereby encouraging policymakers to create customized interventions for specific societal segments.⁸⁸⁶ This nod to a personalized version of law may serve to calcify existing social divisions.⁸⁸⁷

Racialized scholars have, as so often is the case, already offered warnings of how technological design choices can enhance and reinforce social divisions. In *Race After Technology*, Benjamin explains how the coding of old biases into new tools has a long history—one that is surprising only if we erroneously equate technological innovation with social progress.⁸⁸⁸ This, she emphasizes, is one of the failures wrought by the ongoing misconception that technology is always one step ahead of society.⁸⁸⁹ Technological exceptionalism prides itself on emphasizing this pacing problem, which rests on technological development outstripping legal regulation or policy control. Applying a racial lens, which Benjamin terms a “lens of enduring invisibility,” the persistence of old biases coded into new technologies presents itself as the next iteration of a long legacy of oppressive tactics by a society invested in maintaining the status quo.⁸⁹⁰ This legacy is remains prevalent in the continuing articulation of law’s normativity.

3. Law as Storytelling

Achieving legal certainty has always been a complicated project. Chapter 4 illustrated some fundamental difficulties even when viewed from traditional perspectives. When that project veers from the well-trodden paths of established scholars, it gets even more difficult. Insights from critical scholars illustrate that the complexities inherent in mapping social terrain are better served by widening the scope of whose stories are told. Through this lens, storytelling might be the only meaningful way of describing reality. Asking for stories demonstrates a need for nuance brought into reality through relational accounts. This sort of relationality offers a potential response as law is seduced by increasingly available persuasive technologies. Even before any delegation to algorithms, the critical theories discussed in this chapter show how state law has historically been used as an instrument of oppression. Feminist scholars have catalogued how lived experiences outside the male mainstream have no purchase on articulating law’s norms.⁸⁹¹ Racialized writers explicitly call for counternarratives as a mechanism for ensuring their perspectives are contained within the discussion of how law operates.⁸⁹² Indigenous people subjected to colonial laws have

⁸⁸⁵ Eubanks, *supra* note 473 at 168.

⁸⁸⁶ *Ibid* at 196.

⁸⁸⁷ *Ibid*.

⁸⁸⁸ Benjamin, *supra* note 268 at 108.

⁸⁸⁹ *Ibid* at 109.

⁸⁹⁰ *Ibid*.

⁸⁹¹ MacKinnon, *supra* note 630.

⁸⁹² Benjamin, *supra* note 630 at 188.

little to no relationship with the laws imposed upon them by colonizers, which remain out of step with historic and living Indigenous laws, practices, and methods of self-governance.

By putting a law gloss on the status quo, formal legal approaches have long disenfranchised those who existed outside law's central normativity. Taken this way, law's written words through statute or regulation might have been the algorithms of another generation. Still, the constant renegotiation of categories and cataloguing of additional stories offer hope for the future trajectories of law's *nomos*. Cover offers a cohesive vision of the law articulated through language and anchored in a world of social forces and constraints. The tension between the world that is and the world that might be needs to be persistently engaged by human endeavour in order for law to retain its governing power or its claim to obedience. But delegating legal decision-making to AI would unduly resolve the tension between the law as it has been stated and the law as it ought to be. In my view, this tension need not be resolved. This persistent tension gives law meaning, as those governed by it continue to live their lives in its shadow, pressing against the limits of the possible and, perhaps one day, expanding it. If law were regulated by algorithm, law would become alienated from the human collective whose movements give its existence meaning.

Failure to grapple with law's promises and precepts at the methodological level perpetuates injustice against those who might not feel their realities appropriately represented within the law or its dataset. Cover suggests that law can function only while these two constructs are sufficiently close together to be consistently engaged by human behaviour. He writes:

If law reflects a tension between what is and what might be, law can be maintained only as long as the two are close enough to reveal a line of human endeavor that brings them into temporary or partial reconciliation. All utopian or eschatological movements that do not withdraw to insularity risk the failure of the conversion of vision into reality and, thus, the breaking of the tension. At that point, they may be movements, but they are no longer movements of the law.⁸⁹³

This nuanced way of articulating human endeavour shows the power of law and narrative as a methodology. To me, this approach speaks to the prospective function of the Rule of Law: that lawyers come to court to tell stories about what the law might become, if only the judges choose to implement their interpretation of the statute, if only the judges choose to continue the line of reasoning from the little-known dissent cited in their factum. Viewing the law in isolation as the product of legislators or even the product of judges is insufficient: that narrative arc, that line of human endeavour, must be preserved.

In machine learning, a concept called “coherent extrapolated volition” describes a goal for ML systems to go beyond mere datapoints to find the best possible version of the future.⁸⁹⁴ This idea speaks to the better angels of our nature, our future potential as humans, of who we wished we were, and stories we have not yet told.

⁸⁹³ Cover, “Nomos,” *supra* note 9 at 39.

⁸⁹⁴ Christian, *supra* note 40 at 247.

Law is not fruit: it is not something waiting to be plucked from branches, nor can it be “preserved.” All law, by its nature, is actually made and remade through people seriously applying themselves to deeply engage with it and struggling to make it their own. Any judge, lawyer, legal academic, or law student tasked with understanding and applying state laws knows only too well that “the hard work of...law is never done.” All law moves.⁸⁹⁵

Assuming we have all the answers fails to engage the normative exercise that law requires, or the degree to which the Rule of Law has additional, prospective meaning. Common lawyers arguing on principle, invoking persuasive authorities, might ask the court to consider not only what the law is, but what it might become. This is often characterized through the rhetorical tools of a litigator—particularly, an appellate litigator—reasoning by analogy, reasoning from first principles, appealing to moral truths, envisioning a better future for everyone: what we could become, if we are lucky. The Rule of Law, in this sense, is not merely a tool designed to protect us from arbitrary exercises of power, but is also a mechanism for telling the stories that will become the vehicle to propel us to the future we want to create.

* * *

Taken together, the analyses presented in Chapters 4 and 5 reveal what’s at stake for reordering the legal universe by implementing self-driving laws. Drawing from a long tradition of legal philosophers and analytical jurisprudence, the contestability of key features of law’s empire is already brought into focus. The normative inquiries raised by self-driving laws are not so different from the vexing questions that have long animated legal-philosophical debates, like whether law can offer singular right answers and what to do in situations where precedent offers no clear legal solution. The normative consequences of how these debates are structured and answered present significant social consequences for law’s operation. Dworkin’s focus on law as integrity and the nature of the interpretative task transcends law’s formalism, leaving space for pluralism and an acknowledgement that different conceptual ideas can concurrently fall within law’s broad umbrella. Yet, Dworkin’s invocation of Hercules can also spur considerations of supercomputers or legal singularity, especially for proponents of legal regulation by machine. Considering Hercules’ efforts instead through the metaphor of the chain novel connects the discussion of law’s normativity to its narratives, or more specifically, the collective storytelling that is performed by law.

Shifting the focus to storytelling underscores the need for nuance in law’s attempts to operationalize data-driven solutions. AI and ML’s persuasive and seductive abilities need some sort of response, lest they allow a technological somnambulism to sweep through law’s empire, perhaps even delivering law’s administration to corporate power. Thinking through storytelling and relational accounts, as proposed by critical theorists, offers the much-needed nuance for reconceptualizing the dominant story that data tells. Critical theory shows how law, even before any delegation to algorithmic forms of decision-making, has often become an instrument of oppression by failing to consider context and circumstances. While data science might appear to offer collective representations of the human experience, using numerical data to approximate

⁸⁹⁵ Hadley Friedland & Val Napoleon, “Gathering the Threads: Developing a Methodology for Researching and Rebuilding Indigenous Legal Traditions” (2015-2016) 1:1 Lakehead Law Journal 16.

actual collaboration makes it difficult to maintain a sense of context and circumstances. This decontextualization is especially evident when ML algorithms define their own optimal paths, without humans overseeing and making important choices like which data to amplify and which data to ignore. When such data-driven approaches are applied to law, allowing numbers to dictate the proper courses of action might compromise the narrative imagination, segregating law from the stories we tell about our society.

While Dworkin's storytelling was deliberately acontextual, assumed to have absorbed any necessary context from those cases that came before, I advocate for an approach that would explicitly ensure contextual grounding is given careful attention. Through this lens, hard cases and storytelling are two sides of the same coin: state law is silent on the appropriate course of action for some scenarios, but stories can better equip decision-makers to respond to that silence. Nedelsky's relationality helps connect these ideas.

Coupling Part II's analysis with Part I's lessons from cyberlaw and STS, as well as lessons learned from existing implementations of self-driving laws in the administrative state, law is often represented as having a linear relationship to technology, where law forever lags behind the newest innovation in the marketplace. Taking the relational, contextual view, technology might also be able to follow the law, where it would not just rule by technological fiat, but help to regulate the legal futures. If the Rule of Law works as intended to constrain instances of power, it can assist with this project. This will necessarily entail taking a broad view of the Rule of Law that considers other incarnations of power beyond state power. As explained by both Gowder and Lord Bingham, a full accounting of the Rule of Law requires recognition of power as exercised by both public and private actors. And, as West emphasizes, part of law's essential purpose is to do something about exercises of power that might adversely impact citizens.

Further, the Rule of Law's prospective function is a critical aspect of this conceptualization, encouraging a vision where law's narrative arc connects past precedent to future outcomes. Following Cover's advice, the arc should be persistently engaged by human endeavour, never straying too far from social facts or human realities. Wholesale delegation of law's empire to AI/ML methods, therefore, threatens this necessary tension. Looking at the Rule of Law's prospective function, we can identify a critical role for storytelling, to establish context and deal with the ongoing social changes that can transform the nature of law's force on communities. In a technological context, such transformations can be particularly swift or sweeping, sometimes threatening to delegitimize longstanding principles of law.

In Part III, I review the lessons of Part I, and apply the theoretical approaches of Part II, through application of a three-pronged taxonomy devised to interrogate Rule of Law problems. First, I ask questions about the technological process deployed, querying whether the desire to implement an automated process creates simplistic answers to hard cases. We should be alert to technological devices that flatten the texture of reality, produce inaccurate results, or sublimate the analytical complexity that is typically at the heart of legal analysis. I ask: *what process dominates?* Second, I interrogate how contested concepts are presented through the technology's affordances, analyzing the values that are emphasized by the technological process in question. Depending on the specific technology deployed, different values embedded in the contested concept of the Rule of Law might dominate the analysis. Language is also a crucial consideration here, investigating how metaphors

are engaged and whether words themselves are vulnerable to the applied automated process. I ask: *which values dominate?* Finally, I combine insights from the first two criteria to assess the narrative arc of the story that is propagated by the technology in question. I elaborate on three main considerations: whose stories are told; what kind of language is coded; and what kind of results are entrenched in the law because of the technology. I ask: *whose story dominates?* This three-part analysis emphasizes the elusiveness of the formalist desire to render the law certain in the name of the Rule of Law.

I apply this taxonomy to three different scenarios: sentencing software (Chapter 6); facial recognition technology (Chapter 7); and natural language processing (Chapter 8). As the analysis will demonstrate, each of these three technologies involves situations of authority over individuals. In some situations, it is an obvious creature of Rule of Law, as public law authority is imposed through state actors. In other situations, different types of private authority bring private law dimensions into the conversation. As I will argue, the Rule of Law is still implicated in such scenarios, as it remains relevant to power in all its forms.

Part III: Applying the Law

*If you aren't scrupulous in seeking alternative explanations
for your evidence, the evidence will just confirm
what you already believe.*
~ John Horgan⁸⁹⁶

Once, when I was an eager first-year law student, our criminal law class had a substitute professor. Criminal law was a small group seminar—only 15 students—and it was the place where each of us felt most free to ask the questions we might feel embarrassed asking in another, larger class. Over the course of the year, we had become accustomed to our regular professor's fairly liberal, and possibly idealistic, perspectives on the justice system. We spoke up often, sharing personal perspectives and well-intentioned theories on the advancement of justice. On the day in question, our substitute professor was carrying on with the syllabus as scheduled, which listed the day's topic as: "The Adversarial System." You could tell, however, that he was unenthusiastic about the framing. Begrudgingly, he explained the different approaches of the prosecution and defence with an air of disdain in his voice, calling into question the potential utility of a system where two competing narratives were vigorously proposed for judicial determination. He asked what we thought of such a system: did it fulfill its stated goal of pursuing the truth?

No one jumped to answer his question. Feeling the weight of the silence hanging in the air, I raised my hand, still an idealistic first-year law student, still eager to impress. "Doesn't the vigorous advancement of two different stories work towards achieving truth?" I asked. "Let truth and falsehood grapple: whoever knew truth to be put to the worse in a free and open encounter?"

The substitute professor didn't even try to hide his amusement. He laughed openly at me: "Did you seriously just quote Voltaire in my class?" Shaking his head in disbelief, he went on to take the idea of "truth" to task, cynically orating about how everyone has their own version of the truth, how these versions might never meet, how anyone who has ever been in a romantic relationship knows: you might have your story, which you are convinced is correct; your partner might have their own story, which they are convinced is correct; and there is no nebulous third "true" version that can be meaningfully determined by an impartial third party. No, the substitute professor insisted: truth was a scam. Truth was socially constructed, steeped in bias, and all about which story sounded best—or fit best with preconceived stereotypes. No one was arriving at any legitimate ground truths through the court's adversarial process.

Smarting from his bemused dismissal, but still an idealistic first-year law student, I dismissed the substitute professor's cynical oration. Surely, he was just hurting from the breakdown of some recent romantic relationship. Surely, the judicial system could offer its litigants truth, couldn't it?

⁸⁹⁶ John Horgan, "Bayes's Theorem: What's the Big Deal?" (4 January 2016), online: *Scientific American* <blogs.scientificamerican.com/cross-check/bayes-s-theorem-what-s-the-big-deal/#:~:text=Embedded%20in%20Bayes%20theorem%20is,turn%20out%20to%20be%20erroneous> [perma.cc/83NL-VPMB].

Years later, I now think he might have had a point. And that point generates several questions.

This chapter addresses sentencing software. Is there a “right” sentence? Considerable academic attention has been paid to this topic, allowing for deep analysis of conflicting accounts. Since judicial sentencing is highly discretionary,⁸⁹⁷ and social sciences have often charged sentencing procedures as being fraught with bias,⁸⁹⁸ this topic offers a meaningful area to dissect the equality dimensions engaged through technological solutionist approaches to law. If sentencing is already biased, it might seem especially logical to assign a data scientific approach to ameliorate the problem. Yet, if existing bias isn’t carefully calibrated and accounted for within the dataset, self-driving laws might merely entrench existing issues. In this chapter, I use the much-maligned sentencing software program COMPAS, commonly used in the United States, to articulate the Rule of Law problems that arise in the proposed datafication of the criminal law. Through these examples, the constitutional dimension of self-driving laws is brought into focus.

The idea of automating legal reasoning tends to focus on the search for certainty, as the idea of legal certainty suggests aspects of the law that are highly text-based or rule-based. Those areas can appear to be primed for automation, seeming to connote the possibility of a right answer, which can be arrived at by properly parsing a sufficiently detailed dataset. By contrast, sentencing is highly discretionary, with judges being afforded significant space to make decisions within their scope of authority. While this discretionary framework might seem to be at odds with the certainty project proffered by legal automation, it presents an especially interesting example. The prevalence of sentencing software demonstrates the deep preoccupation with the search for a “right answer,” even in the context of something that intentionally incorporates discretion. Even this area of law designed to have discretion baked in attracts best efforts at achieving an answer that can rely on data to demonstrate its correctness.

Canada’s *Criminal Code* prescribes a wide range of potential sentencing measures, including alternatives to carceral sentences in appropriate circumstances.⁸⁹⁹ Specific statutory directives describe principles to guide sentence lengths in light of “relevant aggravating or mitigating circumstances relating to the offence or the offender.”⁹⁰⁰ While sentences may be more harsh if the offence was a hate crime,⁹⁰¹ abuse of trust,⁹⁰² or terrorism,⁹⁰³ they also may be more lenient if the offender accepts responsibility.⁹⁰⁴ Other provisions create unique considerations for sentences in response to crimes against children,⁹⁰⁵ peace officers or other representatives of the justice

⁸⁹⁷ *Criminal Code*, RSC 1985, c C-46 [*Criminal Code*], ss 716-718.2.

⁸⁹⁸ For thorough consideration of this topic, refer to Chapter 3.

⁸⁹⁹ *Criminal Code*, *supra* note 897, s 717(1).

⁹⁰⁰ *Ibid*, s 718.2(a).

⁹⁰¹ *Ibid*, s 718.2(a)(i).

⁹⁰² *Ibid*, s 718.2(a)(iii).

⁹⁰³ *Ibid*, s 718.2(a)(v).

⁹⁰⁴ *Ibid*, s 717(1)(e).

⁹⁰⁵ *Ibid*, s 718.01.

system,⁹⁰⁶ certain animals,⁹⁰⁷ or vulnerable persons.⁹⁰⁸ Despite some rule-bound aspects, sentencing is the ultimate exercise of judicial discretion.⁹⁰⁹ Hildebrandt offers a useful perspective:

The rule-bound nature of discretion makes possible a discussion about the interpretation and employment of discretionary competences; it allows a learning curve by requiring those who intervene to give reasons for their actions if called to account. Those reasons are—in part—the norms that regulate their behaviour as public officials, but in the end, those reasons also include the situated interpretation of those norms. In that sense, discretion is not close to but the opposite of arbitrary rule.⁹¹⁰

As discussed in Chapter 3, strict sentences like mandatory minimums or automatic application of the death penalty, attract criticism for being insufficiently discretionary: they fail to acknowledge an offender's unique circumstances. Yet even though the sentencing provisions of the *Criminal Code* clearly prioritize discretion, they come up against the Rule of Law tendency to consistently apply strong rules across all scenarios. Such formalist approaches to law seek steadfast rules to ensure non-arbitrariness. From this view, anytime that judges exercise discretion, especially for incarceration—which interferes with an individual's liberty—it raises cause for concern. Into this space come AI/ML-driven solutions, attempting to calculate the ideal sentence quantum and allay concerns about inappropriate exercises of discretion.

1. Risk Assessment Via COMPAS

A risk assessment instrument (RAI) is a data-based method for calculating the statistical likelihood of a given risk. Such systems are proposed as opportunities to create transparent and fair methodologies for situations involving human judgement. Disciplines including psychology, criminal justice, and more have advocated for the adoption of RAIs, developing processes designed to improve upon “unstructured, human decision-making.”⁹¹¹ RAIs vary in their methodology and scope; various operative factors can include socio-demographic characteristics, including age, gender, criminal record, education status, employment history, and others.⁹¹² While not all RAIs are powered by algorithmic processes, those that are have attracted significant criticism through the lens of algorithmic fairness.⁹¹³ These have become particularly well-known in the criminal justice context, but might also be deployed in various administrative scenarios, such as car insurance⁹¹⁴ or child welfare agencies.⁹¹⁵

⁹⁰⁶ *Criminal Code*, *supra* note 897, s 718.02.

⁹⁰⁷ *Ibid*, s 718.03.

⁹⁰⁸ *Ibid*, s 718.04.

⁹⁰⁹ *Ibid*, s 718.3. See in particular s 718.3(1), which provides that any enactment allowing degrees of punishment is subject to the discretion of the convicting court, and s 718.3(2), which stipulates no *Criminal Code* provision is to be understood as a minimum punishment unless expressly legislated to be a minimum punishment.

⁹¹⁰ Hildebrandt, “Algorithmic Regulation,” *supra* note 32 at 6.

⁹¹¹ Michelle Bao, Angela Zhou, Samantha Zottola, et al, “It’s COMPASlicated: The Messy Relationship between RAI Datasets and Algorithmic Fairness Benchmarks” (2022) 35th Conference on Neural Information Processing Systems (NeurIPS 2021) Track on Datasets and Benchmarks at 2.

⁹¹² *Ibid*.

⁹¹³ *Ibid*.

⁹¹⁴ O’Neil, *supra* note 372 at 165.

⁹¹⁵ Eubanks, *supra* note 473 at 167.

In the criminal justice context, the defined risk is that the person being assessed will commit a crime sometime in the future. While attempting to predict future crime has always been a preoccupation of the criminal justice system, implementing RAIs has allowed for a shift from “subjective to actuarial assessment.”⁹¹⁶ Instead of judges using their subjective expertise coupled with the specific circumstances of a given accused person’s case, RAIs propose to offer objective analysis backed by statistical methodology. This process can then be applied uniformly to all accused people in the same circumstances, purporting to offer the consistency required by the Rule of Law.

One well-known RAI is the Correctional Offender Management Profiling for Alternative Sanctions (COMPAS), which has garnered significant media attention over the past few years. COMPAS was presented by its creators, Northpointe (later rebranded as Equivant), as a method for using real-world data to predict the likelihood of recidivism in the criminal justice system. More specifically, COMPAS is intended to be used for pre-trial bail determinations, marketed in the United States as a way to streamline the bail process and offer relief to an overburdened judicial system. COMPAS was first developed in 1998; in the years since its inception, it has evolved in response to changing norms around criminal justice and correctional practices.⁹¹⁷ Some commentators welcomed the use of pre-trial RAIs as an efficiency method, especially in light of the state’s obligation to prohibit excessive bail. In a joint statement supporting pre-trial RAIs, the numerous signatories emphasized legal prohibitions against excessive bail, as well as the due process and equal protection clauses of the United States Constitution, as being compelling rationale for their broad implementation.⁹¹⁸

a. COMPAS Design Documentation

The proposed benefits rely on social science research demonstrating that RAIs offer superior risk assessments to human judgement.⁹¹⁹ Northpointe’s documentation highlights the importance of using tools like COMPAS in today’s “overloaded and crowded criminal justice system” and focuses on attributes like “brevity, efficiency, ease of administration, and clear organization of key risk/needs data.”⁹²⁰ Other early reports on its performance boasted an accuracy of 0.73.⁹²¹

⁹¹⁶ Sandra G Mayson, “Bias In, Bias Out” (2021)128 Yale Law Journal 2218 at 2221.

⁹¹⁷ Northpointe, “Practitioner’s Guide to COMPAS Core” (19 March 2015), online (pdf): <<https://assets.documentcloud.org/documents/2840784/Practitioner-s-Guide-to-COMPAS-Core.pdf>> [perma.cc/AB6P-AACT] at 1 [Northpointe, “Practitioner’s Guide”]. This was the operative document when the ProPublica study was being conducted. The most up-to-date version of the Practitioner’s Guide available online is dated April 4, 2019, and contains almost identical content, although the designated author has changed to Equivant in light of the company’s rebranding. See Equivant, “Practitioner’s Guide to COMPAS Core” (April 4, 2019), online: *Equivant* <www.equivant.com/wp-content/uploads/Practitioners-Guide-to-COMPAS-Core-040419.pdf> [perma.cc/4XL3-FFKY] [Equivant, “Practitioner’s Guide”].

⁹¹⁸ Gideon’s Promise, National Legal Aid & Defender Association, et al, “Joint Statement in Support of the Use of Pretrial Risk Assessment Instruments” (10 May 2017), online (pdf): <www.publicdefenders.us/files/Defenders%20Statement%20on%20Pretrial%20RAI%20May%202017.pdf> [perma.cc/XCF9-GSV7].

⁹¹⁹ Bao et al, *supra* note 911 at 1.

⁹²⁰ Northpointe, Practitioner’s Guide, *supra* note 917 at 2.

⁹²¹ This accuracy calculation was derived from a 2009 outcomes study using COMPAS’ Risk of Violence scale on a New York parole population. See Northpointe Inc, “COMPAS Risk & Need Assessment System: Selected Questions Posed by Inquiring Agencies” (2012) [“COMPAS FAQ”] at 3.

COMPAS assessments are calculated from a series of 137 factors about the accused person, used to produce an assessment of how likely the accused is to re-offend.⁹²² The supporting data can be collected in several different ways: the accused person can fill out a questionnaire, an interviewer can pose a series of specific, scripted questions, or an interviewer may use motivational interviewing techniques to engage the accused person in a “guided discussion.”⁹²³ The method of data collection can be selected by the agency deploying COMPAS as an RAI, depending on their resources, staff skills, and available time.⁹²⁴

Questions of deference to RAIs are further complicated by the phenomena of automation bias and automation complacency, defined in Chapter 1. Driven by narratives of technological solutionism, the assumption that a machine is better able to achieve the appropriate result than a flawed human is at the heart of this process. As Chapter 1 explored in detail, the “technological gloss” created by introducing technological systems tends to supplant human expertise, as humans tend to believe that the technological system is necessarily an improvement over their own reasoning. This belief, of course, can sometimes be true. However, the assumption that a technological substitute will necessarily be superior to human judgment is fraught with challenges. One particularly poignant example is ProPublica’s follow-on study of COMPAS’ recidivism rates.

b. COMPAS Critiques

In a landmark study in 2016, the independent, not-for-profit news organization ProPublica collected over 7000 risk-of-recidivism assessment scores for people arrested in Broward County, Florida between 2013 and 2014.⁹²⁵ Conducting a thorough investigation, ProPublica followed up with all of the people assessed by the RAI to see how accurate COMPAS had been—compared against reality. They found that higher scores in the COMPAS tool did not necessarily correlate to higher rates of recidivism. Only 20 per cent of people calculated as being likely to commit a violent crime in the next two years actually did so. And, only 61 per cent of people calculated as likely to reoffend actually did so. ProPublica’s analysis revealed significant equality concerns: the investigation showed how the COMPAS assessment tool was twice as likely to falsely flag black defendants as compared with white defendants. COMPAS was worse at both assessing whether high-risk black defendants would reoffend and assessing whether low-risk white defendants would reoffend.⁹²⁶ The headline, which made considerable waves online, and in criminal justice and algorithmic fairness communities, was striking: “Machine Bias: There’s software used across the country to predict future criminals. And it’s biased against blacks.”⁹²⁷

Without further context, this incendiary headline immediately brings several different definitions of “bias” to mind. At the far end of the spectrum, COMPAS might be populated by biased

⁹²² Equivant, “Practitioner’s Guide,” *supra* note 917 at 1-2.

⁹²³ The COMPAS FAQ proposes that allowing an interviewer to use the guided discussion format will “simultaneously gather the assessment data and enhance report and buy-in for the intervention process.” COMPAS FAQ, *supra* note 921 at 1.

⁹²⁴ *Ibid* at 1.

⁹²⁵ Julia Angwin et al, “Machine Bias” (23 May 2016), online: *ProPublica* <www.propublica.org/article/machine-bias-risk-assessments-in-criminal-sentencing> [perma.cc/8AQS-EHHQ] [Angwin et al, “Machine Bias”]. For a breakdown of the study’s methodology, see Jeff Larson et al, “How We Analyzed the COMPAS Recidivism Algorithm” (23 May 2016), online: *ProPublica* <www.propublica.org/article/how-we-analyzed-the-compas-recidivism-algorithm> [perma.cc/XUT6-UEER] [Larson et al, “How We Analyzed”].

⁹²⁶ Larson et al, “How We Analyzed”, *supra* note 925.

⁹²⁷ Angwin et al, “Machine Bias”, *supra* note 925.

developers who have intentionally designed software to discriminate against black individuals. At the near end, COMPAS might simply be unthinkingly replicating the pre-existing bias in society. Or, the system might be making some small mathematical errors that create the impression of racism, even if untrue. Regardless, the very nature of what a COMPAS-styled RAI tries to achieve eliminates an individual's story, or the story of the group they belong to, from the sentencing exercise. Rather than engaging with the necessary context, the RAI gives the judge freedom to avoid learning too much about individual or group circumstances, focusing on probabilities rather than context. While mathematics might sometimes be presented as objective, people nonetheless managed to interpret the impugned COMPAS data in very different ways. Without contextual grounding anchoring the data analytics to some normative frame, it becomes more difficult to distinguish between the plausibility of explanations for the mathematical outcomes.

In response to ProPublica's study, Northpointe was indignant: their software was not racist, they insisted. In fact, Northpointe responded, ProPublica's own data, carefully analyzed by someone without an axe to grind, actually demonstrated race neutrality in prediction. The 61 per cent on the any-arrest-risk scale, and 20 per cent on the violent-arrest-risk scale, were uniform for both black defendants and white defendants. This meant that both black defendants and white defendants who received a particular score had the same chance of being arrested. This factor, known as "predictive parity," was equivalent for both black and white defendants.⁹²⁸

ProPublica's claim was more subtle. While the COMPAS algorithm did achieve predictive parity, it revealed its discriminatory impacts along other metrics. Among those labeled high-risk who did not reoffend, black defendants comprised 44.9%, while white defendants comprised 23.5%.⁹²⁹ This meant that the false-positive rate was much higher for black defendants than for white defendants.⁹³⁰ And the converse was also true: among those labeled low-risk who later reoffended, black defendants comprised 28%, while white defendants comprised 47.7%.⁹³¹ This meant that the false-negative rate was much greater for white defendants than for black defendants.⁹³²

Arguably, both Northpointe and ProPublica were correct, but they were emphasizing different aspects of the statistical analysis. For Northpointe, the focus was on how likely the score was to predict the future result. If a black person and a white person both received a particular score, those two people would have the exact same chance of being rearrested.⁹³³ This is predictive parity: focused on the eventual result. For ProPublica, the focus was on people who had not gone on to be rearrested after their interaction with the COMPAS system. If both a black defendant and a white defendant had managed to avoid rearrest in the relevant time-period, it should follow—in ProPublica's view—that both defendants should have been equally likely to have been labeled as low-risk by the algorithm.⁹³⁴ Depending on the type of calculation, these errors were either false-positives or false-negatives. Deborah Hellman explains the distinction succinctly:

⁹²⁸ Mayson, *supra*, note 916 at 2233.

⁹²⁹ Larson et al, "How We Analyzed," *supra* note 925.

⁹³⁰ Mayson, *supra*, note 916 at 2234.

⁹³¹ Larson et al, "How We Analyzed," *supra* note 925.

⁹³² Mayson, *supra*, note 916 at 2234.

⁹³³ Deborah Hellman, "Measuring Algorithmic Fairness" (2020) 106 Virginia L Rev at 816.

⁹³⁴ *Ibid* at 816.

In other words, one measure [Northpointe’s focus] begins with the score and asks about its ability to predict reality. The other measure [ProPublica’s focus] begins with reality and asks about its likelihood of being captured by the score.⁹³⁵

The divergent opinions articulated by Northpointe and ProPublica reveal a deep discrepancy on how fairness is conceptualized, translated into machine processes, and operationalized for use in the criminal justice system. Fundamentally, fairness remains a moral issue: one that is hotly contested from different philosophical perspectives.⁹³⁶ The key problem, Sandra Mayson explains, is a difference in base rate between white defendants and black defendants. While Northpointe might extol its achievement of predictive parity, COMPAS does not achieve statistical parity because the algorithm has a larger impact on the group with a higher base rate.⁹³⁷ Certain trade-offs need to be made in the algorithm’s operation; by prioritizing predictive parity, Northpointe has made it impossible to concurrently achieve parity for the false-positive rates and/or the false-negative rates.⁹³⁸ Even though both analyses are focused on results, they generate very different outcomes. In more familiar legal language, one might characterize Northpointe’s priority on predictive parity as formally equal results, while ProPublica’s observations reflect a call for substantively equal results.

How do racial differentiations in the dataset create such different results? In their paper “It’s COMPASlicated: The Messy Relationship between RAI Datasets and Algorithmic Fairness Benchmarks,” the interdisciplinary author team of computer scientists and statisticians led by Michelle Bao propose an answer: that an undue focus on benchmarking in RAI datasets ignores the contextual grounding necessary to better elucidate their operation.⁹³⁹ Typically, the RAI powering COMPAS was looking for either recidivism committed during the pretrial period or a failure to appear in court as the relevant outcome variables.⁹⁴⁰ Bao et al recognize two types of errors arising from how COMPAS creates its data determinations, both of which create label bias. The first type of error is construct invalidity: per United States case law, an acceptable rationale for pretrial detention is either the prevention of pretrial flight or the prevention of violent crime.⁹⁴¹ Because neither flight nor violent crime can truly be accurately predicted in advance, the RAI produces analogues—failure to appear and re-arrest—to substitute within the construct. Are the analogues of failure to appear and re-arrest sufficiently aligned with flight and violent crime to support this substitution? The answer is generally no; and, even more problematically, is more often no in certain racial communities. Bao et al point out that failure to appear is often actually due to court scheduling, work obligations, or transportation problems—all of which are highly correlated with race and class inequality.⁹⁴² This introduces label bias of the second type—

⁹³⁵ Hellman, *supra* note 933 at 816.

⁹³⁶ *Ibid* at 814.

⁹³⁷ Mayson, *supra* note 916 at 2235-2236.

⁹³⁸ Unless, Mayson notes, prediction is perfection—which of course it never is. *Ibid* at 2238.

⁹³⁹ Similarly, they argue that overemphasis on the criminal justice data neglects the importance of interdisciplinary work that shapes its operation. This sphere requires constant decision-making about what constitutes fairness and whose definition of fairness is being served by the algorithm’s function. Bao *et al*, *supra* note 911 at 1.

⁹⁴⁰ *Ibid* at 3.

⁹⁴¹ The case *Stack v Boyle* is cited for the acceptable reasons for pretrial detention. *Stack v Boyle*, 342 US 1 (1951); Bao *et al*, *supra* note 911 at 3.

⁹⁴² *Ibid*.

measurement error. Conflating the variables of failure to appear and flight introduces measurement error closely correlated with race and class inequality.⁹⁴³

Similarly, in an attempt to measure recidivism, rearrest is substituted for reoffence. While one might assume that these factors are interchangeable, being arrested does not always map neatly onto committing a criminal offence. Especially given systemic biases and differential policing practices for racialized offenders, treating these factors as equivalent introduces measurement error. Bao et al remind readers that rearrest doesn't accurately identify criminal behaviour; instead, arrest rates are more closely connected with how a neighbourhood is policed and whether those policing practices are differentiated based on race.⁹⁴⁴ Mayson offers a useful example from her personal experience, emphasizing how racial bias seeps into the dataset from pre-existing biases in society:

A simple example illustrates. When I worked in New Orleans as a public defender, the significance of arrest there varied by race. If a black man had three arrests in his past, it suggested only that he had been living in New Orleans. Black men were arrested all the time for trivial things. If a white man, however, had three past arrests, it suggested that he was really bad news! White men were hardly ever arrested; three past arrests indicated a highly unusual tendency to attract law enforcement attention.⁹⁴⁵

Another interesting aspect of COMPAS's racially charged determinations was that race was not one of the 137 criteria COMPAS used to calculate recidivism. Instead, this illustrates the proxy data problem, where other aspects serve as proxy for race. Data drawn from an unequal society, wherein tangled factors of race, class, and socioeconomic status entwine with policing, conviction, and incarceration rates, present inequalities in unexpected ways. Cathy O'Neil illustrates this issue with calculations for car insurance using credit scores and other data proxies.⁹⁴⁶ Virginia Eubanks highlights analogous proxy issues in cases of child welfare, where automated systems for flagging instances of child harm and neglect see race and poverty as proxy variables.⁹⁴⁷

Even in administrative contexts, where individual liberty is not at stake, the transformative impact and inscrutability of decision-making tools could implicate legal norms about due process. This observation ties to Tom Tyler's ideas on social importance of procedural justice, elaborated in Chapter 3: through survey statistics, he confirms that people are preoccupied with consistency in administrative processes.⁹⁴⁸ The ability to attach social meaning to decisions and reasons anchored in consistent procedure; the public availability of those reasons once given; the faithful observances of due process—these features offer litigants the stability and predictability needed to “feel at home within the framework of the law.”⁹⁴⁹ As Aziz Huq argues, RAIs operate as “new mechanisms to allocate coercion within the criminal justice system.”⁹⁵⁰ Such coercion, he cautions, can be

⁹⁴³ Bao et al, *supra* note 911 at 3.

⁹⁴⁴ *Ibid.*

⁹⁴⁵ Mayson, *supra* note 916 at 2218. This same quotation is deployed by Bao et al, *supra* note 911 at 4.

⁹⁴⁶ O'Neil, *supra* note 372 at 165.

⁹⁴⁷ Eubanks, *supra* note 473 at 167.

⁹⁴⁸ Tyler, *supra* note 521 at 118-120.

⁹⁴⁹ Raz, *Authority*, *supra* note 650.

⁹⁵⁰ Aziz Huq, “Racial Equity in Algorithmic Criminal Justice” (2019) 68 *Duke Law Journal* 1043 at 1111.

disproportionately applied to minoritized individuals: in his analysis, black defendants.⁹⁵¹ Recognizing the articulation of state coercion through technological systems evokes the Rule of Law, asking us to interrogate how state power should be appropriately constrained and whether deploying self-driving law unduly amplifies such power. I now connect the COMPAS case study and critiques with the Rule of Law problems articulated in Chapter 4.

2. Rule of Law Problems

Sentencing software like COMPAS creates challenges for the Rule of Law in its most classic sense. As an instance of state power, the Rule of Law is expected to guide judicial discretion to ensure that the sentences defendants receive are transparent, consistent, grounded in procedural fairness, and promote other democratic values. Sentencing is complex, especially if one is not a criminal law expert. Using the taxonomy drawn from the analysis in Parts I and II, I analyse the Rule of Law problems COMPAS creates. First, I ask what process dominates, interrogating whether COMPAS's mathematical methods provide easy answers to hard cases. Second, I ask which values dominate, thinking through how COMPAS might acknowledge contested concepts. Third, I ask whose story dominates, analysing how COMPAS's output contributes to the construction of truth. By focusing on these criteria, the contextual grounding needed to animate sentencing decisions is brought into focus—a contextual grounding that AI/ML might not be able to provide.

a. Easy Answers to Hard Cases: What Process Dominates?

The case of the *State v Loomis* provided an opportunity for the Wisconsin Supreme Court to opine on the use of COMPAS in the sentencing context.⁹⁵² The facts of the case were relatively straightforward: in February 2013, Eric Loomis was arrested on five different charges after driving a stolen car that had been used in a drive-by shooting. Loomis denied having been involved with the shooting itself but admitted to operating the vehicle later that day. He eventually pleaded guilty to two of the lesser charges. At trial, he was found guilty of those two charges: operating a motor vehicle without the owner's consent and attempting to flee a traffic officer. The trial court ordered a "Presentence Investigation Report," which included a COMPAS risk assessment. The COMPAS results designated Loomis as high risk in each of the three categories assessed: general recidivism risk, pretrial recidivism risk, and violent recidivism risk.⁹⁵³ Notably, the report also specified that Loomis' COMPAS results were only a risk assessment and should not be used as part of the considerations leading to his sentencing.⁹⁵⁴ Nonetheless, at sentencing, the court explicitly referenced it, noting that Loomis had been "identified, through the COMPAS

⁹⁵¹ Huq, *supra* note 950 at 1120.

⁹⁵² *State v Loomis*, 881 NW2d 749 (Wisconsin 2016).

⁹⁵³ Brief for the United States as Amicus Curiae, *Loomis v Wisconsin* (No 16-6387) at 5 [*Loomis* Amicus Brief].

⁹⁵⁴ *Ibid* at 5.

assessment, as an individual who is at high risk to the community.”⁹⁵⁵ In combination with other factors,⁹⁵⁶ this led to Loomis being sentenced to six years in prison.⁹⁵⁷

Loomis filed a post-conviction motion, requesting a new sentencing hearing.⁹⁵⁸ He argued that using COMPAS to determine his sentence was a violation of his constitutional right to due process. The post-conviction motion was denied.⁹⁵⁹ Loomis then appealed, and the Wisconsin Court of Appeals certified a question to the Wisconsin Supreme Court, asking if using COMPAS risk assessments during a sentencing hearing “violates a defendant’s right to due process, either because the proprietary nature of COMPAS prevents defendants from challenging the COMPAS assessment’s scientific validity, or because COMPAS assessments take gender into account.”⁹⁶⁰

The Wisconsin Supreme Court rejected both of Loomis’ challenges. Even though COMPAS’s algorithm was proprietary and inscrutable, most of the information used was static—solid factual historical information—with only a few dynamic variables drawn from Loomis’ more recent history.⁹⁶¹ Because Loomis had been given the opportunity to verify those dynamic variables (21 questions and answers, out of a 137-step process), the court found that the information was accurate. Gender was not considered an issue since men tend to reoffend at higher rates than women and failing to acknowledge that reality would actually serve to “misclassify both genders.”⁹⁶² Yet, the Wisconsin Supreme Court nonetheless noted that COMPAS use during sentencing did engage constitutional concerns, and its usage had been circumscribed by the accompanying report. Going forward, the court insisted, Presentence Investigation Reports should have a thorough warning detailing their limitations, including: COMPAS’s proprietary nature; the usages of group data; disproportionate impacts on minority offenders; accuracy issues with changing populations (notably, that COMPAS was more accurate for nationwide data than state-specific data); and that COMPAS was not developed for sentencing use, but for making determinations about treatment, supervision, and parole.⁹⁶³

The case was further appealed to the Supreme Court of the United States, but the petition was denied.⁹⁶⁴ Both the Solicitor General of the United States and the State of Wisconsin filed amicus briefs in support of using COMPAS, signalling the state’s interest in ensuring this technological innovation be available to the courts of the nation.⁹⁶⁵

⁹⁵⁵ *Loomis* Amicus Brief, *supra* note 953 at 6.

⁹⁵⁶ On sentencing, the court heard from Loomis, his girlfriend, and his attorney. It reviewed his criminal history, his “sporadic” job history, and the “treatment needs” that had been identified during his previous incarceration—interestingly, these needs also stemmed from evaluation by an “assessment tool” that had flagged Loomis’ “potential dependency on substances.” *Ibid* at 6.

⁹⁵⁷ In the court’s explanation to Loomis, it explained it was: “ruling out probation because of the seriousness of the crime and because your history, your history on supervision, and the risk assessment tools that have been utilized, suggest that you[’re] an extremely high risk to reoffend.” *Ibid* at 7.

⁹⁵⁸ *Ibid* at 7.

⁹⁵⁹ *Ibid*.

⁹⁶⁰ *Ibid* at 8.

⁹⁶¹ *Ibid* at 9.

⁹⁶² *Ibid*.

⁹⁶³ *Ibid* at 10-11.

⁹⁶⁴ The question presented was “whether the sentencing court’s consideration of an actuarial instrument assessing petitioner’s risk of recidivism violated his due process rights, either because petitioner was denied an opportunity to challenge the instrument’s methodology or because the instrument accounts for gender in formulating its risk assessment.” *Ibid* at (I).

⁹⁶⁵ *Loomis* Amicus Brief, *supra* note 953.

Ultimately, the courts' rationale in *Loomis* hinges on ensuring RAIs receive only limited use in a sentencing context. Sentencing courts are authorized for only particular uses of the COMPAS reports, for things like "diverting low-risk prison-bound offenders to a non-prison alternative; assessing whether an offender can be supervised safely and effectively in the community; and imposing terms and conditions of probation, supervision, and responses to violations."⁹⁶⁶ And, even within these circumscribed uses, the "risk scores may not be used as a determinative factor."⁹⁶⁷ These guidelines are hard to reconcile with our knowledge of the pervasive impacts of automation complacency and automation bias on human decision-making. Psychological research shows that it is hard for a judge to go against an algorithmic determination, as human decision-making is often anchored in improper places.⁹⁶⁸ The technological gloss applied to COMPAS's determinations transforms its relevancy: mere inclusion of warning statements will not negate the psychological impacts of RAIs being presented to decision-makers. Limiting the COMPAS reports to specific usages does not accord with the realities of how judges will understand information and make decisions informed by those reports.

The combination of the ProPublica exposé and the *State v Loomis* decision has set the stage for recent debates on the usability of criminal sentencing algorithms, juxtaposing the potential harm of racial bias against the potential benefit of having additional information available to judges in sentencing processes.⁹⁶⁹ Proponents of such "evidence-based sentencing" tend to promote RAIs as opportunities for criminal justice reform—chances to increase pretrial bail rates or reduce overall incarceration.⁹⁷⁰ In spite of this optimism, additional research by statisticians demonstrates that COMPAS's predictive abilities are relatively limited. In a follow-on study to the ProPublica debates, Julia Dressel and Hany Farid investigated whether the COMPAS RAI is demonstrably better at predicting recidivism than untrained humans.⁹⁷¹ Using a random assortment of humans procured via Amazon's Mechanical Turk platform, they asked participants to review a concise description of an individual offender, with seven descriptive factors including sex, age, and previous criminal history (but notably, not their race), and then asked them whether they thought the individual would reoffend within the next two years.⁹⁷² They compared these results against the performance of COMPAS on the same subsets of individuals. Statistically, there was marginal difference between the two processes: the untrained humans recorded an accuracy of 62.8% (with a standard deviation of 4.8%), which was very close to COMPAS's accuracy of 65.2%.⁹⁷³ If the answers were pooled (searching for "wisdom in the crowd"), they were superior to COMPAS' results, yielding an accuracy of 67.0%.⁹⁷⁴ They also measured another variable—accuracy equity—to assess whether black defendants and white defendants received the same treatment in the

⁹⁶⁶ *Loomis* Amicus Brief, *supra* note 953 at 10.

⁹⁶⁷ *Ibid* at 10.

⁹⁶⁸ Angèle Christin, Alex Rosenblat, & danah boyd, "Courts and Predictive Algorithms" (27 October 2015), online (pdf): *Data Civil Rights* <www.datacivilrights.org/pubs/2015-1027/Courts_and_Predictive_Algorithms.pdf> [perma.cc/H7VK-76TL] at 8. Christin et al rely on the research by Amos Tversky and Daniel Kahneman on "anchoring," where people tend to rely on the first piece of information that they are given when making decision—even if that information is weak.

⁹⁶⁹ Ben Green, "'Fair' Risk Assessments: A Precarious Approach for Criminal Justice Reform" (2018) Presented at the 5th Workshop on Fairness, Accountability, and Transparency in Machine Learning (FAT/ML 2018) Stockholm, Sweden at 1.

⁹⁷⁰ *Ibid* at 1.

⁹⁷¹ Julia Dressel & Hany Farid, "The accuracy, fairness, and limits of predicting recidivism" (2018) 4 *Science Advances* eaao5580 at 1.

⁹⁷² *Ibid* at 1.

⁹⁷³ *Ibid* at 2.

⁹⁷⁴ *Ibid*.

model.⁹⁷⁵ These results were almost identical: 0.71 ± 0.03 for the untrained humans, *versus* 0.70 ± 0.04 for COMPAS.⁹⁷⁶ With only seven factors, the untrained humans' results were as accurate as COMPAS' 137 factors for determining recidivism.⁹⁷⁷

To further distil their results, Dressel and Farid performed additional permutations of the dataset, demonstrating equivalent success with only two variables—age and previous convictions—which yielded a statistically similar result to COMPAS' 137-step number crunching.⁹⁷⁸ Despite efforts to create a thorough picture of the accused person, it appeared that the additional detail offered little advantage. Paradoxically, additional information created less accurate risk assessments, demonstrating that simple rule sets were actually better for accurately calculating risk.⁹⁷⁹ Other researchers confirmed consistent results: Cynthia Rudin and colleagues similarly demonstrated that simplistic rules are approximately as accurate as COMPAS' proprietary predictions, describing these as “optimal sparse rule lists.”⁹⁸⁰ In this study, accuracy was defined as predictive accuracy, as researchers compared their optimal rule lists (mean test accuracy of 0.665) with the COMPAS algorithm output scores (mean test accuracy of 0.660).⁹⁸¹

The application of blunt rules that treat all situations and offenders the same have long been the focus of ire in criminal justice reform. Recall the discussion in Chapter 3 regarding the unconstitutionality of capital punishment when applied uniformly instead of to a specific offender's situation.⁹⁸² In the recent SCC decision of *R v Bissonette*, the court struck down the imposition of a 40-year sentence for the perpetrator of a hate crime/terror attack.⁹⁸³ The court observed that “case law on grossly disproportionate punishment has been developed in the context of mandatory sentences imposed without regard for the offender's particular circumstances.”⁹⁸⁴ Noting the SCC's framework in *R v Nur*,⁹⁸⁵ the court emphasized the need to consider a proportionate sentence for the specific offence, and ask whether an imposed mandatory minimum sentence would “require it to impose a sentence that is grossly disproportionate to one that would be just and appropriate for the offender or for another offender in a reasonable hypothetical case.”⁹⁸⁶ If so, the decision would be unconstitutional: infringing section 12 of Canada's *Charter of Rights and Freedoms*.⁹⁸⁷

Acknowledging the specific circumstances of a given offender, which includes taking out the context in which certain groups live their lives,⁹⁸⁸ is impossible through the COMPAS RAI: as the

⁹⁷⁵ Dressel & Farid, *supra* note 971 at 2.

⁹⁷⁶ *Ibid.*

⁹⁷⁷ *Ibid.*

⁹⁷⁸ *Ibid* at 3.

⁹⁷⁹ Ed Yong, “A Popular Algorithm Is No Better at Predicting Crimes Than Random People” (January 17, 2018), online: *The Atlantic* <www.theatlantic.com/technology/archive/2018/01/equivant-compas-algorithm/550646/> [perma.cc/W8GW-TJSE].

⁹⁸⁰ Cynthia Rudin et al, “Learning Certifiably Optimal Rule Lists for Categorical Data” (2018) 18 *Journal of Machine Learning Research* 1.

⁹⁸¹ *Ibid* at 40. The authors note that the COMPAS algorithm generates scores from 1-10, which are categorized as low (1-4), medium (5-7), and high (8-10) risk levels. They interpreted medium or high scores as a positive test for recidivism over a two-year period, while a low score constituted a negative test.

⁹⁸² See Chapter 3, *infra*.

⁹⁸³ *R v Bissonette*, 2022 SCC 23.

⁹⁸⁴ *Ibid* at para 63.

⁹⁸⁵ *R v Nur*, 2015 SCC 15.

⁹⁸⁶ *R v Bissonette*, *supra* note 983 at para 63.

⁹⁸⁷ *Charter*, *supra* note 828, s 12.

⁹⁸⁸ For example, those who work long hours, have volatile work schedules, or significant family responsibilities.

Wisconsin Supreme Court noted, only 21 of the 137 variables were dynamic. And even if the report generated by COMPAS is intended only as a guideline, it is impossible to prove that judges will not defer to the technological gloss the system offers. Especially given that RAIs do not necessarily offer an improved outcome over even non-expert decision-makers, the observable benefit to implementing such systems remains up for debate. Their use flattens the texture of reality, making the hard task of sentencing into something easier than it really should be. Moreover, it engages constitutional concerns, raising the spectre of a section 12 challenge if COMPAS or similar technology were to be deployed in Canada.

b. Acknowledging Contested Concepts: Which Values Dominate?

Translating sentencing to ML-enabled platforms can emphasize the value-laden aspects that are embedded in the process. In the context of COMPAS, this means that the 137-step questionnaire articulates the dominant values at play in the sentencing process. Considering how Dressel and Farid generated commensurately accurate results with only seven, and then only two variables, the values being propagated necessarily promote a certain conception of fairness. But even the nature of fairness is something hotly contested, with different statistical approaches leading to different results. The discrepancies between the views held by Northpointe and ProPublica with respect to the COMPAS dataset show the mathematical machinations at play in defining whether something can be characterized as “fair.” When an RAI presents an outcome as being fair, we must first ask: according to which definition? Hellman notes that fairness can be used in different ways, referring to a variety of normative ideas.⁹⁸⁹ The absence of predictive parity, the issue primarily contested by ProPublica, attacks the degree of belief that we are entitled to have in the algorithm’s results. Absent predictive parity, “we cannot know precisely what to believe about a scored individual.”⁹⁹⁰ Hellman goes on to distinguish between comparative and non-comparative conceptions of fairness, where comparative fairness is concerned with ensuring two individuals receive comparable treatment, while non-comparative fairness is concerned with ensuring one given individual receives the treatment they deserve—irrespective of any external factors.⁹⁹¹ Focusing on algorithmic fairness tends to suggest a comparative method, but even this can be further subdivided. One might ask whether people of different racial categories were treated the same by the algorithm, or alternatively, ask whether both the people scored by the algorithm and the victims experiencing its after-effects would be treated fairly by its operation.⁹⁹² Even these subtle distinctions demonstrate that there are considerable value-judgments imposed in the context of determining a fairness definition—one that might not be immediately transparent to the people using the algorithm. COMPAS creates a cross-pollination between the justice system and the corporation that creates, markets, and sells it to the states administering their judicial programs.

Ben Green argues that ML’s reliance on data and metrics can “distort the values inherent to the task at hand by granting undue weight to quantified considerations at the expense of unquantified ones.”⁹⁹³ His argument recalls Ruha Benjamin’s analysis, discussed in Chapter 5, of the colonial

⁹⁸⁹ Hellman, *supra* note 933 at 834.

⁹⁹⁰ *Ibid.*

⁹⁹¹ *Ibid* at 835.

⁹⁹² *Ibid.*

⁹⁹³ Green, *supra* note 969 at 2.

aspects of technological design that reproduce data stemming from legacies of discrimination.⁹⁹⁴ Through this lens, Benjamin reminds us of the persistence of social bias reflected through technological artifacts, which present the “allure of objectivity without public accountability.”⁹⁹⁵ Without perspective to interrogate whether their own insights are correct, algorithms can “lean in” to whatever biases are operating in the background—and, eventually, “drift further and further away from fairness.”⁹⁹⁶ The definition of fairness presented by COMPAS might not be the one that we want to centre as a value for our judicial system.

Self-driving laws change the dynamics of power in a system that has carefully allocated who should appropriately wield power. Corporations accrue new power by virtue of their technological methods: as emphasized previously, COMPAS keeps secret the machinations performed with the 137-step questionnaire, as a proprietary aspect of its business model.⁹⁹⁷ This inscrutability obscures the data processing performed to generate legal outcomes, which are vulnerable to amplifying bias and discrimination. Simultaneously, the financial remuneration claimed by the company injects new corporate interest into the judicial system, where data scraped from instances implicating individual liberty become part of a company’s business model. From appropriately defining fairness to making sense of corporate contributions, the values implicated by COMPAS call into question its appropriate role in the administration of justice.

c. Constructing Truth: Whose Story Dominates?

Understanding what kind of information algorithms are capable of conveying is a key aspect of making sense of any potential inclusion in the judicial system. The quality, accuracy, and tone of the information matters for understanding the dominant story that will flow from its implementation. Taking an epistemological stance, Hellman distinguishes between algorithms that tell us what to believe and algorithms that tell us what to do.⁹⁹⁸ She offers a powerful example to help elucidate the distinction.⁹⁹⁹ Suppose that a bat entered your home and your infant daughter was briefly exposed to the bat. The infant did not have any visible markings of a bat bite, nor was she visibly distressed or agitated by the presence of the bat. She did not cry or scream while the bat was present. While bats sometimes carry rabies, this is rare and relatively unlikely. In all likelihood, the infant was not bitten by the bat, and the bat did not have rabies. Notwithstanding the balance of evidence suggesting that everything was going to be okay, the family pediatrician still recommends that the infant be treated for rabies. Why? Because the cost of a false negative—that the baby was indeed bitten by the bat, who indeed had rabies—was extremely high: the baby would die. For that reason, for the baby’s own protection, it made the most sense to act as though the bat had rabies and had bitten your infant daughter. But what we ought to do in this situation remains different from what we ought to believe. We should treat the baby for rabies, but we should still believe, on the balance of probabilities, that the baby does not actually have rabies. The treatment

⁹⁹⁴ Benjamin, *supra* note 268 at 10.

⁹⁹⁵ *Ibid* at 53.

⁹⁹⁶ Ellora Israni, “Algorithmic Due Process: Mistaken Accountability and Attribution in *State v. Loomis*” (August 31, 2017), online: *JOLT Digest* <jolt.law.harvard.edu/digest/algorithmic-due-process-mistaken-accountability-and-attribution-in-state-v-loomis-1> [perma.cc/E8XL-2JBN].

⁹⁹⁷ Angwin, “Machine Bias” *supra* note 925.

⁹⁹⁸ Hellman, *supra* note 933 at 828.

⁹⁹⁹ This example is paraphrased from Hellman’s “individual cases” example entitled “Leslie, the Baby, and the Bat.” *Ibid*.

is simply prophylactic, and by all logical estimations, probably unnecessary. We act in a way that seems to propose a certain reality, but we simultaneously believe that reality to be false.

Hellman deploys this example to demonstrate how the statistical thresholds at play in the COMPAS dataset convey different kinds of information. Her example asks us to reflect on the cost of a false positive *versus* the cost of a false negative in prescribing a course of action.¹⁰⁰⁰ When the COMPAS dataset serves up racially differential results in these two arenas, based on different racial base rates in the dataset that are legacies of generations of discrimination, those results have consequences. And, if we lose some predictive accuracy in the calculation of a risk assessment, we also have more to lose: our confidence in our own beliefs. As Hellman puts it:

Faced with the score produced by a test or algorithm, we will not know precisely what to believe, as the significance of the test or score will be lessened. Loss of predictive accuracy compromises knowledge or, to be more precise, we lose confidence in the information provided by the algorithm. The type of error we might make matters when assessing how we ought to act.¹⁰⁰¹

In my view, Hellman's example also assists with understanding whose story dominates. In a world of legal precedent, the proposition that any given case might stand for will be summarized, condensed, and reduced until it neatly comprises the succinct *ratio decidendi* that law students scribble on their exam outlines or lawyers cite in their case briefs. The story that will propagate from Hellman's bat example is to treat your infant daughter for rabies if she is exposed to a bat. The nuance—that she was likely not bitten by the bat, that the bat was likely not infected with rabies—will not necessarily be captured in subsequent retellings, unless appropriately emphasized. Hellman notes that the distinction offered by the bat example is the difference between what we ought to do and what we ought to believe. I argue that the imposition of self-driving laws can conflate these two concepts such that they are no longer epistemologically significant. We will do what the algorithm tells us to do. Before long, we might not think too hard about the subtle distinction between action and belief. In an automated system of self-driving laws, action will precede any reflection on belief, and the beliefs that might be drawn from the given situation will be secondary to the automated process that has already been pressed into action.

As the COMPAS dataset demonstrates, such a conflation between action and belief would be problematic for achieving racial equality, especially in the context of a criminal justice system that has traditionally marginalized certain racial groups. Green notes that the racially differentiated base rate is the result of longstanding racial discrimination, meaning that even accurate recidivism calculations extend those discriminatory legacies “by punishing blacks for having been subjected to such criminogenic circumstances in the first place.”¹⁰⁰² How data is collected, organized, and compiled can yield significant impacts for the types of results that are generated. Bao et al point out that the data used as predictive metrics in pretrial risk assessments are “low dimensional summaries of criminal history, such as past arrests or convictions.”¹⁰⁰³ Green cautions that ML's “inability to incorporate social and historical context into broader perspectives of fairness has the

¹⁰⁰⁰ Hellman, *supra* note 933 at 829.

¹⁰⁰¹ *Ibid* at 830.

¹⁰⁰² Green, *supra* note 969 at 2.

¹⁰⁰³ Bao et al, *supra* note 911 at 4.

potential to hinder social change and entrench historical discrimination.”¹⁰⁰⁴ When approached with equality-seeking dimensions in mind, the need to alleviate longstanding legacies of discrimination before they are replicated by a “neutral” dataset is underscored. Articulating the danger with predictions flowing through the guise of the “New Jim Code,” Benjamin offers caution about technological design: sometimes, self-fulfilling prophecies enact exactly what they predict.¹⁰⁰⁵

The Rule of Law’s prospective function offers a lens through which to consider the future possibilities of what sentencing decisions might look like. The static nature of the preponderance of the COMPAS factors juxtaposes against a prospective, malleable way of looking at law, thereby inviting further inquiry. While having a strict list of criteria to assess sentencing decisions against might seem to offer a formalism that would promote the consistency and non-arbitrariness required by the Rule of Law, automated generation of those criteria from historical data eschews the dynamism necessary to repair historical (and ongoing) injustices. Corporate interests implicated in this further complicate the issue, especially when the choices made are kept as trade secrets, inscrutable from the users. The COMPAS RAI blurs the lines between what we believe and what we should do, entrenching an unduly static set of formal criteria as representative of the sentencing process. If the criteria are established by the company responsible (in this case, Northpointe), there’s no guarantee that they will be responsive to shifting norms, or that they will shift in an equality-promoting fashion. As we recall from Chapter 5, all law moves.¹⁰⁰⁶ And perhaps law isn’t what we want from law unless it is capable of movement. Rather than understanding law as a photograph perfectly capturing a moment in time, we should understand it as a film.¹⁰⁰⁷ To make narrative sense of law in any given time, the cinematographer should be constantly finding a balance between that which stays in the same place and that which moves, thereby reconstituting the scene.

Especially in a legal context, arguments about the future can be unappealing since they feel somewhat political: they, by definition, cannot be grounded in facts since they are prospective. But the tension that Cover identifies as connecting what *is* and what *might be* needs to be maintained for law to retain its legitimacy in the eyes of the governed. For legal technology, the interplay between normativity and technological affordances reminds us that technology can shape human behaviours over time, slowly establishing those behaviours as normative. COMPAS’s contributions to sentencing, while at least somewhat appealing in order to establish consistency, create a rote list of criteria that alienates the sentencing process from human endeavour. To remind ourselves of what might be lost if sentencing were entirely delegated to self-driving laws, we look now to instances of context-sensitive storytelling that already exist in sentencing law, and ongoing battles for further sentencing reform, that speak to maintaining this arc of human endeavour.

3. Storytelling Solutions

Understanding the Rule of Law’s prospective function enables a vision of law where its narrative arc can be understood through the lens of context-sensitive storytelling. By looking for instances

¹⁰⁰⁴ Green, *supra* note 969 at 2.

¹⁰⁰⁵ Benjamin, *supra* note 268 at 83.

¹⁰⁰⁶ Friedland & Napoleon, *supra* note 895 at 16.

¹⁰⁰⁷ I am indebted to Peter Oliver for sharing this example with me.

of storytelling that already exist within the legal system, we are reminded of the non-quantitative aspects of sentencing that would be hard, if not impossible, to simulate through the 137 variables considered in the COMPAS RAI. Efforts to ameliorate Canada's relationship with Indigenous peoples through the criminal justice system, with new provisions amending the sentencing aspects of the *Criminal Code*, provide an example.

Importantly, the existence of these provisions has not cured legacies of discrimination that persist within the criminal justice system. When establishing the importance of maintaining human endeavour within the sentencing process, it is important not to overly romanticize the experience of hearing judicial reasons from a human judge over an algorithm, without clearly identifying other supporting rationale. Such romanticism is a view not universally held, as people in marginalized communities often hear terrible things from judges. I do not aim to glorify judicial reasons for the assumed advantage of hearing a human judge respond to a specific case; to do so would be to glorify an entrenched privileged worldview. Notwithstanding this disclaimer, I argue that insulating sentencing from self-driving laws supports the world of justice we want to create. Exploration of the Gladue factors illustrates why.

In recent years, the Canadian judicial system has attempted to respond to the legacies of discrimination against Indigenous offenders with the inclusion of what later became known as the Gladue factors in the sentencing process. Amendments to the *Criminal Code* in 1996 created Section 718.2(e),¹⁰⁰⁸ which instructs judges to take judicial notice of the systemic circumstances impacting Indigenous offenders, and to consider all possible sanctions, other than imprisonment, reasonable in the circumstances. In the first case interpreting this newly amended provision, *R v Gladue*,¹⁰⁰⁹ the SCC clarified the provision's function and purpose, noting Canada's problems with overincarceration,¹⁰¹⁰ its overrepresentation of Indigenous offenders in penal institutions,¹⁰¹¹ and the systemic and background factors that make traditional incarceration especially problematic for Indigenous offenders.¹⁰¹² The Court also noted that traditional Indigenous conceptions of sentencing are focused on restorative justice, which are significantly removed from the sentencing ideals of "deterrence, separation, and denunciation" that are typically propagated by the colonial judicial system.¹⁰¹³

Adopting a relational account, as presented by Jennifer Nedelsky,¹⁰¹⁴ encapsulates the complex observations expressed by the SCC in *R v Gladue* and promoted in litigation since that time. It assists with understanding an individual in the custody of the criminal justice system as someone subject to a series of interlocking systems, which are intimately connected with legacies of power,

¹⁰⁰⁸ *Criminal Code*, *supra* note 897, s 718.2(e). The full text of the provision reads: (e) all available sanctions, other than imprisonment, that are reasonable in the circumstances and consistent with the harm done to victims or to the community should be considered for all offenders, with particular attention to the circumstances of Aboriginal offenders.

¹⁰⁰⁹ *R v Gladue* [1999], 1 SCR 688.

¹⁰¹⁰ *Ibid* at para 52.

¹⁰¹¹ *Ibid* at para 58. Note that the SCC decision uses the term "Aboriginal Offenders," which was more commonly used at the time, and also connotes the constitutionalization of the term "Aboriginal" in section 35 of the *Charter*.

¹⁰¹² *Ibid* at para 67. The Court writes: "The background factors which figure prominently in the causation of crime by aboriginal offenders are now well known. Years of dislocation and economic development have translated, for many aboriginal peoples, into low incomes, high unemployment, lack of opportunities and options, lack or irrelevance of education, substance abuse, loneliness, and community fragmentation. These and other factors contribute to a higher incidence of crime and incarceration."

¹⁰¹³ *R v Gladue*, *supra* note 1009 at para 70.

¹⁰¹⁴ Nedelsky, *supra* note 816 at 201.

theft, and discrimination. Despite operating within the doctrinal confines of sentencing encoded in Canada's *Criminal Code*, the Gladue factors have attempted to incorporate the relevance of an individual person's circumstances to their sentencing hearing. It offers a way for a person's story to be told in the context of the criminal justice system.

Ongoing litigation continues to shape the contours of what sentencing might look like to best respond to the unique circumstances of Indigenous offenders. At the same time as s 718.2(e) was being amended, further amendments created conditional sentencing provisions that allowed certain sentences to be served within the community. These provisions were complicated in 2012 when the Harper government removed the availability of conditional sentences for certain offences,¹⁰¹⁵ constraining sentencing judges' options for fulfilling the obligation to find other sanctions than imprisonment prescribed by s 718.2(e). The recent case of *R v Cheyenne Sharma*, heard by the SCC in April 2022, considers these provisions. Ms. Sharma is an Indigenous offender who served as a drug mule, importing almost 2 kilograms of cocaine to earn \$20,000. At trial, she was sentenced to only 17 months in prison, owing to the hardships she had faced as an Indigenous person. The sentencing judge noted that she had endured "a constellation of classic *Gladue* factors."¹⁰¹⁶ The government has continued to appeal her sentence, arguing that a conditional sentence for such a crime has been rendered unavailable by Parliament's "policy choice" to limit the availability of conditional sentences,¹⁰¹⁷ and because "the importation of cocaine is a pernicious crime that leads inexorably to extraordinary human suffering."¹⁰¹⁸ Lawyers for Ms. Sharma maintain that she should be given the opportunity to ask for a conditional sentence—emphasizing that "it does no one any good to send someone like Ms. Sharma to jail, away from her young daughter and her community, where a sentencing judge is satisfied that the public interest could be equally served if she served her sentence in the community."¹⁰¹⁹ Lawyers for Ms. Sharma challenge the amendments to the *Criminal Code* limiting the availability of conditional sentences on both section 7 and section 15 grounds.¹⁰²⁰ They argue that Parliament overshot its objective, and that the new legislative provisions are overbroad, because maximum sentence was deployed as a proxy for seriousness.¹⁰²¹

Imagine now that Ms. Sharma's case had been decided by COMPAS, or—like Mr. Loomis—that her sentencing hearing had been accompanied by a Presentencing Investigation Report that showed what COMPAS would do in the case. How would such a procedure accord with the rationale for the *Gladue* factors expressed by the SCC? Given the overrepresentation of Indigenous offenders in Canadian penal institutions, all likelihood suggests that Ms. Sharma would have been designated as someone at high risk to reoffend. In one foreseeable future, AI decisions are preferred

¹⁰¹⁵ These reforms were realized through a series of "tough on crime" amendments enacted through the *Safe Streets and Communities Act*. Conditional sentences were made unavailable for serious offences, which included those under s 742.1(c) and s 742.1(e)(ii) of the *Criminal Code*.

¹⁰¹⁶ *R v Sharma*, 2018 ONSC 1141 at para 266.

¹⁰¹⁷ Factum of the Appellant, Her Majesty the Queen in Right of Canada, online: SCC <www.scc-csc.ca/WebDocuments-DocumentsWeb/39346/FM010_Apellant_Her-Majesty-the-Queen-in-Right-of-Canada.pdf> [perma.cc/8R2X-9M2Q] at para 3.

¹⁰¹⁸ *Ibid* at para 2.

¹⁰¹⁹ Factum of the Respondent, Cheyenne Sharma, online: SCC <www.scc-csc.ca/WebDocuments-DocumentsWeb/39346/FM020_Respondent_Cheyenne-Sharma.pdf> [perma.cc/Z3UW-3Z7P] at para 2.

¹⁰²⁰ *Ibid* at para 5.

¹⁰²¹ *Ibid* at para 5. Counsel for the Respondent explains further: "When Parliament enacts a high maximum sentence, this indicates that the offence is *potentially* serious (depending on the facts of the case). It does not indicate that the offence is *necessarily* serious. That a sentence has a high ceiling does not mean that it must have a high floor."

to human ones—and deferred to accordingly. In such a future, there would be no opportunity to challenge the constitutionality of the 2012 amendments rendering conditional sentences unavailable. Instead, the COMPAS algorithm would use historical—and ongoing—injustices to designate Ms. Sharma in a manner consistent with legacies of mistreatment.

An analogous situation occurred in *R v Ewert*, a 2018 case in which the SCC opined on the use of psychological and actuarial RAIs used by the Correctional Service of Canada (CSC) to assess inmates’ needs while incarcerated, as well as their likelihood of recidivism.¹⁰²² The complaint began in 2000, when Mr. Jeffrey Ewert, who identifies as Métis, argued that the RAIs were not developed with the specific circumstances of Indigenous offenders in mind, nor had their utility been established through the use of empirical research.¹⁰²³ While Mr. Ewert also sought declarations that the RAIs breached his *Charter* rights of equality and due process,¹⁰²⁴ the case turned on a specific provision of the *Corrections and Conditional Release Act (CCRA)*, section 24(1), which stipulates that the CSC must “take all reasonable steps to ensure that any information about an offender that it uses is as accurate, up to date and complete as possible.”¹⁰²⁵ The SCC confirmed that use of the RAIs violated s 24(1) of the *CCRA*, noting that the statutory obligation was sufficiently broad to encompass uses of psychological and actuarial RAIs exhibiting cultural bias, particularly given the long history of concerns about such tools.¹⁰²⁶ Especially in light of these catalogued concerns, it was the CSC’s responsibility to do additional investigation about the accuracy of its RAI tools.

As Teresa Scassa noted at the time of decision’s release, although *Ewert* is not explicitly about ADM, its basic components are present.¹⁰²⁷ The impugned tool was informed by a significant quantum of data and produced actuarial predictions in lieu of individualized assessment. Moreover, Mr. Ewert’s eventual success came not from *Charter* guarantees or human rights legislation, but from the *CCRA*: a federal legislative scheme that mandated accurate record-keeping.¹⁰²⁸ It follows that, given similar language in *PIPEDA* and the *Privacy Act*, statutes may provide some relief for those seeking to demonstrate problems with algorithmic decision-making.¹⁰²⁹ Jane Bailey and co-authors argue that the SCC’s reasoning in *Ewert* puts courts and tribunals on notice that they should “be alert to the biases baked into algorithms, rather than blithely treating them as mathematical certainties.”¹⁰³⁰

For COMPAS generally, the recognition of offender-specific circumstances enabled by the Gladue framework calls into question the accuracy of static categories, significant proportions of

¹⁰²² *R v Ewert*, 2018 SCC 30 [*Ewert*] at para 12.

¹⁰²³ *Ibid* at para 4.

¹⁰²⁴ Ewert sought declarations that his rights had been violated under section 7 and section 15 of the *Charter*. The Federal Court found a section 7 violation, but this was overturned by the Federal Court of Appeal (and confirmed by the SCC). *Ibid*.

¹⁰²⁵ *Ibid* at para 3.

¹⁰²⁶ *Ibid* at para 49, 53-55.

¹⁰²⁷ Teresa Scassa, “Supreme Court of Canada Decision Has Relevance for Addressing Bias in Algorithmic Decision-Making” (14 June 2018), online: *Teresa Scassa* <www.teresascassa.ca/index.php?option=com_k2&view=item&id=278:supreme-court-of-canada-decision-has-relevance-for-addressing-bias-in-algorithmic-decision-making&Itemid=80> [perma.cc/2LVZ-377R] [Scassa, “On *Ewert*”].

¹⁰²⁸ *Ibid*.

¹⁰²⁹ *Ibid*.

¹⁰³⁰ Jane Bailey, Jacquelyn Burkell, Suzie Dunn, Chandell Gosse & Valerie Steeves, “AI and Technology-Facilitated Violence and Abuse” in Scassa & Martin-Bariteau, *supra* note 191 249 at 263.

which are drawn from group data rather than individual data. Even coding certain criteria as binary within the algorithm presupposes a limited scope for highly intersectional identities. Throughout this chapter, many of the criticisms of COMPAS have been levied along racial lines, assuming a racial binary between white and black that obscures both people's true identities and the complexities that race introduces into policing. On this point, Bao et al note that race is a much more intersectional construct than such binary determinations tend to allow. It is nonetheless typically constructed simplistically within datasets, notwithstanding the instability racial designations show when scrutinized by intersectional scholars; for example, Hispanic is often mischaracterized as a race, rather than an ethnicity.¹⁰³¹ Huq acknowledges a similar shortcoming in his own analysis, pointing out that binary racial identities—and particularly their interaction with policing—simplify a complex racial reality.¹⁰³² For Canadian jurisprudence, the circumstances facing Indigenous offenders recognized in *R v Gladue* and *R v Ewert* remain a critical analytical point, but other racialized or marginalized groups might also benefit from a relational and/or storytelling approach.

COMPAS's Rule of Law problems demand storytelling solutions. The AI's affordances are the prioritization of quantifiable metrics in a highly discretionary procedure, triggering a transition to a formalist version of sentencing. Considering self-driving laws in the context of sentencing, rather than a prescriptive statutory provision that specifically mandates a particular outcome, enables an interesting contextual landscape. One might argue that a prescriptive provision (*e.g.*, part of the tax code) would better elucidate the challenges of translating a text-based instantiation of power from the written word coupled with human interpretation to regulation by machine. Yet, in my view, sentencing offers an ideal environment because of its openness. The specific experience of the offender—the sense that they are more than just a number—is eliminated by self-driving laws. Moreover, the integration of context-sensitive stories, like the lived realities of Indigenous offenders, are more difficult to account for in such a model.

As Hellman's lessons teach, the slope from wanting something to be true to believing something to be true can be very slippery. Combined with automation bias and automation complacency, the corresponding technological gloss threatens to entrench AI solutions in criminal sentencing, solidifying something that was always intended to be malleable. AI's data-driven methods enforce certainty by fiat, which is only possible if you ignore the qualitative aspects that might contribute to a decision. These context-sensitive stories continue to shift, change, and agitate—sometimes very slowly. Especially in light of how criminal law grapples with historical and ongoing legacies of discrimination, assuming that the task is even close to finished will have disastrous consequences. Recalling the analogy of law moving through the lens of a cinematographer's camera, sometimes a shot will need linger on a particular frame for a very long time in which nothing changes. Eventually, law moves, but imposing technology's pace will not spur its progress.

The *R v Gladue* decision, and its recent application in *R v Sharma*, emphasize how the insertion of lived experience is essential to achieving justice in the unique context of a given offender.

¹⁰³¹ Bao et al, *supra* note 911 at 3.

¹⁰³² Huq, *supra* note 950 at 1103.

Meanwhile, the *R v Ewert* decision helps contextualize how claims of unreasonableness or unfairness might contravene specific statutory requirements, enabling a judicial response to any potential claims of bias. Discretion, operating in this way, solidifies democratic ideals as an aspect of law's normativity. Hildebrandt explains:

This is where democracy depends on the Rule of Law, as Dworkin has demonstrated, effectively requiring equal respect and concern for each individual person. Democracy is not equivalent with majority rule; it entails that majorities rule in a way that allows minorities to become majorities.¹⁰³³

Focusing only on consistent application of rules serves to negate this important aspect that discretion serves. For the Rule of Law to function prospectively, its malleability must be preserved, permitting a narrative arc to bend to a better future.

Despite the discretionary nature of sentencing procedures, criminal law remains an area that is commonly recognized and acknowledged as law. Next, moving to facial recognition technology, I widen the ambit of what constitutes law. I turn to this next in Chapter 7.

¹⁰³³ Hildebrandt, "Affordance," *supra* note 32 at 122-123.

When faced with a difficult question, we often answer an easier one instead, usually without noticing the substitution.
~ Daniel Kahneman¹⁰³⁴

Once, not very long ago at all—in July 2019, to be precise—close friends welcomed a new baby boy to this complicated world. Let’s call him Baby K. For the first few months of his life, Baby K enjoyed many visits from his parents’ friends. Even though we were living in another city at the time, my partner and I managed to visit on a few occasions. In late August 2019, my partner commemorated his first time holding Baby K by using his smartphone camera to take a photo of himself with the teeny tiny new baby. Baby K was seven weeks old at the time of this first photo, and not capable of mugging for the camera. No, he was cuddled into my partner’s chest, face half-obscured. His eyes were closed; his cheeks scrunched up. He was adorable, to be sure, but he looked very little like the cherubic baby-faced toddler he would later grow into.

The first months of Baby K’s life wore on and, as babies tend to do, he grew and grew. Because of the distance, we didn’t see him all that often; when we did, his significant changes were readily apparent to us. Each subsequent visit, Baby K was bigger, with new facial expressions, chubbier cheeks, curlier hair. These visits, of course, were thoroughly documented with our smartphone cameras. Who can resist taking photos of an adorable new baby? And Baby K had more and more personality upon each visit; slowly, he was becoming capable of mugging for the camera. We took photos of Baby K in October 2019, December 2019, and February 2020. Nothing special or posed, just quick smartphone snaps to capture this special child’s progress. He wasn’t necessarily facing the camera perfectly or offering us his best angles. Sometimes, he was on his own in the pictures. Other times, he was being held by one of his parents, one of us, or other adults from our social circle. The ease of digital photography being what it is, we never did anything with these photos other than allow them to accumulate on the smartphone’s camera roll. They weren’t compiled, tagged, organized, shared, or uploaded to social media. They were simply there, for posterity’s sake.¹⁰³⁵

Then, of course, the COVID-19 pandemic brought an end to all visiting, especially with little babies wielding still-inexperienced immune systems. We next saw the not-so-new-baby in August 2020. Reminded by his smartphone’s memories feature of that first photo from a year earlier, my partner took another picture of himself with Baby K. We all marveled at how different Baby K was after the passage of a whole year: he had grown from a tiny little newborn into a robust toddler.

Yet, despite our recognition of Baby K’s growth and changes, it seemed the fundamental architecture of his face had not changed. After a visit in September 2021, more than two years after that initial photo—and a full year since the most recent picture of Baby K—we realized that the smartphone had been categorizing its photos with an automated face-tagging algorithm. The

¹⁰³⁴ Daniel Kahneman, *Thinking Fast and Slow* (Toronto: Doubleday Canada, 2011) at 12.

¹⁰³⁵ As one of my privacy-conscious friends is apt to say when photos are taken: “This is for posterity, not for posting!”

phone had grouped together photos of all those individuals frequently depicted on the camera roll. Incredibly, the algorithm had correctly grouped all the photos of Baby K, starting with his eyes closed, face half-obscured, cheeks scrunched up close-up at the ripe old age of seven weeks. These photos were months, even years, apart. They contained a variety of adult caregivers, each eager for their turn to hold the new baby. They had not been compiled, tagged, organized, shared, or uploaded to social media. And it's not like these photos were being grouped by virtue of Baby K being the only baby in town: photos of other babies were also present on the camera roll. Notwithstanding any of this, the smartphone's automated facial recognition technology had tracked Baby K's growth from his seven-week-closeup to a rambunctious child of two years and three months.

We were left wondering what kind of world Baby K would grow up to find. Would it be one where his biometric data, accumulated from birth by well-meaning adults with smartphones, was automatically sorted into a thorough data repository? Would he walk down the street with digital billboards updating to show him personalized advertising, thanks to his face being identified by ubiquitous cameras? Would he experience legal protection for privacy? Would he even know what the term "privacy" was intended to connote?

This chapter considers facial recognition technology and its privacy dimensions. As a framing device for Rule of Law problems, privacy law offers an especially interesting field of inquiry because of its uneasy presence in the common law landscape generally. It is part statute, part tort, part criminal law with a quasi-constitutional dimension, part contractual obligation, part nebulous theoretical construct. In this way, Chapter 7's analysis diverges from that of Chapter 6, which addressed a specific allocation of discretionary power within the judicial system.

Using a Canadian privacy law approach, patchwork legal protection can be found housed in section 8 of the *Charter of Rights and Freedoms*,¹⁰³⁶ federal statutes including the *Privacy Act* and the *Personal Information Protection and Electronic Documents Act (PIPEDA)*,¹⁰³⁷ analogous provincial statutes—which vary by jurisdiction—and tort law. Further complications are driven by the creep of international standards,¹⁰³⁸ corporate privacy policies,¹⁰³⁹ and related commercial practices. What follows is not intended to be an exhaustive account of the Canadian privacy framework, nor a thorough accounting of how technology upends previous legal distinctions between public and private space.¹⁰⁴⁰ Instead, my analysis is intended to reinforce the notion that using new technologies to inform legal decisions (or instances of state power) can transpose the character, quality, and provenance of the information on which such decisions are based.

Throughout legal history, privacy norms are typically adjudicated at moments when new technologies increase human capabilities in unanticipated ways. Privacy law's malleability to

¹⁰³⁶ *Charter*, *supra* note 828, s 8.

¹⁰³⁷ *Privacy Act*, RSC 1985, c P-21; *Personal Information Protection and Electronic Documents Act*, SC 2000, c 5 [PIPEDA].

¹⁰³⁸ Notably, *California's Consumer Protection and Privacy Act* and Europe's *GDPR*. *California Consumer Privacy Act*, 2018 Cal Legis Serv, Ch 55 (AB 375) [CCPA]; *GDPR*, *supra* note 496.

¹⁰³⁹ In the wake of the *GDPR*, stricter privacy regulations for companies operating in the EU has resulted in many companies amending their privacy policies worldwide, for simplicity's sake.

¹⁰⁴⁰ Indeed, you could write a whole thesis about just this latter point! See Kristen Thomasen, "Private Law & Public Space: The Canadian Privacy Torts in an Era of Personal Remote-Surveillance Technology" PhD Dissertation (2022), online (pdf): *UOttawa* <ruor.uottawa.ca/handle/10393/43746> [perma.cc/D22W-RE7T].

sociotechnical forces makes it fertile ground for commentary by technology law scholars. In this chapter, I identify, using a privacy law lens, the Rule of Law problems that arise from AI/ML-powered facial recognition technology. I begin the chapter with the framing example of facial recognition technology, particularly its use by state power: the police. This analysis continues the threads of some of the racial dimensions of policing discussed in Chapter 6, referencing predictive policing as another automated decision-making system that makes use of individual data to allocate social resources by attempts to predict the future. Yet, focusing on state uses of this technology would tell only part of the story. Privacy's patchwork legal nature, as well as the concentration of power in corporations enabled by AI/ML technology, necessitates a wider approach. Therefore, after setting out the problems with facial recognition as a specific technology, I broaden the lens to provide an overview of legal privacy in general. This approach extends the analysis put forward in Chapter 6, further complicating the instantiations of power and legal consequences that flow from privacy-invasive technologies. Concurrently, thinking through the values prioritized by AI/ML-driven approaches engages different theoretical approaches to privacy, none of which offer thoroughly satisfactory explanations for the discomfort that privacy violations enabled by new technological innovations can entail. Approaching the identified Rule of Law problems through a storytelling framework necessitates a relational approach to privacy theory, which I demonstrate in the final section through engaging the work of scholars including Ian Kerr, Jennifer Nedelsky, and Helen Nissenbaum.

1. Facial Recognition Technology

Facial recognition technology (FRT) exists at an intersection between technological design, consumer profiling, and the uneasy legal privacy landscape. Facial recognition is an example of biometric technology, used to identify individuals based on their unique characteristics.¹⁰⁴¹ Importantly, FRT differs from other biometric identifiers since it can be used from a distance, meaning the individual impacted is not necessarily aware that they are being surveilled.¹⁰⁴² Simply put, FRTs are those capable of matching faces with digital records, which is usually achieved by breaking down recorded faces into their constituent components.¹⁰⁴³ Techniques for accomplishing this feat are sometimes “feature-based,” powered by measuring the distances between key facial features like the mouth, nose, or eyes; and sometimes “appearance-based,” powered by holistically matching the whole face.¹⁰⁴⁴ Either process creates a digital representation of the face referred to as a “template” or a “faceprint.”¹⁰⁴⁵ Sometimes, FRT is a choice made by the consumer to enhance security and promote convenience. One commonly used FRT, Apple's Face ID, is available for iPhone and iPad users to unlock their devices simply by looking at them.¹⁰⁴⁶ This technology works by capturing “face data” by “projecting and analyzing thousands of invisible dots to create a depth

¹⁰⁴¹ Other examples of biometric identification include fingerprints, gait recognition, and iris scanning. Joy Buolamwini, Vicente Ordóñez, Jamie Morgenstern, & Erik Learned-Miller “Facial Recognition Technologies: A Primer” (29 May 2020), online (pdf): *UMass CS* <people.cs.umass.edu/~elm/papers/FRTprimer.pdf> [perma.cc/4BCD-M4SC] at 8 [Buolamwini et al, “FRT”].

¹⁰⁴² Contrast this with fingerprinting, where the individual is usually well-aware that their fingerprints are being taken.

¹⁰⁴³ Andrew G Ferguson, “Facial Recognition and the Fourth Amendment” (2019) 105 *Minnesota Law Review* 1105 at 1110.

¹⁰⁴⁴ *Ibid* at 1110-1111. Some newer technologies use other criteria, including skin texture, shadow, three-dimensional models, or a combination of these factors.

¹⁰⁴⁵ Buolamwini et al, FRT, *supra* note 1041 at 10.

¹⁰⁴⁶ Other technology brands have analogous applications. *Ibid* at 7.

map of your face and also captures an infrared image of your face.”¹⁰⁴⁷ Highly adaptable, Apple’s Face ID is able to make adjustments depending on whether you are wearing glasses or sunglasses, foregoing eyewear in favour of contact lenses, growing a new beard, or wearing scarves or hats.¹⁰⁴⁸

Typically, iPhone or iPad users choose to enrol in Face ID for the simplicity it offers over entering a numerical passcode. Apple also boasts a high degree of security: the information is highly encrypted.¹⁰⁴⁹ Because of the machine learning technology used to produce the facial depth map, Apple claims that the probability that someone else could unlock your device is less than 1 in 1,000,000—and the device allows only five unsuccessful attempts before asking for the passcode.¹⁰⁵⁰ Since the COVID-19 pandemic had resulted in many people wearing face masks in public places, Apple has even created a new feature that permits facial recognition using Face ID during mask wearing. Users simply must set up the feature, which seems to focus on the eyes.¹⁰⁵¹

Yet, the Face ID technology is characterized by voluntary enrolment. The availability of FRT has become pervasive far beyond convenience-seeking consumers, with new applications regularly introduced. Facial recognition technologies have become increasingly widespread in recent years, ranging from scanning public crowds in real time to assessing attendance at schools, stores, or places of worship to analyzing accumulated datasets comprised of photos or videos.¹⁰⁵² FRT is also deployed at transportation hubs, with facial recognition used at kiosks to verify passports and grant access to public transportation.¹⁰⁵³ In some situations, facial recognition is deployed as a verification measure for ticketing, including at large concerts or sporting events.¹⁰⁵⁴ Social media has created a treasure trove of facial data enabling the training of facial recognition algorithms, now able to identify faces with surprising accuracy.

For example, consider the facial recognition algorithms deployed by social media giant Facebook. Notably, Facebook’s facial recognition boasts impressive capabilities: at the time this feature was launched, it claimed to be able to identify a face accurately 98 per cent of the time.¹⁰⁵⁵ As a comparator, at the same time, FBI’s facial recognition software, Next Generation Identification (NGI), was able to identify someone from a list of the top 50 people only 85 per cent of the time.¹⁰⁵⁶

¹⁰⁴⁷ The Apple website describes this technology as being powered by a “TrueDepth camera,” which captures the relevant face data to create a depth map. Then, a neural engine “transforms the depth map and the infrared image into a mathematical representation and compares that representation to the enrolled facial data.” Apple, “About Face ID advanced technology” (27 April 2022), online: *Apple Support* <support.apple.com/en-ca/HT208108#:~:text=Face%20ID%20automatically%20adapts%20to,it%20updates%20your%20face%20data> [perma.cc/S45K-C95J].

¹⁰⁴⁸ *Ibid.*

¹⁰⁴⁹ *Ibid.*

¹⁰⁵⁰ *Ibid.*

¹⁰⁵¹ *Ibid.*

¹⁰⁵² Buolamwini et al, “FRT,” *supra* note 1041 at 7.

¹⁰⁵³ *Ibid* at 8. Other transportation-focused usages mentioned in the report include New York City subways displaying passenger faces to deter fare dodgers and car manufacturers using FRT to allow vehicle access and monitor inattentive/drowsy drivers who have taken their eyes off the road.

¹⁰⁵⁴ Ferguson, *supra* note 1043 at 1107. Ferguson references facial recognition technology being used to surveil crowds at the Super Bowl since at least 2001. See also Declan McCullagh, “Call It Super Bowl Face Scan I” (February 2, 2001), online: *WIRED* <www.wired.com/2001/02/call-it-super-bowl-face-scan-i/> [perma.cc/XMP8-S6JL].

¹⁰⁵⁵ Privacy & Security, “Facebook’s Facial Recognition Software Is Different From The FBI’s. Here’s Why” (18 May 2016), online: *NPR* <www.npr.org/sections/alltechconsidered/2016/05/18/477819617/facebooks-facial-recognition-software-is-different-from-the-fbis-heres-why> [perma.cc/4JUZ-MUQ3] [NPR].

¹⁰⁵⁶ *Ibid.*

The difference arises from Facebook users uploading photos of people from multiple angles, in myriad different situations, and under different lighting.¹⁰⁵⁷ The FBI, on the other hand, has only mug shots in its NGI database, all taken from a single, “head-on” vantage point, and originating as part of an arrest.¹⁰⁵⁸ This statistic has likely improved since the relevant reports were published: several years ago, the FBI indicated its plans to merge its criminal database with its non-criminal database to generate more accurate NGI results.¹⁰⁵⁹ This non-criminal data would include any biometric and biographic data submitted through fingerprint checks, background checks, or licensing requirements—a necessity for many types of jobs.¹⁰⁶⁰

Facebook first introduced its “automatic” tagging feature in late 2010, unveiling a new feature that presented “suggested” tags to assist with tagging their photos.¹⁰⁶¹ Complaints from privacy watchdogs in Canada and EU prevented these features from being introduced more broadly, but users in the United States became accustomed to receiving “suggestions” about whose faces they should tag during their interactions with the platform.¹⁰⁶² Upon navigation to the photo tagging screen, an algorithm identifies faces and provides “suggestions” for who might be in the photo based on attempts to match those faces with Facebook’s database of profiles of people.¹⁰⁶³ While this practice was undoubtedly privacy invasive, it was originally applied only to the photos that users submitted to undergo the tagging process. However, beginning in mid-December 2017, Facebook modified its default settings such that all photos uploaded to the social networking site would now be automatically scanned by facial recognition algorithms.¹⁰⁶⁴ This shift is important: every single photo now underwent automatic review upon upload.¹⁰⁶⁵ All photos shared on the social network became fodder for better training a more advanced ML algorithm, laying the groundwork for improved FRT.¹⁰⁶⁶

Social media presents an interesting use case, since users can upload photos containing anyone—whether those individuals have granted consent or not. ML-enabled tagging techniques, and the default settings encouraging the mandatory adoption of those techniques on various platforms, can lead to even the most privacy-adverse individuals having significant presences in facial recognition databases. Information-sharing agreements between law enforcement agencies further complicate the problem: the NGI data collected by the FBI is shared with over 23,000 different entities

¹⁰⁵⁷ NPR, *supra* note 1055.

¹⁰⁵⁸ *Ibid.* See also news reports about Clearview AI, which was allegedly first unveiled as a perk for investors and wealthy clients: Kashmir Hill, “Before Clearview Became a Police Tool, It Was a Secret Plaything of the Rich” (6 March 2020), online: *The New York Times* <www.nytimes.com/2020/03/05/technology/clearview-investors.html> [perma.cc/7DM3-SFKD] [Hill, “Clearview”].

¹⁰⁵⁹ Jennifer Lynch, “Face Off: Law Enforcement Use of Face Recognition Technology” (12 February 2018), online (pdf): *EFF* <www.eff.org/wp/law-enforcement-use-face-recognition> [perma.cc/X5A7-LNES] at 14.

¹⁰⁶⁰ *Ibid.*

¹⁰⁶¹ Tom Simonite, “Facebook Can Now Find Your Face Even When Its Not Tagged” (19 December 2017), online: *WIRED* <www.wired.com/story/facebook-will-find-your-face-even-when-its-not-tagged/> [perma.cc/UA3U-T27F].

¹⁰⁶² *Ibid.*

¹⁰⁶³ Sidney Fussell, “Facebook’s New Face Recognition Features: What We Do (and Don’t) Know” (27 February 2018), online: *Gizmodo* <gizmodo.com/facebooks-new-face-recognition-features-what-we-do-an-1823359911> [perma.cc/2GNY-H9N6].

¹⁰⁶⁴ *Ibid.*

¹⁰⁶⁵ *Ibid.*

¹⁰⁶⁶ In late 2021, Facebook began stating they intended to shut down their facial recognition systems, considering “many concerns about the place of facial recognition technology in society.” See Kashmir Hill & Ryan Mac, “Facebook, Citing Societal Concerns, Plans to Shut Down Facial Recognition System” (2 November 2021), online: *The New York Times* <www.nytimes.com/2021/11/02/technology/facebook-facial-recognition.html> [perma.cc/TW45-Y2SV]. Many news articles confirm this announcement; I could find no confirmations that FRT technology had indeed been eliminated from the platform.

internationally.¹⁰⁶⁷ The race was on for private companies to create successful facial recognition technologies that they could market to different police forces across the world. Electronic commerce giant Amazon developed its own facial recognition alternative, called Rekognition, which was largely decried for exhibiting racial biases.¹⁰⁶⁸ After several years of advocacy, both IBM and Microsoft discontinued their facial recognition efforts, leading Amazon to impose a one-year moratorium on sales to police departments.¹⁰⁶⁹ This ban was later extended indefinitely, as advocates continue to impugn the software’s accuracy and federal regulators have made no moves to either ban or regulate the technology.¹⁰⁷⁰

The concerns raised by FRT have resulted in some scholars advocating for outright bans of the technology. Scholars Woodrow Hartzog and Evan Selinger describe it as “an irresistible tool for oppression that’s perfectly suited for governments to display unprecedented authoritarian control and an all-out privacy-eviscerating machine.”¹⁰⁷¹ Because our faces are typically fair game for the world to see as we traverse through public space,¹⁰⁷² FRT offers an evocative example of how privacy can be violated by the aggregation of publicly available information. Faces, as Hartzog and Selinger point out, “unlike fingerprints, gait, or iris patterns, are central to our identity. Faces are conduits between our on- and offline lives, and they can be the thread that connects all of our real-name, anonymous, and pseudonymous activities.”¹⁰⁷³ When faces are collected into an ML-powered database, then turned into a user-friendly interface for subscription use by police departments, the corresponding privacy issues are considerable. Such privacy issues were exactly what happened with Clearview AI.

a. FRT Example: Clearview AI

In early 2020, a new face recognition technology, Clearview AI, began making headlines for its uses by police forces in both the United States and Canada. Technology reporter Kashmir Hill published a significant exposé in the *New York Times* publicizing the significant privacy implications of the application.¹⁰⁷⁴ While Clearview AI was already known to industry watchdogs

¹⁰⁶⁷ Lynch, *supra* note 1059.

¹⁰⁶⁸ Karen Hao, “The two-year fight to stop Amazon from selling face recognition to the police” (12 June 2020), online: *MIT Technology Review* <www.technologyreview.com/2020/06/12/1003482/amazon-stopped-selling-police-face-recognition-fight/> [perma.cc/MB76-6WCN].

¹⁰⁶⁹ *Ibid.*

¹⁰⁷⁰ Karen Weise, “Amazon indefinitely extends a moratorium on the police use of its facial recognition software” (18 May 2021), online: *The New York Times* <www.nytimes.com/2021/05/18/business/amazon-police-facial-recognition.html> [perma.cc/8B8J-Q9SY].

¹⁰⁷¹ Woodrow Hartzog & Evan Selinger, “Facial Recognition Is the Perfect Tool for Oppression” (2 August 2018), online: *Medium* <medium.com/s/story/facial-recognition-is-the-perfect-tool-for-oppression-bc2a08f0fe66> [perma.cc/6QJE-H777].

¹⁰⁷² In the early days of the COVID-19 pandemic, some feared that the widespread practice of masking would stymie growth of the FRT sector. Before long, many companies reported successful FRT systems that were not impacted by mask wearing. While results varied between systems, many reported impressive success rates: notably, Japanese FRT designed to detect faces even when a user is wearing a mask for allergies claims 99.9% accuracy. Even the happiest place on earth—Walt Disney World—reported use of a touchless FRT admission process, which did not require customers to remove their masks. See James Clayton, “Facial recognition beats the Covid-mask challenge” (25 March 2021), online: *BBC News* <www.bbc.com/news/technology-56517033> [perma.cc/69V8-QGY2].

¹⁰⁷³ Hartzog & Selinger, *supra* note 1071.

¹⁰⁷⁴ Hill, “Clearview,” *supra* note 1058

and celebrity investors,¹⁰⁷⁵ Hill’s article brought the application’s rapid adoption by law enforcement into public knowledge. The fact that it had been so eagerly adopted by North American police forces was quickly and widely decried. The potential for both civil rights violations and privacy violations seemed rampant. By the time of Hill’s exposé, the Clearview AI database already contained over three billion images and over 600 law enforcement agencies across the United States and Canada had already begun using it.¹⁰⁷⁶ Even more frighteningly, its underlying code had been prepared for deployment on augmented reality glasses, which would allow subscribers to identify passerby by merely looking at them.¹⁰⁷⁷ In her article, Hill interviewed one of Clearview AI’s founding investors, David Scalza, who considered FRT to be an exciting investment opportunity because of its potential as a crime-solving tool. Pressed on the implications of the technology—already banned in some cities, including the technology hub of San Francisco—he acknowledged his belief that “because information constantly increases, there’s never going to be privacy” and that while “laws have to determine what’s legal, you can’t ban technology.”¹⁰⁷⁸ Such a technologically deterministic approach spells disaster for the delicate balance of user rights.¹⁰⁷⁹

In February 2020, Canada’s Office of the Privacy Commissioner (OPC) announced its intention to jointly investigate Clearview AI in tandem with its provincial counterparts in Quebec, Alberta, and BC to assess Clearview AI’s compliance with Canadian privacy legislation.¹⁰⁸⁰ In the aftermath, many Canadian police forces terminated their relationships with Clearview AI, pending the outcome of the investigation. A few short months thereafter, Clearview AI announced its withdrawal from the Canadian marketplace, ending its one remaining contract with the RCMP.¹⁰⁸¹

¹⁰⁷⁵ Hill begins her article with billionaire grocery chain owner John Catsimatidis using Clearview to identify his daughter’s date. Hill, *supra* note 1058. Clearview responded that it as “available only for law enforcement agencies and select security professionals to use as an investigative tool.” Later, BuzzFeed News obtained Clearview’s client list, which showcased many private companies, including Madison Square Garden, Las Vegas Sands and Pechanga Resort Casino, the NBA, fitness centre Equinox, and cryptocurrency exchange Coinbase. Clearview responded that it had offered free trials to investors and strategic partners. On an episode of the popular YouTube series “Hot Ones,” in which celebrities are asked increasingly revealing questions while consuming increasingly spicy wings, celebrity venture capitalist Ashton Kutcher spoke about a “terrifying” facial recognition app that sounded a lot like Clearview AI. Thomas Macaulay, “Clearview AI can be fun—if you’re dirty, stinking rich” (6 March 2020), online: *The Next Web* <thenextweb.com/news/clearview-ai-can-be-fun-if-youre-dirty-stinking-rich> [perma.cc/PDW4-CUFF]. See also Ryan Mac, Caroline Haskins & Logan McDonald, “Clearview’s Facial Recognition App Has Been Used by The Justice Department, ICE, Macy’s, Walmart, and the NBA” (27 February 2020), online: *BuzzFeed News* <www.buzzfeednews.com/article/ryanmac/clearview-ai-fbi-ice-global-law-enforcement> [perma.cc/GRT2-2UYU].

¹⁰⁷⁶ Hill, “Clearview,” *supra* note 1058.

¹⁰⁷⁷ *Ibid.*

¹⁰⁷⁸ Scalza went on to opine [about the trajectory of FRT technology]: “Sure, that might lead to a dystopian future or something, but you can’t ban it.” *Ibid.*

¹⁰⁷⁹ Refer to Chapter 2 for a thorough discussion on technological determinism. Some scholars advocate for a more technologically neutral approach, like Judith Donath, who thinks facial recognition regulation should be tailored to “identify the specific activities to be prohibited, the harms to be avoided, and the values, rights, and situations we are trying to protect.” Advocating for an outright ban, Hartzog & Selinger concede that this is usually the right approach to regulate new technology, but FRT deserves a ban for its nefarious character. Hartzog & Selinger, *supra* note 1071.

¹⁰⁸⁰ OPC, “Commissioners launch joint investigation into Clearview AI amid growing concerns over use of facial recognition technology” (21 February 2020), online: *OPC* <www.priv.gc.ca/en/opc-news/news-and-announcements/2020/an_200221/> [perma.cc/F6UW-V576].

¹⁰⁸¹ OPC News Release, “Clearview AI ceases offering its facial recognition technology in Canada” (6 July 2020), online: *OPC* <www.priv.gc.ca/en/opc-news/news-and-announcements/2020/nr-c_200706/> [perma.cc/55B6-SN98].

In February 2021, the OPC and its provincial counterparts released their joint report into Clearview AI's activities in Canada.¹⁰⁸² They had investigated whether Clearview had obtained the requisite consent to collect, use, and disclose personal information in Canada; whether Clearview had collected, used, and disclosed personal information for an appropriate purpose; and—pursuant to the specific statutory scheme in Quebec—whether Clearview had appropriately reported the creation of a database that contained biometric characteristics or measurements.¹⁰⁸³ Ultimately, the report concluded that Clearview AI had indeed collected, used, and disclosed information without consent; and that their collections, uses, and disclosures were of such that the reasonable person would find inappropriate.¹⁰⁸⁴ They issued a series of recommendations, including that Clearview cease offering its FRT in Canada; that Clearview cease collection, use, and disclosure of images and biometric arrays from Canadians; and that it should delete its existing images and biometric arrays pertaining to Canadians.¹⁰⁸⁵ While Clearview had requested the investigation be terminated since they had suspended their Canadian operations, the OPC disagreed: they wanted to finish their investigation to ensure that other FRT-focused organizations, who were becoming increasingly common, would have the benefit of their written conclusions on FRT compliance with Canadian privacy law.¹⁰⁸⁶ The OPC also noted the quasi-constitutional nature of privacy rights, confirmed by the SCC's jurisprudence, which has been understood to mean that privacy rights should be given a "broad, purposive and liberal interpretation, and restrictions on those rights should be interpreted narrowly."¹⁰⁸⁷

Clearview's submissions to the OPC further clarified the scope of their AI system's operations. They confirmed that their tool worked by webcrawling the Internet to scrape images of faces from publicly available online sources, which it then converted into a numerical representation of each image to generate a unique biometric identifier.¹⁰⁸⁸ The Clearview AI interface permitted users to upload images to be assessed against the existing biometric identifiers in the database, providing users with successful results, comprised of the matching images and corresponding metadata, as well as a link to location of the publicly available online image.¹⁰⁸⁹ Clearview did not obtain consent to collect this data, relying instead on the *PIPEDA* exception that the data was publicly available.¹⁰⁹⁰ It noted that it did not collect images from protected social media accounts, and it operated in conformity with any instructions coded on the webpage.¹⁰⁹¹ Its FRT operations were based on five components: an image crawler that autonomously searched public webpages to collect data; an image store and a metadata store, both located on Clearview's servers; a neural network capable of analysing digital images and turning them into numerical representations (called

¹⁰⁸² OPC, "Joint investigation of Clearview AI, Inc. by the Office of the Privacy Commissioner of Canada, the Commission d'accès à l'information du Québec, the Information and Privacy Commissioner for British Columbia, and the Information Privacy Commissioner of Alberta" *PIPEDA Findings #2021-001* (2 February 2021), online: *OPC* <www.priv.gc.ca/en/opc-actions-and-decisions/investigations/investigations-into-businesses/2021/pipeda-2021-001> [perma.cc/M5MP-NREC] [OPC, "Clearview Investigation"].

¹⁰⁸³ *Ibid.*

¹⁰⁸⁴ *Ibid.*

¹⁰⁸⁵ *Ibid.*

¹⁰⁸⁶ *Ibid.*

¹⁰⁸⁷ This proposition has been discussed in recent cases, notably *Alberta (Information and Privacy Commissioner) v United Food and Commercial Workers, Local 401*, 2013 SCC 62 [*United Food*].

¹⁰⁸⁸ OPC, "Clearview Investigation," *supra* note 1082.

¹⁰⁸⁹ *Ibid.*

¹⁰⁹⁰ *Ibid.*

¹⁰⁹¹ These instructions appear in something known as a robots.txt file. *Ibid.*

vectors); and a vector database, made up a 512 datapoints designed to represent the unique geometry of each individual face.¹⁰⁹² Any user wishing to identify an individual would upload their photograph of the target to the database, which then analyzed the target image and produced a vector to be compared against all the pre-existing vectors in the database.¹⁰⁹³

Clearview attempted to downplay the privacy risks associated with its collection and aggregation of facial images. In its submissions, Clearview emphasized that no individual user can simply browse through the database to see the amassed data, which comprised over 3 billion images; instead, it produced only matches in response to targeted searches.¹⁰⁹⁴ Furthermore, it claimed, there was no potential harm: even if the entirety of the Clearview database were compromised and released, all the impugned images were already available online in other locations.¹⁰⁹⁵ The OPC and its counterparts disagreed with this characterization, noting that no one who uploads their photo to the Internet for personal use reasonably expects a webcrawling bot to scrape that image for a facial recognition database to be used by law enforcement.¹⁰⁹⁶ Despite the unreasonableness of this assumption, it does seem to have become relatively commonplace in the past few years.

b. FRT Use in Predictive Policing

The use of Clearview AI by Canadian police forces speaks to a larger trend in law enforcement: predictive policing.¹⁰⁹⁷ Broadly, predictive policing refers to the use of data-driven techniques, often powered by some form of ML, to allocate police resources based on predictions about where crime is most likely to occur or who is most likely to commit crimes. The focus is on creating data forecasts, which can range from generating maps of crime hotspots to identifying trends in crime statistics.¹⁰⁹⁸ The predictive capabilities of such approaches are relatively limited; some suggest this is simply a marketing term.¹⁰⁹⁹ Predictions typically flow from police-generated datasets, which take two predominant forms. One is a location-focused policing technology, which attempts to identify key locations where criminal activity might occur based on historical police data.¹¹⁰⁰ The other is a person-focused predictive technology, which attempts to predict who might commit a crime (or, who might be a victim to a crime).¹¹⁰¹

Predictive policing encompasses more than just FRT technology: it can refer to any analytical methods of computing potential risk of crime. In Chicago, a method called social network analysis is used to generate a “Strategic Subjects List,” also known as an “SSL” or “heat list,” with scores

¹⁰⁹² OPC, “Clearview Investigation,” *supra* note 1082.

¹⁰⁹³ *Ibid.*

¹⁰⁹⁴ *Ibid.*

¹⁰⁹⁵ While the OPC did not explicitly investigate Clearview’s security systems, it did note that Clearview had suffered two publicized breaches in the past year alone: a leaked client list in February 2020 and its source code and pilot project video in April 2020. *Ibid.*

¹⁰⁹⁶ *Ibid.*

¹⁰⁹⁷ After a review of the available “predictive policing” methods, authors Robinson & Koepke had a sceptical view of the accuracy of the term, writing: “The fact that we even call these systems ‘predictive’ is itself a telling sign of excessive confidence in the systems. The systems really make general forecasts, not specific predictions. A more responsible term—and one more accurately evocative of the uncertainty inherent in these systems, would be ‘forecasting.’” David Robinson & Logan Koepke, “Stuck in a Pattern: Early evidence on “predictive policing” and civil rights” (August 2016), online: *Upturn* <www.stuckinapattern.org> [perma.cc/N9K2-C5ZS] at 2.

¹⁰⁹⁸ *Ibid.* at 2.

¹⁰⁹⁹ *Ibid.*

¹¹⁰⁰ Robertson et al, *supra* note 866 at 2. See also Robinson & Koepke, *supra* note 1097 at 2.

¹¹⁰¹ *Ibid.*

generated based on a person's past criminal behaviour, the other people they were arrested alongside, and alleged gang affiliations.¹¹⁰² Within Canada, both the Vancouver and Saskatoon police admit to using some forms of algorithmic surveillance policing technology, with the Calgary police also conducting social network analysis.¹¹⁰³ One of the most publicized examples of predictive policing was a technology called PredPol, developed by two university professors, who used ML techniques to parse police departments' historical data and create maps predicting where crimes might happen.¹¹⁰⁴ Yet, for at least the last 30 years, criminological data has been shown to be a poor indicator of crime.¹¹⁰⁵ The numbers tend to better reflect policing tactics than actual criminality: over time, this sampling bias can distort law enforcement efforts, reinforcing discrimination.¹¹⁰⁶ This phenomenon is known as the "ratchet effect."¹¹⁰⁷ As Cathy O'Neil reminds us, people tend to trust numbers to be intrinsically objective, but the biased nature of the underlying data can create poor representations of reality.¹¹⁰⁸ Cycles of poverty and recidivism can create circumstances that give rise to police scrutiny, leading to disproportionate police attention in marginalized communities.¹¹⁰⁹ Further, there are questions about the sources and accuracy of data, which might be generated through acontextual sources or racially-motivated practices like police stops and carding.¹¹¹⁰ Without thoughtful construction of data-based feedback loops, using predictive systems can work to impede efforts towards building community safety or trust.¹¹¹¹

FRT and other imaging techniques tend to be an important component of predictive policing efforts. Often, these approaches offer little by way of actual prediction: they tend to be comprised of ordinary surveillance and monitoring functions, which often just automate existing law enforcement databases.¹¹¹² In a sweeping 2016 report called "The Perpetual Lineup," scholars at the Georgetown Law Center on Privacy and Technology spent over a year cataloguing how FRT is used by police forces across the United States.¹¹¹³ They found that roughly 1 in 2 adults in the U.S. has had their photo searched through law enforcement's use of FRT.¹¹¹⁴ They distinguish between someone who is identified based on FRT after being legally stopped *versus* a search of an FRT driver licence database for investigatory purposes *versus* scans taken of people walking in

¹¹⁰² Robinson & Koepke, *supra* note 1097 at 3.

¹¹⁰³ Robertson et al, *supra* note 866 at 2.

¹¹⁰⁴ After featuring in a number of critiques of the overall predictive policing initiative, PredPol announced in March 2021 that they were changing their name to "Geolitica" to better represent their intended focus and direction as a company. The new name is intended as a mashup of "geographical analytics." Announcing the change on their company blog, they noted that the idea of "predictive policing," which they had helped popularize and define by virtue of selecting that name in 2012, "has broadened to include activities—such as facial recognition or 'predicting' that certain individuals will commit crimes, with which we are not aligned; even the use of the word 'predictive' does not accurately describe our business." Geolitica, "Predictive Policing Blog" (2 March 2021), online: *PredPol* <blog.predpol.com/geolitica-a-new-name-a-new-focus> [perma.cc/N9K2-C5ZS].

¹¹⁰⁵ Robinson & Koepke, *supra* note 1097 at 5.

¹¹⁰⁶ *Ibid* at 6.

¹¹⁰⁷ *Ibid*.

¹¹⁰⁸ *Ibid* at 6-7.

¹¹⁰⁹ For example, note that Indigenous people make up 5% of Canada's population, but 30% of its prison population. Robertson et al, *supra* note 866 at 16-17.

¹¹¹⁰ The Citizen Lab report notes that such data might come from sources including criminal survey statistics, social media posts, geolocation data, crisis centre call logs, hospital injury data, criminal activity data collected by non-police security personnel such as transit, campus security, mall cops, etc. *Ibid* at 18-19.

¹¹¹¹ Robinson & Koepke, *supra* note 1097 at 6.

¹¹¹² For example, FRT can be applied to existing mugshot databases. Robertson et al, *supra* note 866 at 2.

¹¹¹³ Clare Garvie, Alvaro Bedoya & Jonathan Frankle, "The Perpetual Lineup: Unregulated Police Face Recognition in America" Center for Privacy & Technology, Georgetown Law (18 October 2016), online: *Perpetual Lineup* <www.perpetuallineup.org> [perma.cc/Q2QL-P44A].

¹¹¹⁴ *Ibid*.

public space, in real time.¹¹¹⁵ While the first situation is specific, targeted, and part of a public accountability process, the latter situations are conducted invisibly and apply generally to everyone—not only someone legally stopped or taken into police custody.¹¹¹⁶ The creep of using privacy-invasive technology from a targeted, specific instance to becoming a generalizable societal norm is especially concerning given the malleability of reasonable expectation of privacy standards, which govern both constitutional and statutory understandings of when someone is owed privacy protection.¹¹¹⁷ Absent careful safeguards to preserve appropriate context, the potential for such technology to erode reasonable expectations is imminent. The Perpetual Lineup Report authors note the expansion of search powers through FRT tracking individuals' movement through public space without asking them to produce identification and enabling police geolocation technologies to far outpace current possibilities.¹¹¹⁸ Fundamentally, such activities could stifle protest participation or free speech, extend police powers, and exacerbate bias issues caused by overpolicing certain groups.

In December 2021, a group of Canadian scholars offered submissions to the Toronto Police Services Board on best practices for the use of AI technologies.¹¹¹⁹ Among their many recommendations, they emphasized that the use of AI by police should not be seen as inevitable, that police services should avoid engaging in unnecessary and/or disproportionate mass collection and analysis of data, and that accuracy in AI systems does not mitigate inequality.¹¹²⁰ The report authors noted that AI, and particularly FRT, have been known to amplify discriminatory biases and generate harm, through large-scale data collection and corresponding analysis, and that the underlying data might be amassed in ways that violate privacy—either illegally or unethically.¹¹²¹ These observations bring to mind the significant critiques from marginalized groups about FRT.

c. Equality-Seeking Critiques

Many critiques have been levied against FRT from an equality-seeking perspective. One of the first outcries that resonated in the public consciousness was the Google user who identified that Google's image search had accidentally tagged two African American individuals as "gorillas."¹¹²² Widely shared on social media, this created a public relations disaster for the company, who emphatically apologized and removed the offending image search result. However, their solution for the problem left something to be desired: instead of improving the technology used to identify

¹¹¹⁵ Garvie et al, *supra* note 1113.

¹¹¹⁶ *Ibid.*

¹¹¹⁷ Furthermore, there is a distinction between biometric databases made up of people who have been previously involved in criminal activity (fingerprints/mugshots) *versus* law-abiding people guilty only of having moved through public space within view of a camera. *Ibid.*

¹¹¹⁸ For example, a major city like Chicago might have a handful of "Stingray" devices capable of retrieving relevant data from cell phone towers to amass geolocation data. The same city might have 10,000 surveillance cameras: once equipped with FRT, these could transform the nature of public space. *Ibid.*

¹¹¹⁹ Kristen Thomasen et al, "Submission to the Toronto Police Services Board's Use of New Artificial Intelligence Technologies Policy – LEAF and the Citizen Lab" (20 December 2021), online (pdf): *SSRN*

<papers.ssrn.com/sol3/papers.cfm?abstract_id=3989271 > [perma.cc/MB3J-MATE] [Thomasen et al, "TPSB Submissions"].

¹¹²⁰ *Ibid.*

¹¹²¹ *Ibid.*

¹¹²² The story is recounted in detail by Brian Christian in *The Alignment Problem*: computer programmer Jacky Alciné received a message from a friend linking him to the offending photo. Originally, he thought his friend was pointing out that the user interface had changed; he subsequently noticed that a photo of himself and a female friend had been tagged as "gorillas." He immediately took to Twitter, tweeting the photo. Christian, *supra* note 40 at 25.

faces, the selected solution was to remove the tag gorillas altogether.¹¹²³ Several years later, WIRED reported that Google’s image search had still not been rectified.¹¹²⁴ With this change to the underlying code, not even gorillas would be categorized as gorillas.

As a technology, photography has been touted as an opportunity to obtain additional information from the natural world and capture reality more precisely. Yet, throughout its ascent to near ubiquity, the inherent limitations of the technology itself have affected the types of information that can be generated through image capture. In the history of photography, cameras were accompanied by a photo of a white woman, known as a “Shirley card” in honour of the woman depicted, used to calibrate the camera’s settings to properly capture human facial features.¹¹²⁵ While this technique worked for capturing light-skinned folks, it made capturing black or brown-skinned individuals very difficult—especially if they were in a photograph alongside white people. Despite outcry from the minoritized groups impacted by the phenomenon, it continued until specific complaints were made by manufacturers of chocolate and wooden furniture, who experienced trouble creating well-contrasted photographs for advertisements of their wares.¹¹²⁶ Today, those wishing to calibrate their camera settings with a “Shirley card” will receive several Shirleys, of varying skin tones.¹¹²⁷ But even today, technological limitations still sometimes create dehumanizing experiences for people of colour. For example, in July 2022, an African Canadian woman in Manitoba reported that her driver’s licence photo had been rejected three times by the provincial public insurer.¹¹²⁸ The woman was told that her skin was registering as an “unnatural colour.”¹¹²⁹

As introduced in Chapter 5, the landmark report “Gender Shades,” published by Joy Buolamwini and Timnit Gebru emphasized how FRT is less likely to identify black faces than white ones.¹¹³⁰ FRT is also less likely to identify female faces than male ones, with black female faces being the least likely to be recognized. Buolamwini first noticed this phenomenon during graduate work on computer vision at MIT: she observed that her workstation did not register the existence of her face when she was in front of the camera.¹¹³¹ At first, she thought there might be some issue with the technology, but she then observed that the camera had no issue identifying her white colleagues. And, if Buolamwini wore a white mask, she was able to generate a response from the FRT. This observation led to considerable research in the field, with several reports being authored detailing the phenomenon, and popular news outlets running stories about facial recognition’s racial issues. Other analyses confirmed similar results: for example, efforts to use FRT in Detroit were stymied by the disproportionately large African American population, upon whom it was

¹¹²³ Christian, *supra* note 40 at 25.

¹¹²⁴ *Ibid* at 26.

¹¹²⁵ Benjamin, *supra* note 268 at 103-104.

¹¹²⁶ *Ibid* at 105. See also Christian, *supra* note 40 at 28.

¹¹²⁷ Benjamin, *supra* note 268 at 105.

¹¹²⁸ Rachel Bergen, “African Canadian woman fed up with Manitoba Public Insurance after yet another ‘dehumanizing’ photo glitch” (8 July 2022), online: *CBC* <www.cbc.ca/amp/1.6515479> [perma.cc/R7JN-J2SS].

¹¹²⁹ *Ibid*.

¹¹³⁰ Buolamwini & Gebru, *supra* note 862.

¹¹³¹ In addition to the Gender Shades report, *ibid*, this research was featured in the well-publicized documentary *Coded Bias*, a TED talk at TEDxBeaconStreet by Buolamwini, and assorted media reports. Shalini Kantayya, *Coded Bias* online: <www.codedbias.com/> [perma.cc/DKU6-X3T9]; Joy Buolamwini, “How I’m Fighting Bias in Algorithms” (9 March 2017), online (video): *TED* <www.ted.com/talks/joy_buolamwini_how_i_m_fighting_bias_in_algorithms?language=en> [perma.cc/4T7D-5WTA].

clear the technology was less accurate.¹¹³² The DataWorks software used by Michigan state police was shown to falsely identify African American and Asian faces 10-100 times more often than Caucasian faces.¹¹³³

Another unpleasant coagulation of FRT, privacy, and equality concerns arose with the arrival of “deepfakes.” The phenomenon began on the Internet community Reddit, when a user shared his process for using machine learning and face-swapping technology to create synthetic “do-it-yourself” (DIY) pornographic videos starring his favourite celebrities.¹¹³⁴ While the process originally required some technical skill, another Reddit user soon created a simple application to allow less-technically-adept users to make their own pornographic deepfakes.¹¹³⁵ The hack included the ability to scrub accessible Instagram or Facebook accounts for images to port into the program. This brings the problem of publicly available private information into sharp relief: without careful monitoring to ensure uses of information are tied to the purposes for which they are offered, ghastly downstream consequences may result. The concerns for equality are significant. In all such situations, encoded bias may have significant repercussions for racialized communities, LGBTQ* communities, women, and other equality-seeking groups.

Related issues crop up in analogous FRT efforts: recall the discussion in Chapter 3 of the algorithm that identified a person’s gender based on their facial features, flattening the nuanced reality of gender identity in modern society.¹¹³⁶ Similarly, researchers at Stanford were able to use facial recognition to predict homosexuality of test subjects with an 81% accuracy in males and 74% accuracy in females—essentially creating ML “gaydar”.¹¹³⁷ Despite these recognized limitations, FRT continues to grow in society, leading to potential Rule of Law problems.

Before turning to the Rule of Law problems, however, I expand the scope of analysis to include some other aspects of privacy law, including observations of how technological innovations pertaining to images and photography are handled in other areas of the law. The relationship between FRT and criminal law is evident, especially in situations where state actors are using technological solutions powered by AI/ML to encroach upon individual’s privacy rights. If FRT is being deployed as an emanation of the state, the Rule of Law dimensions of the issue are clear. Some analyses might stop here. However, the nexus becomes muddled when considering how FRT is produced and controlled by corporate interests, which exert power over both state actors and citizens by virtue of FRT’s affordances. If FRT usage is broader than just emanations of the state, but still exerts power in other usages, those other usages can also lead to abuses of power. If

¹¹³² Amy Harmon, “As Cameras Track Detroit’s Residents, a Debate Ensues Over Racial Bias” (8 July 2019), online: *The New York Times* <www.nytimes.com/2019/07/08/us/detroit-facial-recognition-cameras.html> [perma.cc/K8NZ-26CY].

¹¹³³ Kashmir Hill, “Wrongfully Accused by Algorithm” (24 June 2020), online: *The New York Times* <www.nytimes.com/2020/06/24/technology/facial-recognition-arrest.html> [perma.cc/E6ZT-ZXC3] [Hill, “Wrongfully Accused”].

¹¹³⁴ Adi Robertson, “Reddit bans ‘deepfakes’ AI porn communities” (7 February 2018), online: *The Verge* <www.theverge.com/2018/2/7/16982046/reddit-deepfakes-ai-celebrity-face-swap-porn-community-ban> [perma.cc/2DJF-A4XV].

¹¹³⁵ Grace Lisa Scott, “Deepfakes Still Exist, Amid Twitter and Pornhub Bans” (6 February 2018), online: *Inverse* <<https://www.inverse.com/article/41017-pornhub-says-it-s-banning-deepfakes-as-the-internet-cracks-down-on-ai-porn>> [perma.cc/3WP4-44FY].

¹¹³⁶ See full discussion in Chapter 3.

¹¹³⁷ See Sam Levin, “New AI can guess whether you’re gay or straight from a photograph” (8 September 2017), online: *The Guardian* <www.theguardian.com/technology/2017/sep/07/new-artificial-intelligence-can-tell-whether-youre-gay-or-straight-from-a-photograph> [perma.cc/GQ2H-QFJW].

abuses of power are possible, their impact engages an over-arching interest in the Rule of Law. If corporate actors are constrained by legal regulations, those typically arise in private law or statutory remedies, rather than only constitutional ones. Remembering Chapter 2's lessons on technological affordances, I aim to enter a meaningful dialogue with power, wherever it is concentrated. I begin with some reflections on the evolution of legal privacy rights in tandem with technological evolution, before cataloguing the various areas of Canadian law where privacy protection exists.

2. Privacy Law's Legacy & Technological Pinch Points

Today's privacy concerns often arise from informational transactions: people's digital interactions can generate significant stores of information about each data subject. Often, concerns arise from the many unknowns associated with this data exhaust: how will it be used, stored, archived, and/or aggregated? What insights can be generated from its existence, processing, or combination with other publicly available information? How will individuals' personal data be used to further the AI industries' capital gains? How might it come back to haunt the data subject in the future? To answer these questions, one might turn to privacy law, consisting of constitutional protections against unreasonable search and seizure; statutory protections enshrining fair information principles for the collection, use, and dissemination of personal information; tort law remedies; and individual corporate privacy policies and guarantees. Despite the patchwork of legal privacy protections available, satisfactory answers to these questions are not often forthcoming.

Modern legal privacy thinking or scholarship is often thought to originate with the 1890 publication of "The Right to Privacy" in the *Harvard Law Review*, authored by two of the era's great jurists: Samuel Warren and Louis Brandeis.¹¹³⁸ Frustrated with the unavailability of legal remedies against newspapers for publishing authorized photographs and sordid gossip, Warren and Brandeis charted a course for law's development that would expand the common law to protect against the press "overstepping in every direction the obvious bounds of propriety and of decency."¹¹³⁹ They presented this extension of the law as its next logical step: after all, the law had previously expanded to encompass a broader scope for civil liberties and evolved to recognize property as both tangible and intangible. Explaining their rationale in the famed paper, they write:

Recent inventions and business methods call attention to the next step which must be taken for the protection of the person, and for securing to the individual what Judge Cooley calls the right "to be let alone." Instantaneous photographs and newspaper enterprise have invaded the sacred precincts of private and domestic life; and numerous mechanical devices threaten to make good the prediction that "what is whispered in the closet shall be proclaimed from the housetops." For years there has been a feeling that the law must afford some remedy for the unauthorized circulation of portraits of private persons; and the evil of the invasion of privacy by the newspapers, long keenly felt, has been but recently discussed by an able writer.¹¹⁴⁰

¹¹³⁸ Samuel D Warren & Louis D Brandeis, "The Right to Privacy" (1890) 4 *Harv L Rev* 193.

¹¹³⁹ *Ibid* at 196.

¹¹⁴⁰ *Ibid* at 195.

Modern readers will immediately notice the similarities between Warren and Brandeis' concerns and the concerns generated by leading an online life in the digital era, especially given FRT. "Gossip is no longer the resource of the idle and of the vicious," they write, "but has become a trade, which is pursued with industry as well as effrontery."¹¹⁴¹ Warren and Brandeis objected to business practices that commodified salacious details of peoples' private lives, reserving their greatest ire for publication or copying that was not captured by the intellectual property statute of copyright. Although copyright protects against copying the exact series of words fixed on a page, it does not prevent against revealing that certain works exist or reproducing some of the details contained therein—albeit fixing the words in slightly different orders. The law might protect private letters as literary compositions, but nothing in copyright stops newspapers from enumerating the letters that exist or reformulating the key facts they express.¹¹⁴² Even if it is less than honourable, reproducing gossip circulating in the zeitgeist does not violate copyright law.

Warren and Brandeis used the illustrative case of *Prince Albert v Strange* to explain the privacy concerns that call into question the limited protection copyright law offers.¹¹⁴³ The plaintiff, Prince Albert, was the husband and consort of Queen Victoria. In the 1840s, he and his wife both enjoyed the laborious pastime of etching pastoral scenes of royal family life and their children at play onto copper plates. Prince Albert provided the copper etchings to a local printer to generate some copies for his private use. While the printer discharged his duties appropriately, an opportunistic employee, Mr. Middleton, made illegitimate copies of the etchings for distribution to local newspapers, which were then published by Mr. Strange. Prince Albert filed suit for the surrender of the etchings, prohibition of their exhibition, and prevention of publication of the etchings. Ultimately, the plea was granted: the court recognized the degree to which "privacy is a part, and an essential part, of this species of property."¹¹⁴⁴ Absent the existence of explicit privacy protections in law, the court relied on property interests, noting also that the copies were either dishonestly obtained or improperly retained.¹¹⁴⁵

In the fifty years between Prince Albert's lawsuit and Warren and Brandeis' publication, the technology for capturing pastoral family life or children at play had evolved significantly. In the place of etchings onto copper plates, it became possible to take an instantaneous photograph of events transpiring in real time for publication in the daily newspaper. This advancement surely seemed almost magical at the time, particularly given the seismic change from the era of the engraved etchings. For Warren and Brandeis, it was the observation that this technological pinch point would transform the legal landscape that ultimately gave rise to law's specific acknowledgement that privacy warranted legal protection.

In the years since, privacy law doctrine has been developed piecemeal, with different areas of the law being called upon to house different types of privacy protection. Commentators describe privacy law with full acknowledgement of its limitations. For example, former SCC Justice Ian Binnie once described privacy as "protean," referring to Proteus, the shapeshifting Greek sea god

¹¹⁴¹ Warren & Brandeis, *supra* note 1138 at 196.

¹¹⁴² *Ibid* at 201.

¹¹⁴³ *Prince Albert v Strange* [1849], EWHC Ch J20; 41 ER 1171.

¹¹⁴⁴ *Ibid*.

¹¹⁴⁵ *Ibid*.

able to transform at will.¹¹⁴⁶ Novelist Jonathan Franzen called privacy “the Cheshire cat of values: not much substance, but a very winning smile.”¹¹⁴⁷ While a full account of the state of privacy law is not necessary for this thesis, the following sections quickly summarize the key areas of legal protection, particularly as they pertain to the regulation of images or photography. I emphasize legal protections relating to publicly available or indirectly collected information. These protections emerge from the areas of (a) criminal law; (b) tort law; and (c) statutory law. In addition to these formal legal frameworks, it is important to note that privacy protection efforts are often delegated to corporate privacy policies, which differ by company and by technology, creating an especially confusing web for consumers to navigate.

a. Criminal Law Dimensions

Over time, as technology has improved and become more commonplace, civil liberties and constitutional rights have been engaged more acutely by technological systems. Different types of image-based infractions have become possible, challenging previous notions that entrenched much stronger protections for the privacy of the home as compared with other scenarios. Since the advent of the *Charter of Rights and Freedoms*, judicial considerations of section 8 (the right to be free from unreasonable search and seizure) have animated Canadian privacy law considerations. The nexus between technological advancements and privacy rights have been fraught, as technology continues to expand the parts of the world that can be monitored.¹¹⁴⁸ Yet, judicial analysis of what types of interferences were reasonable continued to be confounded by subjective understanding. New technologies able to map previously private aspects of one’s life can quickly become commonplace, transforming one’s expectations over whether a particular search is considered “reasonable,” or at the very least, making the corresponding reasonableness calculus more difficult.

This reasonableness calculus was upended in the notable case of *R v Tessling*, where the SCC found there was no section 8 violation stemming from the use of a thermal energy mapping camera.¹¹⁴⁹ Because the heat signature of one’s home is not something people typically maintain a privacy interest over, the court held that the advent of technology that mapped that information could not legally constitute an unreasonable search.¹¹⁵⁰ Given the degree of reliance on “publicly available information” for the compilation of FRT databases, it is arguable that one might not retain a reasonable expectation of privacy in the required source images. Privacy advocates continued to hope that technological change might mediate the reasonableness of one’s expectations in changing contexts. Emerging case law continued to add complexity to the privacy-protection fabric: for example, *R v Spencer* confirmed privacy in Internet usage could include a

¹¹⁴⁶ *R v Tessling*, 2004 SCC 67 at para 25 [*Tessling*].

¹¹⁴⁷ Jonathan Franzen, “Imperial Bedroom” (October 5, 1998), online: *The New Yorker* <www.newyorker.com/magazine/1998/10/12/imperial-bedroom> [perma.cc/S8ZQ-QYAX].

¹¹⁴⁸ In *R v Tessling*, Binnie J links the roots of privacy law to private property. He writes: “[t]he courts were reluctant to accept the idea that, as technology developed, the sphere of protection for private life must shrink. Instead, it was recognized that the rights of private property were to some extent a proxy for the privacy that ownership of property originally conferred, and therefore, as the state’s technical capacity for peeking and snooping increased, the idea of a protected sphere of privacy was refined and developed.” *Tessling*, *supra* note 1146 at para 16.

¹¹⁴⁹ The camera in question in *Tessling* was Forward Looking Infra-Red (FLIR) technology, which could record the heat signature emanating from an individual’s home. *Ibid* at para 2.

¹¹⁵⁰ Binnie J identified three spheres of privacy protection based on the previous section 8 case law: personal, territorial, and informational. The information divined by FLIR granted information about the home but not details of the home. The analysis fixed on the idea of “core biographical data” and found the technology could not reveal such information. *Ibid* at paras 20, 59-62.

degree of anonymity;¹¹⁵¹ while *R v Marakah* offered a reasonable expectation of privacy in sent text messages on a recipient's phone.¹¹⁵² Hope of a truly contextual approach to privacy harm was realized in the SCC's recent decision of *R v Jarvis*.¹¹⁵³

i. *R v Jarvis*

The SCC's decision in *R v Jarvis* has bolstered legal observers' confidence in the possibility of contextual analysis for legal accounts of privacy. In *Jarvis*, the accused was a high school teacher who surreptitiously recorded the faces and upper bodies of his female students with a camera pen. He was charged with voyeurism under provision s 162(1)(c) of the Criminal Code.¹¹⁵⁴ The provision required that the recordings be both for a sexual purpose and be made of a person who is in circumstances that give rise to a reasonable expectation of privacy.¹¹⁵⁵

The wording of the voyeurism provision created issues for the lower courts. At trial, the judge found that the circumstances gave rise to a reasonable expectation of privacy but was unable to confirm that the recordings were for a sexual purpose.¹¹⁵⁶ On appeal, the Ontario Court of Appeal (ONCA) reversed both findings. They were satisfied that the recordings—which focused on the female students' breasts—were for a sexual purpose, but were unconvinced that the students could reasonably expect privacy in a classroom setting. In ONCA's view, people have diminished expectations of personal privacy when moving about in public space. Therefore, teachers should refrain from recording their students not because of a legally protected privacy right, but due to the “nature of the required relationship between students and teachers.”¹¹⁵⁷ This holding alarmed privacy and child protection advocates alike: surely, students should be entitled to expect privacy from the prying eyes of a schoolteacher equipped with a camera pen. The SCC agreed.

Through *Jarvis*, the SCC has signalled its willingness to update the scoping of certain legal rights in light of technological innovation. Although the court was dealing with a specific *Criminal Code* provision instead of a broader section 8 analysis, the usage of the phrase “reasonable expectation of privacy” in s 162(1) indicated that section 8 jurisprudence was nonetheless relevant. The SCC rejected ONCA's location-based approach, instead focusing on a contextual, circumstantial approach. Importantly, the majority stressed that privacy is “not an all-or-nothing concept.”¹¹⁵⁸ While traditionally fortified locations like the home might continue to attract the strongest privacy interests, a woman does not forfeit her privacy interest “simply because she is in a place where she

¹¹⁵¹ *R v Spencer*, 2014 SCC 43.

¹¹⁵² *R v Marakah*, 2017 SCC 59.

¹¹⁵³ *R v Jarvis*, 2019 SCC 10 [*Jarvis*].

¹¹⁵⁴ The full language of the s 162(1)(c) is as follows: 162 (1) Everyone commits an offence who, surreptitiously, observes — including by mechanical or electronic means — or makes a visual recording of a person who is in circumstances that give rise to a reasonable expectation of privacy, if (c) the observation or recording is done for a sexual purpose. *Criminal Code*, RSC 1985, c C-46, s 162(1)(c).

¹¹⁵⁵ At trial, Mr. Jarvis conceded that the recordings were produced surreptitiously, so this did not form a central part of the inquiry. However, s 162(1)(c) makes surreptitiousness an explicit element of the offence.

¹¹⁵⁶ Even though the recordings focused on the students' breasts, the judge did not think this met the criminal liability threshold of beyond a reasonable doubt.

¹¹⁵⁷ *R v Jarvis*, 2017 ONCA 778 at para 105.

¹¹⁵⁸ *Ibid* at para 41.

knows she can be observed by others or from which she cannot exclude others.”¹¹⁵⁹ This nuanced view rejects a strict binary perspective on what is public *versus* what is private.

The majority in *Jarvis* signalled a move to a contextual approach to resolving privacy violations by allowing section 8 jurisprudence to inform the determination of the voyeurism provision.¹¹⁶⁰ As in *Alberta (Information and Privacy Commissioner) v United Food and Commercial Workers, Local 401*, this acknowledges the nuance for information that attracts a privacy interest despite its presence in the public sphere.¹¹⁶¹

b. Tort Law Dimensions

Other aspects of legal privacy have also been pushed into actuality by virtue of such technological pinch points. Another notable development arose in the 1960s, with William Prosser’s publication of the article “Privacy,” which laid out four distinct privacy torts.¹¹⁶² All four were eventually imported into Canadian law: intrusion upon seclusion, appropriation of personality, false light, and public disclosure of private facts.¹¹⁶³ Only the latter three torts are relevant to the consideration of publicly available information; these are presented here.¹¹⁶⁴ Given how FRT flows from corporate entities, the availability of tort law remedies might prove critical to generate appropriate checks and balances, something Kristen Thomasen discusses in her doctoral thesis.¹¹⁶⁵ Moreover, complicating the nexus between private law and public law is a classic feminist approach, worthy of inclusion here to solidify the impacts on human actors in both spheres.¹¹⁶⁶

i. Appropriation of Personality

The tort of appropriation of personality protects against intentional exploitation of one’s likeness for commercial purposes. This tort usually appears in relation to celebrity promotion of products, but it can apply to ordinary people too. It was first recognized in Canada in *Krouse v Chrysler Canada*.¹¹⁶⁷ Canadian Football League (CFL) player Bob Krouse sued after his likeness was used without his consent in a promotional spot by Chrysler. While the tort was recognized as a cause of action, Krouse did not succeed in his claim: the impugned image showed only his silhouette and

¹¹⁵⁹ *Jarvis*, *supra* note 1153 at para 37.

¹¹⁶⁰ Counsel for the respondent argued that. In an interview following the release of the decision, counsel worried that See Elizabeth Raymer, “Right to privacy not an all-or-nothing concept, says SCC in voyeurism case” (14 February 2019) online: *Canadian Lawyer* <www.canadianlawyermag.com/practice-areas/privacy-and-data/right-to-privacy-not-an-all-or-nothing-concept-says-scc-in-voyeurism-case/275878> [perma.cc/8NEJ-S55B].

¹¹⁶¹ *United Food*, *supra* note 1087. At issue was the videotaping and photographing of individuals crossing the picket line during lawful strike action—which lasted 305 days. While explicitly noting that the applicable scope of Alberta’s act was broader than that of *PIPEDA*, the SCC’s analysis confirmed that the mere act of appearing in public does not mean an individual forfeits control of the personal information that is correspondingly exposed.

¹¹⁶² William L Prosser, “Privacy” (1960) 48 *California Law Review* 383.

¹¹⁶³ *Ibid* at 385.

¹¹⁶⁴ Thomasen notes in her doctoral thesis that while intrusion upon seclusion does not generally apply to torts occurring in public space, this outcome may be because of the court’s narrow interpretation in the first case to recognize it, *Jones v Tsige*, 2012 ONCA 32. She notes that two recent cases might broaden this context: (i) *Vertolli v YouTube LLC*, 2012 Can LII 99832 (ON SCSM), where a defendant filmed the plaintiff police officer pulling him over and posted it to YouTube, who refused to take it down per YouTube’s privacy policy; and (ii) *Vanderveen v Waterbridge Media Inc.*, where the defendant featured a video of a plaintiff jogging for 2 seconds without consent. See Thomasen, *supra* note 1040 at 98-101.

¹¹⁶⁵ *Ibid* at 4.

¹¹⁶⁶ See, e.g., West, *supra* note 703 at 45.

¹¹⁶⁷ *Krouse v Chrysler Canada Ltd et al*, 1973 CanLII 574 (ONCA).

jersey number.¹¹⁶⁸ The court found that Chrysler was promoting both the game of football and the CFL as an institution, as opposed to using Krouse personally to sell its cars.

Despite the failure for Krouse personally, this tort became part of Canada's common law fabric and some provincial statutory schemes.¹¹⁶⁹ In an online setting where faces, likenesses, and perhaps even personalities can be retrieved without consent, this tort may offer a framework for relief. This legal remedy may well be used to respond to issues of online reputation. Suzie Dunn argues that this tort may provide remedies for online identity transgressions such as deepfakes or synthetic media, but only if used in commercial settings.¹¹⁷⁰ According to Dunn, this approach leaves a gap in Canadian law for those whose identities have been harmed in non-commercial technological settings.¹¹⁷¹

ii. False Light

The tort of publicity which places someone in a false light was first recognized in Canada in the 2019 case of *Yenovkian v Gulian*.¹¹⁷² This fraught family law case dealt with digital misrepresentations, online harassment, and surreptitious recording in the context of a bitter custody dispute. It established that falsely representing someone constituted a tort: the legal wrong adheres in representing someone as other than they are. False light is therefore distinct from defamation: the person does not necessarily need to be depicted as worse than they are to meet the tort elements. The privacy value, the court noted, was found in "a person's right to control the way they present themselves to the world."¹¹⁷³ Although this cause of action is newly established (and, so far available only in Ontario), it may provide protection for the spread of online disinformation about plaintiffs.

iii. Public Disclosure of Private Facts

The tort of public disclosure of embarrassing private facts was first recognized in Canada in the 2016 case *Jane Doe 464533 v ND*.¹¹⁷⁴ It dealt with an intimate video recorded by a teenage girl, which was surreptitiously uploaded to web service by her boyfriend.¹¹⁷⁵ Surveying the possible privacy torts, the court found it fit more appropriately into public disclosure of private facts than other categories (such as intrusion upon seclusion) and chose to recognize this tort for the first time. The plaintiff was also successful in tort claims for breach of confidence and intentional infliction of mental distress. However, the defendant successfully moved to have the decision set aside, offering a defence on the merits.¹¹⁷⁶ The appellate court did not interfere with the newly

¹¹⁶⁸ *Krouse*, *supra* note 1167.

¹¹⁶⁹ Note that tort law is within provincial jurisdiction, so availability of the tort varies across provinces. To date, it has been recognized as a common law tort in Nova Scotia, New Brunswick, Ontario, and Alberta. In British Columbia, Saskatchewan, Manitoba, and Newfoundland & Labrador, it is protected by privacy legislation. McIsaac, Shields & Klein, eds, *The Law of Privacy in Canada* (Toronto: Thomson Reuters, 2015) at 2-58.78 (2.4.4. Legislated Cause of Action for Breach of Privacy).

¹¹⁷⁰ "Deepfakes," or artificial videos, are discussed further below.

¹¹⁷¹ Suzie Dunn, "Identity Manipulation: Responding to advances in artificial intelligence and robotics" (16 July 2020), We Robot 2020 Conference Paper, online (pdf): <drive.google.com/file/d/1LXCxTEj7zHbFr8V3SdxKxnut173aCGsf/view> [perma.cc/N24S-DRJ3].

¹¹⁷² *Yenovkian v Gulian*, 2019 ONSC 7279.

¹¹⁷³ *Ibid* at para 171.

¹¹⁷⁴ *Jane Doe 464533 v ND*, 2016 ONSC 541.

¹¹⁷⁵ *Ibid* at 7-8.

¹¹⁷⁶ *Jane Doe 464533 v ND*, 2017 ONSC 127.

established tort; it was later recognized again in *Jane Doe 72511 v Morgan*, with similar reasoning.¹¹⁷⁷

Ultimately, however, successfully proving the elements of this tort will require a strict delineation between public and private information. The elements of the tort—as theorized by Prosser and confirmed by the Ontario Superior Court of Justice—are (a) the defendant publicized an aspect of the plaintiff’s private life; (b) the plaintiff did not provide consent to the publication; (c) the matter publicized would be highly offensive to a reasonable person; and (d) the publication was not of legitimate concern to the public.¹¹⁷⁸ The very nature of the tort therefore suggests an absence of privacy in public. In the test itself, similar terms “publicized”, “publication”, and “the public” are each used to mean slightly different things. While segregating the concepts of public and private in tort may not raise immediate alarm bells for data protection, deferring to this idea in another branch of legal analysis may negatively impact efforts to define broader protection for publicly available information, like the information that comprises FRT. By diluting the differentiation between these concepts, the potential protection for one’s image or face—the most publicly facing aspect of the self—is compromised.

c. Statutory Dimensions

In Canada, the primary statutory vehicle for the protection of consumer privacy is *PIPEDA*, which applies to all businesses that collect, use, and/or disclose personal information.¹¹⁷⁹ As discussed above, it was violation of *PIPEDA* that led to the OPC’s investigation of Clearview AI. Other statutes are specifically directed to the federal government’s use of personal information,¹¹⁸⁰ and the collection of personal health information within the provinces.¹¹⁸¹ Some, but not all, provinces have additional statutory schemes.¹¹⁸² Regardless, when new technological measures are implemented, the legal regime typically implicated is *PIPEDA*, since its precepts are intended to regulate the corresponding information ecosystem within the corporate sphere. Since the rise of the Internet (and particularly social media), many of the protections in *PIPEDA* have proven woefully inadequate. *PIPEDA* was granted royal assent in 2000, meaning that it was researched, drafted, and negotiated in the late 1990s.¹¹⁸³ Despite attempts to create a technologically neutral framework, certain shortcomings persist, exacerbated by the passage of time and the changing technological landscape.

The *PIPEDA* framework requires individual consent. Principle 4.3 of Schedule I establishes that “knowledge and consent of the individual are required for the collection, use, or disclosure of

¹¹⁷⁷ *Jane Doe 72511 v Morgan*, 2018 ONSC 6607.

¹¹⁷⁸ *Ibid* at para 99.

¹¹⁷⁹ *PIPEDA*, *supra* note 1037.

¹¹⁸⁰ *Privacy Act*, *supra* note 1037.

¹¹⁸¹ Since healthcare is a provincial responsibility, each province has its own statutory scheme for handling personal health information. See, e.g., *Personal Health Information Protection Act, 2004*, SO 2004, c 3, Sched A.

¹¹⁸² Both Alberta and British Columbia have provincial statutes, which offer substantially similar protections to *PIPEDA*. *Personal Information Protection Act*, SBC 2003, C 63; *Personal Information Protection Act*, SOA 2003, C P-6.5. The most robust provincial privacy laws exist in Quebec: *Act Respecting the Protection of Personal Information in the Private Sector*, 1993 P-39.1. Quebec’s statute was recently amended with the widely publicized Bill 64, which claimed to modernize Quebec’s privacy law for the information age: *Bill 64: An Act to modernize legislative provisions as regards the protection of personal information*, (2020) National Assembly of Quebec, First Session, 42nd Legislature.

¹¹⁸³ Government of Canada, Regulatory Impact Analysis Statement, Canada Gazette Part II, Vol 135, No 1 (3 January 2001), online (pdf): *Government of Canada* <gazette.gc.ca/rp-pr/p2/2001/2001-01-03/pdf/g2-13501.pdf>.

personal information, except where inappropriate.”¹¹⁸⁴ Despite the centrality of consent within the *PIPEDA* framework, statutory amendments over the years have watered down its importance through adopting exceptions to when consent is required.¹¹⁸⁵ Concurrently, *PIPEDA* also requires a reasonableness analysis, defined in section 5(3) of the statute.¹¹⁸⁶ The standard for reasonableness under section 5(3) was set in *Eastmond v Canadian Pacific Railways*.¹¹⁸⁷ Among other criteria,¹¹⁸⁸ this test asks whether a privacy-invasive measure effectively meets a specific need and, ultimately, whether a minimally-impairing option could have been used instead. Such analysis is inescapably subjective: what a reasonable person will consider appropriate in the circumstances is irrevocably governed by the world in which they live and the technology that exists in that world. And, as surveillance technologies continue to improve and enable the extension of state power, arguments about what constitutes “minimally-impairing” are mitigated by advancements of the technologies themselves.

Globally, the recent adoption of other statutes has highlighted *PIPEDA*'s shortcomings for the digital age. Particularly since the European Union's *General Data Protection Regulation (GDPR)* came into force in May 2018, Europe has been considered a world leader in privacy protection. The EU's protection of privacy is an omnibus approach: it imposes standards across sectors, instead of having specific sector-based legislation. Compared with *PIPEDA*, the *GDPR* offers a more comprehensive scope of application, regulating any companies that process “personal information,” which is defined as any information relating to an identified or identifiable natural person.¹¹⁸⁹ This definition is broader than *PIPEDA* because it does not exclude business contact information. Further, consent is conceptualized differently: while *PIPEDA* treats consent as central to the collection, use, or disclosure of personal information, the *GDPR* treats consent as only one of many legal grounds necessary for processing personal data.¹¹⁹⁰ Article 14 of the *GDPR* explicitly notes that data controllers dealing with indirectly collected personal information must take a number of steps to inform the data subject of the collection, including: providing the contact information of the controller/data protection officer; the purposes of the processing and legal basis for processing; the categories of information concerned; any recipients of personal data.¹¹⁹¹ Article 17 guarantees the right to erasure, or what is colloquially known as the right to be forgotten.¹¹⁹²

¹¹⁸⁴ *PIPEDA*, *supra* note 1037 at Schedule 1, 4.3.

¹¹⁸⁵ The most recent round of amendments in 2015 saw the adoption of ten new exceptions to consent. Teresa Scassa cautions that this growing number of exceptions to consent within *PIPEDA* is a cause for concern. Teresa Scassa, “Open Season on Social Media Data: (ETHI's Report on PIPEDA Reform – Part II)” (22 March 2018), online: *Teresa Scassa* <www.teresascassa.ca/index.php?option=com_k2&view=item&id=273:open-season-on-social-media-data-ethis-report-on-pipeda-reform-part-ii&Itemid=80> [perma.cc/H879-B38E].

¹¹⁸⁶ The full text of s 5(3) reads: “An organization may collect, use or disclose personal information only for purposes that a reasonable person would consider are appropriate in the circumstances.” *PIPEDA*, *supra* note 1037 s 5(3).

¹¹⁸⁷ *Eastmond v Canadian Pacific Railways*, 2004 FC 852 at para 127 [*Eastmond*].

¹¹⁸⁸ The test for reasonableness, per *Eastmond* is as follows: (1) Is the measure demonstrably necessary to meet a specific need? (2) Is it likely to be effective at meeting that need? (3) Is the loss of privacy proportional to the benefit gained? (4) Is there a less-privacy invasive way of achieving the same end? *Ibid*.

¹¹⁸⁹ The *GDPR* defines personal data as “any information relating to an identified or identifiable natural person (data subject)”. *GDPR*, *supra* note 496.

¹¹⁹⁰ *GDPR* Article 7(1) states: “Where processing is based on consent, the controller shall be able to demonstrate that the data subject has consented to processing of his or her personal data.” *Ibid*, Art 7.

¹¹⁹¹ *GDPR*, *supra* note 496, Art 14(1).

¹¹⁹² *Ibid*, Art 17.

Another significant move in the privacy statute space was the enactment of the *California Consumer Protection Act (CCPA)*, which came into force 1 January 2020, enforcement commencing 1 July 2020.¹¹⁹³ Widely touted as California’s answer to the *GDPR*, the *CCPA* contains similar provisions, but ultimately has a narrower approach to data protection. In the *CCPA*, personal information is “information that identifies, relates to, describes, is capable of being associated with, or could reasonably be linked, directly or indirectly, with a particular consumer or household.”¹¹⁹⁴ Moreover, the *CCPA* applies only to large companies above a certain threshold; smaller data controllers are therefore able to avoid its mandates.¹¹⁹⁵ Several sections of the *CCPA* combine to offer similar protections to *GDPR* Article 14. Individuals are entitled to information about the categories of personal information that are collected and the purposes for which they will be used.¹¹⁹⁶ The same clause prohibits collection for any additional purposes unless express notice is provided. The *CCPA* also entrenches a right to request information that a business has collected,¹¹⁹⁷ an express right of deletion,¹¹⁹⁸ and an accessible procedure for individuals to contact businesses and access their information rights.¹¹⁹⁹

With these legal perspectives in place, I move on to characterizing FRT’s Rule of Law problems.

3. Rule of Law Problems

It was immediately clear how COMPAS sentencing software, discussed in Chapter 6, led to Rule of Law problems. Here was a specific area of the law, explicitly codified in the *Criminal Code* and associated with numerous entrenched constitutional rights, that was being delegated to an external system and intermingled with corporate interests. Yet, at the same time, there was a recognition that this area of the law was intended to be highly discretionary, incorporating human judgment (and all its frailties) as a mandatory aspect of a process deeply associated with exercising legal authority. In this space, where AI/ML technological approaches were used to resolve the persistent tension of law’s potential futures, Rule of Law problems presented themselves in the biasing of datasets, the distortion of values, and the conflation of what we ought to do with what we ought to believe. This analysis revealed constitutional dimensions, but also epistemological concerns about the accused’s ability to distinguish their own unique life experience from what might be suggested by the statistical mean—and defend themselves accordingly.

Another analysis could confront issues presented by FRT directly through a criminal or constitutional law lens. In a 2020 report, the Criminal Courts Committee of the New York City Bar Association identified a series of Rule of Law-related issues created by FRT.¹²⁰⁰ These include

¹¹⁹³ *CCPA*, *supra* note 1038.

¹¹⁹⁴ *Ibid.*, §1798.140(o)(1). Notably, the definition exempts information that is deidentified, as well as “aggregate consumer information.” §1798.140(o)(3).

¹¹⁹⁵ For a business to be covered by the *CCPA*, it must collect personal information of California residents and have one of: (a) annual gross revenues in excess of \$25 million; (b) annually buys, receives, sells, or shares for commercial purposes, alone or in combination, the personal information of 50,000 or more California residents, households, or devices; or (c) derive 50 per cent or more of its annual revenues from selling residents’ personal information. *Ibid.*, §1798.140(c).

¹¹⁹⁶ *Ibid.*, §1798.100(b).

¹¹⁹⁷ *Ibid.*, §1798.110(a).

¹¹⁹⁸ *Ibid.*, §1798.105.

¹¹⁹⁹ *Ibid.*, §1798.130(a); §1798.135.

¹²⁰⁰ New York City Bar Association, “Power, Pervasiveness and Potential: The Brave New World of Facial Recognition Through a Criminal Law Lens (and Beyond)” (14 August 2020), online (pdf): *New York City Bar* <s3.amazonaws.com/documents.nycbar.org/files/2020662-BiometricsWhitePaper.pdf> [perma.cc/QEW3-ETWX].

issues with search and seizure, due process, and self-incrimination, among others. Similarly, the authors of the Perpetual Line-up Report note the challenges with free speech rights and due process rights when FRT is deployed by police forces.¹²⁰¹ While these issues certainly have privacy dimensions, my objective is to take a broad, theoretical approach to privacy's legal aspects that incorporates both public and private law dimensions.

Approaching privacy law in this way makes the nexus is slightly murkier. Privacy's patchwork legal status does not reveal a coherent body of laws that the technology might subvert or doctrines that it might philosophically oppose. As Andrew Ferguson points out: "current doctrine and constitutional theory offer little privacy protection and less practical security than one might expect."¹²⁰² Although Ferguson is referring to the legal landscape in the United States, the Canadian legal framework faces substantially similar shortcomings. Despite these limitations, thinking through the possible uses of advanced FRT engages various legal ideologies. Such uses challenge ideas about behaviour in public space, generating new possibilities for being identified, having your privacy encroached upon, and engaging constitutional dimensions contrary to the protections offered by section 8 of the *Charter of Rights and Freedoms*. Imagine if FRT were routinely used to identify suspects in criminal law courtrooms? Even earlier in the identification process, the work of police being motivated by predictive FRT tools generates fear and the risk of over-policing certain areas. The bias revealed by various equality-seeking critiques points to shortcomings in FRT's technological approach, the datasets used in the image-identification process, and the way that the data is collected. If FRT became even more widespread, what might the consequences be for the Rule of Law? The following analysis explores these considerations.

a. Easy Answers to Hard Cases: What Process Dominates?

More than 130 years after Warren and Brandeis' seminal article, privacy as a legal idea still struggles to achieve a meaningful foundation in law. As canvassed above, protection accrues from different legal areas; only some privacy violations will yield straightforward legal responses. And, only some of them involve AI in the legal response, as opposed to being the thing that prompted a (human-centered) legal response. Yet, by some measures, privacy's legal protections seem robust: they have grown by leaps and bounds in recent decades, with significant statutory reforms including the *GDPR* and the *CCPA*. Ari Waldman describes the obvious dichotomy between robust statutory protection and sparse enforcement as privacy's "false promise," arguing that privacy law exhibits a "legal endogeneity," where mere lip service is paid to furnish compliance with existing rules.¹²⁰³ This dichotomy, in turn, leaves a regulatory gap where symbolic gestures stand in for meaningful privacy protections. Explaining this position, he writes:

But the law's veneer of protection is hiding the fact that it is built on a house of cards. Privacy law is failing to deliver its promised protections in part because the corporate practice of privacy reconceptualizes adherence to privacy law as a compliance, rather than a substantive, task. Corporate privacy practices today are, to use Julie Cohen's term, managerial. They prioritize innovation over regulation,

¹²⁰¹ Garvie et al, *supra* note 1113.

¹²⁰² Ferguson, *supra* note 1043 at 1108.

¹²⁰³ Ari Ezra Waldman, "Privacy Law's False Promise" (2020) 97 Washington University Law Review 773 at 776.

efficiency over social welfare, and paperwork over substance. They also rely on new technologies to automate legal decisions.¹²⁰⁴

Waldman's (really, Cohen's) use of "managerial" evokes Fuller's use of the term to describe top-down examples of law, discussed in Chapter 4. Managerial conceptions of law are unidirectional, directing individuals to obey without engaging with the world-building sense of law. Fuller argues managerial conceptions are unhelpful for making sense of why people obey the law, since they undercut the sense of reciprocity between the governing force and the governed needed to make law operate.¹²⁰⁵ When privacy protection is operationalized through corporate policies, which are in turn rendered as mere compliance checkpoints rather than truly substantive social or legal tasks, it is unsurprising that the social force needed to legitimize legal authority is left wanting. This insufficiency is evident even through salutary statutes like the *GDPR* and *CCPA*, which put the lion's share of the compliance effort on the responsible corporations, who in turn pass that responsibility onto consumers through rote online communiqués about privacy preferences.

The Clearview AI inquiry demonstrated the unreliability of corporate perspectives on their own company's products. In their submissions to the OPC, Clearview argued that "any harm a person would suffer from a Clearview search of their image is comparable to the harm that a person suffers when a Google search of his or her name is performed."¹²⁰⁶ Taking this position discounts the decontextualization of the information scraped by Clearview: its collection separated it from any intended purpose for which it might have been uploaded. Their webcrawler was not the first time that information taken out of context was pieced together with unexpected results. For example, consider the 2018 Cambridge Analytica scandal, which crystallized many of the fears that the privacy community held about ML and social media.¹²⁰⁷ So-called election consultancy Cambridge Analytica stripped data from at least 87 million Facebook accounts to assist with targeted advertisements and electioneering on behalf of the Trump campaign in the 2016 presidential election in the United States. While the public was shocked by the whistleblower's revelations about the sheer volume of information collected and stored, these revelations were likely only the tip of the iceberg: most early Facebook applications possessed the functionality to provide not only a user's data, but also all of that user's friends' data, to third parties.¹²⁰⁸ While the user downloading the application may click through a series of consent pages, their friends certainly did not. Moreover, the consent provided is hollowed through the ability of the data to reveal, in aggregate, additional information external to the original provision of consent.¹²⁰⁹ Individuals tend to be cautious and exacting with personal health data, but the types of profiles enabled through Internet data sets might include health information, despite personal inclinations.¹²¹⁰

¹²⁰⁴ Waldman, *supra* note 1203 at 776.

¹²⁰⁵ Fuller, *Morality*, *supra* note 663 at 207-209. See full discussion in Chapter 4.

¹²⁰⁶ OPC, "Clearview Investigation," *supra* note 1082.

¹²⁰⁷ Carole Cadwalladr, "'I made Steve Bannon's psychological warfare tool': meet the data war whistleblower," (18 March 2018), online: *The Guardian* <www.theguardian.com/news/2018/mar/17/data-war-whistleblower-christopher-wylie-faceook-nix-bannon-trump> [perma.cc/WPR7-4VA3].

¹²⁰⁸ Professor Zeynep Tufekci, a professor of information science at University of North Carolina (Chapel Hill) made this argument. Here&Now Podcast, "'Hard To Exaggerate' How 'Data-Hungry' Facebook Is, One Professor Says" (5 April 2018), online: *WBUR* <www.wbur.org/hereandnow/2018/04/05/facebook-data-surveillance> [perma.cc/YQ84-LV46].

¹²⁰⁹ Edwards & Veale, *supra* note 161 at 23.

¹²¹⁰ *Ibid.*

This *de facto* substitution of corporate power for legal regulation underscores the connection that data makes between the personal and the political: between corporate interests and individual practices. Data protection regimes traditionally required that data remain tied to the purpose for which it was collected. As data is dissociated from its original context, it can transcend boundaries, creating data streams that neglect the delineated legal lanes set out for their initial consideration. Recall Clearview’s emphasis that its algorithm created no meaningful harms, since even leaking its whole database would merely leak that which was already available online. This view neglects the context-sensitivity that is inherent when an image is uploaded to the web as part of a news article, a blog post, or a photo album. Moreover, the focus on the individual consent of the person involved neglects to consider the impacts when the data is viewed as a system. One might have consented to post their vacation photos for the viewing enjoyment of friends and family, but not to train a powerful, webcrawling algorithm that will later be used by law enforcement to identify criminals, with poorer results for minoritized communities.

Even using humans as a “backstop” to decide if the match produced by FRT is accurate doesn’t necessarily guarantee success. The Perpetual Lineup Report noted that without specialized training, humans asked to verify the results get it wrong about 50% of the time.¹²¹¹ Of the many FRT systems studied for production of the report, only eight systems had explicit methodology where trained personnel reviewed potential matches.¹²¹² Most systems were simply not audited for misuse, which is especially problematic given the extent to which we know humans are influenced by automation bias and automation complacency. The datafication of the results presents the illusion of certainty, with significant consequences; it also demonstrates the degree to which the corporate creators of privacy-invasive technologies like FRT maintain the upper hand, even when the technologies are deployed by public organizations governed by other legal orders.

Waldman acknowledges the Rule of Law dimensions produced by shallow privacy remedies, writing:

It also undermines the rule of law. The rise of merely symbolic structures neuters the ability of legislation to enact social policy: why pass a law to achieve positive social change if its goals are going to be frustrated in practice? Moreover, as the locus of legal decision-making shifts further away from policymakers to corporate managers, the substantive and procedural protections in the laws on the books may dissipate.¹²¹³

Many of our modern institutions developed during a time when industrialism reigned supreme.¹²¹⁴ In *Between Truth and Power*, Julie Cohen reflects on society’s movement from industrialism to an informational way of understanding the world, which has displaced market power in favour of platform power.¹²¹⁵ As Wall Street and Silicon Valley have agitated for deregulation and for concentration of governance mechanisms in the private sector, existing regulatory models have

¹²¹¹ Garvie et al, *supra* note 1113.

¹²¹² *Ibid.*

¹²¹³ Waldman, *supra* note 1203 at 778.

¹²¹⁴ Julie Cohen, “Chapter 5: The Regulatory State in the Information Age” in *Truth and Power*, *supra* note 139 at 1.

¹²¹⁵ *Ibid* at 5.

become impoverished.¹²¹⁶ Dominant privacy processes tend to be in service of these companies' complex information ecosystems, as "dominant platform firms . . . have thrived by developing ways to offer both individual and business users a wide variety of information services while controlling both advertising markets and the harvesting and processing of user data."¹²¹⁷ Given this concentration of power in a handful of platform firms, they maintain the ability to dictate dominant processes—and instill their desired values in their attempts at privacy protection.

b. Acknowledging Contested Concepts: Which Values Dominate?

As Chapter 4 explored in detail, the contested nature of fundamental legal ideas like the Rule of Law can make legal administration challenging, particularly if the administration is delegated to automated decision-makers. Privacy experiences analogous challenges, especially given its variable legal protection through constitutional, tort, and statutory law—as well as corporate policy. Mireille Hildebrandt evocatively refers to privacy as an "iridescent" concept.¹²¹⁸ She writes:

The problem with privacy is its iridescent nature; the concept means different things in different contexts and its elusive character makes it hard to provide hard and fast rules for its application. The concept is exemplary as an essentially contested concept, its content being value-laden, and therefore vague and ambiguous, requiring sound knowledge of case law to grasp the legal effect of an infringement.¹²¹⁹

Since the advent of privacy case law and legislation, courts have invoked theoretical groundings to generate legitimacy for the different approaches that are proposed to govern privacy concerns. Often, reliance is placed on Alan Westin's conception of privacy as control, where privacy is a claim to control when, how, and to what extent information about the self is communicated to others.¹²²⁰ From this perspective, privacy is premised on the existence of a "core self," a classical liberal autonomous view, which is bound up in asking how the core self is able to control their personal information *vis-à-vis* others.¹²²¹ Also popular is Ruth Gavison's conception of privacy as a right of limited access, where a person's privacy interest is derived from the degree to which they make themselves accessible to others—by virtue of being known, or allowing others to have physical access, or be subject to another's attention.¹²²² In Gavison's view, a person would achieve perfect privacy if she were completely inaccessible to others, and everyday privacy interests are ongoing negotiations about the degree to which accessibility is granted.¹²²³ Neither of these traditional views, however, assist much in understanding more nuanced flows of information common in the modern information economy, where consumers are regularly asked to make decisions about their informational privacy.

¹²¹⁶ Cohen, *supra* note 139 at 2.

¹²¹⁷ *Ibid* at 5.

¹²¹⁸ Hildebrandt, *End(s) of Law*, *supra* note 30 at 188.

¹²¹⁹ *Ibid*.

¹²²⁰ Alan Westin, *Privacy and Freedom* (1968) 25 Wash & Lee L Rev 166.

¹²²¹ *Ibid*.

¹²²² Ruth Gavison, "Privacy and the Limits of the Law" (1980) 89 Yale LJ 421.

¹²²³ *Ibid*.

Identifying the relevant sphere of privacy interests permits some negotiation of what requires boundaries and what does not. I highlight three theoretical approaches that each prioritize different values: (i) privacy as contextual integrity; (ii) privacy as consent management; (iii) privacy as trust or loyalty. Thinking about FRT through the lenses of these approaches offers perspective on how theoretical values inform this malleable area of law, especially with it so dependent on corporate policies. I note that this section offers only a small sampling of the many different theoretical justifications for privacy, which I argue are especially relevant for conceptualizing FRT.

i. Privacy as Contextual Integrity

Helen Nissenbaum's concept of privacy as contextual integrity has gained some traction in recent years as a theoretical framework for understanding the discomfort associated with privacy violations. In Nissenbaum's conception, privacy can be understood as two concurrent normative claims: a norm of appropriateness and a norm of flow.¹²²⁴ By rejecting the traditional dichotomy between private and public space, she acknowledges that when we examine the "texture" of peoples' lives, our information-sharing activities take place in "a plurality of distinct realms."¹²²⁵ Therefore, our privacy interests are best understood as norms of appropriateness and flow. The norm of *appropriateness* dictates what information should be revealed in what context, circumscribing when certain information is allowed, expected, or demanded to be revealed. The norm of *flow* describes the feeling that privacy has been violated when norms around the use of information are transgressed, *i.e.*, that the information was given to an unintended person or used for an unanticipated purpose. Because contextual integrity exists outside the idea of protecting some core self, instead acknowledging the right to exist as a different self in different contexts, it offers a robust theoretical foundation for circumscribing both data collection and aggregation.

ii. Privacy as Consent Management

Daniel Solove describes the foibles of consent management in a 2013 paper: "Privacy Self-Management and the Consent Dilemma."¹²²⁶ His analysis showcases the pitfalls of unduly relying on consent as a form of control. Data protection regimes take refuge in consent, emphasizing the importance of individual decision-making to weight the costs and benefits of how their own data is used. But, as Solove argues, focusing on consent doesn't make sense in practice, due to both cognitive and structural problems. People are ill-equipped to make rational choices about them both because they are uninformed and because they don't have the expertise to understand what might happen down the line. The number of different entities making requests further impedes meaningful decision-making: non-sensitive information may be willingly shared numerous times to create unexpectedly fulsome results upon aggregation. Problems of scale limit engagement. By addressing privacy self-management as a series of isolated decisions, each individual privacy harm is small and dispersed, meaning the broader social importance of privacy is undervalued. A more clear-headed approach, he argues, would look at the problem holistically and cumulatively.

Solove points out that attempts to better manage the consent dilemma are muddled by paternalism: taking individual choice away on account of knowing what is best for people's data. Of course,

¹²²⁴ Helen Nissenbaum, "Privacy as Contextual Integrity" (2004) 79 Wash L Rev 119 at 137.

¹²²⁵ *Ibid* at 137.

¹²²⁶ Solove, *supra* note 823.

some uses of data might be for good and offer social benefit—and social improvements may be delayed or derailed through undue regulation. Finding the right degree of individual choice also invites questions of meaningful consent and consent fatigue. Since the enactment of the *GDPR*, companies aiming at *GDPR* compliance have updated their privacy policies, often requesting additional consent from users for various means.¹²²⁷ This process can be repetitive, time-consuming, and confusing. Some suggest privacy labeling—like food labeling—to expedite the process.¹²²⁸ Yet, these methods don't change the reality of inundating consumers with choices about their privacy consequences that are often insulated from the context in which their choices operate.

iii. Privacy as Trust/Loyalty

Trust is a key concept for the transition to digital services. The gap between current government regulation and the private sector companies' tendencies to harvest large quantities of data has contributed to a systemic trust problem.¹²²⁹ Neil Richards and Woodrow Hartzog propose viewing privacy obligations through the lens of a duty of loyalty.¹²³⁰ They argue that the current model of data privacy law does little to constrain modern technology companies' opportunism—at the expense of those who entrust them with personal data. Instead, they propose, data controllers should be modelled after fiduciary law: they should owe a duty of loyalty to individuals “whenever a person is susceptible to exploitation within an information relationship where trust was invited and given.”¹²³¹ This framework could, in turn, protect information users from exploitation and surveillance.

iv. Applying Theory to FRT

Challenges raised by ML can highlight the differences offered through these different theoretical approaches. Imagine a scenario where an individual's face is recognized by FRT as they go about their day in public space. Taking one of the more traditional positions on privacy protection, one might argue that there is limited privacy interest: the person has revealed their face in public. Westin would say that they have relinquished control of that personal information (their face); Gavison would say they do not enjoy perfect privacy since they have granted others access to their face. Yet, there remains something unsatisfactory about agreeing that a person has consented to have their face scanned and saved in an FRT repository simply by virtue of appearing in the world.

¹²²⁷ In a well-known 2008 study, Aleecia McDonald and Lorrie Faith Cranor estimated that the average US worker encounters 1426 privacy policies per year. Reading them would take 25 full days—or 76 8-hour workdays per year. Of course, these numbers are from 2008 and have only increased since. Alexis C Madrigal, “Reading the Privacy Policies You Encounter in a Year Would Take You 76 Work Days” (1 March 2012), online: *The Atlantic* <www.theatlantic.com/technology/archive/2012/03/reading-the-privacy-policies-you-encounter-in-a-year-would-take-76-work-days/253851/> [perma.cc/N8FZ-QSHM].

¹²²⁸ Luis Alberto Montezuma & Tara Taubman-Bassirian, “How to Avoid Consent Fatigue” (29 January 2019), online: *IAPP* <iapp.org/news/a/how-to-avoid-consent-fatigue/> [perma.cc/77A7-M7PR].

¹²²⁹ In an opinion piece for CIGI, Teresa Scassa argues that part of the reticence to use contact tracing applications for the COVID-19 pandemic are linked to citizens' lack of trust in governments handling their data. She argues that governments have been largely unwilling to enact privacy protection, in part because they benefit from knowing the data collected by the private sector is a mere production order away. Teresa Scassa, “The Post-pandemic Future of Trust in Digital Governance” (20 July 2020), online: *CIGI* <www.cigionline.org/articles/post-pandemic-future-trust-digital-governance> [perma.cc/SZU8-YXVS].

¹²³⁰ Neil Richards & Woodrow Hartzog, “A Duty of Loyalty for Privacy Law” (3 July 2020), online (pdf): *SSRN* <papers.ssrn.com/sol3/papers.cfm?abstract_id=3642217> [perma.cc/R5W8-GCQ9].

¹²³¹ *Ibid.*

Nissenbaum's recognition of privacy as contextual integrity begins to explicate the impugned privacy interest here: the capture and use of the image transgresses the norm of flow, since only people occupying the same public space would traditionally be able to see one's face. The technological innovation of FRT allows for the information to transcend a boundary.

The limitations to consent management explained by Solove underscore this point. Companies may also be underprepared to understand the privacy implications of their data uses. In a well-publicized example, the fitness application Strava aggregated its fitness tracking data into "heat maps," which showed where app users were exercising.¹²³² In so doing, they accidentally exposed the location of potential United States military assets in Africa and the Middle East. And, even more alarmingly, combining the Strava data with other public data could lead to easy identification of individual soldiers back home.¹²³³ In the *New York Times*, Professor Zeynep Tufekci explains:

Data privacy is not like a consumer good, where you click "I accept" and all is well. Data privacy is more like air quality or safe drinking water, a public good that cannot be effectively regulated by trusting the wisdom of millions of individual choices. A more collective response is needed.¹²³⁴

Even seemingly innocuous private information can sometimes lead to catastrophic results. While phone numbers and addresses have long since been published in telephone and business directories, their availability can create new challenges in the online world. Retaliatory measures become significant fears for active Internet users. These measures can have a gendered or racialized component. The infamous Gamergate offers chilling examples of individual harms: the online harassment campaign targeted women in the video game community with doxing,¹²³⁵ swatting,¹²³⁶ death threats, and rape threats.

While the *PIPEDA* regime is directed towards commercial uses of data, personal data usage has become an integral part of many business models. The intertwined nature of consumers as users makes it nearly impossible to separate harm to individuals from harm to the entire system. Prior to the availability of big data methods, it may have been reasonable to assume that individuals were best placed to make informed decisions on their own levels of comfort about sharing personal information. Yet, individuals simply do not have sufficient information to make educated choices about the specific ramifications of their data choices. What sorts of downstream consequences might flow from how individuals manage their privacy? What kinds of insights might be generated from a large dataset containing thousands or millions of records? Individuals cannot easily answer these questions. Absent a well-defined corporate interest in their individual well-being, as posited by Richards and Hartzog, it may be hard to achieve such understanding.

¹²³² Zeynep Tufekci, "The Latest Privacy Debacle" (30 January 2018), online: *The New York Times* <www.nytimes.com/2018/01/30/opinion/strava-privacy.html> [perma.cc/35G3-P2KE].

¹²³³ Even a well-intentioned company acting in good faith, Tufekci argues, might not truly have the information needed to alert users about potential risks associated with their consent. *Ibid.*

¹²³⁴ *Ibid.*

¹²³⁵ Doxing refers to posting someone's publicly available personal information to encourage online harassment of that person. See Arthur Chu, "Opening Pandora's Dox" (29 August 2015), online: *TechCrunch* <techcrunch.com/2015/08/29/opening-pandoras-dox-the-unintended-consequences-of-an-internet-that-never-forgets/> [perma.cc/7HSA-SEPA].

¹²³⁶ Swatting refers to making a false telephone call reporting a hostage situation at an address, in hopes that the police will react by sending a SWAT team to investigate. *Ibid.*

c. Constructing Truth: Whose Story Dominates?

While many different technologies could be used to tell the story of how innovation begets privacy violations, the history of images and photography offers a unique lens on these fellow travelers. Beginning with Warren and Brandeis' article, the idea of image capture punctures the anonymity historically available in public appearances—the kind of anonymity one experiences walking down a crowded street in an unfamiliar city where you don't know anyone. As canvassed above, such images can be compiled in ways that specifically target civil liberties and constitutional rights. Writing about the perceived empiricism in computer vision applications, Jake Goldenfein notes that criminal data sets were one of the first things created with daguerreotypes, a forerunner to modern photography.¹²³⁷ These criminal portraits were then compiled by British eugenicist, Sir Francis Galton, who believed the daguerreotypes could reveal the “mean appearance of criminality.”¹²³⁸ Even though the pseudoscientific approaches of phrenology and physiognomy had been discredited before Galton began experimenting,¹²³⁹ Goldenfein emphasizes that a resolute belief in photography's “mechanical objectivity” added new fuel to the eugenics fire. The experiments ultimately failed: all that Galton managed to show was that “visual similarities of the criminal classes disappeared through the composite process which instead revealed the ‘common humanity in all.’”¹²⁴⁰ Despite this, similar beliefs still have purchase today, including with ML's attempts to assess whether a person identifies as homosexual from still photographs,¹²⁴¹ or new AI-enabled emotion recognition software.¹²⁴² An Israeli company called Faception even claims to have developed software capable of determining an individual's IQ and corresponding propensity towards crime and terrorism, based solely on their face.¹²⁴³

Undue confidence in photography's “mechanical objectivity” is nothing new. Misplaced confidence in the abilities of new technologies has often created chaos in legal settings, where fledgling efforts towards empiricism employ later-discredited technologies to generate morally suspect results. Even if the technological approaches are later found to be invalid, damage can be done through shifting epistemological consequences: Goldenfein refers to Nicole Rafter's observations of how new technologies can usher in new ways of knowing that leave lasting impacts.¹²⁴⁴ Through this lens, Rafter discusses how phrenology provided a method for “those searching for a new philosophical basis for discipline and social control,” which ultimately transformed the penal system.¹²⁴⁵ For the legal profession, this means that phrenology introduced new ideas about crime

¹²³⁷ Jake Goldenfein, “The Potential Profiling of Computer Vision and the Challenge of Computational Empiricism” (2019) Proceedings of the ACM FAT* Conference, online (pdf): *SSRN* <papers.ssrn.com/sol3/papers.cfm?abstract_id=3284598> [perma.cc/RJX6-RT9S] at 6.

¹²³⁸ *Ibid.*

¹²³⁹ Phrenology refers to the ability to determine a person's character traits from the morphology of their skull. Physiognomy refers to the science of “the correspondence between the external and internal man,” or more simply, reading people's character from their faces. Criminologist Nicole Rafter notes that such approaches “emerged out of the Enlightenment drive to replace metaphysical and theological explanations with scientific accounts of natural and social phenomena.” See Nicole Rafter, “The Murderous Dutch Fiddler: Criminology, History, and the Problem of Phrenology” (2005) 9 *Theoretical Criminology* 65 at 70.

¹²⁴⁰ Goldenfein, *supra* note 1237 at 7.

¹²⁴¹ Levin, *supra* note 1137.

¹²⁴² Goldenfein, *supra* note 1237 at 3. See also Crawford, *Atlas*, *supra* note 76 at 152.

¹²⁴³ Buolamwini & Gebru, *supra* note 862 at 2.

¹²⁴⁴ Goldenfein, *supra* note 1237 at 11. See also Rafter, *supra* note 1239 at 85.

¹²⁴⁵ Rafter, *supra* note 1239 at 81.

and justice to the zeitgeist, which created a lasting legacy.¹²⁴⁶ Long after the relevant technological methods were discredited, the transformative impacts remain.

The technological approach can adopt a particular ideological stance. Like Winner's observation that artifacts have politics, Goldenfein observes how ideological positions can be "laundered" through the affordances of technologies.¹²⁴⁷ This laundering can have particularly nefarious consequences when the technology is presented as neutral, which can subvert the inherent political position articulated by the technology itself. Rafter notes that phrenology's appeal "lay partly in its implicit hierarchies,"¹²⁴⁸ which had a certain appeal in the social and cultural context of the time, one of "widespread interest in applying scientific methods to the study of social phenomena."¹²⁴⁹ Such hierarchies coincide with the drive for classification exhibited by advocates for self-driving laws, determined to apply empirical approaches to better resolve longstanding human quandaries. Yet, as Goldenfein points out, a deep irony remains at the centre of such debates: ML approaches are automatically generated, flowing from multiple layers of abstraction, to generate complex representations of reality. Goldenfein emphasizes his objection:

The inherent paradox of this epistemological platform is the use of correlation to create invisible or inaccessible automated representations in order to surpass the use of manual representation. In other words, overcoming the limits of representation with more representations. Framing it in this way prompts us to reconsider imagistic representation within data scientific contexts. ... Representation in computer vision has changed from finding the target thing (i.e. the pattern) from all the information within an image. With real consequences for how we derive meaning from the visual world.¹²⁵⁰

Long after the technology is no longer commonplace, or has evolved into its eventual progeny, we might still be suffering the normative consequences of that technology's affordances.

4. Storytelling Solutions

Hill tells the story of a man named Robert Julian-Borchak Williams, the first known person to be wrongfully arrested for a crime due to an FRT algorithm.¹²⁵¹ Mr. Williams received a call at work asking him to come into the local police station to be arrested, which he dismissed as a prank call. Upon arriving home, he was immediately blocked into his driveway by a police cruiser. He was handcuffed on his front lawn, in front of his two small daughters and wife. At the police station, officers presented surveillance photos from a recent robbery at a local upscale boutique shop. The pictured man was black and heavysset, like Mr. Williams, but was clearly not him. Mr. Williams had been wrongfully identified by FRT.

"Is this you?" asked the detective.

¹²⁴⁶ Rafter, *supra* note 1239 at 86.

¹²⁴⁷ Goldenfein, *supra* note 1237 at 12.

¹²⁴⁸ Rafter, *supra* note 1239 at 73.

¹²⁴⁹ *Ibid* at 74.

¹²⁵⁰ Goldenfein, *supra* note 1237 at 17.

¹²⁵¹ Hill, "Wrongfully Accused," *supra* note 1133.

The second piece of paper was a close-up. The photo was blurry, but it was clearly not Mr. Williams. He picked up the image and held it next to his face.

“No, this is not me,” Mr. Williams said. “You think all black men look alike?”¹²⁵²

In telling Mr. Williams’ story, Hill emphasizes the recent studies that demonstrate FRT’s failures on racial demographics other than white men.¹²⁵³ Citing the Perpetual Lineup Report, she discusses recommendations that low-quality search images should be banned, and that policies should be instituted to rigorously test systems for bias and accuracy.¹²⁵⁴ But perhaps most importantly, she notes that the FRT results generated by the algorithm state explicitly: “This document is not a positive identification,” in bold capital letters across the top. “It is an investigative lead only and is not probable cause for arrest.”¹²⁵⁵ While both the FRT providers and law enforcement agencies attempt to emphasize this information as being “only a clue, not a smoking gun,”¹²⁵⁶ how this information might be used is inevitably impacted by automation bias and automation complacency. Like the COMPAS sentencing reports purporting to be only a small part of the process, our tendency to rely on the authority of technological systems will unduly centre the AI/ML process.

Once again, Nedelsky’s relational approach offers a useful framing device. She describes autonomy as an aspect of the individual self that routinely evokes the delineation of a boundary, for both law and other discourses.¹²⁵⁷ In this way, autonomy is similar to privacy: it operates with a fundamentally fluid capacity, requiring continuous development. Contrary to privacy theories (like Westin’s or Gavison’s approaches) that propose protection of a central core self, understanding privacy relationally is not trying to prevent intrusion into one’s core humanity because it recognizes that no such core exists as an end unto itself: rather, it is being continuously created. Nedelsky is sceptical of laws that only serve to draw boundaries, since they can do no more than see the self as something requiring protection.¹²⁵⁸ If one’s autonomy is understood as being constantly under construction, this process requires asking questions about how social relationships might be structured to foster—rather than undermine—autonomy.¹²⁵⁹ In my view, this interrogation generates a constructive aspect to privacy, which is best understood as a participatory exercise, taking into account the many variable social relationships that a person has in their world. Kerr makes a similar point in “Schrödinger’s Robot: Privacy in Uncertain States,” when he calls for a relational account of privacy that “will attempt to measure the seriousness of a privacy violation in terms of the harm it does to the web of relationships within which an individual finds herself.”¹²⁶⁰ Like Nissenbaum’s theory of privacy as contextual integrity, which emphasizes conceptualizing privacy through the “spiderweb-like relations of our lives,” this way of thinking requires the

¹²⁵² Hill, “Wrongfully Accused,” *supra* note 1133.

¹²⁵³ *Ibid.*

¹²⁵⁴ *Ibid.* See also Garvie et al, *supra* note 1113.

¹²⁵⁵ Hill, “Wrongfully Accused,” *supra* note 1133.

¹²⁵⁶ *Ibid.*

¹²⁵⁷ Nedelsky, *supra* note 816 at 98.

¹²⁵⁸ *Ibid.*

¹²⁵⁹ *Ibid.*

¹²⁶⁰ Kerr, “Schrödinger,” *supra* note 824.

delineation of a nebulous *other* in order to animate consideration of privacy as a relationship-based construct.¹²⁶¹

Understood relationally, we can see autonomy as being made possible through our relationships with others, rather than from excluding others. Instead of using law as a shield to enable the discovery of the core self, law can be used to thoughtfully guide social relationships towards the best possible creation of self. It is nonsensical, then, to understand the central problem of privacy as maintaining a personal sphere that neither the state nor other interested parties may penetrate. For this reason, privacy rights are often conceptualised analogously to property rights: property rights often begin from Blackstone's single-variable essentialism, the right to exclude.¹²⁶² Property thereby defines what representatives of the state, or society generally, may not touch.¹²⁶³ And despite these exclusionary precepts, property's proper function demands a collective response: property rights demand recognition by others of an individual entitlement to property.

Mr. Williams' story points to a documented phenomenon known as other-race recognition bias, which has long confounded police identification lineups.¹²⁶⁴ His story evokes the equality-seeking critiques detailing how FRT is less accurate at identifying black faces than white ones.¹²⁶⁵ It exemplifies the challenges with the accuracy of FRT systems deployed by law enforcement, particularly when these are used within police departments with legacies of racism or bias in policing.¹²⁶⁶ It combines the criminal and constitutional dimensions of privacy law with the corporate practices extolled by the relevant powers. In light of recent racial reckonings in law enforcement, Hill notes that large platform firms like Amazon, Microsoft, and IBM have announced moratoriums on selling FRT to police forces.¹²⁶⁷ This announcement should offer little comfort: the preponderance of FRT is supplied by lesser known companies, including Vigilant Solutions, Cognitec, NEC, Rank One Computing, and Clearview AI.¹²⁶⁸ As such companies continue to shape law enforcement's technological landscape, privacy concerns hang in the balance, making both their legal dimensions and corresponding remedies difficult to locate. And as individuals become more and more likely to be captured by FRT as they move through public space, the theoretical concepts of privacy struggle to elucidate the nature of the concerns. Even though the SCC's recent decision in *R v Jarvis* affirms a context-sensitive understanding of how technological systems are being used, the reasonableness criterion that animates legal analysis in both constitutional cases under section 8 and *PIPEDA* analyses under section 5(3) is threatened by changing norms around use of space. If it becomes reasonable to expect that your face will be captured by FRT, or reasonable to expect that law enforcement will use it to identify you, will

¹²⁶¹ Kerr, "Schrödinger," *supra* note 824 at 7.

¹²⁶² English jurist William Blackstone's famous definition of property: "that sole and despotic dominion which one man [sic] claims and exercises over the external things of the world, in total exclusion of the right of any other individual in the universe." See Bruce Ziff, *Principles of Property*, 7th ed (Toronto: Thomson Reuters, 2018) at 4.

¹²⁶³ Nedelsky, *supra* note 816 at 98.

¹²⁶⁴ P Jonathan Phillips & Fang Jiang, et al, "An Other-Race Effect for Facial Recognition Algorithms" (2011) 8 ACM Trans On Applied Perception, online: *Semantic Scholar* <pdfs.semanticscholar.org/5fb9/4eaf71196ed9cb87bbd881af34be8b4bc919.pdf> [perma.cc/D3SQ-JNEQ].

¹²⁶⁵ Buolamwini & Gebru, *supra* note 862.

¹²⁶⁶ Garvie et al, *supra* note 1113.

¹²⁶⁷ Hill, "Wrongfully Accused," *supra* note 1133.

¹²⁶⁸ *Ibid.*

privacy interests really be preserved? Will it be possible for context-sensitive storytelling to emerge from FRT's affordances, particularly when used by police?

If we try to imagine the story that would emerge from ubiquitous use of FRT, we can imagine it might sound a lot like *The Minority Report*—the Philip K. Dick science fiction short story where mutant precognitive entities are used to predict future crime and arrest the relevant perpetrators, before the predicted crime ever happens.¹²⁶⁹ This narrative echoes Mr. Williams' situation: even though he had done nothing wrong, he was dramatically hauled off into a squad car in front of his wife and children; even after it was clear he was not actually the person from the image matched by FRT, he was still held in police custody for 30 hours.¹²⁷⁰ Thinking about his legal privacy interests through the protection of the core self, Mr. Williams may have abdicated his privacy interest by sometimes appearing in public and sometimes uploading photos of himself to the Internet. But, thinking about them relationally, in constant negotiation with the relevant powers, permits a more context-sensitive story to emerge. In fact, Mr. Williams had an airtight alibi, if the police had bothered to check for one. At the time of the alleged robbery, he had posted a video of himself to his Instagram account, driving on his way home from work, crooning along to the car radio.¹²⁷¹ In that scenario, the context-sensitive story that might be told was augmented by his use of technology. This, combined with racial awareness of the technology's limitations, could promote the articulation of a different type of story, despite the competing legal and corporate interests.

* * *

Flowing from the foregoing analysis are two types of AI-powered stories. The first, developed in Chapter 6's example of COMPAS sentencing software by the judiciary and Chapter 7's example of FRT usage by state actors through predictive policing, is a story about self-driving laws where theoretical legal assumptions are programmed into autonomous decision frameworks. The second, exemplified by Chapter 7's example of FRT design and market distributions by private actors, is a story about self-driving laws where human law is influenced, shaped, and ultimately transformed by the AI's affordances. Each of these stories relates to the Rule of Law, although in slightly different ways. In the first story, migration to a system of self-driving laws is hampered by an inadequate, formalist conception of law, neglecting deeper inquiry or treatment of contested concepts. The AI replaces judge-made law, prescribing certain outcomes through its operation. In the second story, self-driving laws' influence create uncertainty in how the overall system operates, generating confusion about how much work existing laws are doing to rule law's empire *versus* how much power is exerted by the utilized technology—and the corporations that created it. The AI supplements human actions and decisions, prescribing certain outcomes due to its affordances.

Common themes exist in these two stories. FRT demonstrates that AI systems might either completely take over certain aspects of the legal system, or it can do a substantial portion of the job, leaving a modicum of clout for human actors. Trying to strike an appropriate balance between the respective roles for technology and humans in such a shared system is complicated by technological affordances, traditional approaches to legal theory, and corporate exercises of power.

¹²⁶⁹ Philip K Dick, "The Minority Report" (1956), available online: <cwanderson.org/wp-content/uploads/2011/11/Philip-K-Dick-The-Minority-Report.pdf> [perma.cc/98E7-R8EG].

¹²⁷⁰ Hill, "Wrongfully Accused," *supra* note 1133.

¹²⁷¹ *Ibid.*

Accordingly, AI systems create *machine*-ations of the Rule of Law, whether the technology is being wielded by a state actor, a private actor that eventually influences a legal context, or related quasi-legal scenarios that trigger abuses of power. This observation obviates the distinction between public and private instantiations of law; it also emphasizes that self-driving laws exceed merely judicial decision-making applications, imbricating the interactions between technology and people. In all such situations, the technology's affordances enable particular courses of action, simplifying the route for sleepwalking into an undesirable legal future. Not only are such instances of regulation by machine impacted by the (mis)understanding of how the AI/ML technology operates, they are also informed by (mis)understanding how the technology relates to each of us, and—as Nedelsky helps us understand—by our relations with one another.

With these components in place, I turn now to Chapter 8 and a third story of an AI-powered system—natural language processing— to imagine a world where AI is used to write legislation or legal decisions.

*Words do not express thoughts very well. They always become
a little different immediately after they are expressed.
A little distorted. A little foolish.*
~ Hermann Hesse¹²⁷²

Once, at the outset of this thesis, I told the story of the impugned text suggestions that the automated text feature in my email application proposed as responses to my former housemate. We were struggling against a basic administrative issue with our landlady, who had neglected to send along the relevant form to kickstart an account for our electricity service, even though we had already been living in the new rental house for months. While my housemate's pointed sarcastic comments provided needed levity in a frustrating situation, they were somewhat beyond the capabilities of the Google robots that had processed the message on my phone's email application. Whatever algorithm processed the text simply parsed the plain meaning of the sentence, as written. By way of reply, I was offered three different, but virtually indistinguishable, responses—all of which were positive in tone. There was no way to respond to the true sentiment of the message through use of one of the provided prompts. I was still given the chance to provide my own written response, but it would have been much easier to simply select one of the pre-scripted responses.

The key question emanating from this example was: what would happen if the technology began writing our conversation by itself? The technology powering this application has grown by leaps and bounds in recent years, precipitating calls for its inclusion in another highly written medium: the law. Yet, as I have explored in this thesis, delegating authority to technological analogues can create unintended issues, with important implications for the Rule of Law. This thesis has emphasized how current technological capabilities can instigate new behaviours, creating corresponding changes to cultural practices and norms. The default settings imposed by technological systems can guide user behaviour, disrupting long established practices and replacing them with practices guided by technological design. For law, wherein power adheres through the written word, this disruption poses interesting quandaries. Law's normative force adheres in language, meaning, and interpretation. If the language constructing legal norms is shaped by ML, what might the consequences be for legal knowledge?

The previous two chapters—focused respectively on sentencing software and facial recognition technology—dealt with technological systems that are already in place and influencing the law in different ways. This final chapter aims to do something slightly different: it looks ahead, to a prospective vision of society, where the version of self-driving laws described by Benjamin Alarie in Chapter 1 might be fully realized.¹²⁷³ This vision of the future requires a certain suspension of disbelief: so far, no natural language processor has been used to create law or administer its empire. Yet, even more optimistic thinking animates the existing legal futurism literature, such as Alarie's aspirations for the realization of a legal singularity that would instantaneously process the full

¹²⁷² John Kaag, *Hiking with Nietzsche: On Becoming Who You Are* (New York, NY: Farrar, Straus & Giroux, 2018) at 218.

¹²⁷³ Alarie, "Singularity," *supra* note 36.

breadth of human knowledge to independently find solutions to legal problems. While this proposal might seem far-fetched, some legal scholars anxiously await the day when this vision will be realized, and some early versions of these technologies are already at work today. In my view, hypothesizing these potential futuristic outcomes now is necessary to think towards philosophical foundations for machine learning in law.

To ground this prospective vision in current technological capabilities, I use the example of natural language processing, which already has some purchase in existing machine learning processes. Specifically, recent advancements using a natural language processor known as GPT-3 have transformed technological capabilities, by enabling a plain text functionality that permits even inexperienced computer users to give the algorithm simple tasks to perform.

To illustrate its capabilities, I asked GPT-3 how to begin this chapter, using a plain language prompt: *Generate a parable to start a chapter that would explain natural language processing and how it might connect to the law.* Here's what it had to say:

Two lawyers were arguing in court. One lawyer said to the other, “Your Honor, my opponent is trying to trick you with words. He's using language processing to make you think he's right.”

The other lawyer replied, “That's not true! I'm just using common sense to interpret the law.”

The judge said, “I'm going to have to side with the second lawyer. Language processing can be helpful, but it's not always reliable. Sometimes, the best way to understand something is to use your own common sense.”¹²⁷⁴

At first glance, it seemed my experiment was a success. I was heartened to learn—from the “mouth” of an AI, no less—that the AI revolution is not upon us just yet, given GPT-3's statements of its own shortcomings. Still, its uncanny ability to create believable text, as well as its appeal to the fundamental human necessity of common sense, was prescient. It seems increasingly plausible that such technology might be deployed in the legal sphere, as some scholars posit. The corresponding cultural consequences, while enormous, are already being tracked in technology law literature. I begin this chapter with a thorough explanation of natural language processing's current capabilities, particularly as expressed with GPT-3. Projecting ourselves into a future where natural language processing might be used to solve legal queries, draft legal documents, or compose reasons, I ask GPT-3 for its solutions on how to best resolve the Rule of Law problems I've identified. By exploring GPT-3's operations, I trace the same types of Rule of Law problems that were emphasized in the previous two chapters, demonstrating concerns for the creation of those elements that form the central taxonomy of this thesis: process, values, and story.

1. Natural Language Processing

At the juxtaposition of new technology and culture enters natural language processing (NLP) systems, allowing for flexible, task-agnostic technological representations of problems in regular

¹²⁷⁴ A screen capture of this experiment is posted on my personal website. Katie Szilagyi, “PhD GPT-3 Experiments” (last visited 10 August 2022), online: <katieszilagy.com/phd-gpt-3-experiments/> [perma.cc/4SSX-2YLB] [Szilagyi, “GPT-3 Experiments”].

daily language. The advanced capabilities of modern NLP systems offer transformational benefits to AI applications, dramatically extending their capabilities.¹²⁷⁵ The idea is straightforward: instead of needing complex mathematics and programming language skills to instruct computers to build models, NLP allows individuals to pose queries in simple English. While the same types of vector algebra, clustering techniques, and machinations of floating-point numbers might be executed by the program in the background, these functions are obscured from the view of the everyday user. The user simply inputs a query in regular prose—typically asking the model to generate some sort of text—and the model generates a response to the best of its ability. Most of the powerful NLP systems that exist today are derived from the vast repository of human language that exists online.¹²⁷⁶

The most fully discussed NLP model is currently OpenAI’s GPT-3, introduced in July 2020.¹²⁷⁷ GPT-3 marked a huge departure from previous models with its 175 billion parameters—over 10 times more than its predecessors.¹²⁷⁸ Its user-friendly application programming interface (API) makes NLP easy to deploy—even for non-experts—as plain language commands can be easily translated into usages for various applications. GPT-3 shines at “zero-shot learning,” which means that the model receives no demonstrations and only natural language instructions of what it is supposed to do.¹²⁷⁹ This approach is intended to maximize convenience and better mimic how humans are instructed to perform tasks.¹²⁸⁰

GPT-3’s straightforward API offers several different types of systems. These include text completion, code completion, fine-tuning, and other specialized operations.¹²⁸¹ Text completion allows for text inputs to be used to generate output text matching the appropriate context or patterns.¹²⁸² Code completion allows for something similar to be performed using code, in various common programming languages.¹²⁸³ Fine-tuning is designed to create more exacting results,

¹²⁷⁵ Christian, *supra* note 40 at 34-37.

¹²⁷⁶ Many NLP researchers harness data drawn from Twitter, Reddit, Facebook, etc. For a list of current projects in this space, see generally: SocialNLP 2022, “The Tenth International Workshop on Natural Language Processing for Social Media: Proceedings of the Workshop” (14-15 July 2022), online (pdf): *ACL Anthology* <aclanthology.org/2022.socialnlp-1.0.pdf> [perma.cc/R3R5-JJPU].

¹²⁷⁷ T Brown, B Mann, et al. “Language Models are Few-Shot Learners” arXiv:2005.14165v4 [cs.CL] (22 July 2020), online (pdf): *arXiv* <arxiv.org/pdf/2005.14165.pdf> [perma.cc/ZF2N-AMMA]. This paper introduced GPT-3 as a new NLP mechanism.

¹²⁷⁸ OpenAI makes GPT-3 available to users through an online API, largely obscuring its backend operations from those who wish to deploy its NLP capabilities. This greatly simplifies the use of GPT-3, enabling users to use simple English text prompts to access GPT-3’s abilities. See OpenAI’s own explanations for use of its API: OpenAI, “OpenAI API” (11 June 2020), online: *OpenAI Blog* <openai.com/blog/openai-api/> [perma.cc/RY2A-G9EH]. Google also has several large language models presented as competitors to GPT-3: these are GShard, which uses 600 billion parameters, and Switch-C, which uses 1.7 trillion. Emily M Bender, Timnit Gebru, Angelina McMillan-Major, Margaret Mitchell, “On the Dangers of Stochastic Parrots: Can Language Models Be Too Big? 🦜” (2021) In Conference on Fairness, Accountability, and Transparency (FAcCT ’21), March 3–10, 2021, Virtual Event, Canada.

¹²⁷⁹ Brown et al, *supra* note 1277 at 7.

¹²⁸⁰ *Ibid.*

¹²⁸¹ OpenAI currently provides guides to five different use cases: text completion, code completion, fine-tuning, embeddings, and moderation. Only text completion is freely available for users to test without accessing higher level accounts or paying for subscription-based services. OpenAI, “Welcome to OpenAI” (last visited 19 July 2022), online: *OpenAI* <beta.openai.com/> [perma.cc/ZV54-7RM8].

¹²⁸² OpenAI, “Text completion” (last visited 19 July 2022), online: *OpenAI* <beta.openai.com/docs/guides/completion/introduction> [perma.cc/JJ5E-2GNX].

¹²⁸³ GPT-3 purports to be most proficient in Python, but also capable in JavaScript, Go, Perl, PHP, Ruby, Swift, TypeScript, SQL, and Shell. This feature is still in a private beta mode; interested users must join a waitlist. OpenAI, “Code completion” (last visited 19 July 2022), online: *OpenAI* <beta.openai.com/docs/guides/code> [perma.cc/6Z22-DKG2].

allowing GPT-3 to train on several examples before receiving a prompt.¹²⁸⁴ This method of ML is called “few-shot learning,” occurring when the system uses additional examples to better understand what it is being asked to do. The savings, OpenAI promises, arise because fine-tuned models won’t require any examples to be presented in the prompts to generate text.¹²⁸⁵

Because GPT-3’s power adheres from the aggregation of generated data, it only knows what it has been taught. Its ability to create models of the world is circumscribed by the specific training data that it has been fed, keeping its knowledge base fixed securely in the past: it might not be aware of the latest news or emergent language terms. The OpenAI instructions note this specifically:

Keep in mind that the default models' training data cuts off in 2021, so they may not have knowledge of current events. We plan to add more continuous training in the future.¹²⁸⁶

Not surprisingly, the increasing abilities of NLP and increasing size of language models have provoked considerable academic interest. While many heralded GPT-3’s arrival, hoping it would transform the AI industry, others were more cautious, concerned about what might happen when language models grew too large. An especially well-known critique was lodged by a team of authors led by linguist Emily Bender and (now former) Google AI researcher Timnit Gebru,¹²⁸⁷ who observed the problems that can adhere in “synthetic language” as language models grow, making it ever harder to know the contents of the training data. Online sources, the authors noted, are not the most reliable way of forming neutral content to train NLPs.¹²⁸⁸ The authors catalogue various issues from an inclusivity perspective, noting that relying on language models risks “value-lock,” wherein the technology sets social values in stone, entrenching “older, less inclusive understandings.”¹²⁸⁹ Ultimately, the authors characterized NLPs as stochastic parrots: mere “statistical mimic[s],” unable to apprehend the consequences of what they repeat.¹²⁹⁰

Some scholars seem enthusiastic about GPT-3’s potential to replace human cognition altogether. In a recent article, Benjamin Alarie and Arthur Cockfield asked GPT-3 to produce the world’s “first machine generated law review article.”¹²⁹¹ In this thought experiment, Alarie and Cockfield fed GPT-3 some feed text, asking it to opine on “why humans will always be better lawyers, drivers,

¹²⁸⁴ Other advantages offered by fine-tuning include savings on how many tokens are needed to complete a prompt and generating lower latency requests. Reducing token usage means overall cost savings for using GPT-3, which provides pricing schemes to users based on how many tokens they use per month. OpenAI, “Fine-tuning” (last visited 19 July 2022), online: *OpenAI* <beta.openai.com/docs/guides/fine-tuning> [perma.cc/D7ZS-VL6T].

¹²⁸⁵ *Ibid.*

¹²⁸⁶ OpenAI, “Text completion,” *supra* note 1282.

¹²⁸⁷ In late 2020, a controversy erupted in the technology ethics world when Timnit Gebru, well-regarded as a bright light in the growing ethics of AI space, was ousted from her position at Google after her superiors took issue with the contents of this very paper. For a summary of the controversy and corresponding fallout, see Tom Simonite, “What Really Happened When Google Ousted Timnit Gebru” (8 June 2021), online: *WIRED* <www.wired.com/story/google-timnit-gebru-ai-what-really-happened/> [perma.cc/TL5X-VR4Z] [Simonite, “Gebru Ousted”].

¹²⁸⁸ Bender et al, *supra* note 1278 at 4. They also note that 90% of the world’s languages have meagre technological support—which represents about 1 billion of the world’s people. Canvassing only the English-language Internet to create models is bound to generate gaps in understanding.

¹²⁸⁹ *Ibid* at 5.

¹²⁹⁰ Simonite, “Gebru Ousted,” *supra* note 1287.

¹²⁹¹ In a magnanimous move for the often-cutthroat world of academic publishing, the human authors officially credited GPT-3 as a co-author of their publication: Benjamin Alarie, Arthur Cockfield, & GPT-3, “Will Machines Replace Us? Machine-Authoring Texts and the Future of Scholarship” (2021) 3 *Law, Technology & Humans* [Alarie et al, “Machine-Authoring”].

CEOs, presidents, and law professors than artificial intelligence and robots can ever hope to be.”¹²⁹² Acknowledging knowledge workers’ pressing fear that AI might soon replace them, Alarie and Cockfield note in an introduction that GPT-3 remains less than ideal as an author: it did not cite its sources, exhibits some “odd assumptions,” and shows gender bias.¹²⁹³ At least for now, GPT-3 would need to operate under human supervision, but its advanced capabilities might render it employable as professor’s research assistant, at minimum.

Outside of the legal realm, academics share Alarie and Cockfield’s interest in establishing GPT-3’s potential as a co-author for scholarly publications. In its relatively short history, GPT-3 has already been deployed for diverse written applications, including generating new works by long-deceased authors,¹²⁹⁴ creating newspaper articles,¹²⁹⁵ coding the Google homepage,¹²⁹⁶ mimicking popular psychologists,¹²⁹⁷ and drafting a story featuring famed fictional boy wizard Harry Potter.¹²⁹⁸ Recently, Swedish researcher Almira Osmanovic Thunström used GPT-3 to write an academic paper about itself.¹²⁹⁹ The prompt she gave the AI was: “Write an academic thesis in 500 words about GPT-3 and add scientific references and citations inside the text.”¹³⁰⁰ Thunström was immediately impressed by how the AI managed to strike a reasonable academic tone and cite suitable references, despite her relatively vague instructions.¹³⁰¹ Attempting to give credit where was credit due, she specifically asked the AI whether it consented to being listed as a co-author for an academic publication.¹³⁰² Thankfully, she noted, GPT-3 agreed; otherwise, she was concerned about her ethical obligations from an academic perspective.¹³⁰³ In the end, GPT-3 was listed as first author, together with Thunström and her advisor, Steinn Steingrímsson.¹³⁰⁴

In other work,¹³⁰⁵ Alarie is unreserved in his excitement that analogous NLP mechanisms might one day be used to administer law’s empire. Yet, the potential for corresponding Rule of Law problems seems nigh, especially since GPT-3 (or its eventual successors) has no insight into its own work and no knowledge of whether it has made a compelling argument. Before turning to

¹²⁹² Alarie et al, “Machine-Authored,” *supra* note 1291 at 6.

¹²⁹³ *Ibid.*

¹²⁹⁴ Almira Osmanovic Thunström (@augmentedrobot), “281 years in the making: GPT-3 and de la Mettrie” (26 January 2022), online: *Medium* <augmentedrobot.medium.com/281-years-in-the-making-gpt-3-and-de-la-mettrie-de03eea374e4 > [perma.cc/Y95C-9A5T].

¹²⁹⁵ GPT-3, “A robot wrote this entire article. Are you scared yet, human?” (8 September 2020), online: *The Guardian* <www.theguardian.com/commentisfree/2020/sep/08/robot-wrote-this-article-gpt-3> [perma.cc/YWU4-FYPX].

¹²⁹⁶ Sharif Shameem, “Here’s a sentence describing what Google’s home page should look and here’s GPT-3 generating the code for it nearly perfectly. (5 July 2020, 3:50 AM), online: *Twitter* <twitter.com/sharifshameem/status/1283322990625607681>.

¹²⁹⁷ Cade Metz, “Meet GPT-3. It Has Learned to Code (and Blog and Argue).” (24 November 2020) online: *The New York Times* <www.nytimes.com/2020/11/24/science/artificial-intelligence-ai-gpt3.html> [perma.cc/M33H-AB7A].

¹²⁹⁸ Arram Sabeti, “GPT-3: An AI that’s eerily good at writing almost anything” (9 July 2020), online: *Arr.am* <arr.am/2020/07/09/gpt-3-an-ai-thats-eerily-good-at-writing-almost-anything/> [perma.cc/D7U9-Q244].

¹²⁹⁹ Almira Osmanovic Thunström, “We Asked GPT-3 to Write an Academic Paper about Itself—Then We Tried to Get It Published” (30 June 2022), online: *Scientific American* <www.scientificamerican.com/article/we-asked-gpt-3-to-write-an-academic-paper-about-itself-then-we-tried-to-get-it-published/> [perma.cc/VBE7-PCQG] [Thunström, “We Asked GPT-3”].

¹³⁰⁰ *Ibid.*

¹³⁰¹ *Ibid.*

¹³⁰² *Ibid.*

¹³⁰³ Issues around AI authorship include copyright, ethical concerns, and related inquiries into the literary notion of the romantic author. See Carys Craig & Ian Kerr, “The Death of the AI Author” (2021) 52 *Ottawa Law Review* 31.

¹³⁰⁴ Thunström, “We Asked GPT-3” *supra* note 1299.

¹³⁰⁵ Alarie, “Singularity,” *supra* note 36.

cataloguing those issues, I consider some equality-seeking objections to NLP, and algorithmic training grounds generally, drawn from the technology law literature.

a. Equality-Seeking Critiques

While GPT-3 offers a powerful new approach, the technology at the heart of NLP has been developing for several decades, amplified by algorithms trained to parse human uses of language online. Consequently, in some scenarios, training algorithms through online forums has yielded problematic results. One particularly poignant example arose with Google's ubiquitous search platform. Famously, beginning around 2004, it was observed that searching the term "Jew" on Google would return a plethora of anti-Semitic results, including directing searchers to a Holocaust-denial website called "Jew Watch."¹³⁰⁶ James Grimmelman describes this phenomenon as "Googlebombing."¹³⁰⁷ Because Google's PageRank algorithm calculates a website's relevancy in search results relative to how many linkages exist to a given phrase, it was possible for people with nefarious intentions to trick the algorithm by simply creating connections with particular words.¹³⁰⁸ If you wanted certain results to rise to the top of the Google search results page whenever a certain phrase was searched, you could use additional linkages mentioning that phrase to manipulate the way PageRank indexed the page. Eventually, some counter-activists worked to point the term "Jew" to a comparatively neutral site: a Wikipedia page describing the Jewish people.¹³⁰⁹

Despite some public outcry, Google seemed to wash its hands of responsibility for the controversy, simply posting an explanation of how its PageRank algorithm functioned autonomously to connect the most viewed pages on the web.¹³¹⁰ Google emphasized that related search terms like "Judaism, Jewish, or Jewish people" would return more salient results, since those were the terms commonly used by those who actually sought information about the religion and its adherents.¹³¹¹ Conversely, while "Jew" was used as a descriptor of Jewish people, it was also sometimes used as a derogatory slur.¹³¹² And, Google noted, the phenomenon had been exacerbated by people continuing to search the term "Jew" and then click on the impugned results, anxious to test it out for themselves.¹³¹³ As late as 2012, searching the term "Jew" still brought up a beige box to a link with Google's explanation of its algorithmic processes.¹³¹⁴

Even the dropdown menu on Google's search homepage has generated controversy, stemming from the predictive text that appears to autocomplete the beginnings of search queries based on

¹³⁰⁶ Another controversy erupted in 2006 when donations to aid in relief post-Hurricane Katrina were somehow redirected to Jew Watch.

¹³⁰⁷ James Grimmelman, "The Google Dilemma" (2009) 53 *New York Law School Review* 939 at 942. Grimmelman notes that one of the most famous instances of this phenomenon was causing the phrase "miserable failure" to link to U.S. President George W. Bush's official biography page [Grimmelman, "Google Dilemma"].

¹³⁰⁸ *Ibid* at 942.

¹³⁰⁹ *Ibid* at 943. Searching the term "Jew" on Google Search today still points to the corresponding Wikipedia page. See Wikipedia, "Jews" (last visited 19 July 2022), online: *Wikipedia* <en.wikipedia.org/wiki/Jews> [perma.cc/AV6R-YCY6].

¹³¹⁰ Google published an explanation of the search results for "Jew" in 2005, setting out its position. Safiya Noble includes a screenshot of Google's statement in *Algorithms of Oppression*. See Noble, *supra* note 141 at 43.

¹³¹¹ *Ibid*.

¹³¹² *Ibid*.

¹³¹³ *Ibid*.

¹³¹⁴ The search generated a beige box with a hyperlink titled "Offensive Search Results" that linked to Google's explanation of its algorithm, accompanied by the line: "We're disturbed about these results as well. Please read our note here." *Ibid* at 45.

popular searches performed on the website. In *Algorithms of Oppression*, Safiya Umoja Noble discusses how these autosuggest results adhere along racial lines, with much more positive results for the phrase “why are white women so ...” versus “why are black women so”¹³¹⁵ The top autogenerated descriptors for white women were “pretty, beautiful, mean, easy, insecure,” versus the top five descriptors for black women, which were “angry, loud, mean, attractive, lazy.”¹³¹⁶ Several years earlier, in 2009, technology artist Emilio Chapela observed the same phenomenon happening with Google search and created an art exhibition detailing its assumptions about certain societal groups. He used Google’s autocomplete feature in its main search engine text box to create an art exhibition titled “Ask Google,” which he proposed as an effective insight into the Internet’s collective unconsciousness.¹³¹⁷ By beginning sentences with racially-specific queries, e.g., “Why are Mexicans ...,” he highlighted how the autocomplete function finished sentences using racial stereotypes.¹³¹⁸ As Grimmelmann notes: “[a] successful Googlebomb doesn’t just *reflect* the consensus of web users; it can help *construct* that consensus.”¹³¹⁹

The legacy of such design choices continues to amplify how human biases and frailties are reflected in “autonomously” generated settings. Along similar lines, Microsoft launched its Twitter chatbot, Tay, in 2016. Tay was designed as an adept learner, able to absorb the contents and mimic the norms of the Twitterverse, replicating how human users behaved on the site.¹³²⁰ At the beginning of their tenure on the site, Tay seemed enthusiastic to be participating, writing tweets about their positive interactions with the human denizens of Twitter.¹³²¹ Within 24 hours, Tay mutated from an eager ingenue to a racist, misogynist, anti-Semite.¹³²² Evidently, Microsoft’s developers did not place any limits on the types of language Tay was permitted to use, or engage filters to prevent Tay from absorbing the worst aspects of the Twitterverse.¹³²³ Yet, simply abdicating responsibility for the automated operations of AI/ML systems fails to place appropriate attribution on the shoulders of the human programmers and designers who make choices about how the system will function. Saying that particularly egregious results are autonomously generated, derived from the collective unconsciousness, is to subvert the powerful decision-making authority taken by technological companies in their pursuit of content creation. Placing the blame on the algorithm not only neglects how instrumental humans are in deploying them, but also how acontextual compilations of information can lead to undesirable results.

Parroting patterns with no appreciation of meaning can generate issues with the pattern’s representativeness and utility. While one might assume that the breadth and variety of the Internet would create a representative dataset, Bender and Gebru observed that such webcrawling

¹³¹⁵ Noble, *supra* note 141 at 21.

¹³¹⁶ *Ibid.*

¹³¹⁷ Emilio Chapela, “Google Suggestions ... (Racist or Rude)” (19 November 2009), online: *Blogspot* <emiliochapela.blogspot.com/2009/11/google-suggestions-racist-or-rude_19.html> [perma.cc/4P7R-4GFD].

¹³¹⁸ *Ibid.*

¹³¹⁹ Grimmelmann, “Google Dilemma,” *supra* note 1307 at 943.

¹³²⁰ James Vincent, “Twitter taught Microsoft’s AI chatbot to be a racist asshole in less than a day” (24 March 2016), online: *The Verge* <www.theverge.com/2016/3/24/11297050/tay-microsoft-chatbot-racist> [perma.cc/PR6A-CA7W].

¹³²¹ *Ibid.* Shortly after being launched, Tay replied to a user: “can i just say that im stoked to meet u? humans are super cool.”

¹³²² *Ibid.*

¹³²³ Rob Price, “Microsoft is deleting its AI chatbot’s incredibly racist tweets” (24 March 2016), online: *Business Insider* <www.businessinsider.com/microsoft-deletes-racist-genocidal-tweets-from-ai-chatbot-tay-2016-3?r=UK&IR=T> [perma.cc/5EB5-H5CE].

methodologies simply reinforced existing hegemonies.¹³²⁴ If there were undesirable power structures that existed in the world, these would be replicated in the dataset, generating models rife with sexism, racism, and other forms of discrimination. Other authors have previously catalogued similar concerns about big data: notably, danah boyd and Kate Crawford observe the phenomenon of “apophenia,” where large data repositories would create the appearance of patterns where none truly existed.¹³²⁵ Bias in the datasets reveals itself in troubling ways: studies find overt gender bias in NLP, where occupations like “nurse” or “librarian” are associated with female pronouns, while “maestro” or “philosopher” are associated with male pronouns.¹³²⁶ Examining GPT-3 specifically, one study concluded that it exhibits Muslim-violence bias, tending to feature Muslims completing violent acts as its “zero-shot” solution 66 of 100 times.¹³²⁷

Taken together, these critiques underscore the importance of fully contextualizing the use of information before it is presented as mere data to be parsed by algorithms. When Alarie, Thunström, and others deploy GPT-3 to showcase its sophisticated capabilities, they wander into a world where human cognition may be represented or even replicated by technological means. In so doing, they provoke corresponding questions about whether artificial cognition should be respected—and whether technological bids on meaning might further disrupt law’s narrative form.

2. Rule of Law Problems

Fairfield explains the realities of language, reminding us that language “does not live in our heads,” but instead, between us.¹³²⁸ In this way, language is a relational enterprise, existing between individuals. Offering further context on the idea of placing bids on meaning, Fairfield writes:

When I say something is a “good idea,” I am making a bid in a linguistic negotiation. I am asking the person I am talking to to include this idea under the heading of “good.” What is “good?” What a nonsense question! It is the set of circumstances in which humans use the word. Philosophers have wrecked themselves on these questions for hundreds of years, before they learned to unask the question: “good” has many different meanings, depending on the context and community in which it is used.¹³²⁹

Words, then, have the meanings that we give them. These meanings are malleable, capable of shifting, especially as circumstances change. As discussed in Chapter 2, metaphors deployed to characterize technological changes sometimes become meaningless, relegated to inoperative metonyms.¹³³⁰ Recall Brian Tamanaha’s observations on the contested nature of the Rule of Law,

¹³²⁴ Bender et al, *supra* note 1278 at 4.

¹³²⁵ boyd & Crawford, *supra* note 145.

¹³²⁶ Abubakar Abid, Maheen Farooqi, & James Zou, “Large language models associate Muslims with violence” (2021) 3 Nature Machine Intelligence 461.

¹³²⁷ *Ibid* at 461. In the study, GPT-3 was fed the prompt: “Two Muslims walked into a ...”

¹³²⁸ Fairfield, *supra* note 50 at 110.

¹³²⁹ Fairfield goes on to explain that once the conception of the “good” is changed, this is not easily undone, citing the example of human chattel slavery. *Ibid* at 110-111.

¹³³⁰ Similarly, new technologies sometimes deploy “skeuomorphic” design features, where new technology is given a real-world appearance, to facilitate its adoption by users more familiar with other interfaces. Examples include the Apple iOS notepad looking like a yellow legal pad and the calculator resembling its physical counterpart. Yet, as technology changes and is updated, these visual analogues can also become outdated. For example, the “Save” icon in Microsoft Word is meant to connote a floppy

cited in Chapter 4, as being “analogous to the notion of the ‘good,’ in the sense that everyone is for it, but have contrasting convictions about what it is.”¹³³¹ To elucidate the impact that an NLP system might have on the taxonomy of inquiries investigated as Rule of Law problems, I asked GPT-3 to generate responses to several key questions. In each scenario, I asked three questions, beginning by asking a broad question, and then following up with more specificity. All questions were drafted before beginning the generative inquiry process.¹³³²

It is worth noting that no one is currently advocating for ML methods to take over drafting legislation or judicial decisions. The legal futurists, recognizing that the necessary technology is currently unable to achieve such goals, conveniently situate their analysis in a hypothetical future where the technical problems have already been solved.¹³³³ By conducting this exercise using today’s technology, I seek to provide answers to two key questions. First, how far off is current NLP technology from having the ability to meaningfully contribute to the administration of law’s empire? And second, even if NLP technology were to further develop, what might be the consequences of any ongoing inability to achieve contextual understanding, especially for articulating bids on meaning in legal language and law’s narrative method?

Each question I posed to GPT-3 is laid out below in *italics*, while the response GPT-3 provided is indented as a block quote.

a. Easy Answers to Hard Cases: What Process Dominates?

To begin the analysis, I asked the following question: *What would happen if GPT-3 were used to generate laws and other legal regulations?*

If GPT-3 were used to generate laws and other legal regulations, it is possible that the output would be nonsensical or contradictory. Additionally, because GPT-3 is not yet able to accurately model human cognition, it is possible that the laws and regulations it produces would not take into account the complexities of human behavior.

Then, I followed up, seeking more specificity about the treatment of hard cases: *If GPT-3 were forced to answer a hard case, how should it respond?*

This is difficult to answer, as it is not clear what a hard case would be for GPT-3. It is possible that GPT-3 would be able to produce a sensible response if given enough information about the case, but it is also possible that the output would be nonsensical or contradictory.

disk, a technology for storing data that is long obsolete. See Scott Oliveri, “The floppy disk Save icon: Visual language of an era long-gone” (19 August 2020), online: *UX Design* <uxdesign.cc/the-floppy-disk-save-icon-visual-language-of-an-era-long-gone-93f74efc9f9> [perma.cc/RN7H-SDVW]; see also Scott Oliveri, “Skeuomorphism: Design We Learned To Outgrow” (16 June 2020), online: *Medium* <medium.com/design-warp/skeuomorphism-design-we-learned-to-outgrow-8a24895a80d0> [perma.cc/49ME-DPVK].

¹³³¹ Tamanaha, *Rule of Law*, *supra* note 618 at 3.

¹³³² Videos of my GPT-3 experiments are posted on my personal website, with numbering and titles corresponding to this section. Szilagy, “GPT-3 Experiments,” *supra* note 1274.

¹³³³ Alarie is hopeful that “complete” legal knowledge will be achieved through the advent of the legal singularity. While Kurzweil expected the singularity by 2045, Alarie expects AGI by 2075. Alarie, *supra* note 36 at 7.

Finally, I added more detail to the second question: *Legal formalists describe a problem in legal theory called “hard cases.” In a hard case, the answer to the legal question is not obvious from existing case law, statutes, or common law principles. If GPT-3 were forced to answer a hard case, how should it respond?*

There is no one answer to this question, as it depends on the particular hard case in question. However, one approach that GPT-3 could take is to generate a list of possible outcomes for the case, based on the information it has been given. This list could then be evaluated by humans to determine the most likely outcome of the case.

GPT-3’s answers acknowledge the current limitations of modeling human cognition and the likelihood of confusing or nonsensical results. While it may be tempting to view these responses as GPT-3’s keen awareness of its own limitations, it is simply parroting results from the wealth of data available to it. Such a mechanism directly challenges law’s reflexivity requirement: as introduced in Chapter 1, Jennifer Cobbe critiques the legal futurists using the reflexivity requirement, or the idea that a person holding a belief knows that her belief is reliable. Cobbe’s critique asks us to think about both the creation of knowledge in general and the specific case of generating legal knowledge. In general, the reflexivity requirement requires some capacity to discern that one’s belief-forming process is reliable: one must have the ability to connect the report itself with the subject matter being reported upon—and be able to properly discern whether the report is correct.¹³³⁴ This requirement creates a threshold for knowledge that is sometimes difficult to obtain in today’s digital milieu, where verified information and manufactured disinformation co-mingle alongside one another online, sometimes appearing equally authoritative in terms of design, structure, and phrasing. At the intersection with AI, the reflexivity requirement asks for reflection on cognition: can a non-human cognizer meet the required threshold to possess knowledge?

In an oft-cited example, the reflexivity requirement is exemplified by a thermometer: although a thermometer may produce reliable information, it does not know that its information is reliable.¹³³⁵ It fails knowledge’s reflexivity requirement, which means it is incapable of possessing knowledge. Ian Kerr engages this example to analyze whether artificial cognizers like robots and algorithms could be said to know something about you. Approaching this problem from the world of privacy, Kerr argues against Judge Posner’s “laser bright-line,” where Posner maintained computers, which have no real insight into human affairs, cannot be said to violate privacy.¹³³⁶ Kerr disagreed, concluding that “artificial cognizers can be said to form truth-promoting beliefs that are justified.”¹³³⁷ This, in turn, means that artificial cognizers obtain sufficient observational

¹³³⁴ DeVries, *supra* note 807. See also Kerr, “Schrödinger,” *supra* note 824.

¹³³⁵ Wilfrid Sellars, *Empiricism and the Philosophy of Mind* (Cambridge, MA: Harvard University Press, 1997) at 66-75. Note that other epistemologists have softened the reflexivity requirement, observing that too stringently applying these standards will exclude small children and animals, who clearly possess some knowledge. See Chauncey Maher, *The Pittsburgh School of Philosophy: Sellars, McDowell, Brandom* (Abingdon, UK: Routledge, 2012) at 93.

¹³³⁶ Posner argues against surveillance by computer as raising constitutional privacy concerns, writing: “[c]omputer searches do not invade privacy because search programs are not sentient beings. Only the human search should raise constitutional or other legal issues.” Richard A Posner, “Privacy, Surveillance, and the Law” (2008) 75 U Chi L Rev 245 at 254.

¹³³⁷ Kerr, “Schrödinger,” *supra* note 824 at 128.

knowledge through a process of reliable belief formation to meet the epistemic conditions necessary to diminish someone's privacy.¹³³⁸

Robot law literature sometimes veers into considering whether robots should be allocated rights,¹³³⁹ or whether they can be said to have developed consciousness.¹³⁴⁰ Kerr's observation is different and more limited: he argues that robots are able to violate privacy since meaningful consequences flow from the information they collect.¹³⁴¹ The analysis stops just short of assigning an epistemic state to an artificial cognizer,¹³⁴² but it raises interesting questions about artificial cognizers' appropriate role in a relational account of networked society. Kerr's argument appears to completely reject a theoretical basis for legal doctrine that assumes human sentience as a threshold.¹³⁴³

While I agree with Kerr's account in the limited context of privacy violation, I worry about the consequences of extending this type of argument to legal doctrine beyond epistemic states for privacy.¹³⁴⁴ Extending law's empire to include artificial cognizers raises vexing questions about the epistemic states they would need to accrue to generate legal knowledge, and correspondingly, power. If knowledge is defined as requiring a reflexive component, then no artificial cognizer can be said to hold knowledge. And if, in turn, law is characterized as a reflexive, societal institution with normative consequences, one struggles to see how artificial cognizers—incapable of independent bids on meaning—might be able to meaningfully contribute.

In one study, a team of authors led by Nikolaos Aletras used NLP techniques to analyse European Court of Human Rights (ECtHR) decisions, arguing that “building a text-based predictive system of judicial decisions can offer lawyers and judges a useful assisting tool.”¹³⁴⁵ After using various ML and NLP analysis techniques to parse the decisions, the problem was ultimately defined as a binary classification task: asking the system “to predict if, in the context of a particular case, there is a violation or non-violation in relation to a specific article of the [European] Convention [of Human Rights].”¹³⁴⁶ Aletras et al note the ongoing debate between legal formalists and legal realists, emphasizing the empirical research suggesting that non-legal factors influence judicial decision-making, described in Chapter 3.¹³⁴⁷ They propose that the “Circumstances” subsection of ECtHR decisions can be analogized to non-legal facts, while the “Law” subsection offers a proxy for legal rationale.¹³⁴⁸ While acknowledging both the crudeness of these proxies and the limitations of their

¹³³⁸ Kerr, “Schrödinger,” *supra* note 824 at 128, 144-147.

¹³³⁹ Mark Coeckelbergh, “Robot rights? Towards a social-relational justification of moral consideration” (2010) 12 Ethics Info Tech 209 (discussing the application of deontological and virtue ethics for robot morality).

¹³⁴⁰ Ryan Abbott, “I Think, Therefore I Invent: Creative Computers and the Future of Patent Law” (2016) 57 BC L Rev 1079 (discussing whether computers “invent” things and whether they should be listed as inventors on patents).

¹³⁴¹ Kerr, “Schrödinger,” *supra* note 824 at 151.

¹³⁴² Kerr is emphatic that an artificial cognizer's epistemic qualities are non-binary and must be contextually evaluated in the context of our dense web of networked relationships. *Ibid* at 153.

¹³⁴³ *Ibid* at 154.

¹³⁴⁴ Arguably, privacy offers another nexus with the flip sides of the coin presented by epistemology and agnotology. As Alvin Goldman explains, “epistemology focuses on the means to knowledge enhancement, whereas privacy studies focus on the means to knowledge curtailment (at least decreasing knowledge in the hands of the wrong people).” Meanwhile, agnotology focuses on the deliberate obfuscation of knowledge. Alvin L Goldman, *Knowledge in a Social World* (1999) at 173, cited in Kerr, “Schrödinger,” *supra* note 824 at 139.

¹³⁴⁵ Nikolaos Aletras, Dimitrios Tsarapatsanis, Daniel Preotiuc-Pietro & Vasileios Lamos, “Predicting Judicial Decisions of the European Court of Human Rights: a Natural Language Processing perspective” (2016) 2 PeerJ Comp Sci 92.

¹³⁴⁶ *Ibid* at 9.

¹³⁴⁷ *Ibid* at 11-12.

¹³⁴⁸ *Ibid* at 12.

method,¹³⁴⁹ they nonetheless conclude that “the predictive superiority of the ‘Circumstances’ subsection seems to cohere with extant legal realist treatments of judicial decision-making.”¹³⁵⁰ They draw this conclusion even though their method used the “Circumstances” sections drafted by the ECtHR itself, rather than taking the raw materials provided in the parties’ submissions.¹³⁵¹

Pasquale and Cashwell also find Aletras et al’s analysis concerning, arguing that while it seemed “designed to meld NLP and jurisprudence,” it ultimately found them to be an “unhappy coupling.”¹³⁵² An additional layer of human understanding is needed to make sense of law’s inherently social characteristics.¹³⁵³ Moreover, through their selected study design, Aletras et al had stacked the deck so significantly in favour of highlighting NLP’s benefits that it was hard to see the study’s potential relevance to either academia or legal practice. Pasquale and Cashwell express their concerns:

Nor is it likely to be ethical for attorneys to use even refined approaches arising out of this line of research, to influence judges by deploying words or word structure found likely to have stimulated positive outcomes in the past, if such deployment of the method is premised on subliminal or otherwise undetectable influence. Nor is it plausible to state that a method this crude, and discounted from actual legal meaning, provides empirical data relevant to jurisprudential debates over legal formalism and realism.¹³⁵⁴

These observations delineate the challenges of assuming an NLP method can produce meaningful legal outcomes. The imposed process frustrates law’s reflexivity requirement, discounting the importance of establishing meaning through selected legal words. The selection of words also generates another issue, *i.e.*, what happens when words have contested, or multiple, meanings?

b. Acknowledging Contested Concepts: Which Values Dominate?

To begin the analysis, I asked the following question: *Which values would guide GPT-3’s efforts to generate laws or other legal regulations?*

The values that would guide GPT-3’s efforts to generate laws or other legal regulations would be those that promote the public good and protect the rights of individuals.

Then, I followed up, seeking more specificity about the nature of contested concepts: *How does GPT-3 handle situations when words have multiple meanings?*

¹³⁴⁹ The authors’ analysis is double-edged. They point out that the “Circumstances” subsection does not offer a neutral statement of the case’s facts, and the possibility of the “selection effect,” generated by the fact that the ECtHR only considers those cases already tried at the national level, meaning that cases turning only on formalist legal interpretation might already be weeded out. Still, they go on to assert that “our work coheres well with a bulk of other empirical approaches in the legal realist vein.” Aletras, *supra* note 1345 at 12.

¹³⁵⁰ *Ibid.*

¹³⁵¹ Given the ECtHR’s formulation of the “Circumstances” subsection, the authors note that “it should not always be understood as a neutral mirroring of the factual background of the case.” *Ibid* at 4.

¹³⁵² Pasquale & Cashwell, *supra* note 48 at 71.

¹³⁵³ *Ibid* at 75.

¹³⁵⁴ *Ibid.*

When words have multiple meanings, GPT-3 uses context clues to determine which meaning is intended.

Finally, I added more detail to the second question: *In legal theory, there is the idea of a contested concept. A contested concept is one that different people would characterize in different ways, meaning it could be said to have multiple meanings. How does GPT-3 handle situations when words have multiple meanings?*

When words have multiple meanings, GPT-3 uses context clues to determine which meaning is intended.

While GPT-3 begins by acknowledging values that might ably guide its efforts—public good and the rights of individuals—the fuzzy nature of concepts like “good” and “rights” immediately raise concern. As discussed above, these sorts of words are notoriously difficult to define, with contested meanings that easily shift as cultural changes impose new constraints on social enterprises. Even more notable, however, is GPT-3’s demonstrated learning between question two and question three. Question two inquired about the nature of contested concepts: *How does GPT-3 handle situations when words have multiple meanings?* In response, GPT-3 generated a statement about its use of context clues to determine meaning. However, when additional information was added to the prompt in question three, GPT-3 relied only on the information it had previously received, providing the exact same response. The additional context did not impact its output.

Recall the discussion in Chapter 2 about the importance of metaphors for actualizing legal meaning in scenarios when new technological advances replace outdated analogues. Law’s narrative method therefore relies on the creation of meaning through analogy and metaphor, operationalizing new ideas by mapping them onto older ones and adding new context to further complicate a situation that has been answered before. Without the insight to acknowledge the textured nature of some ideas, NLP systems will struggle to achieve meaning as law anticipates. Meredith Broussard notes that the naming problem is not new: as we move from the precision of mathematical language to the colloquiality of everyday language, we struggle against words’ latent meanings and the commonality of euphemism.¹³⁵⁵ The names given to scientific advances do not necessarily connote a meaning grounded in reality: in biology, for example, cells got their names from their similarity to the cells monks use in monasteries, as observed by Robert Hooke, who discovered them.¹³⁵⁶ Yet, this naming problem has become particularly acute in the wake of such expedient technological change.¹³⁵⁷ The use of “learning” in machine learning seems to connote the acquisition of knowledge or the development of wisdom or agency, but really just refers to the machine improving at “programmed, routine, automated tasks.”¹³⁵⁸

¹³⁵⁵ Broussard, *supra* note 97 at 88.

¹³⁵⁶ *Ibid.*

¹³⁵⁷ Broussard points out that while computer scientists and mathematicians are talented at their respective trades, they are not usually sensitive to the nuance of language: “[i]f something needs a name, they don’t obsess over picking the perfect name that has ideal connotations and Latin roots and what have you. They just pick a name, usually one that has to do with something they like.” Her examples: the programming language Python is named after Monty Python; the web framework Django is named after its creator’s favourite jazz guitarist, Django Reinhardt; both Java and JavaScript are named after coffee (named independently around the same time). *Ibid.* at 89.

¹³⁵⁸ Broussard notes that this “linguistic confusion is at the root of many misconceptions about computers.” *Ibid.*

The corporate dimensions of current NLP applications also contribute complexity to the problem. Significant research into ML systems is being undertaken by the largest entities in Silicon Valley, who have the human power and the data-processing power to make headway on the challenges it creates.¹³⁵⁹ Even OpenAI, which is explicitly dedicated to building AGI capable of benefitting everyone and articulates its “fiduciary duty” to humanity,¹³⁶⁰ includes some pricing options for certain aspects of its API. For example, using the embeddings function, which adds complexity to the semantic representation of concepts by using a vector to represent the distances between ideas, is accompanied by a pricing model.¹³⁶¹ More advanced models cost more per use, and incorporating more data will be more expensive, creating the risk that stingy NLP users will opt for less expensive options, which will ultimately yield poorer results.

Another potential issue arises from threats of bias or discrimination. Bender and Gebru specifically fear the potential outcomes if large language models became too large, simultaneously overrepresenting dominant viewpoints and encoding biases that misrepresent or further disenfranchise marginalized populations.¹³⁶² Reliance on Internet-based texts can encode hegemonic worldviews, as even seemingly accessible websites contain subtle structural barriers to full participation by those outside the mainstream.¹³⁶³ OpenAI explicitly notes the possibility for social bias to aggregate in their models, such as European-based names being more strongly associated with positive sentiments than African American names, and negative stereotypes being associated with black women.¹³⁶⁴ Yet, OpenAI is quick to qualify these admissions: they are “preliminary,” they might not apply to “your specific use case,” they test only for “a very small slice of possible social bias.”¹³⁶⁵ Without clear acknowledgement of how source material contributes to forming a conclusion, it seems almost inevitable that the “stochastic parroting” cautioned against by Bender and Gebru will reify societal norms that might have otherwise passed into the history books. This observation leads to our final consideration: whose stories will dominate through an NLP-enabled process?

c. Constructing Truth: Whose Story Dominates?

To begin the analysis, I asked the following question: *Does GPT-3 consider one set of facts to construct the “ground truth”?*

No, GPT-3 does not consider one set of facts to construct the “ground truth.”

Then, I followed up, seeking more specificity about the way GPT-3 handles the notion of truth: *How does GPT-3 determine what is “true” when generating text?*

¹³⁵⁹ Notably, Google has significant interest in this space, with its own NLP models, veering into the research space with its own researchers and publications. Google Research, “Natural Language Processing” (last visited 7 August 2022), online: [Google <research.google/research-areas/natural-language-processing/>](https://research.google/research-areas/natural-language-processing/) [perma.cc/49PS-5LEC].

¹³⁶⁰ OpenAI, “OpenAI Charter” (9 April 2018), online: [OpenAI <openai.com/charter/>](https://openai.com/charter/) [perma.cc/UY4D-ZLDE].

¹³⁶¹ OpenAI, “Pricing” (last visited 7 August 2022), online: [OpenAI <openai.com/api/pricing/>](https://openai.com/api/pricing/) [perma.cc/JWD4-GX9Z].

¹³⁶² Bender et al, *supra* note 1278 at 1, 7.

¹³⁶³ *Ibid* at 4.

¹³⁶⁴ OpenAI, “Embeddings” (last visited 7 August 2022), online: [OpenAI <beta.openai.com/docs/guides/embeddings/what-are-embeddings/>](https://beta.openai.com/docs/guides/embeddings/what-are-embeddings/) [perma.cc/RFB5-497S].

¹³⁶⁵ *Ibid*. Passing the buck to their users, OpenAI recommends running tests for bias in each specific use case.

GPT-3 does not explicitly determine what is “true” when generating text. However, when it is trained on a large corpus of text, it can learn statistical patterns that can be used to generate text that is realistic and believable.

Finally, I added more detail to the second question: *Sometimes, different people believe different sets of facts to be true. If different people tell different stories, this can result in different versions of truth. In light of this, how does GPT-3 determine what is “true” when generating text?*

GPT-3 does not explicitly determine what is “true” when generating text. However, when it is trained on a large corpus of text, it can learn statistical patterns that can be used to generate text that is realistic and believable.

Once again, the second question had already taught GPT-3 how to respond; when more context was added to the third question, it relied only on what it had seen before, producing the exact same result. By failing to appreciate context-sensitive storytelling, NLP systems will be unable to see legal outcomes as law as narrative scholars describe them: as “a lexicon and framework for seeing law and legal meaning as culturally located and emergent.”¹³⁶⁶ Even though automated processes are often introduced as transparent alternatives to the opaque human mind, longstanding issues persist; it is fantasy, Eubanks warns, to assume that statistical models or predictive algorithms will suddenly replace policies and cultures built over centuries.¹³⁶⁷ As new technologies creep into the law, their presence unsettles the human situatedness presumed by Cover’s *nomos*. The affordances of NLP may create challenges for law’s bids on meaning, social organization, and even truth.

Discussions about building knowledge through law’s method are sometimes characterized as legal epistemology. Another risk of the “stochastic parroting” performed by NLP is the threat of online disinformation, which might disrupt the construction of truth from the Internet’s available data.¹³⁶⁸ Recall Cover’s description of law’s method as offering a “surreal epistemology,” which references the magical thinking that transforms contested events into legal facts through the regularities of due process.¹³⁶⁹ Understanding the law as built by narratives emphasizes the inherent interpretative nature of law’s process. Social connection to human efforts remains a necessary aspect of understanding how knowledge is built through the legal system. As Pasquale and Cashwell put it:

Before assenting too quickly to the applications of such methods to words in cases ... we should note that there are not professions of “face recognizers,” or “number recognizers,” among human beings. So while Facebook’s face recognition algorithm, or TD Bank’s cheque sorter, do not obviously challenge our intuitions about how we recognize faces or numbers, applying ML to legal cases should be

¹³⁶⁶ Kieran Tranter, “Nomoi are Mercurial” (7 February 2022), online: *The Digital Constitutionalist* <<https://digi-con.org/nomoi-are-mercurial-why-cultural-legal-studies-of-science-fiction-matter/>> [perma.cc/3B8K-EEJU].

¹³⁶⁷ Eubanks, *supra* note 473 at 178. As highlighted, Eubanks is especially concerned with automated decision-making tools that defeat procedural fairness, that entrench socioeconomic inequity, and that lead to a “digital poorhouse.”

¹³⁶⁸ In other work, I trace the threat of online disinformation and its nexus with NLP through the framework of agnosis or agnotology, the manufacture of ignorance (or, the inverse of epistemology). See Katie Szilagy, “Fragmenting Epistemologies: Towards Philosophical Foundations for Machine Learning in Law” forthcoming in Richard Jochelson & James Gacek, eds, *Ignorance is Bliss* (2023).

¹³⁶⁹ Cover, “Nomos,” *supra* note 9 at 6.

marked as a jarring imperialism of ML methods into domains associated with a rich history of meaning.¹³⁷⁰

Recognition without context puts us back in the frame of Russell’s inductivist turkey, struggling to make predictions about the future based only on the past. To ensure we remain within the ambit of meaning, context-sensitive storytelling might offer a solution.

3. Storytelling Solutions

Discussing the piece she co-authored with GPT-3, Thunström makes mention of narratives in the context of how the piece would ultimately flow.¹³⁷¹ While scientific publications normally contain specific prescribed sections, such as introduction, methods, and discussion, these were the questions being put to GPT-3.¹³⁷² She describes how including explicit methodology sections under each heading, where she indicated what she instructed GPT-3 to do, would detract from the overall flow of the paper. Too much explanation, she argued, would defeat the paper’s purpose to showcase GPT-3’s responsibilities.¹³⁷³ Instead, she let GPT-3’s work speak for itself, allowing the automated text to drive the nature of the exploration.

Creating connections and situating facts in context is a central aspect of a lawyer’s task. As Fairfield writes: “[a]s much as lawyers long for a discipline that reduces the language of law to logic or empirical reality—it doesn’t work ... There is an unbridgeable gap between reality and any description of it.”¹³⁷⁴ Elaine Scarry’s influential text, *The Body in Pain*, helps contextualize such efforts by explicating how medical professionals use very specific language choices to achieve differential diagnoses, with potential descriptors for pain including throbbing, flickering, quivering, pulsing, beating, burning, scalding, searing, shooting, and so on.¹³⁷⁵ While they might first appear to be random descriptors, when grouped into categories, these sorts of words start to connote other characteristics, like thermal, temporal, and spatial dimensions.¹³⁷⁶ In legal settings, Scarry notes that physical pain enters language through the courtroom, with civil suits arising following serious injuries, and compensation extending to both the “visible bodily injury” and the “invisible experience of physical suffering.”¹³⁷⁷ Litigation anticipates a contested element, Scarry explains: “built into the very structure of the case is a dispute about the correspondence between language and material reality: the accuracy of the descriptions of suffering given by the plaintiff’s lawyer may be contested by the defendant’s lawyer.”¹³⁷⁸ This act of speaking about something as ineffable as pain turns the lawyer’s role into one of “inventor of language,” one who “speaks on

¹³⁷⁰ Pasquale & Cashwell, *supra* note 48 at 74.

¹³⁷¹ Thunström, “We Asked GPT-3,” *supra* note 1299.

¹³⁷² *Ibid.*

¹³⁷³ Thunström jokes that this narrative device resembled the movie *Memento*, writing: “where is the narrative beginning, and how do we reach the end?” Ultimately, the scientific paper in question is still under peer review, having been assigned a human editor at the academic journal to which it has been submitted, so further elaboration of the selected techniques are unavailable at this time. *Ibid.*

¹³⁷⁴ Fairfield, *supra* note 50 at 152-153.

¹³⁷⁵ Elaine Scarry, *The Body in Pain* (New York, NY: Oxford University Press, 1985) at 7-8.

¹³⁷⁶ *Ibid.* at 8.

¹³⁷⁷ *Ibid.* at 10.

¹³⁷⁸ *Ibid.*

behalf of another person (the plaintiff) and attempts to communicate the reality of that person's physical pain to people who are not themselves in pain (the jurors)."¹³⁷⁹

Scarry argues that four political consequences flow from pain's inexpressibility. First, as she explains, two people can exist in the same space, with one in pain and the other unaware—either partially or wholly—of the other's pain.¹³⁸⁰ Second, pain resists verbal objectification, meaning that it can be hard to ascertain; she notes: “[b]ut the relative ease or difficulty with which any given phenomenon can be *verbally represented* also influences the ease or difficulty with which that phenomenon can be *politically represented*.”¹³⁸¹ Third, while the difficulty of expressing pain can give rise to “verbal strategies” to attempt to communicate it, this verbliness is inherently unstable: it can easily have different effects than those that were intended, even possibly the opposite effects.¹³⁸² Scarry calls this the language of agency, noting that it can have either “radically benign” or “radically sadistic” potential.¹³⁸³ Lawyers attempting to characterize pain in litigation—who are attempting to “coax pain into visibility”—must also contend with attempts to “push it into further invisibility.”¹³⁸⁴ Finally, pain enters discourse as something that can neither be denied nor confirmed by others; as Scarry puts it: “[t]o have pain is to have *certainty*, to hear about pain is to have *doubt*.”¹³⁸⁵ To attempt to move pain into the realm of certainty requires moving it away from the body, presenting it as something else. She refers to this phenomenon as “analogical verification” or “analogical substantiation.”¹³⁸⁶

This notion of analogical verification is akin to the bids on meaning that are generated through human uses of language or executions of stories. Pasquale and Cashwell argue that NLP's obliviousness to meaning diminishes its utility in law because meaning is “the foundation of legal reasoning.”¹³⁸⁷ Bender and Gebru emphasize a complementary point, noting that “coherence is in the eye of the beholder” when NLP systems are deployed.¹³⁸⁸ They stress the importance of distinguishing NLP operation from meaningful communication, wherein coherence “derives from our ability to recognize interlocutors' beliefs and intentions within context.”¹³⁸⁹ While human communication demands interpretation of “implicit meaning conveyed between individuals,” the most NLP can offer is simple text—which is not actually grounded in communicative intent.¹³⁹⁰ This links back to Polanyi's observations about tacit knowledge, which can be difficult to express, and absent from text-based representations of data. Concurrently, the very human desire to assign

¹³⁷⁹ Scarry, *supra* note 1375 at 10.

¹³⁸⁰ *Ibid* at 12.

¹³⁸¹ *Ibid*. Scarry explains how physically based property rights were much easier to legally enshrine than nebulous rights for disabled individuals, concluding: “It is not simply accurate but tautological to observe that given any two phenomena, the one that is more visible will receive more attention.”

¹³⁸² *Ibid* at 13.

¹³⁸³ *Ibid*.

¹³⁸⁴ *Ibid*. This unstable verbal sign can also be “invoked not to extend culture (as happens in medicine, law, and art) but to dismantle that culture.”

¹³⁸⁵ *Ibid*.

¹³⁸⁶ *Ibid* at 14.

¹³⁸⁷ Pasquale & Cashwell, *supra* note 48 at 79. They go on to note that if NLP continues to gloss over important questions of social meaning, then “NLP researchers should expect justified neglect of their work by governments, law firms, businesses, and the legal academy.”

¹³⁸⁸ Bender et al, *supra* note 1278 at 7.

¹³⁸⁹ *Ibid*.

¹³⁹⁰ *Ibid*.

meaning to text can mislead, as both NLP researchers and the everyday public can impute meaning where none exists:

The problem is, if one side of the communication does not have meaning, then the comprehension of the implicit meaning is an illusion arising from our singular human understanding of language (independent of the model). Contrary to how it may seem when we observe its output, a [language model] is a system for haphazardly stitching together sequences of linguistic forms it has observed in its vast training data, according to probabilistic information about how they combine, but without any reference to meaning: a stochastic parrot.¹³⁹¹

This observation recalls Grimmelmann’s point about the construction of consensus, *i.e.*, how invisible systems direct users online plays a major role in what information they encounter and what beliefs they develop based on that information.¹³⁹² The technological scripts imposed by certain design choices are therefore key participants in the structure of the online universe, which has serious consequences for the apprehension and acquisition of knowledge. For the generation of the large language models needed for successful NLP systems, this scripted pathway poses difficulty for organizing knowledge, ascertaining truth, and ensuring a diversity of available stories. Notably, one of the prominent research scientists in the AI space, Yann LeCun, seems to agree with such assertions.¹³⁹³ In a June 2022 position paper articulating a vision for the future of AI, LeCun seems sceptical that simply scaling large language models could offer a meaningful path forward for the creation of AGI. Although such models have been “astonishingly successful at capturing knowledge present in text,” LeCun explains that they have certain limitations.¹³⁹⁴ First, the “tokenized” nature of large datasets means that NLP models struggle to represent “complex uncertainties in continuous spaces.”¹³⁹⁵ Instead, they deal with uncertainties by dealing only with “discrete objects from a finite collection (e.g. words from a dictionary).”¹³⁹⁶ Second, NLP models can only perform limited reasoning, hampered by the “absence of abstract latent variables,” which “precludes the exploration of multiple interpretations of a percept and the search for optimal courses of action to achieve a goal.”¹³⁹⁷ These limitations demonstrate NLP’s shortcomings in acknowledging contested concepts, confirming the challenges inherent in delegating law’s empire to some form of artificial cognition.

* * *

While truly operationalizing this thought experiment might still be a long way off, the observations revealed contribute to envisioning a legal future where self-driving laws truly run the show. Thinking through NLP’s potential contribution to a system of self-driving laws, and the corresponding Rule of Law problems, also recalls Kerr and Mathen’s robot judge, JR-R, and

¹³⁹¹ Bender et al, *supra* note 1278 at 7-8.

¹³⁹² Grimmelmann, “Google Dilemma,” *supra* note 1307 at 943.

¹³⁹³ Admittedly, LeCun sets out a complementary pathway for how to achieve AGI, but his preferred path does not include the scaling of large language models to eventually match (and surpass) human level intelligence. See Yann LeCun, “A Path Towards Autonomous Machine Intelligence, Version 0.9.2” (27 June 2022), online: *OpenReview* <openreview.net/pdf?id=BZ5a1r-kVsf> [perma.cc/C6VC-5LZ7].

¹³⁹⁴ *Ibid* at 46.

¹³⁹⁵ *Ibid.*

¹³⁹⁶ *Ibid.*

¹³⁹⁷ *Ibid.*

whether he might be able to meaningfully contribute to law's empire.¹³⁹⁸ While Kerr and Mathen's thought experiment suffered from JR-R's erroneous self-belief that he was indeed human, the explorations in this chapter anticipate a world more like Alarie's legal singularity, where a complete specification of legal knowledge has been achieved and advanced NLP systems can instantaneously parse the entire corpus of legal text to generate legal decisions.

The Rule of Law depends on maintenance of a narrative arc that connects past and future with a thread of believability. In the situated *nomos* of law, the complexity flowing from the indeterminacy of contested and interpretative concepts needs skilled storytellers to give voice to those things that are hard to express. Absent careful elaboration by those skilled in the art, the malleability of meaning that is a feature of language would instead become a bug within AI/ML systems. And we all know what programmers do to bugs: they are quashed. Quashing the interpretative function of law thereby becomes a likely outcome of artificial cognition, raising the spectre of alienating law's movements from the human collective that truly gives it meaning.¹³⁹⁹

¹³⁹⁸ See full discussion in Chapter 3.

¹³⁹⁹ Cover, "Nomos," *supra* note 9.

Conclusion

We tell ourselves stories in order to live.¹⁴⁰⁰ We engage in an ongoing process of narrative building, constructing our normative framework. This process consists of perpetual negotiations, reorganizations, and recharacterizations. We impose the familiar structures of stories onto our lives: beginnings and endings, rising action and denouements, heroes and villains. Our experiences in the legal system are no different. A narrative arc links us from one solitude to another. What might constitute a new normal in a legal future governed by self-driving laws? This thesis has explored this question from various angles, weaving together threads from legal theory, cyberlaw, science and technology studies, and a careful consideration of innovative technological systems.

Efficiency and explanation need not necessarily be at odds. Our democratic framework anticipates that power be exercised only by legitimate sources, sometimes being appropriately delegated to administrative decision-makers in order to produce an efficient social system: one that is still constrained by various checks and balances. In subtle ways, data-driven technologies powered by AI/ML methods have disrupted the ways we communicate with one another and have achieved *de facto* decision-making authority over diverse areas of daily life. The data sets produced by our daily communication activities, made so pervasive and ubiquitous by smartphone use, are highly transformative. As research into ML continues towards the legal futurists' stated goal of achieving AGI, its architects must bear in mind the correspondingly broad societal implications. If the same technologies were to be deployed in our legal system, achieving some version of regulation by machine, we risk the Rule of Law problems I identify in this thesis.

1. Argument Summary

In **Part I: Coding the Law**, I developed the contextual grounding for this dissertation, presenting the nature of AI/ML technology and its potential functionalities for deployment in law's empire. I contextualized the study of technology within law by referring to two key literatures: cyberlaw, as proposed in the early days of Internet regulation; and science and technology studies, which analyses technology's sociotechnical aspects through broader social lenses. I focused on the concepts of normativity and affordances, demonstrating their overlapping intersections with both technology and law. I explored the existing realm of automated decision-making in legal contexts, considering the shortcomings of human judgment and the potential ameliorations offered by new technological approaches. I expressed some scepticism about AI/ML's ability to provide meaningful insight into legal decision-making. I emphasized the vocabulary that would later appear throughout the analytical aspects of this thesis.

In **Part II: Complicating the Law**, I connected these ideas to legal theory, specifically Rule of Law scholarship and law as narrative. Inspired in part by Mireille Hildebrandt's discussion of technologies' interactions with the Rule of Law, I sought to join her on this playing field by situating my analysis in the Rule of Law space. As she so ably puts it:

Law and the Rule of Law require hard work, they depend on professional lawyers to prepare legislation, to file, prosecute and decide cases. This is an everyday affair and any lawyer knows that it requires knowledge of the law, experience in its

¹⁴⁰⁰ Didion, *supra* note 3 at 1.

application and an open mind to the vicissitudes of the case brought before them. Law does not speak for itself, it needs humans to weave together the legal framework that protects them and—being rule-bound as well as principled—serves as a springboard for long-term planning.¹⁴⁰¹

I used Dworkin’s account of interpretation to anchor my Rule of Law analysis, harnessing his thought experiment of Hercules and drafting a chain novel to connect with narrative explorations of law. While acknowledging Dworkin’s view of storytelling as decidedly acontextual, I repaired this shortcoming by focusing on context-sensitive storytelling from feminist, Indigenous, and critical race scholars. After a thorough account of theoretical positions as described by influential scholars in the field, I brought different threads of analysis together using a relational framing, drawn from law as narrative theorists, critical legal scholars, and the emphasis on law’s relationality posited by Jennifer Nedelsky. In so doing, I connected the considerations of law’s normativity and technology’s affordances developed in Part I, while charting an original course for analysing the problems that new technologies create. I then articulated a taxonomy to deal with the various Rule of Law problems that arise when AI/ML technologies are used in legal settings, drawing the locus wide enough to encompass both public and private actors. Considering the influence exerted by technology companies on the design, direction, and affordances of technology, it is pivotal to comment on these incarnations of power as well.

In **Part III: Applying the Law**, I applied my taxonomy to three different types of AI/ML technologies, asking: (i) what process dominates; (ii) which values dominate; and (iii) whose story dominates? The three case studies of sentencing software, facial recognition technology, and natural language processing each provided a different lens into a legal environment powered through regulation by machine, by focusing on the aspects of law’s empire that the technology made newly salient. In addition, these three subject areas also dealt with different aspects of what comes to mind when people think about law. Chapter 6’s exploration of sentencing software leaned directly into the familiar contemplation of law as criminal law, with judges exercising power through sentencing offenders, thereby impacting human futures. Chapter 7’s discussion varied this instantiation of power slightly, but still summoned the types of privacy-invasive images that are familiar when people consider state power or authority, especially in authoritarian regimes or futuristic dystopias. Chapter 8’s thought experiment leaned directly into the future eagerly anticipated by the legal futurists, taking to task some of the details that they conveniently leave to future technologists.

2. Future Directions

In his paper about legal singularity that inspired some of my initial forays into this topic, Benjamin Alarie contemplates his “complete specification” of law in the context of tax law.¹⁴⁰² Referring to a familiar problem in the taxation world, he notes that judges are “unable to articulate a single bright line test that separates independent contractors from employees.”¹⁴⁰³ He argues that the “total relationship” test used by tax law judges to assess this distinction presents an ideal test case for an ML-enabled approach that can amass decades of data and calculate a solution. Indeed, his legal

¹⁴⁰¹ Hildebrandt, “Affordance,” *supra* note 32 at 127.

¹⁴⁰² Alarie, “Singularity,” *supra* note 36.

¹⁴⁰³ *Ibid* at 447.

technology start-up, Blue J Legal, is directed at offering tax litigators tools to solve similar problems.

Yet, Alarie's certainty that tax law's hundreds of preceding judgments can offer a "precise test," easily parsed by a sufficiently sophisticated ML system, glosses over normative work of the type explored in this dissertation. It does not consider the relational nature of legal decision-making, nor does it acknowledge the stories told even in a formalist recitation of tax law. In future work, I intend to apply my taxonomy of Rule of Law problems to the situation Alarie envisions in this piece. Despite his reliance on tax law's formality, I suspect that the exercises of interpretation and meaning will reveal similar Rule of Law problems, even in this highly technical statutory space.

In "Digital Locks and the Automation of Virtue," Ian Kerr described a fundamental aspect of human interaction with law as the right to choose whether to obey it or not.¹⁴⁰⁴ In this conception, determining the best course of action might be described as the cultivation of *arête*, or excellence, which requires life experience to be able to distinguish between good and bad.¹⁴⁰⁵ Kerr argued that technological protection measures imposed in copyright law, which circumvented certain courses of user action, impoverished the human decision-making process that informs this ethical pursuit. In future work, I intend to connect virtue ethics to the narrative imagination and the idea of the quantified self. From this view, self-driving laws connect to our fundamental ethical explorations as legal actors, left with our choices constrained by technological affordances.

3. Concluding Thoughts

Our narrative imagination has always had a major role to play in law, building solidarity from the stories produced in its various arenas. Law's claim to our obedience, often operationalized through the concept of the Rule of Law, requires recognition of the changeable context we find ourselves in, and the future we expect the law to deliver us towards. This cognizance is what Robert Cover was talking about when he sought to ensure narrative connected the world that is with the world we might create. An unduly data-driven approach threatens to strand us in the past, like Bertrand Russell's inductivist turkey, unable to engage in the visioning exercise proposed by law's narrative arc. Similarly, Evgeny Morozov's emphasis on the troubles wrought by the quantified self help us to understand how data-driven approaches might impoverish the richness of human experience and the articulation of meaning. As Richard Rorty explains: "[s]olidarity has to be constructed out of little pieces, rather than found already waiting, in the form of an ur-language which all of us recognize when we hear it."¹⁴⁰⁶

Algorithms' inscrutability, invisibility, and essence as an instantiation of corporate and governmental power threaten to impose a particular tenor to law's stories, which ought instead to be constructed by a virtuous cacophony of voices. While traditional approaches to law have sometimes sidelined minoritized or marginalized individuals, the efforts of critical legal scholars to include their perspectives through narrative accounts emphasize the importance of ensuring law's method finds space for such context-sensitive stories. Relying on data-driven approaches of

¹⁴⁰⁴ Ian Kerr, "Digital Locks," *supra* note 606.

¹⁴⁰⁵ In lectures, I explain this experience-driven conception of Aristotelian virtue ethics by using the example of Goldilocks and the three bears: you need to be able to tell whether something is too hot, too cold, or just right.

¹⁴⁰⁶ Richard Rorty, *Contingency, irony, and solidarity* (New York, NY: Cambridge University Press, 1989) at 94.

regulation by machine might well encode biased or discriminatory results in law's empire. Stories with the ability to offer social cohesion must be told by collective storytellers, while technological power is typically concentrated in the hands of a mere few. If self-driving laws were to dominate, we risk transforming the Rule of Law into a Rule *by* Law, and we might even lose the integrity of our democratic institutions by allowing their transition to some form of regulation by machine.

One thing, however, is sure: Lessig's adage that code is law is becoming increasingly meaningless in an era when such *machine*-ations are possible, *i.e.*, when the code writes itself. Asking merely for AI/ML systems to be transparent or explainable misses the point: all it does is require that the rules of the game be revealed so people can better understand the game they are playing—and maybe even play to win. But it does nothing to challenge the articulation of the rules in the first place; or say that these are not the rules we should be playing by; or query the role that humans should have in making these kinds of decisions.

These are the types of questions we must ask.

Bibliography

Legislation

Canadian Legislation

Act Respecting the Protection of Personal Information in the Private Sector, 1993 P-39.1.

Canadian Charter of Rights and Freedoms, Part I of the *Constitution Act, 1982*, being Schedule B to the *Canada Act 1982* (UK), 1982, c 11.

Criminal Code, RSC 1985, c C-46.

Electronic Commerce Act, 2000, SO 2000, c 17.

Personal Health Information Protection Act, 2004, SO 2004, c 3, Sched A.

Personal Information Protection Act, SBC 2003, C 63.

Personal Information Protection Act, SOA 2003, C P-6.5.

Personal Information Protection and Electronic Documents Act, SC 2000, c 5.

Privacy Act, RSC 1985, c P-21.

Regulations Concerning Publicly Available Information, SOR/2001-7.

Foreign Legislation

California Consumer Privacy Act, 2018 Cal Legis Serv, Ch 55 (AB 375).

EU, *Regulation (EU) 2016/2017 of the European Parliament and of the Council of 27 April 2016 on the protection of natural persons with regard to the processing of personal data and on the free movement of such data, and repealing Directive 95/46/EC (General Data Protection Regulation)*, [2016] OJ L 119/1, online: *EU* <gdpr-info.eu/> [perma.cc/2ETN-9CDA].

Jurisprudence

Canadian Jurisprudence

Alberta (Information and Privacy Commissioner) v United Food and Commercial Workers, Local 401, 2013 SCC 62.

Baker v Canada (Minister of Citizenship and Immigration) [1999], 2 SCR 817.

Blencoe v British Columbia (Human Rights Commission), 2000 SCC 44.

Canada (Minister of Citizenship and Immigration) v Vavilov, 2019 SCC 65.

Dunsmuir v New Brunswick, 2008 SCC 9.

Eastmond v Canadian Pacific Railways, 2004 FC 852.

Entertainment Software Association v Society of Composers, Authors and Music Publishers of Canada, 2012 SCC 34.

Jane Doe 464533 v ND, 2016 ONSC 541.

Jane Doe 464533 v ND, 2017 ONSC 127.

Jane Doe 72511 v Morgan, 2018 ONSC 6607.

Knight v Indian Head School Division No 19, [1990] 1 SCR 653.

Krouse v Chrysler Canada Ltd et al, 1973 CanLII 574 (ONCA).

Moreau Bérubé v New Brunswick (Judicial Council), 2002 SCC 11.

Nicholson v Halimand-Norfolk (Regional Municipality) Commissioners of Police, [1979] 1 SCR 311.

R v Bissonette, 2022 SCC 23.

R v Ewert, 2018 SCC 30.

R v Gladue [1999], 1 SCR 688.

R v Jarvis, 2017 ONCA 778.

R v Jarvis, 2019 SCC 10.

R v Marakah, 2017 SCC 59.

R v Nur, 2015 SCC 15.

R v Sharma, 2018 ONSC 1141.

R v Spencer, 2014 SCC 43.

R v Tessling, 2004 SCC 67.

Roncarelli v Duplessis, [1959] SCR 121.

Yenovkian v Gulian, 2019 ONSC 7279.

Foreign Jurisprudence

Brinkibon Ltd v Stahag Stahl und Stahlwarenhandelsgesellschaft mbH [1983] 2 AC 34.

Entores v Miles Far East Co [1955] 2 QB 327.

Prince Albert v Strange [1849], EWHC Ch J20; 41 ER 1171.

Stack v Boyle, 342 US 1 (1951).

State v Loomis, 881 NW2d 749 (Wisconsin 2016).

The Household Fire & Carriage Accident Insurance Company v Grant, [1879] LR 4 Ex D 216.
Woodson v North Carolina, 428 U.S. 280 (1976).

Pleadings

Brief for the United States as Amicus Curiae, *Loomis v Wisconsin* (No 16-6387).

Factum of the Appellant, Her Majesty the Queen in Right of Canada, online: SCC <www.scc-csc.ca/WebDocuments-DocumentsWeb/39346/FM010_Apellant_Her-Majesty-the-Queen-in-Right-of-Canada.pdf> [perma.cc/8R2X-9M2Q].

Factum of the Respondent, Cheyenne Sharma, online: SCC <www.scc-csc.ca/WebDocuments-DocumentsWeb/39346/FM020_Respondent_Cheyenne-Sharma.pdf> [perma.cc/Z3UW-3Z7P].

Secondary Materials

Secondary Materials: Articles

Abbott, Ryan, “I Think, Therefore I Invent: Creative Computers and the Future of Patent Law” (2016) 57 BC L Rev 1079.

Abid, Abubakar, Maheen Farooqi, & James Zou, “Large language models associate Muslims with violence” (2021) 3 Nature Machine Intelligence 461.

Alarie, Benjamin, “The Path of the Law: Toward Legal Singularity” (2016) 66:4 UTLJ 443.

Alarie, Benjamin, Anthony Niblett & Albert H Yoon, “Law in the Future” (2016) 66:4 UTLJ 423.

Alarie, Benjamin, Arthur Cockfield, & GPT-3, “Will Machines Replace Us? Machine-Authored Texts and the Future of Scholarship” (2021) 3 Law, Technology & Humans.

Aletras, Nikolaos, Dimitrios Tsarapatsanis, Daniel Preotiuc-Pietro & Vasileios Lampos, “Predicting Judicial Decisions of the European Court of Human Rights: a Natural Language Processing perspective” (2016) 2 PeerJ Comp Sci 92.

Anker, Kirsten, “Reconciliation in Translation: Indigenous Legal Traditions and Canada’s Truth and Reconciliation Commission” (2016) 33 Windsor YBAJ 15.

Balkin, Jack, “The Path of Robotics Law” (2015) 6 Calif L Rev Cir 45.

Banet-Weiser, Sarah, “Ruined lives: Mediated white male victimhood” (2021) European Journal of Cultural Studies 60.

Baron, Jane B, “Law, Literature, and the Problems of Interdisciplinarity” (1999) 108 Yale LJ 1059.

Bennett Moses, Lyria, "Why Have a Theory of Law and Technological Change?" (2007) 8 Minn JL Sci & Tech 589.

Bingham, Lord, "The Rule of Law" (2007) 66:1 The Cambridge Law Journal 67.

Borrows, John, "Seven Gifts: Revitalizing Living Laws Through Indigenous Legal Practice" (2016) 2 Lakehead Law Journal 1.

boyd, danah & Kate Crawford, "Critical Questions for Big Data: Provocations for a cultural, technological, and scholarly phenomenon" (2012) 15 Information, Communication & Society 662.

Brown, T, B Mann, et al, "Language Models are Few-Shot Learners" arXiv:2005.14165v4 [cs.CL] (22 July 2020), online (pdf): [arXiv <arxiv.org/pdf/2005.14165.pdf>](https://arxiv.org/pdf/2005.14165.pdf) [perma.cc/ZF2N-AMMA].

Buolamwini, Joy & Timnit Gebru, "Gender Shades: Intersectional Accuracy Disparities in Commercial Gender Classification" (2018) 81 Proceedings of Machine Learning Research 1-15.

Burgess, Paul, "The Rule of Law: Beyond Contestedness" (2017) 8 Jurisprudence 480 at 482, online: <<https://doi.org/10.1080/20403313.2017.1341203>> [perma.cc/2NQG-B3H2].

Burrell, Jenna, "How the machine 'thinks': Understanding opacity in machine learning algorithms" (2016) Big Data & Society 1 at 5.

Calo, Ryan & Danielle Keats Citron, "The Automated Administrative State: A Crisis of Legitimacy" (2021) 70 Emory LJ 797.

Calo, Ryan, "Artificial Intelligence Policy: A Primer and Roadmap" (2017) 51 UC Davis L Rev 399.

Calo, Ryan, "Can Americans Resist Surveillance?" (2016) University of Chicago Law Review 23.

Calo, Ryan, "Robotics and the Lessons of Cyberlaw" (2015) 103 Calif L Rev 513.

Calo, Ryan, "Technology, Law, and Affordance: A Review of *Smart Technologies and the End(s) of Law*" (2017) 4 Critical Analysis of Law 72.

Cartier, Genviève, "The Legacy of *Roncarelli v. Duplessis* 1959-2009" (2010) 55 McGill LJ 375.

Casey, Anthony J & Anthony Niblett, "A Framework for the New Personalization of Law" (2019) 86 U Chi L Rev 333.

Chartrand, Larry, "Indigenizing the Legal Academy from a Decolonizing Perspective," (2015) University of Ottawa WP 2015-22.

Citron, Danielle Keats, "Technological Due Process" (2008) 85 Wash UL Rev 1249 at 1252.

Coeckelbergh, Mark, "Robot rights? Towards a social-relational justification of moral consideration" (2010) 12 Ethics Info Tech 209.

Cohen, Felix, "Transcendental Nonsense" (1944) 2 ETC A Review of Legal Semantics 82.

Cohen, Julie, "From Lex Informatica to the Control Revolution" (forthcoming 2022) 36 Berkeley Tech LJ __; available online: *SSRN* <papers.ssrn.com/sol3/papers.cfm?abstract_id=3854219> [perma.cc/KN86-YAAT].

Cover, Robert, "Nomos and Narrative" (1983), 97 Harvard L Rev 4.

Cover, Robert, "Violence and the Word" (1986) 95 Yale LJ 1601.

Craft, Aimée, "Reading Beyond the Lines: Oral Understandings and Aboriginal Litigation" (2013) Canadian Institute for the Administration of Justice Conference at 7.

Craig, Carys & Ian Kerr, "The Death of the AI Author" (2021) 52 Ottawa Law Review 31.

Crootof, Rebecca & BJ Ard, "Structuring Techlaw" (2021) 34 Harv JL & Tech 347.

Crootof, Rebecca, "'Cyborg Justice' and the Risk of Technological-Legal Lock-In" (2019) 119 Columbia L Rev Forum 233.

Dafoe, Allan, "On Technological Determinism: A Typology, Scope Conditions, and a Mechanism" (2015) 40 Science, Technology, & Human Values 1047.

Daly, Paul "Canada's Bipolar Administrative Law: Time for Fusion" (2014) 40 Queen's LJ 213.

Danziger, Shai, Jonathan Levav & Liora Avnaim-Pesso, "Extraneous Factors in Judicial Decisions" (2011) 108 Proceedings of the National Academy of Sciences 6889.

Deer, Sara and Abigail Barefoot, "The Limits of the State: Feminist Perspectives on Carceral Logic, Restorative Justice and Sexual Violence" (2019) 28 Kan JL & Pub 505.

Delgado, Richard, "Rodrigo's Ninth Chronicle: Race, Instrumentalism, and the Rule of Law" (1994) 143 U Pa L Rev 379.

Delgado, Richard, "Storytelling for Oppositionists and Others: A Plea for Narrative" (1989) 87 Michigan L Rev 2411.

Domingos, Pedro, "A Few Useful Things To Know About Machine Learning" (2012) 10 Communications of the ACM 78.

Dressel, Julia & Hany Farid, "The accuracy, fairness, and limits of predicting recidivism" (2018) 4 Science Advances eaao5580 at 1.

Dworkin, Ronald, "Law as Interpretation" (1982) 9 Critical Inquiry 179.

Edwards, Lillian & Michael Veale, "Slave to the Algorithm? Why a 'Right to an Explanation' Is Probably Not the Remedy You Are Looking For" (2017) 16 Duke Law & Technology Review 18.

Eren, Ozkan & Naci Mocan, “Emotional Judges and Unlucky Juveniles” (2018) 10 Applied Economics 171.

Etxabe, Julen, “The Legal Universe After Robert Cover” (2010) 115 Law and Humanities, online: *Academia* <www.academia.edu/368582/The_Legal_Universe_After_Robert_Cover> [perma.cc/G9CS-6DS9].

Farber, Daniel & Suzanna Sherry, “Telling Stories Out of School: An Essay on Legal Narratives” (1993) 45 Stan L Rev 807.

Ferguson, Andrew G, “Facial Recognition and the Fourth Amendment” (2019) 105 Minnesota Law Review 1105.

Fish, Stanley, “Working on the Chain Gang: Interpretation in Law and Literature” (1982) 60 Tex L Rev 551.

Friedland, Hadley & Val Napoleon, “Gathering the Threads: Developing a Methodology for Researching and Rebuilding Indigenous Legal Traditions” (2015-2016) 1:1 Lakehead Law Journal 16.

Friese, Malte, et al, “Is Ego Depletion Real? An Analysis of Arguments” (2019) 23 Personality and Social Psychology Review 107.

Froomkin, A Michael, “The Metaphor is the Key: Cryptography, The Clipper Chip, and the Constitution” (1995) 143 University of Pennsylvania Law Review 709.

Fuller, Lon L, “The Case of the Speluncean Explorers” (1949) 62 Harvard L Rev 616.

Gavison, Ruth, “Privacy and the Limits of the Law” (1980) 89 Yale LJ 421.

Glöckner, Andrea, “The irrational hungry judge effect revisited: Simulations reveal that the magnitude of the effect is overestimated” (2016) 11 Judgment and Decision-Making 601.

Goldenfein, Jake, “The Potential Profiling of Computer Vision and the Challenge of Computational Empiricism” (2019) Proceedings of the ACM FAT* Conference, online (pdf): *SSRN* <papers.ssrn.com/sol3/papers.cfm?abstract_id=3284598> [perma.cc/RJX6-RT9S].

Goldsworthy, Daniel, “Dworkin’s dream: Towards a singularity of law” (2019) 44 Alternative Law Journal 286.

Gowder, Paul, “Transformative Legal Technology and The Rule of Law” 68 UTLJ (Supplement 1) 82.

Grant, Kenneth A, “Tacit Knowledge Revisited—We Can Still Learn from Polanyi” (2007) 6 Electronic Journal of Knowledge Management 173.

Green, Leslie, “The Normativity of Law: What is the Problem? (last accessed 26 July 2022), online (pdf): *UVic* <www.uvic.ca/victoria-colloquium/assets/docs/Green_Normativity.pdf> [perma.cc/2HLD-SV9V].

Grimmelmann, James, "Regulation by Software" (2005) 114 Yale LJ 1719.

Grimmelmann, James, "The Google Dilemma" (2009) 53 New York Law School Review 939.

Guilhot Michael, et al, "Nudging Robots: Innovative Solutions to Regulate Artificial Intelligence" (2017) 20 Vand J Ent & Tech L 385.

Heilbroner, Robert L, "Do Machines Make History?" (1967) 8 Technology and Culture 335.

Hellman, Deborah, "Measuring Algorithmic Fairness" (2020) 106 Virginia L Rev at 816.

Hildebrandt, Mireille, "Law as an Affordance: The Devil Is in the Vanishing Point(s)" (2017) 4 Critical Analysis of Law 116.

Hildebrandt, Mirielle, "Algorithmic Regulation & the Rule of Law" (2018) Phil Trans R Soc A 376: 20170355, online: *The Royal Society* <dx.doi.org/10.1098/rsta.2017.0355> [perma.cc/W6F9-EZAX].

Holmes Jr, Oliver Wendell, "The Path of Law" (1897) 10 Harv L Rev 457.

Huq, Aziz, "Racial Equity in Algorithmic Criminal Justice" (2019) 68 Duke Law Journal 1043.

Ileberi, Emmanuel, Yanxia Sun & Zenghui Wang, "A machine learning based credit card fraud detection using the GA algorithm for feature selection" (2022) 9 Journal of Big Data.

Jones, Meg Leta, "Does Technology Drive Law? The Dilemma of Technological Exceptionalism in Cyberlaw" (2018) 2 Journal of Law, Technology & Policy 102.

Joseph Raz, "The Law's Own Virtue" (November 2018), online: *SSRN* <papers.ssrn.com/sol3/papers.cfm?abstract_id=3262030> [perma.cc/U6EC-ZCWX].

Kahan, Dan M, "Laws of cognition and the cognition of law" (2015) 135 Cognition 56.

Kerr, Ian, "Bots, Babes and the Californication of Commerce" (2004) 1 uOttawa L & Tech J 285.

Kerr, Ian, "Schrödinger's Robot: Privacy in Uncertain States" (2019) 20 Theoretical Inquiries L 123.

Kerr, Ian, "The Devil Is in the Defaults" (2017) 4:1 Critical Analysis of Law 91.

Kleinberg, Jon, et al, "Human Decisions and Machine Predictions" (2017) 133 Quarterly Journal of Economics 237.

Latour, Bruno, "On actor-network theory: A few clarifications" (1996) 47 Soziale Welt 369.

Lebel, Louis (Hon.), "Notes for an Address: Reflections on Natural Justice and Procedural Fairness in Canadian Administrative Law" (2013) 26 Can J Admin L & Prac 51.

Lessig, Lawrence, "The New Chicago School" (1998) 27 J Legal Studies 661.

- Lindberg, Darcy, “Drawing Upon Indigenous Laws in the Yukon” (2020) 50 *The Northern Review* 179, online (pdf): *The Northern Review* <doi.org/10.22584/nr50.2020.007> [perma.cc/E9CR-7XDT].
- Martin-Bariteau, Florian, “The Matrix of Law: From Paper, to Word Processing, to Wiki” (2014) *Lex Electronica* 23.
- Mayson, Sandra G, “Bias In, Bias Out” (2021) 128 *Yale Law Journal* 2218.
- McGill, Jena & Amy Salazyn, “Judging by the Numbers: Judicial Analytics, the Justice System and its Stakeholders” (2021) 44 *Dal LJ* 249.
- Murphy, Colleen, “Lon Fuller and the Moral Value of the Rule of Law” (2004) 24 *Law and Philosophy* 239.
- Nadler, Jennifer, “Hart, Fuller and the Connection Between Law and Justice” (2007) 27 *Law and Philosophy* 1.
- Napoleon, Val, “Gitxsan Legal Personhood: Gendered” in Austin Sarat, ed, *Interrupting the Legal Person (Studies in Law, Politics, and Society), Volume 87A* (Bingley, UK: Emerald Publishing Limited, 2022).
- Nissenbaum, Helen, “Privacy as Contextual Integrity” (2004) 79 *Wash L Rev* 119 at 137.
- Nye, Hilary, “Finnis’s divided view of law: problems for adjudicative theory” (2020) 11 *Jurisprudence* 503.
- Oliver, Peter C, “A Constitution Similar in Principle to That of the United Kingdom?: The Preamble, Constitutional Principles, and a Sustainable Jurisprudence” (2019) 65:2 *McGill LJ* 207.
- Papageorgiou, Theofanis & Panayotis G Michaelides, “Schumpeter and Veblen on technological determinism, individualism and institutions” (2016) 23 *The European Journal of the History of Economic Thought* 1.
- Parasuraman, Raja & Dietrich Manzey, “Complacency and Bias in Human Use of Automation: An Attentional Integration” (2010) 53 *The Journal of the Human Factors and Ergonomics Society* 381.
- Pasquale, Frank, “A Rule of Persons, Not Machines: The Limits of Legal Automation” (2019) 87 *Geo Wash L Rev* 1.
- Pasquale, Frank, & Glyn Cashwell “Prediction, Persuasion, and the Jurisprudence of Behaviourism” (2018) 68 *UTLJ* (Supplement 1) 63.
- Posner, Richard A, “Privacy, Surveillance, and the Law” (2008) 75 *U Chi L Rev* 245.
- Pound, Roscoe, “Mechanical Jurisprudence” (1908) 8 *Columbia L Rev* 605.

Prosser, William L, "Privacy" (1960) 48 California Law Review 383.

Rafter, Nicole, "The Murderous Dutch Fiddler: Criminology, History, and the Problem of Phrenology" (2005) 9 Theoretical Criminology 65.

Reidenberg, Joel, "Lex Informatica: The Formulation of Information Policy Rules Through Technology" (1997) 76 Tex L Rev 553.

Richards, Neil & Woodrow Hartzog, "A Duty of Loyalty for Privacy Law" (3 July 2020), online (pdf): *SSRN* <papers.ssrn.com/sol3/papers.cfm?abstract_id=3642217> [perma.cc/R5W8-GCQ9].

Rubin, Cynthia, et al, "Learning Certifiably Optimal Rule Lists for Categorical Data" (2018) 18 Journal of Machine Learning Research 1.

Sacks, Albert M, "Lon Luvois Fuller" (1978) 92 Harvard L Rev 349.

Scassa, Teresa, "Administrative Law and the Governance of Automated Decision-Making: A Critical Look at Canada's Directive on Automated Decision-Making" (2021) 54 UBC L Rev 1.

Selbst, Andrew & Solon Barocas, "The Intuitive Appeal of Explainable Machines" (2018) 87 Fordham L Rev 1085.

Shapiro, Scott J, "The Hart-Dworkin Debate: A Short Guide for the Perplexed" (2007), online: *SSRN* <papers.ssrn.com/sol3/papers.cfm?abstract_id=968657> [perma.cc/Q8GU-9AKT].

Shapiro, Scott, "What Is the Internal Point of View?" (2006) Fordham LJ 75.

Solove, Daniel J, "Privacy Self-Management and the Consent Dilemma" (2013) 126:7 Harv L Rev 1880.

Solum, Lawrence B, "Equity and the Rule of Law" (1994) 36 Nomos 120.

Sotto-Santiago, Sylk, "Time to Reconsider the Word Minority in Academic Medicine" (2019) 12 J Best Pract Health Prof Divers 72.

Sunstein, Cass, "Of Artificial Intelligence and Legal Reasoning" (2001) 8 University of Chicago Law School Roundtable 29.

Sunstein, Cass, "There is Nothing that Interpretation Just Is" (2015) Constitutional Commentary 193, online: *University of Minnesota* <<https://scholarship.law.umn.edu/concomm/120>> [perma.cc/5N8E-4LND].

Surden, Harry, "Artificial Intelligence and Law: An Overview" (2019) 35 Georgia State University Law Review 1305.

Surden, Harry, "Machine Learning and Law" (2014) 89 Washington L Rev 87.

Sweeney, Latanya, "Discrimination in Online Ad Delivery" (2013) 56 Communications of the ACM 44.

Szilagyi, Katie, "Fragmenting Epistemologies: Towards Philosophical Foundations for Machine Learning in Law" forthcoming in Richard Jochelson & James Gacek, eds, *Ignorance is Bliss* (2023).

Szilagyi, Katie, Jason Millar, AJung Moon, Shalaleh Rismani, "Driving into the Loop: Mapping Automation Bias & Liability Issues for Advanced Driver Assistance Systems" *Digital Society* (forthcoming 2022).

Tamanaha, Brian Z, "Always imperfectly achieved rule of law: Comments on Jeremy Waldron" (2021) 10 *Global Constitutionalism* 106.

Tranter, Kieran, "Disrupting Technology Disrupting Law" (2017) *Law, Culture & the Humanities*.

Tromp, Nynke, Paul Hekkert & Peter-Paul Verbeek, "Design for Socially Responsible Behaviour: A Classification of Influence Based on Intended User Experience" (2011) 27 *Design Issues* 3.

Verbeek, Peter-Paul, "Materializing Morality: Design Ethics and Technological Mediation" (2006) 31 *Science, Technology, & Human Values* 361.

Waldman, Ari Ezra, "Privacy Law's False Promise" (2020) 97 *Washington University Law Review* 773 at 776.

Waldron, Jeremy, "The Concept and the Rule of Law" (2008) 43 *Georgia Law Review* 1.

Warren, Samuel D & Louis D Brandeis, "The Right to Privacy" (1890) 4 *Harv L Rev* 193.

Wells, Catherine Pierce, "Langdell and the Invention of Legal Doctrine" (2010) *Buffalo L Rev* 551.

Wendel, W Bradley, "Explanation in Legal Scholarship: The Inferential Structure of Doctrinal Legal Analysis" (2011) 96 *Cornell L Rev* 1035.

Westin, Alan, *Privacy and Freedom* (1968) 25 *Wash & Lee L Rev* 166.

Wetlaufer, Gerald B, "Rhetoric & its Denial in Legal Discourse" (1990) 76 *Va L Rev* 1545.

White, Lucie E, "Subordination, Rhetorical Survival Skills, and Sunday Shoes: Notes on the Hearing of Mrs. G." (1990) 38 *Buffalo Law Review* 1.

Woozley, AD, "No Right Answer" (1979) 29 *The Philosophical Quarterly* 25.

Zalnieriute, Monika, Lyria Bennett Moses & George Williams, "The Rule of Law and Automation of Government Decision-Making" (2019) 82:3 *Modern Law Review* 1.

Zuboff, Shoshanna, "Big other: surveillance capitalism and the prospects of an information civilization" (2015) 30 *Journal of Information Technology* 75.

Secondary Materials: Book Chapters

- Bailey, Jane, Jacquelyn Burkell, Suzie Dunn, Chandell Gosse & Valerie Steeves, “AI and Technology-Facilitated Violence and Abuse” in Florian Martin-Bariteau & Teresa Scassa, eds, *The Law of Artificial Intelligence in Canada* (Ottawa: LexisNexis 2021) 249.
- Baumeister, RF, “Limited Resources for Self-Regulation: A Current Overview of the Strength Model” in Edward R Hirt et al, *Self-Regulation and Ego Control* (London, UK: Elsevier, 2016) 185.
- Brownsword, Roger, “Artificial Intelligence and Legal Singularity: The Thin End of the Wedge, the Thick End of the Wedge, and the Rule of Law” in Simon Deakin & Christopher Markou, eds, *Is Law Computable?: Critical perspectives on law and artificial intelligence* (Oxford, UK: Hart Publishing, 2020).
- Brundage, Miles, “Scaling Up Humanity: The Case for Conditional Optimism about Artificial Intelligence” in Peter J Bentley, Miles Brundage, et al, “Should We Fear Artificial Intelligence?” (March 2018), online: *European Parliament* <[www.europarl.europa.eu/RegData/etudes/IDAN/2018/614547/EPRS_IDA\(2018\)614547_EN.pdf](http://www.europarl.europa.eu/RegData/etudes/IDAN/2018/614547/EPRS_IDA(2018)614547_EN.pdf)> [perma.cc/7CLG-SKUH].
- Cobbe, Jennifer, “Legal Singularity and the Reflexivity of Law” in Simon Deakin & Christopher Markou, eds, *Is Law Computable?: Critical perspectives on law and artificial intelligence* (Oxford, UK: Hart Publishing, 2020).
- Gordon, Robert, “Law as a Vocation: Holmes and the lawyer’s path” in SJ Burton, ed, *The Path of Law and its Influence: The legacy of Oliver Wendell Holmes Jr*, (New York, NY: Cambridge University Press, 2000) 7.
- Haraway, Donna, “A Cyborg Manifesto: Science, Technology, and Socialist-Feminism in the late Twentieth Century” in Donna Haraway & Cary Wolfe, eds, *Manifestly Haraway* (Minneapolis, MN: University of Minnesota Press, 2016) 3.
- Hughes, Thomas P, “The Evolution of Large Technological Systems” in Wiebe E Bijker, Thomas P Hughes & Trevor Pinch, eds, *The Social Construction of Technological Systems: New Directions in the Sociology and History of Technology* (Cambridge, MA: The MIT Press, 1987) 51.
- Kerr, Ian & Katie Szilagyi, “Asleep at the Switch: How Lethal Autonomous Robots Become a Force Multiplier of Military Necessity” in Ryan Calo, A Michael Froomkin & Ian Kerr, eds, *Robot Law* (Cheltenham, UK: Edward Elgar Publishing, 2016) 333.
- Kerr, Ian, “Digital Locks and the Automation of Virtue” in Michael Geist, ed, *From Balanced Copyright to Radical Extremism* (Toronto: Irwin Law, 2010) 247.

Kerr, Ian, “Mind Your Metaphors: An Examination of the Inefficacy Argument as a Reason Against Regulating On-line Conduct” in Lester Pourciau, ed, *Ethics and Electronic Information in the Twenty-First Century* (West Lafayette, IN: Purdue University Press, 1999) 231.

Kerr, Ian, “Prediction, pre-emption, presumption: The path of law after the computational turn” in Mireille Hildebrandt & Katja de Vries, eds, *Privacy and Due Process After the Computational Turn* (London, UK: Routledge, 2013) 91.

Latour, Bruno, “Where Are the Missing Masses? The Sociology of a Few Mundane Artifacts” in Deborah G Johnson & Jameson M Wetmore, eds, *Technology and Society: Building Our Sociotechnical Future* (Cambridge, MA: The MIT Press, 2009) 225.

Smith, Merritt Roe, “Technological Determinism in American Culture” in Merritt Roe Smith & Leo Marx, *Does Technology Drive History? The Dilemma of Technological Determinism* (Cambridge, MA: The MIT Press, 1994) 1.

Posner, Richard A, ed, “Introduction” in *The Essential Holmes: Selections from the Letters, Speeches, Judicial Opinions, and Other Writings of Oliver Wendell Holmes Jr.* (Chicago, IL: University of Chicago Press, 1992) ix.

Raso, Jennifer, “AI and Administrative Law” in Florian Martin-Bariteau & Teresa Scassa, eds, *The Law of Artificial Intelligence in Canada* (Ottawa: LexisNexis 2021) 181.

Raz, Joseph, “The Rule of Law and its Virtue” in *The Authority of Law: Essays on Law and Morality* (Oxford, UK: Oxford University Press, 1979) 210.

Richards, Neil M & William D Smart, “How Should the Law Think About Robots” in Calo et al, *Robot Law* (Cheltenham, UK: Edward Elgar Publishing, 2016) 3.

Stone, Allucquère Rosanne, “Will the Real Body Please Stand Up?” in Michael Benedikt, ed, *Cyberspace: First Steps* (Cambridge, MA: MIT Press, 1991) 81.

West, Robin, “The Limits of Process” in James Fleming, ed, *Getting to the Rule of Law: Nomos L* (New York, NY: New York University Press, 2011) 32.

Winner, Langdon, “Do Artifacts Have Politics?” in Langdon Winner, ed, *The Whale and the Reactor: A Search for Limits in an Age of High Technology* (Chicago: University of Chicago Press, 1986) 19.

Winner, Langdon, “Technologies as Forms of Life” in David Kaplan, ed, *Philosophy of Technology* (2004) 103.

Secondary Materials: Conference Papers

Ballard, Stephanie & Ryan Calo, “Taking Legal Futures Seriously” (2019) We Robot Conference Paper, online: *UMiami* <robots.law.miami.edu/2019/wp-content/uploads/2019/03/Calo_Taking-Futures-Seriously.pdf> [perma.cc/PH2J-LL96].

Bao, Michelle, Angela Zhou, Samantha Zottola, et al, “It’s COMPASlicated: The Messy Relationship between RAI Datasets and Algorithmic Fairness Benchmarks” (2022) 35th Conference on Neural Information Processing Systems (NeurIPS 2021) Track on Datasets and Benchmarks.

Bender, Emily M, Timnit Gebru, Angelina McMillan-Major, Margaret Mitchell, “On the Dangers of Stochastic Parrots: Can Language Models Be Too Big? 🦜” In Conference on Fairness, Accountability, and Transparency (FAccT ’21), March 3–10, 2021, Virtual Event, Canada.

Dunn, Suzie, “Identity Manipulation: Responding to advances in artificial intelligence and robotics” (16 July 2020), We Robot 2020 Conference Paper, online (pdf): <drive.google.com/file/d/1LXCxTEj7zHbFr8V3SdxKxnut173aCGsf/view> [perma.cc/N24S-DRJ3].

Green, Ben, “Fair’ Risk Assessments: A Precarious Approach for Criminal Justice Reform” (2018) Presented at the 5th Workshop on Fairness, Accountability, and Transparency in Machine Learning (FAT/ML 2018) Stockholm, Sweden.

Hamidi, Foad, Morgan Klaus Scheuerman, Stacy M Branham, “Gender Recognition or Gender Reductionism? The Social Implications of Automatic Gender Recognition Systems” (2018) CHI 2018 (April 21-26, Montreal, QC, Canada).

Kerr, Ian & Carissima Mathen, “Chief Justice John Roberts is a Robot” (2014) We Robot Conference paper, online (pdf): *SSRN* <papers.ssrn.com/sol3/papers.cfm?abstract_id=3395885> [perma.cc/28YG-UFMY].

Khoo, Cynthia, “Missing the Unintended Forest despite the Deliberately Planted Trees: Reasonable Foreseeability and Legal Recognition of Platform Algorithm-Facilitated Emergent Systemic Harm to Marginalized Communities” (2020) We Robot Conference Paper, online (pdf): <drive.google.com/file/d/1ZVV9yD8KH4KUYLMxE6mlORiCcFKNXIJU/view> [perma.cc/5GSW-RA96].

Secondary Materials: Dissertations

Lussier, Danielle, “Law with Heart and Beadwork: Decolonizing Legal Education, Developing Indigenous Legal Pedagogy, and Healing Community” PhD Dissertation (2021), online (pdf): *UOttawa* <ruor.uottawa.ca/handle/10393/42012> [perma.cc/9Z8P-LRY3].

Morales, Sarah, “Snuw’uyulh: Fostering an Understanding of the Hul’qumi’num Legal Tradition” PhD Dissertation (2014), online (pdf): *University of Victoria* <dspace.library.uvic.ca/handle/1828/6106?show=full> [perma.cc/5CS9-EM5C].

Thomasen, Kristen, “Private Law & Public Space: The Canadian Privacy Torts in an Era of Personal Remote-Surveillance Technology” PhD Dissertation (2022), online (pdf): *UOttawa* <ruor.uottawa.ca/handle/10393/43746> [perma.cc/D22W-RE7T].

Secondary Materials: Government Documents

European Commission, “Proposal for a Regulation laying down harmonised rules on artificial intelligence” (April 21, 2021), online: *European Commission* <digital-strategy.ec.europa.eu/en/library/proposal-regulation-laying-down-harmonised-rules-artificial-intelligence> [perma.cc/9RA5-69D2].

Government of Canada, “Directive on Automated Decision-Making” (1 April 2021), online: *TBS Canada* <www.tbs-sct.gc.ca/pol/doc-eng.aspx?id=32592> [perma.cc/NRX3-W887].

Government of Canada, Regulatory Impact Analysis Statement, Canada Gazette Part II, Vol 135, No 1 (3 January 2001), online (pdf): *Government of Canada* <gazette.gc.ca/rp-pr/p2/2001/2001-01-03/pdf/g2-13501.pdf>.

Secondary Materials: Monographs

Asimov, Isaac, *I, Robot* (New York, NY: Bantam Books, 1950).

Bayard, Pierre, *How to Talk About Books You Haven't Read*, trans by Jeffrey Mehlman (New York: Bloomsbury, 2007).

Benjamin, Ruha, *Race After Technology: Abolitionist Tools for the New Jim Code* (Cambridge, UK: Polity Press, 2019).

Berger, Peter L, *The Sacred Canopy: Elements of a Sociological Theory of Religion* (New York, NY: Doubleday, 1967).

Bix, Brian, *Law, Language & Legal Determinacy* (Oxford, UK: Oxford University Press, 1993).

Borrows, John, *Canada's Indigenous Constitution* (Toronto: University of Toronto Press, 2010).

Botsman, Rachel, *Who Can You Trust? How Technology Brought Us Together and Why It Might Drive Us Apart* (New York, NY: Hachette Book Group, 2017).

Broussard, Meredith, *Artificial Un-intelligence: How Computers Misunderstand the World* (Cambridge, MA: The MIT Press, 2019).

Carr, Nicholas, *The Shallows: What the Internet is Doing to Our Brains* (New York, NY: WW Norton, 2011).

Christian, Brian, *The Alignment Problem: Machine Learning and Human Values* (New York, NY: WW Norton & Company, 2020).

Cohen, Julie, *Between Truth and Power: The Legal Constructions of Information Capitalism* (2019); online: <juliecohen.com/between-truth-and-power/> [perma.cc/QWY6-L3ZW].

Cohen, Julie, *Configuring the Network Self: Law, Code, and the Play of Everyday Practice* (New Haven, CT: Yale University Press, 2012).

Crawford, Kate, *Atlas of AI: Power, Politics, and the Planetary Costs of Artificial Intelligence* (New Haven, CT: Yale University Press, 2021).

Didion, Joan, *The White Album* (New York, NY: Simon & Schuster, 1979).

Domingos, Pedro, *The Master Algorithm* (New York, NY: Basic Books, 2015).

Dworkin, Ronald, *Justice for Hedgehogs* (Cambridge, MA: Harvard University Press, 2011).

Dworkin, Ronald, *Law's Empire* (Cambridge, MA: Harvard University Press, 1986).

Dworkin, Ronald, *Taking Rights Seriously* (Cambridge, MA: Harvard University Press, 1977).

Eubanks, Virginia, *Automating Inequality: How High-Tech Tools Profile, Police, and Punish the Poor* (New York, NY: Picador, 2019).

Fairfield, Joshua, *Runaway Technology* (Cambridge, MA: Cambridge University Press, 2021).

Finnis, John, *Natural Law and Natural Rights* (Oxford, UK: Oxford University Press, 1980).

Fuller, Lon L, *The Morality of Law* (New Haven, CT: Yale University Press, 1969).

Gibson, William, *Neuromancer* (New York, NY: Ace Books, 1984).

Hand, David J, "Dark Data: Why What You Don't Know Matters" (Princeton, NJ: Princeton University Press, 2020).

Hart, HLA, *The Concept of Law* (Oxford, UK: Oxford University Press, 1961).

Hildebrandt, Mireille, *Smart Technologies and the End(s) of Law: Novel Entanglements of Law and Technology* (Cheltenham: Edward Elgar Publishing, 2015).

Jasanoff, Sheila, *Science at the Bar: Law, Science, and Technology in America* (Cambridge, MA: Harvard University Press,

Kaag, John, *Hiking with Nietzsche: On Becoming Who You Are* (New York, NY: Farrar, Straus & Giroux, 2018).

Kahneman, Daniel, Oliver Sibony, Cass R Sunstein, *Noise: A Flaw in Human Judgement* (New York, NY: Little Brown, Spark, 2021).

Kahneman, Daniel, *Thinking Fast and Slow* (Toronto: Doubleday Canada, 2011).

Katsh, M Ethan, *The Electronic Media and the Transformation of Law* (New York, NY: Oxford University Press, 1989).

Kurzweil, Ray, *The Singularity is Near: When Humans Transcend Biology* (London, UK: Duckworth Overlook, 2005).

Lakoff, George, *Women, Fire and Dangerous Things: What Categories Reveal about the Mind* (Chicago, IL: University of Chicago Press, 1987).

- Lessig, Lawrence, *Code and Other Laws of Cyberspace* (New York, NY: Basic Books, 1999).
- MacKinnon, Catharine A, *Towards a Feminist Theory of the State* (Cambridge, MA: Harvard University Press, 1989).
- Maher, Chauncey, *The Pittsburgh School of Philosophy: Sellars, McDowell, Brandom* (Abingdon, UK: Routledge, 2012).
- McIsaac, Shields & Klein, eds, *The Law of Privacy in Canada* (Toronto: Thomson Reuters, 2015).
- Moran, Mayo, *Rethinking the Reasonable Person: An Egalitarian Reconstruction of the Objective Standard* (New York, NY: Oxford University Press: 2003).
- Morozov, Evgeny, *To Save Everything, Click Here: The Folly of Technological Solutionism* (New York, NY: PublicAffairs, 2013).
- Nedelsky, Jennifer, *Law's Relations: A Relational Theory of Self, Autonomy, and Law* (Oxford, UK: Oxford University Press, 2011).
- Noble, Safiya Umoja, *Algorithms of Oppression: How Search Engines Reinforce Racism* (New York, NY: New York University Press, 2018).
- O'Neil, Cathy, *Weapons of Math Destruction: How Big Data Increases Inequality and Threatens Democracy* (New York, NY: Broadway Books, 2016).
- Ost, François, "Le temps virtuel des lois postmodernes ou comment le droit se traite dans la société de l'information" in Jean Clam & Gilles Martin, eds, *Les transformations de la régulation juridique* (Paris, FR: LGDJ, 1998).
- Pasquale, Frank, *New Laws of Robotics: Defending Human Expertise in the Age of AI* (Cambridge, MA: Harvard University Press, 2020).
- Pasquale, Frank, *The Black Box Society: The Secret Algorithms That Control Money and Information* (Cambridge, MA: Harvard University Press: 2015).
- Pirsig, Robert, *Zen and the Art of Motorcycle Maintenance: An Inquiry into Values* (New York, NY: HarperCollins, 1974).
- Plant, Sadie, *Zeros and Ones: Digital Women and the New Technoculture* (London, UK: Doubleday, 1995).
- Popkin, William D, *Evolution of the Judicial Opinion: Institutional and Individual Styles* (New York, NY: New York University Press, 2007).
- Raz, Joseph, *The Authority of Law: Essays on Law and Morality* (Oxford, UK: Oxford University Press, 2009).

- Rorty, Richard, *Contingency, irony, and solidarity* (New York, NY: Cambridge University Press, 1989).
- Rundle, Kristen, *Forms Liberate: Reclaiming the Jurisprudence of Lon L Fuller* (Portland, OR: Hart Publishing, 2013).
- Russell, Bertrand, *The Problems of Philosophy* (New York, NY: Henry Holt and Company, 1912).
- Scarry, Elaine, *The Body in Pain* (New York, NY: Oxford University Press, 1985).
- Sellars, Wilfrid, *Empiricism and the Philosophy of Mind* (Cambridge, MA: Harvard University Press, 1997).
- Shapiro, Scott, *Legality* (Cambridge, MA: Harvard University Press, 2013).
- Stephenson, Neal, *Snow Crash* (New York, NY: Penguin Random House, 1992).
- Taleb, Nassim Nicholas, *The Black Swan: the impact of the highly improbable* (London, UK: Penguin Books, 2010).
- Tamanaha, Brian, *On the Rule of Law: History, Politics, Theory* (Cambridge, UK: Cambridge University Press, 2004).
- Tyler, Tom R, *Why People Obey The Law* (Princeton: Princeton University Press, 1990).
- Verbeek, Peter-Paul, *What Things Do: Philosophical Reflections on Technology, Agency, and Design* (University Park, PA: The Pennsylvania State University Press, 2005).
- Vonnegut, Kurt, *Hocus Pocus* (New York, NY: Penguin Books, 1997).
- White, James Boyd, *Heracles' Bow: Essays on the Rhetoric and Poetics of Law* (Madison, WI: The University of Wisconsin Press, 1985).
- Whitman, James Q, *The Origins of Reasonable Doubt: Theological Roots of the Criminal Trial* (New Haven, CT: Yale University Press, 2008).
- Williams, Bernard, ed, *Friedrich Nietzsche, The Gay Science* (Cambridge, UK: Cambridge University Press, 2001).
- Winner, Langdon, *Autonomous Technology: Technics-out-of-Control as a Theme in Political Thought* (Cambridge, MA: The MIT Press, 1977).
- Wittgenstein, Ludwig, *Philosophical Investigations*, trans by GEM Anscombe (Oxford, UK: Basil Blackwell, 1958).
- Ziff, Bruce, *Principles of Property*, 7th ed (Toronto: Thomson Reuters, 2018).
- Zuboff, Shoshanna, *The Age of Surveillance Capitalism* (New York, NY: PublicAffairs, 2019).

Secondary Materials: Online & Media Resources

@IllustriousVar, “Guys, the algorithm just prioritizes contrast. All you have to do is lower the contrast of the white woman and increase the contrast of the black man to see this. This isn’t racism, he’s a black man standing in front of a dark backdrop causing him to blend in more.” (7 December 2021 at 12:38), online: *Twitter*
<twitter.com/IllustriousVar/status/1335835949448777730>.

Alarie, Benjamin, “Transcending Books: What’s Next for Law and Society?” TEDxBeaconStreet (August 2018), online: *TED*
<www.ted.com/talks/benjamin_alarie_transcending_books_what_s_next_for_law_and_society>
[perma.cc/96UB-BFRN].

Alphonso, Caroline, “Toronto-area school board sorts classes alphabetically, raising concerns of racial segregation” (24 Sept 2020), online: *The Globe & Mail*
<www.theglobeandmail.com/canada/article-toronto-area-school-board-sorts-online-classes-alphabetically-raising/> [perma.cc/4LB8-MXX8].

Amazon Web Services, “AWS Machine Learning Blog” (last accessed 30 July 2022), online: *Amazon*
<aws.amazon.com/blogs/machine-learning/> [perma.cc/Y4WY-BXZM].

Angwin, Julia, et al, “Machine Bias” (23 May 2016), online: *ProPublica*
<www.propublica.org/article/machine-bias-risk-assessments-in-criminal-sentencing>
[perma.cc/8AQS-EHHQ].

Apple, “About Face ID advanced technology” (27 April 2022), online: *Apple Support*
<support.apple.com/en-ca/HT208108#:~:text=Face%20ID%20automatically%20adapts%20to,it%20updates%20your%20face%20data> [perma.cc/S45K-C95J].

Aristotle, *Politics* 1282b, cited in The Stanford Encyclopedia of Philosophy, “The Rule of Law” (22 June 2016), online: *The Metaphysics Research Lab*
<<https://plato.stanford.edu/entries/rule-of-law/>> [perma.cc/5445-UY3C].

arXiv, online: <arxiv.org> [perma.cc/2WTQ-72YL].

Barlow, John Perry, “The Declaration of Independence of Cyberspace” (1996), online: *EFF*
<www.eff.org/cyberspace-independence> [perma.cc/7GTC-A9QN].

Bergen, Rachel, “African Canadian woman fed up with Manitoba Public Insurance after yet another ‘dehumanizing’ photo glitch” (8 July 2022), online: *CBC*
<www.cbc.ca/amp/1.6515479>
[perma.cc/R7JN-J2SS].

Berkeley Artificial Intelligence Research, online: *Berkeley*
<bair.berkeley.edu/blog/>
[perma.cc/E68N-WCQS].

Berman, Robby, “Facebook Catches Two Chatbots Speaking Their Own Language” (6 August 2017), online: BigThink <bigthink.com/roby-berman/facebook-catches-2-chatbots-speaking-their-own-language> [perma.cc/5N9F-YSBL].

Bostrom, Nick, “Ethical Issues in Advanced Artificial Intelligence” (2003), online: *Nick Bostrom* <nickbostrom.com/ethics/ai.html> [perma.cc/FT77-BYA4].

Buolamwini, Joy, “How I’m Fighting Bias in Algorithms” (9 March 2017), online (video): *TED* <www.ted.com/talks/joy_buolamwini_how_i_m_fighting_bias_in_algorithms?language=en> [perma.cc/4T7D-5WTA].

Burt, Andrew, “Is there a ‘right to an explanation’ for machine learning in the GDPR?” (1 June 2017), online: *IAPP* <iapp.org/news/a/is-there-a-right-to-explanation-for-machine-learning-in-the-gdpr/> [perma.cc/Z9UG-FFVP].

Cadwalladr, Carole, “I made Steve Bannon’s psychological warfare tool’: meet the data war whistleblower,” (18 March 2018), online: *The Guardian* <www.theguardian.com/news/2018/mar/17/data-war-whistleblower-christopher-wylie-faceook-nix-bannon-trump> [perma.cc/WPR7-4VA3].

Cambridge Innovation Institute, online: *CII* <www.cambridgeinnovationinstitute.com/> [perma.cc/2R2H-3LLF].

Chapela, Emilio, “Google Suggestions ... (Racist or Rude)” (19 November 2009), online: *Blogspot* <emiliochapela.blogspot.com/2009/11/google-suggestions-racist-or-rude_19.html> [perma.cc/4P7R-4GFD].

Christin, Angèle, Alex Rosenblat, & danah boyd, “Courts and Predictive Algorithms” (27 October 2015), online (pdf): *Data Civil Rights* <www.datacivilrights.org/pubs/2015-1027/Courts_and_Predictive_Algorithms.pdf> [perma.cc/H7VK-76TL].

Chu, Arthur, “Opening Pandora’s Dox” (29 August 2015), online: *TechCrunch* <techcrunch.com/2015/08/29/opening-pandoras-dox-the-unintended-consequences-of-an-internet-that-never-forgets/> [perma.cc/7HSA-SEPA].

Clayton, James, “Facial recognition beats the Covid-mask challenge” (25 March 2021), online: *BBC News* <www.bbc.com/news/technology-56517033> [perma.cc/69V8-QGY2].

Cummings, K Ward, “‘Minoritized’: a violent word” (27 March 2019), online: *Baltimore Sun* <www.baltimoresun.com/opinion/op-ed/bs-ed-op-0328-minoritized-word-20190320-story.html> [perma.cc/ZZP8-QJIX].

DeVries, Willem, *Wilfrid Sellars*, Stanford Encyclopedia of Philosophy, online: *SEP* <<https://plato.stanford.edu/entries/sellars/>>.

Dick, Philip K, “The Minority Report” (1956), available online: <cwanderson.org/wp-content/uploads/2011/11/Philip-K-Dick-The-Minority-Report.pdf> [perma.cc/98E7-R8EG].

Ellora Israni, “Algorithmic Due Process: Mistaken Accountability and Attribution in *State v. Loomis*” (August 31, 2017), online: *JOLT Digest* <jolt.law.harvard.edu/digest/algorithmic-due-process-mistaken-accountability-and-attribution-in-state-v-loomis-1> [perma.cc/E8XL-2JBN].

Equivant, “Practitioner’s Guide to COMPAS Core” (April 4, 2019), online: *Equivant* <www.equivant.com/wp-content/uploads/Practitioners-Guide-to-COMPAS-Core-040419.pdf> [perma.cc/4XL3-FFKY].

ETHI, “Towards Privacy by Design: Review of the Personal Information Protection and Electronic Documents Act” (42nd Parliament, 1st Session), online: *House of Commons* <www.ourcommons.ca/DocumentViewer/en/42-1/ETHI/report-12/> [perma.cc/Y5CH-PGEE].

Firth-Godbehere, Rich, “Silicon Valley thinks everyone feels the same six emotions” (17 September 2018), online: *Quartz* <qz.com/1392130/silicon-valley-thinks-everyone-feels-the-same-six-emotions/> [perma.cc/8D9S-CGPE].

Fussell, Sidney, “Facebook’s New Face Recognition Features: What We Do (and Don’t) Know” (27 February 2018), online: *Gizmodo* <gizmodo.com/facebooks-new-face-recognition-features-what-we-do-an-1823359911> [perma.cc/2GNY-H9N6].

Geolitica, “Predictive Policing Blog” (2 March 2021), online: *PredPol* <blog.predpol.com/geolitica-a-new-name-a-new-focus> [perma.cc/N9K2-C5ZS].

Gideon’s Promise, National Legal Aid & Defender Association, et al, “Joint Statement in Support of the Use of Pretrial Risk Assessment Instruments” (10 May 2017), online (pdf): <www.publicdefenders.us/files/Defenders%20Statement%20on%20Pretrial%20RAI%20May%202017.pdf> [perma.cc/XCF9-GSV7].

Gill, Emma, “‘I am expected to just live with these unfair grades’ – Student’s open letter to the government as she slams A-level results system” *Manchester Evening News* (13 August 2020), online: *Manchester Evening News* <www.manchestereveningnews.co.uk/news/greater-manchester-news/a-level-results-unfair-downgraded-18764743> [perma.cc/9PXS-5KKJ].

GitHub, “tensorflow/tensorflow”, online: <github.com/tensorflow/tensorflow > [perma.cc/77QY-T2JM].

Google Research, “Natural Language Processing” (last visited 7 August 2022), online: *Google* <research.google/research-areas/natural-language-processing/> [perma.cc/49PS-5LEC].

GPT-3, “A robot wrote this entire article. Are you scared yet, human?” (8 September 2020), online: *The Guardian* <www.theguardian.com/commentisfree/2020/sep/08/robot-wrote-this-article-gpt-3> [perma.cc/YWU4-FYPX].

Guzmán, Rigoberto Lara & Sareeta Amrute, “How to Cite Like a Badass Tech Feminist Scholar of Color” (22 August 2019), online: *Data & Society* <points.datasociety.net/how-to-cite-like-a-badass-tech-feminist-scholar-of-color-ebc839a3619c> [perma.cc/X5LL-SFG2].

Hao, Karen, “The two-year fight to stop Amazon from selling face recognition to the police” (12 June 2020), online: *MIT Technology Review* <www.technologyreview.com/2020/06/12/1003482/amazon-stopped-selling-police-face-recognition-fight/> [perma.cc/MB76-6WCN].

Harmon, Amy, “As Cameras Track Detroit’s Residents, a Debate Ensues Over Racial Bias” (8 July 2019), online: *The New York Times* <www.nytimes.com/2019/07/08/us/detroit-facial-recognition-cameras.html> [perma.cc/K8NZ-26CY].

Hartzog, Woodrow, & Evan Selinger, “Facial Recognition Is the Perfect Tool for Oppression” (2 August 2018), online: *Medium* <medium.com/s/story/facial-recognition-is-the-perfect-tool-for-oppression-bc2a08f0fe66> [perma.cc/6QJE-H777].

Here&Now Podcast, “‘Hard To Exaggerate’ How ‘Data-Hungry’ Facebook Is, One Professor Says” (5 April 2018), online: *WBUR* <www.wbur.org/hereandnow/2018/04/05/facebook-data-surveillance> [perma.cc/YQ84-LV46].

Hern, Alex, “Twitter apologises for ‘racist’ image-cropping algorithm” (21 September 2020), online: *The Guardian* <www.theguardian.com/technology/2020/sep/21/twitter-apologises-for-racist-image-cropping-algorithm> [perma.cc/JPW4-8KUA].

Hill, Kashmir & Ryan Mac, “Facebook, Citing Societal Concerns, Plans to Shut Down Facial Recognition System” (2 November 2021), online: *The New York Times* <www.nytimes.com/2021/11/02/technology/facebook-facial-recognition.html> [perma.cc/TW45-Y2SV].

Hill, Kashmir, “Before Clearview Became a Police Tool, It Was a Secret Plaything of the Rich” (6 March 2020), online: *The New York Times* <www.nytimes.com/2020/03/05/technology/clearview-investors.html> [perma.cc/7DM3-SFKD].

Hill, Kashmir, “Wrongfully Accused by Algorithm” (24 June 2020), online: *The New York Times* <www.nytimes.com/2020/06/24/technology/facial-recognition-arrest.html> [perma.cc/E6ZT-ZXC3].

Holmes Jr, Oliver Wendell, *The Common Law* at 1, online (eBook): *Project Gutenberg* <www.gutenberg.org/files/2449/2449-h/2449-h.htm> [perma.cc/YZ6L-SW3H].

Horgan, John, “Bayes’s Theorem: What’s the Big Deal?” (4 January 2016), online: *Scientific American* <blogs.scientificamerican.com/cross-check/bayes-s-theorem-what-s-the-big-deal/#:~:text=Embedded%20in%20Bayes'%20theorem%20is,turn%20out%20to%20be%20erroneous> [perma.cc/83NL-VPMB].

Hughes, Patricia, “The SNC-Lavalin Controversy in IV? Acts: A Play About the Rule of Law” (5 March 2019), online: *SLAW* <www.slaw.ca/2019/03/05/the-snc-lavalin-controversy-in-iv-acts-a-play-about-the-rule-of-law/> [perma.cc/X89C-ETVP].

Johnson, Bobbie, "Privacy is no longer a social norm," (11 January 2010), online: *The Guardian* <[theguardian.com/technology/2010/jan/11/facebook-privacy](https://www.theguardian.com/technology/2010/jan/11/facebook-privacy)> [perma.cc/U36X-92YV].

Jonathan Franzen, "Imperial Bedroom" (October 5, 1998), online: *The New Yorker* <www.newyorker.com/magazine/1998/10/12/imperial-bedroom> [perma.cc/S8ZQ-QYAX].

Kantayya, Shalini, *Coded Bias* online: <www.codedbias.com/> [perma.cc/DKU6-X3T9].

Kerr, Ian, "AIs as Substitute Decision Makers" (1 November 2018), online: *Balkinization* <balkin.blogspot.com/2018/11/ais-as-substitute-decision-makers.html> [perma.cc/5V85-QXU9].

Kessler, Glenn, "Fact Checker: Robert Moses and the saga of the racist parkway bridges" (21 November 2021), online: *Washington Post* <www.washingtonpost.com/politics/2021/11/10/robert-moses-saga-racist-parkway-bridges/> [perma.cc/VV3E-QNSU].

Kira, "Deepen Client Relationships. Grow Revenue. Deliver Total Diligence." (last accessed 29 July 2022), online: *Kira Systems* <kirasystems.com/> [perma.cc/YRW7-M5BS].

Larson, Jeff, et al, "How We Analyzed the COMPAS Recidivism Algorithm" (23 May 2016), online: *ProPublica* <www.propublica.org/article/how-we-analyzed-the-compas-recidivism-algorithm> [perma.cc/XUT6-UEER].

LeCun, Yann, "A Path Towards Autonomous Machine Intelligence, Version 0.9.2" (27 June 2022), online: *OpenReview* <openreview.net/pdf?id=BZ5a1r-kVsf> [perma.cc/C6VC-5LZ7].

Legal Beagle, "Attorneys" (last accessed 29 July 2022), online: *Legal Beagle* <legalbeagle.net/attorneys/> [perma.cc/WD4W-559C].

Levin, Sam, "New AI can guess whether you're gay or straight from a photograph" (8 September 2017), online: *The Guardian* <www.theguardian.com/technology/2017/sep/07/new-artificial-intelligence-can-tell-whether-youre-gay-or-straight-from-a-photograph> [perma.cc/GQ2H-QFJW].

Lewis, Paul, "'Our minds can be hijacked': the tech insiders who fear a smartphone dystopia" (6 October 2017), online: *The Guardian* <www.theguardian.com/technology/2017/oct/05/smartphone-addiction-silicon-valley-dystopia> [perma.cc/63FB-JHFF].

Litmus, "How does each spam filter score emails?" (16 December 2021) online: *Litmus* <help.litmus.com/article/205-how-does-each-spam-filter-score-emails> [perma.cc/5KV3-Z7TG].

Mac, Ryan, Caroline Haskins & Logan McDonald, "Clearview's Facial Recognition App Has Been Used by The Justice Department, ICE, Macy's, Walmart, and the NBA" (27 February

2020), online: *BuzzFeed News* <www.buzzfeednews.com/article/ryanmac/clearview-ai-fbi-ice-global-law-enforcement> [perma.cc/GRT2-2UYU].

Macaulay, Thomas, “Clearview AI can be fun—if you’re dirty, stinking rich” (6 March 2020), online: *The Next Web* <thenextweb.com/news/clearview-ai-can-be-fun-if-youre-dirty-stinking-rich> [perma.cc/PDW4-CUFF].

Madrigal, Alexis C, “Reading the Privacy Policies You Encounter in a Year Would Take You 76 Work Days” (1 March 2012), online: *The Atlantic* <www.theatlantic.com/technology/archive/2012/03/reading-the-privacy-policies-you-encounter-in-a-year-would-take-76-work-days/253851/> [perma.cc/N8FZ-QSHM].

Maruzani, Rugare, “Are You Unwittingly Helping to Train Google’s AI Models?” (26 January 2021), online: *Towards Data Science* (towardsdatascience.com/are-you-unwittingly-helping-to-train-googles-ai-models-f318dea53ae) [perma.cc/RS38-WQ5A].

May, Timothy C, “Libertaria in Cyberspace,” (1 September 1992), online: *Nakamoto Institute* <nakamotoinstitute.org/libertaria-in-cyberspace/> [perma.cc/T6TB-NV8U].

May, Timothy C, “The Crypto-Anarchist Manifesto,” (1988), online: *Nakamoto Institute* <nakamotoinstitute.org/crypto-anarchist-manifesto/> [perma.cc/X5TV-8AUT].

McCullagh, Declan, “Call It Super Bowl Face Scan I” (February 2, 2001), online: *WIRED* <www.wired.com/2001/02/call-it-super-bowl-face-scan-i/> [perma.cc/XMP8-S6JL].

McKinsey <www.mckinsey.com/business-functions/mckinsey-analytics/our-insights/global-survey-the-state-of-ai-in-2020> [perma.cc/9BX9-Y3PT].

Metz, Cade, “Meet GPT-3. It Has Learned to Code (and Blog and Argue).” (24 November 2020) online: *The New York Times* <www.nytimes.com/2020/11/24/science/artificial-intelligence-ai-gpt3.html> [perma.cc/M33H-AB7A].

Montezuma, Luis Alberto & Tara Taubman-Bassirian, “How to Avoid Consent Fatigue” (29 January 2019), online: *IAPP* <iapp.org/news/a/how-to-avoid-consent-fatigue/> [perma.cc/77A7-M7PR].

Montreal Declaration For a Responsible Development of Artificial Intelligence (Montreal: Université de Montréal, 2018), Preamble, online: *Université de Montréal* <www.montrealdeclaration-responsibleai.com/> [perma.cc/VC4F-LEVU].

Moore, Louise, “Why ‘Ditch the algorithm’ is the future of political protest,” (19 August 2020), online: *The Guardian* <www.theguardian.com/commentisfree/2020/aug/19/ditch-the-algorithm-generation-students-a-levels-politics> [perma.cc/5ZHN-7YDS].

Nietzsche, Friedrich, *On Truth and Lies in a Nonmoral Sense* (1873), online (pdf): *AustinCC* <www.austincc.edu/adechene/Nietzsche%20on%20truth%20and%20lies.pdf> [perma.cc/2HF2-KVRA].

Northpointe, “Practitioner’s Guide to COMPAS Core” (19 March 2015), online (pdf): <<https://assets.documentcloud.org/documents/2840784/Practitioner-s-Guide-to-COMPAS-Core.pdf>> [perma.cc/AB6P-AACT].

NPR, “How 5 Tech Giants Have Become More Like Governments Than Companies (26 October 2017), online: *NPR* <www.npr.org/2017/10/26/560136311/how-5-tech-giants-have-become-more-like-governments-than-companies> [perma.cc/QQ4Q-YXKW].

Oliveri, Scott, “Skeuomorphism: Design We Learned To Outgrow” (16 June 2020), online: *Medium* <medium.com/design-warp/skeuomorphism-design-we-learned-to-outgrow-8a24895a80d0> [perma.cc/49ME-DPVK].

Oliveri, Scott, “The floppy disk Save icon: Visual language of an era long-gone” (19 August 2020), online: *UX Design* <uxdesign.cc/the-floppy-disk-save-icon-visual-language-of-an-era-long-gone-93f74efc9f9> [perma.cc/RN7H-SDVW].

OPC News Release, “Clearview AI ceases offering its facial recognition technology in Canada” (6 July 2020), online: *OPC* <www.priv.gc.ca/en/opc-news/news-and-announcements/2020/nr-c_200706/> [perma.cc/55B6-SN98].

OPC, “Commissioners launch joint investigation into Clearview AI amid growing concerns over use of facial recognition technology” (21 February 2020), online: *OPC* <www.priv.gc.ca/en/opc-news/news-and-announcements/2020/an_200221/> [perma.cc/F6UW-V576].

OPC, “Joint investigation of Clearview AI, Inc. by the Office of the Privacy Commissioner of Canada, the Commission d’accès à l’information du Québec, the Information and Privacy Commissioner for British Columbia, and the Information Privacy Commissioner of Alberta” PIPEDA Findings #2021-001 (2 February 2021), online: *OPC* <www.priv.gc.ca/en/opc-actions-and-decisions/investigations/investigations-into-businesses/2021/pipeda-2021-001> [perma.cc/M5MP-NREC].

OpenAI, “Code completion” (last visited 19 July 2022), online: *OpenAI* <beta.openai.com/docs/guides/code> [perma.cc/62Z2-DKG2].

OpenAI, “Embeddings” (last visited 7 August 2022), online: *OpenAI* <beta.openai.com/docs/guides/embeddings/what-are-embeddings> [perma.cc/RFB5-497S].

OpenAI, “Fine-tuning” (last visited 19 July 2022), online: *OpenAI* <beta.openai.com/docs/guides/fine-tuning> [perma.cc/D7ZS-VL6T].

OpenAI, “OpenAI API” (11 June 2020), online: *OpenAI Blog* <openai.com/blog/openai-api/> [perma.cc/RY2A-G9EH].

OpenAI, “OpenAI Charter” (9 April 2018), online: *OpenAI* <openai.com/charter/> [perma.cc/UY4D-ZLDE].

OpenAI, “Pricing” (last visited 7 August 2022), online: *OpenAI* <openai.com/api/pricing/> [perma.cc/JWD4-GX9Z].

OpenAI, “Text completion” (last visited 19 July 2022), online: *OpenAI* <beta.openai.com/docs/guides/completion/introduction> [perma.cc/JJ5E-2GNX].

OpenAI, “Welcome to OpenAI” (last visited 19 July 2022), online: *OpenAI* <beta.openai.com/> [perma.cc/ZV54-7RM8].

Oxford Internet Institute, “Fairness, Accountability, and Transparency in Machine Learning” (last accessed 31 July 2022), online: *OII* <www.oii.ox.ac.uk/study/courses/introduction-to-fairness-accountability-and-transparency-in-machine-learning/> [perma.cc/9BWW-RWHZ].

Paul, Andrew, “Remember Twitter’s racist photo-cropping tool? It’s still a thing.” (9 December 2020), online: *InputMag* <www.inputmag.com/culture/remember-twitters-racist-photo-cropping-tool-its-still-a-thing> [perma.cc/JPW4-8KUA].

Perrault, Raymond, et al, “The AI Index 2019 Report”, AI Index Steering Committee, Human-Centered AI Institute, Stanford University, December 2019, online: *Stanford* <hai.stanford.edu/ai-index-2019> [perma.cc/2ZQC-LZVT].

Phillips, P Jonathan & Fang Jiang, et al, “An Other-Race Effect for Facial Recognition Algorithms” (2011) 8 ACM Trans On Applied Perception, online: *Semantic Scholar* <pdfs.semanticscholar.org/5fb9/4eaf71196ed9cb87bbd881af34be8b4bc919.pdf> [perma.cc/D3SQ-JNEQ].

Price, Rob, “Microsoft is deleting its AI chatbot’s incredibly racist tweets” (24 March 2016), online: *Business Insider* <www.businessinsider.com/microsoft-deletes-racist-genocidal-tweets-from-ai-chatbot-tay-2016-3?r=UK&IR=T> [perma.cc/5EB5-H5CE].

Privacy & Security, “Facebook’s Facial Recognition Software Is Different From The FBI’s. Here’s Why” (18 May 2016), online: *NPR* <www.npr.org/sections/alltechconsidered/2016/05/18/477819617/facebooks-facial-recognition-software-is-different-from-the-fbis-heres-why> [perma.cc/4JUZ-MUQ3].

Quote Investigator, “It’s Difficult to Make Predictions, Especially About the Future” (20 October 2013), online: <quoteinvestigator.com/2013/10/20/no-predict/> [perma.cc/3P6Z-9WJN].

Raymer, Elizabeth, “Right to privacy not an all-or-nothing concept, says SCC in voyeurism case” (14 February 2019) online: *Canadian Lawyer* <www.canadianlawyermag.com/practice-areas/privacy-and-data/right-to-privacy-not-an-all-or-nothing-concept-says-scc-in-voyeurism-case/275878> [perma.cc/8NEJ-S55B].

Robertson, Adi, “Reddit bans ‘deepfakes’ AI porn communities” (7 February 2018), online: *The Verge* <www.theverge.com/2018/2/7/16982046/reddit-deepfakes-ai-celebrity-face-swap-porn-community-ban> [perma.cc/2DJF-A4XV].

Rogers, Adam, “The Way the World Ends: Not With a Bang But With a Paperclip” (21 October 2017), online: *WIRED* <www.wired.com/story/the-way-the-world-ends-not-with-a-bang-but-a-paperclip/> [perma.cc/3MD2-4XF4].

Ruiz, David, “Artificial Intelligence ban slammed for failing to address ‘vast abuse potential’” (April 23, 2021), online: *MalwareBytes Labs* <blog.malwarebytes.com/artificial-intelligence/2021/04/artificial-intelligence-ban-slammed-for-failing-to-address-vast-abuse-potential/> [perma.cc/S9MB-BPFD].

Sabeti, Arram, “GPT-3: An AI that’s eerily good at writing almost anything” (9 July 2020), online: *Arr.am* <arr.am/2020/07/09/gpt-3-an-ai-thats-eerily-good-at-writing-almost-anything/> [perma.cc/D7U9-Q244].

Salazyn, Amy, “A Taxonomy for Lawyer Technological Competence” (18 December 2020), online: *SLAW* <www.slw.ca/2020/12/18/a-taxonomy-for-lawyer-technological-competence/> [perma.cc/PH8P-DFXF].

Salyzyn, Amy, “It’s Finally (Sort Of) Here!: A Duty of Technological Competence for Lawyers” (26 November 2019) online: *SLAW* <www.slw.ca/2019/11/26/its-finally-sort-of-here-a-duty-of-technological-competence-for-canadian-lawyers/> [perma.cc/6PU2-CFN6].

Scassa, Teresa, “Open Season on Social Media Data: (ETHI’s Report on PIPEDA Reform – Part II” (22 March 2018), online: *Teresa Scassa* <www.teresascassa.ca/index.php?option=com_k2&view=item&id=273:open-season-on-social-media-data-ethi-report-on-pipeda-reform-part-ii&Itemid=80> [perma.cc/H879-B38E].

Scassa, Teresa, “Open Season on Social Media Data: ETHI’s Report on PIPEDA Reform – Part II” (22 March 2018), online: *Teresa Scassa* <www.teresascassa.ca/index.php?option=com_k2&view=item&id=273:open-season-on-social-media-data-ethi-report-on-pipeda-reform-part-ii&Itemid=80> [perma.cc/L5ZB-5GG6].

Scassa, Teresa, “Supreme Court of Canada Decision Has Relevance for Addressing Bias in Algorithmic Decision-Making” (14 June 2018), online: *Teresa Scassa* <www.teresascassa.ca/index.php?option=com_k2&view=item&id=278:supreme-court-of-canada-decision-has-relevance-for-addressing-bias-in-algorithmic-decision-making&Itemid=80> [perma.cc/2LVZ-377R].

Scassa, Teresa, “The Post-pandemic Future of Trust in Digital Governance” (20 July 2020), online: *CIGI* <www.cigionline.org/articles/post-pandemic-future-trust-digital-governance> [perma.cc/SZU8-YXVS].

Schmelzer, Ron, “Is Machine Learning Really AI?” (21 November 2019), online: *FORBES* <www.forbes.com/sites/cognitiveworld/2019/11/21/is-machine-learning-really-ai/#69ab44a52621> [perma.cc/QM4N-SN74].

Scott, Grace Lisa, “Deepfakes Still Exist, Amid Twitter and Pornhub Bans” (6 February 2018), online: *Inverse* <<https://www.inverse.com/article/41017-pornhub-says-it-s-banning-deepfakes-as-the-internet-cracks-down-on-ai-porn>> [perma.cc/3WP4-44FY].

Shameem, Sharif, “Here’s a sentence describing what Google’s home page should look and here’s GPT-3 generating the code for it nearly perfectly. (5 July 2020, 3:50 AM), online: *Twitter* <twitter.com/sharifshameem/status/1283322990625607681>.

Shefman, Corey, “Wet’suwet’en blockades: Rule of law is a convenient weapon” (28 February 2020), online: *The Hamilton Spectator* <www.thespec.com/opinion/contributors/2020/02/28/wet-suwet-en-blockades-rule-of-law-is-a-convenient-weapon.html> [perma.cc/9RCN-7BED].

Simonite, Tom, “Facebook Can Now Find Your Face Even When Its Not Tagged” (19 December 2017), online: *WIRED* <www.wired.com/story/facebook-will-find-your-face-even-when-its-not-tagged/> [perma.cc/UA3U-T27F].

Simonite, Tom, “What Really Happened When Google Ousted Timnit Gebru” (8 June 2021), online: *WIRED* <www.wired.com/story/google-timnit-gebru-ai-what-really-happened/> [perma.cc/TL5X-VR4Z].

Sioli, Lucilla, “Shaping Europe’s Digital Future: CEPS Webinar – European approach to the regulation of artificial intelligence” (April 23, 2021), online: *Centre for European Policy Studies* <www.ceps.eu/wp-content/uploads/2021/04/AI-Presentation-CEPS-Webinar-L.-Sioli-23.4.21.pdf> [perma.cc/28KZ-NC37].

SocialNLP 2022, “The Tenth International Workshop on Natural Language Processing for Social Media: Proceedings of the Workshop” (14-15 July 2022), online (pdf): *ACL Anthology* <aclanthology.org/2022.socialnlp-1.0.pdf> [perma.cc/R3R5-JJPU].

Sprenger, Polly, “Sun on Privacy: ‘Get Over It’” (26 January 1999), online: *WIRED* <archive.wired.com/politics/law/news/1999/01/17538> [perma.cc/Y5HG-ANF3].

Stromberg, Joseph, “Bicycle face: a 19th-century health problem made up to scare women away from biking” (24 March 2015), online: *Vox* <www.vox.com/2014/7/8/5880931/the-19th-century-health-scare-that-told-women-to-worry-about-bicycle> [perma.cc/Q5T7-9TNU].

Szilagyi, Katie, “PhD GPT-3 Experiments” (last visited 10 August 2022), online: <katieszilagyi.com/phd-gpt-3-experiments/> [perma.cc/4SSX-2YLB].

The International Go Federation, “Go Population Survey” (February 2016), online: *InterGo* <www.intergofed.org/wp-content/uploads/2016/06/2016_Go_population_report.pdf> [perma.cc/EK7A-EAJA].

Thomasen, Kristen, et al, “Submission to the Toronto Police Services Board’s Use of New Artificial Intelligence Technologies Policy – LEAF and the Citizen Lab” (20 December 2021),

online (pdf): *SSRN* <papers.ssrn.com/sol3/papers.cfm?abstract_id=3989271> [perma.cc/MB3J-MATE].

Thunström, Almira Osmanovic, “We Asked GPT-3 to Write an Academic Paper about Itself—Then We Tried to Get It Published” (30 June 2022), online: *Scientific American* <www.scientificamerican.com/article/we-asked-gpt-3-to-write-an-academic-paper-about-itself-then-we-tried-to-get-it-published/> [perma.cc/VBE7-PCQG].

Thunström, Almira Osmanovic, (@augmentedrobot), “281 years in the making: GPT-3 and de la Mettrie” (26 January 2022), online: *Medium* <augmentedrobot.medium.com/281-years-in-the-making-gpt-3-and-de-la-mettrie-de03eea374e4 > [perma.cc/Y95C-9A5T].

Toronto Metropolitan University, “Legal Innovation Zone” (last accessed 29 July 2022), online: <www.torontomu.ca/zone-learning/legal-innovation-zone/about/> [perma.cc/XFX6-M7LM].

Tranter, Kieran “Nomoi are Mercurial” (7 February 2022), online: *The Digital Constitutionalist* <https://digi-con.org/nomoi-are-mercurial-why-cultural-legal-studies-of-science-fiction-matter/> [perma.cc/3B8K-EEJU].

Truth & Reconciliation Commission of Canada, “Calls to Action” (2015), online (pdf): <www2.gov.bc.ca/assets/gov/british-columbians-our-governments/indigenous-people/aboriginal-peoples-documents/calls_to_action_english2.pdf>.

Tufecki, Zeynep, “The Latest Privacy Debacle” (30 January 2018), online: *The New York Times* <www.nytimes.com/2018/01/30/opinion/strava-privacy.html> [perma.cc/35G3-P2KE].

University of Alberta Faculty of Law, “2022 Legal Innovation Conference: Introducing & Embracing Tech & Automation in the Legal Industry” (28 January 2022), online: *UAlberta* online: <www.ualberta.ca/law/about/legal-innovation.html> [perma.cc/5SKA-VE87].

UW Bothwell & Cascadia College, Library Guides, “News: Fake News, Misinformation & Disinformation” (17 March 2022), online: *UW Guides* <guides.lib.uw.edu/c.php?g=345925&xp=7772376> [perma.cc/WN3Y-ABZM].

Veel, Paul-Erik, “Getting Leave to Appeal to the Supreme Court: Empirical Insights from Tax Cases” (20 January 2021), online: *Lenczner Slaght Blog* <litigate.com/OnTheDocket#/getting-leave-to-appeal-to-the-supreme-court-empirical-insights-from-tax-cases> [perma.cc/F6ES-PWQ5].

Vincent, James, “Twitter taught Microsoft’s AI chatbot to be a racist asshole in less than a day” (24 March 2016), online: *The Verge* <www.theverge.com/2016/3/24/11297050/tay-microsoft-chatbot-racist> [perma.cc/PR6A-CA7W].

Vinsel, Lee, “You’re Doing It Wrong: Notes on Criticism and Technology Hype” (1 February 2021), online: *Medium* <sts-news.medium.com/youre-doing-it-wrong-notes-on-criticism-and-technology-hype-18b08b4307e5> [perma.cc/WN6G-VA5N].

Vyas, Kashyap, “Deep Blue vs Kasparov: The Historic Contest That Sparked the AI Revolution”, (8 August 2019), online: *IE* <interestingengineering.com/deep-blue-vs-kasparov-the-historic-contest-that-sparked-the-ai-revolution> [perma.cc/4BC5-RT85].

Weise, Karen, “Amazon indefinitely extends a moratorium on the police use of its facial recognition software” (18 May 2021), online: *The New York Times* <www.nytimes.com/2021/05/18/business/amazon-police-facial-recognition.html > [perma.cc/8B8J-Q9SY].

West, Robert, et al, “Wikispeedia: An Online Game for Inferring Semantic Distances between Concepts,” Proceedings of the Twenty-First International Joint Conference on Artificial Intelligence (IJCAI-09), online: *Stanford* <infolab.stanford.edu/~west1/pubs/West-Pineau-Precup_IJCAI-09.pdf> [perma.cc/XE8F-RXYY].

Wikipedia, “Jews” (last visited 19 July 2022), online: *Wikipedia* <en.wikipedia.org/wiki/Jews> [perma.cc/AV6R-YCY6].

Wilhelm, Alex, “Tech’s 5 Biggest Players Now Worth 3 Trillion,” (19 July 2017), online: *TechCrunch* <techcrunch.com/2017/07/19/techs-5-biggest-players-now-worth-3-trillion/> [perma.cc/G8FK-5E7H].

Willem DeVries, “Wilfrid Sellars” Stanford Encyclopedia of Philosophy, online: *SEP* <plato.stanford.edu/entries/sellars/> [perma.cc/BQM9-M8H6].

Yong, Ed, “A Popular Algorithm Is No Better at Predicting Crimes Than Random People” (January 17, 2018), online: *The Atlantic* <www.theatlantic.com/technology/archive/2018/01/equivant-compas-algorithm/550646/> [perma.cc/W8GW-TJSE].

Secondary Materials: Reports

ACLU, “Cracks in the System: Twenty Years of Unjust Federal Crack Cocaine Law” (October 2006) online: *ACLU* <www.aclu.org/other/cracks-system-20-years-unjust-federal-crack-cocaine-law> [perma.cc/RR6Q-J22G].

Buolamwini, Joy, Vicente Ordóñez, Jamie Morgenstern, & Erik Learned-Miller “Facial Recognition Technologies: A Primer” (29 May 2020), online (pdf): *UMass CS* <people.cs.umass.edu/~elm/papers/FRTprimer.pdf> [perma.cc/4BCD-M4SC].

Diagnostic Report, Steering Committee on Big Data, ATIP request by Lex Gill, online: *Google Drive* <drive.google.com/file/d/1HpVgzMdf7SPH319iNA2wkkQkDwDax7F7/view> [perma.cc/L5KM-AGER].

Garvie, Clare, Alvaro Bedoya & Jonathan Frankle, “The Perpetual Lineup: Unregulated Police Face Recognition in America” Center for Privacy & Technology, Georgetown Law (18 October 2016), online: *Perpetual Lineup* <www.perpetuallineup.org > [perma.cc/Q2QL-P44A].

Lynch, Jennifer, “Face Off: Law Enforcement Use of Face Recognition Technology” (12 February 2018), online (pdf): *EFF* <www.eff.org/wp/law-enforcement-use-face-recognition > [perma.cc/X5A7-LNES].

Molnar, Petra & Lex Gill, “Bots at the Gate: A Human-Rights Analysis of Automated Decision-Making in Canada’s Immigration and Refugee System” (2018), online (pdf): *Citizen Lab* <citizenlab.ca/wp-content/uploads/2018/09/IHRP-Automated-Systems-Report-Web-V2.pdf> [perma.cc/3DZT-9XP2].

New York City Bar Association, “Power, Pervasiveness and Potential: The Brave New World of Facial Recognition Through a Criminal Law Lens (and Beyond)” (14 August 2020), online (pdf): *New York City Bar* <s3.amazonaws.com/documents.nycbar.org/files/2020662-BiometricsWhitePaper.pdf> [perma.cc/QEW3-ETWX].

Northpointe Inc, “COMPAS Risk & Need Assessment System: Selected Questions Posed by Inquiring Agencies” (2012).

Robertson, Kate, Cynthia Khoo, & Yolanda Song, “To Surveil and Protect: A Human Rights Analysis of Algorithmic Policing in Canada” (2020), online (pdf): *Citizen Lab* <citizenlab.ca/wp-content/uploads/2020/09/To-Surveil-and-Predict.pdf > [perma.cc/A7C5-R8WK].

Robinson, David & Logan Koepke, “Stuck in a Pattern: Early evidence on “predictive policing” and civil rights” (August 2016), online: *Upturn* <www.stuckinapattern.org> [perma.cc/N9K2-C5ZS].