Tech, Trust, and Legitimacy: Discourse Ethics and the Adoption of Artificial Intelligence Technologies in Policing

Janis Steinbachs

MA Thesis Submitted to the Faculty of Philosophy,
School of Public Affairs, Saint Paul,
Saint Paul University,
in partial Fulfillment of the requirements for the
degree of Master of Arts in Public Ethics

Ottawa, Canada
August 2025

Contents

Chapter One: Introduction: AI, Challenges to Police Trust and Legitimacy, and Habermas

In recent years, rapidly developing artificial intelligence (AI) technologies have been recognized for their potential to advance and improve the tactics and strategies of police services in Canada and worldwide (Emerson and Epstein IACP). They hold promise for improved public safety outcomes by offering transformational solutions and potentially enhancing how police investigate and prevent crime, gather data, and identify actionable patterns to promote community safety (BBC).

AI technologies have also provoked significant debate on issues related to privacy, bias, and accidental or purposeful misuse (Ezzeddine et al.). Globally, security services increasingly use these technologies to control dissent and populations (Zhang).

Canadian police services have also begun exploring potential uses for these technologies, developing policies and plans for their future adoption as part of a debate that focuses on commercial availability, efficiency, and effectiveness (Toronto Police Service [TPS]). At the same time, police departments delivering local or municipal service must continue building trust and legitimacy within the communities they serve, especially as public institutions face serious community concerns, declining trust levels, and increasing questions about their legitimacy. Traditionally, police and other institutions have sought legitimacy by accessing the will-formation aspects of the public sphere to understand the opinions and views of the people they serve (Habermas, *Between* 304). However, in recent years, the public sphere has become disrupted by new technologies, such as social media, which have separated the "journalistically institutionalized" public sphere from "semi-public spheres" (Habermas, *New Structural* 54–56).

The media system, for example, which is crucially important to the public sphere and the formation of public opinion, is under threat as new media and increased corporate ownership

reduce its reach and ability to develop competing public opinions, which are central to deliberative politics.

Problem Statement

The promise of AI technology and its accompanying risks have provoked a rich debate among the community, police, philosophers, and ethicists focused on issues such as autonomy, fairness, accountability, and systemic bias. These are all important issues. My thesis focuses on the manner in which public institutions like the police can adopt AI technology. How can Canadian police adopt AI technologies in an ethical manner that aligns with public expectations and demands in a time when the public sphere is fragmented and disrupted?

Purpose and Objective of the Study

This thesis will examine how Habermas's theories on discourse ethics and the public sphere can help Canadian society and the police can achieve an ethical application of AI technologies, supporting its legitimate use while ensuring trust and legitimacy in policing. My goal is not to evaluate the ethics or effectiveness of AI; rather, I aim to discuss how police can approach the adoption of AI for ethical use through discourse. The objective of my study is to demonstrate that Habermas's discourse ethics, his views on the public sphere, and the connections he draws between deliberative democracy and communicative rationality provide a viable theoretical framework for Canadian police to adopt in critical discussions about key topics such as the use of AI. This study is meant to assist all Canadian police at the federal, provincial, and municipal levels. By adopting these frameworks and developing practical systems that ensure citizen participation in policy-setting and decision-making, the police can more effectively make decisions that align with, or at least demonstrably sensitive to, community

expectations, while advancing the interests of public safety. By engaging in these dialogues, police can also strengthen the public sphere's influence on the application of AI technologies.

The study is focused on answering key questions such as:

- How can Canadian police adopt AI technologies in a manner that is ethical and meets public expectations and demands in a time when the public sphere is fragmented and disrupted?
- How can Habermas's theories assist police with the ethical adoption of AI while building public trust and legitimacy?
- What deliberative systems can be implemented to approximate an ideal speech situation and enhance community understanding of AI?
- How can discourse ethics provide a community deliberation model that assists the police and the people they serve to understand and mitigate issues related to AI adoption?
- How can communicative rationality through discourse address inherent issues with instrumental reason in AI?

The Relevance of Jürgen Habermas

The theoretical basis of this study will build on the theories of Jürgen Habermas. Habermas's wide-ranging scholarship has illuminated and informed our understanding of the processes by which humans build their understanding of the world and form public opinion. He has informed our understanding of how humans can achieve emancipation through normative and rational decisions, while also focusing on the processes by which citizens demand that states act on their behalf, in accordance with their views and needs, and against "all forms of institutional repression" (Rasmussen 7–9).

I have chosen Habermas's work as the theoretical framework for this thesis based on his understanding of the classical public sphere, communicative rationality, discourse ethics, and deliberative processes that strengthen democracy. The themes relevant in the debate of policing and AI—ethical adoption, transparency, societal norms and expectations—mirror those Habermas explores in his work. Although Habermas's account of the classical public sphere has generated significant theoretical and empirical criticism, I intend to demonstrate its pertinence to the emergent debates around AI. Such an approach is after all consistent with Habermas himself, who recently (*New Structural*) updated his understanding of the public sphere to reflect new realities prompted by new and emerging media technologies. I believe that Habermas's theories on the public sphere and deliberative democracy, especially given his recent works, remain relevant in our media and communications centred society. Many of the most pertinent criticisms of his work expand, clarify, and sharpen his work as he works to incorporate alternative publics, pluralism, digital media, and other communications phenomena that have emerged in this century like "alternative facts".

Section One: Theoretical Framework

Habermas's writings include many of the philosophical themes and discussions in Western societies and his theories allow us to rationally criticize and analyze societal issues. Habermas's exhaustive understanding of Western thought and ongoing scholarship offer valuable insights that contribute to solutions for improving society.

Habermas is widely recognized for his ability to identify and highlight the significant connection between successful deliberative democracy and free and unrestricted debate and discussion (Chambers, "Deliberative" 310). This discussion and consensus leads to societal integration and legitimation. He is also clear that instrumental or strategic communication, with

its orientation toward success (unlike communicative action's orientation toward reaching understanding), can negatively impact and disrupt humans' ability to communicate and act on their will (Habermas, *Theory* 1: 286).

It was Habermas who identified the public sphere as a historically contingent realm of discourse, where public opinion could be formed about the issues of the day (*The Structural* 27–30). This in turn leaves the modern democratic state in a position where it must be responsive to these opinions and adapt them into decisions and policy-making.

Habermas's conception and critique of the public sphere will be central to my thesis (*The* Structural 27–30). It has been updated by Habermas and refined through criticism. I will argue that institutions like the police require a functioning, healthy public sphere where will formation exists to ensure legitimacy by responding to those wills. Institutions in a democratic state must find legitimacy in their own resources through democratic will formation (Habermas, *New Structural* 12). I will also analyze the impact of a disrupted public sphere and raise concerns about the impact it has on institutions like the police in seeking legitimation and building trust in the community that it is acting according to their will.

The thesis will argue that Habermas's theories on the classical public sphere, the lifeworld, and his discourse ethics are important and relevant tools for understanding community concerns and expectations in relation to AI. This thesis also proposes a way forward in these discussions that supports deliberative democracy, communicative rationality, and the legitimacy of the police in serving the community by providing a theory of ethics that connects to the politics and deliberation of our time.

The theory of communicative action addresses the use of speech acts to reach a mutual understanding that plays into action coordination (Habermas, *Theory* 1: 101). Habermas's

discourse ethics connects communicative action to real-world decisions, founded on the idea that through communicative rationality, citizens can reach solutions using principles that ensure all affected parties can participate in discourse with the goal that they can, or would, agree to the norms decided upon. Habermas argues that discursively attained consensus, attained through enabling conditions, allows humans to take propositions as true (*Moral* 58). I will argue that discourse ethics provides the basis for community policing models that lead to normative and accepted decisions on issues like the use of AI.

Other Relevant Scholarship on Habermas

There is significant scholarship and discussion on Habermas, with many scholars recognizing the relevance of his theories in the public sphere, the use of language, deliberative democracy, and discourse on AI. In recent years, several scholars have applied Habermas's theories to issues related to policing and the use of AI by administrative agencies. McCandless and Vogler argue that Habermas's theories on communicative action could help explain distrust in police by "a lack of communicative acts in favor of strategic acts emphasizing crime control that maintains unjust systems" (103). Paolo Monti applies a Habermasian assessment of large language models as "atypical participants" in discourse, advancing our understanding of AI's role in human discussion (Monti 62). Dongwoo Kim discusses the use of AI by administrative agencies, focusing on the threat these technologies pose to democratic systems and human autonomy. They argue that communicative action can assist with the development of AI governance (83). Xivuri and Twinomurinz argue that Habermas can be used to develop process frameworks to ensure fairness is embedded within AI algorithm development before and after implementation (336). Higgitt explores the alignment between Habermas's views on human reason with the connectionist approach of AI development (83). Rasmussen argues that discourse

ethics provides a "procedural justification" for truth and validity claims (63). Chambers also credits Habermas's contribution to deliberative democracy with its focus on rational debate, equality, intersubjective understanding, and transparency (*Deliberative* 310).

There is also notable criticism of Habermas, which this thesis will explore and address. Scholars such as Nancy Fraser have critiqued Habermas, arguing that he has idealized the public sphere and fails to address real inequalities that exist in society (136–137). According to her argument, Habermas's conception fails to address the lack of equal voice and the gendered nature of the public and private divide. She also notes that in any discussion in the sphere, power imbalances result in some voices being amplified while others are silenced. Finally, she argues that Habermas underplays and fails to identify what she terms "subaltern counter publics." She believes these mini-public spheres are "discursive arenas where members of subordinated social groups invent and circulate counter discourses, which in turn permit them to formulate oppositional interpretations of their identities, interests, and needs" (Fraser 67).

Section Two: The Public Sphere and the Risk of a Declining Media System

In Section Two, I will discuss the current state of the public sphere in Canada, including the introduction of social and new media. I will argue that the ongoing colonization of the public sphere by strategic or instrumental rationality presents a threat to both the community and the police, hindering the ability to build rational consensus on the adoption of AI. A disrupted public sphere challenges the public's ability to develop a will and for institutions like the police to hear, understand, and act on it. I will use a variety of quantitative and qualitative sources to explain the current disrupted nature of the public sphere. I will also discuss how new communication mediums such as social media shape will formation and undermine ethical dialogues by blurring the public and private spheres.

Habermas continues to publish on this topic and is preoccupied by what he sees as a decline in the attention paid to political news and the analysis of those events (*New Structural* 43). He is further concerned by the emergence of issues like fake news and declining trust in publicly funded broadcasters. The purpose of this analysis is to highlight Habermas's concerns about the decline of the media system and its impact on institutions like the police in understanding the community's will formation. Habermas is concerned that in some subcultures, the public sphere is no longer seen as inclusive or a place "for a generalization of interests that includes all citizens" (54). I will suggest how his recent concerns align with the Canadian reality in general and the discourse on policing in particular.

Section Three: AI Technologies in Policing and Risks

Police services are well-known for recording and maintaining vast amounts of data related to their daily operations, such as community concerns, crime reports, informants, evidence, and other intelligence. In the last few decades, increasing efforts have been made to improve the storage and accessibility of this data, with stated goals of efficiency and community safety, among others. AI has become recognized as a powerful new tool for these efforts.

The focus of this section will be on police decisions to adopt these technologies, particularly in the realms of:

- Predictive policing and crime prevention tools
- Facial recognition
- Crime analysis and prevention
- Data analysis

I will argue that by relying on the instrumental and strategic rationality of these technologies, the police risk negatively impacting trust and legitimacy.

This section will also provide a brief review of the key ethical issues and high-level risks related to the police adoption of AI tools including:

- Impact on privacy
- Systemic bias and inherent bias in data
- Technochauvinism
- Infringements on human and charter rights
- Public distrust

I will discuss how these technologies are an emerging priority but are largely commercially driven, and the risk that even well-intentioned algorithmic systems could have disproportionate consequences on vulnerable and marginalized communities (Marda 4).

I will also explore the risk that adopting AI could lead police to further misunderstand the lifeworld. The fact that a decision is "data-driven" does not mean it is fair, just, accurate, or appropriate (6).

Section Four: Applying Habermas

In this section, I will discuss my main arguments. I will argue that while AI technologies hold significant promise for enabling the police to improve community safety, these technologies must be legitimatized through discourse in line with Habermas's model for police to maintain public trust. I will argue that key to this model is an open and transparent dialogue in the public sphere, using experts and non-experts from the community, especially marginalized communities.

I will argue that current police–community models for deliberation, including the Philadelphia model, provide guidelines on how these discussions can be modelled and actioned. I

will argue that poor AI adoption could negatively impact and further colonize the lifeworld, and that governing discourse could be a first significant step in community–police dialogue.

Chapter Two: A Habermasian Compass for the Ethical Adoption of AI

Over the last few decades, police services have seen advancements in computers, data collection, and communication tools that were unheard of and impossible just one generation ago (Johnson 1-1). The rapid pace of change has brought new capabilities to a data-rich industry like policing, with new abilities such as facial recognition, automatic license plate reading, body-worn cameras, crime prediction/prevention technologies, advanced crime mapping, data mining, and other machine learning (ML) or AI technologies that promise to increase efficiency and improve public safety outcomes (Chen). Police services have long been known to collect massive amounts of data but have struggled to find solutions to sorting and using that data due to staffing and technical limitations. AI-based tools developed by large technology companies are a new solution to that problem, offering the ability to collect, analyze, and utilize data cheaply and efficiently (Thomson Reuters 3). In a recent interview in *Police Chief Magazine*, a tech expert summarized the current debate in AI and policing: "For the law enforcement community, technology can help make all the things police departments need to do better, more effective, and more efficient" (Chen).

This thesis will review the current debate on AI adoption by police services in Canada and caution against an instrumental approach. For my purposes, and in the spirit of the Frankfurt School of critical theory, an instrumental approach is defined as the most efficient and effective ways or means to achieve an end and does not place "importance on whether the purposes are reasonable" (Horkheimer, *Eclipse* 3-5).

Instead, I will engage with the work of Jürgen Habermas, specifically his views on the public sphere, communicative action, and discourse ethics. My argument is twofold. I will argue that Habermas's work provides a framework for adopting this technology that engages the

public, builds trust and legitimacy in policing, and fosters public support for the technologies being adopted. Furthermore, Habermas's theories on the public sphere, along with his updated work, offer valuable insights for institutions like the police, which rely on the public sphere to understand the will and opinion of the public. A disrupted public sphere threatens police legitimacy; therefore, police must ensure that the frameworks developed to understand public opinion on important topics like AI are ethical and deliberative, ensuring that decisions regarding its use reflect and contribute to legitimacy.

In this normative analysis, I argue that AI adoption has significant potential to improve public safety, but its use by the police must be guided by ethical and democratic discourse with the public they serve. While AI has significant potential to assist police, ethicists are increasingly examining the issue of its adoption to understand its societal impact. Valid concerns remain about AI and "discrimination and fairness, privacy and human autonomy in semi-automated decision-making, risks of individual and social surveillance or threats to democracy through dynamic misinformation in social media" (Häußermann and Lütge 341). The emerging field of AI ethics seeks to establish normative approaches, both in theory and practice, that minimize or mitigate risks and enhance the benefits of the technology. In this regard, I seek to examine the potential of Habermas's theories to boost the deliberative and democratic nature of dialogue with the people police serve and develop a framework and potential approaches to ensure that ethical discourse guides police adoption of AI.

Trust and Legitimacy

The terms trust and legitimacy are often used in policing. Before continuing, it is worth briefly touching on the foundation of these terms. It is recognized that police services rely on the voluntary cooperation of the public to maintain order, prevent and solve crime, and carry out

their duties (Mazerolle and Ransley 4). A defining feature of police legitimacy is that "people feel obliged to voluntarily comply" without fear of punishment or expectations of reward. In their study "Legitimacy and Cooperation: Why Do People Help the Police Fight Crime in Their Communities?," Tom R. Tyler and Jeffrey Fagan argue that "legitimacy is a normative judgment." In modern society, authorities can benefit from voluntary cooperation that is not obtained through power, sanctions, or incentives (239). This legitimacy is built through "process-based policing," where the police exercise their authority in "ways that members of the public evaluate as fair" (264). For police to be successful, they must have active public cooperation, often attained by demonstrating "the fairness of the processes the police use when dealing with members of the public (267).

Police in Canada currently enjoy a high level of trust, according to several studies, including one by Statistics Canada. From October 2022 to January 2023, about two-thirds (67%) of Canadians reported having high confidence in the police. This makes policing one of the most trusted institutions in Canada, higher than the Federal Parliament, the justice system and courts, and the Canadian media. Confidence in these institutions reflects the sense that they are safe, effective, transparent, and accountable ("Confidence in the Police").

While trust remains high, many Canadian researchers have recognized that long-identified discriminatory practices have disproportionately affected marginalized and racialized communities, particularly in the indigenous and Black communities. This includes racial profiling, harsher arrest decisions, as well as obtaining citizens' personal information through "carding or street checks. Evidence suggests that these practices have a significant impact on perceptions and support for the police, particularly from racialized communities" (Samuels-Wortley 138).

This dialogue involves police services across Canada that are taking steps through recruitment, training, policy change, and other evaluations to develop systems aimed at building trust. In a special study conducted for the Canadian Association of Chiefs of Police, researchers comprising senior Canadian police leaders recognized the unique situation Canada is in. The "Canadian policing community may be in good shape concerning the trust we enjoy from our public ... However … If we rest on our past laurels, we will almost certainly see a continuous decline in the trust that Canadians place in their police" (Canadian Association of Chiefs of Police 8).

Utilizing Habermas

This chapter provides a theoretical basis for applying Habermas's theories to the adoption of AI technologies by police in Canada. I will argue that Habermas's approach provides a unique perspective on the issue that can help society bridge the gap between groundbreaking technology, commercialization, technocracy, and public service. My goal is to highlight emancipatory portions of Habermas's views on the public sphere and discourse ethics to provide a normative critique and framework to guide the police and community in developing an AI adoption process that is ethical, transparent, deliberative, and viewed as legitimate by the population. This process should be reproducible. Habermas's critique and understanding of the bourgeois public sphere is central to my arguments, as is his discourse ethics, which is grounded in communicative rationality. I will discuss the impact of the disintegration of the public sphere and its influence on policing and issues of privacy. I argue that Habermas's theoretical framework and his work on discourse ethics provide an approach focused on consensus-building and rational decision-making. By emphasizing an inclusive, rational, and transparent approach, Habermas's work can help ensure that the will of the public is reflected in the adoption of new AI technology and that

critical technocratic decisions are determined with public input in an ethical manner that is socially acceptable and does not further colonize the lifeworld.

Chapter Outline

In this chapter, I relay my understanding of Habermas's theories and relate them to issues regarding the adoption of AI through ethical discourse.

I will argue that:

The *public sphere,* defined by Habermas, provides a conceptual area where inclusive, open, and ongoing dialogue can occur among members of the public. Within this sphere, both the police and society benefit when the use of AI technologies is transparently discussed, and the public will on how to proceed is formed through rational-critical debate. This debate and will formation are vital to circumventing corporate and technocratic interests and decision-making and building trust in police adoption of AI.

This approach will also assist in preventing the further colonization of the *lifeworld* by bureaucracy and technocracy (definition of technocracy). The approach defended herein allows the lived experiences, values, and societal background represented in the lifeworld to influence the decisions that affect it.

For Habermas, the lifeworld is a background in which all humans live, where shared understandings of culture, knowledge, and social norms shape daily interactions. "The shared lifeworld offers a storehouse of unquestioned cultural givens" that provides the background that we draw on to reach understanding and interpretation (*Moral Consciousness* 135).

Bureaucracy is a system that concentrates "the collectivity's capacity for action," designed to steer and administrate certain societal functions through instrumental rationality, rules, and procedures focused on efficiency (*Theory* 2: 171). It is not part of the lifeworld, and

while it can be effective, it also threatens the lifeworld through colonization, disruption, and ignoring the needs of individuals (Habermas, *Truth* 81).

Technocracy arises from the system. It is an overreliance on technical rationality, influencing society by depoliticizing the public sphere, reducing political issues meant for public discussion into something to be solved by technical experts. This leads to the "elimination of the distinction between the practical and technical" (Habermas, *Toward* 112–113).

Habermas's discourse ethics provides a framework for police to engage all stakeholders and individuals in justifying societal norms related to AI through critical-rational dialogue. The concept of his discourse ethics is focused on success based on the universal acceptability of decisions. This approach involves all who are affected but particularly helps address the concerns of those with the most to lose through AI usage, such as marginalized people and racialized groups, by hearing and addressing the issues and concerns raised. This discourse can provide a legitimized voice to individuals often excluded from decision-making, fostering ethical discourse that is inclusive and fair. Habermas's views on deliberative democracy argue for consensus-building through public deliberation. Roughly, the legitimacy of institutions backed by force or compulsion is at once *specified* and *shored up* by aligning them with the norms expressed and argued through a critical mass of citizen participation (*Between 110,301*). This emancipatory process helps ensure decisions are made through democratic processes.

The rationality derived through *communicative action* is based on developing mutual understanding rather than defining success through strategic outcomes. The Habermasian approach avoids the potential or real situation of top-down police decision-making, communications, and consultation that seek to inform the public about decisions rather than involve the public in deliberating on them. By engaging the public sphere and individuals and

groups through discourse, the police and the public can co-create a framework for AI adoption and governance. This approach more effectively aligns with the public will and reflects the demands of the public for greater police accountability.

Finally, there is increasing recognition of the need for transparency in AI adoption and a clearer understanding of how the technology works. Habermas's validity claims are crucial to this discussion. Truth, rightness, and truthfulness (*Moral* 58) must be satisfied. Corporate interests and proprietary ownership issues threaten transparency in this dialogue, and it is vital that the public understand how these tools function to determine if they can trust its usage.

Habermas's Public Sphere

Academics and theorists have long used Habermas's theories to provide a theoretical framework for discussing issues in multiple industries, communications, government, and public administration. Habermas himself has applied his theories to the legitimacy of law and other major issues in philosophy and contemporary discussion in works such as *Between Facts and Norms*. Even in the narrow context of policing, Schneider argues that using Habermas's theories could help address and eliminate power relationships, replacing them with more empathetic relationships between the police and the public (347). In 2019, McCandless and Vogler argued that Habermas's theory could help explain distrust in police due to the lack of communicative acts used by police "in favor of strategic acts emphasizing crime control that maintains unjust systems" (370). These theorists have provided a basis for discussing how to apply these theories in Canadian policing.

I approach this issue as a police civilian professional who has seen an increasing number of challenges to the current policing model. To address the complex challenges police services

face, including multiple jurisdictions and changing demographics, in subsequent chapters I will propose that the police require a framework that includes communicative rationality.

Habermas's Public Sphere

Habermas's classical description and definition of the public sphere is critical to this thesis and my understanding of the issue. In *The Structural Transformation of the Public Sphere*, Habermas describes the public sphere as an intermediary space between private people and the state, where citizens participate in rational-critical dialogue through various processes and media (*The Structural* 27). In his multiple works on this subject, Habermas notes the rise of the public sphere, beginning in the coffee houses and salons of Western Europe, followed by its disintegration and colonization through the system of government, power, mass media and other subsystems that represent power and money (*The Structural* 31-43, 141-157, 195-202). The critical literature has emphasized the need to adjust and flesh out the empirical, historical claims made by Habermas in the 1960s and further de-idealize his conception of the public sphere. These criticisms were made in light of gendered exclusions, new social movements, counterpublics, and marginalized voices raised by writers like Barker, Benhabib, and Ryan (Calhoun 1992). Nonetheless, I align with Habermas and, notably, even some of his most trenchant sympathetic critics in asserting that the concept of a public sphere in general retains its normative force (Fraser 109–142). In the following paragraphs, I will outline my understanding of the public sphere and its relation to the discussion of police adoption of AI technologies.

Development of the Public Sphere

Prior to the public sphere, public authority was embodied by nobility and monarchs, who presented themselves to the people through personal excellence and with "some sort of higher power" (*The Structural* 7). Habermas calls this "*representative* publicity" (8). However, a new

form of *critical* publicness began to form with the emergence of finance and trade, guided by a bourgeois elite (14). This early capitalism resulted in "the sphere of public authority" (18–21), with discussions taking place in political journals and other correspondence, where topics deemed important by the public were debated. In these discussions, authors used reason "to think their own thoughts," which were sometimes "directed against the authorities" (25). These discussions led to the formation of public opinion (26). In short, Habermas describes a sphere where individuals come together as equals to participate in critical and rational debates, providing a foundation for democratic societies "by articulating the needs of society with the state" (*The Structural* 176).

The public sphere is part of the lifeworld, the background of societal understandings, and cultural norms and traditions, and is separate from the state and administrative power. The concept of the public sphere aligns with the bourgeois approach to bind the state to "a system of norms legitimated by public opinion" (*The Structural* 82). Within the political realm, the public sphere served as a means of legitimizing law through the force of the better argument agreed upon by those in the public sphere. "Public debate was supposed to transform *voluntas* into a ratio that in the public competition of private arguments came into being as the consensus about what was practically necessary in the interests of all" (83). A key focus of the sphere was the need for equality among participants. However, only propertied private individuals were admitted to this critical debate (109). Access to this realm must be granted to all citizens and "a portion of the public sphere comes into being in every conversation in which private individuals assemble to form a public body" (Habermas *Encyclopedia* 49). I will discuss the limitations and critiques of Habermas's view on the inclusivity of the classic public sphere later in this chapter. Within this sphere, people participate as "equals in pursuit of truth and the common good"

(Finlayson 12). In his idealized view of the public sphere, Habermas identifies it as "the self-interpretation of the function of the bourgeois public sphere crystalized in the idea of public opinion" (*The Structural* 89).

Erosion of the Public Sphere

Habermas then discusses the erosion of the public sphere as it is colonized by multiple strategic-rational or instrumental interests, including economic, political, and mass media. These interests distort the public sphere and manipulate public opinion. Instead of debating culture, the focus became on consuming culture through mass media. The public sphere in the world of letters was "replaced by the pseudo-public or sham private world of culture consumption" (*The Structural* 160). Where at one time, the distribution of information connected the private and public spheres, the new press focused on publishers focused on the commercial aspect of the enterprise" (181–185). This eventually led the press to become a capitalist undertaking "enmeshed in a web of interests extraneous to business that sought to exercise influence upon it" (185). The "new media" of radio, television, and film also saw "economic concentration and technological-organizational" coordination (187). These media saw their original basis reversed. Once they were focused on the critical-rational debate, economic interests forced these media to return to the private sphere of commodity (188). "The more their effectiveness in terms of publicity increased, the more they became accessible to the pressure of certain private interests, whether individual or collective (188). The commercialization of the media, as well as economic and political influence, was joined by a rise in bureaucracy and the use of publicity by special interest associations and political parties (197). This "public relations" allows these groups to "manipulate 'public opinion' without themselves being controlled by it" (200). These factors have resulted in a weakened public sphere, compromised of its ability to work through

information, create critical-rational debate, and produce legitimate public opinion to guide decision-making in democratic societies. Habermas, as early as 1962, goes so far as to warn of a refeudalization of the public sphere, meaning a return of the "representative," uncritical publicity that characterized pre-capitalist social forms (*The Structural* 176, 197-199).

Now, I will focus on the work Habermas produced to update his thinking on this subject through *A New Structural Transformation of the Public Sphere and Deliberative Politics*. Habermas's new work seeks to explain his views on how the digitization of the media is transforming its structure as well as the political process (*New Structural* 3). The booklet provides an enlightening update on the current state of the public sphere, which is facing a crisis of legitimacy as it faces political influence, algorithmic control, private networks and ownership, and social media. Habermas paints a bleak view of the current state, highlighting the erosion and alteration of the public sphere through digitalization, globalization, monopolization, and commercialization.

Habermas believes democracies must enable a media structure that is inclusive and includes a "deliberative character" of will formation and public opinion (59). "The point of deliberative politics is, after all, that it enables us to improve our beliefs in political disputes and gets closer to correct solutions to problems" (17*).*

But that is the current state of the public sphere. The media system is crucial for the political public sphere, as it allows for generating competing public opinions and condensing them into "effective public opinions" (30–31). A professional staff trained in mass media, comprising writers and journalists, is necessary to mediate this discussion (31). The new media represents a "caesura in the historical development" of the media, which is comparable to the introduction of printing (33). The new media has disrupted the model of a media system

moderated by professionals. The owners of these networks, such as Meta, X, and others, neither produce nor edit nor select, but act "without responsibility" by allowing information to spread across networks guided by algorithms (37). The promise of an egalitarian approach to communication is "drowned out, at least in part, by the desolate cacophony in fragmented, self-enclosed echo chambers" (37). Social media promised emancipation by allowing everyone to be an author, but the skills necessary for "professional selection and discursive examination of contents" were not provided (*The Structural* 38–39). This concerns Habermas, as a critical component of the media system has been lost. I argue that this further disintegration and transformation of the sphere should be concerning for organizations such as the police, who rely on the sphere to understand how to best serve the public in line with its expectations. "It is harmful for a democratic system as a whole when the infrastructure of the public sphere is no longer able to direct the citizens' attention to the relevant issues that need to be decided or to ensure the formation of the competing public opinions—and that means, qualitatively filtered opinions" (*New Structural* 57). This situation has been accompanied by a rise of "fake news" and reduced trust in the state-financed media, with a perception that elites are corrupt or suspect (43–45). The triumph of the attention economy has had an impact on all media, with a focus on entertainment, personalization, and emotional responses to media information rather than rational debate (48–49).

Habermas fears that what has resulted is a fragmented and isolated public sphere, where communications for the "generalization of interests that includes all citizens" is no longer the goal, nor is it inclusive (54). "It is now commonplace to speak of disrupted public spheres that have become detached" from one another (57). He refers to it as a "plebiscitary public sphere," based on user input grounded in technical and economic infrastructure, where information is

shared that cannot be defined as either private or public, thereby provoking attention (54–55). As a result, the "political sphere of constitutional democracies loses the appearance of an inclusive space for a possible discursive clarification of competing claims to truth and a general equal consideration of interests" (55). This results in competing public spheres where "communicative contents could no longer be exchanged in the currency of criticizable validity claims" (56). In this situation, it is not the proliferation of fake news that is most concerning but our inability to identify fake news as such (57). This combination of media commodification, re-feudalization of the sphere by elites, technocratization of politics and decision-making, and digitization through advancements like social media, has eroded the sphere. According to Habermas, this has resulted in polarization, depoliticization and disengagement, and erosion of trust and legitimacy of institutions (45).

Connection to AI Adoption

Open deliberation in a democratic space is vital to ensuring legitimacy in institutions like the police that support democratic principles. The inability to develop a strong public sphere that reinvigorates portions of the system like the police is vital to Habermas. Without a robust public sphere, there is a risk that police may ignore or misunderstand public will and rely on technocratic decisions that drive instrumental decisions instead of developing decisions in consultation with the public. Habermas is also concerned about technocratic knowledge driving decisions rather than public reason (*Toward* 112). For Habermas, technocracy is a style of rule and decision-making that is unresponsive to the public sphere and can actively subvert public will and colonize it" (*Theory v2* 318). The organizational form of bureaucracy or technocracy

gains wide acceptance and considerable permanency on the strength of their greater effectiveness and superior level of integration (*Theory v2* 321).

Technocracy is a threat to democracy because it silences the public will and instead focuses on experts making decisions; Habermas discusses, for example, technocratic management in the European Union to substantiate this general claim (*The Lure*). In the world of AI, the current discussion is driven by technology companies and management consultants that are focused on sales strategies (81). Without developing mechanisms to strengthen democratic participation in these discussions, the police, and other governmental organizations, risk allowing unequal power dynamics and corporate interests to dominate the debate, focused on instrumental reasoning and accumulating profits through the sale of AI products to police departments (Bloch-Wehba 9-10).

A recent World Economic Forum report highlighted the current "black box" nature of some AI technologies, where the public, and sometimes the developers of the tools, do not know why they produce the results they do. "Modern AI's complexity makes it highly effective in low-risk scenarios, but almost impossible to trust in high-risk scenarios, such as health care, criminal justice, finance and more" (Zuccarelli 1). Again, the report reflects on the need "to foster public–private cooperation across different stakeholders to leverage a wide range of expertise" (Zuccarelli 1). A great deal of research focuses on the need for expert panels to observe these technologies, but Habermas can assist by connecting these technocrats with the lifeworld.

Some research focused on the health care field argues that policy issues are well-suited to public deliberation when they include conflicting public values, high controversy, combined expert and real-world knowledge, and low trust in government. In these circumstances, the deliberation process can help members of the public work through the complexity of the issues

and build trust in the policy-making process (Solomon and Abelson 1). Such research is promising in suggesting there is potential to use deliberative dialogue to understand the views and needs of the people the system is designed to serve.

Lifeworld

The lifeworld is an essential underlying concept in Habermas's work: "a storehouse of unquestioned cultural givens from which those participating in communication draw agreed upon patterns of interpretation for use in their interpretive efforts" (*Moral Consciousness* 135). Within the lifeworld, the actor stands "face to face" with issues that they must resolve intersubjectively. This shared lifeworld provides the cultural context in which all moral and practical discourse takes place, and allows for communicative action (*Moral* 135–136). Moral norms and ethical standards are rooted in the common experience of the lifeworld, providing information that members of the public can discuss. The shared lifeworld "offers a storehouse of unquestioned cultural givens from which those participating in communication draw agreed upon patterns of interpretation for use in their interpretive efforts" (*Moral Consciousness* 135).

This interpretation offers an "intuitively pre-understood context." However, participants must reach an understanding "about something in the world" if they want to keep their "action plans on a consensual basis with a jointly defined action situation" (136).

For Habermas, we gain our ability to be objective by recognizing that there is a "community of speaking and acting subjects" who can reach an understanding about what takes place in or is affected by it (*Theory* 1: 12–13). Through "communicative practice," we confirm our understanding of the intersubjectively shared lifeworld—the sum of interpretations and understandings shared by community members "as background knowledge." This concept of the lifeworld, an objective world with a shared understanding of truth, is taken from Husserl's study

and developed by others, including Heidegger. For Husserl, the lifeworld is a realm of "original self-evidences." It is the world of our "shared, self-evidently perceived" intersubjective experiences, such as seeing or hearing the same things, and other manners in which we build our shared understanding of life (37). Habermas argues that "subjects acting communicatively" come to an understanding in the lifeworld where rationalization occurs through "normatively ascribed agreement" as opposed to communicatively achieved understanding (70). This concept of the lifeworld is important for policing, as it is where people observe their lives and form opinions against a background of common sense.

Relation to AI

The promise of AI lies in the perceived new ability to use computer systems to conduct analyses and make decisions automatically using collected data (West). There is concern that this use of data could become the default way organizations using AI technology interpret the needs of the community, rather than connecting with the lifeworld. This could occur through various means, including the technocratic lure of instrumental use of the data or outright colonization of the lifeworld by corporate interests.

use of the data or outright colonization of the lifeworld by corporate interests.

The "autonomy and complexity" of these systems means that while people often do not know how they work, they still may trust their decision-making based on their past track record (König and Wenzelburger 134–135). There is also the concern that corporate interests may purposefully drive this discussion.

Paul Nemitz argues that the same tech companies moving forward with AI emerged out of a rejection of centralized laws. They consistently make "explicit or implicit claims that parliamentarians and governments do not understand the Internet and new technology such as

AI" (5). Their approach speaks to a suspicion or misunderstanding of liberal democracy. Nemitz argues that technologists actually misunderstand democracy, the rule of law, and the need to protect fundamental rights in a technological world where "these three pillars of constitutional democracy" are under threat. "On the contrary, the figures of argumentation presented by tech corporations and activists alike against new law over and over demonstrate that still today they put technology before and above democracy" (5).

System

To assist in navigating the lifeworld, humans develop organizations to act as "processes of rationalization" that help "steer" the social world. Weber assists us in understanding how this tendency toward "bureaucratization" ensures humans come under an autonomous social system (Habermas, *Theory* 2: 305). These systems act in their own self-interest or strategically, beginning to assume autonomy and become "peculiarly indifferent to culture, society and personality" (311).

This bureaucratization and system-building leads to an "irresistible tendency to an ever-expanding bureaucratization" which "colonizes" more and more of the lifeworld in representing power, money and administration-like bureaucracies (311). These systems, while in some cases helpful, can have the undesirable effect of taking over decision-making, stifling intersubjective dialogue. Habermas notes that Luhmann's systems theory assumes that the lifeworld has already been driven back by the system and only operates in "niche" areas. The most important of these systems for Habermas are the ones that include "rule-governed deployment of coercive power" or the administrative state (Russell 14). These systems coordinate action and hold society together, but they do so strategically and are not necessarily focused on understanding and consensus but rather the instrumentalization of communication to reach certain ends (Finlayson

54–55). The subsystems cause important symbolic structures of the lifeworld to be severed from cultural and political life. The subsystems are unable to connect action-coordinating mechanisms of mutual understanding (Habermas, *Theory* 2: 322).

Habermas defines three parts of the system: sociocultural, political, and economic. Police are contained within a subset of the political system as an institution and part of the distribution of legitimate power (*Legitimation* 6). These systems are prone to crises that must be resolved by the actors within them. Within the political system, there is a requirement for "an input of mass loyalty that is as diffuse as possible" (46).

As an institution and in practice, police are focused on community safety. They apply and investigate laws, address community issues, and investigate and prevent crime (Policing in Ontario). According to the model laid out by Habermas in *Between Facts and Norms*, a key source of the legitimacy of law in general is citizens' belief that they have imposed such laws upon themselves rationally and through democratic participation (*Between* 83-84). However, a key focus of police is the deterrence model of crime control, which secures instrumental compliance whereby the threat of police action and court sanction prevents people from taking certain actions. Deterrence models are based on the idea "that offenders and would be offenders are responsive primarily to the risk of punishment" (Jackson 3). Police in Canada are also independent agencies, governed by policing acts and varying types of oversight, none of which prescribe policing "operational" approaches such as focus, deployment, investigations, or other activities. Instead, these priorities are passed to the police through formal or informal networks of communication (McCandless and Vogler). Police then determine which operations require their attention and what actions should be taken. This has the effect of separating police as a system from the lifeworld and the common intersubjective understandings within it. For

Habermas, rationality is connected to how people "acquire and use knowledge" (*Theory v1* 8). In a rapidly changing world, with demographic and attitude shifts, police and the community risk being separated from one another and misinterpreting or misunderstanding norms and values. Should the background of the lifeworld change and shift, police could find themselves in a place where the lifeworld agrees with the validity claim that "you cannot trust the police."

Theory of Communicative Action

A key concept for Habermas is *communicative action*, an expanded concept of rationality based on dialogue and the ideal of a rational consensus. He writes that communicative action is a consensual form of social coordination within which people "mobilize the rationality potential" (*Theory v1* 99). In communicative action, intersubjective social interactions occur, within which language is used to reach mutual understanding, which plays the role of action coordination." Through this, people enter into action-oriented decisions toward reaching mutual understanding (*TJ* 110).

In communicative action, people come to an understanding through discussions of utterances. Through these discussions, they reaffirm their membership in social groups and their own identities (Habermas, *Theory* 2: 138). These statements also represent socialization and integration. "Communicative action takes place within a lifeworld that remains at the backs of participants in communication" (1: 335).

Habermas argues that we can reach rational understanding through communication and speech acts guided by rules of discourse aimed at reaching consensus and coordination (*Theory v1* 101). For Habermas, language is a "medium for a kind of reaching understanding." In reaching this understanding, participants raise "validity claims" that are accepted or contested by the listener (99). At the centre of this theory are the concepts of autonomy and solidarity (Rehg

137). Habermas recognizes the responsible nature of humans in positions where they participate in larger discourse on important topics. Through "intersubjective discussion and engagement" with others, we can reach a consensus where everyone's autonomy as a contributor is respected while a consensus is developed rationally (137).

Strategic Action

Strategic action differs from communicative action by instrumentalizing speech, using influence to achieve some goal or end (Russell 11). In strategic action, "one actor seeks to influence the behaviour of another by means of the threat of sanctions or the prospect of gratification in order to cause the interaction to continue as the first actor desires" (Habermas, *1990* 58). Habermas believes strategic action is oriented toward success rather than reaching understanding. In this scenario, the strategic actor will attempt to influence their opponent's definition of a situation, as well as their decisions and motives, by using threats and enticements (*1990* 133). The coordination of action and cooperation that follows depends on the interests of the various actors (*1990* 134).

Habermas focuses on universality, equal participation, and shared understanding and provides a framework to reach agreement on norms through democratic deliberation rather than a prescribed list of moral norms (*Moral* 57-58). "I call interactions communicative when the participants coordinate their plans of action consensually, with the agreement reached at any point being evaluated in terms of the intersubjective recognition of validity claims" (*Moral* 58).

Rules of Discourse

Habermas's discourse ethics is founded on his arguments that moral norms are derived through inclusive, non-coercive, and rational dialogue. Discourse ethics provides a "procedure" "for testing the validity of norms that are being proposed" for adoption (103). This approach

relies on communicative action, where rational arguments on issues of mutual concerns are achieved through "action oriented toward reaching an understanding" (102). This approach to the development of moral norms is meant to free humanity from the burdens of instrumentalism by adopting an emancipating approach to the development of reason. Habermas's principle of universality is focused on the individual's ability to "freely accept the consequences and side effects," with observing the norm vital to his thought (93). Only those norms can claim to be valid that meet (or could meet) with the approval of all affected in their capacity as participants in a practical discourse" (93). A presupposition for this discourse is an "ideal speech situation," a term Habermas had used but later retracted, which is "immune to repression and inequality" and follows a determined set of rules that ensure a fair process for communication (88).

If a person is seeking to use a validity claim to make themselves understood, they must follow certain rules. According to Habermas, they must "utter" something understandable, giving the hearer something to understand and making themselves understandable.

Within communication, Habermas argues for a principle of universalization where every person entering the discourse accepts the "universal and necessary presupposition of argumentative speech" (87). He adopts rules of discourse, which have no ethical content, as laid out by Robert Alexy. While these rules refer to ideal situations, they also highlight the key characteristics of empirical instances of communication that help reduce barriers and distortions created by power relationships and language:

1.1 No speaker may contradict himself.

1.2 Every speaker who applies predicate F to object A must be prepared to apply F to all other objects resembling A in all relevant aspects.

1.3 Different speakers may not use the same expression with different meanings.

2.1 Every speaker may assert only what he really believes.

2.2 A person who disputes a proposition or norm not under discussion must provide a reason for wanting to do so.

3.1 Every subject with the competence to speak and act is allowed to take part in the discourse.

3.2 a) Everyone is allowed to question any assertion whatsoever.

b) Everyone is allowed to introduce any assertion whatsoever to the discourse.

c) Everyone is allowed to express his attitudes, desires and needs.

3.3 No speaker may be prevented, by internal or extensive coercion, from exercising his rights as laid down in (1) and (2) above (Habermas, *Moral Consciousness* 87– 89).

For Habermas, these rules are "inescapable." They are transcendental rules of communication, such that contradicting any one of them is a "performative contradiction," defined as a situation where the content of a statement contradicts the "performative part of the speech act" (*Moral* 89) (Apel 265). As such, they are necessary for a "search for truth" (*Moral* 87), defining the participants and guaranteeing all participants an equal opportunity to contribute and access the discussion. These rules of discourse direct participants toward an "ideal speech situation" of rationally motivated consensus. They are necessary because in a modern society, there is no alternative to communication and discourse to resolving disputes (Finlayson 44).

As I noted at the beginning of this section, Habermas's theories recognize the autonomous nature of individuals and our need for collective action. It is "insight," developed through ethical discourse, that allows individuals to take action in line with universal agreements made collectively and based in rationality (*Moral Consciousness* 162). For Habermas, discourse theory is a dialogical approach designed to bring people together for meaningful discussion or argumentation based on the concept that democratic procedure provides a legitimizing force for

ideas in society. It only intends to determine how "norms of action" are impartially justified (*Between* 108–109). "Democratic procedure makes it possible for issues and contributions, information and reasons to float freely; it secures a discursive character for political will formation; and it thereby grounds the fallibilist assumption that results issuing from proper procedure are more or less reasonable" (*BFN* 448). This process gives us the "communicative freedom" to say yes or no. It also allows us to reason and question and accept ideas only when we believe the idea, statement, or offering is acceptable (Russell 25).

Validity Claims

According to Habermas, in daily life, humans interact with one another and make "validity claims" based on their understanding of the world around them (*Theory v1* 10, 39). There are three types of validity claims: truth, which relates to the objective world; rightness, which relates to the shared social world; and truthfulness, sometimes referred to as sincerity or authenticity, which is related to the speaker's own subjective world (*Moral Consciousness* 58). A simple statement like "the grocery store was busy" or "traffic was slow" is backed up by the knowledge of the speaker. When someone makes a validity claim, it falls to the hearer of the claim to determine whether to accept it (Russell 8). It is the hearer's acceptance that allows for some type of unified action to occur. When the hearer rejects the claim, it is because it does not conform to the hearer's "subjective experience," "interpersonal relations," or the "world of existing affairs" (*Moral* 137).

Adoption of AI

Acceptance and understanding are examples of communicative action, and this shared acceptance binds us as humans into a shared experience that Habermas calls the "lifeworld." The point of the argument or discourse is that we should try to resolve our differences through non-

manipulative and non-coercive strategies. It is the acceptance of this speech that allows us to coordinate our actions. When we don't accept a validity claim, we have several options, including withdrawing from the discussion, beginning a process of argumentation, or moving to strategic action (Russell 9).

Nancy Fraser's Critique

Habermas's work has faced criticism from a range of theorists. These criticisms have focused on an overreliance on rationality, a failure to recognize the exclusion of marginalized people, neglect of the working class, ignoring a plurality of identities, and his historical accuracy (*Habermas and The Public Sphere*). This criticism enriches our engagement and brings a new lens to his work. However, since *radical* critiques of his framework fall victim to the charge of performative contradiction (Habermas, *The Philosophical Discourse of Modernity*), I will take it for granted herein that Habermas's overall approach is compelling, and that friendly critiques are the most interesting and productive for my purposes.

In her work "Rethinking the Public Sphere: A Contribution to the Critique of Actually Existing Democracy," Nancy Fraser provides a unique critique and expansion of Habermas's theory on the public sphere that is worth mentioning here, as it adds to this thesis. Fraser notes that Habermas's theory is "indispensable" for social critiques and for "democratic political practice" (Fraser 57), but his approach and views on the public sphere are too idealized and do not accurately reflect current and existing democratic practices (61).

In particular, Habermas idealizes the liberal public sphere by failing to recognize that it was not accessible to all, excluding women and non-propertied men (Fraser 63). While he celebrated the inclusivity of the sphere, he failed to fully explain the power imbalances that would exclude people like women from discussions. Habermas was wrong in his assessment of a

single public sphere, Fraser argues. “Moreover, not only were there always a plurality of competing publics but the relations between bourgeois publics and other publics were always conflictual” (61).

Fraser does not argue that these “subaltern counterpublics” are virtuous or democratic—some are not—but they do function with a dual character of “withdrawal and regroupment” and “agitational activities directed toward wider publics” (68). They offer the ability to communicate in “one’s own voice” (69). They offer “emancipatory potential,” with the ability to offset the participatory privileges offered by other spheres (68). Fraser further argues that publics can be defined as strong or weak. Strong publics are those whose discourse encompasses both opinion formation and decision-making (75), while weak publics are those whose deliberative practices exist solely in opinion formation and not decision-making (74).

Habermas’s Response

In 1992, Habermas responded to the criticisms his book had produced over the three decades since its publishing, rethinking and clarifying his views. He refocused on the disintegration of the sphere as “it changed with the rise of the electronic mass media, the new relevance of advertising, the increasing fusion of entertainment and information, the greater centralization in all areas, the collapse of the liberal associational life, the collapse of surveyable public spheres on the community level, etc.” (“Further” 436).

Overall, Habermas reaffirmed his thinking to support the political public sphere as a foundational piece of a healthy democratic system but acknowledged the existence of subaltern publics, though he never used the term, and recognized the role of women and marginalized people (441). “Both women and the other groups were denied equal active participation in the formation of political opinion and will” (428). He argued that “it is wrong to speak of one single

public" and recognized that these spheres existed and were "the co-existence of competing public spheres that takes account of the dynamics of those processes of communication that are excluded from the dominant public sphere" (424–25).

Habermas agrees that upon further review, the exclusion results in the same subaltern publics that Fraser references … a "counterproject to the hierarchical world of domination" (427). Habermas accepted the criticism on the power imbalances and their impact on discussion. "Only a stereoscopic view of this sort reveals how a mechanism of exclusion that locks out and represses at the same time calls forth countereffects that cannot be neutralized" (427). These criticisms sharpened Habermas's theory and deepened our understanding of the public sphere.

Adoption of AI

Fraser's critique is helpful to the discussion on AI adoption and strengthens the use of Habermas. Her work reinforces the issue of power and the need for police to overcome power imbalances and reach out to audiences and publics that are traditionally excluded from larger societal debates. Fraser's approach helps us understand that various groups communicate in different ways, that existing power imbalances might have a suppressive effect on their communication, and that police approaches must respect, embrace, and correct for those differences to ensure critical-rational debate.

Conclusion

In this chapter, I have presented my understanding of Habermas's theories and related them to issues regarding the adoption of AI.

I have shown that Habermas's public sphere is one where equal, inclusive, and transparent dialogue can occur on topics concerning the public, including the adoption of AI. The public sphere is vital for will formation by the public, ensuring that public concerns are discussed

without the influence of state, corporate, or technocratic inputs. These discussions ensure the lifeworld is not further colonized.

Habermas's discourse ethics provides a framework for police to engage all stakeholders and individuals to justify societal norms related to AI through critical-rational dialogue. The concept of his discourse ethics is focused on success based on the universal acceptability of decisions. This approach helps address the issues of those with the most to lose through AI usage, such as marginalized people and racialized groups, by hearing and addressing the issues and concerns raised. This discourse can provide a legitimized voice to individuals often excluded from decision-making dialogue, with ethical discourse that is inclusive and fair. Habermas's views on deliberative democracy argue for consensus-building through public deliberation. This emancipatory process helps ensure decisions are made through democratic processes.

The rationality derived through *communicative action* is based on developing mutual understanding rather than defining success through strategic outcomes. The Habermasian approach avoids the potential or real situation of top-down police decision-making, communications, and consultation that seeks to inform the public about decisions rather than involving it in deliberation. By engaging the public sphere and individuals and groups through discourse, the police and the public can co-create a framework for AI adoption and governance. This approach ensured enhanced alignment with the public will and better reflects the demands of the public for greater police accountability.

Chapter Three: Fragmentation and the Canadian Public Sphere

In this section, I will discuss the role of the Canadian public sphere as a place for rational-critical debate and public will formation, the importance of the sphere in terms of the ability of institutions to comprehend and take action on issues and concerns raised by the public, and the challenge to that comprehension and action with the disruption and colonization of the public sphere in an age of social media, decreased trust in media, and media monopolization by market actors. My objective is to show that institutions that support democracy in a liberal democratic system, like the police, should recognize the impact of the disruption of the media system and develop strategies that allow for the public will to be co-formed and understood. These abilities are necessary for police to maintain public trust, especially in marginalized and diverse communities. I will detail Habermas's critique in this chapter, which is relevant to the Canadian public sphere and Canadian institutions and democracy.

Canadian Public Sphere

The Canadian public sphere is changing as it struggles with further fragmentation and colonization by instrumental or strategic reason, through market forces, conglomeration of media ownership, and digitization (Edge). Media ownership continues to be highly concentrated within a few large corporations, reducing the number of trained journalists and limiting voices and viewpoints (Mills). Market forces, driven by several multinational companies like Meta, are furthering the fragmentation and colonization of the public sphere by driving social media use that is often based on private forces, whether through the platform owner or the influencers, creating echo chambers where true discourse cannot or does not take place (Ottawa Declaration).

This work is driven by a focus on datafication and surveillance economy factors, which weaken the public sphere and make it more difficult for democratic institutions to respond to the

needs of citizens. Later in this chapter, I will discuss trust in Canadian society and media, media ownership, diversity in published opinions, the fragmentation of the sphere and alt spheres, and the representation of diverse and marginalized peoples to provide insight into the current state of the political public sphere. I will attempt to demonstrate that the Canadian political public sphere is showing the same issues and raising the same questions that Habermas and others recognize in the broader Western world.

Trust in Canadian Institutions and Media: A Link

According to analysis conducted by Statistics Canada, confidence in Canadian institutions and trust in media are interrelated. When it comes to institutions, confidence refers to the intuitional performance where "trust in the media is more related to expectations and perceptions of truthful and unbiased reporting" ("Confidence in Institutions").

When measuring institutional confidence, close to two-thirds of Canadians report high confidence in the police – a higher proportion than for any other institution. Canadian media ranked at 37%, behind other institutions such as schools and the health care system. In the study, 58% of those who reported high confidence in four or more institutions also reported a high level of trust in news or information from the media. In contrast, 15% of those who reported low confidence in institutions reported high trust in news or information from the media ("Confidence in Institutions").

Public Sphere and Habermas

Habermas's classical definition of the public sphere refers to a space where citizens and private individuals can come together to engage in critical-rational discussions about issues of common concern in a space free from state or market domination. It allows for an unfiltered discussion focused on autonomy, reason, and the force of a better argument, and enables the

formation of public will or opinion. The public sphere is critical to democracy as it holds the political system accountable for decisions or needs that must be addressed, provides the basis of political legitimacy to institutions, and makes possible citizens' civic participation. By allowing for this debate and discussion, the sphere helps mediate between the state and people by developing and expressing public will to enable actors in the system to understand and address or implement that will (*The Structural* 27–30).

Habermas acknowledges that this definition of the public sphere is somewhat idealized (*Further* 422) and, empirically, has been undermined due to issues related to societal inequalities, state interventions (welfare state), and private interests such as the market or media conglomeration (*Further* 428).

Habermas's original definition of the public sphere is dated, and recently, he took on the challenge of reviewing that work with an understanding of the current state of the sphere. In his *A New Structural Transformation of the Public Sphere and Deliberative Politics,* he addresses the impact of new media, including digital and social media, which he believes has had an impact similar to the printing press (*New Structural* 33). Habermas recognizes that the public sphere is fragile and expresses concerns about further disruptions that introduce even more private interests, which disrupt dialogue and further its colonization.

In his newer work, Habermas details the further fragmentation of the public sphere driven by societal and institutional distrust, hyper-commercialization of communications networks, and fragmentation fuelled by digital and social media. Habermas outlines how digitalization has transformed communication and, in turn, the public sphere, first by removing boundaries to authorship and publishing information and then through fragmenting the sphere (3). In the decades since his first work, the Internet has grown, television and radio continue, but print and

magazine readership and publishing have plummeted. The rise of new media is occurring in the shadow of the commercial exploitation of Internet communications that are currently virtually unregulated (3). “Democracies require discursive deliberations to replenish their legitimacy, which allow for decisions that are in principle agreeable to all parties affected, grounded in rationality, and “cognitively correct and viable” (13). These deliberations are the “central role of the public sphere” to allow for all relevant opinions to be heard and discussed in a public manner that enables will formation (13).

“One circumstance is especially relevant in our context: the overall influence that the will of the citizens, hence of the sovereign, acquires over the decision of the political system depends essentially on the enlightening quality of the contribution of the mass media to this formation of opinion” (15).

Media System

Habermas believes that the media system plays a critical role in the political public sphere in generating, refining, and circulating “competing public opinions that satisfy the standards of deliberative politics” (30). He argues that trained journalists, editors, publishing houses, and media organizations form the media system’s infrastructure and are required to define the “scope and the deliberative quality of the offerings” (31). This work traditionally provided a forum that systems like the government could use to understand will formation.

“Public opinions are only relevant if opinion makers from the ranks of politics, as well as the lobbyists and PR agencies of the functional subsystems of society and, finally, the various actors from civil society, are sufficiently responsive to discover the problems in need of regulation and then to ensure the correct input” (30).

However, the introduction of digital and social media, with its multiple offerings like X, Facebook, Instagram, and others, has created a "caesura" in the historical development of media and communication (31). This break was at first sold as a promise of emancipation, based on the idea that everyone would have the opportunity to participate in a worldwide dialogue. The nature of networked media radically altered the communication pattern, "empowering all potential users in principle to become independent and equally entitled authors" (36). Habermas believes there are several issues with the new media that should cause concern. First, the companies themselves "solidified in Silicon Valley," driven by profits as they manage platforms for distribution. They control but take no responsibility for their content. They manage algorithms and other technology that drive usage and revenue by promoting the most profitable content, regardless of its discursive value (38). "The platforms do not offer their emancipated users any substitute for the professional selection and discursive examination of contents based on generally accepted cognitive standards" (38–39). This emancipation of the author role has come with no training, and combined with the technology based on algorithms, has created "fragmentation in conjunction with a simultaneously unbounded public sphere" (39). This situation leads to a "dynamic that counteracts the integrating power of the communication context of the nationally centred public spheres" previously established through media like print, radio, and television.

Media Consumption Habits and Trust

Habermas is also concerned about civic participation rates in media and new media. Using EU media statistics, he notes that the proportions of the time devoted to different media have shifted, with newspapers being hit the hardest. Other media, like television, radio, and books have continued, but reading intensity and other factors have declined (41). Distrust in public media has also grown, with most citizens reporting having recognized fake or distorted

news. That distrust extends to state-funded media, which arguably has not seen the same level of market-driven cuts as privately owned media. “The growing doubts about the quality of state-financed media presumably go hand in hand with the increasingly widespread conviction that the political class is either unreliable or corrupt, or at any rate suspect” (45). The result of this situation is a large and diverse media that allows for a plurality of viewpoints but where individuals can “retreat into shielded echo chambers of the like-minded.” This polarization of individuals into “competing public spheres” has been recognized by many scholars and experts (45).

Habermas is troubled by trends in new media, recognized by scholars, that have blurred the line between private and public. “In certain subcultures, the political public sphere is no longer seen as a communication space for generalizing interests that includes all citizens” (54). Habermas refers to the work of Bennett and Pfetsch, who argue in their article “Rethinking Political Communication in a Time of Disrupted Public Spheres” that social divides and declining central parties fragment discourse and challenge the ability to form collective public opinion. While these platforms seemingly increase access to media, they simultaneously create echo chambers and isolated hubs (Bennett and Pfetsch 243).

Habermas believes a media structure founded in a deliberative character that ensures inclusiveness is critical for will formation and is a “constitutional imperative” (*New Structural* 59). For now, the turn toward social media provides the risk of “deception-prone, communicative content,” with some social media–focused politicians participating in a “plebiscitary public sphere” based on likes and dislikes (54–56). As an increasing number of people begin to adopt social media as their preferred method of receiving information, the risk increases that “communicative contents could no longer be exchanged in the currency of criticizable validity

claims" (56). It is damaging to democracy when the public sphere is unable to direct individuals' views to matters of common interest or concern (57).

Habermas offers a call to action to Western governments to explore ways to preserve the public sphere, including regulating media companies like Meta, focusing on digital literacy, regulating digital platforms, and encouraging civic engagement (57–58).

Using market data, political commentary, and academic sources, I will now provide a short analysis of Canadian politics to demonstrate that the same complex of factors Habermas identified in Western countries is also occurring in Canada. This section aims to highlight the challenges institutions like the police have in connecting with the public when the intermediary of the public sphere is disrupted. Concerns about the decline of the Canadian public sphere have prompted political debates for decades. In its 2009 *Final Report on the Canadian News Media*, the Standing Senate Committee on Transport and Communications spoke of some of the same challenges highlighted in Habermas's work.

> To make informed decisions, citizens need a wide range of news and information. They also need access to a broad and diverse array of opinions and analyses about matters of public interest. Journalists, as are the information media that transmit such information, are important providers. This is why the freedom of the press is widely recognized as a central pillar of any democracy (1).

In 2009, the disruption of the public sphere by digital media had only just begun, and already the Senate Committee had recognized that few online services provided the quantity and quality of original reporting offered by the traditional news media" (5). Even in 2009, the Senate Committee had recognized that the advent of a "news on demand culture" with online newspapers, advanced cell phone technology, 24-hour news channels, blogs with personal and

political opinions, and other advances had resulted in significant changes, noting that "consumer markets for all forms of news media have fragmented dramatically."

This work built on other government studies, like the Davey Report in 1970 and the Kent Commission in the early 1980s, which focused on a growing concern in government and the population "that concentration and conglomerate ownership had increased" (Standing Senate Committee 57) and was hurting trained journalists, local news, diverse opinions, and media literacy. The Senate Committee came to several conclusions, including that information and opinions are fundamental to democracies within modern societies. Everyone need not agree, but there should be mechanisms that allow for participation in "debates of the larger society," and newspapers and broadcasters were important to this ongoing debate. The report focused on the plurality of ownership. "The public interest in healthy and vibrant news media is as important as the public interest in the rights and freedoms of individual citizens."

Media Ownership

Media and Internet Concentration in Canada, a 1984–2022 study conducted by the Global Media and Internet Concentration (GMIC) Project at Carleton University, tracked key issues related to media and the Internet. According to the study, the top six Canadian media-owning companies—Bell, TELUS, Rogers, Shaw, Quebecor, and the CBC—accounted for 64.9% of all media revenue in Canada ($67.8 billion) in 2022, while the top ten "US- and foreign-based Internet and media giants—i.e. Google, Meta, Amazon, Netflix, Apple, Microsoft, Sony, Disney, Spotify, and TikTok in Canada accounted for 18.4% of the total" (Winseck 8).

The study found that Canadians still have media choice, with a plurality of online media including CBC, Bell, the *Toronto Star*, and new media in the form of products like the National Observer, Narcity, Canadaland, and Village Media. They also have access to MSN News, CNN,

CBS, NBC, *The New York Times*, BBC, and *The Guardian*. However, like Habermas, the study recognizes the impact felt by Canada's largest newspaper chains, including Postmedia, Torstar, and Quebecor, which continue to see plummeting revenue, with revenue share dropping by more than 30% between 2010 and 2022.

The study warns that debates over media ownership are highly politicized and sometimes "lacking evidence" to ground discussions, as massive corporate interests hire experts to argue for profitable positions (1). Still, media ownership matters, and the introduction of a complex web of Internet service providers, social platforms, legacy media, and handheld technology has created a range of online media services that are increasingly important to the economy. "We must seek to better grasp how gatekeeper power works at the communications network and digital platform levels to shape people's access to news and media content" (9–10).

Digital Media Use

According to polls and studies conducted by Stats Can and Statista, four out of five Canadians regularly use social and digital channels for information. Statista reports that digital media consumption doubled between 2019 and 2022, driven by Internet television usage (Guttman). Television viewing remains common, with radio continuing to be the second most popular medium.

A 2023 Environics study found that 88% of those polled had visited a social media site at least once a week, with the usage of major platforms like Facebook, X, YouTube, and TikTok showing statistically significant increases in 2024 (Denham).

Echo Chambers or Confrontation

Since Habermas's earliest description of the public sphere, the changes driven by digital media have led to a "dispersion and cacophony of public voices" and an "inability to communicate across differences" (Bennett and Pfischer 245).

It is worth briefly touching on the debate over the polarization of society and how digital platforms drive this discussion. In some academic circles, the "selective exposure hypothesis" is used to suggest that a "high choice" media environment allows individuals to avoid those who disagree with them and their worldviews and retreat into "echo chambers" (Tornberg 10). "The resulting lack of exposure to competing perspectives is, in turn, said to lead to more extreme issue positions as interacting with opposing viewpoints is seen as central for moderating opinions" (Tornberg 3). However, recent empirical research on this topic suggests that the digital sphere exposes people to others with diverse and opposing opinions, "often in contentious ways" (Tornberg 3).

> At the same time, such exposure does not seem to lead to political moderation. Digital media intensify polarization not as echo chambers but as a sorting machine, fueling a runaway social process that destabilizes plural societies by drawing more and more issues into a single expanding social and cultural divide. This suggests that the attempts of media platforms to reduce polarization by acting against echo chambers—algorithmically increasing exposure to opposing ideas—may backfire. (Tornberg 10)

As Habermas notes, we are still reckoning with the changes being caused by social media and working to understand how the technology changes communication and society (*New Structural* 33).

Aengus Bridgman, with the Centre for Media, Technology and Democracy, conflates these two thoughts by arguing that "it can be simultaneously true that Canadians are generally not living in information bubbles, being radicalized by social media, and being convinced by disinformation campaigns; AND a sizeable population of Canadians are" (Bridgman).

The Digital Media and Trained Journalists

While the media sphere has grown in the diversity of products and platforms, journalism and journalists, which Habermas calls the "infrastructure" of the media system, have gone through significant changes in Canada (Pereira 2533). Changes and financial reductions at major media organizations have resulted in fewer job openings and an increase in temporary contracts or freelance positions. Workloads have increased, leading to an accumulation of tasks, deteriorating working conditions, and a lack of training (2533).

In a 2017 study titled "The Shattered Mirror," the Public Policy Forum detailed a declining legacy media and the increased role of digital media. "The 20th-century news media are less and less prominent, except to provide grist for a public conversation they no longer control" (14). The report found that the reduction of Canadian media, the disappearance of newspapers, the increase in digital media, and the decreasing number of journalists resulted in a situation where sources of opinion were flourishing but sources of fact were shrinking (50). The situation was even worse in areas outside of major news centres, like the prairies, where beat reporters in places like legislatures have all but disappeared (51). The report found that digital news operations were not growing to fill the void caused by the disappearance of news agencies (62).

Datafication and the Public and Private

When digital media first arrived, their promise was focused on democratizing media and ensuring access for all people. Many observers, including Slakoe Splicahl, note that we are actually witnessing further colonization of the media system and the subordinating of "freedom of communication to capital" (Splichal 87). The "networked public sphere" was supposed to lead to a situation where public discourse, alongside traditional media, moved past the Habermasian model by systematically increasing communication at every level, including the political system, civil society, and the lifeworld (Splichal 89). He argues there "are many reasons why it is difficult to imagine that the design and development of new technologies would be a major factor in the emergence of a democratic communication system." A major reason is that the driving force of technological development continues to be capital, or those in control of the means of production (Splichal 111).

Conclusion

The purpose of this chapter was to show that the Canadian public is being affected and disrupted by the same issues facing other public spheres in the Western world, as Habermas has detailed in his recent works. These include the increasing dominance of capital, the disruption of digital media, reduced public trust in the media system, and fewer trained journalists. The next step in the argument will be to discuss community and expert concerns related to the implications of AI on policing.

Chapter Four: Fear the Machine? AI, Bias, Surveillance, and Community Concerns

"Technique tolerates no judgment from without and accepts no limitation" (Ellul 134).

It is important to briefly discuss some of the key criticisms of the use of AI in law enforcement to aid and inform the development of policy recommendations for Canadian policing related to the ethical adoption of AI. This chapter aims to define and outline the key issues that the policy suggestions in the next chapter are designed to address.

A review of current literature and media debate reveals several questions and concerns about AI adoption by the police that could impact public trust. These include the threat of technochauvinism or overreliance on technology, the opacity of AI systems, the potential for inherent bias in algorithms and data that could reinforce and worsen systemic issues, privacy concerns, and a lack of accountability measures that could lead to misuse or abuse. By addressing these concerns through a Habermasian lens, we can ensure these technologies are deployed in line with community expectations and for societal good that is recognized by the public.

Questions about the use of technology and its impact on society are not uncommon in philosophy and ethics, and the issues highlighted in this chapter are no exception. They revolve around the instrumental threat that AI poses to humanity and the need for careful consideration in its implementation. As French philosopher Jacques Ellul notes in *The Technological Society*, humans have long struggled with technology. Unchecked, technology acts autonomously in line with efficiency and independent of human morality or values. As it progresses, it has the ability to shape human institutions, behaviour, and humanity itself.

> A principal characteristic of technique ... is its refusal to tolerate moral judgments. It is absolutely independent of them and eliminates them from its domain. Technique never observes the distinction between moral and immoral use. It tends, on the contrary, to create a completely independent technical morality (Ellul 97).

Ethics and philosophy have an important role in this debate. AI is being called on to answer many of society's moral questions, informing decisions about crime, parole, and imprisonment, and discussing issues like who should get a loan or is eligible for welfare (O'Neil and Gunn 265). Current literature and discussion have posed the argument that AI could promote a "fairer criminal judicial system," with machines that could analyze and weigh relevant factors faster and better than humans based on its ability to ingest massive amounts of data. The hope is that AI "would therefore make decisions based on informed decisions devoid of any bias and subjectivity" (UNESCO).

Institutional adopters who make up the system, like police services, should understand that technologies like AI provide the ability to shape the future, both positively and negatively, solving problems but creating them at the same time with intelligible decisions that potentially create new concerns and risks for human rights and fairness (UNESCO). However, the warning from Paul Virilio remains: "When you invent the ship, you also invent the shipwreck; when you invent the plane you also invent the plane crash; and when you invent electricity, you invent electrocution. Every technology carries its own negativity, which is invented at the same time as technical progress" (89).

In the case of the police, that unintended consequence could be an impact on the service it provides society and the perception of police fairness and trust.

Technochauvinism

There is the potential for police services to over-rely on technology, leading to the devaluation of human reasoning, experience, and social context (Stanley). Meredith Broussard studies humanity's reliance on technology and the challenges it brings in its implementation. In her work, she details the ways fairness and bias found in society manifest in the technology and social systems humans produce. For her, "digital technology is wonderful and world-changing; it is also racist, sexist, and ableist" (*More* 4). New technology promising dramatic solutions for industries like policing is nothing new, she argues. As police review these technologies, they should understand the biases within technology and its potential to further harm marginalized people. There is a tendency to refer to situations where bias appears in digital technology as a "glitch" or something that can be solved. The issues found "show how long-standing social problems are reproduced and amplified inside algorithmic systems" (4).

In recent years, there has been a turn toward the "professionalization" of policing, often highlighted by the term intelligence-led and predictive policing. This idea "emphasizes the use of intelligence collection and data analysis to guide the selection and implementation of police policies (Sklansky 3). The strategies aim to steer the direction of policing by making the objective analysis of crime data and intelligence the central component of police thinking (3). "The guiding philosophy is bureaucratic and technocratic rather than collaborative and community-based" (4). There is also a strong tendency for the benefits of technology to be overemphasized, with limited success stories in policing (8).

This mode of thinking, as described by Michelle Broussard, can lead to technochauvinism, a kind of bias that considers computational solutions superior to other solutions. "Embedded in this bias is an a priori assumption that computers are better than

humans—which is actually a claim that the people who make and program computers are better than other humans" (*More* 2). Broussard notes that technochauvinism is often based on human factors and frailties such as "self-delusion, racism, bias, privilege, and greed." In many cases, it is the people selling the technology who celebrate its benefits, with arguments that the "algorithms are unbiased" or "computers make neutral decisions because their decisions are based on math" (2).

For some, the human element of policing can never be lost, even when AI guides police. "Discretion and caution are still required; a well-trained and reasonable human evaluator must always be the final decision-maker. Failing to adhere to this basic standard not just reflects poorly on the involved police department but also erodes community trust in police around the world" (Jenkins and Shields).

The Black Box and Transparency

In order to trust a system, people must understand it and how it works. Nevertheless, there is a growing realization of the opacity surrounding the inner workings of some AI systems. This lack of transparency, sometimes referred to as a "black box," can take many forms, including corporate or institutional concealment (including proprietary reasons), technical issues such as the code being difficult to read and only available to experts, and the opacity that arises from the "high-dimensionality characteristic of machine learning and the demands of human-scale reasoning" (Burrell 1–2). This last characteristic is particularly concerning, as "modern AI technologies have been based on neural networks that closely mimic human learning." Similar to human intelligence, it is difficult to fully understand how the system reaches conclusions, making it difficult to adjust or fix systems when their outputs are suboptimal (Rawashdeh).

"Machine optimizations based on training data do not naturally accord with human semantic explanations" (Burrell 10).

In his book *The Black Box Society*, Frank Pasquale discusses the increasing opacity between the public and the institutions that serve it, arguing that this issue matters because algorithms are increasingly expressing power and control in bureaucratic settings. The decisions that were once made only by humans poring over information or considering the circumstances are now made in an instant by a machine (8). As these decisions increasingly focus on more moral questions and enter areas of the system like banking, loans, the judicial system, and the police, the question must be asked: "Are these algorithmic applications fair?" (9).

While this does create concern, there are potential solutions.

Burrell argues that discussions between "legal scholars, social scientists, domain experts, along with computer scientists may chip away at these challenging questions of fairness in classification in light of the barrier of opacity. Additionally, user populations and the general public can give voice to exclusions and forms of experienced discrimination (algorithmic or otherwise) that the 'domain experts' may lack insight into" (10).

Nicholas Diakopoulos argues that full transparency of AI systems may be both impossible and unnecessary. "Pragmatically, transparency is merely about producing information that promotes the effective governance and accountability of a system … there is still plenty of information that can be disclosed about algorithms" (212). Diakopoulos notes that systems that make "high-stakes decisions" on issues like criminal justice where individual liberties are at stake should see a higher level of transparency despite potential costs (210). In these high-stakes contexts, he advocates for the design of transparency policies and approaches that are

understandable to stakeholders, with clear "context-specific ethical issues" and monitoring systems that would enumerate violations to share with the stakeholders (210).

Bias in AI

AI systems can inherit or reflect the biases present in the data they are trained on and use to perform tasks. If not addressed, this issue has the potential to create further systemic issues and even more inequity (National Association for the Advancement of Colored People [NAACP]).

This problem does not only apply to police data.

"Bias in, bias out," said one Yale University health researcher, who found that biases in medical AI compounds in the AI life cycle and can have negative consequences in clinical decision-making and exacerbate disparities (Yale School of Medicine; Cross et al. 1).

Technology companies and researchers have demonstrated that algorithms are more efficient at solving many problems. However, the question of bias that negatively impacts racialized, indigenous, disabled, and marginalized people, and how it runs counter to the public interest, is still a major concern for policy-makers and legislators. Every dataset is different but, generally speaking, data are collected in real time from users through interactions with government, social media, location information, websites, e-commerce, and cell phone data. These small pieces of data are gathered and compiled and eventually mined for government, commercial, or public use (Turner Lee 253). This is particularly concerning because at every stage of the criminal justice system, Black Canadians and indigenous people have been historically overrepresented (Foster and Jacobs 227). According to the Ontario Human Rights Commission, "disproportionate arrests and charges of Black people have been the subject of multiple reports and studies for over 30 years" (Ontario Human Rights Commission).

In its report, *To Surveil and Protect*, CitizenLab researchers note that populations that are disproportionately affected by the criminal justice system should be a key consideration during the adoption of these technologies, as the use of the technologies may enhance and further those disproportions" (Robertson et al. 8).

In typical AI applications, the systems are trained on information uploaded by the designers and intended users. Police data such as traffic stops, arrests, crime reports, gang member information, and convictions will have inherent issues regarding overrepresentation and other discriminatory information.

"However, the fact that all of the most common sources of police data are affected by problems that may distort the reliability of those sources should give rise to significant questions and apprehension about using big data-driven policing techniques such as algorithmic policing" (Robertson et al. 25).

Datasets can exclude populations, and when data is misapplied, online users can be tracked or profiled based on their online activities and behaviours. This could lead to key decisions about the background or future activity of individuals (Turner Lee 253). If misused, it can lead to the inequitable treatment of people based on their race, gender, or abilities (253).

There is already a recognition of the potential and real discriminative impacts of AI-based decision-making. Certain AI systems responsible for predicting the risk that offenders will reoffend have been found to predict higher rates of risk for Black people and less for white defendants. Google's Ads tool was found to show fewer ads for high-paying jobs to women compared to men (Ntoutsi et al. 2).

Data Quality

Another major issue is the quality of police data. Data quality is important for any AI system to ensure it provides the intended outcomes (Ataman). Ensuring high quality data, including accuracy and thoroughness, allows systems to make better predictions, avoid bias, and foster trust amongst users (Ataman).

A 2022 study that surveyed police crime analysts in Canada raised important questions about the current state and quality of police crime data. While the study's size was not generalizable to all police services, the analysts identified data quality as a major issue in policing, one that was not receiving enough attention.

"Policing is increasingly being shaped by data collection and analysis. However, we still know little about the quality of the data police services acquire and utilize" (O'Connor 637). Data and data collection have been an increasing portion of policing in Canada as police services seek to access new technologies like big data (638). This move toward evidence-based policing was made partly to justify the public expenditure on police budgets and partly as a turn toward a more data-driven approach. As a result, policing has become more reliant on information technologies such as GiS crime mapping and hiring analysts tasked with turning increasing amounts of data into information that can be used by officers to solve and prevent crime (639). "Good data"—that is, accurate and complete data—is vital to a system that relies on data to function (639).

The analysts reported several concerns "including 'poor report writing' (e.g., 'omissions in narrative and data fields'), 'incomplete reports,' 'missing information,' 'inconsistent information,' 'wrong codes,' and incorrect information (e.g., 'errors'). They also identified the need for better data quality oversight from police supervisors (647).

When asked direct questions in a survey, 77% of the sixty-seven respondents agreed with the statement that “data collected at my police service is inconsistent” and 26% agreed with the statement that “the records management system we use ensures data quality.”

However, decision-making that is based on faulty data or fails to recognize the limitations of collected data could result in the misallocation of valuable police resources, wasted public dollars, and even worse, including more crime and victimization. Interpreting accurate or inaccurate data incorrectly could lead to similar results. These potential issues impacting how data is utilized in police decision-making are best understood through a lens of data quality. (O’Connor 639).

Accountability Gaps

Police oversight, which has been increasing for several decades in Canada, is designed to ensure police are “subordinate to the rule of law” and to foster confidence in the criminal justice system (Puddister 391). A key factor in trust in the administration of justice is the presence of accountability frameworks for police misconduct and independent oversight of the police (Puddister and McNabb 832). The community must be given a role in this process to ensure that police are answerable to the communities they serve and serve them well (832).

Since the 1970s, there has been a movement in Canada to increase police oversight, often driven by the concerns of Black and indigenous community members and often after allegations of police wrongdoing and brutality (Kwon and Wortley 647). This has led to a series of efforts, each of them increasing civilian police oversight (648).

Policing is now a heavily regulated level of government that relies on legislation, governance, oversight, and policy to administer its duties in protecting communities. In the province of Ontario, police services boards, the Inspector General of Policing, and Chiefs of

Police are responsible for monitoring and ensuring compliance with legislation, training, standards, and regulations (Government of Ontario). Through regular reporting, data production, investigations, audits, and reviews, they are expected to oversee the equitable, effective, and adequate delivery of police service. Other actors and agencies like the Law Enforcement Complaints Agency, Ontario Special Investigations Unit, professional standards units, and the court system also provide oversight and review of police actions. The policing industry has been created through a series of public inquiries and discussions over time, where the public and legislators have sought more transparency and understanding of how police do their work (Government of Ontario). Police services have a strong focus on community, with whole sections and employees dedicated to relations with various groups to build an understanding of how police do their work (Inspectorate of Policing Ontario). This regulation took decades to implement after significant dialogue in the community.

AI Outpaces Regulation

AI's rapid development and subsequent adoption has challenged organizations across the world who are working to develop regulatory frameworks to govern its use (de Almeida et al). While the potential of AI seems immense, there is a recognition that governance techniques should be applied to develop these regulatory models for "social values legitimated via dialogue and scientific research" (de Almeida et al).

Still, AI is outpacing laws and ethical guidelines as it challenges social and regulatory standards.

"Despite a decade-plus exposure to the emergence and growth of big data, these regulatory frameworks are not prepared to respond to and oversee the data-intensive aspects of AI systems" (King and Meinhardt 29).

Even without those measures, governments continue to purchase these systems and deploy them, which "raises concerning questions about the potential for surveillance and the impact on civil liberties" (32).

In Canada, AI technologies developed by Clearview AI created a major public confidence issue when media learned that police were using the technologies without the necessary authority or oversight (Gillis). Officers across the country were found to be using the tool to identify suspects in cases, without the knowledge or approval of their organizations (Gillis).

"While police agencies like the RCMP and Toronto Police had been using facial recognition technologies before Clearview AI, the company's technology was found to be scraping images off of the internet and using them in its database." Of interest in the Clearview case is how a US-based company, scraping user images from US-based social media platforms, could operate for a considerable period in contravention of Canadian laws (as well as those from other non-US jurisdictions) in amassing its database of over three billion images" (Shepherd 180).

A federal Privacy Commission Report found that Clearview's business in Canada "represents the mass identification and surveillance of individuals by a private entity in the course of commercial activity" (Gillis). In response, the Toronto Police Services Board established its governance "for the use of new and/or enhanced Artificial Intelligence (AI) technology," including previously improved technology. The policy establishes an assessment and accountability framework for the use and purchasing of AI technology. The Toronto Police Service then adopted its own procedure to govern the use of the technology by police members including a "risk assessment, costs, and benefits related to the Service's use of AI technology" (TPS).

The Royal Canadian Mounted Police (RCMP) have also adopted a framework as part of its National Technology Onboarding Program. The *Transparency Blueprint: Snapshot of Operational Technologies* recognizes that the use of AI by law enforcement raises privacy concerns and issues related to potential bias. To ensure AI is used "legally, ethically and responsibly," the RCMP should consider issues like transparency, understandability, accountability measures and measures to prevent bias (RCMP 18–19).

The Clearview AI case is not the only time public confidence was shaken by police data collection.

In 2019, the Ontario Government asked Justice Michael Tulloch to review the street check system ("An Ontario Judge"). His review was focused on the use of street checks by Ontario police, a system that allowed officers to gather intelligence on individuals for future use. The street checks system, now referred to as regulated interactions, was originally intended as an investigative tool to capture information about people suspected of being involved in criminal activity. Instead, the system resulted in low-quality intelligence that alienated the community through random stops ("An Ontario Judge").

In his *Report of the Independent Street Checks Review*, Tulloch discussed the role of police in society and crime prevention. "Crime prevention is essential to the maintenance of public safety, and the police must have proper tools in order to undertake this work. However, the public's trust in police is the bedrock on which police legitimacy is built: without it, police lose authority and the ability to do their jobs" (6). He found that a major concern of people who participated in consultations was how police would use the information and whether it would lead to people being labelled as "usual suspects" or "known to police" (79). Tulloch noted that the relationship between the police and racialized and indigenous communities "is a complex

one" and that building "respectful relationships" takes time and commitment (93). Part of that relationship must include clear, articulable, and understandable rules on how information is gathered and used. Procedural justice is an important theme in policing (48).

To hold police accountable and assess the fairness their actions, decision-making processes must be known and scrutinized (Vestby and Vestby 45).

Conclusion

In this Chapter, I have outlined key questions that the police and the public are struggling with as the discussion on AI adoption in policing continues. It is clear that issues such as technochauvinism, AI opacity, data bias and quality, accountability, regulation, and the potential misuse or abuse of the technology are part of this discussion. The police and the community must work together to address these issues and determine whether and how AI should be used in policing. The rapid expansion of AI is expected to continue as industries, governments, and individuals seek to utilize these tools. This is expected to increase the demand for data from private companies (King and Meinhardt 17). Police sectors recognize AI as an important new technology that can enhance police service capabilities by "ingesting, analyzing, and visualizing data to provide insights into multiple options for resource deployment" (Jenkins and Shields). This is not a conversation the police can avoid, as the public will expect these tools to be used in some capacity (Jenkins and Shields).

Chapter Five: Communicative Rationality and a Fragmented Public Sphere: How Police Can Close the Gaps

Like any technology, AI is neither inherently positive nor negative to society but its potential impact in areas of high societal importance such as policing prompts questions about how it can be adopted in an ethical manner that builds and sustains trust and legitimacy (Megdad 24). This thesis provides normative recommendations on how Canadian police services and the communities they serve can develop a framework, founded in Habermasian discourse ethics, to consider and adopt AI technologies designed to ensure public safety and enhance crime prevention and investigation. The framework is based on Habermas's theories and critiques of the public sphere, political legitimation, communicative action, and deliberative democracy. It is designed to enhance transparency, understanding, trust, and legitimacy in AI adoption. The study is meant to assist police in avoiding a narrow focus on instrumental reasoning and administrative efficiency in favour of an approach that balances efficiency with alignment with public opinion and community values and expectations. This thesis argues that police can utilize Habermas's theories to develop structures for public engagement and deliberation founded in communicative rationality and in recognition of the challenges of a disintegrating public sphere. This approach has the potential to further democratize policing and build trust.

Chapter Overview

In this chapter, I will gather the key arguments of the thesis related to the challenges of AI, Habermas's theories of discourse ethics, communicative rationality, and the fragmented public sphere and provide practical approaches for police to adopt these technologies ethically. I will then propose a framework and approach for the police and the community, based on

Habermas's theories, designed to develop public engagement tools that address the challenges of a fragmented public sphere.

The Promise of AI

I have highlighted the discussion surrounding the promise of AI technology and its accompanying risks related to policing in the Canadian context. As these technologies advance, community, police, philosophers, and ethicists have focused on issues such as autonomy, fairness, accountability, and systemic bias. In this thesis, I have discussed the increasing use and discussion of AI technologies by police services in Canada and abroad. Police services are interested in the technology and its capabilities in various police functions, including crime investigation, prevention, and prediction, with "a view to increase efficiency and reduce resource demands on policing time and personnel" (Ezzeddine et al.). However, this interest in the advancing field of AI "also creates conflicts between their potential security benefits and concerns about their accuracy and fairness" including issues of privacy, discrimination, and the reinforcement of social inequities (Ezzeddine et al.). Glikson and Woolley note that research on people's trust in AI technologies is limited, but their review reveals the "important role of AI's tangibility, transparency, reliability, and immediacy behaviours in developing cognitive trust" (627).

For police to be successful, they must have active public cooperation, often attained by demonstrating "the fairness of the processes the police use when dealing with members of the public (Tyler and Fagan 267). While trust remains high in the overall community, police continue to face concerns from racialized and marginalized communities about their treatment and the fairness of their processes (Wortley and Owusu-Bempah 570).

Criticisms of AI Adoption

To understand how Habermas's theories can assist police and the community in mediating understanding, I have highlighted some key concerns raised in the public sphere related to AI adoption. These include:

Technochauvinism:

There is the potential for police services to over-rely on technology, leading to the devaluation of human reasoning, experience, and social context (Stanley). Meredith Broussard studies humanity's reliance on technology and the challenges it brings in its implementation. In her work, she details the ways fairness and bias found in society manifest in the technology and social systems humans produce. For her, "digital technology is wonderful and world-changing; it is also racist, sexist, and ableist" (Broussard 4). New technology promising dramatic solutions for industries like policing is nothing new, she argues. As police review these technologies, they should understand the biases within technology and its potential to further harm marginalized people. There is a tendency to refer to situations where bias appears in digital technology as a "glitch" or something that can be solved. The issues found "show how long-standing social problems are reproduced and amplified inside algorithmic systems" (4).

In recent years, there has been a turn toward the "professionalization" of policing, often highlighted by the term intelligence-led and predictive policing. This idea "emphasizes the use of intelligence collection and data analysis to guide the selection and implementation of police policies (Sklansky 3). The strategies aim to steer the direction of policing by making the objective analysis of crime data and intelligence the central component of police thinking (Sklansky 3). "The guiding philosophy is bureaucratic and technocratic rather than collaborative

and community-based" (Sklansky 4). There is also a strong tendency for the benefits of technology to be overemphasized, with limited success stories in policing (Sklansky 8).

This mode of thinking, as described by Michelle Broussard, can lead to technochauvinism, a kind of bias that considers computational solutions superior to other solutions. "Embedded in this bias is an a priori assumption that computers are better than humans—which is actually a claim that the people who make and program computers are better than other humans" (Broussard 2). Broussard notes that technochauvinism is often based on human factors and frailties such as "self-delusion, racism, bias, privilege, and greed." In many cases, it is the people selling the technology who celebrate its benefits, with arguments that the "algorithms are unbiased" or "computers make neutral decisions because their decisions are based on math" (2).

For some, the human element of policing can never be lost, even when AI guides police. "Discretion and caution are still required; a well-trained and reasonable human evaluator must always be the final decision-maker. Failing to adhere to this basic standard not just reflects poorly on the involved police department but also erodes community trust in police around the world" (Jenkins and Shields).

Transparency and Black Box:

The black box nature of AI, which can include corporate or institutional concealment (including proprietary reasons), technical issues such as the code being complex to read and only available to experts, and the opacity that arises from the "high-dimensionality characteristic of machine learning and the demands of human-scale reasoning" (Burrell 1–2).

Bias in Data:

AI systems can reflect or adopt the bias in the data they are trained on and use to perform tasks. If not addressed, this issue has the potential to exacerbate systemic issues and even create more inequity (NAACP). This is particularly concerning because Black Canadians and indigenous people have been historically overrepresented at every stage of the criminal justice system (Foster and Jacobs 227).

Over-Surveillance:

Datasets can exclude populations and misapply data. Individuals can be tracked or profiled based on their online activities and behaviours, leading to key data-based decisions about their backgrounds or future activity (Turner Lee 253). If misused, this can lead to inequitable treatment based on race, gender, or ability (253).

Data Quality:

Policing is increasingly being shaped by data collection and analysis, but we still know little about the quality of the data police services acquire and utilize" (O'Connor 637). Poor-quality data can lead to unintended results and consequences.

Accountability and Oversight:

To trust technology, it is important for people to see the results of its use. A key factor in trust in the administration of justice is the presence of accountability frameworks for police misconduct and independent oversight of the police (Puddister and McNabb 832). The community must be given a role in this process to ensure that police are answerable to the communities they serve and serve them well (832).

Habermas's Theories

"I do not see deliberative politics as a far-fetched ideal against which sordid reality must be measured, but as an existential precondition in pluralistic societies of any democracy worthy of the name" (Habermas, *New Structural* 10).

Habermas's theories are uniquely positioned to assist police services in addressing the challenges of building trust and legitimacy amid the continued disintegration of the public sphere. He offers a hopeful and normative analysis that can assist police services and the people they serve in developing deliberative approaches that can assist with the common identification and thematizing of problems, as well as possible solutions.

In a Habermasian sense, the discussion of AI in policing raises the spectre of further colonization of the lifeworld through a narrowly instrumental approach to data (Schuilenburg 9-11, Habermas, *Theory v2* 322). Police services risk focusing on serving society through the lens of AI, thereby datafying and skewing its view on the values, needs, and situations of the people they serve. This process of colonization, grounded in strategic action and instrumental rationality, could reduce individuals' ability to participate in genuine discourse and deliberation on the police service they want.

Applying Habermas

Habermas's work assists us in building a framework for adopting AI technology that engages the public and builds trust and legitimacy in policing and public support for the technologies being considered. Furthermore, Habermas's theories on the public sphere and his updated work provide valuable insights for institutions like the police, which draw upon the sphere to understand the will of the public and public opinion. A disrupted public sphere threatens police legitimacy and can skew police services' picture of public opinion. Police must

therefore ensure that the frameworks developed to understand public opinion on important topics like AI are representative, ethical, and deliberative to ensure that decisions made on its use contribute to legitimacy.

An Orientation to Understanding

An important precondition for this framework is the ability and willingness to be oriented toward understanding, participate in discussions, and listen directly to the community. Police community discussions can include various issues affecting police operations, usually the domain of police professionals. As Sklansky notes, there is a tension in policing between "professionalism" and community policing.

Professionalism was blamed for making police departments insular, arrogant, resistant to outside criticism and feckless in responding to social ferment. Community policing was very consciously a reaction against police professionalism: emphasizing the plurality of police functions instead of a single-minded focus on crime control, prioritizing community input and involvement over expertise and technical analysis, and favouring decentralized over centralized authority and locally tailored rather than globally rationalized solutions (Sklansky 5).

This would require opening up discussions on matters that affected the community and listen to the problems, concerns, and proposed solutions from the community and co-produce solutions." In a Habermasian sense, these questions should be debated in venues in which equal, well-informed actors forward truth claims and arrive at a common meaning. This dynamic, of course, assumes that those in power acknowledge that those not in power are well-informed actors with a say" (McCandless and Vogler 448).

According to data researcher Jeni Tennison, the benefit of including the people affected by technology is not just one of justice but also a practical strategy for including "stakeholders

who may be otherwise unconvinced or unconcerned," thereby reducing the risk of failures, building support for adaptive regulation and "building public understanding and sustainable adoption" (Tennison). Tennison argues that "we need to identify methods of involving multiple diverse stakeholders in these decisions that are practical, cost-effective and provide real power to everyone involved" (Tennison).

The Canadian Public Sphere

In the third chapter of this work, I demonstrated that the disintegration of the public sphere Habermas discussed in his *A New Structural Transformation of the Public Sphere* is evident in the Canadian political public sphere, with the growing use of social media and the blurring of boundaries between the public and private spheres. This should concern leaders of police services, as the sphere is crucial to its ability to hear and understand the community's needs. The Canadian public sphere is changing as it struggles with further colonization through market forces, conglomeration of media ownership, and digitization that drives colonization and fragmentation (Nieborg 3, 8-10). Media ownership continues to be highly concentrated within a few large corporations, reducing the number of trained journalists and limiting voices and viewpoints (Nieborg 3). Market forces, driven by several multinational companies like Meta, are furthering the fragmentation and colonization of the public sphere by driving social media use that is often based on private forces, whether through the platform owner or the influencers, creating echo chambers where true discourse cannot or does not take place.

A Habermasian Framework for AI Adoption

"Policing is an integral part of social integration and any model for adoption must start with an assumption of transparency and explanation in decision-making and not succumb to the

"temptation to outsource aspects of the decision-making process, and thus responsibility and accountability for the decision itself, to technological tools" (Moses and Chan 817).

This section highlights the framework, based on Habermas's theories, that Canadian police services could utilize to support the adoption of AI technologies. I drafted this approach to reflect the current environment in which police and the community find themselves, where questions of police authority, bias, and racism persist, and in keeping with the concerns raised by scholars, experts, and citizens regarding AI and its lack of transparency, black box nature, and potentially harmful impacts on society. The model is also meant to support police services and the community to ensure decision-making is transparent, reasoned, and inclusive.

With these overarching themes related to openness, inclusion, absence of coercion, and critical-rational debate, a focus must be placed on the functions of society that contribute to will formation. I propose a three-pronged approach for police and the community that focuses on:

- Ensuring the health of the public sphere and its actors
- Public engagement through civil society
- Ensuring widespread consultation and discussion with all of society, particularly with the marginalized and others who would be most affected
- Digesting and making operational decisions through a formalized police decision-making process
- Regular and joint reporting of results of AI technologies to ensure consistent feedback loops

The goal of this approach is to allow the public sphere to assist police in identifying problems, thematizing those problems, and proposing possible solutions.. The police, as an institution, would benefit from this discussion, especially as it faces critical questions such as:

- What input data are used?
- What dataset has been used to train the model?
- When and where were the data collected?
- Does input data capture features (directly or indirectly) that should not be relevant to the decision? For example, are any input features correlated with gender in such a way that model decisions are different if you are male or female?
- Is the data representative of the field that the model decisions affected? For example, has the model been tested in the setting where it is applied?
- What are the most obvious differences between the training setting and the current setting?
- Do we need to make any adjustments for particular groups or decisions? How are these data collected? For example, were they collected with the intention of being used for these kinds of decisions?
- Do we know of any selection biases (either by design or due to practical issues) in the data collection?
- Who collects the data?" (Vestby and Vestby 52).

In *Moral Conscious and Communicative Action,* Habermas lays out clear rules of discourse that should act as guiding principles. Applied to the context of community consultation over the use of AI in policing, they have obvious advantages in emphasizing the *rationality*, *representativeness*, and *unforced nature* of public discussion on the issue. The rules are:

1.1 No speaker may contradict himself.

If the police are to enter these discussions in the community, they must ensure consistency in their communication to foster trust and meet expectations. For example, if a

commitment is made to engage diverse communities, they must be engaged. Police should engage the community consistently and maintain dialogue and approaches to avoid contradictions regarding the purpose of AI tools and their intended contribution to public safety. They should begin the discussions with principles focused on openness, equity, and understanding and then follow those principles. For example, if AI is presented as a tool that maintains public safety in an equitable way, it is vital for them to explain how potential inequities are addressed. Additionally, a commitment to listen must be accompanied by a genuine effort to consider feedback, take meaningful action, and provide clear explanations.

1.2 Every speaker who applies predicate F to object A must be prepared to apply F to all other objects resembling A in all relevant aspects.

Assertions or decisions made in these processes must be applied consistently. If issues related to bias or privacy are raised about one technology, they should be applied to others. If AI is deemed overly intrusive in one area of the city, that should apply to other portions of the city as well. Safeguards should be applied consistently across various societal groups.

1.2 Different speakers may not use the same expression with different meanings.

AI technology can include multiple terms proposed by police, IT companies, and others. There should be a clear and meaningful lexicon used by experts and the community in the same manner to ensure understanding. Terms like equity, bias-neutral, data, and public safety should have general meanings.

2.1 Every speaker may assert only what he really believes.

It is vital that police approach the discussion with beliefs that are consistent, clear, and provable. For example, if police believe a tool can improve crime prevention or crime solvency in a bias-neutral or impartial manner, they should offer proof. If proof does not exist, the

statement should not be made. The benefits and risks should be presented to the community so everyone, including marginalized communities, can understand what is at stake and how those risks can be addressed or mitigated rather than ignored. Claims about AI's effectiveness in solving the specific problems it is designed for should be explained.

2.2 A person who disputes a proposition or norm not under discussion must provide a reason for wanting to do so.

Issues raised by the community, such as the potential harms of AI, should be addressed without offering assurances or promises. Claims of efficacy or harm should be discussed using evidence and proof with both qualitative and quantitative data.

3.1 Every subject with the competence to speak and act is allowed to take part in a discourse.

A wide swath of the community should be engaged to ensure the participation of civil society and individuals who represent indigenous, racialized, and lower-income people. This includes addressing barriers to inclusion by meeting the community where it lives, providing opportunities for online discussion, town halls, multiple language venues, and other approaches to ensure the broadest interpretation of this point.

3.2a Everyone is allowed to question any assertion whatever.

The power dynamics involved when police participate in discussion with the community, some of whom could be considered marginalized, must also be acknowledged and addressed. It is important to consider approaches that allow those voices to be heard, including independent AI assessments that are available to the community, financial support for community-led research, or the creation of community boards or panels with the authority to represent community voices.

3.2b Everyone is allowed to introduce any assertion whatever into the discourse.

Discussions must value the diverse views and opinions in the community and allow those views to be reflected in the discussion. A strategic communication approach would be for the police to design and present the discussion agenda. However, a more communicative approach would see diverse and marginalized communities raising specific issues. Official agendas focused on an instrumental approach to a decision should be avoided. Community-driven views, lived experiences, and the difficult relationship some people and communities have with the police should be discussed and reflected.

3.2c Everyone is allowed to express his attitudes, desires, and needs.

Any discussion on technology will require experts to explain how it works and what it does. However, these technical aspects should not guide the conversation. Instead, it should be focused on addressing the demands and views of the community to help understand the background of the lifeworld and the public sphere**.**

3.3 No speaker may be prevented, by internal or external coercion, from exercising his rights as laid down in (3.1) and (3.2).

Power dynamics should be addressed to ensure neither police nor technology companies influence the discussions. The community should be an equal partner in the discussion, and venues and approaches should be selected to ensure people can participate without these efforts being labelled as “anti-police” (Habermas, *Moral* 87–89).

The Public Sphere

“The democratic quality of policing is among its important moral dimensions” (Vestby and Vestby 55).

In line with Habermas’s approach, I propose an approach that focuses first on the public sphere, with concerted efforts, based on transparency, to highlight the question of AI adoption in

the political public sphere. This should include traditional media, social media, and other more traditional approaches, such as community gatherings and town halls. Police should clearly communicate their intention to establish a process for the adoption of AI software and tools. They should seek out public forums where issues related to AI adoption can be discussed, including universities, community meetings, and organization meetings. Guided by the discourse principles, police should explain the rationale, purposes, and goals of the technology.

The goal of this approach should be to allow the public sphere to assist police in identifying problems, thematizing those problems, and proposing possible solutions (BFN Page). The police, as an institution, would benefit from this discussion, especially as it faces the critical questions previously described.

Experts and community members should review the information being shared to ensure it is factual, valid, and understandable. The discussion should also be undertaken neutrally to avoid strategic action.

The public sphere, as defined by Habermas, provides a conceptual area where inclusive, open, and ongoing dialogue can take place among members of the public. Within this sphere, the police and society benefit when AI technologies are transparently discussed, and the public will on how to proceed is formed through rational-critical debate. This debate and will formation are vital to circumventing corporate and technocratic interests and decision-making and will build trust in AI adoption by police.

This approach will also help prevent the further colonization of the lifeworld by bureaucracy and technocracy. It allows the lived experience, values, and societal background represented in the lifeworld to influence the decisions that affect it.

Civil Society

Civil society should be a key focus, as the police formally and transparently engage with human rights, advocacy, and community groups and formalize the feedback received.

"Civil society is uniquely positioned to detect problems with technology early," writes Tennison. Civil society is often the first to learn and understand emerging patterns affecting the people they represent and are "more directly exposed to the here-and-now effects of data than governments or companies and can bring this experience to the discussion" (Tennison).

Informed decision-making should be a key focus, and police should bring in experts representing different groups to help people understand what is being proposed.

The communication style should include mechanisms such as surveys, conversations, and other tools but should also be deliberative and focus on public forums and debate to ensure interactivity. Working with the community, the police should gather and interpret the information received, validating it before reporting back. An ongoing process should also be developed to ensure the public sphere and residents are provided with information about what was done and the impact of the work. During this phase, public understanding and awareness can be enhanced to allow for issue identification and solution-oriented proposals. This phase would rely on police and expert technical briefings, utilizing media awareness and civil society to enhance understanding. All discussions should be framed communicatively to discuss validity claims rather than strategic action.

Critical-Rational Debate

This would then lead to public discussions and critical-rational debate supported through public forums and assemblies that allow for pluralistic viewpoints and diverse opinions. Ongoing

engagement and mediated discussions with civil society and experts and non-experts from related fields, including human rights and AI, would enhance this understanding.

The police should work with the community to ensure the broadest swath possible of citizens and residents are included in the discussion, whether through direct discussion, dialogue in the public sphere, or through advocates, community groups, and civil society.

A key concern raised by Habermas's critics (and acknowledged by Habermas) is the ability of marginalized or less enfranchised people to access the public sphere and participate in deliberative discussions. This would include an intentional focus to ensure members of racialized, indigenous, and marginalized communities are included.

"Any discussion of police–community relations needs to consider how distrust is often a manifestation of years of disproportions in stop, search, and arrest rates" (McCandless and Vogler 444). The dialogue should include police, community, and independent experts who can interact with non-experts focused on issues such as "(1) the type of data we use to learn; (2) the learning goal we set; and (3) how later actions affect subsequent training data" (Vestby and Vestby 49).

While experts would be crucial to this discussion, it is also important to ensure the information is understandable to all participants. While the information contains technical aspects, the key ethical issues and concerns can likely be communicated to residents: "datasets are created in political contexts, the choice of what data to collect, the level of accuracy, the level of consistency across time, and the choices of variables, all have resulted in inherent capabilities and limitations of the uses of the dataset" (Sengupta 2024).

In their article "Algorithmic Prediction in Policing: Assumptions, Evaluation, and Accountability," Moses and Chan draw attention to critical issues in big data–style policing,

including data quality, the importance of independent valuations, and design results that are not solely focused on outcomes. They also discuss the issue of transparency, which is a continuing theme in this work.

Habermas's discourse ethics provides a framework for police to engage all stakeholders and individuals in justifying societal norms related to AI through critical-rational dialogue. The concept of his discourse ethics is focused on success based on the universal acceptability of decisions. This approach helps address the issues facing those with the most to lose through AI usage, such as marginalized people and racialized groups, by hearing and addressing the issues and concerns raised. This discourse can provide a legitimized voice to individuals often excluded from the decision-making dialogue, with ethical discourse that is inclusive and fair. Habermas's views on deliberative democracy argue for consensus-building through public deliberation. This emancipatory process will help ensure decisions are made through democratic processes.

Public Discourse and Institutional Deliberations

After gathering this information and public input, the police should share the main public concerns, areas of consensus, and areas of key dispute with the community. By conducting these discussions in public, the police can ensure their decisions are oriented toward communicative action rather than bureaucratic or strategic action.

In a study of eighty-four Canadian governmental frameworks for AI governance, Attard-Frost et al. found that providing the public with "greater access to information about AI governance initiatives and their outcomes, policymakers and public servants can reduce a significant barrier to public trust" (10). They noted that if the public cannot see its direct contribution in final decisions related to governance, the "outcomes may be viewed as tokenizing, performative, or self-legitimizing, and ultimately jeopardize public trust rather than

secure it" (9). To address this, policy-makers and public servants should apply best practices and principles for public participation in AI design and governance and policy-makers should "specify precise strategic goals, success measures, and performance targets for their initiatives, as well as routinely publish information on their degree of success in achieving their performance targets and strategic goals" (10).

The proposal laid out in this thesis presents an approach to inclusive, open, and reasoned dialogue, offering a justifiable, rational, and overarching framework for the police. This approach should be guided by the need for rational-critical debate, power equalization, freedom from influence, autonomy in will formation, and the principles of universalization and discourse. It is vital to this discussion that participants reach agreement through practical discourse, with decisions that are or could be universally understood and agreed to by all parties (Habermas, *Moral* 65–66). Police should work to create forums that are inclusive, diverse, and representative of the populations most affected. Any discussion on police operations and approaches is potentially fraught with issues of technocratic knowledge as well as power imbalances. The police could work with third parties to develop these approaches to ensure compliance with the rules of discourse.

Reporting Back

The final step in this process is for police to report back on the tools being used, including costs, issues, successes, and failures. Police must take accountability for the processes they present, legitimizing those processes through justified decision-making and providing evidence of their effectiveness.

AI is a tool police can use to serve the community. The challenge after selection or implementation is the temptation to outsource aspects of the decision-making process, and thus

responsibility and accountability, to technological tools. However, as we have seen, AI tools can introduce assumptions that decision-makers may be unaware of. Ongoing monitoring and reporting is therefore necessary.

Police should report these decisions both through formal oversight processes and in the public sphere where relationships have been established. The reports should detail the use of AI tools, how effective they were in achieving the goals set out by the community, and whether adjustments to policy or decisions are required.

The police must be able to demonstrate how the discourse in the public sphere, along with its deliberative processes, have led to decision-making that the public can recognize is in line with its will. This dialogue should be formed as a feedback loop and discussed in the public sphere, with civil society and the public.

Conclusion

In this chapter, I have outlined a framework for dialogue that police can follow to implement AI. The instrumental reasons for the utilization of AI are apparent to most observers. Habermas's model allows the police and the community to work together to explore the issues, concerns, and benefits of AI in a way that supports communicative rationality. This ensures that AI, rather than the community, serves as a means to enhance public safety.

Chapter Six: Conclusion: Legitimacy, Discourse, and AI in Public Safety

> "Technocratic consciousness reflects not the sundering of an ethical situation but the repression of 'ethics' as such as a category of life" (Habermas, *Knowledge* 112).

AI continues to alter the landscape of modern policing, with new offerings and tools that propose more efficient and effective approaches to public safety. From a technical perspective, the future appears bright. Facial recognition, real-time crime analysis, predictive policing, and other tools promise to improve policing and police services' ability to solve and prevent crime while limiting the need for significant budget increases (Epstein and Emerson). However, as we saw in this thesis, while police "can perform their traditional functions not just on a faster and larger scale" with AI, how that instrumental progress is achieved raises concerns about bias, privacy, and fairness (Joh 263). This has prompted a discussion about rules governing the use of AI in critical industries and services that make up the system of money and power, such as banking, government, and civil administration (263–64).

Contribution

This discussion began with the questions of:

- How Canadian police can adopt AI technologies in a manner that is ethical and meets public expectations and demands in a time when the public sphere is fragmented and disrupted?
- How can Habermas's theories assist police with the ethical adoption of AI while building public trust and legitimacy?

- What deliberative systems can be implemented to approximate an ideal speech situation and enhance community understanding of AI?
- How can discourse ethics provide a community deliberation model that assists the police and the people they serve understand and mitigate issues related to AI adoption ?
- How can communicative rationality through discourse address inherent issues with instrumental reason in AI?

The aim was to show how technology could be adopted without falling prey to a "technocratic consciousness" that depoliticizes decision-making and removes them from the public sphere (Habermas, *Toward* 112). The normative recommendations contained in this thesis are designed to help police services and the communities they serve navigate a world with a public sphere that is fragmented, challenged, and disrupted by digital media, while ethically approaching the adoption of AI technologies in a manner that builds trust and legitimacy and adheres to public opinion or will.

Habermasian Compass

My study's theoretical basis was built on Habermas's theories. Habermas's scholarship has illuminated and informed our understanding of how humans build their understanding of the world and form public opinion (Chambers, "Navigating" 300). He provides valuable insights into how humans can find emancipation through normative and rational decisions, emphasizing the processes by which citizens demand that states act on their behalf, in line with their views and needs, and against "all forms of institutional repression" (Rasmussen 7–9). The promise and risks of AI technology are central to an ongoing debate among the community, police, philosophers, and ethicists, focusing on issues such as agency, privacy, fairness, accountability,

and systemic bias. These concerns are critical for police services and the communities that rely on them, as they directly impact trust in the police.

As stated in the introduction, the goal of this work was not to evaluate the ethics or effectiveness of AI but rather to explore how police can approach the adoption of AI ethically through discourse ethics.

In Chapter Two, I discussed Habermas's scholarship in relation to the public sphere and its fragmentation, discourse ethics, his views on the public sphere and its fragmentation, and the connections he draws between deliberative democracy and communicative rationality. I argue that this work provides a viable theoretical framework for Canadian police to adopt in critical discussions, such as the use and adoption of AI in policing technologies. I argued that by adopting these frameworks and developing practical systems that ensure citizen participation in policy-setting and decision-making, the police will be better equipped to make decisions that align with, or are at least demonstrably sensitive to, community expectations, while simultaneously advancing the interests of public safety. By engaging in these dialogues, police can also revitalize the public sphere's engagement with the issue of AI technology application.

In Chapter Three, I discussed Habermas's updated work, a *New Structural Transformation of the Public Sphere,* where he discusses the topics of polarization, digital fragmentation, and its impact on dialogue in the sphere. I related that work to the Canadian public sphere, which has experienced media consolidation, fewer trained journalists, an increase in social and digital media, and a reduction in print and long-form media. I also discussed trust in Canadian society and media, media ownership, diversity in published opinions, fragmentation of the sphere and alt spheres, and the representation of diverse and marginalized peoples to provide insight into the current state of the political public sphere.

In Chapter Four, I discussed the serious concerns being raised internationally and in Canada on critical ethical issues related to the adoption of AI in policing. These include technochauvinism and the devaluation of human reasoning, transparency and the black box nature of the technology, the issues of data quality and inherent bias in AI, the unpredictability of the technology, and the lack of accountability processes when errors are made. In a system where racialized and marginalized people are overrepresented in the criminal justice system (Kwon and Wortley 647), addressing these concerns, often by those potentially most affected by the technology, is crucial to maintaining public trust. This includes accountability frameworks and transparency in reporting back to the community on the positive and negative impacts of these technologies. A key factor in trust in the administration of justice is the presence of accountability frameworks for police misconduct and independent oversight of the police (Puddister and McNabb 832).

In Chapter Five, I outlined a framework based on Habermas's theories for the ethical adoption of this technology that engages the public, builds trust and legitimacy in policing, and garners public support for the technologies being adopted. The framework included using Habermas's rules of discourse to help the police address power imbalances and ensure inclusion in deliberations. Guided by the rules of discourse, the framework's goal was to engage all potential parties in the discussion – including experts, the community, civil society representing affected groups, the police, and the media – in critical discussions on AI. This would inform and understand will formation related to the use of AI. Furthermore, it would begin to create an ongoing discussion with police on the efficacy of AI and its potential positive and negative impacts by incorporating reporting mechanisms to inform the public sphere.

Habermas's theories on the public sphere and his updated work offer valuable insights for institutions like the police, which draw on the public sphere to understand the will of the public and public opinion. A disrupted public sphere threatens police legitimacy. Therefore, the police must ensure that the frameworks developed to understand public opinion on important topics like AI are ethical and deliberative, so that decisions made on its use reflect and contribute to legitimacy. Habermas's discourse ethics provides a framework for police to engage all stakeholders and individuals in justifying societal norms related to AI through critical-rational dialogue. The concept of his discourse ethics is focused on success based on the universal acceptability of decisions. This approach helps address the issues of those with the most to lose through AI usage, such as marginalized people and racialized groups, by hearing and addressing the issues and concerns raised.

Glossary

Artificial intelligence: Artificial intelligence (AI) "is considered the capability of a computer to do things that are normally associated with human cognition, such as reasoning, learning, and self-improvement" (Government of Canada).

Communicative Rationality: A form of rationality derived through mutual understanding through linguistic communication and oriented toward reaching consensus and action coordination rather than achieving strategic outcomes (*Truth and Justification* Habermas 110).

Media Fragmentation: The diversification and proliferation of media channels and platforms, leading to segmented audiences and fragmented public discourse. Scholars often link this phenomenon to the decline of a unified public sphere and the rise of multiple, niche publics (Strani, 245, 258).

Media Subsystem (within Habermas's System–Lifeworld Theory): According to Habermas, the media system is an essential part of the public sphere that is made up of mass media companies, journalists, broadcasters, opinion makers and other actors who edit news and develop opinions for mass consumption bridging public sphere with the political sphere (Political Communication 417). Ideally, the media system should mobilize and pool relevant issues and required information" and allow for information to be discussed in and "generate rationally motivated yes and no attitudes that are expected to determine the outcome of procedurally correct decisions." (Political Communication 417)

.

New media: New media is generally considered any media, from journal articles to podcasts to magazines, that are delivered to end users through digital technology. This could also include websites, email and other internet related forms of communication. (*What is new media?* SMHU)

Public Sphere: *In Habermas' original conception,* the public sphere is a space in social life where private individuals come together to take part in rational-critical debate to engage in open, rational-critical debate, oriented toward the public interest and capable of influencing political action (*The Structural* 27-30). Originally coined by Habermas to describe "private people gathered together as a public" engaged in debate (Soules). It is shaped by communication and includes mass media, scholarly discourse, social media, and public policy channels.

Social media: Defined as internet based mass communication, including social networking and blogging, where users share information, images, opinion and other content in networks and in order to build audiences (*Social Media* Britannica). A subset of new media focused on platforms enabling user-generated content, social interaction, and networked communication—typical examples include Facebook, Twitter, and Instagram. Again, while not directly defined by Habermas, you can draw on sociological or media studies literature for academically established definitions.

System (Habermas's Theory of Communicative Action): In Habermas's theory, the *system* refers to the institutional organization of domains of society such as the administration, military and judiciary that is organized through instrumental and strategic forms of coordination such as money and power The system and its subsystems operates according to its own functional and instrumental logic. This is often sperate from and in tension with the lifeworld. (*Theory 2* 171-172)

Technochauvinism: Technochauvinism is the assumption that computers are superior to people, or that a technological solution is superior to any other (Broussard, "Letting Go"). It is the belief that technological solutions are inherently superior to human or cultural approaches—that "tech

is always the solution." Meredith Broussard defines it as the conviction that "technology is always the highest and best solution" (Thierer).

Works Cited

"An Ontario Judge Says Carding Doesn't Work. But Will Politicians Listen?" *CBC News*, 21 Dec. 2018, www.cbc.ca/news/canada/toronto/ontario-carding-review-michael-tulloch-1.4964768. Accessed 1 Mar. 2025.

Ataman, Altay. "Data Quality in AI: Challenges, Importance & Best Practices." *AI Multiple: Research*, 9 July 2025, research.aimultiple.com/data-quality-ai/. Accessed 12 Mar. 2025.

Attard-Frost, Blair, et al. "The Governance of Artificial Intelligence in Canada: Findings and Opportunities from a Review of 84 AI Governance Initiatives." *Government Information Quarterly*, vol. 41, no. 2, 2024, Article 101929. https://doi.org/10.1016/j.giq.2024.101929.

Bennett, W. Lance, and Barbara Pfetsch. "Rethinking Political Communication in a Time of Disrupted Public Spheres." *Journal of Communication*, vol. 68, no. 2, 2018, pp. 243–53, https://doi.org/10.1093/joc/jqx017.

Birhane, Abeba, et al. "Power to the People? Opportunities and Challenges for Participatory AI." *EAAMO '22: Proceedings of the 2nd ACM Conference on Equity and Access in Algorithms, Mechanisms, and Optimization*, Article no. 6, 17 Oct. 2022, pp. 1–8. ACM Digital Library, https://doi.org/10.1145/3551624.3555290.

Bridgman, Aengus. "The Role of Social Media in Polarizing Canadians." *Centre for Media, Technology and Democracy*, 29 May 2024, www.mediatechdemocracy.com/all-work/the-role-of-social-media-in-polarizing-canadians. Accessed 18 Aug. 2025.

Broussard, Meredith. "Letting Go of Technochauvinism." *Public Books*, 17 June 2019, www.publicbooks.org/letting-go-of-technochauvinism/. Accessed 18 Aug. 2025.

---. *More than a Glitch: Confronting Race, Gender, and Ability Bias in Tech*. MIT Press, 2023.

Burrell, Jenna. “How the Machine ‘Thinks’: Understanding Opacity in Machine Learning Algorithms.” *Big Data & Society*, vol. 3, no. 1, 2016. https://doi.org/10.1177/2053951715622512.

Canadian Association of Chiefs of Police. *Trust Matters: A Summary Report on the Proceedings and Outcomes of the 10th Cohort*. 2017, https://www.cacpglobal.ca/images/2017cohort/CACP_Global_Studies_2017_-_Summary_Report_-_Trust_Matters_July_18.pdf.

Calhoun, Craig, editor. *Habermas and the Public Sphere*. MIT Press, 1992.

Chambers, Simone. “Deliberative Democratic Theory.” *Annual Review of Political Science*, vol. 6, 2003, pp. 307–26. https://doi.org/10.1146/annurev.polisci.6.121901.085538.

---. “Navigating Pluralism: Beiner on Habermas.” *The Review of Politics*, vol. 78, no. 2, 2016, pp. 299–302. https://doi.org/10.1017/S0034670516000097.

Chen, Frank. “Exploring AI for Law Enforcement: Insight from an Emerging Tech Expert.” Interview conducted by Chris Hsiung. *Police Chief Magazine*, 20 Sept. 2023, www.policechiefmagazine.org/exploring-ai-law-enforcement-interview/. Accessed 1 Mar. 2025.

Cote, Joe. “What Is New Media?” *Southern New Hampshire University Newsroom*, 23 Sept. 2022 (updated 30 Aug. 2023), Southern New Hampshire University, https://www.snhu.edu/about-us/newsroom/liberal-arts/what-is-new-media. Accessed 21 Aug. 2025.

Cross, James L., et al. “Bias in Medical AI: Implications for Clinical Decision-Making.” *PLOS Digital Health*, vol. 3, no. 11, 2024, Article e0000651. https://doi.org/10.1371/journal.pdig.0000651.

De Almeida, Patricia G. R., et al. “Artificial Intelligence Regulation: A Framework for Governance.” *Ethics and Information Technology*, vol. 23, no. 1, 2021, Article 23. https://doi.org/10.1007/s10676-021-09593-z.

Denham, Brittany. “2024 Trends: Social Media in Canada.” *Environics: Research*, 9 May 2024, environics.ca/insights/articles/2024-trends-social-media-in-canada/. Accessed 10 Feb. 2025.

Diakopoulos, Nicholas. “Transparency.” *The Oxford Handbook of Ethics of AI,* edited by Markus D. Dubber, et al. Oxford Academic, 2020, pp. 197–213. https://doi.org/10.1093/oxfordhb/9780190067397.013.11.

Ellul, Jacques. *The Technological Society*.

Epstein, Brandon, and James Emerson. “Navigating the Future of Policing: Artificial Intelligence (AI) Use, Pitfalls, and Considerations for Executives.” *Police Chief Magazine*, 3 Apr. 2024, Police Chief Online. *Police Chief Magazine*.

Epstein, Brandon, and James Emerson. “Navigating the Future of Policing: Artificial Intelligence (AI) Use, Pitfalls, and Considerations for Executives.” *Police Chief Magazine*, 3 Apr. 2024, https://www.policechiefmagazine.org/navigating-future-ai-chatgpt/. Accessed 18 Aug. 2025.

Ezzeddine, Yasmine, et al. “Citizen Perspectives on Necessary Safeguards to the Use of AI by Law Enforcement Agencies.” *arXiv*, 2023. https://doi.org/10.48550/arXiv.2306.01786.

Finlayson, James G. *Habermas: A Very Short Introduction*. Oxford UP, 2005.

---, and Dafydd Huw Rees. “Jurgen Habermas.” *Stanford Encyclopedia of Philosophy*, 15 Sept. 2023, https://plato.stanford.edu/entries/habermas.

Flew, Terry. *New Media: An Introduction*. Oxford UP, 2002.

Fraser, Nancy. “Rethinking the Public Sphere: A Contribution to the Critique of Actually Existing Democracy.” *Social Text*, no. 25/26, 1990, pp. 56–80. https://doi.org/10.2307/466240.

Gillis, Wendy. “144 Toronto Police Officers Signed Up to Use Clearview AI Mass Surveillance Tech.” *Toronto Star*, 29 Mar. 2024, www.thestar.com/news/gta/144-toronto-police-officers-signed-up-to-use-clearview-ai-mass-surveillance-tech/article_2cf7befc-7c22-5b47-bcb7-ab365d426569.html. Accessed 1 Mar. 2025.

Glikson, Ella, and Anita Williams Woolley. “Human Trust in Artificial Intelligence: Review of Empirical Research.” *The Academy of Management Annals*, vol. 14, no. 2, 2020, pp. 627–60. https://doi.org/10.5465/annals.2018.0057.

Government of Canada. “What is Artificial Intelligence (AI)?” *The Department of National Defence and Canadian Armed Forces Artificial Intelligence Strategy*, 11 Mar. 2024, https://www.canada.ca/en/department-national-defence/corporate/reports-publications/dnd-caf-artificial-intelligence-strategy/what-is-ai.html. Accessed 18 Aug. 2025.

Gunthro, Clem. Review of Broussard, Meredith, *Artificial Unintelligence: How Computers Misunderstand the World*. *Journal of Intellectual Freedom & Privacy*, vol. 3, no. 2–3, 2018.

Guttman, A. “Media Usage in Canada - Statistics & Facts.” *Statista*, 24 Apr. 2024, www.statista.com/topics/1612/media-in-canada/. Accessed 18 Aug. 2025.

Habermas, Jürgen. *A New Structural Transformation of the Public Sphere and Deliberative Politics*. Translated by Ciaran Cronin, English edition, Polity, 2023.

---. *Between Facts and Norms: Contributions to a Discourse Theory of Law and Democracy*. MIT Press, 1996.

---. "The Public Sphere: An Encyclopedia Article (1964)." *New German Critique*, no. 3, 1974, pp. 49–55. *JSTOR*, https://doi.org/10.2307/487737. Accessed 14 Apr. 2025.

---. "Further Reflections on the Public Sphere." *Habermas and the Public Sphere*, edited by Craig Calhoun. MIT Press, 1992, pp. 421–61.

---. *Knowledge and Human Interests*. Polity Press, 1991.

---. *Justification and Application*. MIT Press, 1994.

---. *Legitimation Crisis*. Translated by Thomas McCarthy. Beacon Press, 1975.

---. *Moral Consciousness and Communicative Action*. Translated by Christian Lenhardt and Shierry Weber Nicholsen. Polity Press, 2007.

---. "Political Communication in Media Society: Does Democracy Still Enjoy an Epistemic Dimension? The Impact of Normative Theory on Empirical Research." *Communication Theory*, vol. 16, no. 4, 2006, pp. 411–26. https://doi.org/10.1111/j.1468-2885.2006.00280.x.

---. *Toward a Rational Society: Student Protest, Science, and Politics*. Translated by Jeremy J. Shapiro, Beacon Press, 1970. Chapter: "Technology and Science as 'Ideology'", pp. 81–122.

---. *The Philosophical Discourse of Modernity: Twelve Lectures*. Translated by Frederick G. Lawrence, MIT Press, 1987.

---. *The Structural Transformation of the Public Sphere: An Inquiry into a Category of Bourgeois Society*. Translated by Thomas Burger, with the assistance of Frederick Lawrence, MIT Press, 1989.

---. *The Theory of Communicative Action. Volume 1: Reason and the Rationalization of Society*. Translated by Thomas McCarthy. Polity Press, 1984.

---. *The Theory of Communicative Action. Volume 2: Lifeworld and System: A Critique of Functionalist Reason*. Translated by Thomas McCarthy. Polity Press, 1987.

---. *Toward a Rational Society*. Beacon Press, 1970.

---. *Trust and Justification*. Polity Press, 2017.

Häußermann, Johann Jakob, and Christoph Lütge. "Community-in-the-Loop: Towards Pluralistic Value Creation in AI, or—Why AI Needs Business Ethics." *AI and Ethics*, vol. 2, no. 2, 2022, pp. 341–62. https://doi.org/10.1007/s43681-021-00047-2.

Higgitt, Ryan. "Habermas' Communicative Rationality and Connectionist AI." *Culture, Theory and Critique*, vol. 52, no. 1, 2011, pp. 83–100. https://doi.org/10.1080/14735784.2011.621664.

Inspectorate of Policing Ontario. "Ontario's Inspector General of Policing." Inspectorate of Policing, n.d., https://www.iopontario.ca/en#:~:text=Ontario's%20Inspector%20General%20of%20Policing,Ontario's%20policing%20legislation%20and%20regulations.

Jackson, Jonathan, et al. *Policing by Consent: Understanding the Dynamics of Police Power and Legitimacy*. 2012, http://eprints.lse.ac.uk/47220/1/Policing%20by%20consent(lsero).pdf.

Jenkins, Michael, and Chris Shields. "A Principled Approach to Artificial Intelligence for Policing." *Police Chief Magazine*, 10 Apr. 2024, www.policechiefmagazine.org/principled-approach-ai/. Accessed 1 Mar. 2025.

Joh, Elizabeth E. "Ethical AI in American Policing." *Journal on Emerging Technologies*, vol. 3, no. 2, 2022, 262–87.

Kim, Dongwoo. “Artificial Intelligence Governance through Communicative Action: A Habermasian Perspective.” *University of Toronto Faculty of Law Review*, vol. 82, no. 1, 2024, pp. 103–25.

King, Jennifer, and Caroline Meinhardt. *Rethinking Privacy in the AI Era*. 2024, hai-production.s3.amazonaws.com/files/2024-02/White-Paper-Rethinking-Privacy-AI-Era.pdf.

König, Pascal D., and Georg Wenzelburger. “Toward a Theory of Political Strategy in Policy Analysis.” Politics & Policy, vol. 42, no. 3, 2014, pp. 400–30. https://doi.org/10.1111/polp.12076.

Kwon, Jihyun, and Scot Wortley. “Policing the Police: Public Perceptions of Civilian Oversight in Canada.” *Race and Justice*, vol. 12, no. 4, 2022, pp. 644–68, https://doi.org/10.1177/2153368720924560.

Marda, Vidushi. “Artificial Intelligence Policy in India: A Framework for Engaging the Limits of Data-Driven Decision-Making.” *Philosophical Transactions of the Royal Society A*, vol. 376, no. 2133, 2018. https://doi.org/10.1098/rsta.2018.0087.

Mazerolle, Lorraine, and Janet Ransley. *Third Party Policing*. Cambridge UP, 2006.

McCandless, Sean, and G. M. Vogler. “‘Habermasville’: Police–Community Intersections and Communicative Rationality.” Administrative Theory & Praxis, vol 42, no. 4, 2020, 443–58. https://doi.org/10.1080/10841806.2019.1678350.

Monti, Paolo. “AI Enters Public Discourse: A Habermasian Assessment of the Moral Status of Large Language Models.” *Etica & Politica/Ethics & Politics*, vol. 26, no. 1, 2024, pp. 61–80.

Moses, Lyria Bennett, and Janet Chan. “Algorithmic Prediction in Policing: Assumptions, Evaluation, and Accountability.” *Policing and Society*, vol. 28. no. 7, 2018, pp. 806–22. https://doi.org/10.1080/10439463.2016.1253695.

National Association for the Advancement of Colored People. “Artificial Intelligence in Predictive Policing Issue Brief.” *NAACP*, n.d., naacp.org/resources/artificial-intelligence-predictive-policing-issue-brief. Accessed 12 Mar. 2025.

Nemitz, Paul. “Constitutional Democracy and Technology in the Age of Artificial Intelligence.” *Philosophical Transactions of the Royal Society A*, vol. 376, no. 2133, 2018. https://doi.org/10.1098/rsta.2018.0089.

Ntoutsi, Eirini, et al. “Bias in Data-Driven Artificial Intelligence Systems—An Introductory Survey.” *WIREs Data Mining and Knowledge Discovery*, vol. 19, no. 3, 2020, Article e1356. https://doi.org/10.1002/widm.1356.

O’Connor, Christopher D., et al. “Thinking About Police Data: Analysts’ Perceptions of Data Quality in Canadian Policing.” *The Police Journal,* vol. 95, no. 4, 2022, pp. 637–56. https://doi.org/10.1177/0032258X211021461.

O’Neil, Cathy, and Hanna Gunn, “Near-Term Artificial Intelligence and the Ethical Matrix.” *Ethics of Artificial Intelligence*, edited by S. Matthew Liao. Oxford Academic, 2020, 237–70. https://doi.org/10.1093/oso/9780190905033.003.0009.

Ontario Human Rights Commission. *From Impact to Action: Final Report into Anti-Black Racism by the Toronto Police Service*. 2023, www3.ohrc.on.ca/en/impact-action-final-report-anti-black-racism-toronto-police-service.

Pasquale, Frank. *The Black Box Society: The Secret Algorithms That Control Money and Information.* Harvard UP, 2015. https://doi-org.proxy.bib.uottawa.ca/10.4159/harvard.9780674736061.

Pereira, Fábio Henrique. "Media Crisis and Its Effects on Digital Journalism Careers in Canada." *Journalism*, vol. 24, no. 11, 2023, pp. 2533–50. https://doi.org/10.1177/14648849221116186.

Public Policy Forum. *The Shattered Mirror: News, Democracy and Trust in the Digital Age*. 2017, https://ppforum.ca/wp-content/uploads/2017/01/theShatteredMirror.pdf.

Puddister, Kate. "Oversight and Accountability for Serious Incidents in Canada: Who Polices the Police?" *Canadian Public Administration,* vol. 66, pp. 390–408. https://doi.org/10.1111/capa.12538.

Puddister, Kate, and Danielle McNabb. "Police Oversight in Practice: The Special Investigations Unit and Civilian Police Oversight in Ontario, Canada." *Policing & Society*, vol. 34, no. 8, 2024, pp. 831–45. https://doi.org/10.1080/10439463.2024.2339364.

Rasmussen, David M. *Reading Habermas*. Blackwell, 1990.

Rawashdeh, Samir. "AI's Mysterious 'Black Box' Problem, Explained." *University of Michigan-Dearborn News*, 6 Mar. 2023, umdearborn.edu/news/ais-mysterious-black-box-problem-explained. Accessed 18 Aug. 2025.

Rehg, William. *Insight & Solidarity: The Discourse Ethics of Jürgen Habermas*. University of California Press, 1994.

Robertson, Kate, et al. *To Surveil and Predict: A Human Rights Analysis of Algorithmic Policing in Canada*. 2020, https://citizenlab.ca/wp-content/uploads/2020/09/To-Surveil-and-Predict.pdf.

Royal Canadian Mounted Police. *Transparency Blueprint: Snapshot of Operational Technologies*. N.d., rcmp.ca/sites/default/files/doc/national-technology-onboarding-program-transparency-blueprint.pdf.

Russell, Matheson. *Habermas and Politics: A Critical Introduction*. Edinburgh UP, 2019.

Samuels-Wortley, Kanika. "To Serve and Protect Whom? Using Composite Counter-Storytelling to Explore Black and Indigenous Youth Experiences and Perceptions of the Police in Canada." *Crime & Delinquency*, vol. 67, no. 8, 2021. https://doi.org/10.1177/0011128721989077.

Sengupta, Ushnish. "Black Box Algorithmic Decision-Making and Transparency Challenges in Policing Practice: Lessons from Implementation of New Technologies by the Toronto Police Service." *SSRN*, 2024. https://dx.doi.org/10.2139/ssrn.5000333.

Shepherd, Tamara. "The Canadian Clearview AI Investigation as a Call for Digital Policy Literacy." *Surveillance & Society*, vol. 22, no. 2, 2024. https://doi.org/10.24908/ss.v22i2.16300.

Should AI Play an Ever-Growing Role in Tackling Crime? BBC News, 26 Jan. 2024, section *Business*. BBC. URL: https://www.bbc.com/news/business-68067081.

Sklansky, David A. *Democracy and the Police.* Stanford UP, 2007. https://doi.org/10.1515/9780804763226.

"Social Media." *Encyclopaedia Britannica*, 2025. https://www.britannica.com/topic/social-media. Accessed 18 Aug. 2025.

Solomon, Stephanie, and Julia Abelson. "Why and When Should We Use Public Deliberation?" *The Hastings Center Report*, vol. 42, no. 2, 2012, pp. 17–20. https://doi.org/10.1002/hast.27.

Soules, Marshall. “Jürgen Habermas and the Public Sphere.” *Media Studies*, 2007, https://www.media-studies.ca/articles/habermas.htm. Accessed 18 Aug. 2025.

Splichal, Slavko. *Datafication of Public Opinion and the Public Sphere: How Extraction Replaced Expression of Opinion*. Anthem Press, 2022

---. *Principles of Publicity and Press Freedom*. Rowman & Littlefield Publishers, 2002.

Standing Committee on Transport and Communications. *Final Report on the Canadian News Media*. Vol. 2, 2006.

Stanley, Jay. “ACLU White Paper on Police Departments’ Use of AI to Draft Police Reports.” American Civil Liberties Union, 10 Dec. 2024, www.aclu.org/documents/aclu-on-police-departments-use-of-ai-to-draft-police-reports.

Statistics Canada. “Confidence in Institutions and the Media, 2023.” The Daily, 13 Feb. 2024, https://www150.statcan.gc.ca/n1/daily-quotidien/240213/dq240213a-eng.htm. Accessed 18 Aug. 2025.

---. “Confidence in the Police, the Justice System and Courts, the Federal Parliament, and the Canadian Media Varied across Racialized Groups.” *The Daily*, 31 Mar. 2023, https://www150.statcan.gc.ca/n1/daily-quotidien/230331/dq230331c-eng.htm. Accessed 18 Aug. 2025.

Strani, Aikaterini. *A Critical Study of Communicative Rationality in Habermas’s Public Sphere*. 2011. Heriot-Watt University, dissertation.

Tennison, Jeni. “Why We Need Inclusive Data Governance in the Age of AI.” CIGI, 12 Nov. 2024, https://www.cigionline.org/articles/why-we-need-inclusive-data-governance-in-the-age-of-ai/.

Therier, Adam. “Is it ‘Techno-Chauvinist’ & ‘Anti-Humanist’ to Believe in the Transformative Potential of Technology?” *The Technology Liberation Front*, 18 Sept. 2018, https://techliberation.com/2018/09/18/is-it-techno-chauvinist-anti-humanist-to-believe-in-the-transformative-potential-of-technology/. Accessed 18 Aug. 2025.

Törnberg, Petter. “How Digital Media Drive Affective Polarization through Partisan Sorting.” *Proceedings of the National Academy of Sciences*, vol. 119, no. 42, 2022, Article e2207159119. https://doi.org/10.1073/pnas.2207159119.

Toronto Police Service. “Artificial Intelligence.” *Police Reform*, n.d., www.tps.ca/police-reform/artificial-intelligence/. Accessed 1 Mar. 2025.

Tulloch, Michael H. *Report of the Independent Street Checks Review*. 2018, www.ontario.ca/page/report-independent-street-checks-review.

Turner Lee, Nicol. “Detecting Racial Bias in Algorithms and Machine Learning.” *Journal of Information, Communication and Ethics in Society*, vol. 16, no. 3, 2018, pp. 252–60.

Tyler, Tom R., and Jeffrey Fagan. “Legitimacy and Cooperation: Why do People Help the Police Fight Crime in Their Communities?” *Ohio State Journal of Criminal Law*, vol. 6, 2008, 231–75.

United Nations Educational, Scientific and Cultural Organization. “Artificial Intelligence: Examples of Ethical Dilemmas.” *UNESCO*, 21 Apr. 2023, www.unesco.org/en/artificial-intelligence/recommendation-ethics/cases. Accessed 12 Mar. 2025.

Vestby, Annette, and Jonas Vestby. “Machine Learning and the Police: Asking the Right Questions.” *Policing*, vol. 15, no. 1, 2021, pp. 44–58. https://doi.org/10.1093/police/paz035.

Virilio, Paul, *Politics of the Very Worst: An Interview with Philippe Petit*, edited by Sylvère Lotringer. Translated by Michael Cavaliere. Semiotext(e), 1999.

Winseck, Dwayne. *Media and Internet Concentration in Canada, 1984–2022*. 2023, https://doi.org/10.22215/gmicp/2023.2.

Wortley, Scot, and Akwasi Owusu-Bempah. "Race, Police Stops, and Perceptions of Anti-Black Police Discrimination in Toronto, Canada over a Quarter Century." *Policing*, vol. 45, no. 4, 2022, pp. 570–85. https://doi.org/10.1108/PIJPSM-11-2021-0157.

Xivuri, Khensani, and Hosanna Twinomurinzi. "A Habermasian Approach to Fair Processes in AI Algorithms." *Artificial Intelligence Research*, vol. 1551. Springer International Publishing, 2022, pp. 335–43. https://doi.org/10.1007/978-3-030-95070-5_22.

Yale School of Medicine. "'Bias In, Bias Out': Yale Study Identifies Bias in Medical AI." *NewsWise*, 22 Nov. 2024, www.newswise.com/articles/bias-in-bias-out-yale-study-identifies-bias-in-medical-ai. Accessed 18 Aug. 2025.

Zhang, Angela H. "Authoritarian Countries' AI Advantage." *Project Syndicate*, 4 Sept. 2024, www.project-syndicate.org/commentary/how-china-uae-became-artificial-intelligence-powerhouses-by-angela-huyue-zhang-2024-09. Accessed 18 Aug. 2025.

Zuccarelli, Eugenio. "Building Trust in AI Means Moving Beyond Black-Box Algorithms: Here's Why." *World Economic Forum*, 2 Apr. 2024, https://www.weforum.org/stories/2024/04/building-trust-in-ai-means-moving-beyond-black-box-algorithms-heres-why/. Accessed 18 Aug. 2025.