

L'intelligence artificielle comme assistante dans les revues systématiques

Repenser la collaboration humain-machine dans la synthèse des connaissances en santé

Vincent Martin-Schreiber

2025-05-07

Université d'Ottawa

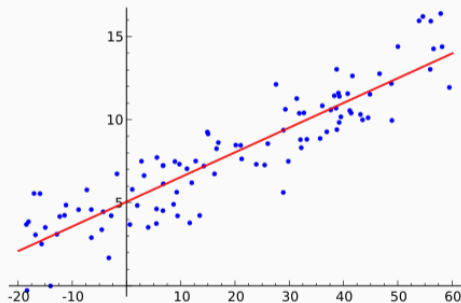
Perspective historique



Grattoir double sur lame, grotte d'Aurignac (Haute-Garonne, France).

Bible de Gutenberg (*Bible de Gutenberg*, 1455)

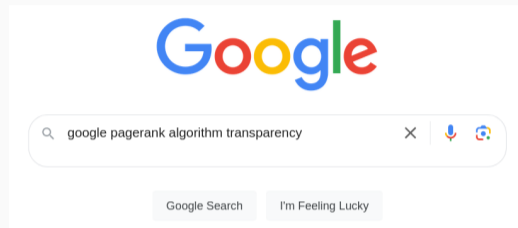
“in machine learning decisions are based on algorithms built with data” (Irizarry, 2019)



And also:

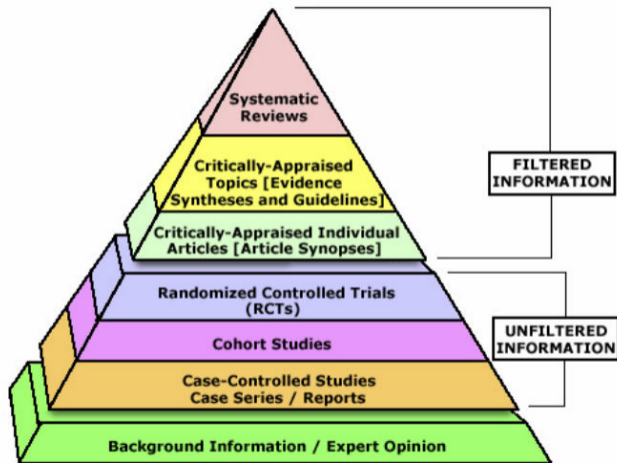
- spam filtering
- credit scoring
- Text prediction and autocorrect on your smartphone
- music streaming recommendations

Boîtes noires au quotidien



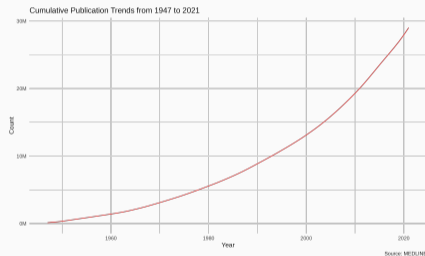
Le défi croissant des revues systématiques

Médecine basée sur des données probantes

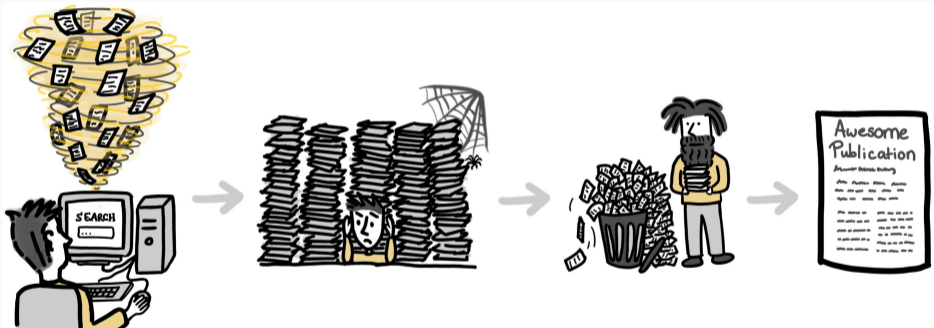


L'explosion de la littérature scientifique

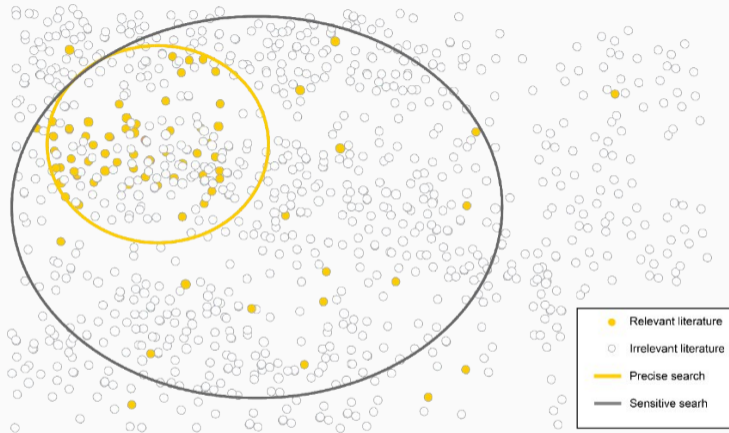
- Plus de 36 millions d'articles dans PubMed (2023)
- Comment synthétiser ces connaissances en expansion ?



Processus des revues systématiques



Choisir la taille du filet : un choix cornélien



Comparaison hypothétique entre une recherche précise et une recherche sensible, montrant que même une approche sensible peut ignorer des articles pertinents.

Processus classique:

1. Définition de la question de recherche
2. Recherche exhaustive dans les bases de données
3. Sélection manuelle basée sur titres/résumés (>95% d'exclusions)
4. Lecture des textes complets
5. Extraction et synthèse des données

Limitations:

- Chronophages
- Risque d'erreurs humaines dû à la fatigue et au volume
- Tendance à restreindre les critères de recherche pour limiter le volume

Un processus de synthèse fiable et reproductible ?

La production de revues systématiques et de méta-analyses a atteint des proportions épidémiques. Il est possible que la grande majorité de ces revues soient inutiles, trompeuses et/ou contradictoires.
(J. P. a. Ioannidis, 2016)

Voir aussi : J. P. A. Ioannidis (2005)

Exemples :

- inconsistance des résultats des requêtes dans l'espace et dans le temps
- La plupart des requêtes de revues systématiques ne sont pas reproductibles.

(Pozsgai et al., 2021; Rethlefsen et al., 2024)

Exemple

	2023-12	2025-02	DIFF
2022	4802	4787	-15
2021	4378	4347	-31
2020	3831	3809	-22
2019	3343	3328	-15
2018	2964	2944	-20
2017	2738	2718	-20
2016	2693	2685	-8
2015	2472	2471	-1

Requête : ("bayesian"[All Fields] OR "bayesianism"[All Fields] OR "bayesians"[All Fields]) AND ("method s"[All Fields] OR "methods"[MeSH Terms] OR "methods"[All Fields] OR "method"[All Fields] OR "methods"[MeSH Subheading]) dans la base Pubmed

Le problème actuel:

- Processus chronophage et laborieux
- Taux d'erreur humaine non négligeable - **10.76%** (95% CI: 7.43%-14.09%) (Wang et al., 2020)
- Délais de publication des synthèses de plus en plus longs
- Volume croissant de données versus capacités humaines limitées

Conséquences directes:

- Risque accru de retards dans l'adoption des meilleures pratiques en santé
- Décalage entre l'état des connaissances et la pratique clinique

Tension fondamentale:

- **Rigueur méthodologique:** nécessité d'une évaluation critique humaine
- **Efficacité du processus:** besoin d'effectuer la synthèse des connaissances
- **Transparence:** maintenir la traçabilité des décisions

Questions émergentes:

- Comment préserver la qualité tout en gagnant en efficacité?
- Quel équilibre entre automatisation et expertise humaine?
- Comment rendre ce processus transparent et reproductible?

L'automatisation dans la synthèse des connaissances: risques et opportunités

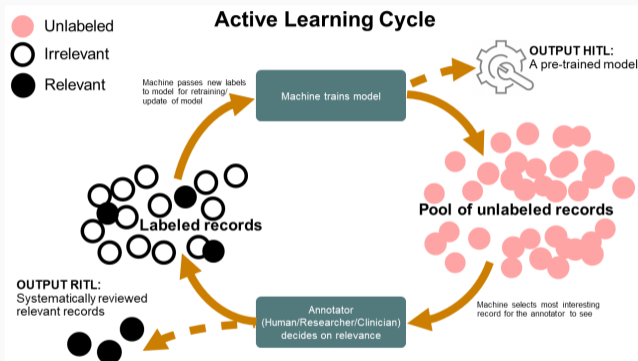
Des outils variés mais avec des limitations:

- **Abstrackr, Rayyan, Colandr:** outils populaires mais souvent en “boîte noire”
- **Covidence:** intégration récente d'outils d'IA mais limites en transparence
- **LLMs:** tentant, mais...

Risques liés à l'automatisation complète:

- Manque de transparence algorithmique
- Risque de biais dans les modèles
- Perte de la nuance et de l'expertise critique humaine

L'apprentissage actif: un paradigme alternatif



Apprentissage actif (Active Learning):

- Approche où le modèle d'IA sélectionne les documents à étiqueter
- Interaction continue entre l'algorithme et le chercheur
- Optimisation progressive du modèle de prédiction
- Focalisation sur les cas les plus informatifs

Principes d'une collaboration humain-machine efficace et éthique

Le concept de “researcher-in-the-loop”

Au-delà du “human-in-the-loop”:

- Conception centrée sur le chercheur, pas seulement sur l’algorithme
- L’humain reste l’autorité décisionnelle finale
- Collaboration plutôt que délégation ou remplacement

“Dans l’approche researcher-in-the-loop, le produit principal du processus est une sélection de références, pas un modèle d’apprentissage machine. Toutes les références pertinentes sont examinées par un humain, et le processus exige une transparence et une reproductibilité complètes.” (van de Schoot et al., 2021)

Fondements d'une collaboration humain-machine éthique:

- **Transparence** des algorithmes et des décisions
- **Équité** dans la représentation des connaissances
- **Respect de l'autonomie** du chercheur
- **Bénéfice** pour la communauté scientifique et les patients
- **Reproductibilité** des résultats

Implications pratiques:

- Nécessité de documenter les paramètres des modèles
- Partage des données d'entraînement
- Compréhension des limites de l'outil par les utilisateurs

Barrières à l'adoption:

- Scepticisme légitime face aux “boîtes noires”
- Cultures disciplinaires diverses face aux IA
- Questions sur la validité scientifique des résultats
- Responsabilité professionnelle et légale

Facteurs favorisant la confiance:

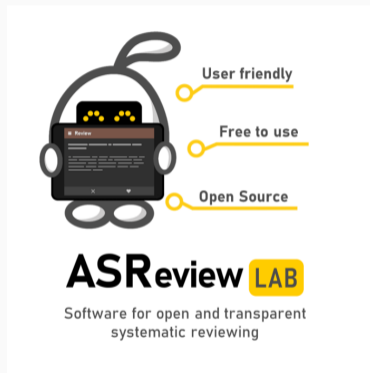
- Transparence des algorithmes
- Validation empirique des performances
- Contrôle maintenu par le chercheur
- Ouverture du code source
- Documentation complète du processus

Le cas d'ASReview: un exemple d'approche “researcher-in-the-loop”

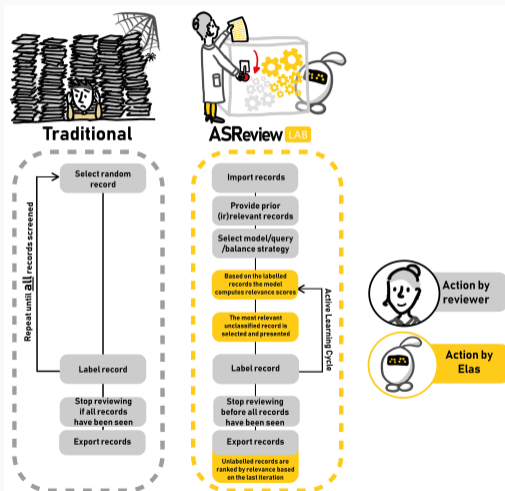
Qu'est-ce qu'ASReview?

Caractéristiques :

- Logiciel libre et open source développé à l'Université d'Utrecht
- Approche “researcher-in-the-loop” (chercheur dans la boucle)
- Utilise l'apprentissage actif pour prioriser les références pertinentes
- Performance démontrée par simulation sur des datasets variés
- Flexibilité des modèles d'IA utilisés

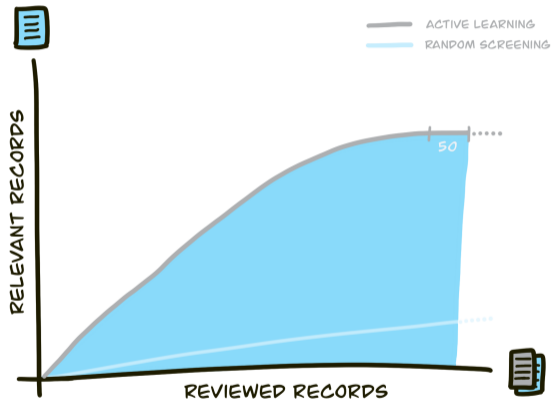


Le cycle d'apprentissage actif dans ASReview



1. Feature Extractor algorithm
2. Classification algorithm
3. Query Strategy
4. Balancing strategy
5. Stopping rule

Recall



Résultats des études de simulation:

- 95% des articles pertinents identifiés après screening de 8% à 33% des références
- 70% à 100% des articles pertinents trouvés dans les premiers 10% examinés
- Économie significative de temps et d'effort

Systematic Review	WSS@95	WSS@100	RRF@10
	Mean [95%CI]	Mean [95%CI]	Mean [95%CI]
Ribbing	72 [71-74]	72 [71-73]	79 [78-81]
RSA	72 [72-73]	62 [61-63]	70 [69-71]
MOM	50 [50-51]	37 [36-38]	58 [56-60]

Figure 1: Courbe de performance typique

WSS@95 is the percentage work saved (over sampling) at 95% recall. WSS@100 is the percentage work saved (over sampling) at 100% recall. RRF@10 is the percentage of relevant references identified after having screened the first 10% of the records.

(Pijls, 2024)

4 ESSENTIAL FREEDOMS OF SOFTWARE

0 To run the software when ever you wish & for what ever purpose.

1 To study the source code & make modifications to the software.

2 To give or sell copies of the software to other people.

3 To give or sell copies of your modified versions of the software.

You have the 4 essential freedoms with other useful items that belong to you. Clothing, Food, Simple Electrical Devices. But most software companies do not want you to have these essential freedoms with software, running on your various devices. Taking away your control over your own devices.

SWITCH INSTEAD TO FREE SOFTWARE!

www.GNU.org

L'importance de l'open source:

- Code entièrement accessible et modifiable
- Documentation complète des algorithmes
- Possibilité d'audit par la communauté scientifique
- Sauvegarde de l'état des modèles et décisions

Fonctionnalités favorisant la reproductibilité:

- Export des décisions de screening avec horodatage
- Documentation des paramètres des modèles
- Possibilité de simulation pour valider les approches
- Pas de "propriété" des données par un tiers

Aspects éthiques et implications pour l'avenir de la recherche en santé

Considérations fondamentales:

- Risque de “délégation excessive” aux algorithmes
- Nécessité d'une expertise pour interpréter les résultats
- Questions sur la responsabilité des décisions
- Risque d'exacerber certains biais dans la littérature

Réponses potentielles:

- Formation des chercheurs à l'utilisation critique des outils
- Documentation explicite du rôle de l'IA dans le processus
- Vérifications humaines stratégiques
- Transparence méthodologique complète

Avantages spécifiques de l'approche open source:

- Auditabilité complète des algorithmes
- Indépendance vis-à-vis d'intérêts commerciaux
- Développement communautaire et pluridisciplinaire
- Évolutivité en fonction des besoins scientifiques
- Préservation de la souveraineté des données

Différence fondamentale avec les approches propriétaires:

- Contraste avec les “boîtes noires” commerciales
- Autonomie des institutions de recherche
- Pérennité et indépendance des outils

Transformations méthodologiques:

- Vers des revues systématiques “vivantes” et collaboratives
- Accélération de la mise à jour des connaissances
- Possibilité d'élargir la portée des revues
- Démocratisation de l'accès aux méthodes systématiques

Impact sur les standards de publication:

- Nouvelles exigences de transparence algorithmique
- Documentation des processus assistés par IA
- Évolution des critères PRISMA et AMSTAR
- Besoins en compétences hybrides chez les chercheurs

**Conclusion: vers une
complémentarité réfléchie et
maîtrisée**

Une distribution optimale des rôles:

- **IA comme assistante:** priorisation, identification de patterns, gestion du volume
- **Humain comme expert:** jugement critique, interprétation contextuelle, décisions finales
- **Collaboration dynamique:** adaptée aux différentes phases et contextes

Un changement de paradigme:

- De l'opposition "homme vs machine" à la collaboration "homme avec machine"
- Des rôles complémentaires plutôt que concurrents
- Une redéfinition de la valeur ajoutée humaine

- collaboration à grande échelle
- “living systematic reviews”

Questions?

Cette présentation est disponible en ligne :

Martin-Schreiber, V. (7 mai 2025). L'intelligence artificielle comme assistante dans les revues systématiques : repenser la collaboration humain-machine dans la synthèse des connaissances en santé. 92e Congrès de l'Acfas (ACFAS 2025), Montréal QC (Canada). Université d'Ottawa. <https://doi.org/10.5281/zenodo.15341589>

DOI 10.5281/zenodo.15341589

Ce contenu est sous licence CC BY-SA 4.0.



Pour en savoir plus:

- Site officiel d'ASReview: asreview.nl
- Code source: github.com/asreview/asreview
- Documentation: asreview.readthedocs.io

Références

Bible de Gutenberg : [Biblia latina]. Ex. sur vélin - Volume 3. (1455).

Ioannidis, J. P. A. (2005). Why Most Published Research Findings Are False. *PLOS Medicine*, 2(8), e124. <https://doi.org/10.1371/journal.pmed.0020124>

Ioannidis, J. P. a. (2016). The Mass Production of Redundant, Misleading, and Conflicted Systematic Reviews and Meta-analyses. *The Milbank Quarterly*, 94(3), 485–514. <https://doi.org/10.1111/1468-0009.12210>

Irizarry, R. A. (2019). *Introduction to data science: Data analysis and prediction algorithms with R*. CRC Press.

- Pijls, B. G. (2024). Machine Learning assisted systematic reviewing in orthopaedics. *Journal of Orthopaedics*, 48, 103–106. <https://doi.org/10.1016/j.jor.2023.11.051>
- Pozsgai, G., Lövei, G. L., Vasseur, L., Gurr, G., Batáry, P., Korponai, J., Littlewood, N. A., Liu, J., Móra, A., Obrycki, J., Reynolds, O., Stockan, J. A., VanVolkenburg, H., Zhang, J., Zhou, W., & You, M. (2021). Irreproducibility in searches of scientific literature: A comparative analysis. *Ecology and Evolution*, 11(21), 14658–14668. <https://doi.org/10.1002/ece3.8154>

- Rethlefsen, M. L., Brigham, T. J., Price, C., Moher, D., Bouter, L. M., Kirkham, J. J., Schroter, S., & Zeegers, M. P. (2024). Systematic review search strategies are poorly reported and not reproducible: A cross-sectional meta-research study. *Journal of Clinical Epidemiology*, *166*, 111229.
<https://doi.org/10.1016/j.jclinepi.2023.111229>
- Schoot, R. van de, Coimbra, B. M., Evenhuis, T., Lombaers, P., Weijdema, F., Bruin, L. de, Neeleman, R., Grandfield, E. M., Sijbrandij, M., Teijema, J. J., Jalsovec, E., Bron, M. P., Winter, S., Bruin, J. de, & Zuiden, M. van. (2025). *The Hunt for the Last Relevant Paper: Blending the best of humans and AI*. OSF.
https://doi.org/10.31234/osf.io/p4xm5_v2

- van de Schoot, R., de Bruin, J., Schram, R., Zahedi, P., de Boer, J., Weijdema, F., Kramer, B., Huijts, M., Hoogerwerf, M., Ferdinands, G., Harkema, A., Willemsen, J., Ma, Y., Fang, Q., Hindriks, S., Tummers, L., & Oberski, D. L. (2021). An open source machine learning framework for efficient and transparent systematic reviews. *Nature Machine Intelligence*, 3(2), 125–133. <https://doi.org/10.1038/s42256-020-00287-7>
- Wang, Z., Nayfeh, T., Tetzlaff, J., O'Blenis, P., & Murad, M. H. (2020). Error rates of human reviewers during abstract screening in systematic reviews. *PLOS ONE*, 15(1), e0227742. <https://doi.org/10.1371/journal.pone.0227742>