





# Intelligence artificielle invisible: Présentation sur le cas de PubMed

---

Clara Turp et Lucy Kiester, Bibliothèques de McGill  
Congrès des professionnel.le.s de l'information, 3 Novembre 2020





# Introduction

# Agenda

- PubMed
- Implications éthiques
- Enseignement



PubMed

# Pourquoi PubMed?

- 2020: Nouvel interface et Best match par défaut
  - Il utilise l'intelligence artificielle
- Financé par les États-Unis: statistiques, algorithmes, décisions sont publiques.
- Correspond aux critères sur les algorithmes invisibles

# Qu'est-ce que PubMed

- Base de données gratuite américaine produite par NCBI
- Donne accès à MEDLINE, et les périodiques biomédicaux centraux à travers le monde.
- Indexation avec MeSH (Vedettes-matières médicales)

# PubMed en chiffres

- 3.3 milliards de recherches en 2017
- 80% des clics arrivent sur la première page des recherches



# Algorithme de PubMed



# Méthodes classiques



- Ordonne les résultats en donnant un poids aux documents en fonction des termes de la question et des termes indexés.
- Combien de fois un terme apparaît dans le document et dans la collection complète?
- TF-IDF, BM25
  - Manières différentes de calculer les poids

# Learning-to-rank: LambdaMART

- BM25 est la première couche qui ordonne les résultats.
- LambdaMART est la seconde couche qui utilise l'intelligence artificielle pour réordonner les 500 premiers résultats.
- Utilisation d'un idéal (les clics) pour entraîner l'algorithme
- Utilisation d'éléments: date, format, ...

# Utilisation de l'intelligence artificielle

- Manipulation des questions:
  - Suggestions en fonction de recherches communes
  - Ajouts à la question avec des termes de vedettes-matières.
  - Correction des erreurs
- Différentiation des noms d'auteurs
  - Regarder les dates et les co-auteurs.



# Implications éthiques de l'intelligence artificielle

# Transparence

- Complexité des algorithmes: effet boîte noire
- Présente un point de vue traditionnel ou populaire
- Transparence = Accessibilité et compréhension
- Manipulations des questions des usagers

# Médecine Factuelle

- Date de l'article et de l'information
- Hiérarchie des formats d'articles (revue systématique, essai clinique, etc.)
- L'algorithme prend en considération la date et le format... mais comment?
  - On ne sait pas!
- Prise de décision clinique
  - Recherche vite, page 1 des résultats (souvent 1er article)

# Expérience des usagers

- “Best Match” est fait pour aider les usagers dans l’océan des milliers d’articles
  - usagers ne passent quasiment jamais la première page des résultats
- Pertinence des résultats perçue vs réelle
  - Éducation de l’algorithme (clics)
  - Usagers qui sont pas des chercheurs experts
- IA est le défaut, mais pas possible de savoir facilement qu’il est IA



# Enseignement des compétences de recherche



# Enseignement de PubMed (et autres!)

- Aborder l'intelligence artificielle dans l'instruction
- PubMed est souvent utilisé dans l'instruction, parce que c'est gratuit et accessible, mais pas la seule avec IA
- Encourager les pensées critiques - dans le cycle de recherche évaluer comment vous avez trouvez les articles, pas seulement les articles eux-mêmes.

# Exemple personnel

- J'enseigne PubMed souvent
  - Plusieurs contextes (recherche, question cliniques)
- Enseigner comment évaluer le biais potentiel d'une recherche, même si c'est pas super évident
  - Applicable au delà du PubMed




# Conclusion

Pensées critiques dans l'instruction

Évaluer les *sources* de l'informations

Soyez conscients comment IA peut faire partie d'une  
base de données, que vous le sachiez ou pas



# Bibliographie

- Ananny M. Toward an Ethics of Algorithms: Convening, Observation, Probability, and Timeliness. *Science, Technology, & Human Values*. 2016;41(1):93-117. doi: 10.1177/0162243915606523.
- Beck JB, Tieder JS. Electronic resources preferred by pediatric hospitalists for clinical care. *Journal of the Medical Library Association* : JMLA. 2015;103(4):177-83.
- Fiorini N, Canese K, Starchenko G, Kireev E, Kim W, Miller V, et al. Best Match: New relevance search for PubMed. *PLoS biology*. 2018;16(8):e2005343. Epub 2018/08/29. doi: 10.1371/journal.pbio.2005343. PubMed PMID: 30153250.
- Fiorini N, Leaman R, Lipman DJ, Lu Z. How user intelligence is improving PubMed. *Nature Biotechnology*. 2018;36(10):937-45. doi: 10.1038/nbt.4267.
- Liu T-Y. Learning to rank for information retrieval. *Learning to Rank for Information Retrieval*. Berlin, Heidelberg: Springer Berlin Heidelberg; 2011. p. 3-30.
- Mittelstadt BD, Allo P, Taddeo M, Wachter S, Floridi L. The ethics of algorithms: Mapping the debate. *Big Data & Society*. 2016;3(2):2053951716679679. doi: 10.1177/2053951716679679.
- National Library of Medicine. PubMed Overview: National Library of Medicine; [updated September 6, 2019; cited 2020 March 23]. Available from: <https://www.nlm.nih.gov/bsd/pubmed.html>.
- NCBI. Machine-learning based pipeline relying on LambdaMART currently used in PubMed for relevance (Best Match) searches GitHub [updated April 30, 2018; cited 2020 March 22]. Available from: <https://github.com/ncbi-nlp/PubMed-Best-Match>.
- Noble SU. *Algorithms of Oppression: How Search Engines Reinforce Racism*: NYU Press; 2018.