

MÉTRIQUES D'ÉVALUATION EN RI DE LA QUALITÉ DES LIENS HYPERTEXTE DU WEB

Dominique Loyer

DIC-9345 TALN

27 mars 2025

INTRODUCTION

- -Nous sommes de plus de plus «noyé» par le surplus d'information. (des centaines de zéta-octets en 2025. 10^{21} ->Googol c'est 10^{100})
- -Ainsi, le défi de nos jours n'est pas de trouver de l'information mais de pouvoir connaître quelle information est plus véridique
- **Les LLMS sont très opaques (boite noire) malgré les modèles ouverts comme DeepSeek, le modèle est ouvert mais les données ne le sont pas.**

PROBLÉMATIQUES ADRESSÉES:

- -Comment aider l'utilisateur lambda à se démêler dans tout ça et pouvoir bénéficier d'un
- système avec une vocation de «pensée critique» et pouvoir s'informer de façon éclairée
- non seulement avec des points de vue qui vont dans le sens de l'utilisateur mais aussi
- avec des points de vue divergents et documentés

ÉTAT DE L'ART

- Liao, Q. V. et Wortman Vaughan, J. (2024). AI Transparency in the Age of LLMs: A Human-Centered Research Roadmap. *Harvard Data Science Review*, (Special Issue 5). <https://doi.org/10.1162/99608f92.8036d03b>
- Mahari, R., South, T. et Pentland, S. (2023). *TRANSPARENCY BY DESIGN FOR LARGE LANGUAGE MODELS*.
- Zou, A., Phan, L., Chen, S., Campbell, J., Guo, P., Ren, R., Pan, A., Yin, X., Mazeika, M., Dombrowski, A.-K., Goel, S., Li, N., Byun, M. J., Wang, Z., Mallen, A., Basart, S., Koyejo, S., Song, D., Fredrikson, M., ... Hendrycks, D. (2023, 10 octobre). *Representation Engineering: A Top-Down Approach to AI Transparency*. arXiv. <https://doi.org/10.48550/arXiv.2310.01405>
- Acar, G., Eubank, C., Englehardt, S., Juarez, M., Narayanan, A. et Diaz, C. (2014). The web never forgets: Persistent tracking mechanisms in the wild. Dans *Proceedings of the ACM Conference on Computer and Communications Security* (p. 674-689). Association for Computing Machinery. <https://doi.org/10.1145/2660267.2660347>
- Yang, T., Zhu, X. et Gurevych, I. (2024, 16 juillet). *Robust Utility-Preserving Text Anonymization Based on Large Language Models*. arXiv. Récupéré le 2 novembre 2024 de <http://arxiv.org/abs/2407.11770>
- Jaff, E., Wu, Y., Zhang, N. et Iqbal, U. (2024, 23 août). *Data Exposure from LLM Apps: An In-depth Investigation of OpenAI's GPTs*. arXiv. Récupéré le 2 novembre 2024 de <http://arxiv.org/abs/2408.13247>
- Staab, R., Vero, M., Balunovic, M. et Vechev, M. (2024). *LARGE LANGUAGE MODELS ARE ANONYMIZERS*.
- Chaudhary, G. (2024). Unveiling the Black Box: Bringing Algorithmic Transparency to AI. *Masaryk University Journal of Law and Technology*, 18(1), 93-122. <https://doi.org/10.5817/MUJLT2024-1-4>
- Diakopoulos, N. (s. d.). *Algorithmic Accountability: On the Investigation of Black Boxes*. Columbia Journalism Review. Récupéré le 2 novembre 2024 de https://www.cjr.org/tow_center_reports/algorithmic_accountability_on_the_investigation_of_black_boxes.php/

ÉTAT DE L'ART (EN RÉSUMÉ...)

- En somme:
 - Besoin de reconciliation des deux paradigmes que sont les systemes logiques et les réseaux neuronaux (connectionnistes)
 1. À base de règles (logics: if-then, probabilist, fuzzy, predicate) pour la transparence
 2. RN pour la robustesse et «scalability»
- Imputabilité des géants comme Google, Meta, etc. pour leurs algorithms
- «humain in the loop»-> AI4people, AI4good

LE GAP À COMBLER

- -Plusieurs grands modèles de langues fournissent déjà des sources mais très peu de modèles ne donnent une note (score) ou ne donnent les «benchmark» des métriques même si ils évaluent les sources.
- -Quelle est la qualité des sources?
- -Qui et combien de personnes partagent le point de vue des sources? (difficile à dire avec le référencement non-organique e.g. de Google)
- -A-t' on des sources avec des théories en contradictions dans nos sources? Quelle proportion?

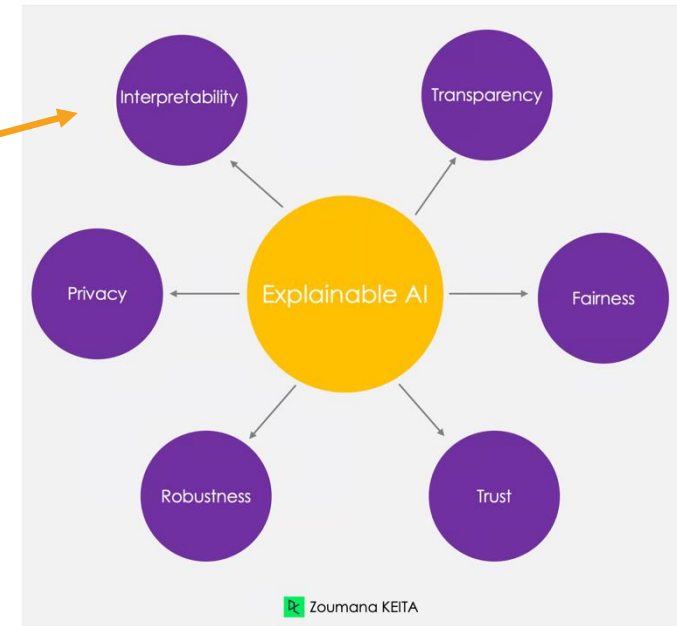
OBJECTIFS:

Clique sur
le mot
Rappel!

- -Mieux comprendre les résultats obtenus d'une recherche d'information (RI)
- Rappel Pertinence, R-Precision, F1 score, BM-25 et plusieurs autres métriques éventuellement (expliqués en termes simples à **l'utilisateur λ**)

PRINCIPES GUIDANT L'EXPLICABILITÉ DE L'AI (XAI)

Nous allons nous concentrer sur la partie : Interprétabilité



- Source:
- <https://www.datacamp.com/tutorial/explainable-ai-understanding-and-trusting-machine-learning-models>

POURQUOI LA PRECISION N'EST PAS SUFFISSANTE?

(EXPLIQUÉE POUR \wedge)

Clique sur le tableau!

Situation	Nombre d'oranges	Nombre de pommes	Précision annoncée	Remarque
Cas équilibré	10	10	70 %	Bon équilibre
Cas déséquilibré	9	1	90 % (oranges)	Pomme ignorée

MÉTHODOLOGIE

- Deux modèles «suborates»¹ implémentés en Python

1. -LIME

- Local Interpretable Model-agnostic Explanations
 - Dans le cas du texte, des mots sont aléatoirement enlevés
 - Dans le cas des données tabulaires «data frames», les caractéristiques «features» sont aléatoirement un peu variées ou carrément éteintes
- Les perturbations instanciées sont rajoutées au modèle prédictif et les instanciations perturbées sont enregistrées et peuvent être interpréter par des modèles transparents comme des arbres de décision ou régressions linéaires.

2. -SHAP

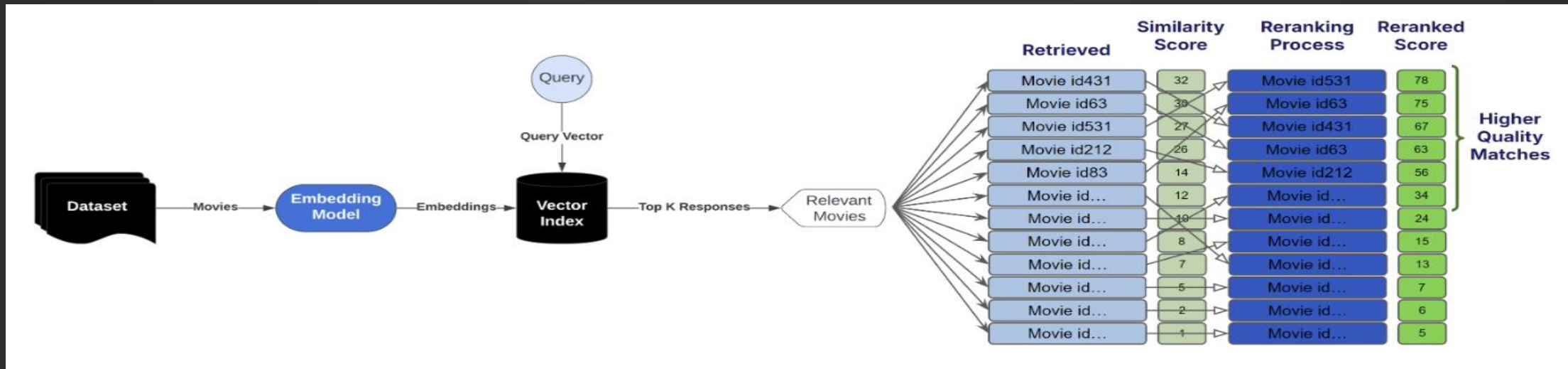
- SHapley Additive exPlanations
 - Largement inspirée de la théorie des jeux et prend en considération la contribution de chaque caractéristique «feature» au modèle prédictif.
 - Contrairement à LIME qui donne une interprétabilité locale, ici SHAP en donne une locale mais aussi globale

CHALLENGES

- Constat: prend beaucoup de CPU,TBU
- Avoir accès à des serveurs qui permettent la parallélisation est essentielle
 - E.g. Même les 25000 documents du TREC peuvent prendre au delà de 48 heures pour l'indexation si on utilise les stopwords seulement et 3j pour les lemmes et cela , en utilisant un système distribué.

RÉSULTATS ET LIMITATIONS

- Résultats d'un «semantic search reranker» dans le contexte d'un RAG pipeline : SALES FORCE



Role of a Reranker in RAG Systems



Reranker's Role:

- Refines initial retrieval results for improved relevance
 - Enhances quality of information for generation
- Uses language models to capture nuanced relevance beyond semantic similarity

source: <https://docs.together.ai/docs/how-to-improve-search-with-rerankers>

source: <https://www.salesforce.com/blog/llamarank/>

RÉSULTATS ET LIMITATIONS

- **Limites envisagées**
 - Coût: Les GPU voire TBU finissent par coûter chers
 - Coût : temporel (beaucoup long que les micro-secondes de Google)
 - Complément aux moteurs de recherche ou outils RI
 - Les autres facettes de l'xAI doivent être aussi intégrées (Vie privée, Transparence des algorithmes, Éthique et équité, la confiance «trust», etc.)

RÉFÉRENCES:

- Bérubé, N. (2022, 20 novembre). Identité numérique: Une solution, mille interrogations. *La Presse*, Contexte. <https://www.lapresse.ca/contexte/2022-11-20/identite-numerique/une-solution-mille-interrogations.php>
- Cagan, A. (2024). L'IA générative a été surestimée. *L'Express*, (3819), 58-60.
- Identité numérique | Une solution, mille interrogations | La Presse*. (s. d.). Récupéré le 22 novembre 2024 de <https://www.lapresse.ca/contexte/2022-11-20/identite-numerique/une-solution-mille-interrogations.php>
- Jaff, E., Wu, Y., Zhang, N. et Iqbal, U. (2024, 23 août). *Data Exposure from LLM Apps: An In-depth Investigation of OpenAI's GPTs*. arXiv. Récupéré le 2 novembre 2024 de <http://arxiv.org/abs/2408.13247>
- Jobin, A., Ienca, M. et Vayena, E. (s. d.). *Artificial Intelligence: the global landscape of ethics guidelines*.
- Lemire, S. G., Franck Jovanovic, Benjamin Lehaire, Daniel. (2022, 10 juin). *L'identité numérique: quels enjeux et quels risques?* Le Journal de Montréal. <https://www.journaldemontreal.com/2022/06/10/identite-numerique-quels-enjeux-et-quels-risques>
- l'information, C. d'accès à. (s. d.). *Loi 25 - Nouvelles dispositions protégeant la vie privée des Québécois - Certaines dispositions entrent en vigueur aujourd'hui*. Récupéré le 22 novembre 2024 de <https://www.newswire.ca/fr/news-releases/loi-25-nouvelles-dispositions-protegeant-la-vie-privee-des-quebecois-certaines-dispositions-entrent-en-vigueur-aujourd-hui-850231083.html>
- Loi 25 - Nouvelles dispositions protégeant la vie privée des Québécois - Certaines dispositions entrent en vigueur aujourd'hui*. (s. d.). Gouvernement du Québec. Récupéré le 2 novembre 2024 de <https://www.quebec.ca/nouvelles/actualites/details/loi-25-nouvelles-dispositions-protegeant-la-vie-privee-des-quebecois-certaines-dispositions-entrent-en-vigueur-aujourd'hui-43212>
- Mahari, R., South, T. et Pentland, S. (2023). *TRANSPARENCY BY DESIGN FOR LARGE LANGUAGE MODELS*.
- Staab, R., Vero, M., Balunovic, M. et Vechev, M. (2024). *LARGE LANGUAGE MODELS ARE ANONYMIZERS*.
- Yang, T., Zhu, X. et Gurevych, I. (2024, 16 juillet). *Robust Utility-Preserving Text Anonymization Based on Large Language Models*. arXiv. Récupéré le 2 novembre 2024 de <http://arxiv.org/abs/2407.11770>

CONCLUSION ET DISCUSSION

- Merci!