TECHNISCHE UNIVERSITÄT WIEN

CD Lab RecSys

18th ACM Conference on Recommender Systems | Bari, Italy, 14-18 October 2024

Enhancing Cross-Domain Recommender Systems with LLMs: Evaluating Bias and Beyond-Accuracy Measures

Thomas E. Kolb, PhD Student & Project Assistant

TU Wien Informatics
Institute of Information Systems Engineering
Research Unit: Data Science
Supervisor: Assistant Prof. Dr. Julia Neidhardt

What Is This All About?

Generative AI (LLMs)

Cross-Domain

Beyond Accuracy

Long-Running A/B Test

Aim of This Work

- Investigate Traditional vs. LLM-Enhanced Cross-Domain Recommender Systems
- Detect and Mitigate Bias
- Evaluate Beyond Accuracy
- Applied within **a real-world setting** in the **domain of news, books, and various lifestyle areas.**

Research Questions

RQ1: To what extent do **LLM-based recommendation methods** improve **cross-domain recommender systems**, and which aspects like user personalization and relevance benefit the most?

RQ2: To what extent do **beyond-accuracy metrics** (novelty, diversity, serendipity) **impact fairness and accuracy in cross-domain news and lifestyle recommender systems**, and how do these impacts differ between traditional and LLM-based systems?

RQ3: To what extent can **bias in cross-domain recommender systems** be **identified, measured, and reduced** with **LLM-based methods** compared to **traditional methods**, and **what evaluation strategies can be developed** to handle challenges like decreased diversity and hallucination?

Methodology & Experiments

Phase 1

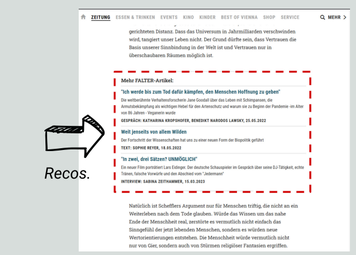
Investigate Domains Separately

Understanding of Stakeholder Perspectives

Experiments with Offline Data

Developing a Modular Recommendation Framework (for a Live System)

Phase 2



Long-Running A/B Test

Traditional Algorithms

LLM Based Algorithms

Phase 3

Evaluation Beyond Accuracy

Bias And Evaluation Strategies

Current Findings

Modular recommender system framework:
Developed a modular framework to analyze and improve news recommendation algorithms (Fig. 1)

What is a good reading recommendation?
Conducted a editor and user study
(Kolb et al. [1], Fig. 2)

Investigated news bias during COVID-19:
Through an offline dataset from derStandard
(Kolb et al. [2], Fig. 5)

Potentials of combining local knowledge and LLMs for recommender systems (Kolb et al. [3])

Several co-supervised bachelor- and master thesis projects:
Content-based recommendations (Dante et al. [4])
Group fairness in Recommender Systems (Blake et al. [5])
User roles in online news forums (Scholz et al. [6])

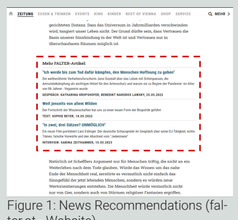


Figure 1: News Recommendations (falter.at - Website)

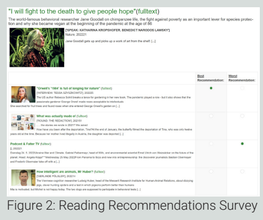


Figure 2: Reading Recommendations Survey

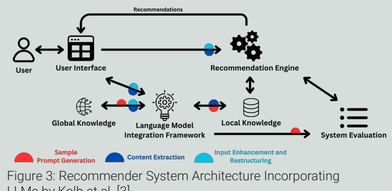


Figure 3: Recommender System Architecture Incorporating LLMs by Kolb et al. [3]

Planned Research

Long-Running A/B Test

Conducting a **long-term A/B test** on FALTER's (Fig. 4) live website using anonymized user behavior data to **investigate bias mitigation strategies and beyond-accuracy measurements**.
Evaluating various recommendation algorithms and **exploring cross-domain recommendations** (news-to-podcast, news-to-restaurant, news-to-book) by e.g. integrating FALTER's book catalog.

Evaluation

Leverage **LLMs to enhance feature extraction and support evaluation of study results**.
Assess article rankings from participants to understand perceptions of a "**good reading recommendation**".

User Study

Insights into the **multi-stakeholder environment** (editors vs. users).
Investigate **beyond-accuracy metrics.**

FALTER

Figure 4: Partner Company
(Source: https://falter.at)

DERSTANDARD

Figure 5: Partner Company
(Source: https://derstandard.at)

Expected Outcomes

A **modular recommendation framework** enabling the integration and evaluation of various recommendation algorithms, including LLM-based approaches.

Deployment within the **news ecosystem allows for real-world testing**, which overcomes the problem of solely evaluating against offline data.

New insights into which aspects of **traditional approaches within cross-domain recommender systems** can be **improved by LLMs**.

New **approaches to identify and mitigate bias**, as well as to better understand fairness (e.g., the problem of decreased diversity, hallucinations of LLMs), within the interplay of traditional methods and LLMs.

Conclusion

Aligned with the **principles of Digital Humanism (Fig. 6)**, this doctoral research aims to develop **new methods and insights to address unfairness and bias in both traditional and LLM-based cross-domain recommender systems**. While generative AI offers significant potential, it also presents **challenges such as hallucinations and reduced diversity**. By comparing novel LLM-based approaches with traditional methods, this work seeks to **enhance the quality and fairness of cross-domain recommendations**, emphasizing the importance of **keeping AI in the loop rather than the human**.

DIGHUM

Figure 6: Digital Humanism at TU Wien
(Source: https://caiml.org/dighum/)

Contact

Dipl.-Ing. Thomas E. Kolb
thomas.kolb@tuwien.ac.at
https://www.linkedin.com/in/kolb-thomas/
https://recsys-lab.at/team/thomas-kolb/

funded by:
Christian Doppler Forschungsgesellschaft

Publications

[1] Kolb, T., Nalis-Neuner, I., & Neidhardt, J. (2023). Like a Skilled DJ -an Expert Study on News Recommendations Beyond Accuracy. In Proceedings of the International Workshop on News Recommendation and Analytics co-located with the 2023 ACM Conference on Recommender Systems (RecSys 2023). CEUR-WS.org.
[2] Thomas Elmar Kolb, Irina Nalis, Mete Sertkan, and Julia Neidhardt. 2022. The Role of Bias in News Recommendation in the Perception of the Covid-19 Pandemic. FAccTRec Workshop '22, in conjunction with RecSys 2022, Seattle, WA, USA.
[3] Kolb, T. E., Wagne, A., Sertkan, M., & Neidhardt, J. (2023). Potentials of Combining Local Knowledge and LLMs for Recommender Systems. In V. W. Anelli, P. Basile, G. De Melo, F. Donini, A. Ferrara, C. Musto, F. Narducci, A. Ragone, & M. Zanker (Eds.), Proceedings of the Fifth Knowledge-aware and Conversational Recommender Systems Workshop co-located with 17th ACM Conference on Recommender Systems (RecSys 2023) (pp. 61–64). CEUR-WS.org. https://doi.org/10.34726/5334
[4] Godolja, D., Kolb, T., & Neidhardt, J. (2024). Unlocking the Potential of Content-Based Restaurant Recommender Systems. In Information and Communication Technologies in Tourism 2024 (pp. 239–244). Springer Nature Switzerland.
[5] Huebner, B., Kolb, T., & Neidhardt, J. (2024). Evaluating Group Fairness in News Recommendations: A Comparative Study of Algorithms and Metrics. In Adjunct Proceedings of the 32nd ACM Conference on User Modeling, Adaptation and Personalization (UMAP Adjunct ,24) (pp. 10). ACM.
[6] Scholz, F., Kolb, T., & Neidhardt, J. (2024). Classifying User Roles in Online News Forums: A Model for User Interaction and Behavior Analysis. In Adjunct Proceedings of the 32nd ACM Conference on User Modeling, Adaptation and Personalization (UMAP Adjunct ,24) (pp. 10). ACM.