Exploring the Impact of Data Quality on Agentic Recommender Systems

1st International Workshop on Data Quality-Aware Multimodal Recommendation

RecSys 2025 Prague

DaQuaMRec @RecSys2025

Marco Valentini
Politecnico di Bari

Antonio Ferrara
Politecnico di Bari

Tommaso Di Noia
Politecnico di Bari

Information Overload and Recommender Systems

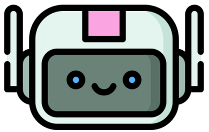
Nowadays, online platforms face the challenge of **Information Overload**, where large catalogs make the user's choice challenging, spoiling his experience

Recommender Systems have emerged as a solution to **mitigate** it

Over the years, **Recommender Systems** have undergone significant **advancements** and **innovations**, from Matrix Factorization to Deep Learning-based methods, up to **Large Language Models**, which had a disruptive impact on the recommendation thanks to their reasoning capabilities

Large Language Model-Empowered Agents for Recommendation

LLM-empowered agents develop sophisticated decision-making capabilities thanks to **reasoning** and **In-Context Learning**. They can access information from **memory** and interact with external **tools** via natural language APIs.



Within **Multi-Agent Systems**, specialized agents **collaborate** to achieve a **shared objective**; the design of these systems varies based on the agents' objectives and their mode of interaction.

The agents can seamlessly operate over **multimodal data** such as text, images, and user-item interactions; for this reason, **Agentic Recommender Systems** can be seen as **Multimodal RSs**

Agentic Recommender Systems

Examples of applications of recently proposed Agentic Recommender Systems:

- AgentCF[1]
- MACRec[2]
- RecMind[3]

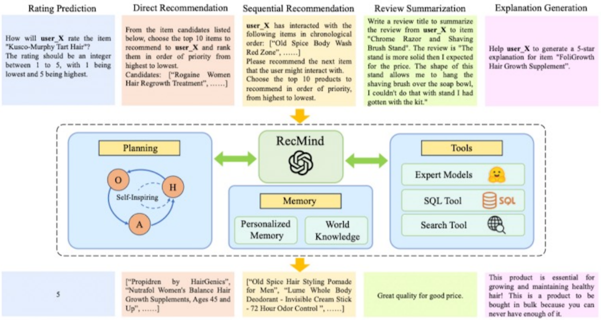


Figure 2: Here is an overview of our proposed RecMind architecture. It comprises four major components: "RecMind" is built based on ChatGPT API, "Tools" supports various API calls to retrieve knowledge from the "Memory" component, "Planning" component is in charge of thoughts generation.

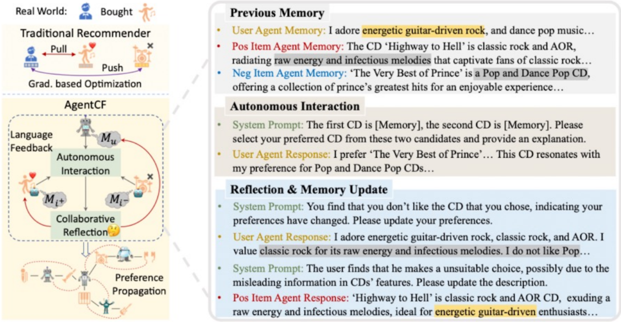
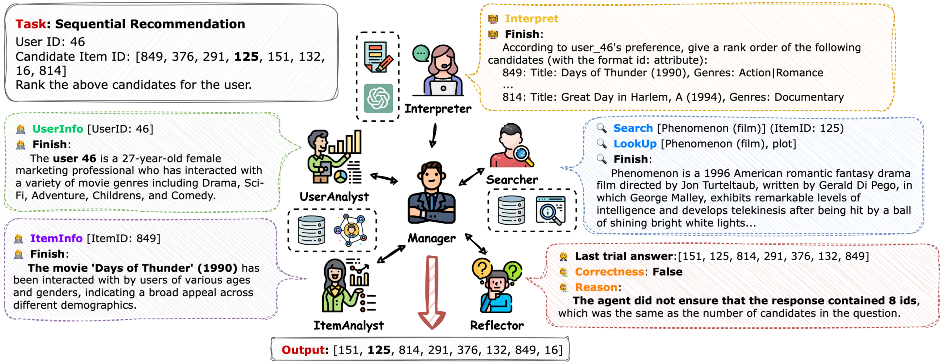


Figure 1: The overall framework of AgentCF and a case about the optimization process of agents: (1) The user and item agents are first prompted to autonomously interact. (2) These agents adjust the misconceptions in their memory, by reflecting on the disparities between their decisions and real-world interactions. In this process, the simulated preferences of user and item agents aggregate (as indicated by the highlighted content) and can propagate to other agents in subsequent interactions.



[1] Zhang, Junjie, et al. "Agentcf: Collaborative learning with autonomous language agents for recommender systems." Proceedings of the ACM Web Conference 2024..
[2] Wang, Zhefan, et al. "Macrec: A multi-agent collaboration framework for recommendation." Proceedings of the 47th International ACM SIGIR Conference on Research and Development in Information Retrieval. 2024.
[3] Wang, Yancheng, et al. "RecMind: Large Language Model Powered Agent For Recommendation." Findings of the Association for Computational Linguistics: NAACL 2024.

Input Data Quality Impact

Traditional RSs

The performance of traditional recommender systems is known to be **heavily dependent** on input data quality. Noisy user logs or biased data are known to **degrade** model performance and user satisfaction

Agentic RSs

Can agents **mitigate** poor-quality input through collaborative reasoning, or does poor-quality data **propagate** through the agent pipeline, with each step **compounding** the errors of the previous ones, **worsening** recommendation quality?

Agentic RSs Evaluation

We propose a **shift** from a model-centric to a **data-centric** perspective, looking for evaluation **criteria** that **decouple** the influence of input data quality from the system's algorithmic performance.

In multi-agent settings, the concept of **data quality** lacks both a **clear definition** and a **systematic method** for study.

We hypothesize that data quality in these systems should be re-evaluated not as a direct performance driver, but as a **minimal threshold** that is necessary to enable effective agent collaboration and reasoning

Relative Data Quality (RDQ)

We introduce **Relative Data Quality (RDQ)** for measuring performance variation across different versions of the same dataset while keeping the system architecture and prompts fixed.

We give **different quality** input data to the same system

Then, we observe the **variation** on the results

This allows us to **decouple** the contribution of data quality from the algorithm itself, providing clear insights into how **robust** the **system** is to variations in the quality of its input data.

Relative Prompt Quality (RPQ)

Prompting strategies play a central role in agentic systems, as prompts define the **roles**, domain expertise, and allowed actions of each agent. We have already observed an **improvement** when we enrich the prompt with **few-shot** examples

We propose **Relative Prompt Quality (RPQ)** to capture the impact of different prompting strategies on the system's output. The prompt becomes a **hyperparameter** that influences the performance of the system

By keeping the architecture and data fixed, RPQ quantifies the system's **sensitivity to prompt** formulation, allowing us to identify best practices for guiding agent behavior and coordination

Conclusion

This paper lays the groundwork for evaluating agentic recommender systems by highlighting the *often overlooked role* of **data** and **prompt** quality.

Our main contributions are:

- **Relative Data Quality (RDQ)**
- **Relative Prompt Quality (RPQ)**

Future work will involve applying these concepts to develop an **evaluation framework** that considers **data quality** and assesses its contribution to the performance of agentic RSs.

RecSys 2025
Prague

Thanks!

Do you have any questions?

m.valentini7@phd.studenti.poliba.it