



# Potentials of Combining Local Knowledge and LLMs for Recommender Systems

Collaborating with AI workshop @ TU Wien

Thomas Kolb, Ahmadou Wagne  
PhD Student & Project Assistant

September 9, 2024

# Definitions and Technologies: Large Language Models

Large pre-trained models based on the **transformer** architecture that can perform various **natural language processing** (NLP) tasks  
(Devlin et al., 2018)

Focus on generative models like GPT<sup>1</sup> or Llama<sup>2</sup>

Application in recommender systems: Early studies employ LLMs **directly as the recommender system** or use them to provide “**explainable**” recommendations (Gao et al., 2023; Liu et al., 2023; Cui et al., 2022)

<sup>1</sup> <https://openai.com/gpt-4> <sup>2</sup> <https://ai.meta.com/llama/>



## Definitions and Technologies: Local Knowledge

**External metadata** of items in the recommendation catalogue

- Examples: **text**, **structured data**, **images**

➡ Use LLMs to embed metadata

Perform search on local knowledge (e.g. vector store) to retrieve **factual items**

Mitigate problems of LLMs like **hallucination**

Incorporated with a **language model integration framework** (e.g. LangChain<sup>3</sup>)

<sup>3</sup><https://python.langchain.com/>



# Introduction

**Motivation:** Exploit the language understanding capabilities of large language models (LLMs) to **enhance various components** of the recommender systems **ecosystem**

- Traditional recommender systems are effective in clearly defined domains; LLMs show **contextual awareness** and good **adaptability** to unseen data (beneficial for cold-start problems)
- Include **local knowledge** to ensure factual recommendations

**But:**

- Literature is often quite abstract. How can we operationalize it?
- Is the “hype” already over, or are we just at the beginning?

# Recent Trends in LLMs and Recommender Systems

A Google Scholar search for 'LLMs in recommender systems' yields the following top results:

- Recommender Systems in the Era of Large Language Models (LLMs) (22.4.2024) (TKDE'24)
- OpenP5: An Open-Source Platform for Developing, Training, and Evaluating LLM-based Recommender Systems (11.7.2024) (SIGIR'24)
- Large Language Models meet Collaborative Filtering: An Efficient All-round LLM-based Recommender System (24.8.2024 – KDD'24)
- All Roads Lead to Rome: Unveiling the Trajectory of Recommender Systems Across the LLM Era (24.7.2024 – preprint/arxiv)

# Recent Trends in LLMs and Recommender Systems

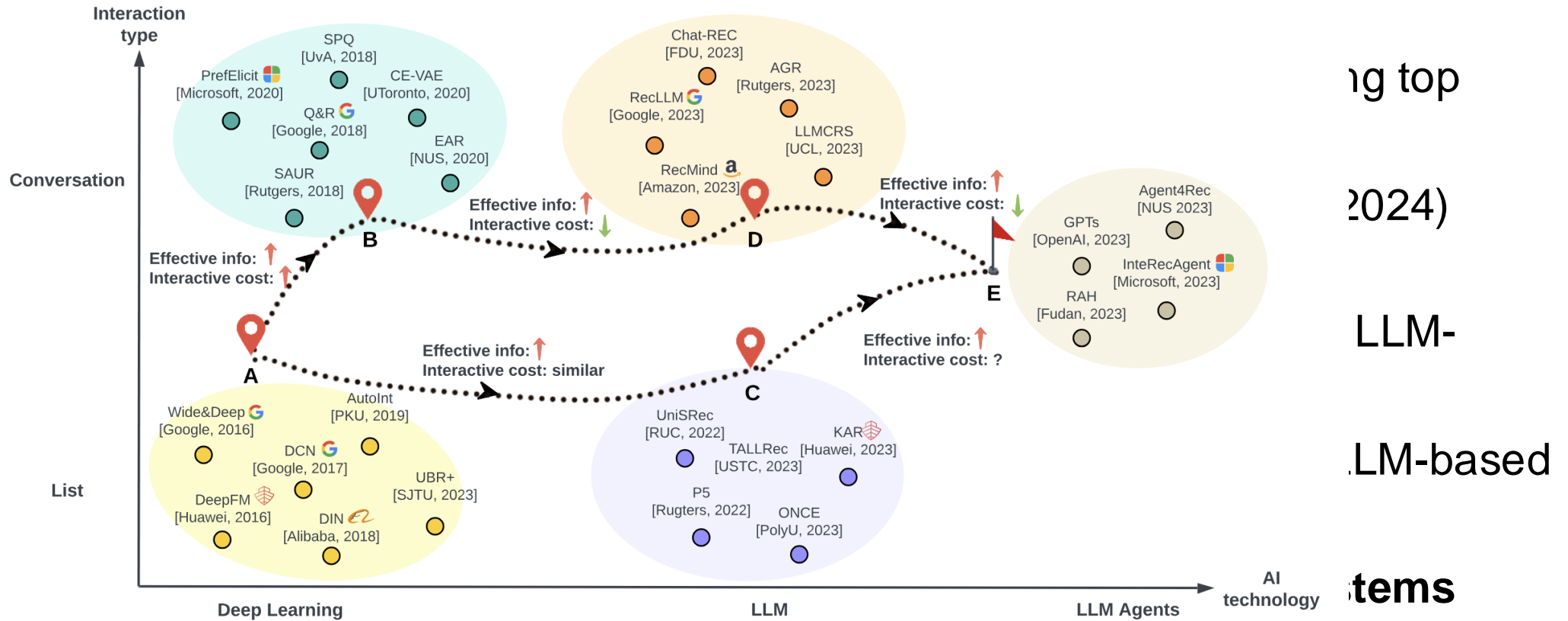
A Google Scholar search for 'LLMs in recommender systems' yields the following top results:

- Recommender Systems in the Era of Large Language Models (LLMs) (22.4.2024) (TKDE'24)
- OpenP5: An Open-Source Platform for Developing, Training, and Evaluating LLM-based Recommender Systems (11.7.2024) (SIGIR'24)
- Large Language Models meet Collaborative Filtering: An Efficient All-round LLM-based Recommender System (24.8.2024 – KDD'24)
- **All Roads Lead to Rome: Unveiling the Trajectory of Recommender Systems Across the LLM Era (24.7.2024 – preprint/arxiv)**

# Recent Trends in LLMs and Recommender Systems

A Google Scholar results:

- Recommender Systems (TKDE'24)
- OpenP5: A Large Language Model-based Recommender System
- All Roads Lead to Rome: Unveiling the Trajectory of Recommender Systems Across the LLM Era (24.7.2024 – preprint/arXiv)



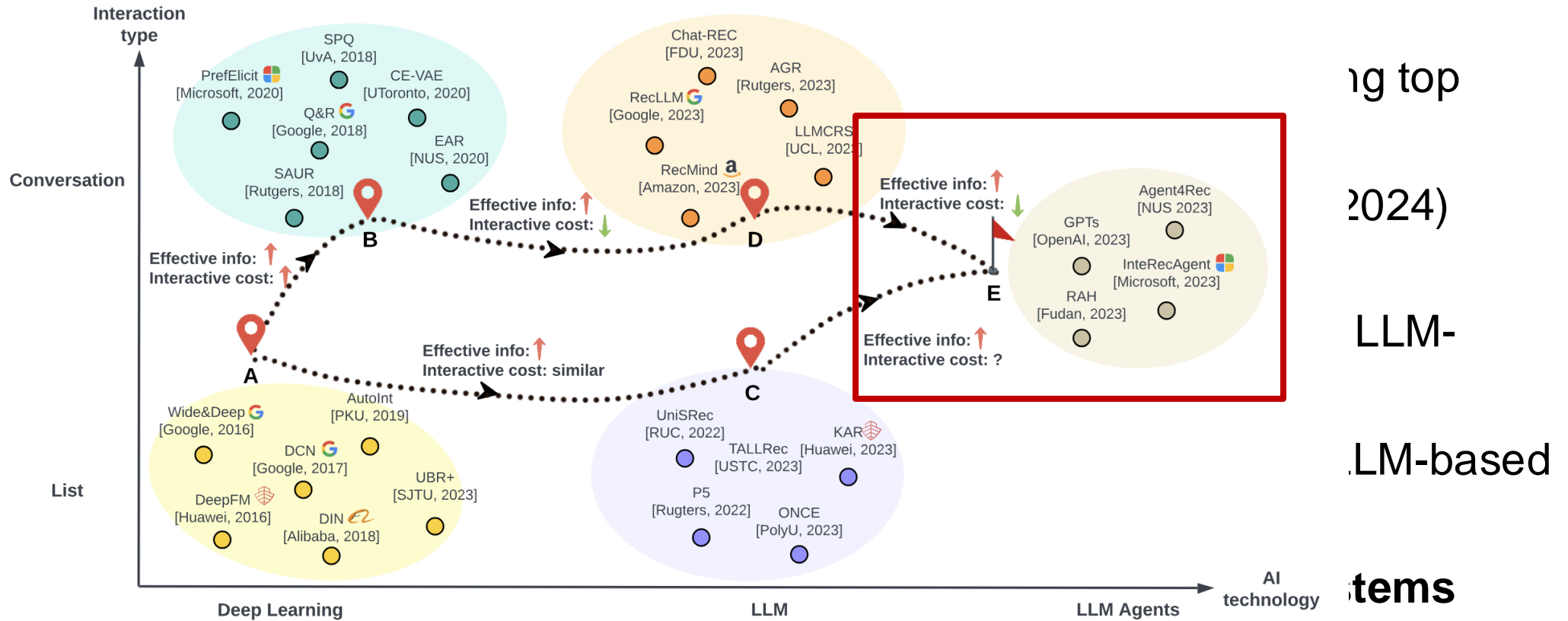
## Across the LLM Era (24.7.2024 – preprint/arXiv)

Bo Chen, Xinyi Dai, Huifeng Guo, Wei Guo, Weiwen Liu, Yong Liu, Jiarui Qin, Ruiming Tang†, Yichao Wang, Chuhan Wu, Yaxiong Wu, and Hao Zhang. 2024. All Roads Lead to Rome: Unveiling the Trajectory of Recommender Systems Across the LLM Era. 1, 1 (July 2024), 33 pages. <https://doi.org/10.1145/nnnnnnnn.nnnnnnnn> (note: all authors are researchers of Noah's Ark Lab, Huawei, China)

# Recent Trends in LLMs and Recommender Systems

A Google Scholar results:

- Recommender Systems (TKDE'24)
- OpenP5: A Large Language Model-based Recommender System
- All Roads Lead to Rome: Unveiling the Trajectory of Recommender Systems Across the LLM Era (24.7.2024 – preprint arXiv)

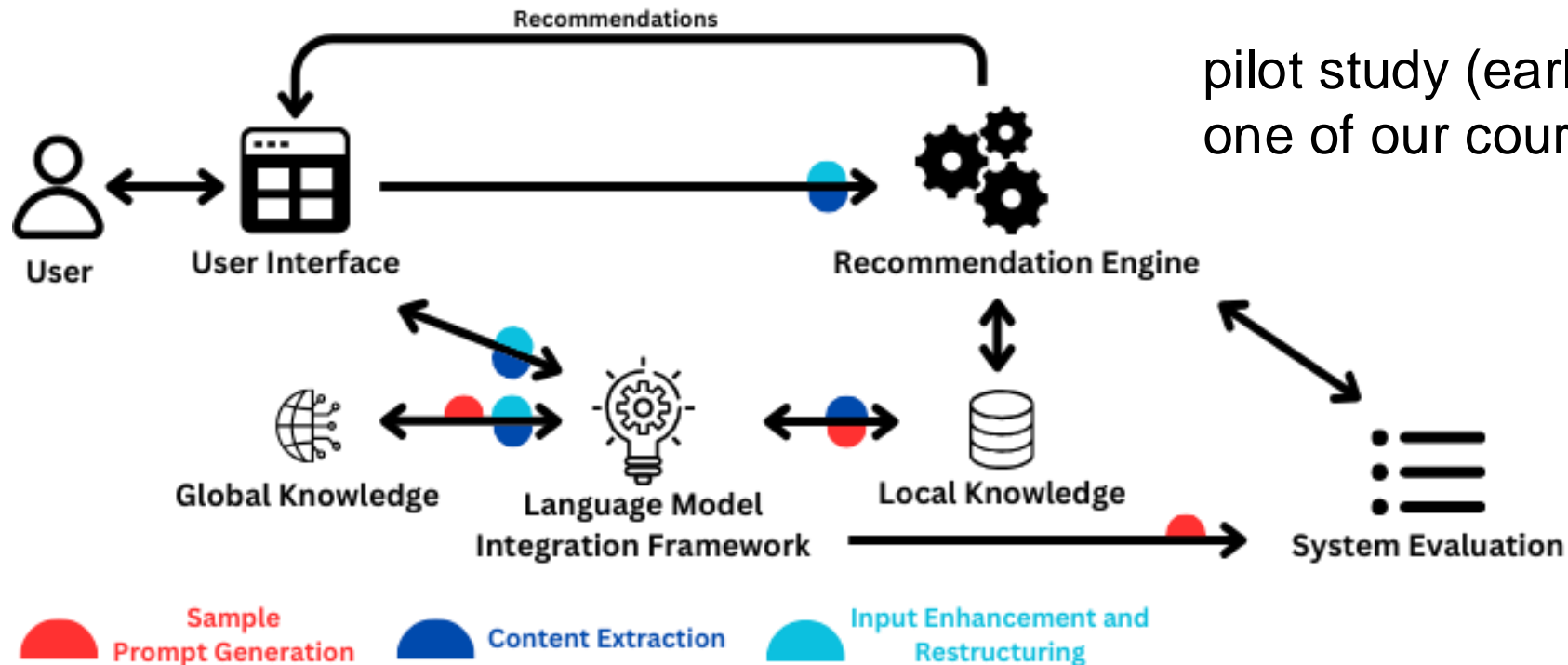


## Across the LLM Era (24.7.2024 – preprint arXiv)

Bo Chen, Xinyi Dai, Huifeng Guo, Wei Guo, Weiwen Liu, Yong Liu, Jiarui Qin, Ruiming Tang†, Yichao Wang, Chuhan Wu, Yaxiong Wu, and Hao Zhang. 2024. All Roads Lead to Rome: Unveiling the Trajectory of Recommender Systems Across the LLM Era. 1, 1 (July 2024), 33 pages. <https://doi.org/10.1145/nnnnnnnn.nnnnnnnn> (note: all authors are researchers of Noah's Ark Lab, Huawei, China)



# Combining Local Knowledge and LLMs



pilot study (early in 2023) in one of our courses

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>



## Combining Local Knowledge and LLMs

### Insights (**Input Enhancement and Restructuring**):

- Use LLMs to **restructure data** (e.g. JSON from textual resumes)
- **Fill missing data** based on the context and/or global knowledge
- Use LLMs for **prompt recommendations** based on a certain query



## Combining Local Knowledge and LLMs

Insights (**Content Extraction**, **Sample Prompt Generation**):

- **Topic extraction** (of user queries and news articles)
- Extraction of **reading motives**
- Build user **preferences during** a session
- Prompt generation (based on items) for **system evaluation**



## Practical Impact: How to make use of that? ... an example based on our partner company



- Literature is often quite abstract. How can we put it into practice?
- What is its impact on companies and their processes?
  - Latency is crucial
  - High user traffic
  - Dynamic item catalog
- Companies are often skeptical about the effect of such new technologies on their business
  - Resource demand
  - Cost vs. profit
  - Accurate processing and communication of product information critical for various stakeholders

*Geizhals (<https://geizhals.at/>) is one of the largest price-comparison platforms in the German-speaking region, with over 4.3 million monthly users and a turnover of approximately 20 million euros.*



# Practical Impact: How to make use of that? ... an example based on our partner company



The screenshot shows the Geizhals website homepage. At the top, there's a search bar with the text 'Suche ...' and a 'Hilfe' link. Below the search bar is a navigation menu with categories: Hardware, Telefon, Video, Foto & TV, Audio & HiFi, Haushalt, Drogerie, Sport & Freizeit, Baumarkt & Garten, Auto & Motorrad, Spielzeug & Modellbau, Games, Filme, Software, Büro & Schule, and Vergleichsrechner. The main content area features a 'Gaming. Ganz viel besser.' banner with a sub-header 'Alles für dein liebstes Hobby!' and a button 'Zu Games, Konsolen und Zubehör'. Below this are four product cards: Microsoft Xbox Series S - 512GB weiß (ab € 259,00), Microsoft Xbox Series X Wireless Controller shock blue (Xbox SX/Xbox One/PC) (ab € 49,90), Sony DualSense Edge Controller wireless weiß (PS5) (ab € 219,90), and Sony PlayStation 5 Slim - 1TB weiß (ab € 447,73). To the right of these cards is a section titled 'Was kann ein NAS?' featuring a Ugreen NASync DX-P4800 Plus NAS drive and a 'Jetzt lesen' button. Below the main banner, there's a horizontal menu with categories: Handys ohne Vertrag, Fernseher, Notebooks, Monitore, Smartwatches & Sportuhren, Tablets, and PCIe. The bottom section is a grid of category tiles: Hardware (Grafikkarten, Monitore, Festplatten & SSDs, Notebooks, Prozessoren (CPUs), Mainboards, Eingabegeräte, Gehäuse, Tablets, Arbeitsspeicher (RAM), Netzwerk, Luftkühlung, Netzteile & USV, Systeme, Wasserkühlung), Telefon (Handy & Smartphones, Smartwatches), Video, Foto & TV (Fernseher, Fotografie, Foto-/Videozubehör), Audio & HiFi (Kopfhörer & Headsets, HiFi-Komponenten, Professional Audio), Haushalt (Staubsaugen & Reinigen, Küchenkleingeräte, Kühlen & Gefrieren, Kochen & Backen, Waschen & Trocknen, Kaffee & Tee, Besteck & Geschirr, Geschirrspülen, Heizen & Kühlen), Drogerie (Medikation & Nahrungsergänzung, Rasur & Haarentfernung, Parfümerie, Mund- und Zahnpflege, Gesundheit, Gesicht- & Körperpflege, Haarstyler), Sport & Freizeit (Outdoor, Sportschuhe, Funktionsbekleidung, Fahrradzubehör, Fahrräder, Fahrradkomponenten, Wintersport, Rollsport, Sportuhren), Auto & Motorrad (Autoreifen & Felgen, E-Ladesysteme, Motorrad), Baumarkt & Garten (Maschinen, Gebäudeautomation & Sicherheit, Gartenmaschinen, Stromerzeugung & -speicherung, Werkstattausstattung, Werkzeugzubehör & Verbrauchsmaterial, Tierbedarf, Elektroinstallation), Büro & Schule (Drucker & Scanner, Büromöbel, Bürogeräte), and Software & Filme. There's also a 'Tarife vergleichen' button in the bottom right corner.

Geizhals (<https://geizhals.at/>) is one of the largest price-comparison platforms in the German-speaking region, with over 4.3 million monthly users and a turnover of approximately 20 million euros.



## Practical Impact: How to make use of that? ... an example based on our partner company

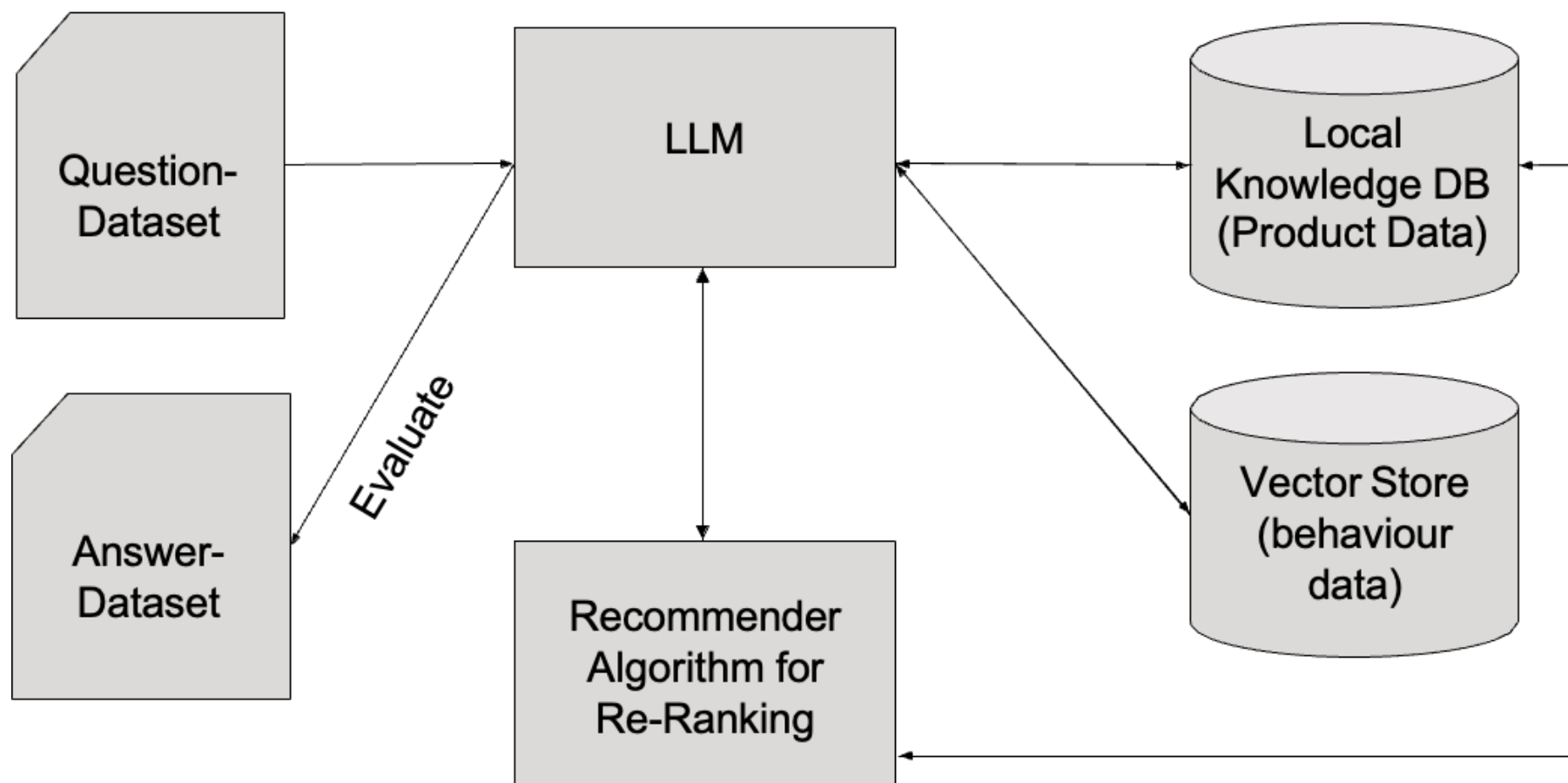


Master Thesis Michael Schmiedmayer: "The impact of using Large Language Models on the performance of Conversational Recommender Systems"

- Utilize an LLM to **produce user prompts** based on behavioural data to simulate users of a CRS
- Setup a pipeline that translates the user's demand **into a query to the local knowledge**
- Provide recommendations based on the extracted constraints
- Investigate the impact of the size (n-parameters) of an LLM among other parameters on the system performance

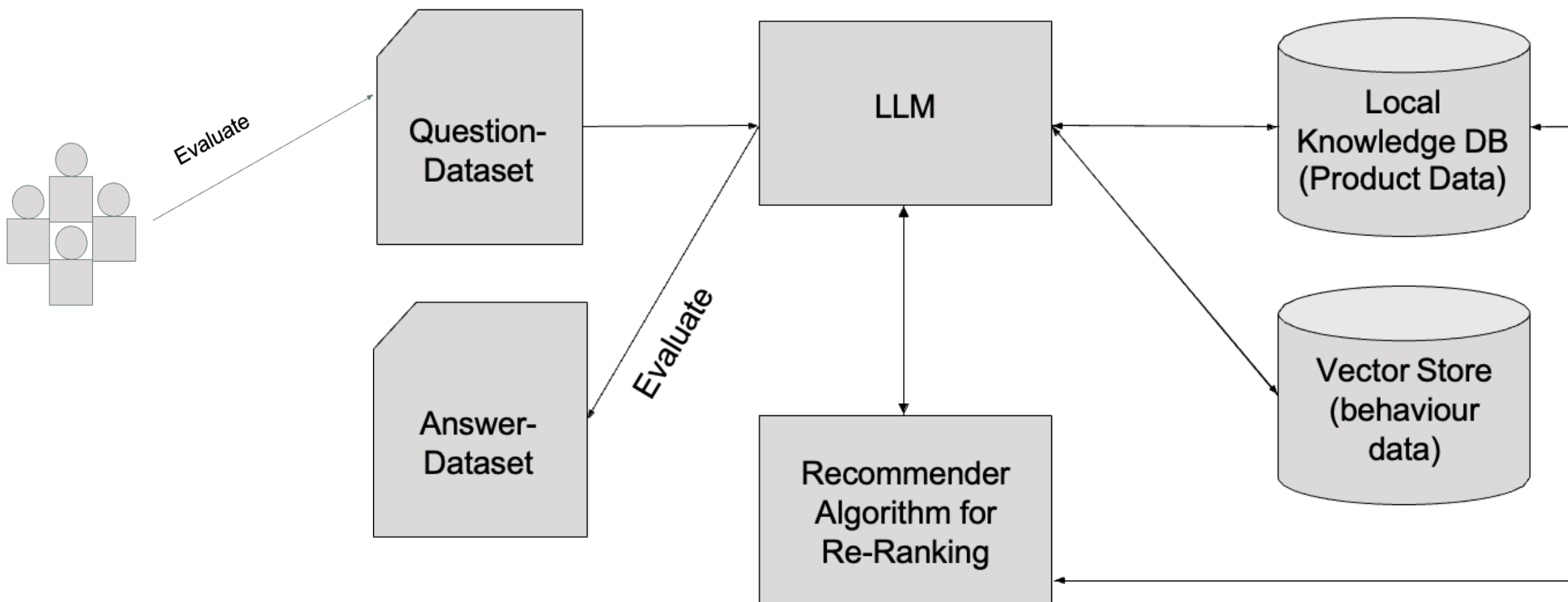


## Practical Impact: How to make use of that? ... an example based on our partner company





# Practical Impact: How to make use of that? ... an example based on our partner company





# Practical Impact: How to make use of that? ... an example based on our partner company



User study to assess the quality of synthetic prompts

- **Incorrect Number:** Numerical values in the question that do not match the input data
- **Incorrect Entity:** Incorrect entities (e.g., product names, categories) mentioned in the question
- **Missing Number:** Essential numerical information missing from the question
- **Missing Entity:** Important entities from the input data not included in the question
- **Hallucination:** Information present in the question that was not part of the input data
- **Not Checkable:** Aspects of the question that cannot be verified against the input data
- **Other:** Any other inaccuracies not covered by the above categories

Howcroft, D. M., Belz, A., Clinciu, M., Gkatzia, D., Hasan, S. A., Mahamood, S., ... & Rieser, V. (2020, December). Twenty years of confusion in human evaluation: NLG needs evaluation sheets and standardised definitions. In 13th International Conference on Natural Language Generation 2020 (pp. 169-182). Association for Computational Linguistics.



## Practical Impact: How to make use of that? ... an example based on our partner company



### System Approach

Can you recommend me a 3D-Printer under 300€?



LangChain<sup>4</sup>

**Action:** sql\_db\_list\_tables

**Action Input:** ""

**Observation:** 3m10\_19\_Schränke\_schr19, 19\_Barebones\_bare19,  
19\_Gehäuse\_geh19, 19\_Rack\_hd19rack, 19\_Rackmount\_sys19rack,  
3D\_Drucker\_dr3d, 3D\_Druckstifte\_dr3dst, 3D\_Scanner\_scan3d, ...

**Thought:** The question is asking for a 3D printer recommendation that costs less than 300 (best\_price). To answer this question, I need to query the "3D\_Drucker\_dr3d" table in the database.

<sup>4</sup><https://python.langchain.com/>



## Practical Impact: How to make use of that? ... an example based on our partner company



**Action:** sql\_db\_query

**Action Input:** SELECT \* FROM 3D\_Drucker\_dr3d WHERE best\_price < 300  
ORDER BY best\_price ASC LIMIT

**Observation:**

[('dr3d', '2023-12-12T08:40:14+01:00', 'Creality Ender 3 NEO', 142.32, ...),  
('dr3d', '2023-12-12T08:40:14+01:00', 'Creality Ender 3', 195.9, ...), ...]

**Thought:** The query returned 10 results, but the question asks for a specific recommendation. I will look at the results and provide the best option based on the given information.

**Final Answer:** The Creality Ender 3 NEO is a recommended 3D printer that costs less than 300 (best\_price).



## Recap & Summary

- How can LLMs be effectively combined with local knowledge to improve the accuracy and relevance of recommendations?
- What ethical considerations and potential biases should be taken into account when using LLMs in conjunction with local data?
- What strategies can be employed to mitigate the risks of hallucination and ensure that recommendations are factually accurate?
- How can users and items be represented in such systems?
- In what ways can LLMs help address the cold start problem in recommendation systems, and what limitations might arise?
- How can we harness the dynamic capabilities of LLMs to refine recommendations based on real-time user feedback?



DI Thomas E. Kolb  
Research Unit of Data Science



[thomas.kolb@tuwien.ac.at](mailto:thomas.kolb@tuwien.ac.at)



<https://www.linkedin.com/in/kolb-thomas/>



<https://recsys-lab.at>



DI Ahmadou Wagne  
Research Unit of Data Science



[ahmadou.wagne@tuwien.ac.at](mailto:ahmadou.wagne@tuwien.ac.at)



<https://www.linkedin.com/in/ahmadouwagne/>



<https://recsys-lab.at>