CD Lab RecSys

Christian Doppler Laboratory for Advancing the State-of-the-Art of Recommender Systems in Multi-Domain Settings

TU WIEN Informatics

© 358108785/Alex - stock.adobe.com

Opportunities of LLMs in the Research Area of Recommender Systems and Information Retrieval

Session 1

Digital Humanism Summer School 2023

Thomas E. Kolb

PreDoc Researcher & PhD Student

CDL - RecSys @ TU Wien

thomas.kolb@tuwien.ac.at

The Laboratory

CD Lab RecSys

FALTER MEDIA
FALTER VERLAG
DIE BESTEN SEITEN ÖSTERREICHS

Domain: News, E-Commerce, Culture, Lifestyle

YouKnowMeBest

Domain: Fashion

Geizhals Preisvergleich

Domain: Preference Elicitation, E-Commerce


Julia Neidhardt


Mete Sertkan


Irina Nalis-Neuner

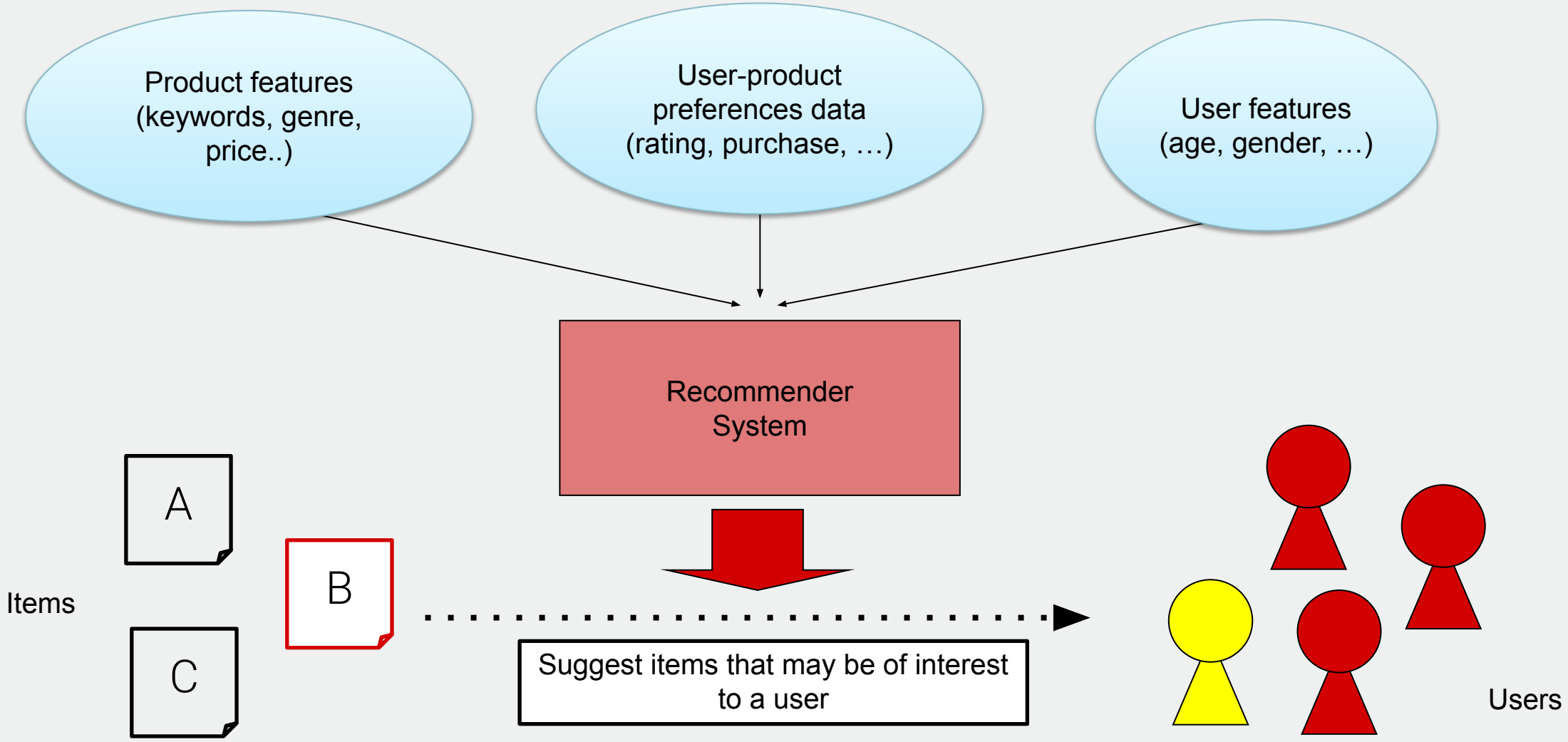

Thomas E. Kolb


Aayesha Aayesha


Ahmadou Wagne

CD Lab RecSys

Recommender Systems



CD Lab
RecSys

Recommender Systems – Social Impact

Recommender systems can make online experiences more enjoyable and efficient for users – but they

- Can limit users' exposure to diverse perspectives and ideas, creating “filter bubbles”
- Require collecting and analyzing large amounts of personal data, raising privacy concerns
- May perpetuate or amplify biases present in the data, leading to unfair recommendations
- Can give certain companies a competitive advantage and lead to market monopolies
- Can be highly engaging and lead to addiction to online platforms

Acknowledgements

Christian Doppler Research Association
Public Funding Body

Research Partner

Program of Today's Morning Sessions

Session 1 (09:00 - 10:30)

- Introduction
- Grouping
- Hands-On Session

Break (10:30 - 11:00)

Session 2 (11:00 - 12:30)

- Group Task Introduction
- Working Phase
- Presentations

Program of Today's Morning Sessions

Session 1 (09:00 - 10:30)

- Introduction
- Grouping
- Hands-On Session

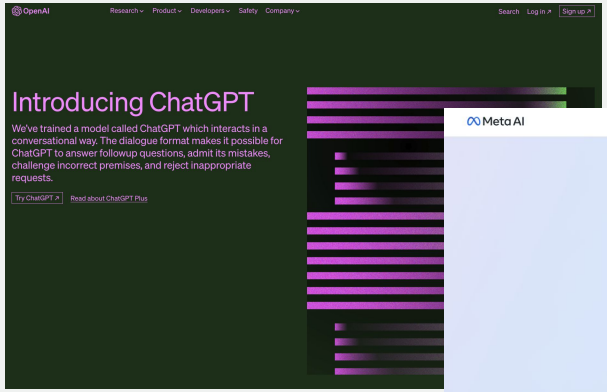
Break (10:30 - 11:00)

Session 2 (11:00 - 12:30)

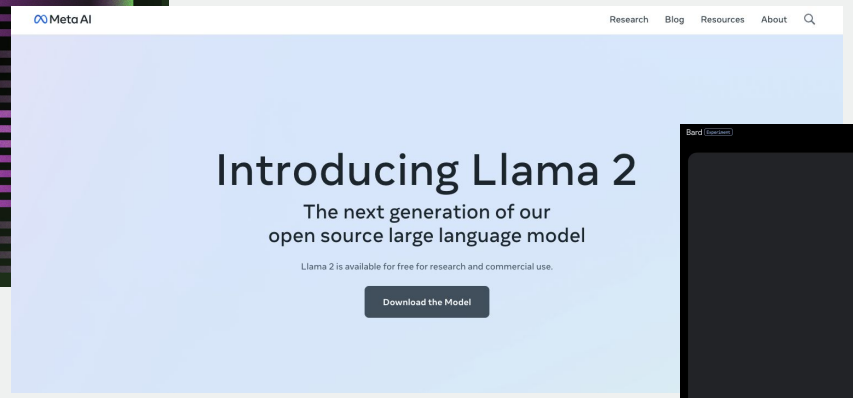
- Group Task Introduction
- Working Phase
- Presentation

Session 1
Introduction

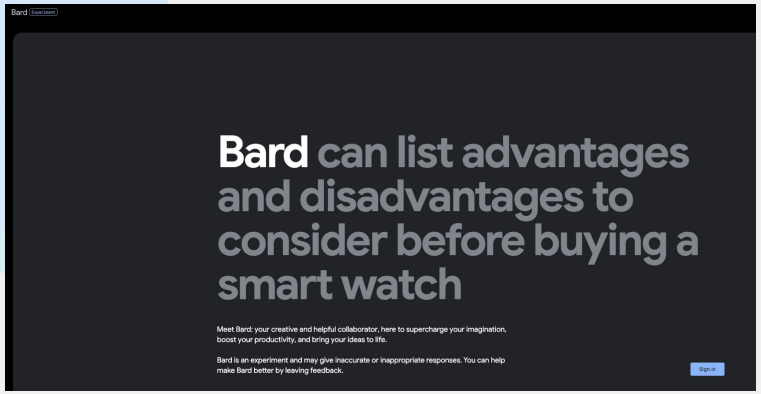
Introduction



https://openai.com/blog/chatgpt



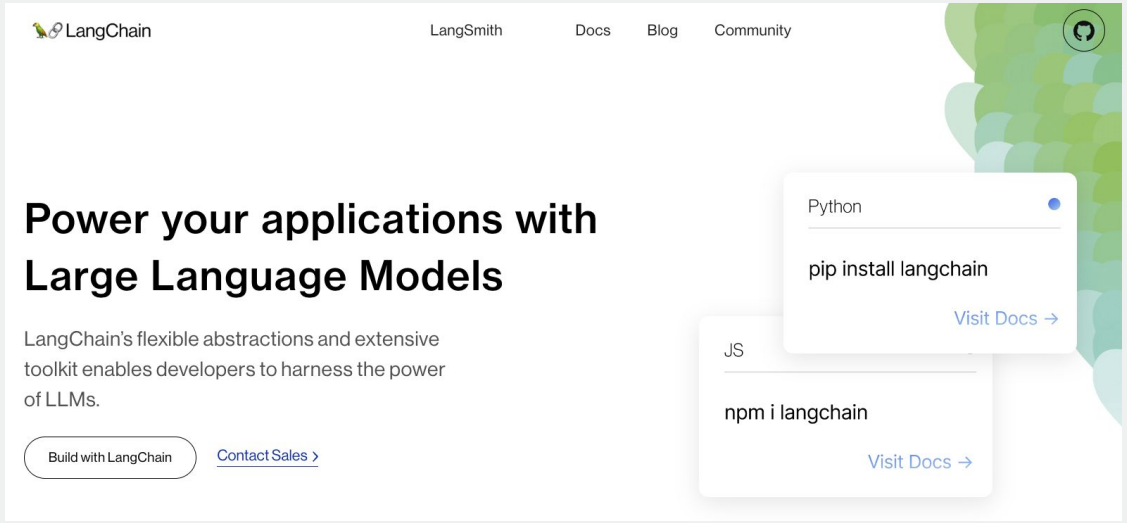
https://ai.meta.com/llama/



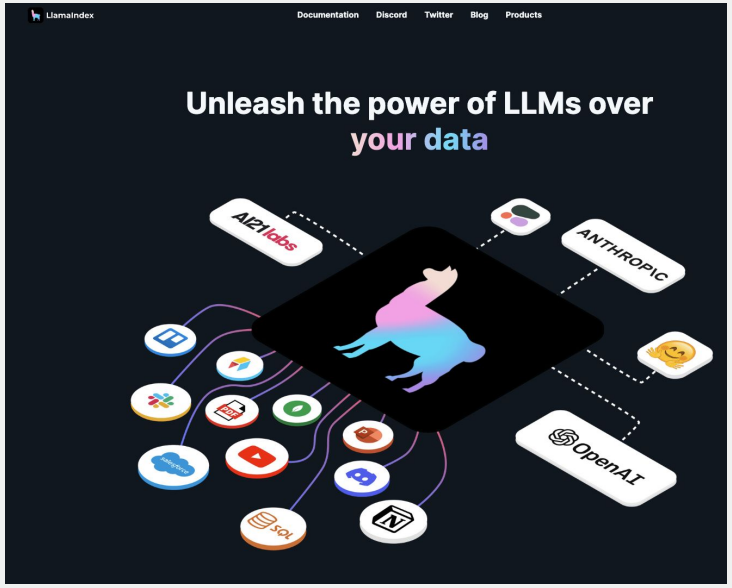
https://bard.google.com/

Many new proprietary and open-source approaches have been introduced to the public recently!

Introduction (cont.)



https://www.langchain.com/



https://www.llamaindex.ai/

The tooling and framework support have significantly increased, in addition to the various LLMs.

Why Is Everyone Interested in LLMs?

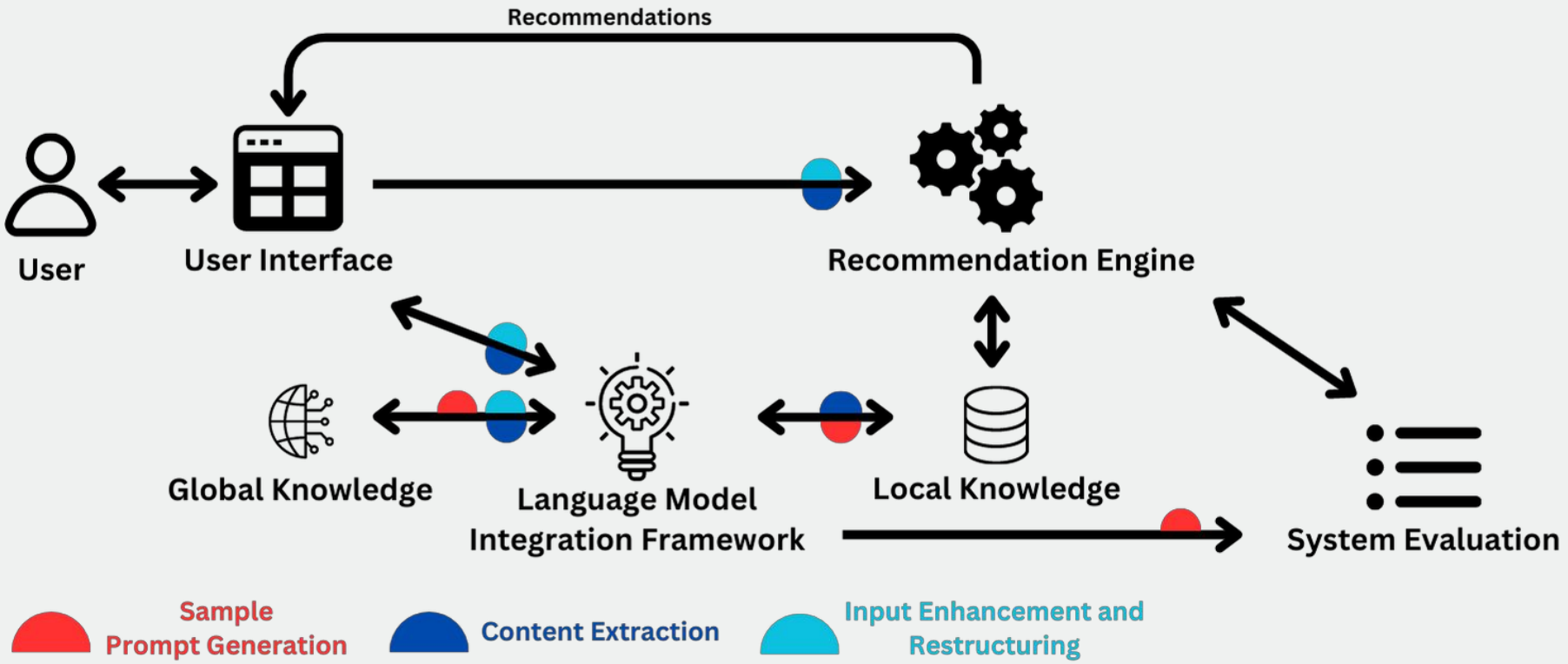
LLMs revolutionized language understanding and generation, enhancing recommendation systems …

… but prior studies used LLMs mostly for global knowledge-based zero- or few-shot recommendations.

How Can LLMs Improve Our Research Field?

- Leverage LLMs alongside local knowledge for better recommender systems.
- We see LLM-recommender integration as a powerful enhancement, not just a replacement.
- This approach complements existing methods, boosting the overall recommendation process.

System Architecture



Thomas E. Kolb, Ahmadou Wagne, Mete Sertkan, Julia Neidhardt. 5th Edition of Knowledge-aware and Conversational Recommender Systems (KaRS) @ RecSys 2023, 18th-22nd September 2023, Singapore (forthcoming)

How Does That Look Like?

Use the LangChain[1] framework to fuse the capacities of GPT with domain's information.

1) https://python.langchain.com/

Relevant Technologies:

GitHub

Streamlit

Session 1
Grouping for the Hands-On / Group Task

Grouping for the Hands-On & Group Task

Goal: Build diverse and interdisciplinary groups with a size of 5 members.

Nice to have: At least one group member should have basic programming skills (in the best case some experience with the python programming language).

Time: 15 min

As soon as your group is final, define a group leader & name and collect the group OpenAPI key for your group (at the presenter desk).

Session 1:
Hands-On Session

Hands-On Project

Please scan this qr code to get access to the project material!

https://github.com/ThomasEKolb/dighum-gpt-colab
https://github.com/ThomasEKolb/dighum-gpt

CD Lab RecSys

Hands-On Task: Goal

Build a tool which provides the user with an easy way to access information within the book “Perspectives on Digital Humanism”.

In this session, we are going to have a look at the DigHum GPT tool.

https://link.springer.com/book/10.1007/978-3-030-86144-5

CD Lab RecSys

Hands-On Task: API Usage

Each group received an OpenAI API key worth of $10,-.

Examples (per 1k tokens):

- Ada v2 (used for embeddings): $0.0001
- GPT 3.5 (prompts): 4K context $0.0015 (input) / $0.002 (output)

If you are unsure about how much budget you already used, we can check that for you!

Hands-On Task: API Usage (cont.)

Keep you budget in mind and plan your workflow accordingly.

You can look up the cost per token here:

https://openai.com/pricing

Save your embeddings to not be required to redo all of them each time you run your code!

Hands-On Task: How to Build a Chat Bot?

We are mainly going to work with Google Colab.

Please check that at least one member of your group has access to:

- Google Colab (coding)
- Github (storing your results)
- Streamlit (creating a web application / optional but recommended)

Coding Session

Summary

Introduction to the practical application of:

- LLMs
- Frameworks

Coding session:

- Technologies (Google Colab, GitHub, Streamlit)
- Creation of a DigHum GPT tool

CD Lab RecSys
Dipl.-Ing. Thomas E. Kolb
Research Unit of Data Science
thomas.kolb@tuwien.ac.at
https://recsys-lab.at

Break

Opportunities of LLMs in the Research Area of Recommender Systems and Information Retrieval

Session 2

Digital Humanism Summer School 2023

Ahmadou Wagne

PreDoc Researcher & PhD Student

CDL - RecSys @ TU Wien

ahmadou.wagne@tuwien.ac.at

Program of Today's Morning Sessions

Session 1 (09:00 - 10:30)

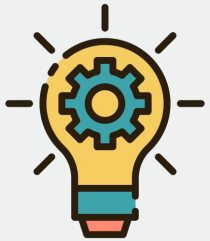
Break (10:30 - 11:00)

Session 2 (11:00 - 12:30)

- Group Task Introduction
- Working Phase
- Presentations

Session 2:
Group Task

Build your own GPT



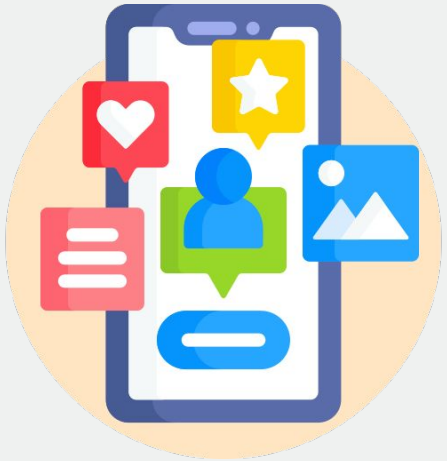
Goal: Build upon the hands-on example from the previous session and create your own application and assess its capabilities by creating and evaluating **5 related prompts** from a possible users perspective

Resources: Choose a digital humanities domain or a use case that promises societal benefits and collect textual resources necessary (possible formats: PDF, XML, txt etc.) and load them into your system

Presentation: Prepare a short presentation (**max. 3 minutes**) that introduces your domain, evaluates your queries and a reflection on potential benefits and threats to society that come with the implementation of such systems

Time: until 11:45

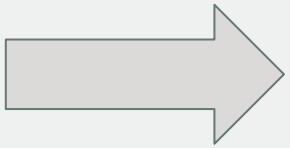
Possible Domains



This presentation used Icons by Freepik, from www.flaticon.com

Resources: Heterogeneous data and prompt templating

- https://python.langchain.com/docs/get_started/introduction.html
- https://python.langchain.com/docs/modules/model_io/prompts/prompt_templates/
- https://python.langchain.com/docs/modules/data_connection/document_loaders/



Back to the notebook

Presentations

- Short introduction into your topic/data
- How did you query the system?
- Qualitative evaluation of responses
- Reflection on potentials and dangers of the implementation of such systems

Submission

Before the presentations start, please submit your code, slides, and a link to your data via mail!

recsys-lab@ec.tuwien.ac.at

Floor plan

Rooms FAVEG B/C

CD Lab RecSys

Floor plan

Room FAV01 A

Floor plan

Room HF0311

Questions?

CD Lab RecSys
Dipl.-Ing. Ahmadou Wagne
Research Unit of Data Science
ahmadou.wagne@tuwien.ac.at
https://recsys-lab.at