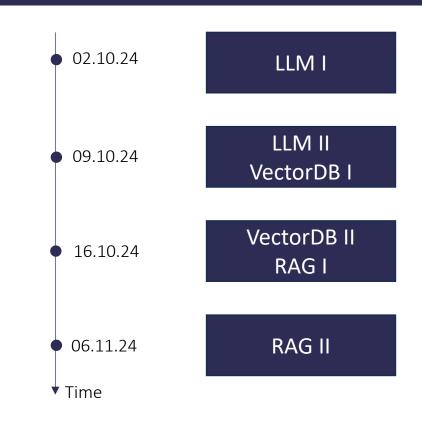
Course Schedule



Overview





LLM Part I

02.10.2024

- What are Large Language Models?
- How are they trained?
- Which providers can be used?
- System Setup (VS Code, Python 3.12, API keys, etc.)
- How can you interact with an LLM via Python?
- What are system-user-Al messages?
- How to use prompt templates?
- How to use chains?



LLM Part II / Vector DB Part I

09.10.2024

- What is a Vector Database?
- What is the difference between a classical and a vector DB?
- How does the data ingestion process work?
- How can data be loaded?
- How are chunks created?
- What are embeddings?
- How do you create chunk embeddings?
- Which database providers can be chosen? How do you select a specific provider?
- How do you store data in a vector database?



Vector DB Part II / RAG Part I

16.10.2024

- What is RAG?
- How does it work?
- Implementing RAG in our custom project.
- If time allows, how to make it part of a web application.

RAG Part II

06.11.2024

- Technical questions of workshop participants related to details of their own RAG system implementation.
- Basic techniques for tuning and managing the quality of RAG systems in the context of selected domains.
- Initiating discussions about possible further joint research projects.



Buffer for technical workshops

13.11.2024

- Technical questions of workshop participants based on identified issues related to details of their own RAG system implementation.
- Security, regulatory compliance, and scalability of RAG systems.
- Short demos of participants' own RAG systems.
- Buffer for technical questions and workshops based on identified issues.



Buffer for technical workshops

20.11.2024

- Technical questions of workshop participants based on identified issues related to details of their own RAG system implementation.
- Security, regulatory compliance, and scalability of RAG systems.
- Short demos of participants' own RAG systems.
- Buffer for technical questions and workshops based on identified issues.



Buffer for technical workshops

27.11.2024

 Short demos of participants' RAG systems. Buffer for technical questions about building RAG systems. (Discussion)



Buffer for technical workshops

04.12.2024

Short demos of participants' RAG systems. Buffer for technical questions about building RAG systems. (Discussion)

