MLOps for Small Language Model Applications

Winter

(in German: MLOps for Small Language Model Applications)

Module-ID: FIN-INF-120515

Abbreviation	Credit Points	Semester	Term	Duration	Language	Level
LLMOps	9	Winter	starting at 1.	1 Semester	english	Master

Link:	LSF
Limited accessability:	true
Responsibility:	Prof. Dr. Klaus Turowski, Dr.-Ing. Abdulrahman Nahhas
Lecturer:	Prof. Dr. Klaus Turowski, Dr.-Ing. Abdulrahman Nahhas
Classes:	Vorlesung + Übung + Tutorial
Applicability in curriculum:	- M.Sc. INF: Informatik - M.Sc. INGINF: Informatik - M.Sc. WIF: Wirtschaftsinformatik - M.Sc. WIF: Informatik - M.Sc. DKE: Applied Data Science - M.Sc. DE: Methoden der Informatik - M.Sc. VC: Computer Science

Intended learning outcomes:
Students ...

- know and understand specialized subject content that expands and deepens Large Language Models (LLM) and machine learning development and operation.
- can independently analyze problems and identify alternative solutions.

- know how to work with large language models following proper machine learning development and operation practices.
- are experienced in the theory and handling of container-based virtualization technologies such as Docker.
- know how to use higher programming languages (Python).
- know how to use various configuration and automation tools and languages to automate the development and operation of LLM solutions.
- have practical and theoretical knowledge of testing LLM-solutions, particularly in relation to essential DevOps concepts such as Continuous Integration and Continuous Delivery, LLMOps, testing, and system landscape.
- have the ability to design, create, test, commission, and support LLM solutions for business problems.

Content:
The generalization and standardization of the Large Language Model Operations (LLMOps) life cycle is crucial for the effective adoption and management of Large Language Models (LLMs) in a business context. In many cases, the lifecycle of an LLM solution does not end with the delivery an initial solution; rather, subsequent model improvement, operation, monitoring, and maintenance have become essential parts of this process. At this point, sophisticated paradigms and methods are required to facilitate the continuous development and operation of these systems while preventing errors, failures, and other disruptions. The course is designed to teach the fundamentals, as well as provide initial practical experience in the development of LLM solutions and technologies, in conjunction with dedicated cloud-native technologies and an emphasis on small LLMs (SML) that are edge-suitable, may be operated on premise, or are ressource-efficient. In addition to the theoretical basics, essential concepts and technologies are discussed and applied that enable the continuous improvement, integration, delivery, and testing of LLM solutions for business problems.

Workload:
Attendance:
21 h lecture / 63 h exercise Independent work: 207 h

1. Preparation and follow-up of lecture & exercise
2. Development of an LLM-based sotution following LLMLOps practicies to address a business problems

9 x30h (63 h attendance time + 207 h independent work)

Pre-examination requirements:	Type of examination:	Teaching method / lecture hours per week (SWS):
	Report	Vorlesung (2 SWS) Übung (2 SWS)

Prerequisites according to examination regulations:	Recommended prerequisites:

1. VLBA: Systemarchitectures or 2. VLBA Cloud DevOps Technologies or 3. Version control systems, DevOps practicies (CI/CD), Containerization technologies, and python.	1. VLBA: Systemarchitectures or 2. VLBA Cloud DevOps Technologies or 3. Version control systems, DevOps practicies (CI/CD), Containerization technologies, and python.

Media:	**Literature:**

Comments:

Revision #1
Created 2026-11-19 09:31:29 UTC by Fakultät für Informatik
Updated 2026-11-19 09:31:29 UTC by Fakultät für Informatik