

## Session 3b: Machine Learning Operations - MLOps

### Short presentation and topics

In this workshop, participants will explore the data management aspects of the Retrieval-Augmented-Generation (RAG) framework as applied to Large Language Models. The primary objective is to convert a machine learning experiment provided in a Jupyter Notebook into a production-ready application by integrating a Data Version Control (DVC) pipeline for efficient data management. This workshop aims to provide participants with the technical skills needed to manage the data lifecycle of Machine Learning models, promote collaboration among teams, and automate operational processes.

### Planning and structure:

#### Goal:

The goal of the workshop is to provide participants with tools and skills permitting to ensure the reproducibility and quality of a model and to help the team collaborate in its development, in order to move it out of the experimental context.

#### Content:

1. Initial setup
2. Exploration of the RAG experiment
3. Transformation of the experiment to reproduce and evaluate the quality of the model
4. Effective collaboration to build a reliable and functional model over time

#### Needs :

Personal laptop with WSL2 / Linux Ubuntu installed and ready to use.

#### Detailed schedule:

13h00-14h00: Initial setup and exploration of the Jupyter Notebook

14h00-17h00: Transformation and integration of the DVC pipeline

17h00-17h30: Time for questions and comments

### Speakers and committee

Last Name	First Name	Institution	e-mail address
Chapuis	Bertil	HEIG-VD	<a href="mailto:bertil.chapuis@heig-vd.ch">bertil.chapuis@heig-vd.ch</a>
Marquis	Rémy	HEIG-VD	<a href="mailto:remy.marquis@heig-vd.ch">remy.marquis@heig-vd.ch</a>