AI-DAYS@HES-SO 2025 – GENEVA & LAUSANNE – JANUARY 27- JANUARY 29, 2025

WORKSHOP DAY JANUARY 27

Session 3a: RAG- Unveiling the Power of Retrieval-Augmented Generation

The Retrieval-Augmented Generation (RAG) framework represents a groundbreaking approach that seamlessly integrates two fundamental techniques, retrieval and generation, within a large language model (LLM). The result is the generation of more context-aware and informative responses, making RAG a valuable tool for companies with extensive documentation but lacking an efficient means to access specific information.

This workshop endeavors to provide a comprehensive understanding of the RAG technology, emphasizing its applications and advantages. Through a technical introduction accompanied by concrete examples, participants will gain insights into how RAG can be effectively employed to address challenges related to information retrieval and contextual generation. The workshop will also facilitate discussions on the practical implementation of RAG in real-world scenarios, exploring its potential in enhancing knowledge management systems.

Furthermore, the workshop will delve into the realm of self-hosted Large Language Models (LLMs), shedding light on the importance of data privacy and security in the deployment of generative AI technologies. Participants will be equipped with knowledge about the intricacies of hosting LLM models independently.

By the conclusion of the workshop, participants will possess the skills to proficiently interact with a LLM, querying it about the contents of its associated documents. The overarching goal is to empower individuals with the expertise.

This workshop endeavors to provide a comprehensive understanding of the RAG technology, emphasizing its applications and advantages.

Content and schedule: 8h30-12h

	Topics
8h30-9h00	Technical background
9h00-9h45	RAG overview and Self Hosting LLM with Ollama
9h45-10h30	Hands on in Google collab - part 1
10h30-10h45	Break
10h45-11h30	Hands on in Google collab-part 2
11h30-12h00	Time for discussion

Needs :

Personnal laptop with internet connexion

Workshop speakers and committee

Last Name	First Name	Institution	e-mail address
Guerne	Jonathan	HE-Arc	jonathan.guerne@he-arc.ch
Marques Reis	Henrique	HE-Arc	henrique.marquesreis@he-arc.ch
Donzé	Célien	HEIA-FR	celien.donze@hefr.ch

swiss ai center

Haute école d'ingénierie et d'architecture Fribourg
Hochschule für Technik und Architektur Freiburg

HEIG-VD Haute école d'ingénierie et de gestion du canton de Vaud

arc ingénierie

hepia Haute école du paysage, d'ingénierie et d'architecture de Genève

Hes·so Valais Wallis

Swiss AI Center for SMEs

CSIA-PME – Centre Suisse d'Intelligence Artificielle pour les PMEs