



UNIVERSITÀ  
DI PAVIA

Ph.D. Program in Electronics, Computer Science and Electrical Engineering

SEMINAR

# Beyond ChatGPT: RAG and Fine-Tuning

*Emanuele Fabbiani*

Tuesday, June 4 – H. 4:00 pm

Magenta Seminar Room, D Floor

Department of Electrical, Computer and Biomedical Engineering

**Abstract:** In most real-world applications, ChatGPT alone is insufficient. Businesses seek to utilize their own private documents to obtain factually accurate answers. Over the past year, two techniques have emerged to address this issue. Retrieval Augmented Generation (RAG) employs text embedding to identify relevant snippets and incorporate them into a prompt for a Large Language Model (LLM) to expand upon. Conversely, fine-tuning involves updating the weights of the LLM with training episodes based on specific documents. Since training LLMs is notoriously costly, fine-tuning often incorporates advanced methods such as low-rank adaptation and quantization. This lecture will delve into both RAG and fine-tuning, discussing the latest techniques for achieving optimal results. We will examine the pros and cons of each technique and discuss real-world applications for both.

**Bio:** Engineer, researcher, entrepreneur. Emanuele earned his PhD in AI by researching time series forecasting in the energy field. He published 8 papers in international journals, presented and organized tracks and workshops at international conferences, including AMLD Lausanne, ODSC London, WeAreDevelopers Berlin, PyData Berlin, PyCon Florence, and lectured in Italy, Switzerland, and Poland. He was a guest researcher at EPFL Lausanne, and he's now the Head of AI at xstream, where he solves business problems with AI. In xstream, his team designed, built, and deployed more than 15 high-profile AI systems, dealing with demand forecasting, credit rating, marketing media mix, and content generation.

## Organizers

Prof. Marco Piastra  
Prof. Marco Porta

## Ph.D. Coordinator

Prof. Ilaria Cristiani