

## INFN Cloud Kubeflow as a Platform (KaaP) and a ChatBot use case

*Thursday, 29 May 2025 12:15 (15 minutes)*

Kubeflow is an open-source Machine Learning (ML) platform built on Kubernetes providing a comprehensive set of tools to manage every step of the ML lifecycle.

With Kubeflow, you can design, build, deploy and maintain an ML solution, therefore managing the end-to-end ML lifecycle.

INFN Cloud offers a “Kubernetes Cluster” PaaS service, we aim to enrich this service with an additional option to configure and deploy a Kubeflow instance on top of it, adding therefore “Kubeflow as a Platform” - **KaaP** - to the INFN Cloud portfolio.

In this presentation we provide a general introduction to Kubeflow, then we present a use case that we are developing within the INFN Information Systems Department (DSI): a fully hosted QA system - **ChatBot** - powered by a Large Language Model (LLM) that assists users with questions about the INFN LibroFirma application.

The ChatBot is built using KaaP components:

- Kubeflow Notebooks: JupyterLab instances to develop and test python code;
- Kubeflow Pipelines: a multi-stage pipeline that periodically fetches and indexes new tickets from LibroFirma ServiceDesk to keep the ChatBot knowledge base up to date;
- KServe: model serving platform that exposes text embeddings and chat completion endpoints;
- Kotaemon: an open-source project to implement the QA system.

The LibroFirma use-case, developed on top of KaaP, provides a reference implementation that can be customized to implement a QA system on any knowledge base.

**Primary author:** GATTARI, Mauro (Istituto Nazionale di Fisica Nucleare)

**Presenter:** GATTARI, Mauro (Istituto Nazionale di Fisica Nucleare)

**Session Classification:** Intelligenza artificiale

**Track Classification:** Intelligenza Artificiale