Open-source AI: Frameworks, Tools

Introduction:
Artificial Intelligence is rapidly being democratized, and Free and Open Source Software (FOSS) is the engine driving this revolution. This talk will be a deep dive into the practical aspects of building, training, and deploying Machine Learning (ML) solutions entirely within the open-source ecosystem.

Our goal today is to examine the entire lifecycle of an ML project, from data preparation to production, using only open source components. This is the **Open MLOps pipeline.**

The Open-Source AI Foundation (Frameworks)

- Deep Learning Frameworks:
    - PyTorch
    - TensorFlow/Keras

- Traditional ML & Scientific Computing:
    - Scikit-learn
    - NumPy/SciPy/Pandas

Open-Source AI Tools

- The Model Hub:
    - Hugging Face Transformers: How to leverage and fine-tune pre-trained models (LLMs, Vision, etc.) using the open Hugging Face ecosystem (models, datasets, spaces).

- Experiment Tracking and MLOps:
    - MLflow: Tracking experiments, managing model versions, and packaging models for deployment. Focus on the four core components: Tracking, Projects, Models, and Registry.
    - DVC (Data Version Control): Managing large datasets and ML models as code (Git-style versioning) to ensure reproducibility.

- Data Annotation & Preparation: Mention of tools like Label Studio for collaborative, open-source data labeling.