



Evangelos Dellis, Cahya Wirawan
Comprehensive Nuclear-Test-Ban Treaty Organization (CTBTO)

P3.5-569

- We built an **end-to-end LLM engineering platform** that can be deployed in our GPU infrastructure
- You will learn that the LLM platform supports:
 - Popular open-source tools for inference, fine-tuning, orchestration
 - LLM application observability (traces & logs)
 - LLM evaluations like LLM-as-a-Judge
 - DeepEval → RAG metrics, relevancy, hallucination, safety
- We explore tools for LLM **fine-tuning** and **creating datasets** using CTBTO data:
 - How we implemented a **training system** using open-source tools
 - What is LLM fine-tuning?
 - Which method to use for fine-tuning?
 - How to create question/answer pairs (datasets)?
- Use case: **Paper Reviewer**
 - Architecture of the system
 - Dataset creation
 - Data processing using local GPUs

