Federated Learning Simulation Engine

Borja Arroyo Galende (UPM), Juan Mata Naranjo (Cineca)
borja.arroyog@upm.es, j.matanaranjo@cineca.it
Structure

• What is federated learning (FL)?
• How does FL work?
• Why FL?
• Proposed simulation tool
• Conclusions
What is FL?
• Is a learning algorithm that allows to train models on decentralized data.
• Allows to train models without the exchange of private data.
• Involves two different entities:
  • Central node stores the global model and acts as aggregator.
  • Data nodes store local copies of the global model: local learners.

Nvidia blog: What is federated learning?
How does FL work?
Runs iteratively such that:

1. The local learners get trained on local data.
2. The aggregator combines the updates received from all data nodes and broadcasts a new version of the global model.
Why FL?
Slide #3.1

- Complies with data protection regulations regarding sensitive data.
- Supports novel approaches → Deep Learning.
- Can leverage on further privacy mechanisms. E.g. SMPC and DP
• Complies with data protection regulations regarding sensitive data.
• Supports novel approaches → Deep Learning.
• Can leverage on further privacy mechanisms. E.g. SMPC and DP
• But it comes with a cost:
  • Leap in complexity w.r.t. traditional ML.
  • In general, humble hardware resources in data nodes. E.g. phones.
  • Unpredictable behaviour in production scenarios.
Proposed simulation tool
Slide #4.1

- **Features:**
  - Allows for the definition of use cases
  - These use cases can be defined by the end user
  - They consist of a set of scenarios
  - A model can be simulated over meaningful data for a specific task
  - FLOps → level 2 of maturity. Automatic training and evaluation + automatic verification of the models
Slide #4.2

**Benefits:**
- Validate that a FL model runs error free prior to production scenario
- Optimizes resources usage in the production environment
- Gives a sense of the performance of the FL model over use cases
- Runs on a multicontainer setup
• Offers a tool that can be used both standalone or embedded within a larger architecture, it can be deployed anywhere and allows for concurrent, scalable, and highly available V&V assessment support for FL models

• Supports AI practitioners in the process of integrating FL driven model designs and to grasp model performance prior to production scenarios, allowing for a boost in trustworthiness towards ethical AI
Thank You