ART-IE Federated Learning (AI)

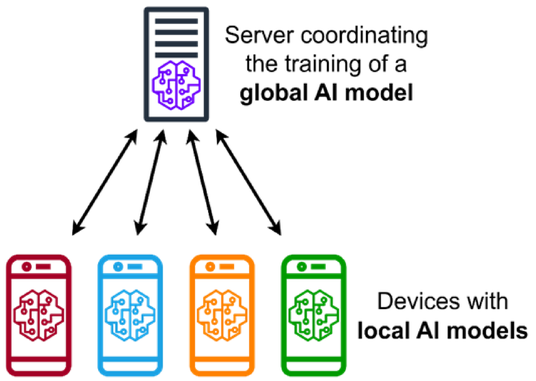
Group 1 RB02 – Meda, Huyen, Mario, Dimitar, Boris and Hristo
Advanced Software

Interreg Vlaanderen-Nederland
Gefinancierd door de Europese Unie

Abstract

Our project focuses on Federated Learning—an AI approach that trains models across multiple devices or organizations without sharing raw data. This method ensures data privacy and compliance with regulations like GDPR. We designed a system that allows decentralized training, with a central server aggregating model updates.

The result is a secure, scalable AI solution suitable for sensitive sectors like healthcare and finance.



Introduction

Sharing data across organizations or devices can lead to privacy concerns, especially when dealing with sensitive information like medical or financial records. Federated Learning provides a secure alternative by keeping data on local devices and training models at the source.

Instead of transferring raw data, each participant trains a model locally and sends only the updates to a central server. These updates are aggregated to form a global model, allowing collaboration without compromising privacy.

This approach enhances security and supports compliance with regulations such as GDPR.

Methods and Materials

This section outlines the key questions and tools that guided the development of our federated learning platform.

Main Question:

- How to design a federated learning platform to enable employees in enterprise companies to train AI models with ensuring data privacy?

To explore and address this question, we followed the **DOT Framework**, which helped us structure our development process through focused sub-questions, such as:

- How can the Art-IE FL Lab platform be architected to balance performance and scalability for federated learning Applications?
- How can the federated learning platform efficiently scale to support thousands of employees or devices within large enterprise environments?
- How can the system monitor and adjust resource usage (CPU, GPU, network) dynamically to maintain performance while respecting enterprise resource constraints?
- How can we ensure privacy-preserving training while aligning with enterprise compliance requirements (e.g., GDPR, HIPAA, ISO 27001)?

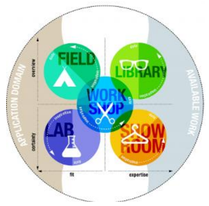


Image 1: DOT Framework

Image 2: Azure

Image 3: C#

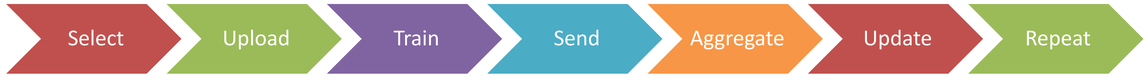
Image 4: TensorFlow

Technologies and Tools
To implement our solution and address these sub-questions, we used:

- **C#** – For developing the client-side components.
- **Azure** – To deploy and manage the centralized aggregation server and supporting infrastructure.
- **TensorFlow** – A versatile open-source library that simplifies the process of training and deploying machine learning models at scale.

Results

Our federated learning platform enables collaborative model training across multiple clients while keeping all data local and secure. The process is designed to ensure privacy, efficiency, and scalability in enterprise environments.
The training follows a clear step-by-step cycle:



Each client selects a model, uploads a local dataset, and performs training on their own device. The trained model updates are then sent to the server, which aggregates them and sends back an improved global model. This process repeats over several rounds. See Image 5.

Conclusions

This study presents a federated learning platform designed for enterprise use, focusing on scalability, performance, privacy, and regulatory compliance. Key features include a modular microservices architecture, performance management through auto-scaling and telemetry, enhanced privacy with techniques like SMPC and differential privacy, and built-in compliance with GDPR, HIPAA, and ISO standards.

The result is a scalable, privacy-preserving, and legally robust system for collaborative AI model training in enterprise environments.

References

- Kairouz, P., McMahan, H. B., et al. (2021). *Advances and Open Problems in Federated Learning*. Foundations and Trends® in Machine Learning.
- Bonawitz, K., Eichner, H., et al. (2019). *Towards Federated Learning at Scale: System Design*. Proceedings of MLSys.
- European Union. (2016). *General Data Protection Regulation (GDPR)*.
- Microsoft Azure Documentation. (n.d.). *Deploying Machine Learning Models on Azure Kubernetes Service (AKS)*.
- Goodfellow, I., Bengio, Y., & Courville, A. (2016). *Deep Learning*. MIT Press. (For general model understanding)

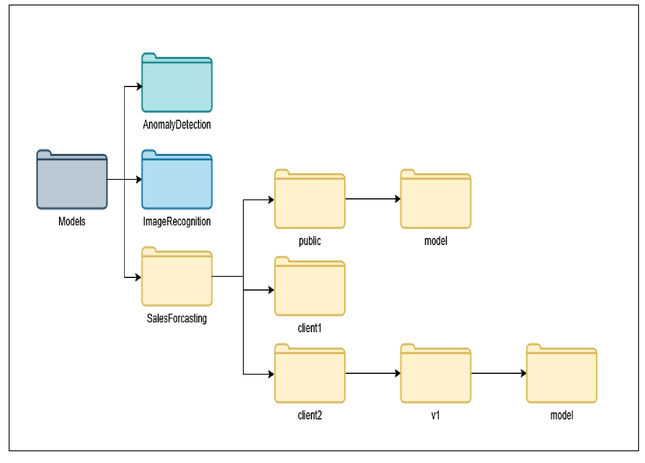


Image 5: File Structure