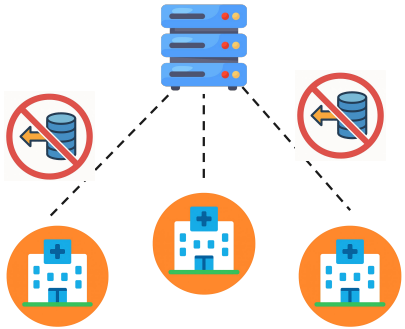
Fair Federated Learning based on Multi-Objective Optimization

DAYstributed, 29/05/2025

Michele Fontana

The Challenge

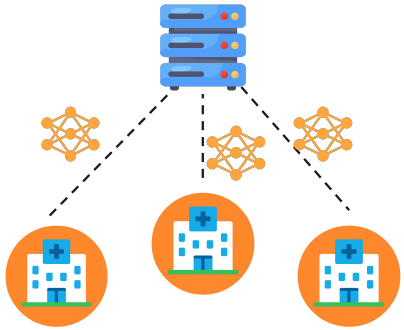


Problem

A consortium of hospital cooperate to build a shared AI model to predict patient rehospitalization within 30 days.

Hospitals contain **sensitive** data that cannot be shared (GDPR)

The Challenge



Problem

A consortium of hospital cooperate to build a shared AI model to predict patient rehospitalization within 30 days.

Hospitals contain **sensitive** data that cannot be shared (GDPR)

Solution

Federated Learning allow to train models without sharing the raw data

Clients train the model on their local data while the server coordinates the learning process

The Fairness Problem

95%

Accuracy

The Fairness Problem

95%

Accuracy

70%

White Patients

Positive prediction rate for rehospitalization

10%

Black Patients

Positive prediction rate for rehospitalization

0.5%

Hispanic Patients

Positive prediction rate for rehospitalization

The Fairness Problem

95%

Accuracy

70%

White Patients

Positive prediction rate for rehospitalization

10%

Black Patients

Positive prediction rate for rehospitalization

The system discriminates individuals based on their ethnicity

0.5%

Hispanic Patients

Positive prediction rate for rehospitalization

Fairness and the EU AI Act

Legal Implications

Fairness is not optional. Non-discrimination is a **strict** legal requirement. Models have to be **fair by-design**

Article 9

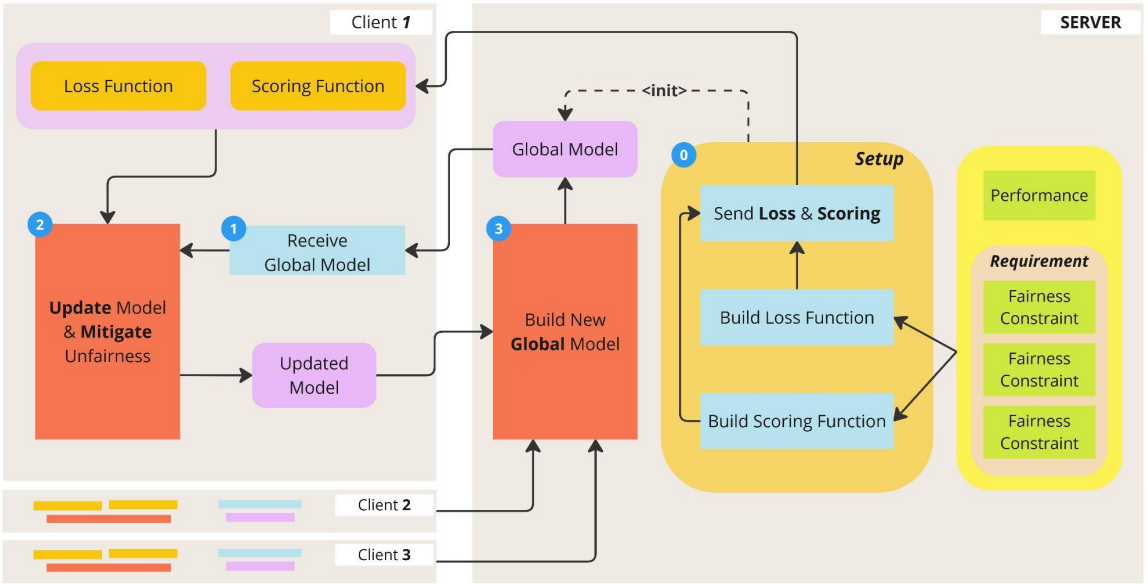
The risk management system [...] shall identify and analyze the known and foreseeable risks to health, safety and fundamental rights [...] i**ncluding risks related to discriminatory outcomes**

Recital 15

It is important to ensure that AI systems are used in a manner that is consistent with Union values, including the protection of fundamental rights, **non-discrimination**, and the protection of vulnerable groups.

GLOFAIR (Global-Local Optimization for Fairness in FL)

- FL method to learn neural networks that are fair and with high performance without pre-processing the training data
- MOO allows to map the performance-fairness tradeoff into a loss function



Fontana, Naretto, Monreale, Nanni, *"Fair Federated Learning methodology based on Multi-Objective Optimization"*, Submitted to Artificial Intelligence Journal, 2025

GLOFAIR vs State-of-the-Art

Feature	State-of-the-Art	GLOFAIR
Metrics Support	Single Metric	**Multiple** Metrics (DP, EOD ,...)
Attribute Handling	Often only binary	Binary and **non**-binary
Fairness Constraints	One constraint at a time	**Finite set** of constraints at the same time
Performance-Fairness Trade-off	Not implemented	**Explicit** control at training time

Take-home messages

- Performance is no longer the only goal.
- Fairness is now a legal and ethical imperative in high-risk domains (healthcare, finance…)
- GLOFAIR enables a new powerful solution for real-world problems:
  - High performance
  - Privacy-preserving
  - Simultaneously satisfy multiple fairness constraints

🧪 Try GLOFAIR on GitHub

QUESTIONS?