Towards Trustworthy AI in Next Generation Wireless Networks

Francesco Marcelloni
DII AI Group Coordinator, IT2PAO Lab Coordinator,
GoodAI Lab Coordinator
FAIR spoke 1 CO-PI
Dipartimento di Ingegneria dell'Informazione
Università di Pisa

July 3, 2024
School of Engineering
Largo Lucio Lazzarino 1
PISA
E-mail: francesco.marcelloni@unipi.it

Syllabus

1 Introduction

2 Federated Learning

3 Explainable Artificial Intelligence

4 Federated Learning of XAI models for Regression

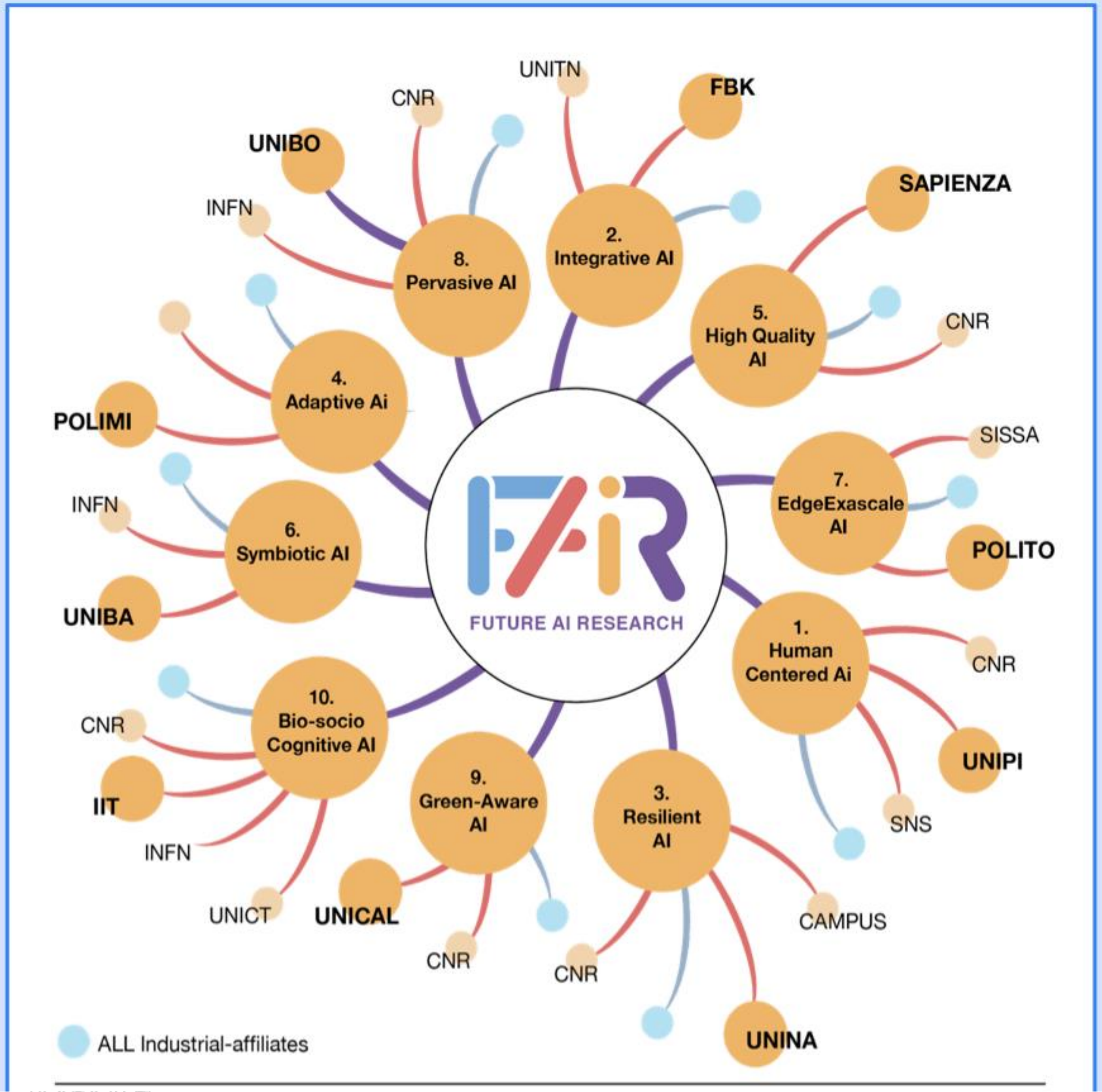
5 Federated Post-hoc Explainability

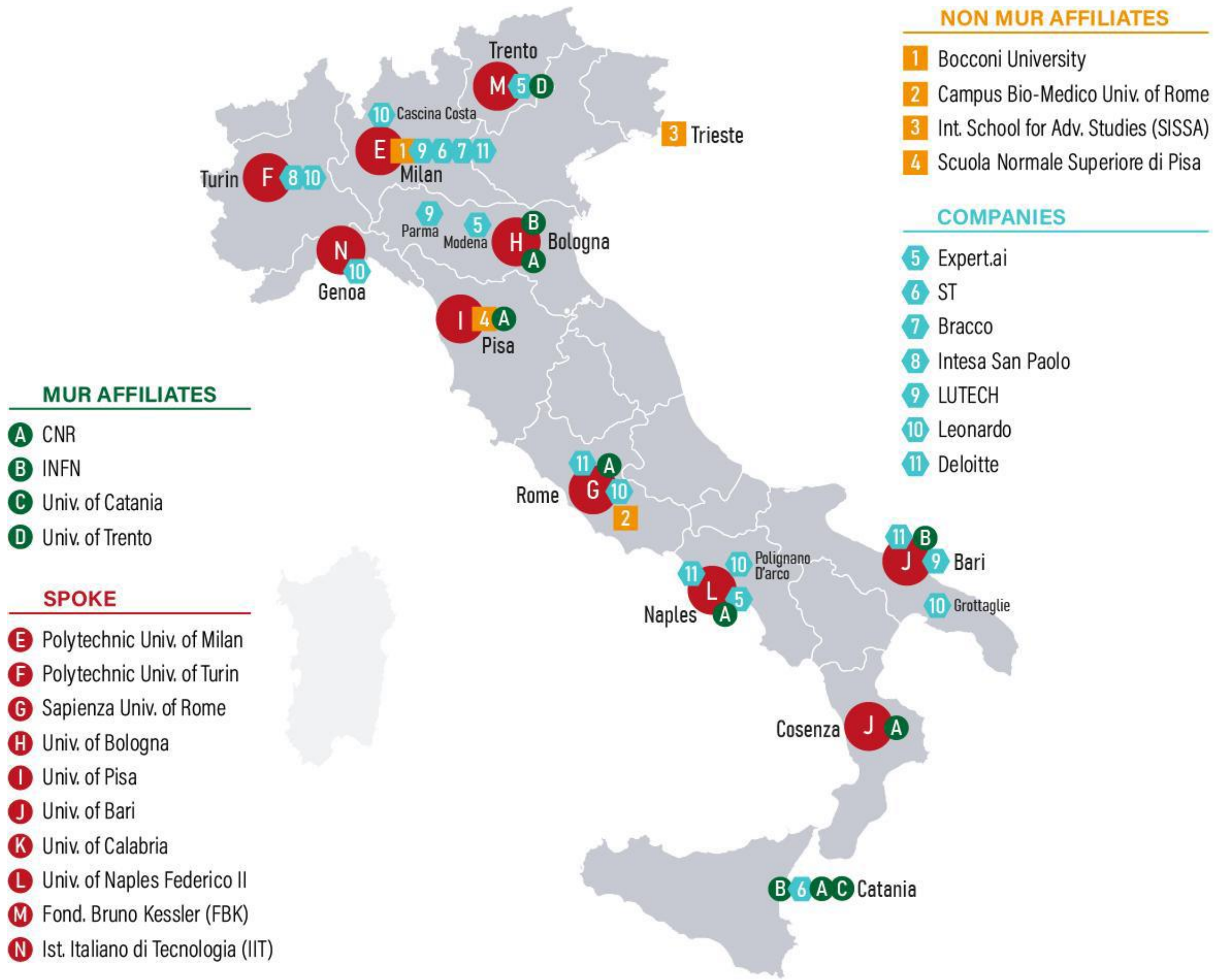
6 Conclusions

Future Artificial Intelligence Research (FAIR)

- FAIR - Future Artificial Intelligence Research (122 million)
- Call **‘Extended Partnerships: Artificial Intelligence. Fundamental aspects’ of the National Recovery and Resilience Plan**
- Started at the beginning of 2023 and will terminate at the end of 2025
- Theoretical, modelling and engineering aspects of modern Artificial Intelligence
- The FAIR project brings together **350 researchers** and is developed in **10 spokes.**

FUTURE AI RESEARCH


FAIR
FUTURE AI RESEARCH
1. Human Centered Ai
2. Integrative AI
3. Resilient AI
4. Adaptive Ai
5. High Quality AI
6. Symbiotic AI
7. EdgeExascale AI
8. Pervasive AI
9. Green-Aware AI
10. Bio-socio Cognitive AI
UNITN
FBK
CNR
UNIBO
INFN
SAPIENZA
CNR
POLIMI
SISSA
POLITO
INFN
UNIBA
CNR
UNIPI
CNR
IIT
SNS
INFN
UNICT
UNICAL
CAMPUS
CNR
CNR
UNINA
ALL Industrial-affiliates


Trento
Cascina Costa
Milan
Turin
Trieste
Parma
Modena
Bologna
Genoa
Pisa
Rome
Polignano D'arco
Naples
Bari
Grottaglie
Cosenza
Catania
MUR AFFILIATES
A CNR
B INFN
C Univ. of Catania
D Univ. of Trento
SPOKE
E Polytechnic Univ. of Milan
F Polytechnic Univ. of Turin
G Sapienza Univ. of Rome
H Univ. of Bologna
I Univ. of Pisa
J Univ. of Bari
K Univ. of Calabria
L Univ. of Naples Federico II
M Fond. Bruno Kessler (FBK)
N Ist. Italiano di Tecnologia (IIT)
NON MUR AFFILIATES
1 Bocconi University
2 Campus Bio-Medico Univ. of Rome
3 Int. School for Adv. Studies (SISSA)
4 Scuola Normale Superiore di Pisa
COMPANIES
5 Expert.ai
6 ST
7 Bracco
8 Intesa San Paolo
9 LUTECH
10 Leonardo
11 Deloitte

Future Artificial Intelligence Research (FAIR)

Spoke 1 “Human-centered AI”

Dino Pedreschi and Francesco Marcelloni

The study of **AI systems that cooperate synergistically, proactively and purposefully with humans** at individual and collective scale

- **amplifying instead of replacing human intelligence**
- **maximizing benefits** while **preventing and minimizing risks**

Spoke 1 - Critical Mass
39 multi-disciplinary scientists
9 UNIPI Departments
2 CNR Institutes
2 SNS Classes

Spoke 1 “Human-centered AI”

1) human-centered **machine learning and reasoning**:
   how humans and AI models interact synergistically, continuously co-evolving together (WP 1.1, 1.2, 1.3)

2) **social-aware AI**:
   how to understand and govern the dynamics and societal outcomes of large-scale socio-technical systems of humans and AIs (WP1.4, 1.5)

3) **responsible design of trustworthy AI systems**:
   how to responsibly (co-)design, develop, validate and use trustworthy AI systems (WP1.6)

Extensive **experiments, case studies and pilots** of Human-centered-AI systems (WP1.7).

FAIR FUTURE AI RESEARCH

Project FoReLab – Department of Excellence

Future-Oriented Research Laboratory (Forelab) aims to focusing towards new methodologies, paradigms, and enabling technologies for Industry 5.0

- **Trustworthy Artificial/Embodied Intelligence (TAEI)**: The activity focuses on frontier aspects of **trustworthy AI** and on the ability of systems to develop **intelligent behaviours** as emerging from the interactions between artificial agents (e.g. robots) and the environment, through the body of the agent itself (Embodied Intelligence), with the aim of promoting widespread, reliable and integrated use of both.

Good AI Lab

Fair Spoke 1's and FoReLab's spin-off: Good AI Lab

FUTURE AI RESEARCH

FoReLab Rethinking the future

How to design Trustworthy, Transparent, Robust, Safe, and Ethically-aligned AI Systems?

- Develop independent methods and benchmarks for assessing the compliance with the principles of "good AI".
- Provide consultancy, evaluation, and certification services aimed at companies and public administration
- Deliver multidisciplinary training courses

Location: Department of Information Engineering, UNIPI

Coordinator: Francesco Marcelloni

Scientific Board with members from UNIPI, CNR and SNS

https://goodailab.unipi.it/

The EU view of AI

AI ACT (21 May 2024):

to improve the functioning of the internal market by laying down a **uniform legal framework** in particular for the development, placing on the market, putting into service and the use of artificial intelligence systems in the Union in conformity with Union values, **to promote the uptake of human centric and trustworthy artificial intelligence** while ensuring **a high level of protection of health, safety, fundamental rights** enshrined in the Charter, including democracy and rule of law and environmental protection**,** against harmful effects of artificial intelligence systems in the Union **and to support innovation.**

The EU view of AI

Seven requirements toward Trustworthy AI

- Human agency and oversight
- Accountability (inc. auditability)
- Technical robustness and safety
- Societal and environmental wellbeing
- Diversity non-discrimination and fairness
- Privacy and data governance
- Transparency

Privacy and data governance → Need to collect data to train accurate AI models clashes with need to preserve privacy of data owners.

Transparency → "AI systems and their decisions should be explained in a manner adapted to the stakeholder concerned."

Fed-XAI
Federated Learning of eXplainable AI models

The EU view of AI

Human agency and oversight

Accountability

Technical robustness and safety

Societal and environmental wellbeing

Seven requirements toward Trustworthy AI

Diversity non-discrimination and fairness

The University of Pisa with the idea of Federated Learning of Explainable Artificial Intelligence Models (Fed-XAI) has been selected as **Key Innovator** by the European Commission's Innovation Radar.

Need to collect data to train accurate AI models clashes with need to preserve privacy of data owners.

"AI systems and their decisions should be explained in a manner adapted to the stakeholder concerned."

Fed-XAI
Federated Learning of eXplainable AI models

Privacy and Data Governance

- Increasing attention towards privacy preservation
  - EU General Data Protection Regulation (**GDPR**)
- Novel learning paradigm: **Federated Learning**
  - Training a *centralized model* on *decentralized data*
  - Participants share model updates, not data

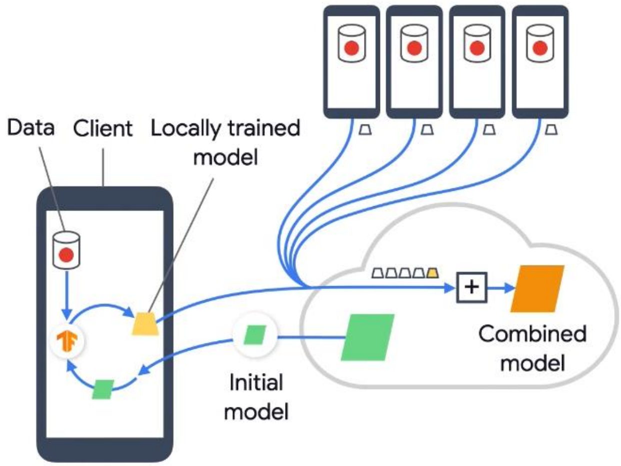
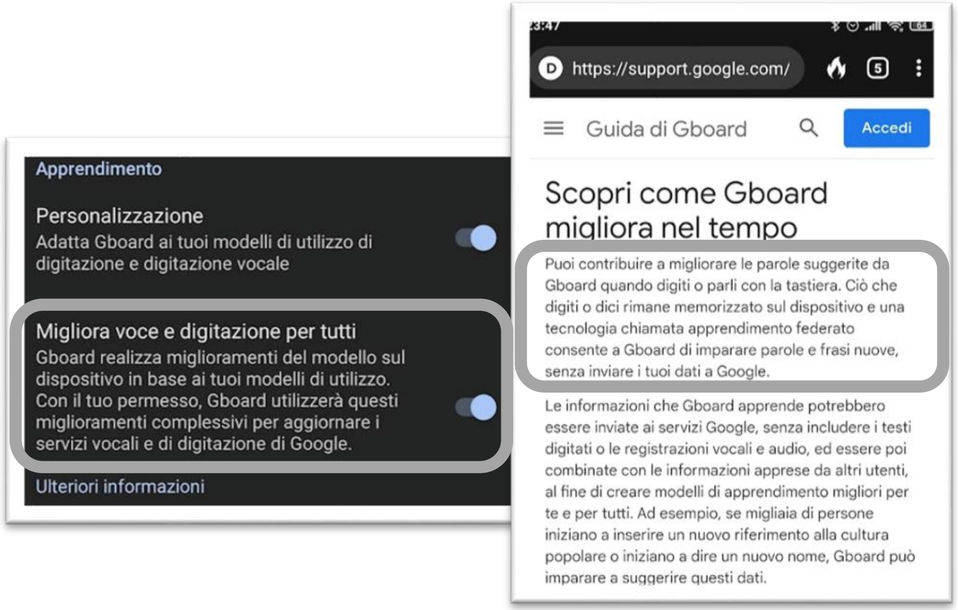
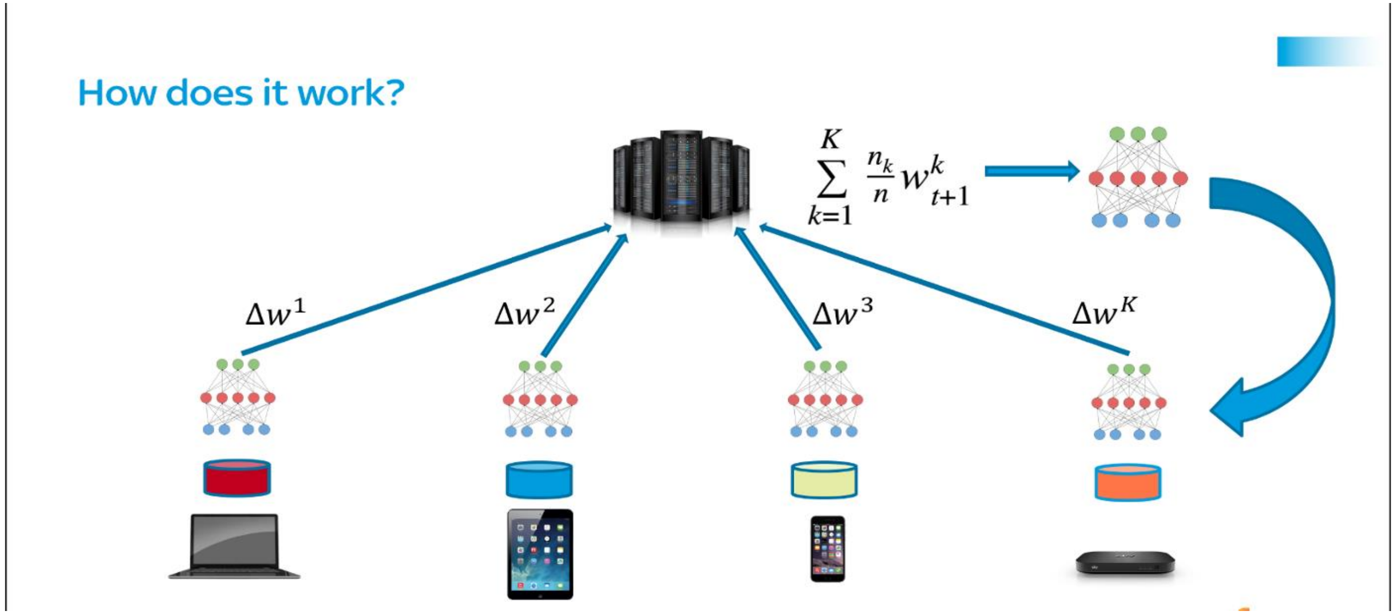


Figure from https://ml.berkeley.edu

- Examples:
  - Gboard
  - Voice Assistant



Popular approaches (Federated Averaging)



Federated Learning (Source: https://proandroiddev.com/federated-learning-e79e054c33ef)

- C = fraction of clients that participates in each federated round
- K = total number of clients (indexed by k)
- E = number of training passes each client makes over its local dataset on each round
- B = local minibatch size used for the client updates (B = ∞ indicates that the full local dataset is treated as a single minibatch
- P_k = set of indexes of data points on client k, with $n_k = |P_k|$

Algorithm 1 `FederatedAveraging`. The K clients are indexed by k; B is the local minibatch size, E is the number of local epochs, and η is the learning rate.

Server executes:
initialize w_0
for each round $t = 1, 2, \ldots$ **do**
  $m \leftarrow \max(C \cdot K, 1)$
  $S_t \leftarrow$ (random set of m clients)
  for each client $k \in S_t$ **in parallel do**
    $w_{t+1}^k \leftarrow$ ClientUpdate(k, w_t)
  $w_{t+1} \leftarrow \sum_{k=1}^{K} \frac{n_k}{n} w_{t+1}^k$

ClientUpdate(k, w): *// Run on client k*
$\mathcal{B} \leftarrow$ (split $\mathcal{P}_k$ into batches of size B)
for each local epoch i from 1 to E **do**
  for batch $b \in \mathcal{B}$ **do**
    $w \leftarrow w - \eta \nabla \ell(w; b)$
return w to server

Zhu H., Xu J., Liu S., Jin Y.Federated learning on non-IID data: A survey (2021) Neurocomputing, 465, pp. 371 - 390

Popular approaches

Federated Stochastic Gradient Descent (FedSGD) vs Federated averaging (FedAVG):

In **FedSGD** each client *k* computes the gradient on its local data at the current model w_t and the central server aggregates these gradients and updates the global model.
Note that FedSGD coincides to FedAvg with
$C = 1$, $B = \infty$, $E = 1$

In **FedAVG** each client locally takes one or multiple steps of gradient descent on the current model w_t using its local data, and the server then takes a weigthed average of the resulting models.

- C = fraction of clients that participates in each federated round
- K = total number of clients (indexed by *k*)
- E = number of training passes each client makes over its local dataset on each round
- B = local minibatch size used for the client updates
  (B = ∞ indicates that the full local dataset is treated as a single minibatch
- P_k = set of indexes of data points on client *k*, with $n_k = |P_k|$

Algorithm 1 `FederatedAveraging`. The K clients are indexed by k; B is the local minibatch size, E is the number of local epochs, and η is the learning rate.

Server executes:
 initialize w_0
 for each round $t = 1, 2, \dots$ **do**
  $m \leftarrow \max(C \cdot K, 1)$
  $S_t \leftarrow$ (random set of m clients)
  for each client $k \in S_t$ **in parallel do**
   $w_{t+1}^k \leftarrow \text{ClientUpdate}(k, w_t)$
  $w_{t+1} \leftarrow \sum_{k=1}^{K} \frac{n_k}{n} w_{t+1}^k$

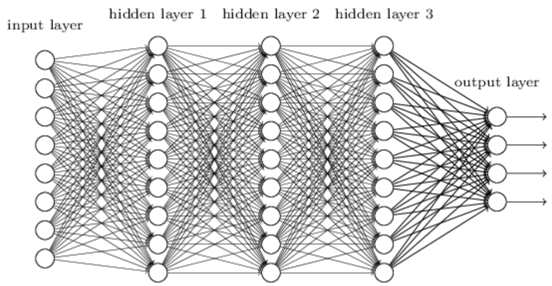
ClientUpdate(k, w): // *Run on client* k
 $\mathcal{B} \leftarrow$ (split $\mathcal{P}_k$ into batches of size B)
 for each local epoch i from 1 to E **do**
  for batch $b \in \mathcal{B}$ **do**
   $w \leftarrow w - \eta \nabla \ell(w; b)$
 return w to server

Zhu H., Xu J., Liu S., Jin Y. Federated learning on non-IID data: A survey (2021) Neurocomputing, 465, pp. 371 - 390

Transparency

- Communication
  - Humans have the right to be informed that they are interacting with an AI system.
- Traceability
  - Data gathering/labelling and algorithms should be documented to the best possible standard
- **Explainability**
  - Systems and decisions should be explained in a manner adapted to the stakeholder concerned
    - *Easy* for decision trees, *critical* for deep neural networks

Deep neural network



Deep neural networks
- Multiple layers of non-linear information processing
- Often referred to as ***opaque*** or ***black-box*** models

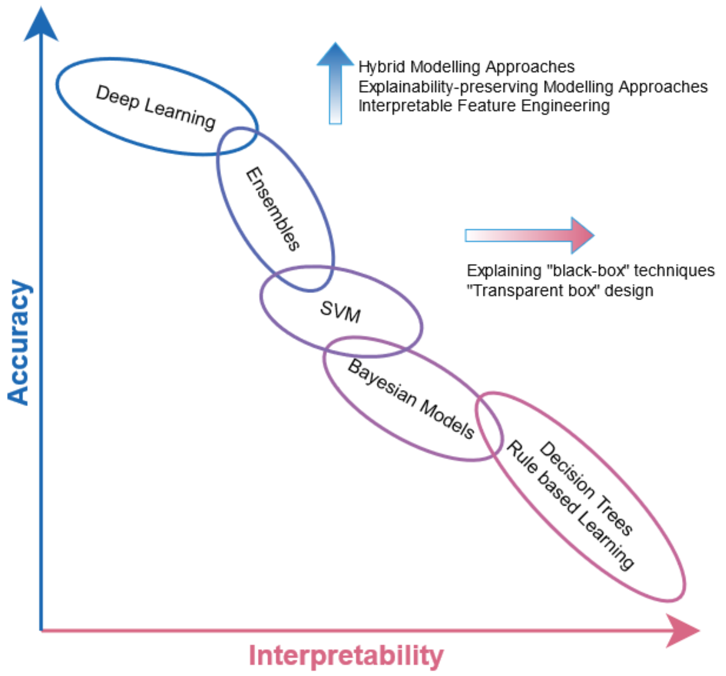
Explainability

- How to achieve explainability?
  - **Post-hoc Explainability** Techniques
  - Design of **Inherently Interpretable Models**

- How to characterize inherent interpretability?
  - There exists a **trade-off*** between
    - model **accuracy**
    - model **interpretability**

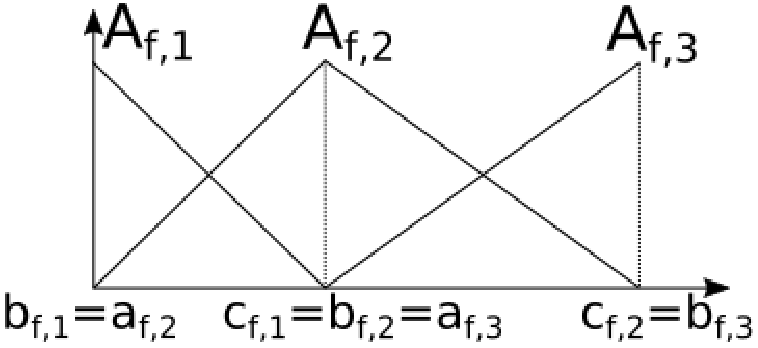
*given

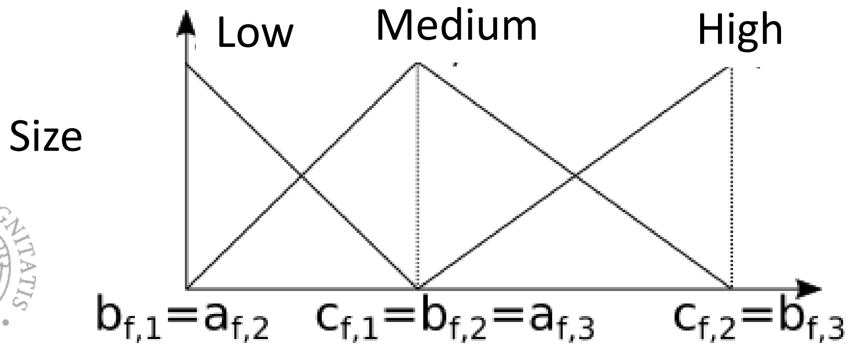
- a target function of a certain complexity, and
- a suitable amount of avaliable data

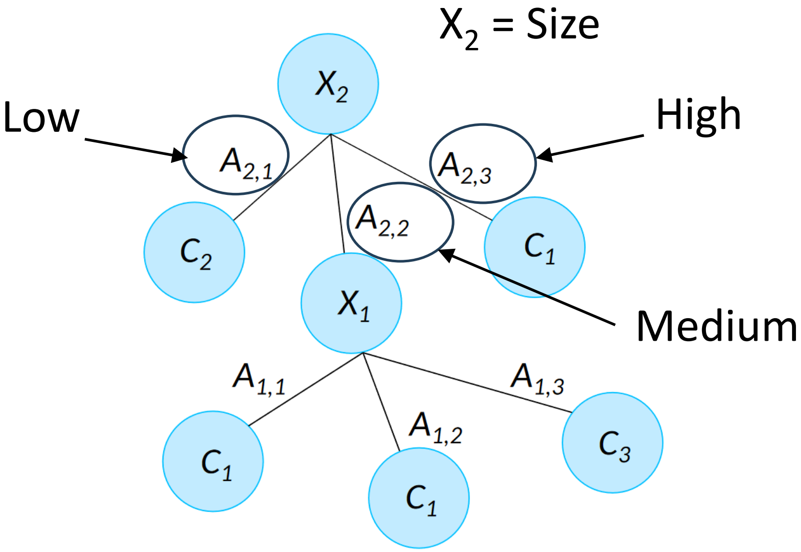


Interpretable Models: Fuzzy Decision Tree (FDT)

- Directed acyclic graph
- Generated in a top–down way by performing **recursive partitions of the attribute space**.
- Typically, requires a **fuzzy partition defined upon each continuous attribute**.







Multi-way FDT

IF X_1 **IS** $A_{1,j_{k,1}}$ … **AND** X_F **IS** $A_{F,j_{k,F}}$

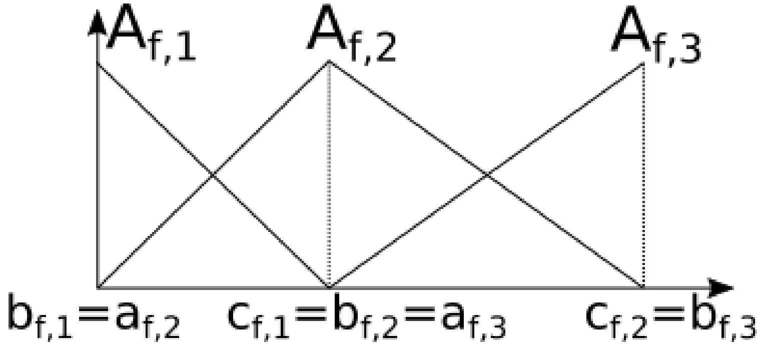
THEN $y_k(\mathbf{x}) = \gamma_{k,0} + \sum_{i=1}^{F} \gamma_{k,i} \cdot x_i$

Interpretable Models: Fuzzy Rule-Based Systems

- The model consists of a rule-base, i.e., a collection of rules in the form

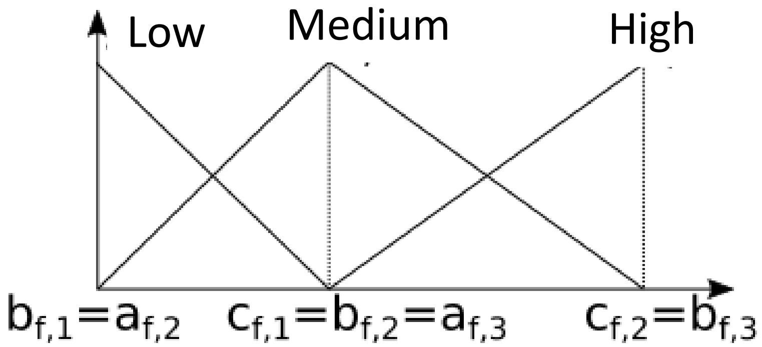
  if «antecedent» then «consequent»

- Example of rules in the form *first-order Takagi-Sugeno-Kang Fuzzy Rule-Based Systems*



$$\textbf{IF } X_1 \textbf{ IS } A_{1,j_{k,1}} \ \ldots \ \textbf{AND } X_F \textbf{ IS } A_{F,j_{k,F}}$$

$$\textbf{THEN } y_k(\mathbf{x}) = \gamma_{k,0} + \sum_{i=1}^{F} \gamma_{k,i} \cdot x_i$$



Size

$$\textbf{IF } \mathit{Size} \textbf{ IS } \mathit{low} \ \ldots \ \textbf{AND } X_F \textbf{ IS } A_{F,j_{k,F}}$$

$$\textbf{THEN } y_k(\mathbf{x}) = \gamma_{k,0} + \sum_{i=1}^{F} \gamma_{k,i} \cdot x_i$$

The Traditional TSK FRBS

Let

- $X = \{X_1, X_2, \dots, X_F\}$, be a set of input variable
- U_f, be the universe of discourse of variable X_f
- Y, be a continuous output variable
- $P_f = \left\{A_{f,1}, A_{f,2}, \dots, A_{f,T_f}\right\}$, be a fuzzy partition over U_f with T_f fuzzy sets

The generic k^{th} rule, R_k, of the rule base is in the form:

IF X_1 **IS** $A_{1,j_{k,1}}$ … **AND** X_F **IS** $A_{F,j_{k,F}}$

THEN $y_k(\mathbf{x}) = \gamma_{k,0} + \sum_{i=1}^{F} \gamma_{k,i} \cdot x_i$

Inference stage:

Given input pattern $\mathbf{x}$, compute strength of activation of each rule:

$$w_k(\mathbf{x}) = \prod_{f=1}^{F} \mu_{f,j_{k,f}}(x_f) \quad \text{for } k = 1,2,\dots,K$$

Estimation of antecedent parameters:

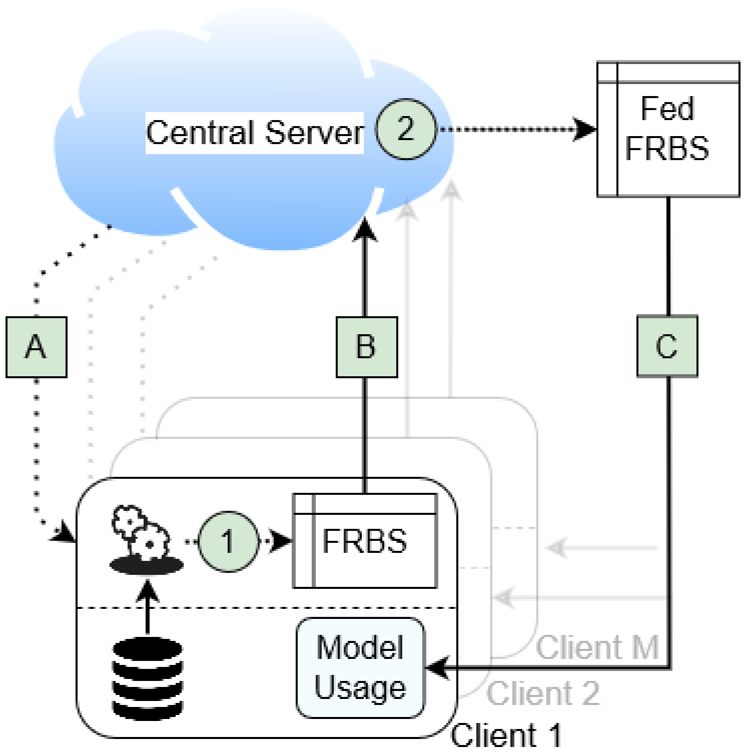
- Clustering in the input-output product space
- Fitting convex envelop of the projected membership values for each discovered cluster

$$\hat{y}(\mathbf{x}) = \sum_{k=1}^{K} \left(\frac{w_k(\mathbf{x})}{\sum_{h=1}^{K} \left(w_h(\mathbf{x})\right)} \right) \cdot y_k(\mathbf{x})$$

Estimation of consequent parameters:

- Weighted Least Squared method

Federated Learning of TSK FRBS



A	**Configuration:** central server configures the learning process
1	**Local learning of TSK-FRBSs**
B	**Transmission** of **local models** to the central server
2	**Federated learning of the global TSK-FRBS:** aggregation of the models
C	**Transmission** of the **aggregated model** to the clients

J. L. Corcuera Bárcena et al. , "*An Approach to Federated Learning of Explainable Fuzzy Regression Models,*" *IEEE Int'l Conf. on Fuzzy Systems (FUZZ-IEEE)*, 2022

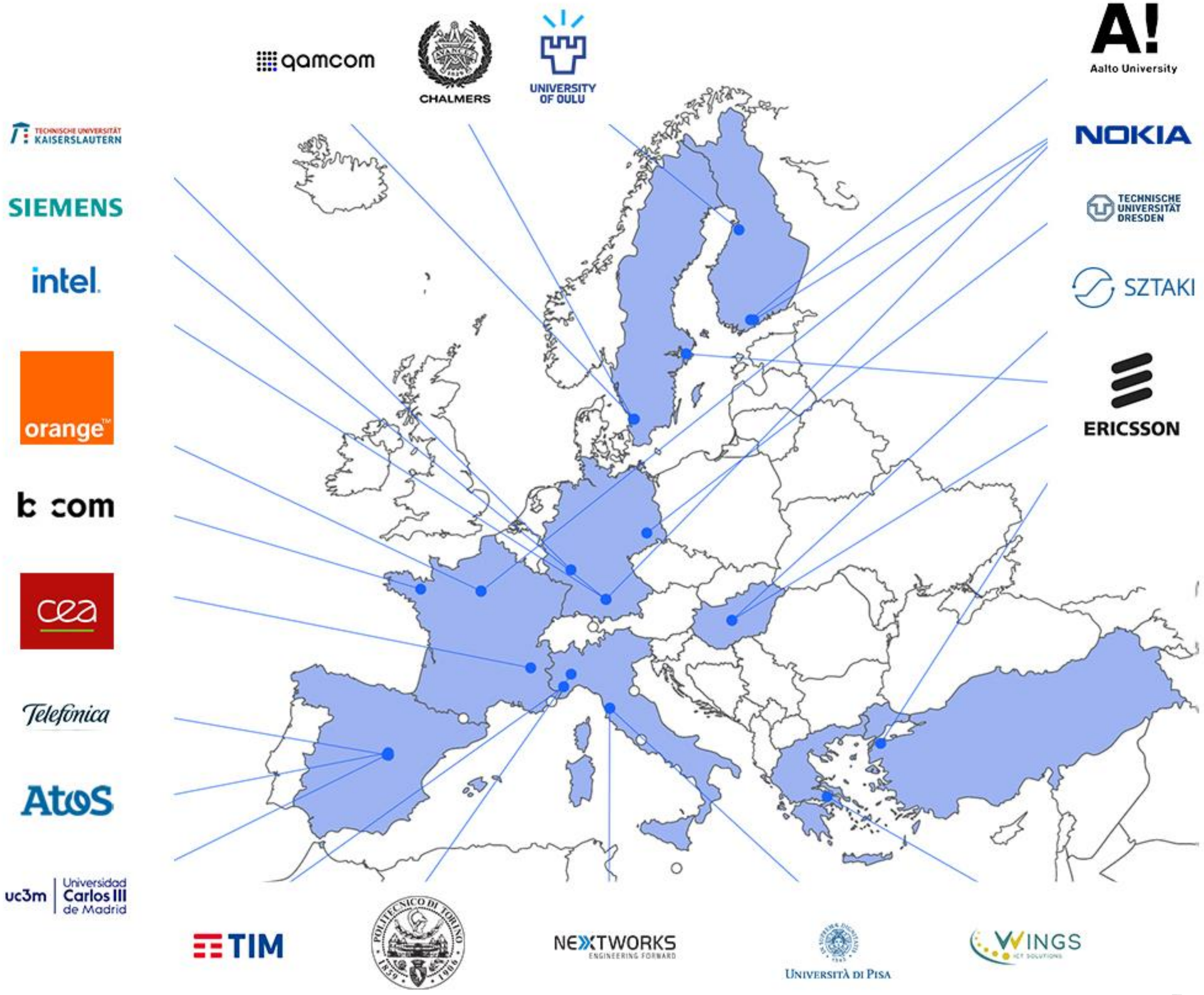
HEXA-X: The European 6G flagship project

Hexa-X

A **flagship for B5G/6G vision** and intelligent fabric of technology enablers connecting human, physical, and digital worlds.

Jan 2021 – June 2023

EU project: HEXA-X - **Programme**: Horizon 2020 - **Grant Agreement ID**: 101015956



6G Research Challenges

Connecting intelligence: 6G shall enable **real-time and vital and fully trusted AI/ML technologies** for significantly improved efficiency and service experience, with the human factor ("**human in the loop**") integrated.

Trustworthiness: 6G shall ensure the **confidentiality**, **integrity** and **availability** of end-to-end communications, and guarantee **data privacy**, **operation resilience** and **security**, building trust of wireless networks as well as its enabled applications among consumers and enterprises.

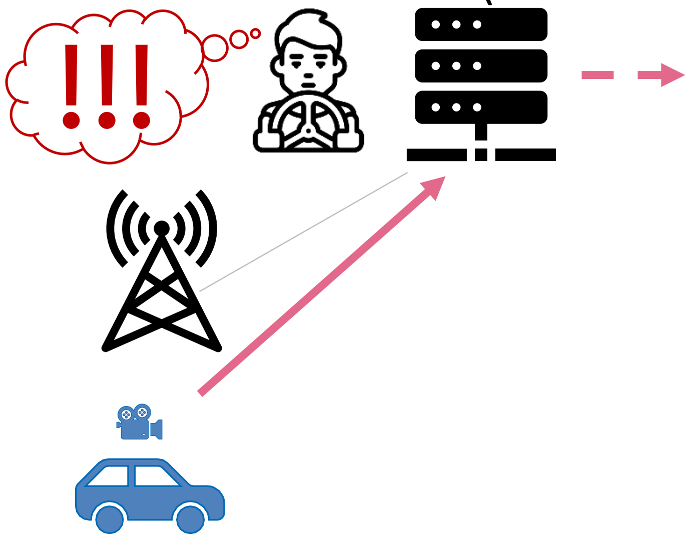
Image extracted from: Hexa-X Deliverable D1.2, "Expanded 6G vision, use cases and societal values". Online: https://hexa-x.eu/wp-content/uploads/2021/05/Hexa-X_D1.2.pdf

Francesco Marcelloni

Vehicular network case study

Tele-operated driving (ToD): one of the innovative services envisioned in 6G systems

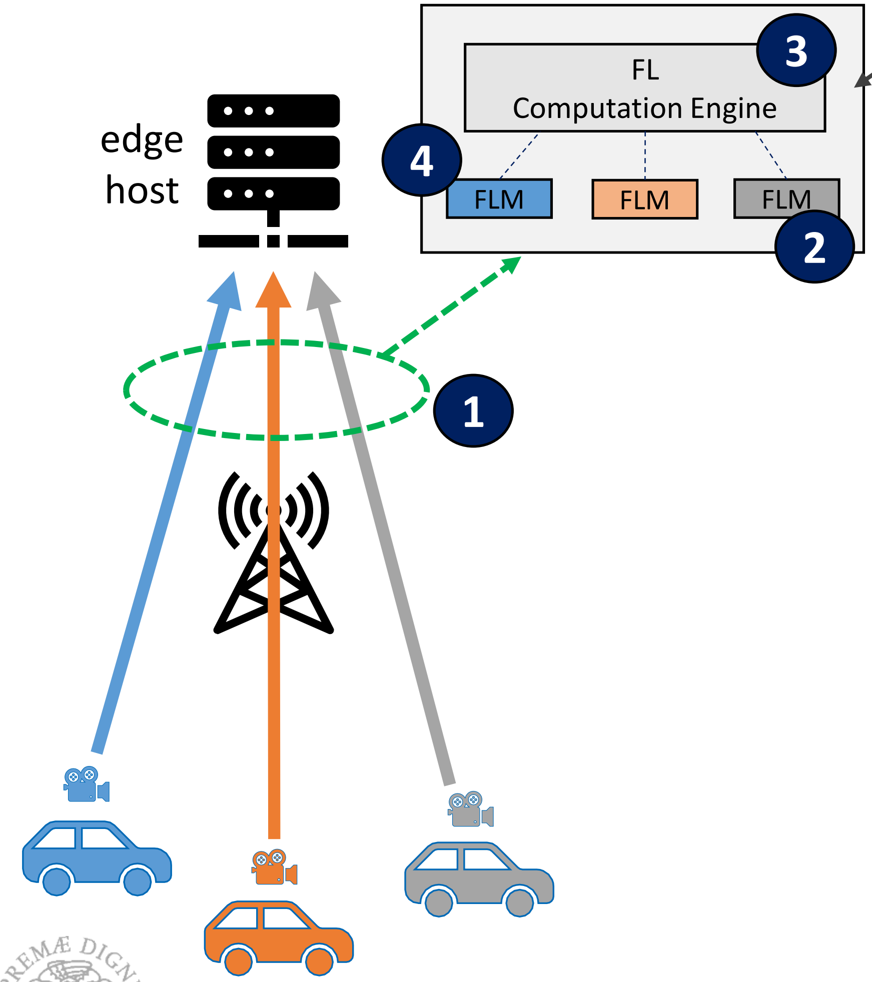
- Connected cars send real-time video streams representing the pilot-seat view to a remote driver at the edge
- The remote driver (human or machine) can control the car by sending commands based on the video



- It is crucial to be able to **predict any fluctuations** of video quality **in advance**
- A simulation campaign generates the training dataset for 15 clients, based on realistic traffic data from TIM
- **Regression task**: Predict *future* Quality of Experience given *historical values* of Quality of Service and contextual metrics

J. L. Corcuera Bárcena et al. , "*Enabling federated learning of explainable AI models within beyond-5G/6G networks*" Computer Communications (2023).

Vehicular network case study



Edge-based **F**ederated **L**earning **as a S**ervice (**FLaaS**) framework

Training phase

1. While User Equipments' (UEs') video streams are being transmitted, **QoS metrics are collected** from the receiving application (e.g., delay) and the network (e.g., cell load)
2. For each UE, a FL Local Manager (FLM) at the UE/edge **learns a local XAI model**
3. A FL Computation Engine **builds a global XAI model** by aggregating local ones received from FLMs. The global XAI model is then provided to the FLMs

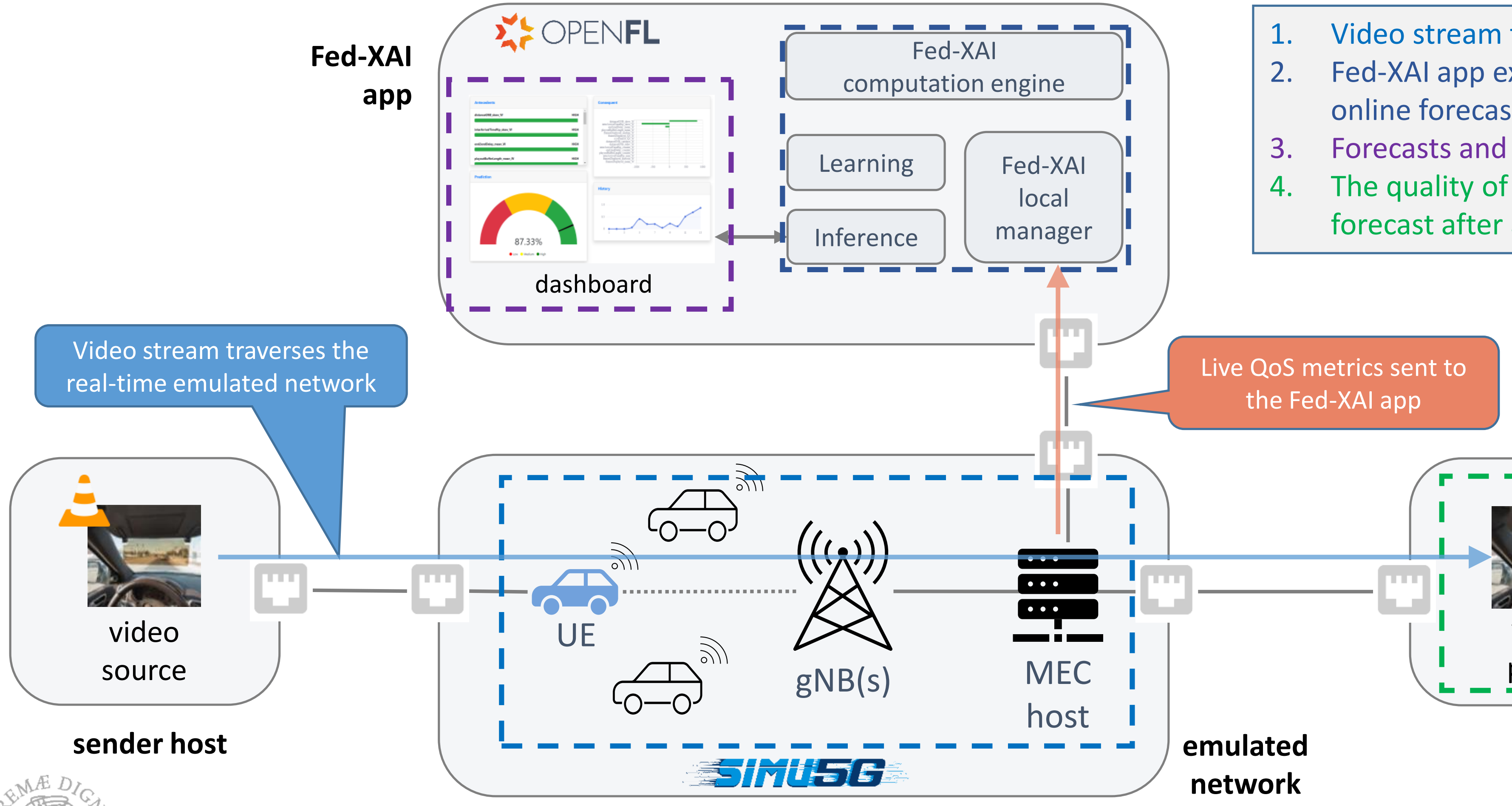
Inference phase

4. Using the live QoS metrics and the global XAI model, each FLM **predicts** the future video quality for the UE, showing it on a **dashboard** in real time

J. L. Corcuera Bárcena et al. , "*Enabling federated learning of explainable AI models within beyond-5G/6G networks*" Computer Communications (2023).

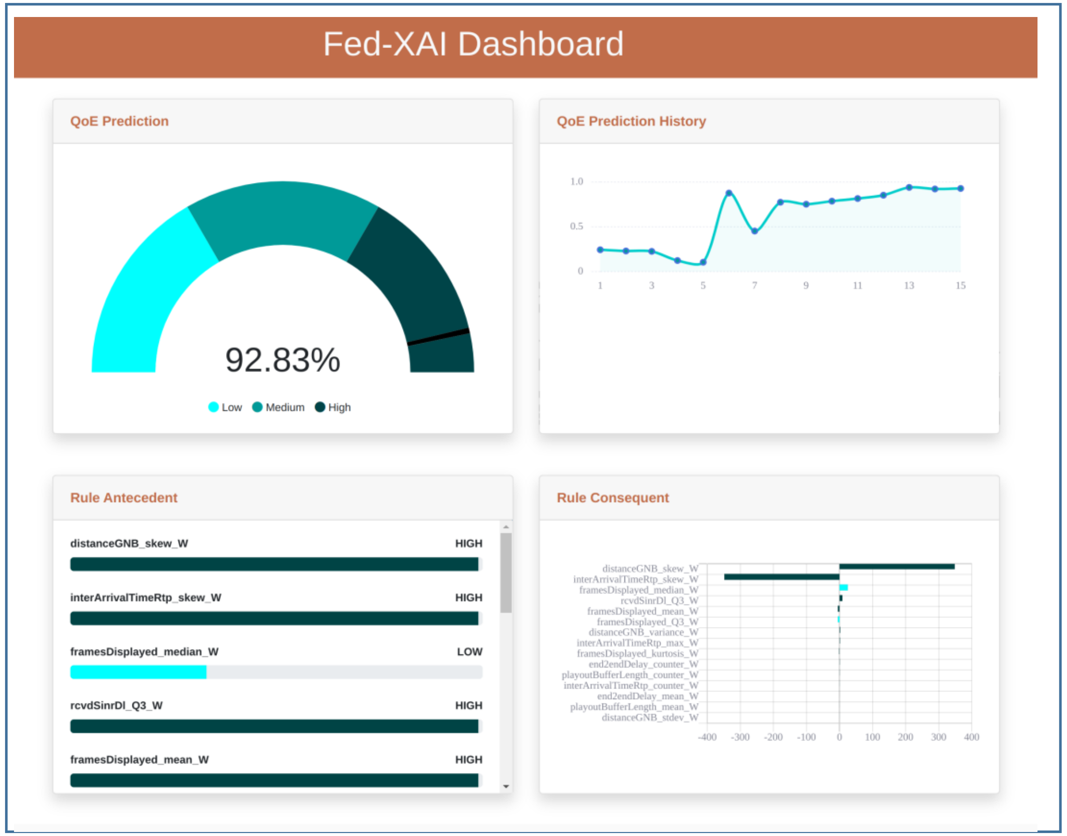
Vehicular network case study

Real-time testbed composed of real devices running real apps, and a network emulator



1. Video stream flows through the emulated network
2. Fed-XAI app exploits the pre-trained XAI model to make online forecasts using live QoS data from the network
3. Forecasts and root causes are shown on a dashboard
4. The quality of the video at the receiver matches the forecast after a few seconds

Fed-XAI Dashboard



View at inference time

- **Prediction**: predicted value
- **History**: track of past prediction values
- **Antecedents**: antecedent values, sorted by weights in consequent
- **Consequent**: weights of each feature for current prediction

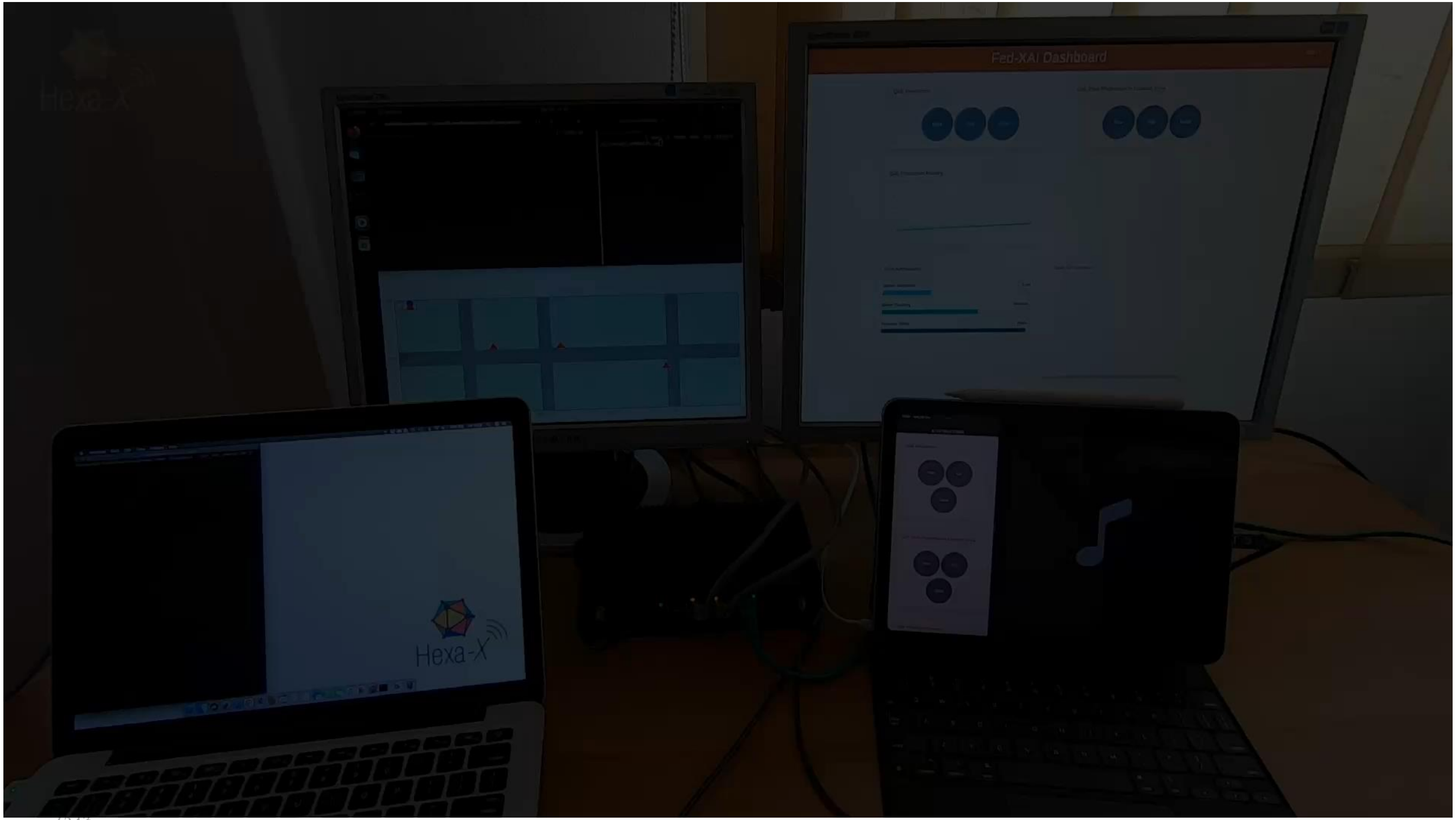
explainability information can be leveraged to identify **potential countermeasure** to be taken

$$
\begin{aligned}
R_k:\ & \textbf{IF}\ framesDisplayed\_Q3\ is\ High \\
& \textbf{AND}\ framesDisplayed\_mean\ is\ Medium \\
& \textbf{AND}\ playoutBufferLength\_mean\ is\ Medium \\
& \textbf{AND}\ interArrivalTimeRtp\_max\ is\ High \\
& \textbf{AND}\ framesDisplayed\_median\ is\ Low \\
& \textbf{AND}\ playoutBufferLength\_counter\ is\ Medium \\
& \textbf{AND}\ distanceBS\_variance\ is\ Low \\
& \textbf{AND}\ distanceBS\_stdev\ is\ Low \\
& \textbf{AND}\ interArrivalTimeRtp\_counter\ is\ High \\
& \textbf{AND}\ interArrivalTimeRtp\_skew\ is\ High \\
& \textbf{AND}\ framesDisplayed\_kurtosis\ is\ Low \\
& \textbf{AND}\ end2endDelay\_mean\ is\ High \\
& \textbf{AND}\ rcvdSinrDl\_Q3\ is\ Medium \\
& \textbf{AND}\ end2endDelay\_counter\ is\ High \\
& \textbf{AND}\ distanceBS\_skew\ is\ Medium \\
& \textbf{THEN}:\ QoE = -0.210 \\
& +0.246 \cdot framesDisplayed\_Q3 \\
& +0.465 \cdot framesDisplayed\_mean \\
& +0.636 \cdot playoutBufferLength\_mean \\
& -0.291 \cdot interArrivalTimeRtp\_max \\
& +0 \cdot framesDisplayed\_median \\
& +0.293 \cdot playoutBufferLength\_counter \\
& +0.001 \cdot distanceBS\_variance \\
& +0.019 \cdot distanceBS\_stdev \\
& +0.223 \cdot interArrivalTimeRtp\_counter \\
& -0.21 \cdot interArrivalTimeRtp\_skew \\
& +0 \cdot framesDisplayed\_kurtosis \\
& -0.257 \cdot end2endDelay\_mean \\
& +0.454 \cdot rcvdSinrDl\_Q3 \\
& +0.223 \cdot end2endDelay\_counter \\
& -0.031 \cdot distanceBS\_skew
\end{aligned}
$$

A. Bechini et al. , "*An Application for Federated Learning of XAI Models in Edge Computing Environments*" IEEE International Conference on Fuzzy Systems (2023).

Fed-XAI demo

https://www.youtube.com/watch?v=azuTyB-LdmQ



Fed-XAI application - Implementation Details

OPENFL

Federated Learning Framework

- **OpenFL** framework**,** developed by Intel and now hosted by The Linux Foundation
  - Seamless integration with containers paradigm
  - Highly flexible, although designed for the aggregation of models like NNs, (i.e. via FedAvg)
- Extended to support FL of inherently interpretable models
  - **OpenFL-XAI** just released https://github.com/Unipisa/OpenFL-XAI
- Actively employed within Hexa-X European project
- Supports research, development, and demonstration activities concerning the FL of XAI models

Post-Hoc Explainability: SHAP

- Alternative approach: Generate a black box model and then try to explain why the inputs produce that output
- Shapley values quantify the impact of each feature on model prediction

Shapley Values in XAI

$$\hat{y}_i = f(x_i) = \phi_0 + \sum_{j=1}^{F} \phi_j$$

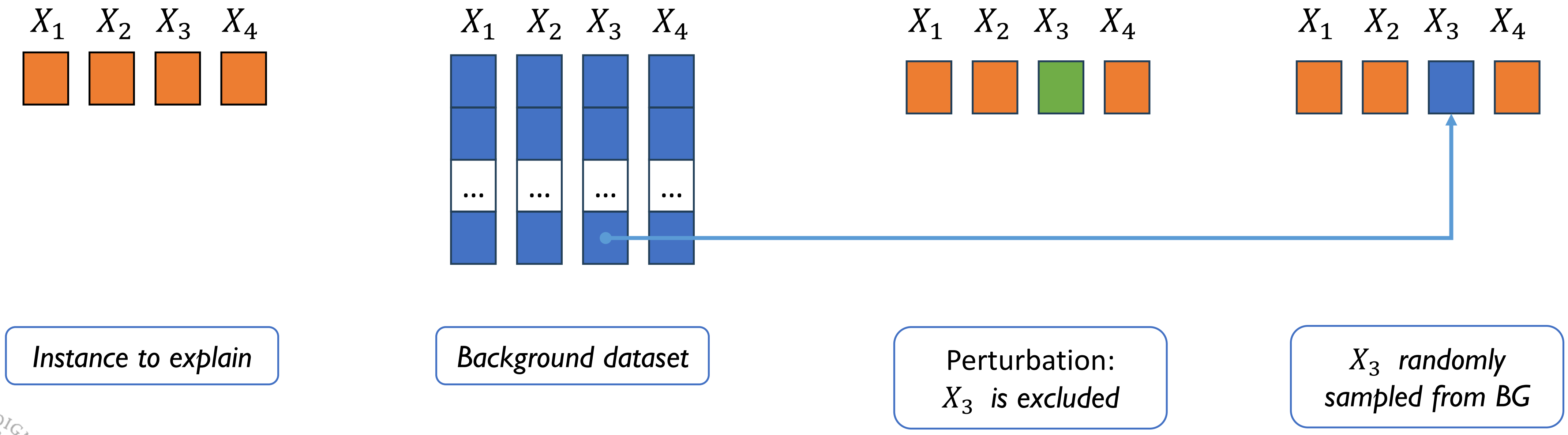
- f Predictive model
- x_i Generic F-dimensional input instance
- ϕ_0 Average of the predictions from a *background dataset*
- ϕ_i Shapley values

- **Local,** i.e., explains individual predictions
- **KernelShap** variant: linear regression-based approximation
  - More **efficient** than naive calculation
  - **Model-agnostic,** suited for both classification and regression tasks

P. Ducange et al. “Consistent Post-Hoc Explainability in Federated Learning through Federated Fuzzy Clustering” IEEE WCCI 2024 - World Congress on Computational Intelligence

Challenge of adopting SHAP in the FL setting

- Estimation of Shapley values for the explanations for x_i involves testing *coalitions* of features by *perturbing* x_i
- A **background dataset (BG)** is exploited for perturbing x_i
  - Replace features excluded from a coalition with those of instances randomly sampled from BG
  - The BG should coincide with the set of data used for learning the f model (i.e., the **training set**)
  - It is a common practice to reduce the numerosity of the BG (e.g., through sampling)



P. Ducange et al. "Consistent Post-Hoc Explainability in Federated Learning through Federated Fuzzy Clustering" IEEE WCCI 2024 - World Congress on Computational Intelligence

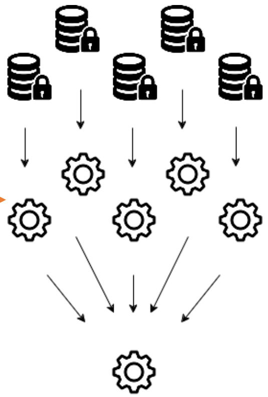
Challenge of adopting SHAP in the FL setting

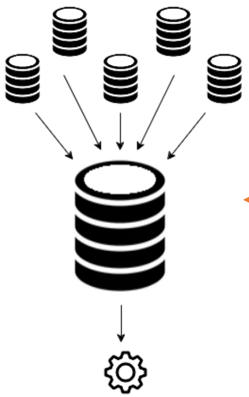
Challenges

- The choice of the **background dataset impacts the resulting explanations**
- In the FL setting the **training set is not available in its entirety** to any party

Desiderata

- **Privacy preservation:** the explainability process should not violate privacy (as a constrain of the FL setting)
- **Consistency**: explanations of the same data instance for the FL model are identical for different participants
- **Accuracy:** explanations in FL match those that would be obtained in the traditional *centralized* setting

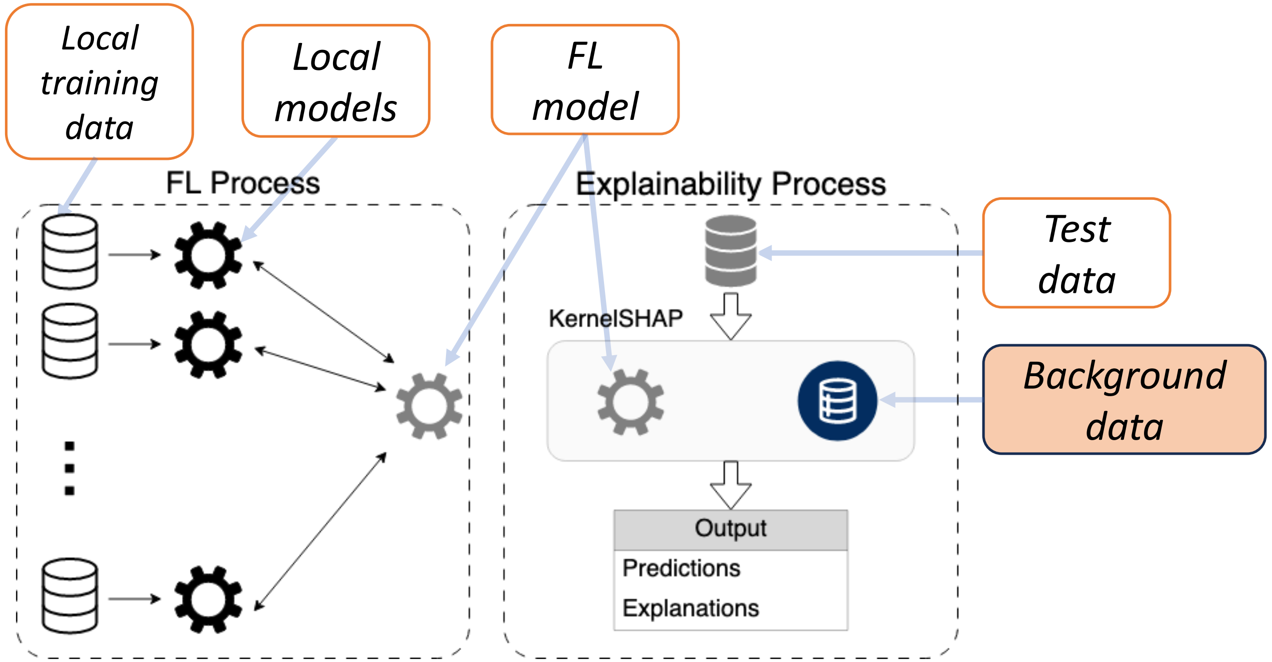




P. Ducange et al. “Consistent Post-Hoc Explainability in Federated Learning through Federated Fuzzy Clustering” IEEE WCCI 2024 - World Congress on Computational Intelligence

Federated SHAP – how to design a *proper* and *common* background dataset

- **Start** communication topology with **horizontally** partitioned data
- The model learned in a federation fashion is ***opaque*** (it requires post-hoc techniques)
- ***non-i.i.d.* setting**: local data follow distributions different from each other and from the overall distribution



Background dataset generation through Federated Fuzzy Clustering

- **Privacy preserving** summarization of scattered data
- **Cluster centers** are exploited as background
  - **common**, i.e., shared to all participants
  - **representative** of the entire data distribution
- Federated-FCM* is adopted but the choice of the clustering algorithm is not critical for our objective

*Corcuera Bárcena et al. ***A federated fuzzy c-means clustering algorithm.*** (2021)

P. Ducange et al. “Consistent Post-Hoc Explainability in Federated Learning through Federated Fuzzy Clustering” IEEE WCCI 2024 - World Congress on Computational Intelligence

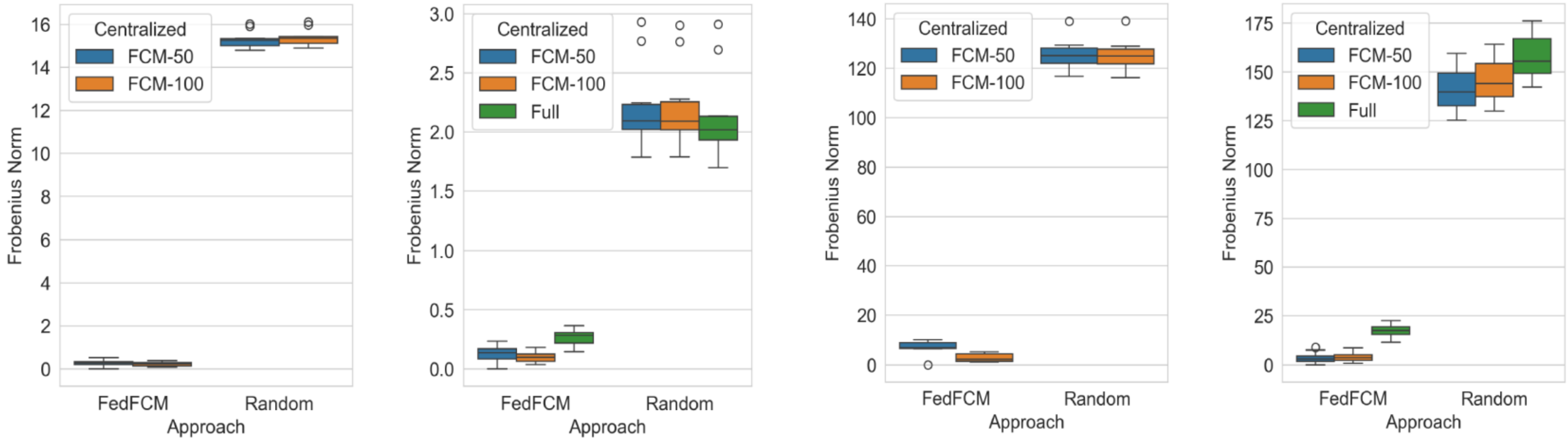
Experimental setup – Baseline approaches

*BG = Background dataset	**Ensure consistency** (same background for all participants)	**Ensure accuracy** (represent the actual overall data distribution)	**Preserve privacy**
Federated SHAP **BG** ← $\boldsymbol{K}$ cluster centers obtained through Federated FCM	✓	✓	✓
Centralized **BG** ← union of the data locally stored in the clients	✓	✓	✗
Random **BG** ← randomly sampling $\boldsymbol{K}$ instances from a uniform distribution over the input space	✓	✗	✓
Local^{m} **BG^{m}** ← $\boldsymbol{K}$ cluster centers obtained through local FCM on the m-th participant	✗	✗	✓

P. Ducange et al. “Consistent Post-Hoc Explainability in Federated Learning through Federated Fuzzy Clustering” IEEE WCCI 2024 - World Congress on Computational Intelligence

Accuracy of explanations

- **Accuracy** $\stackrel{\text{def}}{=}$ explanations match those that would be obtained in the traditional *centralized* setting
- Three *centralized* versions
  - **BG** ← Full training
  - **BG** ← FCM, 50 centers
  - **BG** ← FCM,100 centers
- Ten values for each approach with different random seed



(a) Magic (b) Rice (c) California (d) Abalone

Discrepancy of both the ***Federated SHAP (FedFCM)*** *and the* ***Random*** *approach with the baseline centralized approaches in terms of Frobenius norm of the pairwise difference of* $\boldsymbol{\phi}$ *matrices*

- **Federated SHAP:** low discrepancy with the *centralized* case, low variability
- **Random:** high discrepancy with the *centralized* case, high variability

Challenges and future directions

- Not Independent and Identically distributed (**Non-IID**) data
- Federated Learning with **Streaming Data**
- **Privacy protection**: It is still not clear to what extent these methods harm the data privacy, and there is no quantitative measures to identify the degree of privacy leakage.
- **Large number of hyperparameters** (total number of clients, number of local epochs, client dropout probability)
- Lack of **universally recognized benchmark datasets**
- Research on **Federated Clustering** is still very limited although there exist a number of interesting application domains

I would like to thank the members of the AI group at the Department of Information Engineering

Questions?

For questions and details, please write to Francesco Marcelloni (francesco.marcelloni@unipi.it)