Finanziato dall'Unione europea
NextGenerationEU

Ministero dell'Università e della Ricerca

Italiadomani
PIANO NAZIONALE DI RIPRESA E RESILIENZA

Future Artificial Intelligence Research

Spoke 1
Human-Centered AI

Dino Pedreschi

Università di Pisa

Future Artificial Intelligence Research

Spoke 1 - Human-centered AI - Partners

UNIVERSITÀ DI PISA

Consiglio Nazionale delle Ricerche

SCUOLA NORMALE SUPERIORE

UNIPI (25)
Bacciu Davide
Bernardeschi Cinzia
Cimino MGC Antonio
Cococcioni Marco
De Santis Federica
Ducange Pietro
Fabris Adriano
Fantoni Gualtiero
Gattiglia Gabriele
Greco Alberto
Guidotti Riccardo
Lazzerini Beatrice
Lazzeroni Michela
Lenci Alessandro
Lomonaco Vincenzo
Malizia Alessio
Marcelloni Francesco
Micheli Alessio
Monreale Anna
Neri Veronica
Pedreschi Dino
Pellecchia Enza
Rizzi Andrea
Romito Marco
Ruggieri Salvatore

SNS (5)
Benzi Michele
Giannotti Fosca
Pellungrini Roberto
Squartini Tiziano
Trenz Hans-Joerg

CNR (9)
Boldrini Chiara
Conti Marco
Fabrizio Falchi
Nanni Mirco
Passarella Andrea
Perego Raffaele
Rinzivillo Salvatore
Sebastiani Fabrizio
Straccia Umberto

Spoke 1 - Critical Mass
39 persons
9 UniPI Departments
2 CNR Institutes
2 SNS Classes

Finanziato dall'Unione europea
NextGenerationEU
Ministero dell'Università e della Ricerca
Italiadomani
PIANO NAZIONALE DI RIPRESA E RESILIENZA
Future Artificial Intelligence Research

Spoke 1: the research questions

1) **"human-in-the-loop" machine learning and reasoning**:

how humans and AI interact synergistically in complex (decision making) tasks (WP1.1-2-3)

2) **social-aware AI**:

how to understand and govern the societal outcomes of large-scale socio-technical systems of humans and AIs (WP1.4-5)

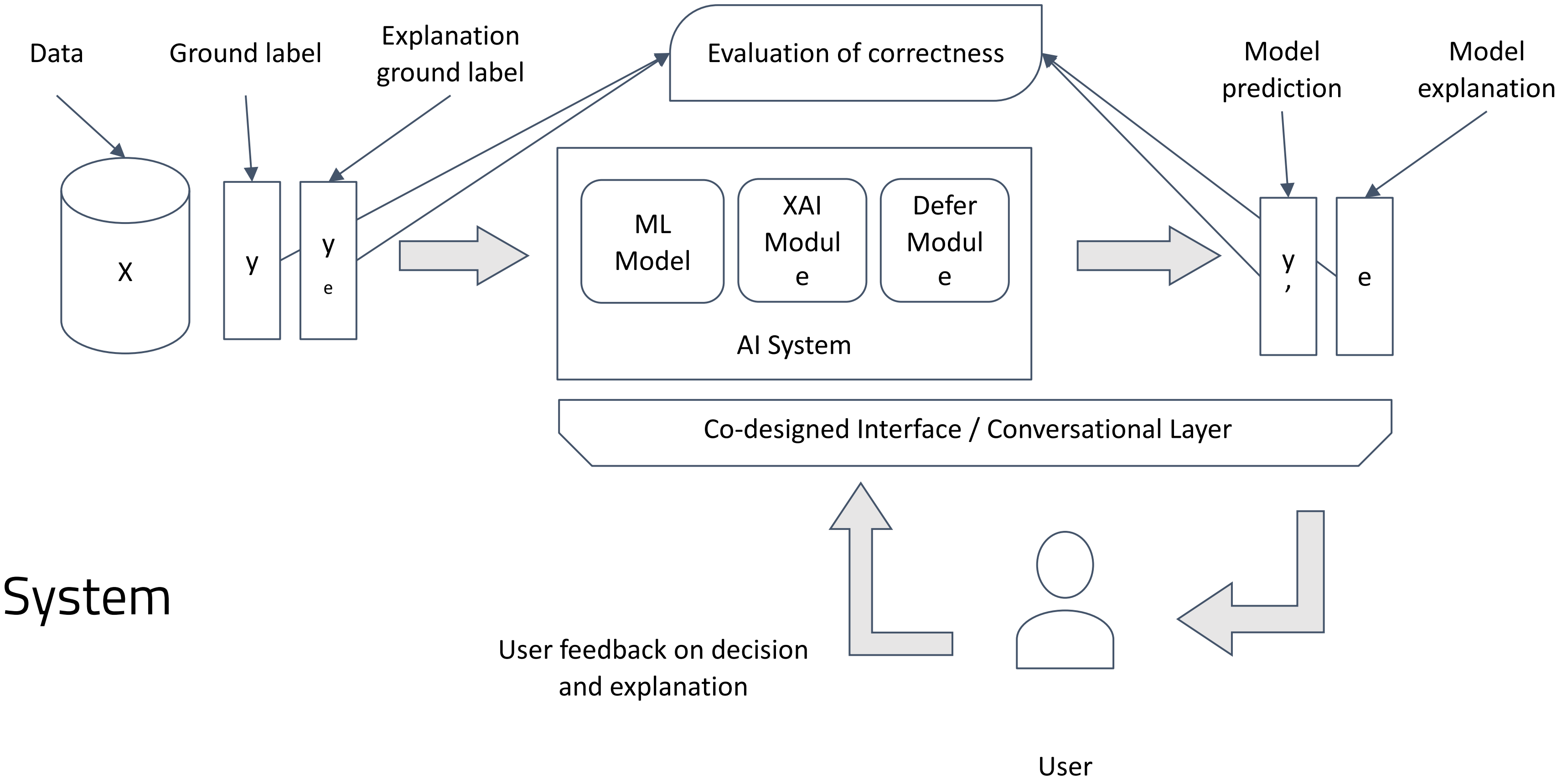
3) **design of trustworthy AI systems**:

how to the responsibly (co-)design, develop, validate and use trustworthy AI systems (WP1.6)

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

WP1.1 - Explainable AI for synergistic Human-AI collaboration

Task lead: SNS, co-PIs: Fosca Giannotti, Riccardo Guidotti.



Synergistic Interpretable
Hybrid Decision Support System

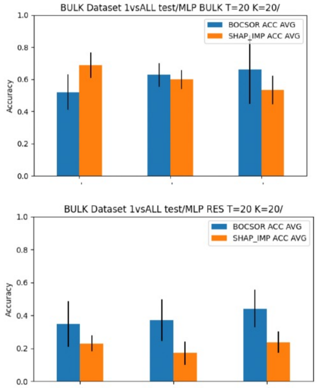
WP1.1 - Explainable AI for synergistic Human-AI collaboration

Task lead: SNS, co-PIs: Fosca Giannotti, Riccardo Guidotti.

Industrial Case Study

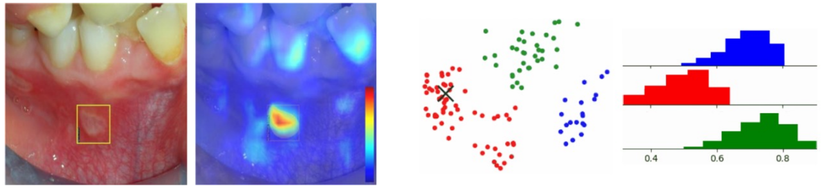
- Machines to manufacture tissue paper.
- Production features can be used to predict paper bulk and resistance.
- Experts provided importance of these features as high, medium, low.

Attribute	Units	Imp. for Bulk	Imp. for Res
ID	Integer	-	-
STRLA	N/m	LOW	MEDIUM
STRLO	N/m	LOW	MEDIUM
ELOLA	%	LOW	LOW
ELOLO	%	LOW	LOW
WEIGHT	gr/m^2	MEDIUM	MEDIUM
THICK	mm	-	LOW
TRH	ShA	LOW	LOW
BRH	ShA	-	LOW
COUPL	Category	MEDIUM	LOW
EMB	Category	MEDIUM	LOW
REW	Category	MEDIUM	LOW
ETR	Category	MEDIUM	MEDIUM
EBR	Category	MEDIUM	MEDIUM
TYPRO	Category	HIGH	HIGH
LAYERS	Integer	HIGH	HIGH
DRYRAT	Real	LOW	MEDIUM
STRCT	Boolean	HIGH	HIGH



Medical Case Study

- Case-Based Reasoning + ML models + XAI for decision support system
- Final decision taken by resident doctors or specialized ones depending on XAI impact.



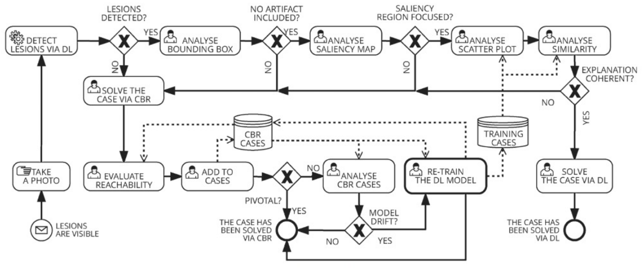


Figure 5 – Workflow design of the proposed method

WP1.2 – "System1 and System2" Machine Learning and Reasoning

Task lead: CNR, co-PIs: Umberto Straccia, Salvatore Ruggieri

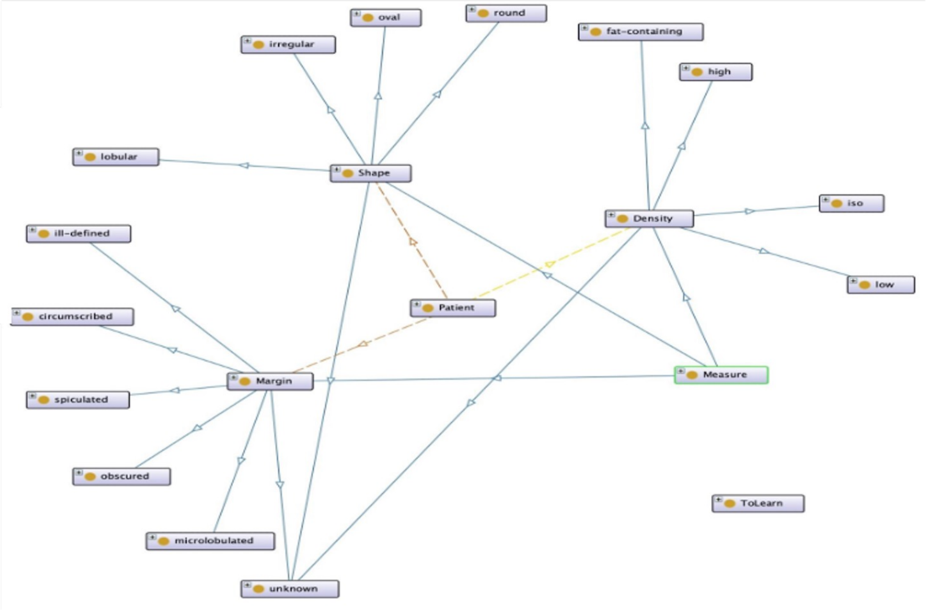
Ontology-based learning and reasoning

Mammography patient data

Patient	hasDensity	hasShape	hasMargin	hasBiRads	hasAge	...
p0	low	lobular	spiculated	5	67	...
p10	high	irregular	spiculated	5	76	...
p102	–	irregular	ill-defined	4	58	...
p108	low	round	circumscribed	4	57	...
p109	–	irregular	ill-defined	5	33	...
p110	low	irregular	ill-defined	4	45	...
p111	low	irregular	ill-defined	5	71	...
...	...	...	...	...	...	...

Learn Rules to Classify:

- **Does a patient have breast cancer?**



Mammography ontology

Example of learnt rule:

(hasMargin Ill-defined) AND (hasShape Irregular) AND (hasAge Old) ==> Cancer, 0.853

***IF** there is an image region whose margin is ill-defined **AND** whose shape is irregular **AND** the person is old*
***THEN** the mammography is about a tumor*

Finanziato dall'Unione europea
NextGenerationEU

Ministero dell'Università e della Ricerca

Italiadomani
PIANO NAZIONALE DI RIPRESA E RESILIENZA

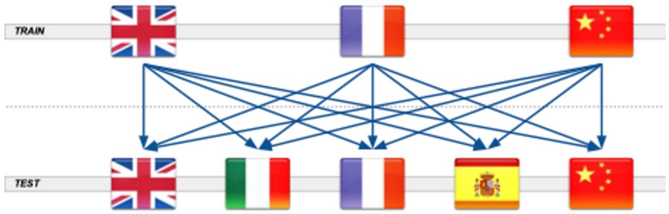
FAIR Future Artificial Intelligence Research

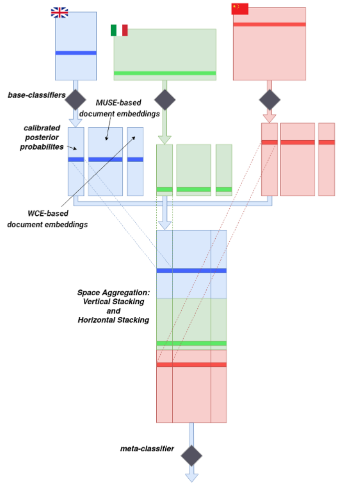
WP1.3 - Human centered Lifelong Learning for Complex Data

Task Lead: UNIPI, co-PIs: Alessio Micheli, Fabrizio Sebastiani

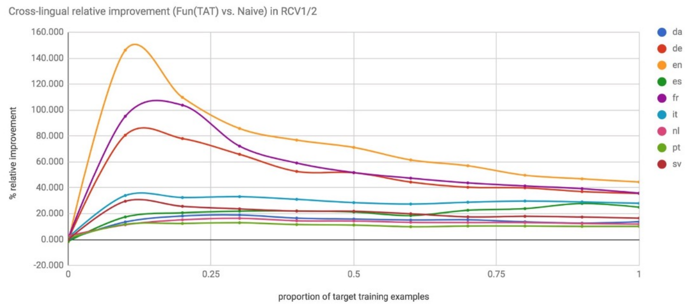
Cross-Lingual Text Classification (CLTC):

- Goal: leverage training data in a source language (e.g., EN) for classifying data in a target language (e.g., IT) when training data in the target language are scarce (zero-shot CLTC) or few (few-shot CLTC)





- Developed Generalized Funnelling, a “heterogeneous transfer learning” architecture that allows this
- In the zero-shot case, allows training a classifier for the target language
- In the few-shot case, radical improvements in classification accuracy
- Will be tested in “continual learning” scenarios



WP1.3 - Human centered Lifelong Learning for Complex Data

Task Lead: UNIPI, co-PIs: Alessio Micheli, Fabrizio Sebastiani

- Large Language Models (LLMs) exhibit impressive abilities on several tasks (despite not being directly trained on them)
- LLMs still face issues on domain-specific applications (e.g., legal, medical, financial etc.), which can be overcome with domain adaptation via further-pretraining and fine-tuning

WP1.4 - Human-AI Socio-technical Complex Systems

Task lead: UNIPI, co-PIs: Dino Pedreschi, Chiara Boldrini

Social networks & AI

- Echo chamber detection, detection of influential nodes, spot the presence of emerging collective phenomena, simulation of human and AI agents
- FOCUS: understand how AI-mediated information flows in a STS

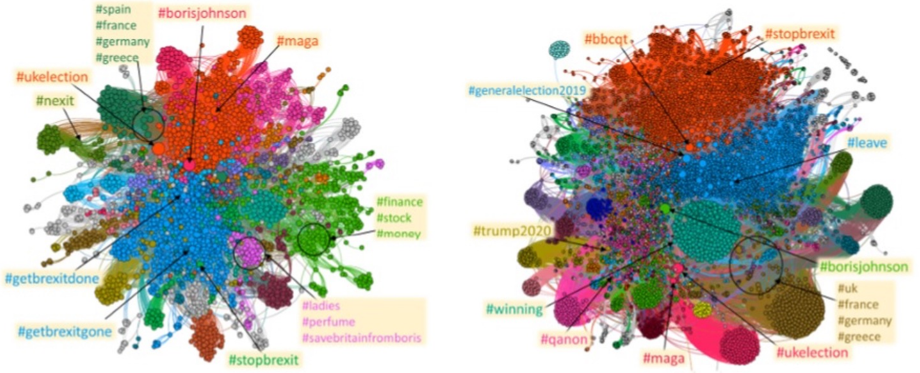


Figure 8: **The backbone of the network of hashtags and bots (left) or suspended users (right) linked by retweets for the whole period of our dataset.** Both networks show a modular structure probably due to the coordination of the automated users: the depicted partitions have a modularity of 0.73 (left) and 0.78 (right). In the bots' network the Brexit discussion appears together in the blue community, while for the suspended users two separate groups are pro-Euro (orange) and pro-Brexit (blue). In both cases, Trump-related hashtags are very common.

Finanziato dall'Unione europea
NextGenerationEU

Ministero dell'Università e della Ricerca

Italiadomani
PIANO NAZIONALE DI RIPRESA E RESILIENZA

Future Artificial Intelligence Research

WP1.4 - Human-AI Socio-technical Complex Systems

Task lead: UNIPI, co-PIs: Dino Pedreschi, Chiara Boldrini

Self-organization & individual vs collective goals in STS

- Impact of GPS navigation apps on urban emissions: egoistic vs collective approach

- Decentralized learning in a network of smartphones with no central server

Rome – 0% vs 100% routed with gmaps

	Method	**Avg accuracy**
Standalone baselines	Centralised	0.9824
	ISOL	0.6473
Distributed baseline	FED	0.9410
Decentralised SOTA	DecHetero	0.9071
	CFA	0.8975
	CFA-GE	0.9460
Decentralised Proposal	DecDiff+VT	**0.9530**

better than Federated!

Finanziato dall'Unione europea
NextGenerationEU
Ministero dell'Università e della Ricerca
Italiadomani
PIANO NAZIONALE DI RIPRESA E RESILIENZA
FAIR Future Artificial Intelligence Research

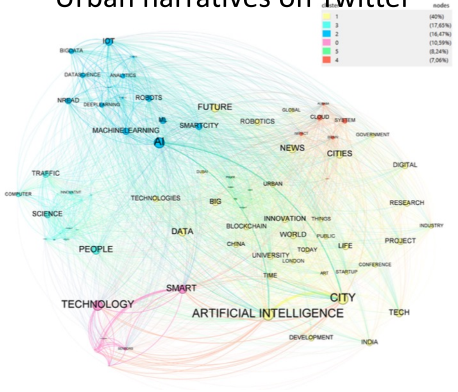
WP1.4 - Human-AI Socio-technical Complex Systems

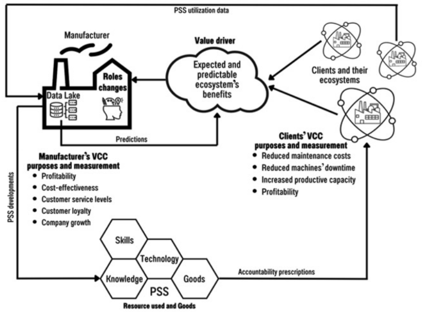
Task lead: UNIPI, co-PIs: Dino Pedreschi, Chiara Boldrini

Social and economic impact of AI-STS

- *AI & sustainable cities*
    - citizens' opinions on urban experiences and visions using big data and AI applications
- *AI & affordable and clean energy*
- *AI for business and economics*
    - interpretive framework to understand how machine learning (ML) affects the way companies interact with their ecosystem and how the introduction of digital technologies affects the value co-creation (VCC) process
    - how AI algorithms can be used to predict resolution of bankruptcy procedures

Urban narratives on Twitter

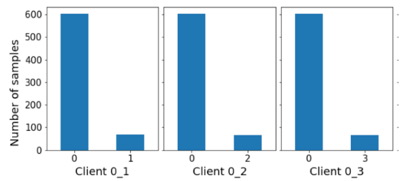




WP1.5 – Decentralized, Cooperative Learning

Task lead: CNR, co-PIs: Raffaele Perego, Francesco Marcelloni

Data distribution across nodes

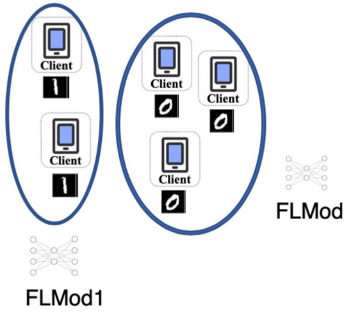


Learning on resource-constrained devices (small memory, battery-operated)

- **Energy-aware learning of neural networks at the edge**
  - We contributed a solution for memory-aware pruning of neural networks at learning time
  - Working on a novel "energy-aware" technique for NN pruning
    - Modeling the energy footprint of different neural network architectures
    - Pruning as the problem of finding the network that maximizes the accuracy with the smallest amount of energy used
    - Experiments on NVIDIA Jetson on a image classification task

- Anomaly Detection with Decentralised Unsupervised Federated Learning
  - System properties
    - *Local anomalies might not be global anomalies*: the "normal" behaviour for a node could be anomalous of another one;
    - *Data Locality*: Nodes can access only a limited portion of data;
    - Nodes have limited resources and train tiny ML models (e.g., Autoencoders) with limited capacity for identifying the anomalous patterns.
  - **Goal**: to go beyond the local representations by extending the generalisation capability of local models.
  - **How**: by identifying which other nodes share the same categories to initiate a federated collaboration for improving the local model

1. Find communities of peers
2. Perform per-community Federated Learning



WP1.5 – Decentralized, Cooperative Learning

Task lead: CNR, co-PIs: Raffaele Perego, Francesco Marcelloni

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

WP1.5 – Decentralized, Cooperative Learning

Task lead: CNR, co-PIs: Raffaele Perego, Francesco Marcelloni

docker

Deployment of components

- **Containers** as de-facto standard for Lightweight Virtualization
- Compliance with edge-computing / MEC-enabled environments

{ REST }

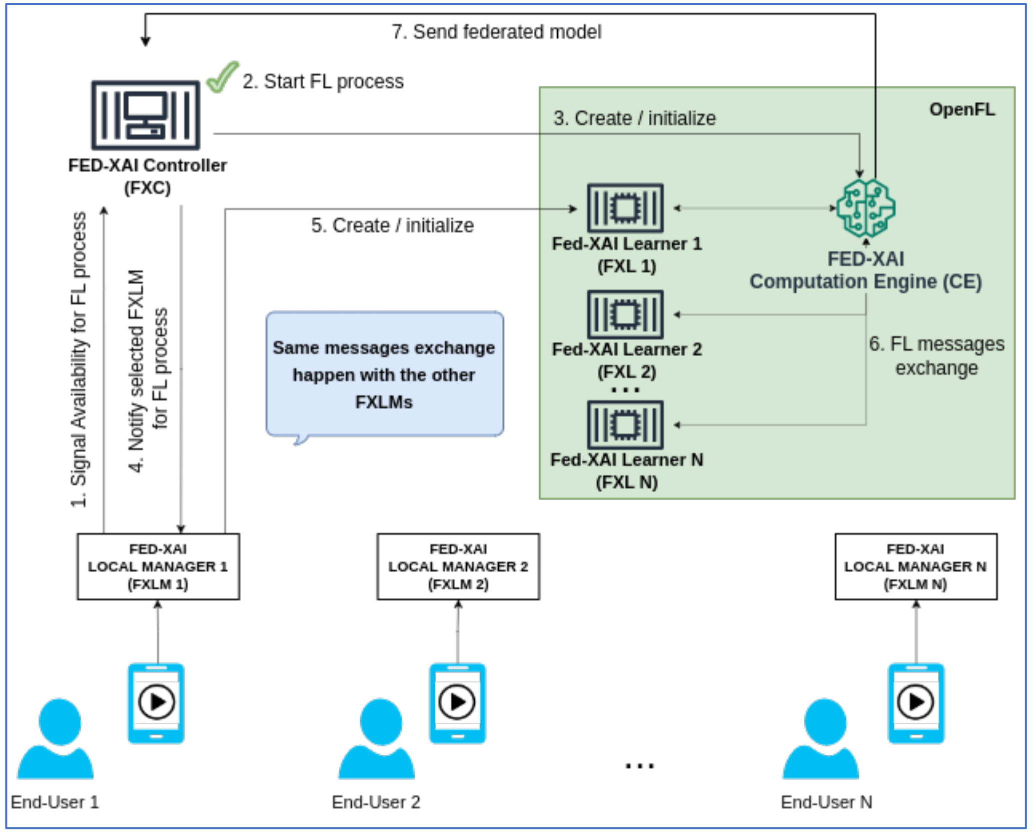
Message exchange

- **RestAPIs** for handling and integrating app microservices
- Over HTTPS: encryption for secure communication

intel/openfl

Federated Learning Framework

- Intel **OpenFL**
- Seamless integration with containers paradigm
- Extended to support FL of inherently interpretable models



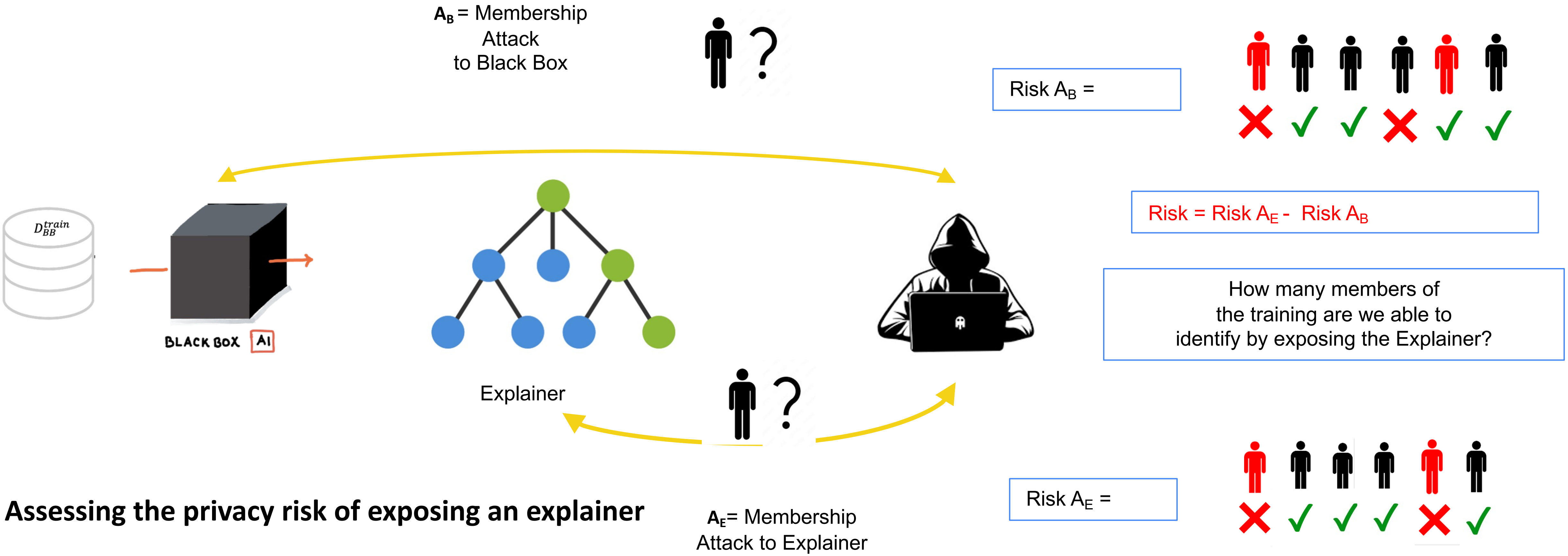
WP1.6 - Co-design methodologies for trustworthiness-by-design

Task lead: UNIPI, co-PIs: Adriano Fabris, Anna Monreale

- Development of **specific codes of conduct**, depending on the needs of companies, regarding the use in their areas of AI programs or devices
- **Ethics and legal counseling** with respect to compliance with European regulations (AI Act)
- **Advice regarding risk assessment** related to privacy compliance
- **Multidisciplinary Laboratory for the Study and Certification of Safe, Trustworthy, and Ethical AI Systems**

WP1.6 - Co-design methodologies for trustworthiness-by-design

Task lead: UNIPI, co-PIs: Adriano Fabris, Anna Monreale



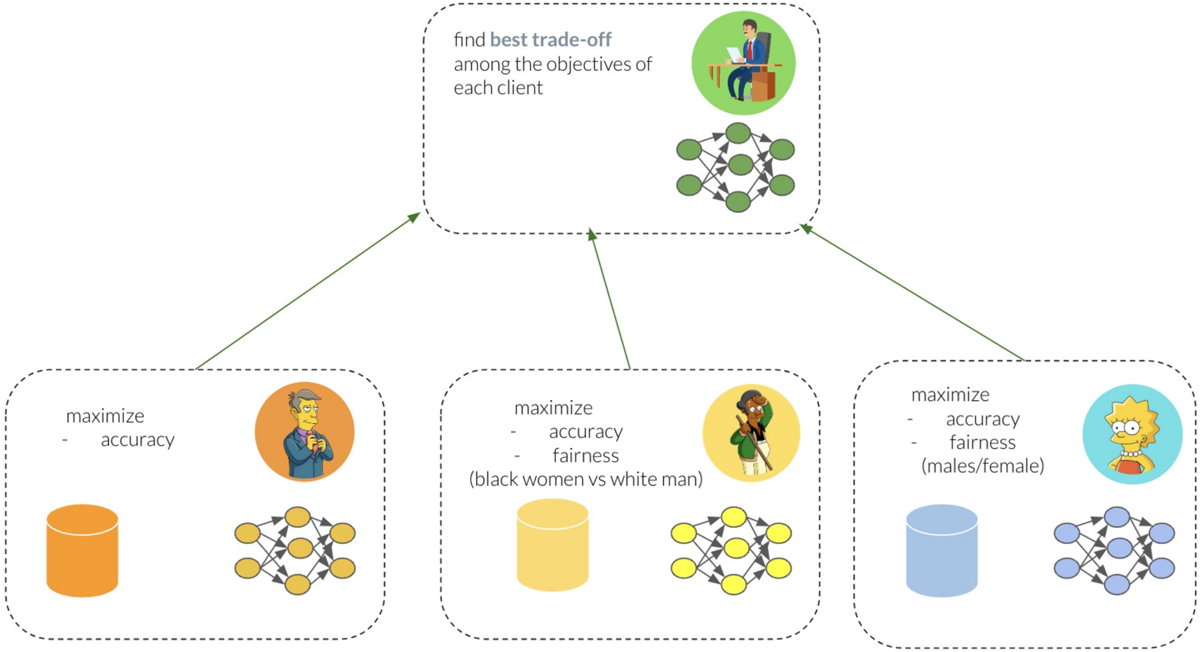
Assessing the privacy risk of exposing an explainer

Finanziato dall'Unione europea NextGenerationEU
Ministero dell'Università e della Ricerca
Italiadomani PIANO NAZIONALE DI RIPRESA E RESILIENZA
FAIR Future Artificial Intelligence Research

WP1.6 - Co-design methodologies for trustworthiness-by-design

Task lead: UNIPI, co-PIs: Adriano Fabris, Anna Monreale

GLOFAIR (Global-Local Optimization for Fairness in Federated Learning)



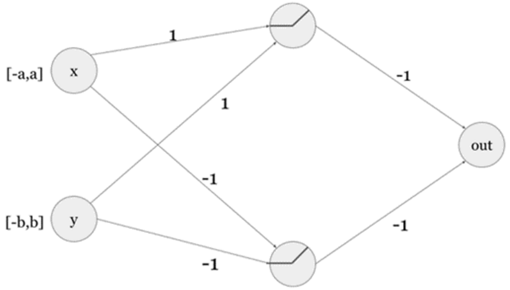
Designing **collaborative learning** where:

- each predictive local model is the result of a **multi-objective** learning process
- each local client has its **own** set of objectives (fairness, utility, …)
- Train the models without modifying or moving the private training data

Finanziato dall'Unione europea
NextGenerationEU
Ministero dell'Università e della Ricerca
Italiadomani
PIANO NAZIONALE DI RIPRESA E RESILIENZA
FAIR
Future Artificial Intelligence Research

WP1.6 - Co-design methodologies for trustworthiness-by-design
Task lead: UNIPI, co-PIs: Adriano Fabris, Anna Monreale

Predictability and reliability of multi-layer perceptron models for non-linear regression applications
Example: bounds on the output of a cruise controller for Advanced Driver Assistance Systems (ADAS)



x ∈ [-a, a]

y ∈ [-b, b]

relu(x) = x if x >0, 0 otherwise

⇩

f(x,y) >= - (a+b)

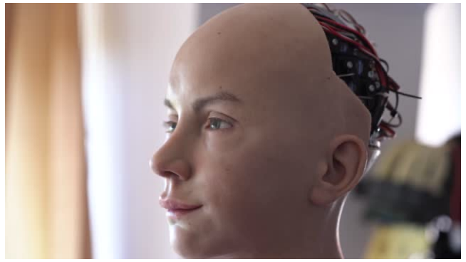
```
gfcnn [a: { x: real | x>0 }, b: { x: real | x>0 }]: THEORY
  BEGIN

    xinreal: TYPE = { r: real | r>=-a AND r<=a}
    yinreal: TYPE = { r: real | r>=-b AND r<=b}

    relu (x: real): real = IF x > 0 THEN x ELSE 0 ENDIF

    input_neuron (x: xinreal, y: yinreal): real = relu(x+y)
    hidden_neuron (x: real, y: real): real = relu(x+y)
    output_neuron (x: real, y: real): real = -(x+y);

    network (x: xinreal, y: yinreal): real =
output_neuron(input_neuron(x,y),input_neuron(-x,-y))

    network_bounds: THEOREM
       FORALL (x:xinreal, y: yinreal): network(x,y) >= -(a+b)

 END gfcnn
```

WP1.7 – Empirical studies and pilots of human centered AI

Task lead: UNIPI, co-PIs: Francesco Marcelloni, Raffaele Perego, Salvatore Ruggieri, Dino Pedreschi

Industrial applications

- Increase **workplace safety** through sensors and artificial intelligence (in collaboration with INAIL)
- Extracting and identifying **engineering relations** between Engineering Design entities from technical documents
- Extraction of **flowcharts** from patent descriptions and images
- **Reducing the gap between the academic world and the job market** by improving the matching between job advertisements and offerings
- Evaluation of an expressive social robot cognitive system by means of **natural human-robot interaction and dialogue**



Finanziato dall'Unione europea
NextGenerationEU

Ministero dell'Università e della Ricerca

Italiadomani
PIANO NAZIONALE DI RIPRESA E RESILIENZA

FAIR
Future Artificial Intelligence Research

WP1.7 – Empirical studies and pilots of human centered AI

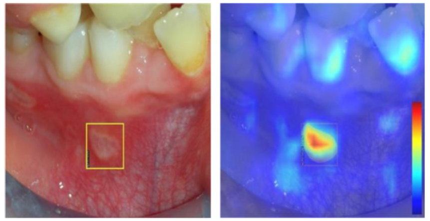
Task lead: UNIPI, co-PIs: Francesco Marcelloni, Raffaele Perego, Salvatore Ruggieri, Dino Pedreschi

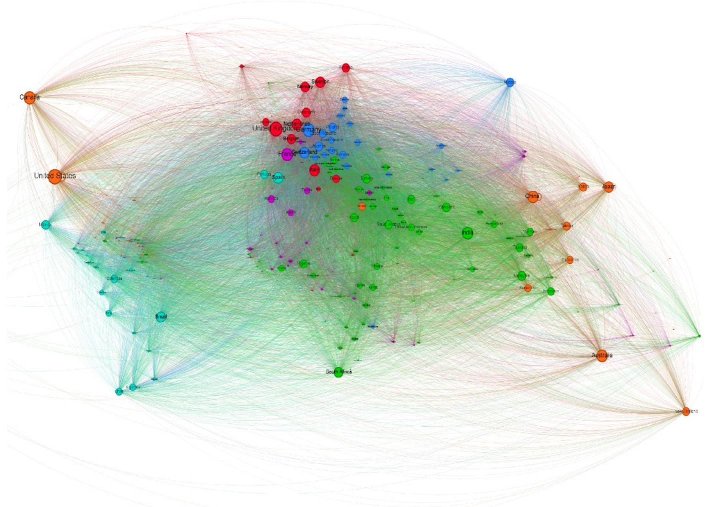
Medical Applications

- Oral squamous cell carcinoma detection and classification
- Analysis of the effects of different diets on 24-hour core body temperature and energy metabolism

Society

- Analysis of the impact of AI systems in territorial and urban contexts, along with other socio-economic and environmental factors
- Exploring urban mobility through GeoAI





Finanziato dall'Unione europea NextGenerationEU
Ministero dell'Università e della Ricerca
Italiadomani PIANO NAZIONALE DI RIPRESA E RESILIENZA
FAIR Future Artificial Intelligence Research

WP1.7 – Empirical studies and pilots of human centered AI

Task lead: UNIPI, co-PIs: Francesco Marcelloni, Raffaele Perego, Salvatore Ruggieri, Dino Pedreschi

Archaeology

- **Urban Archaeology**: AI-controlled robotic arm for classification and selection of archaeological ceramic fragments
- **Environmental Archaeology**: AI approaches for animal bones recognition and landscape reconstruction

- **Sub-AI**: Project to interpret, digitally share, and document underwater archaeological heritage, as well as determine its ecological effects on local biological communities
- **Contemporary Archaeology**: Digital traces of migration in Lampedusa