

Towards Responsible Data Science: Insights from EU-funded Research

Francesca Pratesi

Institute of Science and Technology, Italian National Council of Research

francesca.pratesi@isti.cnr.it

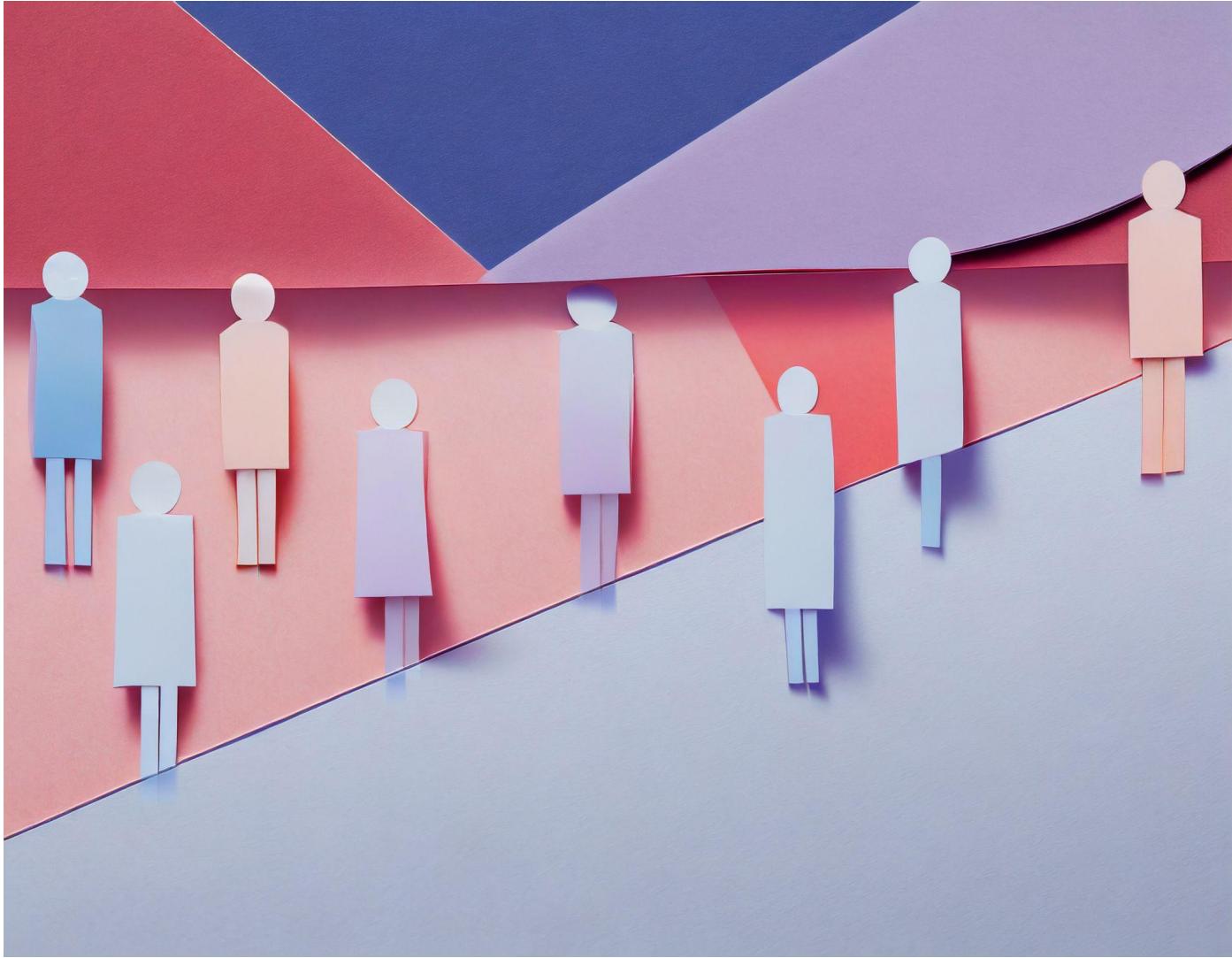
Agenda: The point of view of two EU-funded projects

- SoBigData RI
 - EU principles for Open Science
 - New generation of Data Scientists
- TAILOR
 - Research applied to the various Ethical dimensions



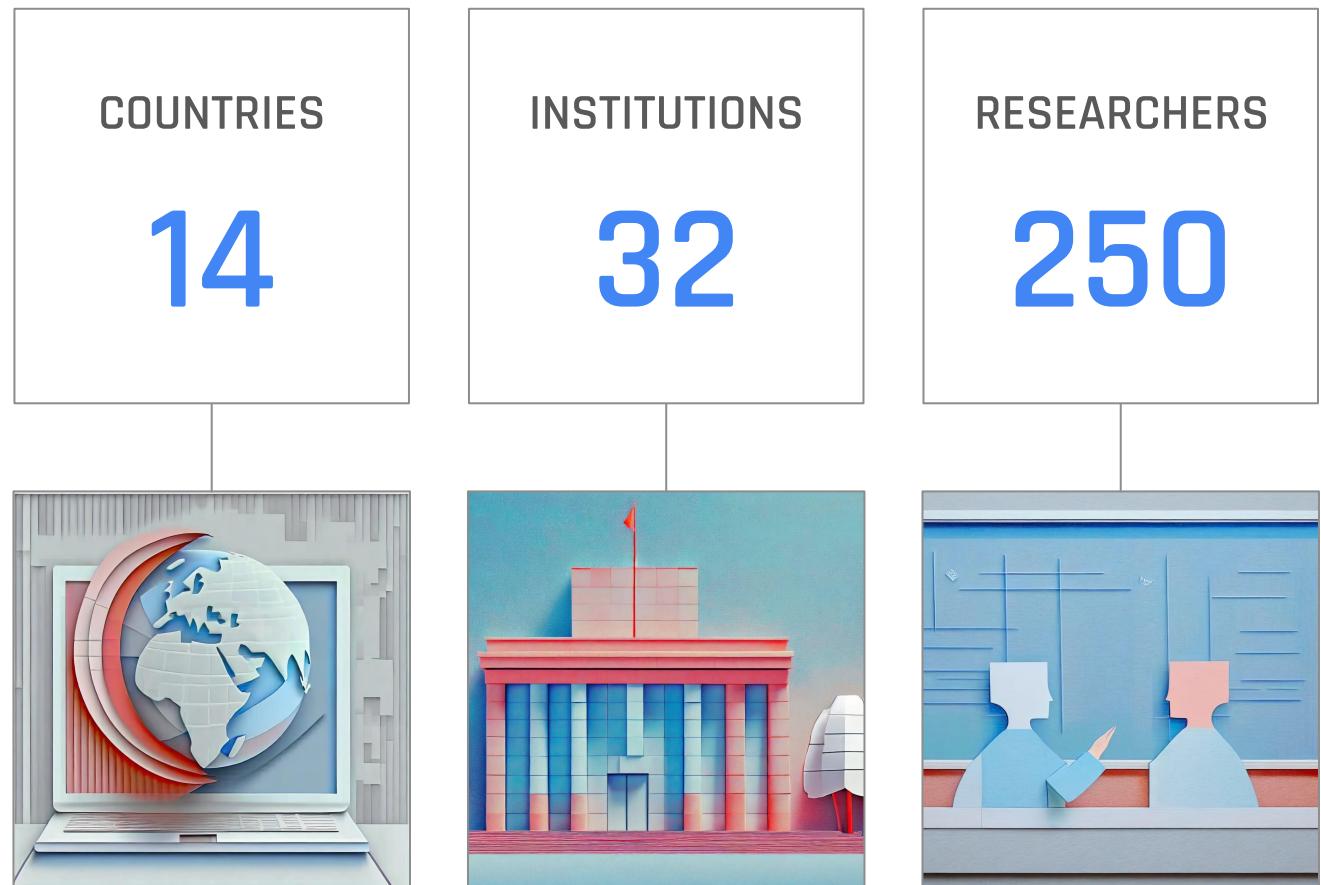


Integrated Infrastructure for Social Mining and Big Data Analytics



What is SoBigData RI

A distributed, Pan-European,
multi-disciplinary research
infrastructure aimed at using **social**
mining and **big data** to understand the
complexity of our contemporary, globally
interconnected **society**



Istituto di Scienze e Tecnologie
dell'Informazione "A. Faedo"
Consiglio Nazionale delle Ricerche



The three pieces of SoBigData RI

EU PROJECTS
3

RI



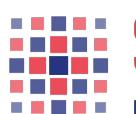
ESFRI
Roadmap



Italian
Node
(PNRR)



 **TAILOR**

 **SOBIGDATA**
RESEARCH INFRASTRUCTURE

 **KDD** Lab



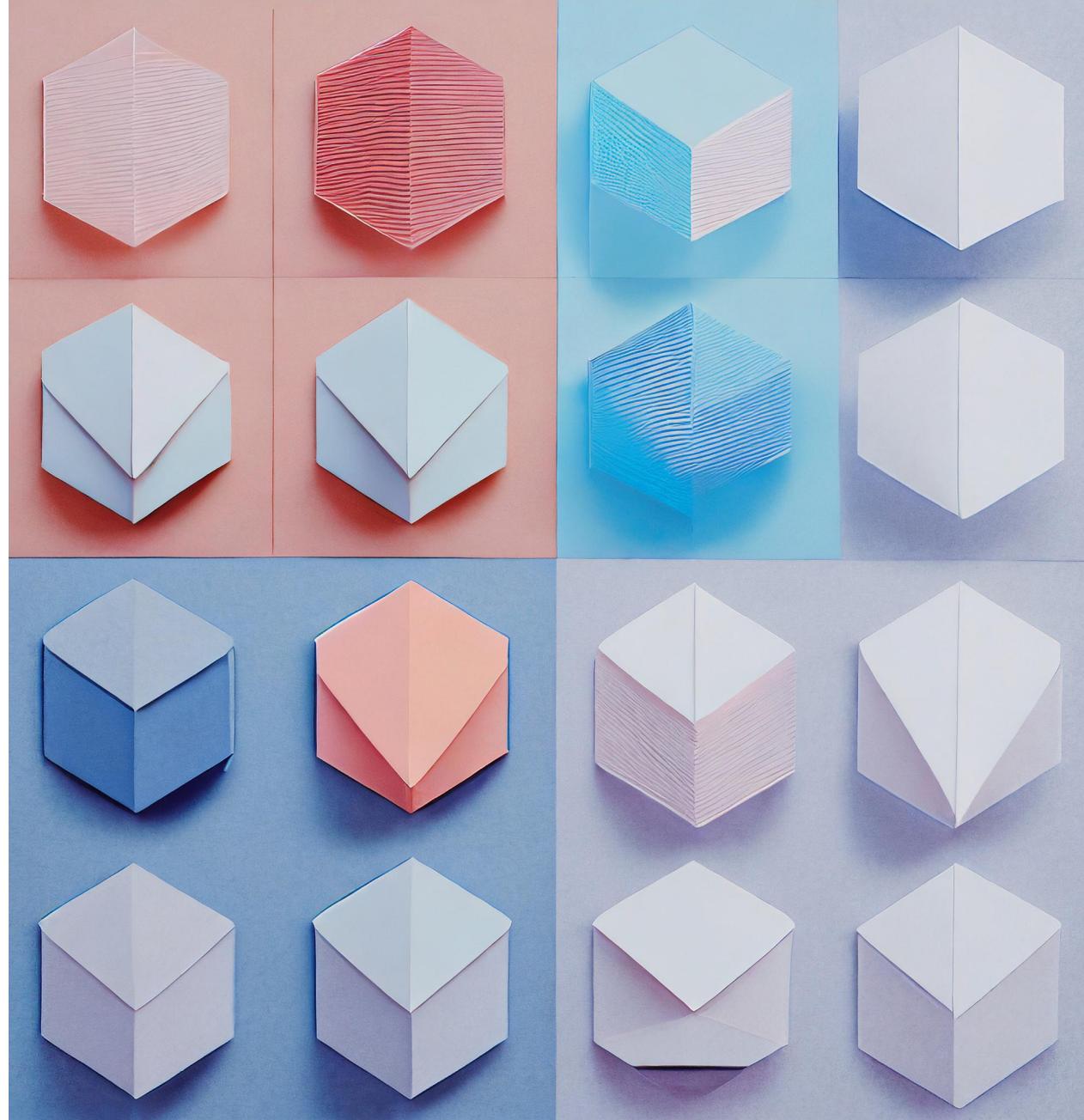
Istituto di Scienze e Tecnologie
dell'Informazione "A. Faedo"
Consiglio Nazionale delle Ricerche



Why SoBigData RI

Responds to the rising demand
for cross-disciplinary data-driven
research and innovation

- **Democratising the benefits of data science** and Big Data within an ethical framework that harmonizes individual rights and collective interest
- Focus on **Social Mining** to understand and model complex social phenomena
- **Open Data ecosystem:** adequate means for accessing big social data together with algorithms for extracting knowledge from them



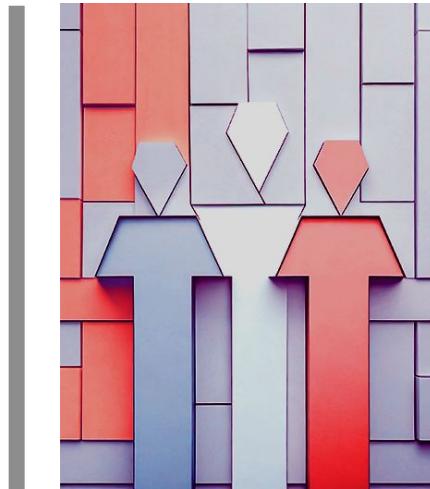
SoBigData RI Overview

INFRASTRUCTURE



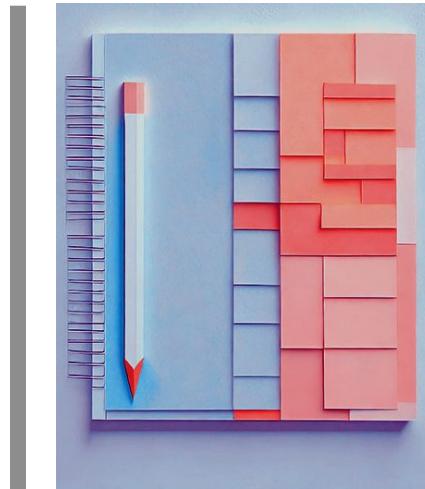
- Computational Power
- Open Science tools
- Resources
- Datasets
- Methods
- Technologies
- Training Materials

COMMUNITY BUILDING



- Researcher communities
- Industries and PA involvement
- Connecting side projects
- Connecting to other realities

GUIDELINES



- Ethical aspects
- Legal Aspects
- Open Science



Istituto di Scienze e Tecnologie
dell'Informazione "A. Faedo"
Consiglio Nazionale delle Ricerche



Virtual Access

SoBigData Catalogue

- Methods,
- Libraries,
- Applications,
- Data

The screenshot shows the SoBigData Catalogue interface. At the top, there are navigation links: 'Items', 'Activity Stream', and 'About'. Below this is a search bar with placeholder text 'Search items...' and a magnifying glass icon. A checkbox labeled 'Include Sub-Organizations' is present. A dropdown menu 'Order by:' is set to 'Relevance'. The main area displays a list of items:

- MANILA**: Described as a low-code web application for machine learning fairness evaluations.
- Twitter Newcomers Dataset**: Twitter accounts detected right after registration and monitored for 21 days, available as a ZIP file.
- 64-tiles tessellation of Manhattan**: A dataset.

On the left sidebar, there is a summary of user statistics: 'Followers' (14) and 'Items' (384), with a 'Follow' button. On the right sidebar, there is a section titled 'SoBigData Services and Products' with a link to the project site: <http://www.sobigdata.eu/read-more>.

SoBigData Lab

- Jupyter Hub
- Data Space
- Execute an Experiment

The screenshot shows the SoBigData Lab interface. It features a large 'jupyter' logo at the top. Below it are three main sections:

- Default Standard - 2GB RAM / 2 cores**: Includes a description of the server setup and a link to 'Saltata in modalità presentazione'.
- Default Small - 4GB RAM / 4 cores**
- Default Medium - 8GB RAM / 4 cores**
- Master in Big Data Analytics & Social Mining notebook server - 2GB RAM / 2 cores**: A note about its preconfiguration for the Master course.

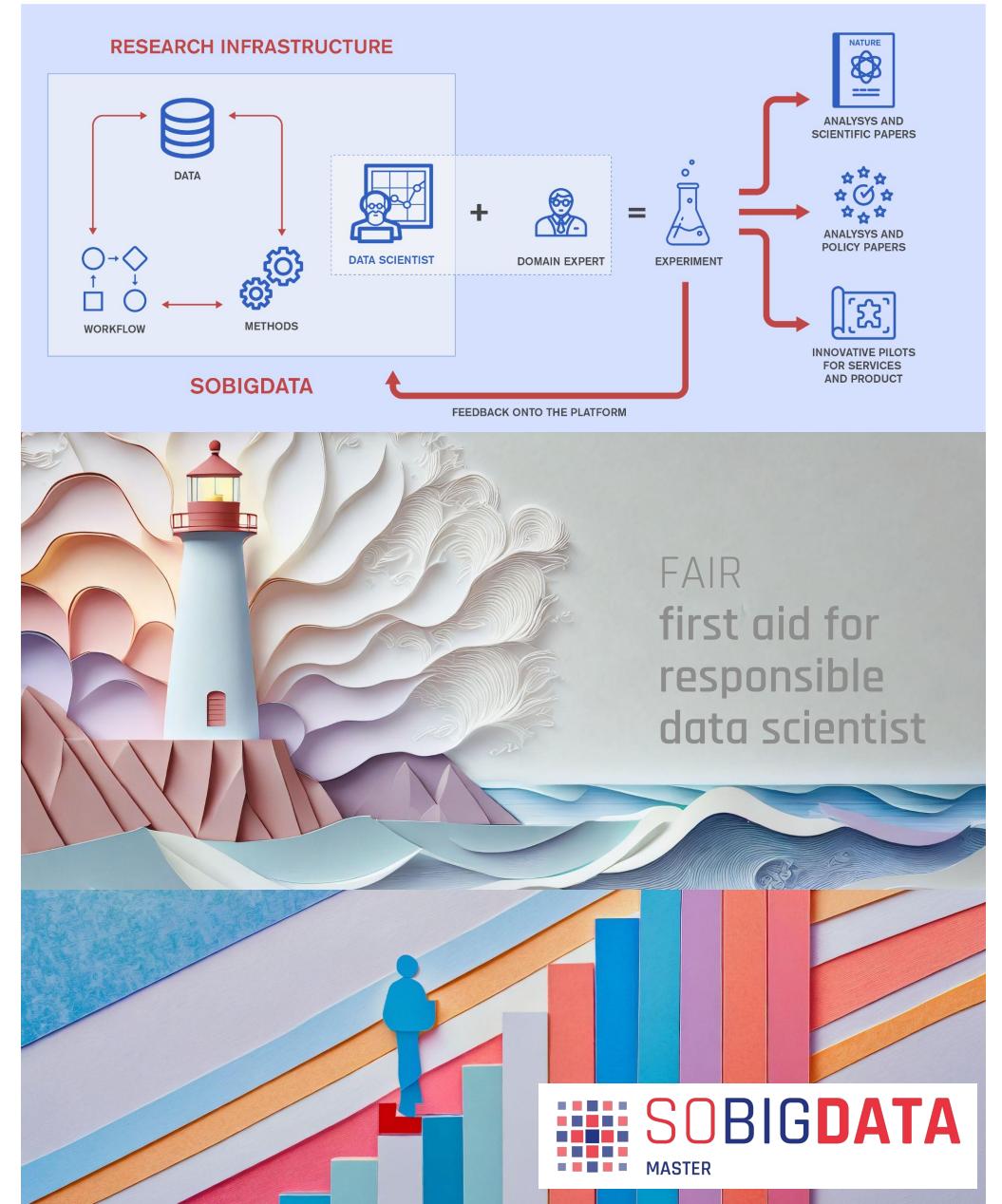
A large orange 'Start' button is located below these options. Further down are sections for 'Access to the Data Space', 'Execute an Experiment', and 'Check the Computations'. To the right, there is a grid of icons representing various research domains: Social Demands, Demography, Economy & Finance 2.0, Sustainable Cities for Citizens, Migration Studies, Sport Data Science, Social Impacts of AI and Explainable Machine Learning, Text and Social Media Mining (TSMM), Complex Network Analysis (CNA), Human Mobility Analytics (HMA), Web Analytics (WA), Visual Analytics (VA), and Privacy Enhancing Technology (PET).



Focus

Responsible data science principles

- To act as a Wind Tunnel for innovative and responsible data science science: **a place to find data, methods, experiment**, and to share ideas and learn big data data analysis.
- **Responsible open Data Science is FAIR + FACT** (ethics by design)
- An “operational” Ethical and Legal framework designed to function as a factory for sensing, understanding ethico-legal issues and data literacy in digital society
- Training Events organized by the RI (Master, Summer Schools, Datathon, Awareness Panel, etc)



Istituto di Scienze e Tecnologie
dell'Informazione "A. Faedo"
Consiglio Nazionale delle Ricerche



FAIR principles

Findable

- The first step in (re)using data is to find them.
- Metadata and data should be easy to find for both humans and computers.

Accessible

- The user need to know how data can be accessed, possibly including authentication and authorisation.

Interoperable

- The data usually need to be integrated with other data.
- The data also need to interoperate with applications or workflows for analysis, storage, and processing.

Reusable

- Metadata and data should be well-described so that they can be replicated and/or combined in different settings.



Istituto di Scienze e Tecnologie
dell'Informazione "A. Faedo"
Consiglio Nazionale delle Ricerche



FACT principles

Fair

- Could data be without prejudice?
- How to avoid unfair conclusions even if they are true?

Accurate

- Could data be without guesswork?
- How to answer questions with a guaranteed level of accuracy?

Confidential

- How to answer questions without revealing secrets?

Transparent

- How to clarify answers such that they become indisputable?



Istituto di Scienze e Tecnologie
dell'Informazione "A. Faedo"
Consiglio Nazionale delle Ricerche



First Aid for Responsible data scientist

TRY IT ON:
fair.sobiqdata.eu/moodle

A blue umbrella with yellow stars and the text "GDPR" on it, set against a background of binary code.



The SoBigData online course developed to ensure that people are familiar with the basic elements about ethics, data protection, and intellectual property law

First Aid for Responsible Data Scientists

Welcome to FAIR

Here you can find the list of available courses. These courses should be attended by every new user of the SoBigData Research Infrastructure, but they can be interesting also for other purposes. For each course, you will find symbols that suggest the potential users interested in it.

 researcher  company  student

Available courses



First Aid for Responsible Data Scientists

The data scientist using the SoBigData infrastructure has the responsibility to get acquainted with the fundamental ethical aspects relating to his/her research and to be aware that using the data he/she can find here will make him/her a "data controller".

The course purpose is to show you the responsibilities, the possibilities and the boundaries you have in data manipulation.





First Aid for GDPR & Data in Business

The course purpose is to show you the specific impact the General Data Protection Regulation will have on companies' duties and how these European companies should comply with the limits in data manipulation the Regulation requires.





First GDPR Aid for Researchers

The course purpose is to show you the specific impact the General Data Protection Regulation will have on researchers' duties and to highlight articles and recitals which specifically mention and explain how research is intended in GDPR's legal system.



- 3 core lessons:
 - ethics
 - data protection law
 - intellectual property law
- 2 additional lessons:
 - GDPR for researchers
 - GDPR for industries
- 57 questions
 - (3 about ethics, 50 about DP, 4 about IP)
- 2 case studies
 - (about real sentences)
- THE GOAL OF THE COURSES
IS TO BRING AWARENESS
AND INFORMATION TO CITIZENS, COMPANIES, INSTITUTIONS
AND RESEARCHERS



- FAIR - First Aid for Responsible Data Scientists by SoBigData

- First Aid for Responsible Data Scientists
- Participants
- Badges
- Competencies
- Home
- Dashboard
- Calendar
- Private files
- My courses
- First Aid for Responsible Data Scientists

UNIT 1 – ETHICAL FRAMEWORK FOR DATA SCIENTIST



UNIT 2- A JOURNEY THROUGH THE NEW EUROPEAN DATA PROTECTION LAW



UNIT 3 – INTELLECTUAL PROPERTY RIGHTS AND SOCIAL MEDIA CONTENT



MY COURSES

- First Aid for Responsible Data Scientists
- All courses ...

COURSE SUMMARY

The data scientist using the SoBigData infrastructure has the responsibility to get acquainted with the fundamental ethical aspects relating to his/her research and to be aware that using the data he/she can find here will make him/her a "data controller".

The course purpose is to show you the responsibilities, the possibilities and the boundaries you have in data manipulation.

-  First Aid for Responsible Data Scientists
-  Participants
-  Badges
-  Competencies
-  Home
-  Dashboard
-  Calendar
-  Private files
-  My courses
-  First Aid for Responsible Data Scientists

First Aid for Responsible Data Scientists

[Home](#)[My courses](#)[First Aid for Responsible Data Scientists](#)[Unit 1- Ethical Framework for Data Scientist](#)[Ethical Framework for Data Scientist](#)

Ethical Framework for Data Scientist

Data Ethics Basics

LET'S START FROM BASICS!

The transformation Social Science is going through leads the data scientist to have at disposal enormous quantities of data. As a data scientist, you can now have enormous quantities of data at your disposal. Often the data used is personal data: data related to individual people.

It is fundamental to remember that in many cases this data are not consciously given by participants to researchers, but **passively** generated as a byproduct of daily activities, such as:

-  use of a mobile phone

-  making a paying in a shop

-  posting something on social media

[Previous](#)[Next](#)

You have completed 6% of the lesson

LESSON MENU

- Welcome!
- SoBigData ethics policy
- Data Ethics importance
- Data Ethics Basics
- Ethical considerations
- Ethics and use of personal data
- Ethics in SoBigData
- Ethics in SoBigData
- Ethics in SoBigData
- Principles
- 1 - Create awareness
- 2 - Critical questions
- 3 - Trade-offs public discussions
- 4 - Privacy preserving techniques
- 5 - Inform the data subjects
- The End



First Aid for Responsible Data Scientists

[Home](#)[My courses](#)[First Aid for Responsible Data Scientists](#)[Unit 2- A journey through the new European Data Protection Law](#)[Unit Quiz Number 2!](#)

Question 9
Incorrect
Marked out of 1
[Flag question](#)

Is pseudonymisation the only measure of data protection allowed by the GDPR?

Select one:

- Yes, because it helps controllers and processors to meet their data protection obligations 
- Your answer is incorrect. Remember what we say about [data protection law principles](#) and the measures implemented by this RI. You can also read the [GDPR Recital 28](#) about the use of pseudonymisation.
- Yes, because it is the only way to protect sensitive data
- No, the GDPR allows any safeguards method

Question 10
Not complete
Marked out of 1
[Flag question](#)

According to the General Data Protection Regulation, is it possible to transfer [personal data](#) outside the EU to third countries?

Select one:

- It is possible, but an adequate level of protection must be guaranteed
- It is possible only in case the Authority of the Member State have a specific regulation
- It is possible only in case the [data subject](#) gave a specific consent

[Check](#)

[Finish attempt ...](#)

QUIZ NAVIGATION



[Finish attempt ...](#)

EU Legislation for protection of personal data

- European Law:
 - Data protection directive (95/46/EC)
 - ePrivacy directive (2002/58/EC)
 - GDPR - current EU Regulation
 - AI Act (Proposed: April 2021)
- Opinions by EU Article 29 Data Protection Working Party
- Ethics Guidelines for Trustworthy AI - High Level Expert Group on AI

Lawful AI

Ethical AI

Robust AI



TAILOR – Unique Selling Point

Actively **bringing together** communities,
especially in **reasoning and learning**,
in an **academic-industrial** network
with the **vision** and **capability**
of developing the **scientific foundations**
for realising the **European vision** of
human-centred **Trustworthy AI**.



Istituto di Scienze e Tecnologie
dell'Informazione "A. Faedo"
Consiglio Nazionale delle Ricerche



A European approach to Artificial Intelligence

.....

A STRATEGY FOR EUROPE TO LEAD THE WAY

Boost
technological
and industrial
capacity
& AI uptake

Prepare for
socio-
economic
changes
Jobs / Skills

Ensure an
appropriate
ethical & legal
framework

AI FOR GOOD AND FOR ALL



European
Commission



TAILOREU

SOBIGDATA
RESEARCH INFRASTRUCTURE

KDD
Lab



Istituto di Scienze e Tecnologie
dell'Informazione "A. Faedo"
Consiglio Nazionale delle Ricerche



Ethics Guidelines for Trustworthy AI – Overview

Human-centric approach: AI as a means, not an end

Trustworthy AI as our foundational ambition, with three components

Lawful AI

Ethical AI

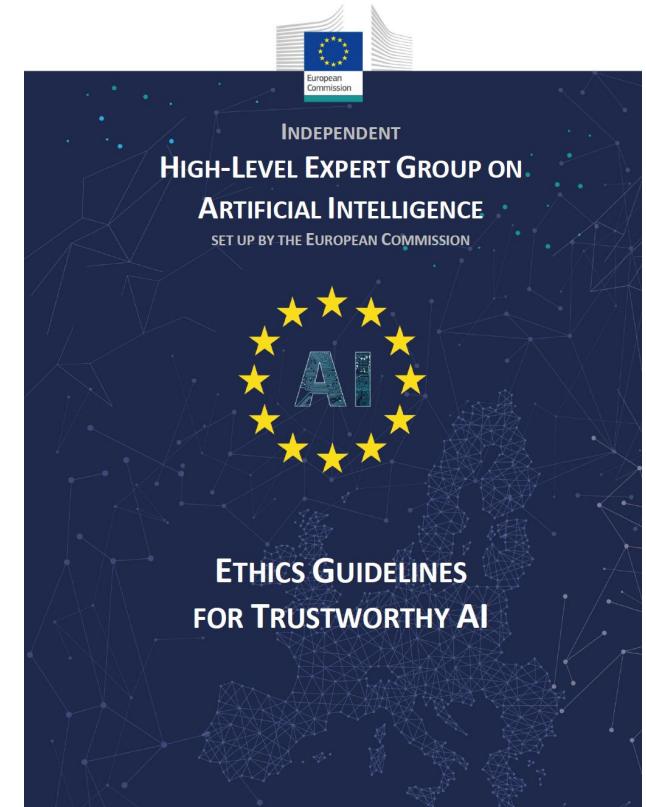
Robust AI

Three levels of abstraction

from principles
(Chapter I)

to requirements
(Chapter II)

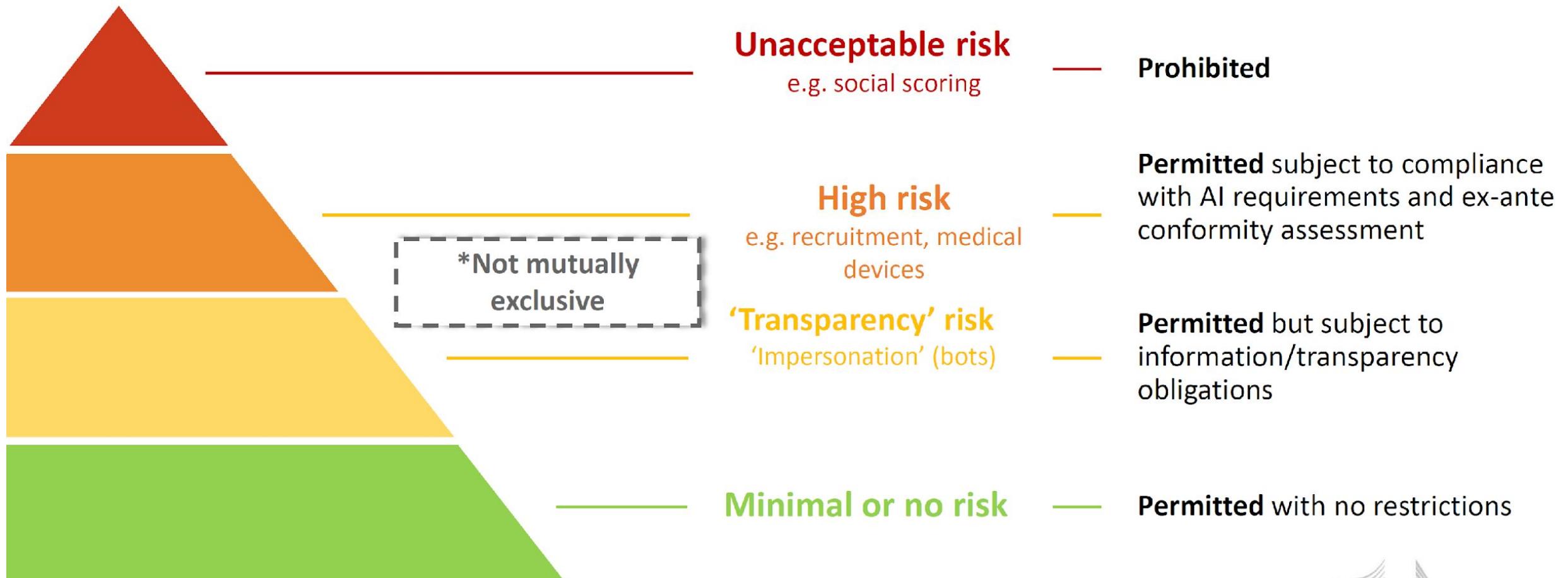
to assessment
list (Chapter III)



<https://ec.europa.eu/digital-single-market/en/news/ethics-guidelines-trustworthy-ai>

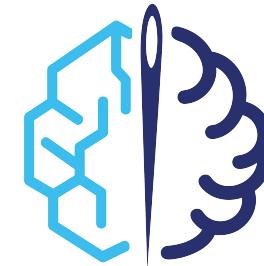


The AI Act: the new EU Regulation on AI



TAILOR

**Foundation of Trustworthy AI:
Integrating Learning, Optimisation and
Reasoning**



**Overview of TAILOR
Francesca Pratesi**

Boosting Capacity to Tackle Major Scientific Challenges

- *How will you boost the research capacity of the AI excellence research centres in Europe by tackling major scientific or technological challenges and by putting in place mechanisms for more efficient collaboration?*
- A **core network** of outstanding AI research centres and major European companies (partners) plus **mechanisms for extending** the network (network members and connectivity fund) to be adaptive and inclusive.
- Five **virtual research environments** to address the **major scientific challenges** required to achieve Trustworthy AI supported by **AI-based** network **collaboration tools**.
- **Strategic** research and innovation **roadmap** to drive the long-term **scientific vision** combined with **bottom-up coordinated actions** collaboratively addressing specific research questions.

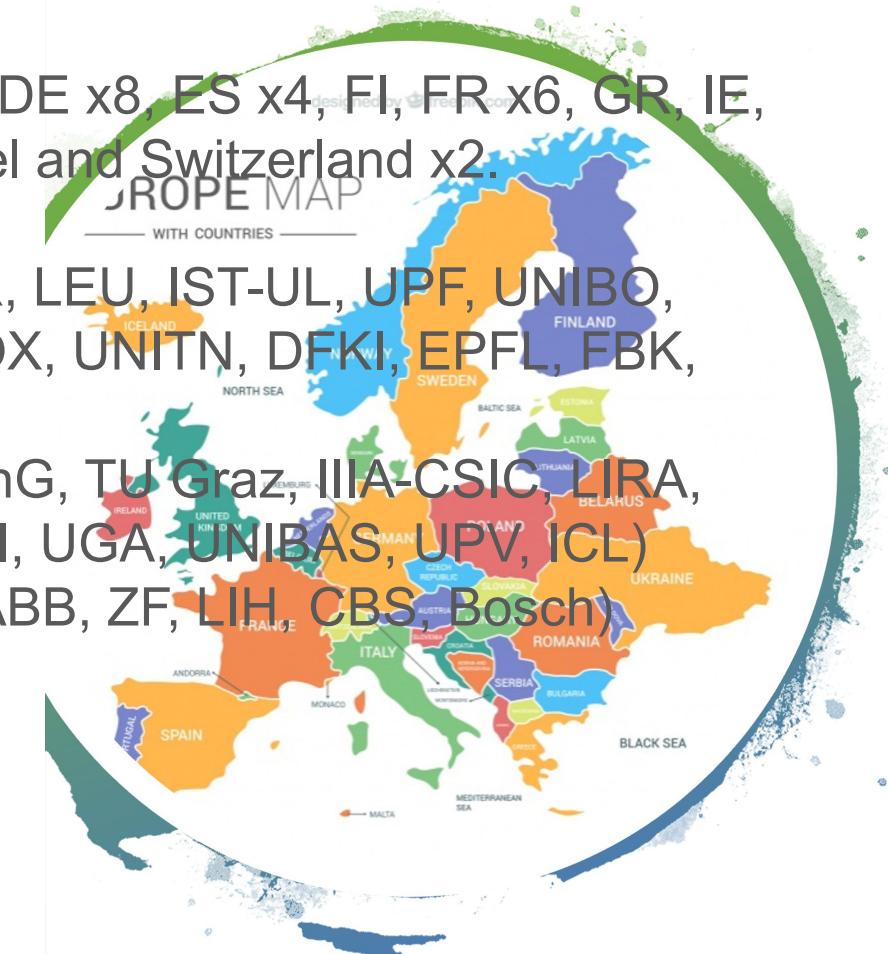


Istituto di Scienze e Tecnologie
dell'Informazione "A. Faedo"
Consiglio Nazionale delle Ricerche



TAILOR Consortium

- 54 partners from 18 EU countries (AT, BE x2, CZ x2, DE x8, ES x4, FI, FR x6, GR, IE, IT x8, LU, NL x6, PL, PT, SE x2, SI, SK, UK x4), Israel and Switzerland x2.
- More than 60 network members.
- 23 Core partners (LiU, CNR, INRIA, UCC, KUL, UOR, LEU, IST-UL, UPF, UNIBO, BIU, TUE, CNRS, JSI, TUDA, UNIBRIS, ALU-FR, UOX, UNITN, DFKI, EPFL, FBK, CINI)
- 21 Partners (VUB, CUNI, CEA, CRIL, CVUT, TUD, FhG, TU Graz, IIIA-CSIC, LIRA, UOA, NEO-UMA, PUT, RWTH, slovak.AI, TNO, UniPI, UGA, UNIBAS, UPV, ICL)
- 10 Industry partners (VW, ENG, Tieto, Philips, EDF, ABB, ZF, LIH, CBS, Bosch)



CLAIRe



TAILOR

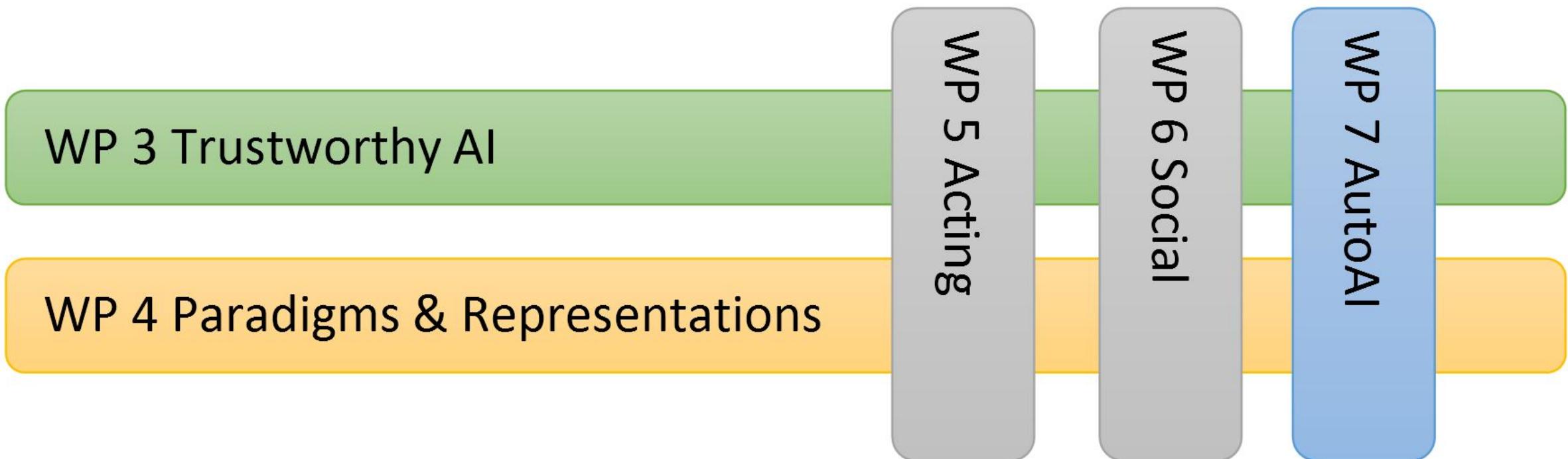
SOBIGDATA
RESEARCH INFRASTRUCTURE

KDD
Lat

Istituto di Scienze e tecnologie
dell'Informazione "A. Faedo"
Consiglio Nazionale delle Ricerche



TAILOR Basic Research Program



Cooperation with Industry and EU Projects

How are you cooperating with the other ICT-48 projects and how are you cooperating with industry & will continue to do so?

- Actively contributing and engaging in the activities and coordination provided by VISION (conferences, coordination meetings, joint mechanisms)
- Supporting and engaging in AIDA (started by AI4Media and co-led by VISION)
- Major overlap and interactions in WP4 (paradigms & representations) with researchers in ELISE
- Many researchers are also involved in HumanE-AI-Net, ELISE, and VISION
- Coordinated actions inspired by HumanE-AI-Net microprojects
- Theme development workshop targeting industry sector needs co-organized with VISION, HumanE-AI-NET and potentially AI4Media
- Working with AI-on-demand platform to provide a Trustworthy AI Repository
- Giving talks at events and conferences organized by AI4EU, BDVA, euRobotics, etc
- Connectivity fund can support joint research and workshops



Istituto di Scienze e Tecnologie
dell'Informazione "A. Faedo"
Consiglio Nazionale delle Ricerche



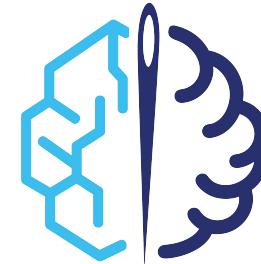
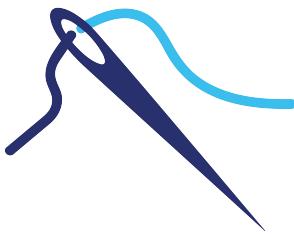
TAILOR Objectives

O1: Establish	O2: Define and maintain	O3: Create	O4: Build	O5: Progress	O6: Increase
O1: Establish a strong pan-European network of research excellence centers on the Foundations of Trustworthy AI	O2: Define and maintain a unified strategic research and innovation roadmap for the Foundations of Trustworthy AI	O3: Create the capacity and critical mass to develop the scientific foundations for Trustworthy AI	O4: Build sustained collaborations with academic, industrial, governmental, and community stakeholders on the Foundations of Trustworthy AI	O5: Progress the Scientific State-of-the-Art for the Foundations of Trustworthy AI	O6: Increase Knowledge and Awareness of the Foundations of Trustworthy AI across Europe



TAILOR

**Foundation of Trustworthy AI:
Integrating Learning, Optimisation and
Reasoning**



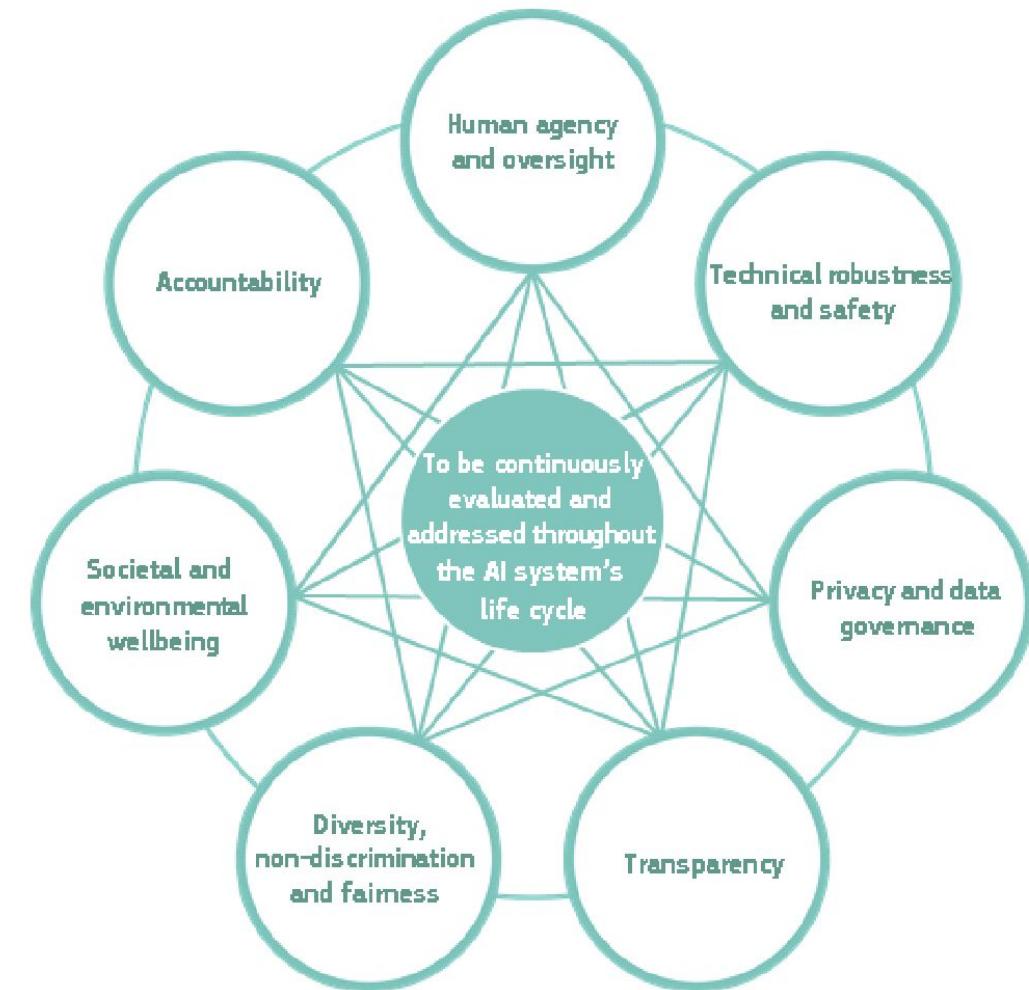
The dimensions of Trustworthy AI
Francesca Pratesi



TAILOR is an ICT-48 Network of AI Research Excellence Centers funded by
EU Horizon 2020 research and innovation programme GA No 952215

Trustworthy AI

- Goal
 - establish a continuous interdisciplinary dialogue for investigating methods and methodologies
 - ***To create AI systems that incorporate trustworthiness by design***
- Organized along the 6 dimensions of Trustworthy AI:
 - Explainability,
 - Safety and Robustness,
 - Fairness,
 - Accountability,
 - Privacy, and
 - Sustainability
- One transversal task that links the 6 dimensions among and ensures coherence and coordination across the activities.



Istituto di Scienze e Tecnologie
dell'Informazione "A. Faedo"
Consiglio Nazionale delle Ricerche



What (ethical) decisions?

- Can we teach ethics to AI?
- Should we teach ethics to AI?
- Understanding ethics
 - Which values? Whose values?
 - Who gets a say?
- Using ethics
 - What is proper action given values?
 - Are ethical theories of use?
 - How to prioritise values?
- Being ethical
 - Is knowing ethics enough?



COMPAS

recidivism black bias

DYLAN FUGETT

Prior Offense
1 attempted burglary

Subsequent Offenses
3 drug possessions

LOW RISK

3

HIGH RISK

10

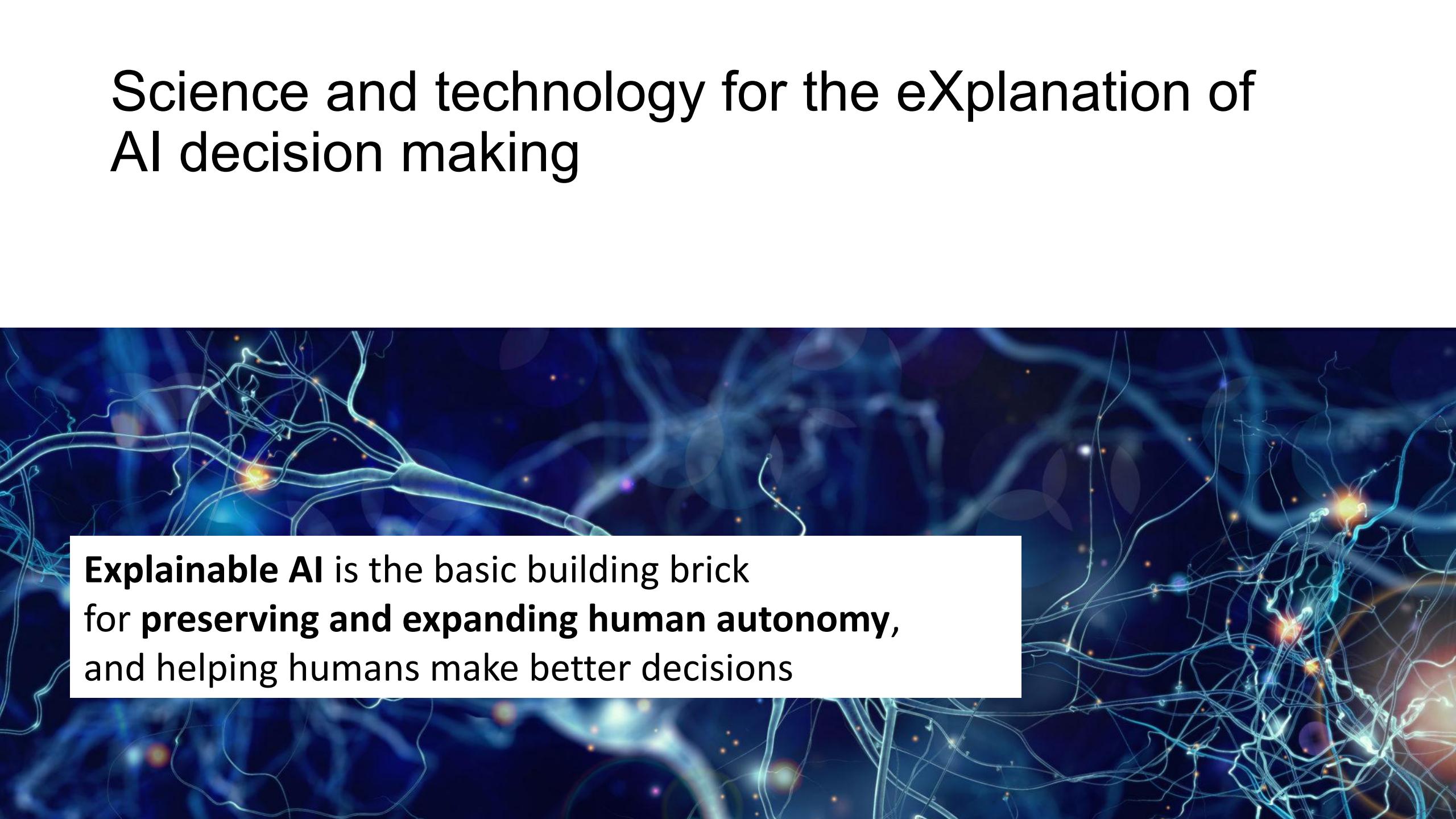
BERNARD PARKER

Prior Offense
1 resisting arrest
without violence

Subsequent Offenses
None

Fugett was rated low risk after being arrested with cocaine and marijuana. He was arrested three times on drug charges after that.

Science and technology for the eXplanation of AI decision making



**Explainable AI is the basic building brick
for preserving and expanding human autonomy,
and helping humans make better decisions**

Science and technology for the eXplanation of AI decision making

- To empower individual against undesired effects of automated decision making
- To implement the “right of explanation” (GDPR)
- To help people make better decisions
- To preserve (and expand) human autonomy



What is a Black Box Model?



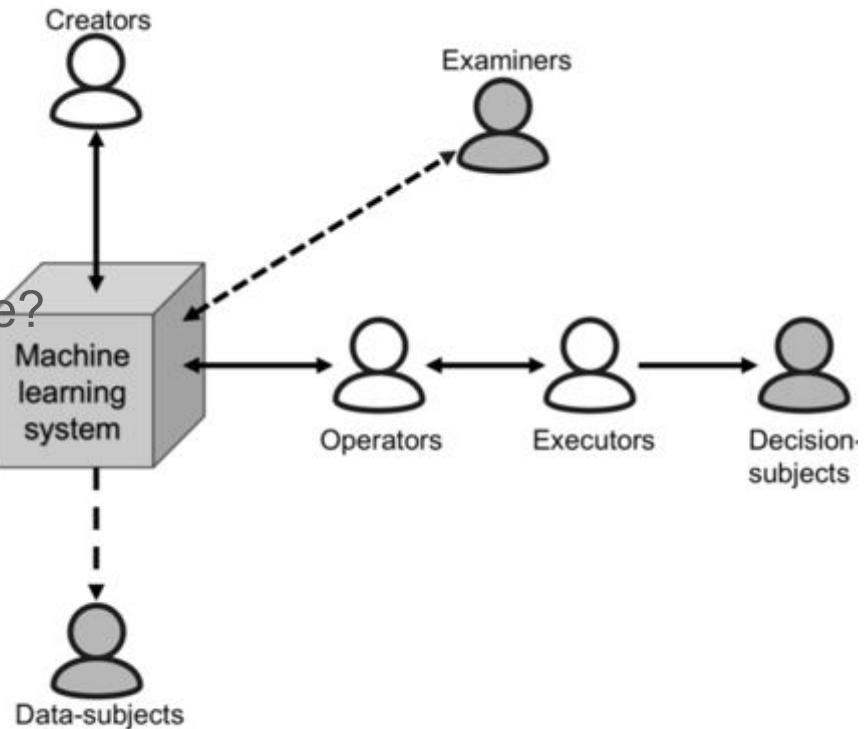
be either **unknown to the observer** or **humans**.

Guidotti, Monreale, Ruggieri, Turini, Giannotti, Pedreschi. *A survey of methods for explaining black box models*. ACM Computing Surveys (CSUR), 51(5), 93 (2018).



What is a good explanation, and for whom

- **End users:**
 - “Am I being treated fairly?”
 - “Can I contest the decision?”
 - “What could I do differently to get a positive outcome?”
- **Engineers, data scientists:**
 - “Is my system working as designed?”
- **Regulators:**
 - “Is the system compliant with the law?”



**An ideal explainer should
model the *user background***

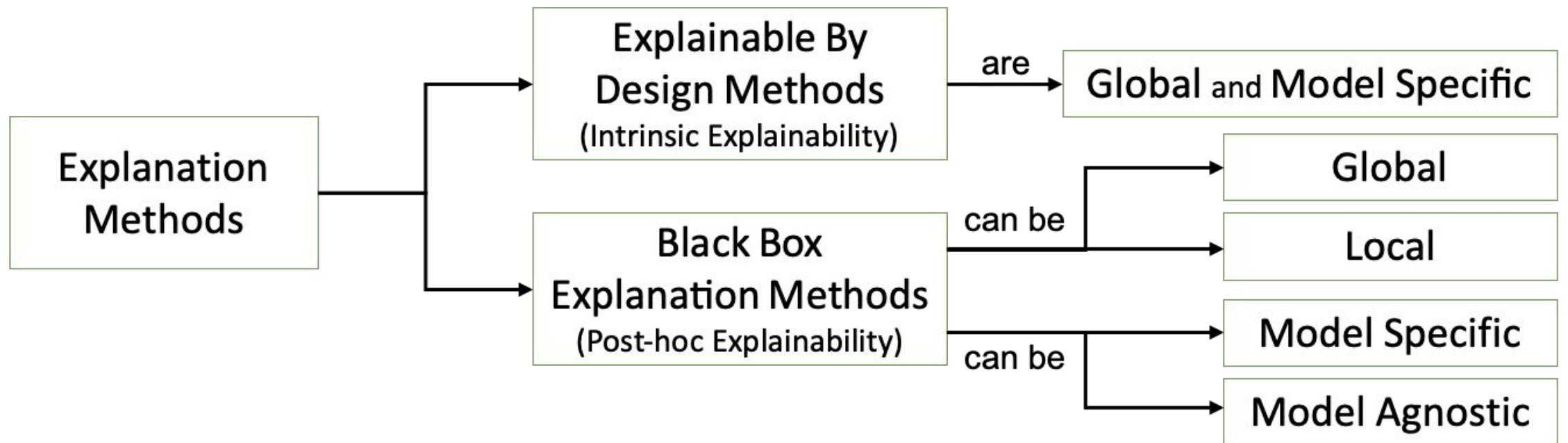
[Tomsett et al. 2018, Weld and Bansal 2018, Poursabzi-Sangdeh 2018, Mittelstadt et al. 2019]



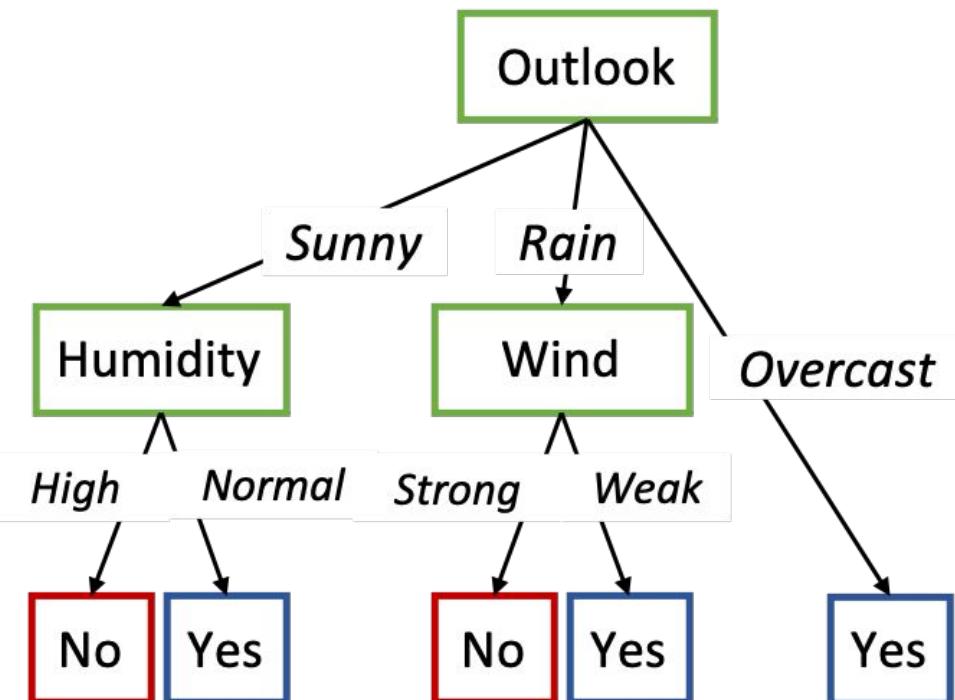
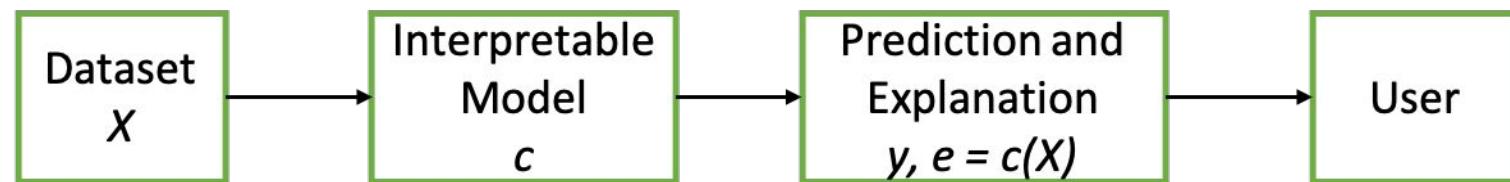
Istituto di Scienze e Tecnologie
dell'Informazione "A. Faedo"
Consiglio Nazionale delle Ricerche



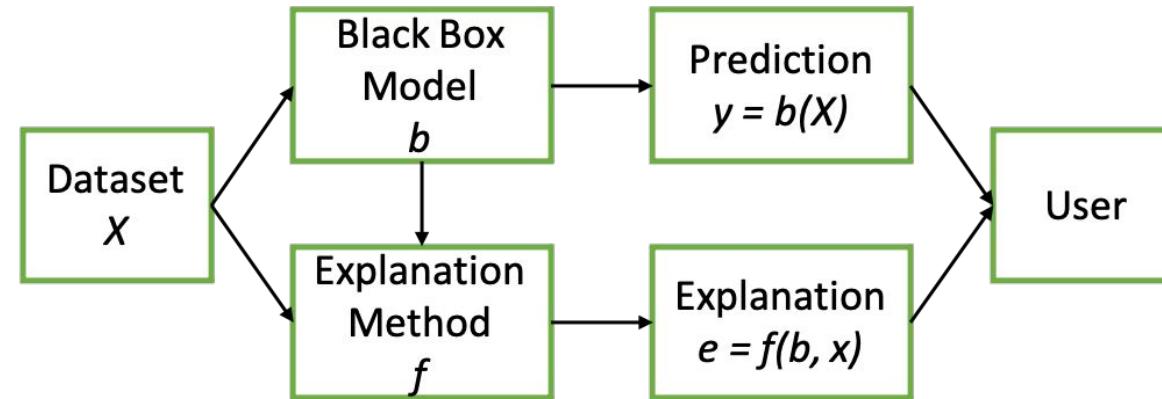
A way to think about Explainability



Explainable by Design Method



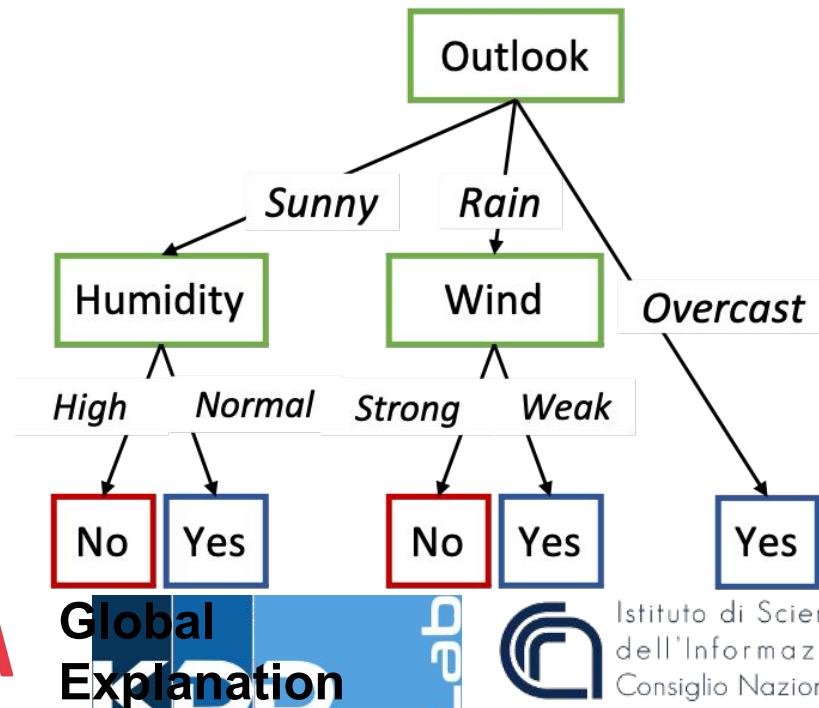
Black Box Explanations: Global vs Local



If Outlook = *Sunny* and Humidity = *Normal*
then Play Tennis = **Yes**

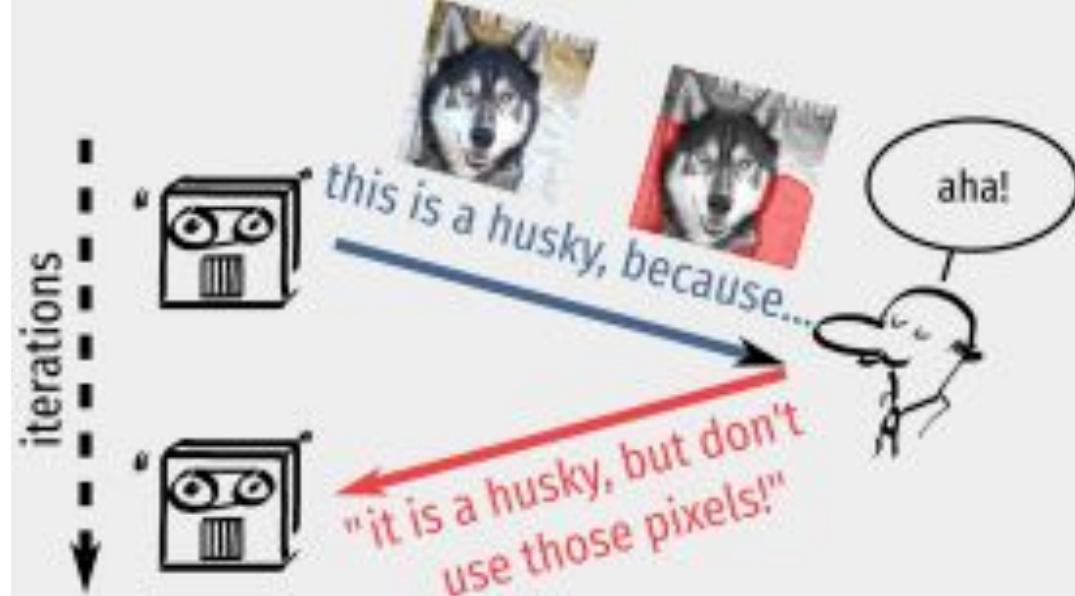
- Outlook: 0.7
- Humidity: -0.4
- Wind: 0.0

Local
Explanations



Research challenges on Explainable AI

- Explanatory Active Learning/Human-Machine Interaction
 - How to combine interaction and explanations to provide better results?



<https://github.com/stefanoteso/awesome-explanatory-supervision>

(Also workshop on Interactive Machine Learning @ AAAI-22)

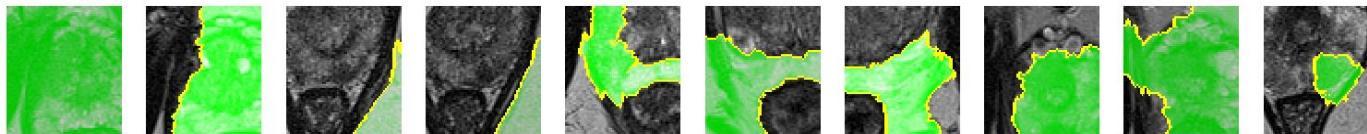


Istituto di Scienze e Tecnologie
dell'Informazione "A. Faedo"
Consiglio Nazionale delle Ricerche



Research challenges on Explainable AI

- We would like to apply, to the **same case study**, different Explainability techniques (e.g., LIME, LORE, ABELE, CRUSADE), providing different outputs, such as:
 - saliency maps (possibly with integrated gradients to be more precise):



- factual and counterfactual rules

$$r = \{ 29 > 0.50, 46 > 0.50, 47 > 0.50, 25 > 0.50, 58 > 0.50, 56 > 0.50, 40 > 0.50 \} \rightarrow \{ \text{class: 1} \}$$
$$c = \{ \{ 46 \leq 0.50 \} \rightarrow \{ \text{class: 0} \}, \{ 40 \leq 0.50 \} \rightarrow \{ \text{class: 0} \}, \{ 58 \leq 0.50 \} \rightarrow \{ \text{class: 0} \} \}$$

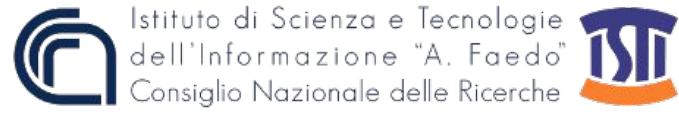
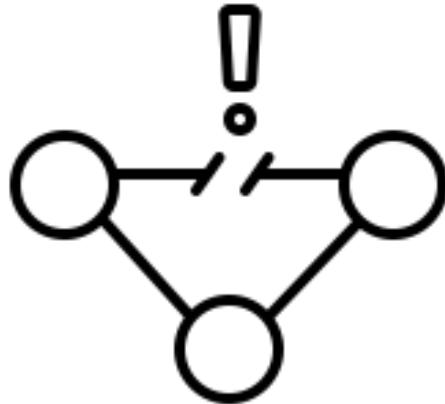
- exemplars and counter-exemplars images

- Then, we must to find a reliable way to **compare the results**

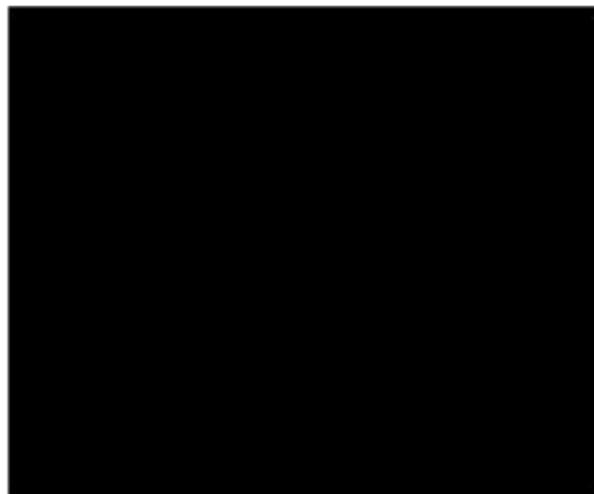
Guidotti, Monreale, Ruggieri, Pedreschi, Turini, Giannotti. *Local Rule-Based Explanations of Black Box Decision Systems*. arXiv preprint arXiv:1805.10820(2018)
Guidotti, Monreale, Giannotti, Pedreschi, Ruggieri, Turini. *Factual and Counterfactual Explanations for Black Box Decision Making*. IEEE Intelligent Systems, 34(6)
Guidotti, Monreale, Matwin, Pedreschi, *Black Box Explanation by Learning Image Exemplars in the Latent Feature Space*. ECML PKDD 2019



Safety and Robustness

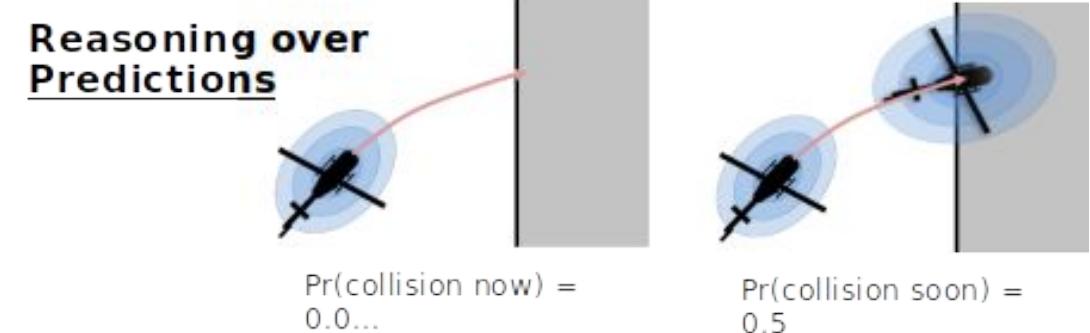
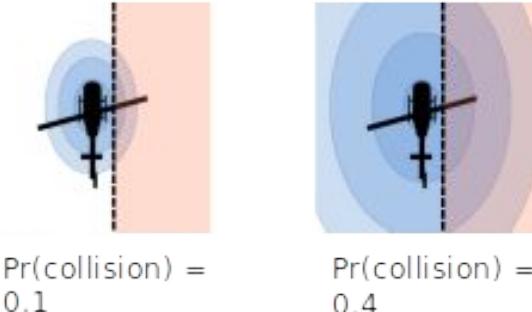
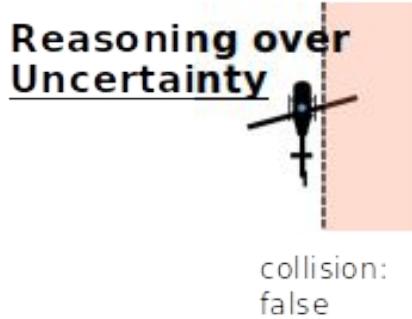


Safe Autonomous Systems / AI



If things can go wrong, they probably will!

This implies the need for continual monitoring of an autonomous system and its environment in a principled, contextual, task specific manner which can be specified by the system itself!



Probabilistic Predictive Stream Reasoning [Tiger and Heintz TIME
2016, IJAR 2020]



SOBIGDATA
RESEARCH INFRASTRUCTURE



Istituto di Scienze e Tecnologie
dell'Informazione "A. Faedo"
Consiglio Nazionale delle Ricerche

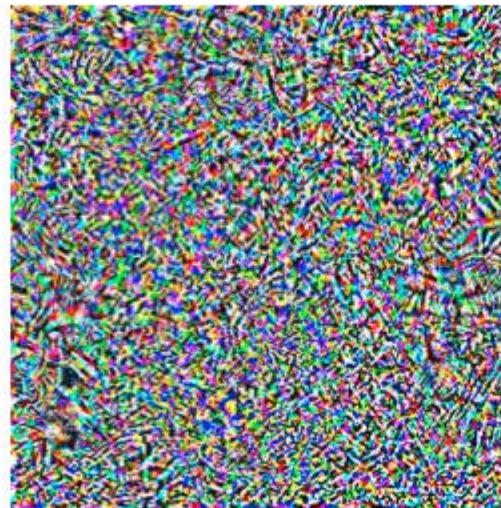


Dealing with truly Adversarial Examples

“pig”



+ 0.005 x



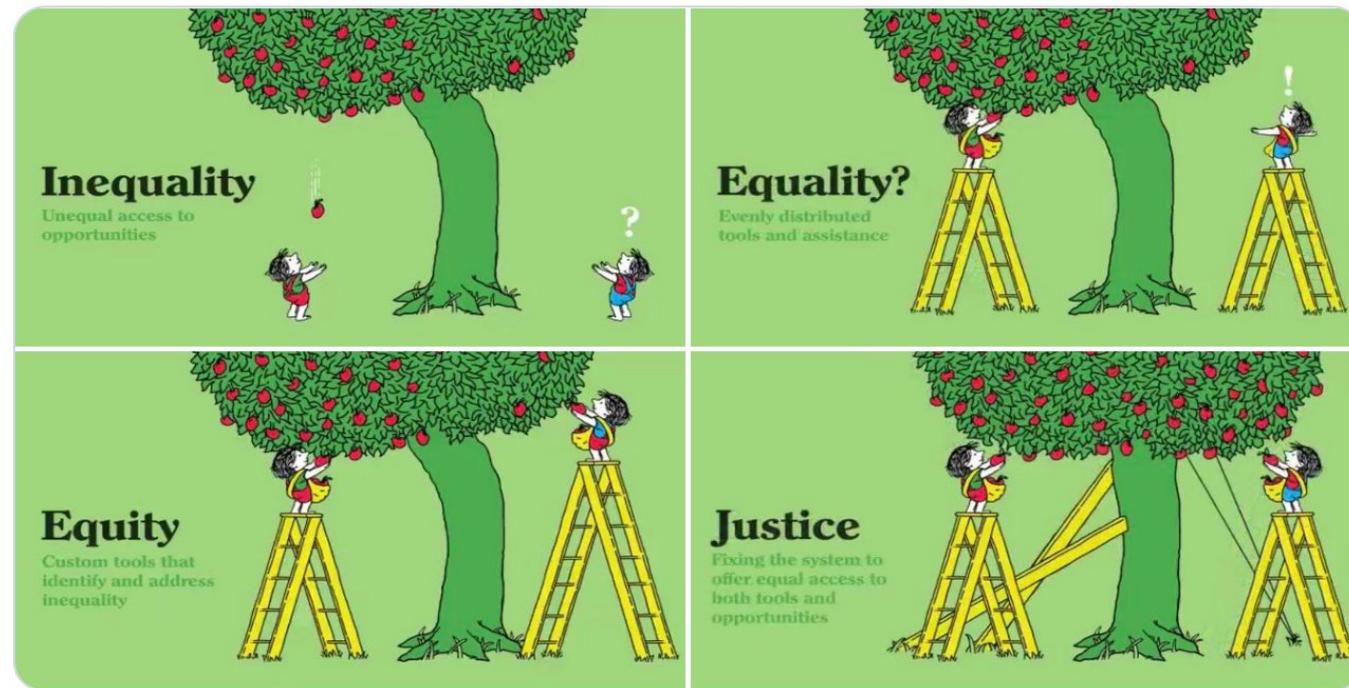
=

“airliner”



What is Fairness and Fairness-Aware ML?

- Impartial and just treatment or behavior without favoritism or discrimination.

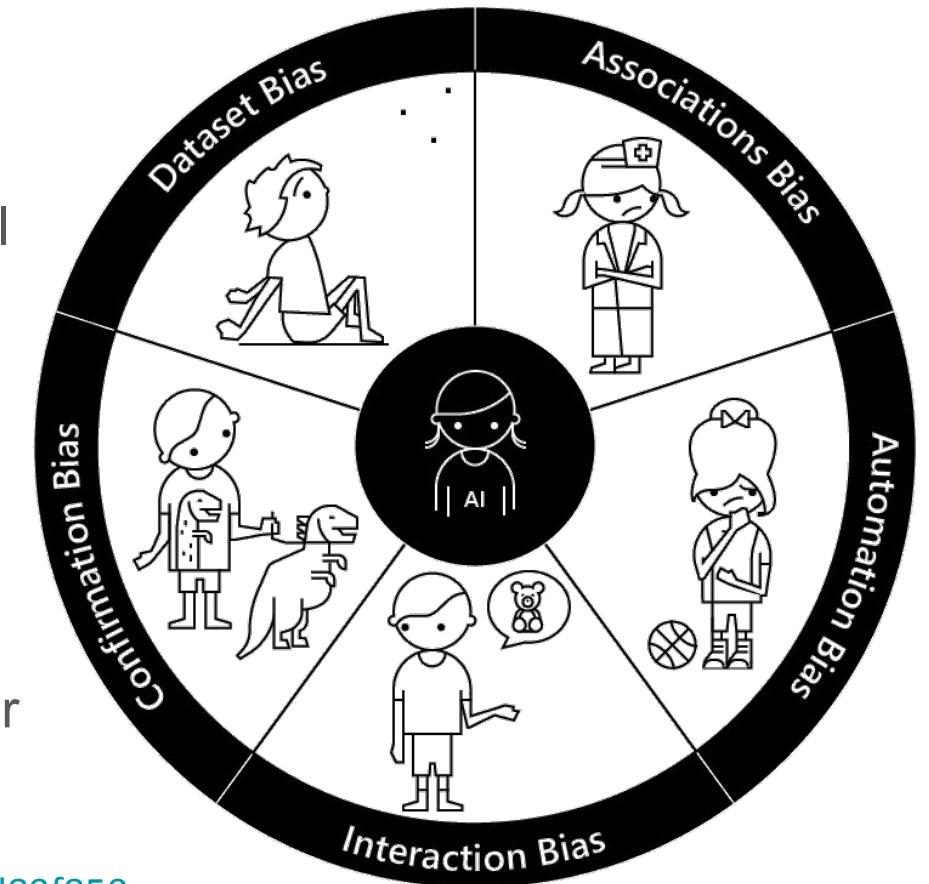


- *Fairness-aware machine learning* algorithms seek to provide methods under which the predicted outcome of a classifier operating on data about people is fair or non-discriminatory.



Bias

- **Dataset bias** – When the data used to train machine learning models doesn't represent the diversity of the customer base.
- **Association bias** – When the data used to train a model reinforces and multiplies a cultural bias.
- **Automation bias** – When automated decisions override social and cultural considerations.
- **Interaction bias** – When humans tamper with AI and create biased results.
- **Confirmation bias** – When oversimplified personalization makes biased assumptions for a group or an individual.



<https://medium.com/microsoft-design/how-to-recognize-exclusion-in-ai-ec2d6d89f850>



Accountability and Reproducibility

Two big issues towards
Trustworthiness
Just think to AI in biomedicine



La radiologia medica (2020) 125:517–521
<https://doi.org/10.1007/s11547-020-01135-9>

EDITORIAL

Check for updates

Artificial intelligence: Who is responsible for the diagnosis?

Emanuele Neri¹ ● - Francesca Coppola² - Vittorio Miele³ - Corrado Bibbolino⁴ - Roberto Grassi⁵

Received: 13 December 2019 / Accepted: 16 January 2020 / Published online: 31 January 2020
© Italian Society of Medical Radiology 2020

Radiomics and Machine Learning

Repeatability and Reproducibility of Radiomic Features: A Systematic Review

International Journal of Radiation Oncology biology • physics

Alberto Traverso, PhD, *† Leonard Wee, PhD, *† Andre Dekker, PhD, *† and Robert Gillies, PhD†

*Department of Radiation Oncology, MAASTRO Clinic, Maastricht, The Netherlands; †School for Oncology and Developmental Biology (GROW), Maastricht University, Maastricht, The Netherlands; and ‡Moffitt Cancer Center, Tampa, Florida

Received Nov 20, 2017, and in revised form May 15, 2018. Accepted for publication May 20, 2018.



TAILORE

SUBIG
RESEARCH INFRASTRUCTURE

logie
edo" reche

ISI

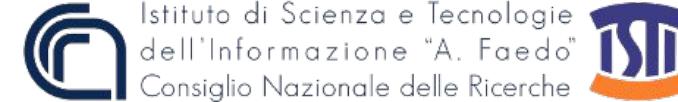
Accountability

Accountability regards the decisions on the attribution of **moral responsibility** and **who is responsible** (and in what sense) for untoward outcomes.

AI systems offer unprecedented opportunities to develop faster and more reliable systems, also helping humans to take decisions, but they also introduce more gray areas.



European Group on Ethics in Science and New Technologies, *Artificial Intelligence, Robotics and 'Autonomous' Systems*, Brussels 9 March 2018, doi:10.2777/531856



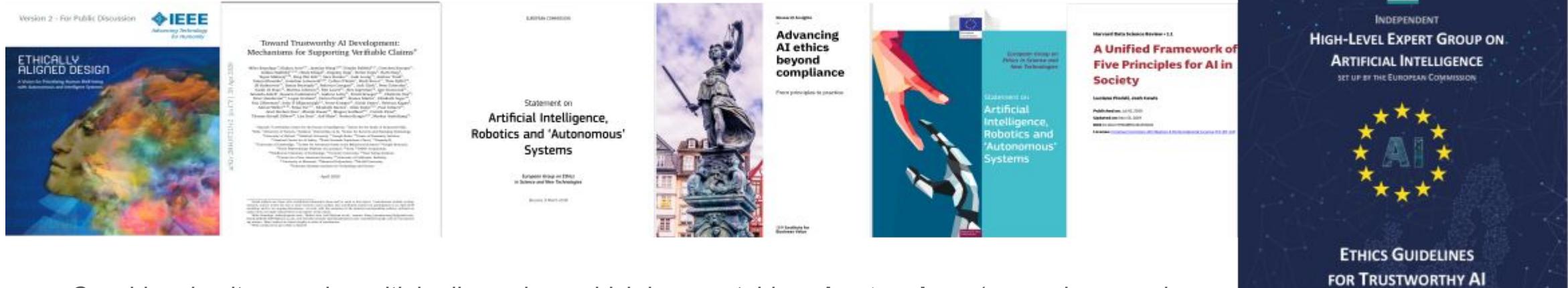
Reproducibility

- **Reproducibility of methods:** the ability to implement, as exactly as possible, the experimental and computational procedures, with the same data and tools, to obtain the same results
- **Reproducibility of results:** the production of corroborating results in a new study, having used the same experimental methods
- **Reproducibility of inference:** the drawing of qualitatively similar conclusions from either an independent replication of a study or a reanalysis of the original study



Research challenges on Accountability and Reproducibility

- Propose methods to **model and analyze** emergent effects and diffusion of responsibility in **multi-agent systems** (problem of many hands and many things)



- Consider simultaneously multiple dimensions which have notable **value tensions** (e.g., privacy and security), especially in opaque and non-reproducible applications.
- Take a **multidisciplinary perspective** on evaluating how to avoid accountability gaps given a broad spectrum of dimensions (e.g., explainability, robustness, and fairness).
- Evaluate **metadata models** on a list of competency questions identified as pivotal for ensuring reproducibility and accountability. It will identify gaps in the schemas and how these schemas might complement each other to document the different dimensions of trustworthiness



Respect for Privacy



Data Protection
Officer (DPO)

Compliance

25 May 2018

Data Breaches

Personal Data



Istituto di Scienze e Tecnologie
dell'Informazione "A. Faedo"
Consiglio Nazionale delle Ricerche



Respect for Privacy



Istituto di Scienze e Tecnologie
dell'Informazione "A. Faedo"
Consiglio Nazionale delle Ricerche



Research Challenge on Privacy

- There is still the need to investigate new methodologies and approaches for:
 - automatically detecting **privacy risks**
 - designing robust **data anonymization algorithms**
 - designing AI algorithms that respect **by design** privacy constraints
 - collaborative projects between academia and **industry** for a continuous transfer of anonymization techniques



Research Challenge on Privacy

- Study the tensions and the **interplay with other aspects and human values**

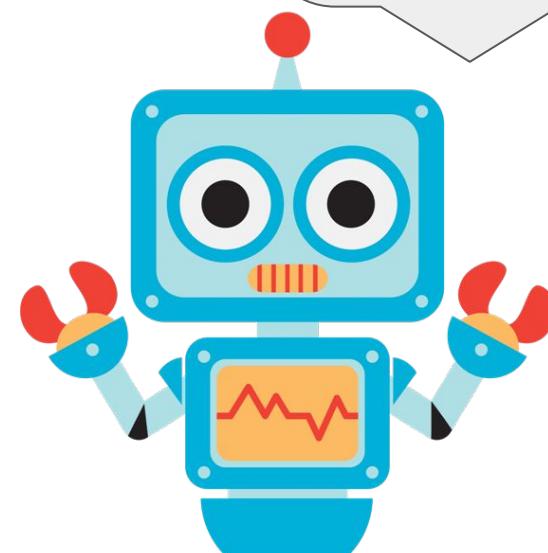


Research Challenge on Privacy

Provide global or local **explanations on privacy risks**



Your risk is supposed to be high because your daily distance is greater than 80% of other drivers.

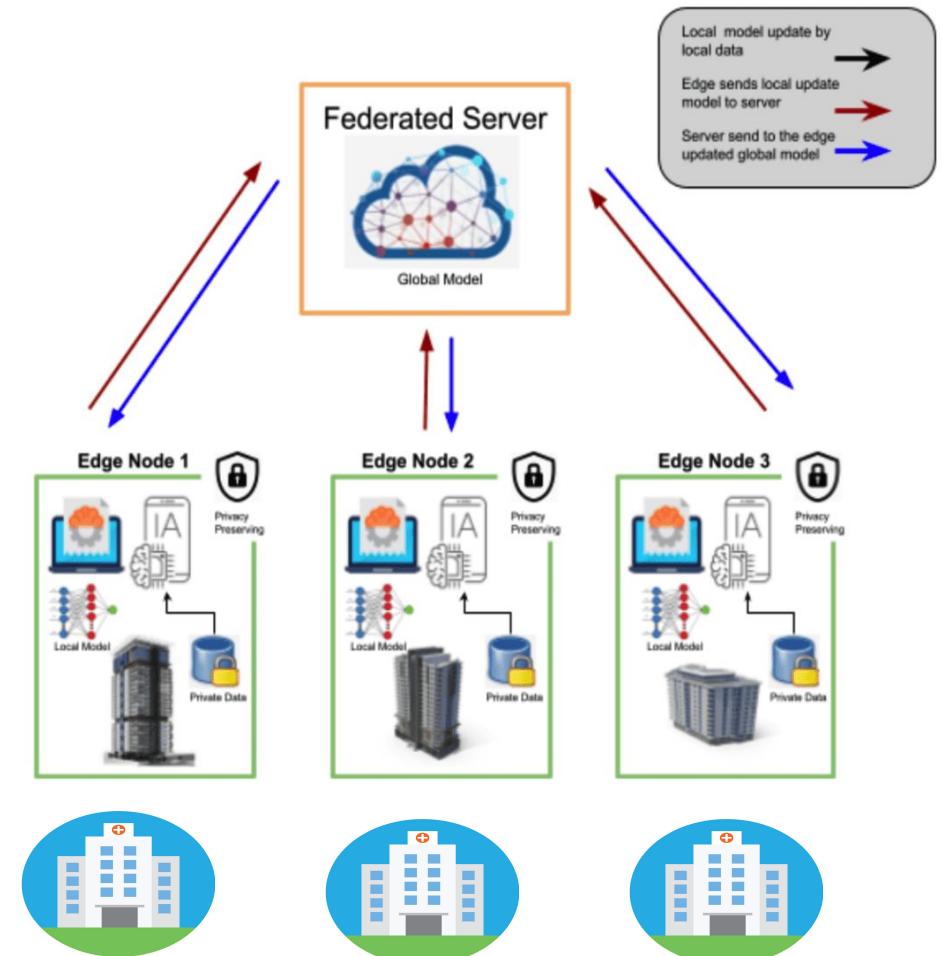


Istituto di Scienze e Tecnologie
dell'Informazione "A. Faedo"
Consiglio Nazionale delle Ricerche



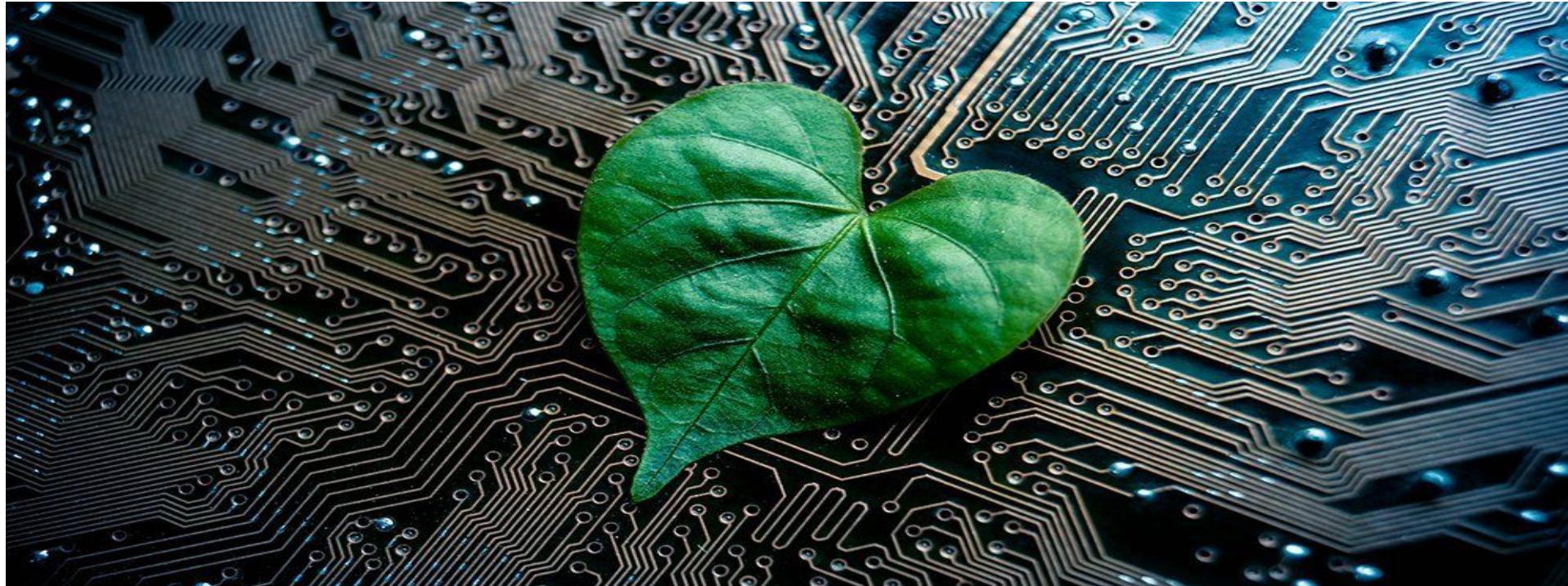
Federated Learning: Classical Setting

“Federated Learning is a ML setting where multiple distributed parties, called **clients**, under the orchestration of a main **server**, **cooperate** to train a **shared global model**, while keeping their **data private**”



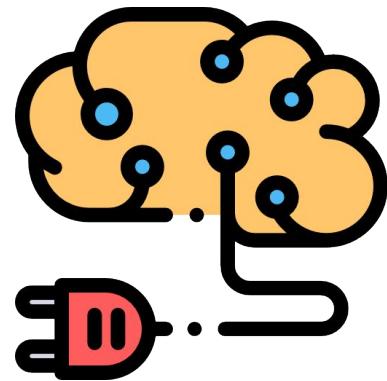
Sustainability

Energy consumption and carbon footprint of AI systems and solution like deep networks is extremely high.



Research Challenge on Sustainability

- Reduce the energy needed to train AI solutions
 - Making them learning faster
 - Recycle learning from different applications (transfer learning)
- Optimise the data centres in which the AIs are trained
 - Reduce the waste of resources

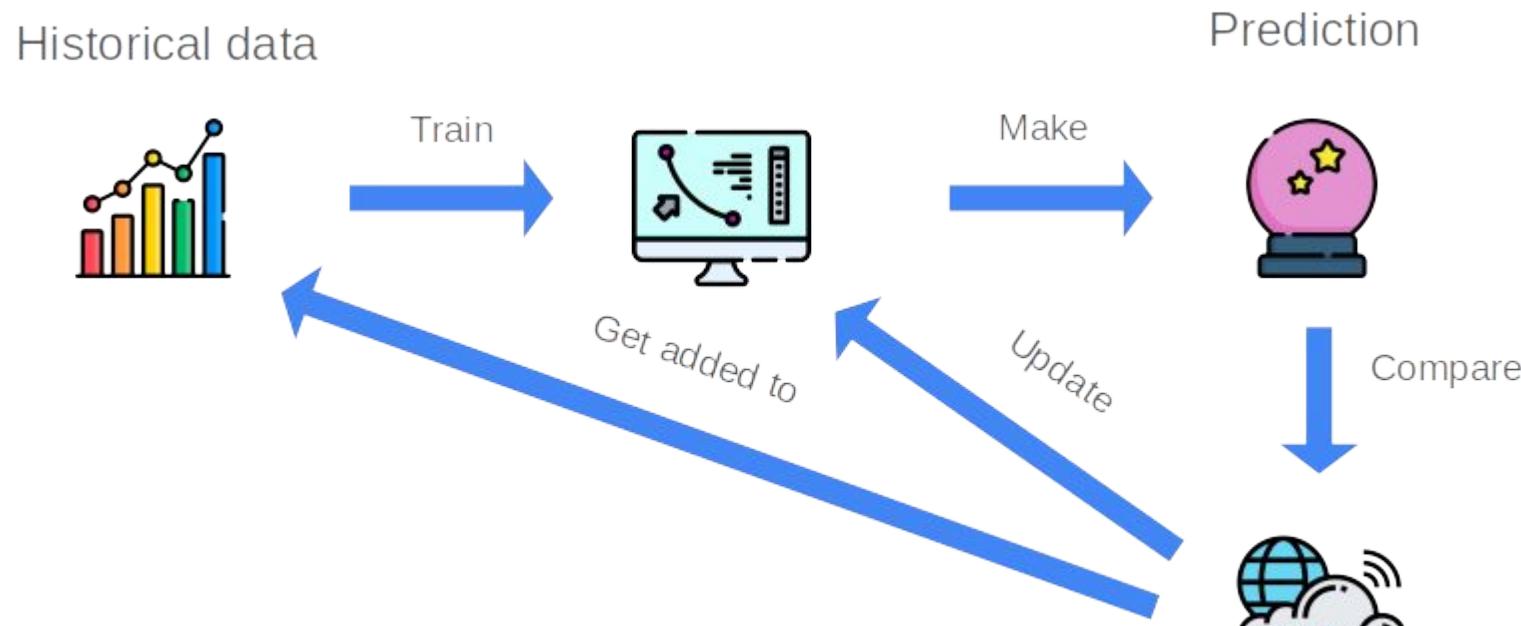


Istituto di Scienze e Tecnologie
dell'Informazione "A. Faedo"
Consiglio Nazionale delle Ricerche



Research Challenge on Sustainability

- Predict the expected behavior of a deep NN both at training and inference time, to steer the **dynamic reconfiguration and adaptation** of the training/inference process



thank
you