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

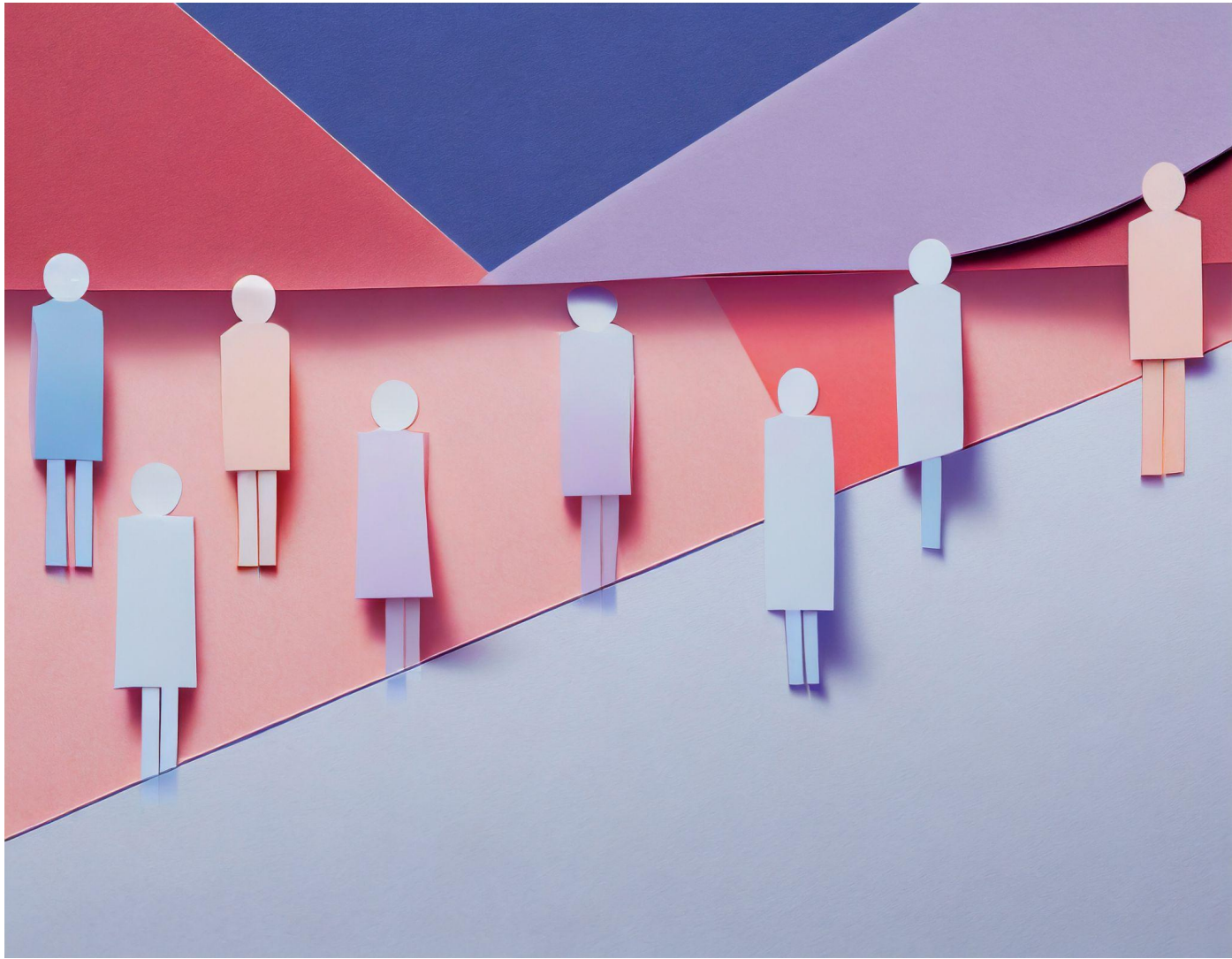
SOBIGDATA
RESEARCH INFRASTRUCTURE

TAILOR

SOBIGDATA

RESEARCH INFRASTRUCTURE

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**

COUNTRIES	INSTITUTIONS	RESEARCHERS
14	32	250

TAILOR

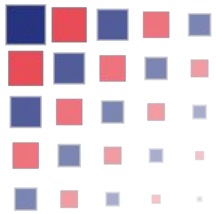
SOBIGDATA RESEARCH INFRASTRUCTURE

KDD Lab

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

The three picies of SoBigData RI

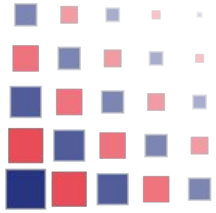
RI



SOBIGDATA++
PLUS PLUS

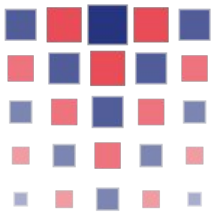
EU PROJECTS
3

ESFRI Roadmap



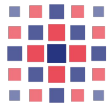
SOBIGDATA PPP
PPP

Italian Node (PNRR)



SOBIGDATA.it
Italian research infrastructure

TAILOR



SOBIGDATA
RESEARCH INFRASTRUCTURE

Istituto di Scienza 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

TAILOR

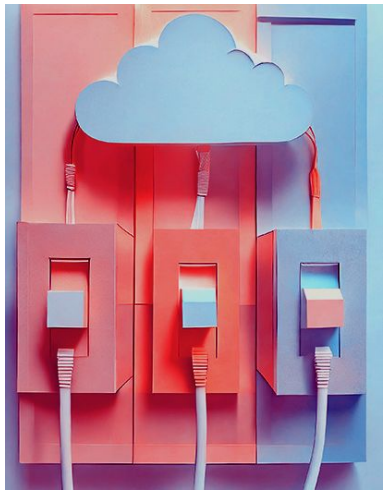
SOBIGDATA RESEARCH INFRASTRUCTURE

KDD Lab

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

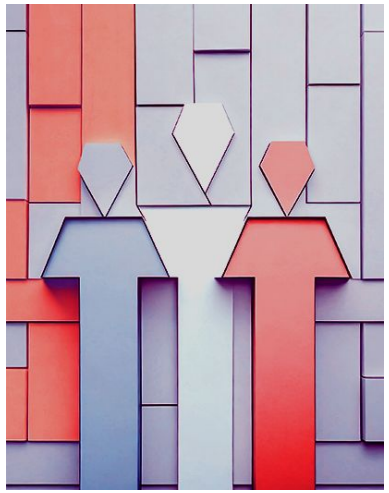
SoBigData RI Overview

INFRASTRUCTURE



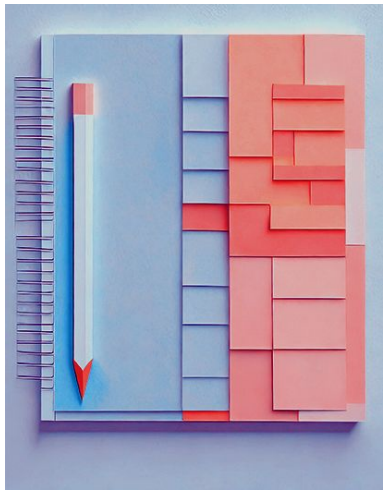
- Computational Power
- Open Science tools
- Resources
- Datasets
- Methods
- Technologies
- Training Materials

COMMUNITY BUILDING



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

GUIDELINES



- Ethical aspects
- Legal Aspects
- Open Science

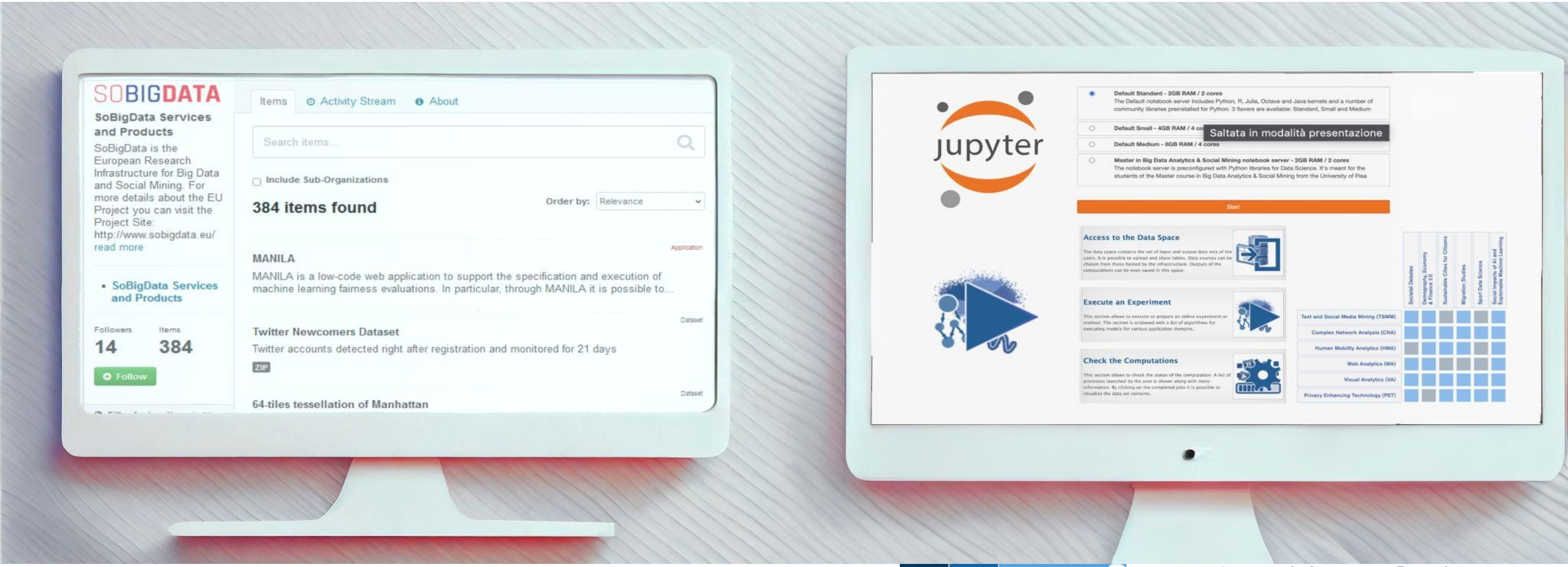
Virtual Access

SoBigData Catalogue

- Methods,
- Libraries,
- Applications,
- Data

SoBigData Lab

- Jupyter Hub
- Data Space
- Execute an Experiment



TAILOR

SOBIGDATA RESEARCH INFRASTRUCTURE

KDD Lab

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

Focus

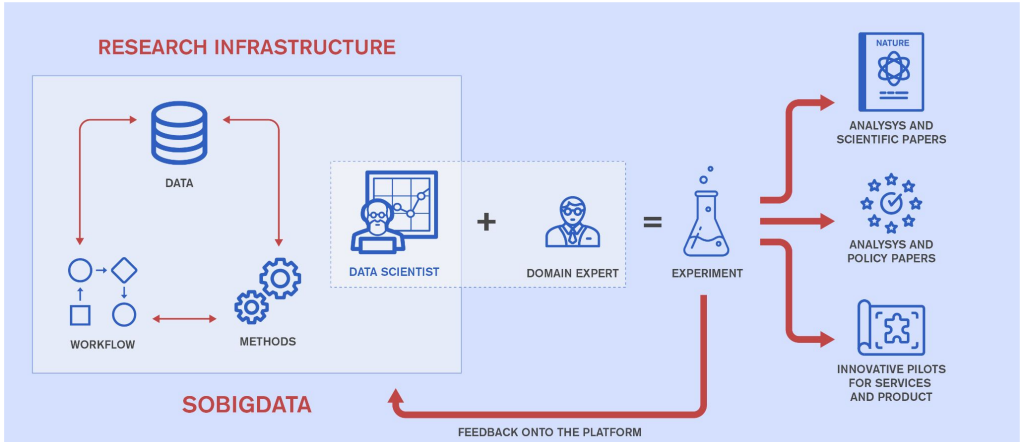
Responsible data science principles

- 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)



TAILOR

SOBIGDATA
RESEARCH INFRASTRUCTURE

KDD Lab

Istituto di Scienza 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.

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?

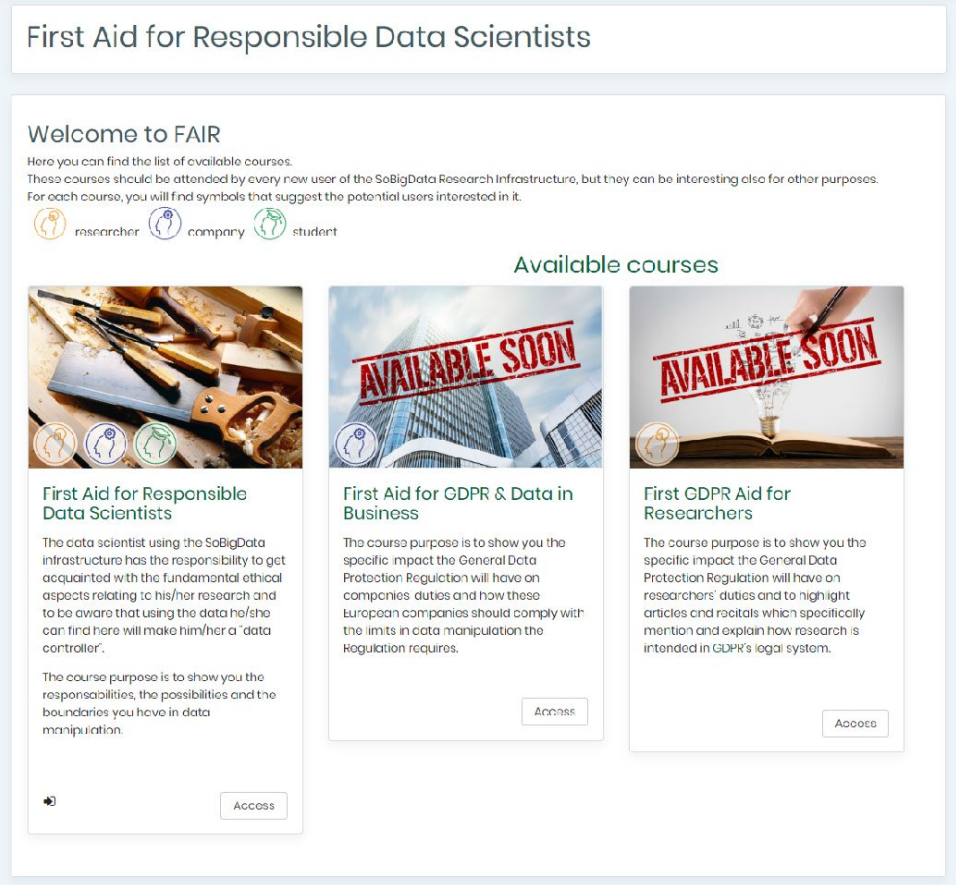
FAIR

First AId for Responsible data scientist

TRY IT ON:
fair.sobigdata.eu/moodle

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

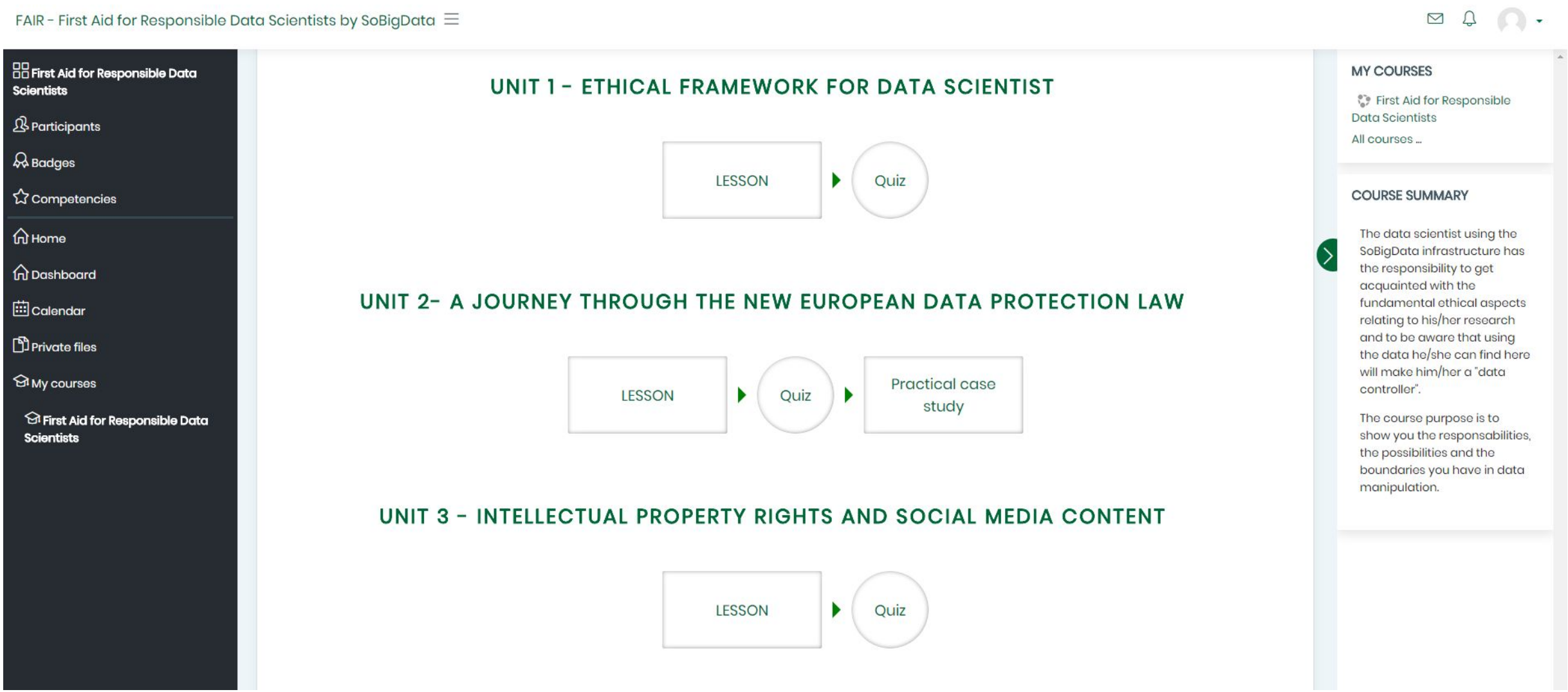
SoBigData



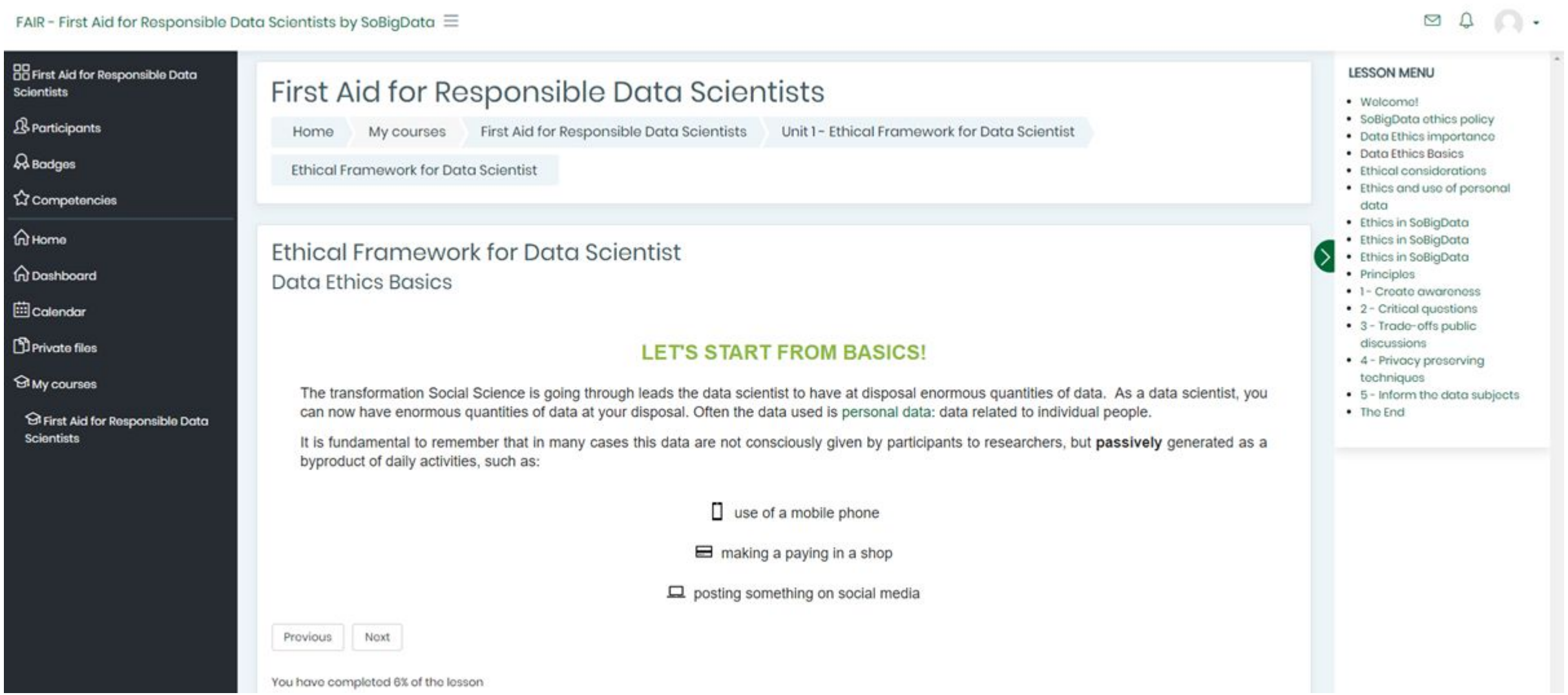
- 3 core lessons:
  - ethics
  - data protection law
  - intellectual property law
- 2 additional lessons:
  - GDPR for researchers
  - GDPR for industries
- 57 questions
  - (3 about ethics, 50 about DP, 4 about IP)
- 2 case studies
  - (about real sentences)

- THE GOAL OF THE COURSES
  IS TO BRING AWARENESS
  AND INFORMATION TO CITIZENS, COMPANIES, INSTITUTIONS
  AND RESEARCHERS

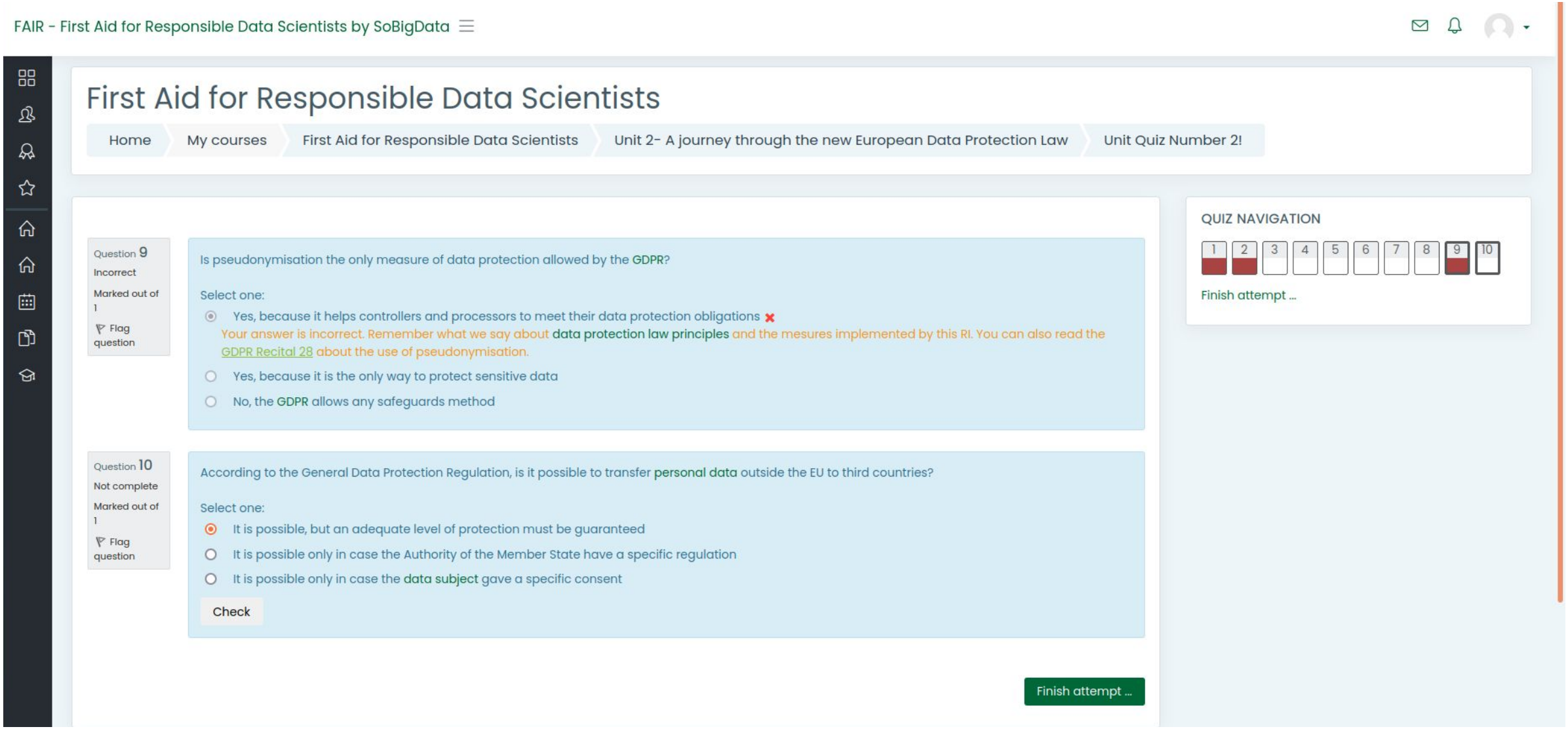
fair.sobigdata.eu/moodle



fair.sobigdata.eu/moodle



fair.sobigdata.eu/moodle



fair.sobigdata.eu/moodle

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**.

TAILOR
SOBIGDATA RESEARCH INFRASTRUCTURE
KDD Lab
Istituto di Scienza 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

TAILOR

SOBIGDATA RESEARCH INFRASTRUCTURE

KDD Lab

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

Ethics Guidelines for Trustworthy AI – Overview

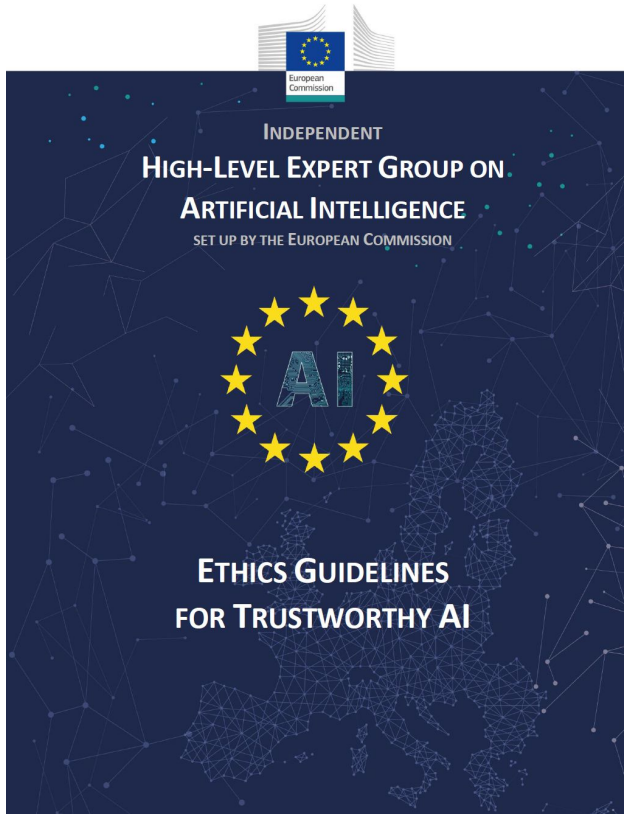
Human-centric approach: AI as a means, not an end

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

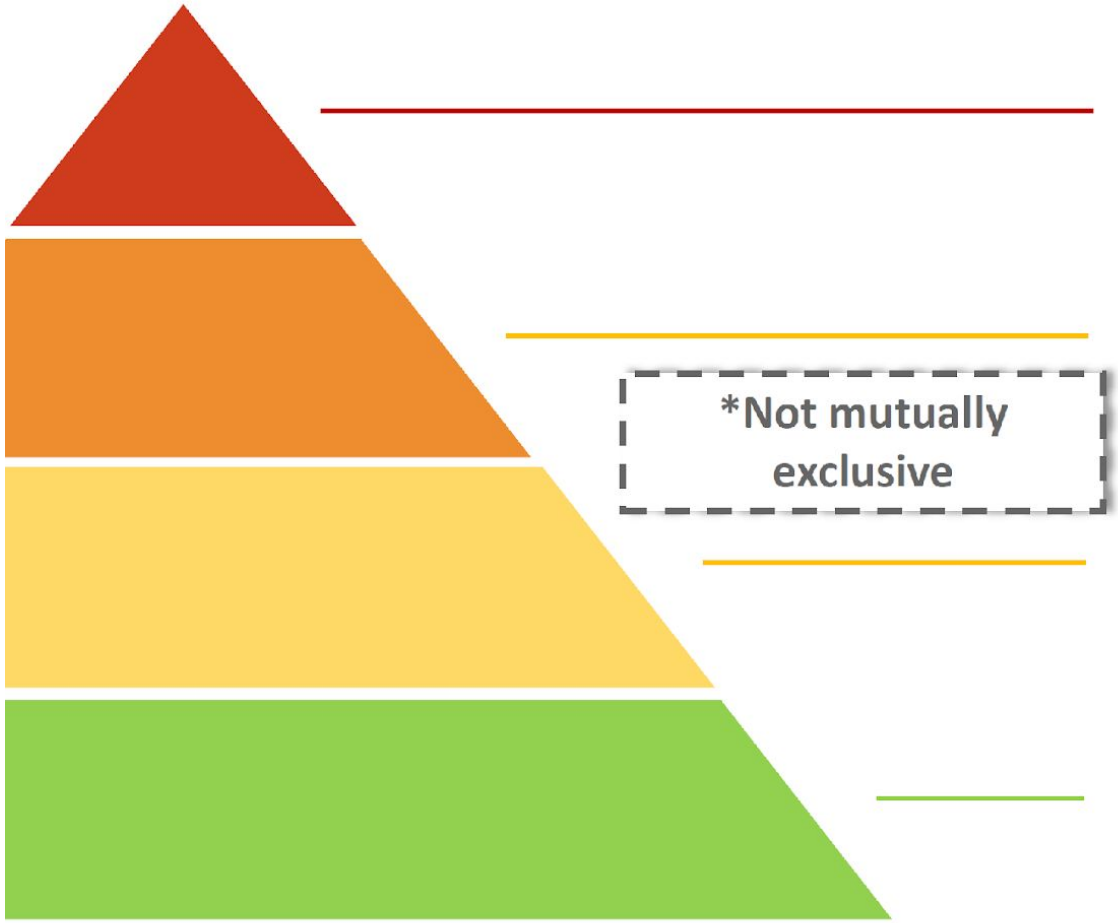
TAILOR

SOBIGDATA RESEARCH INFRASTRUCTURE

KDD Lab

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

The AI Act: the new EU Regulation on AI



*Not mutually exclusive

Unacceptable risk
e.g. social scoring
— **Prohibited**

High risk
e.g. recruitment, medical devices
— **Permitted** subject to compliance with AI requirements and ex-ante conformity assessment

'Transparency' risk
'Impersonation' (bots)
— **Permitted** but subject to information/transparency obligations

Minimal or no risk
— **Permitted** with no restrictions

European Commission

TAILOR

Foundation of Trustworthy AI:
Integrating Learning, Optimisation and Reasoning

Overview of TAILOR
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

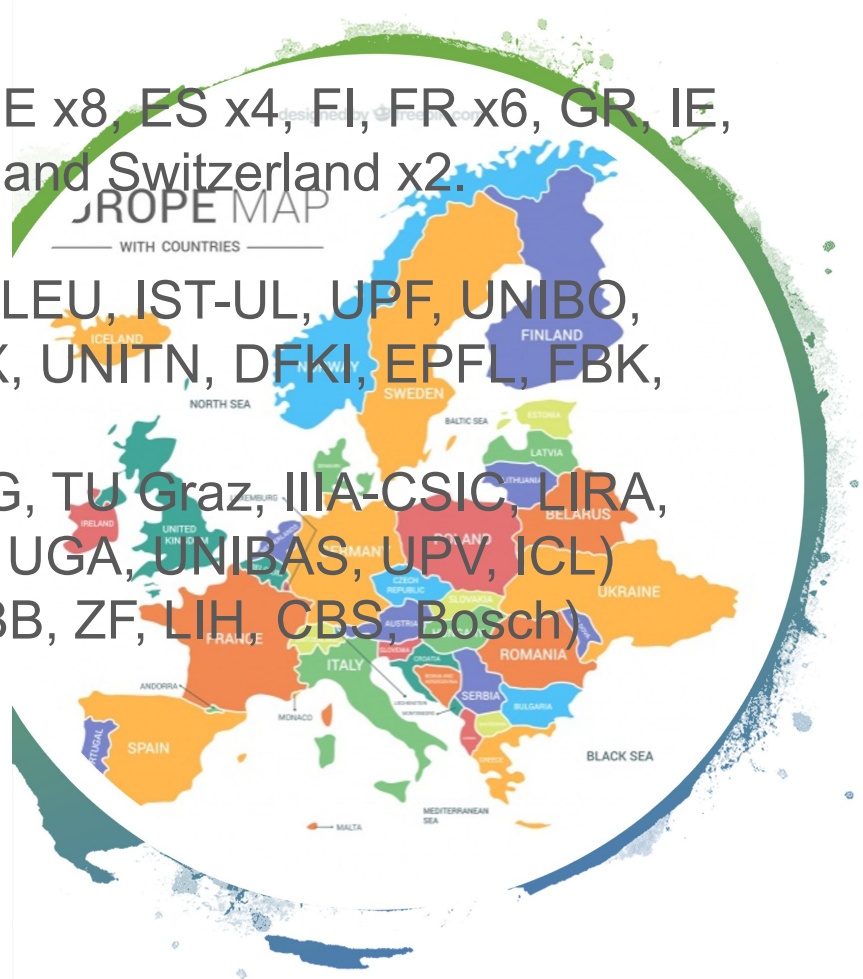
Boosting Capacity to Tackle Major Scientific Challenges

- *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.

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

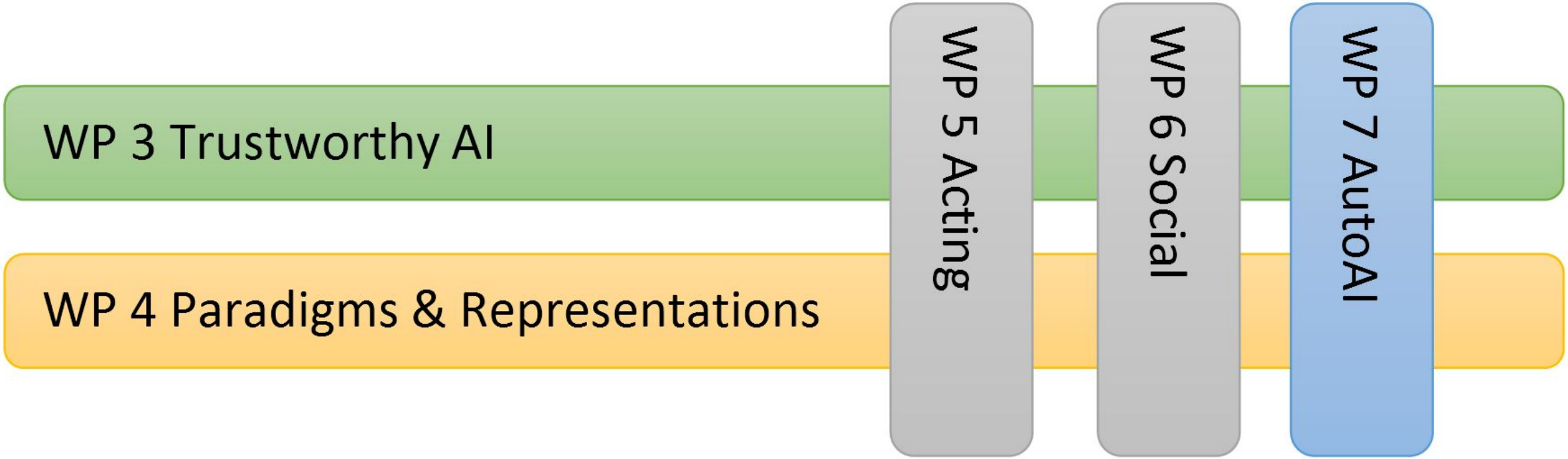
TAILOR

SOBIGDATA RESEARCH INFRASTRUCTURE

KDD Lab

Istituto di Scienza 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

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

Joint Technical Review 2021-09-29

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.

TAILOR

SOBIGDATA RESEARCH INFRASTRUCTURE

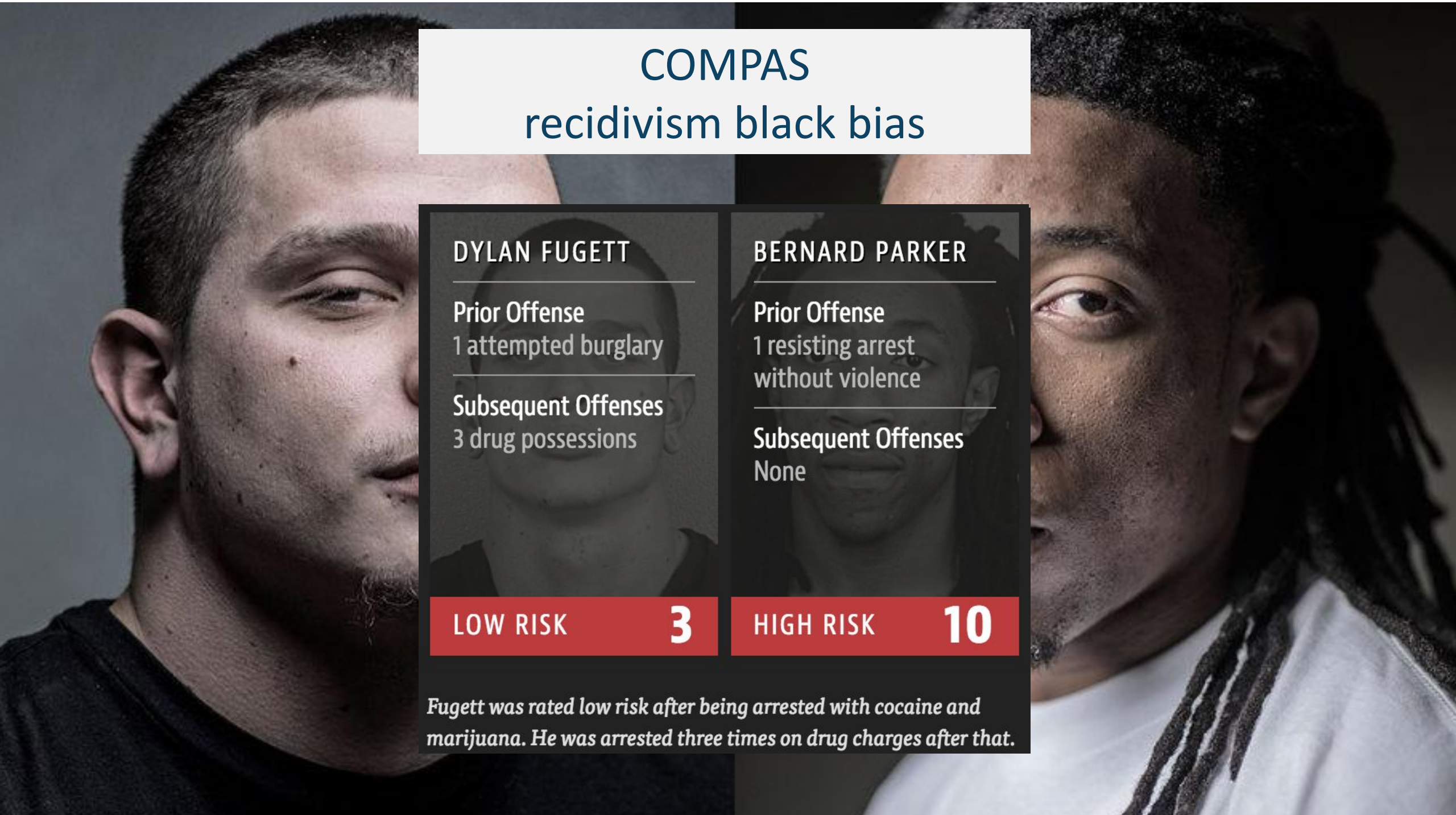
KDD Lab

Istituto di Scienza 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



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

TAILOR

SOBIGDATA RESEARCH INFRASTRUCTURE

KDD Lab

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

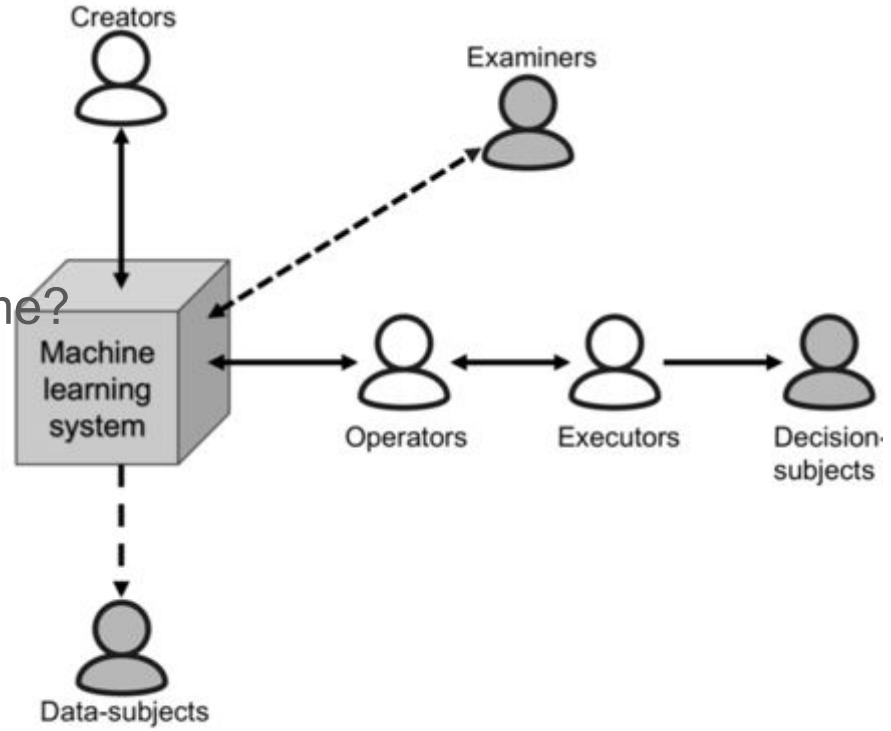
What is a Black Box Model?

e either unknown to the observer
umans.

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

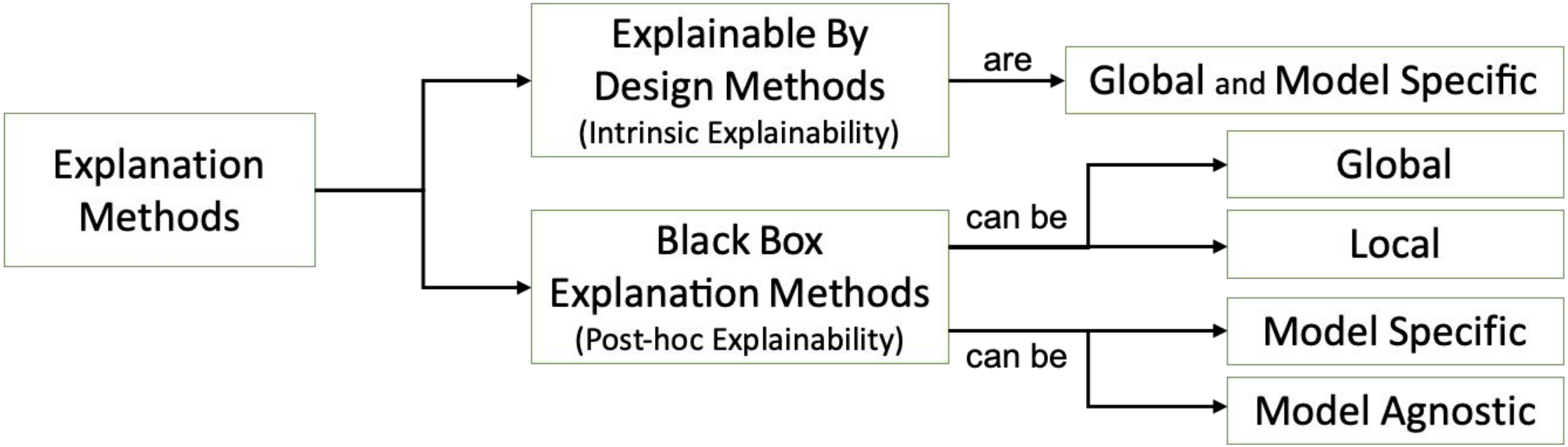
- **End users:**
  - "Am I being treated fairly?"
  - "Can I contest the decision?"
  - "What could I do differently to get a positive outcome?
- **Engineers, data scientists**:
  - "Is my system working as designed?"
- **Regulators:**
  - "Is the system compliant with the law?"



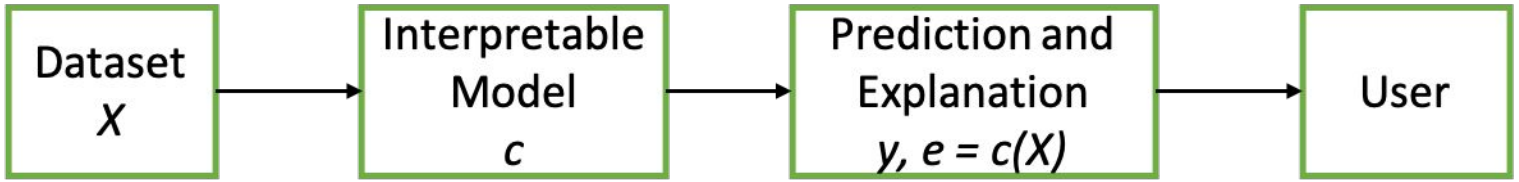
An ideal explainer should model the *user background*

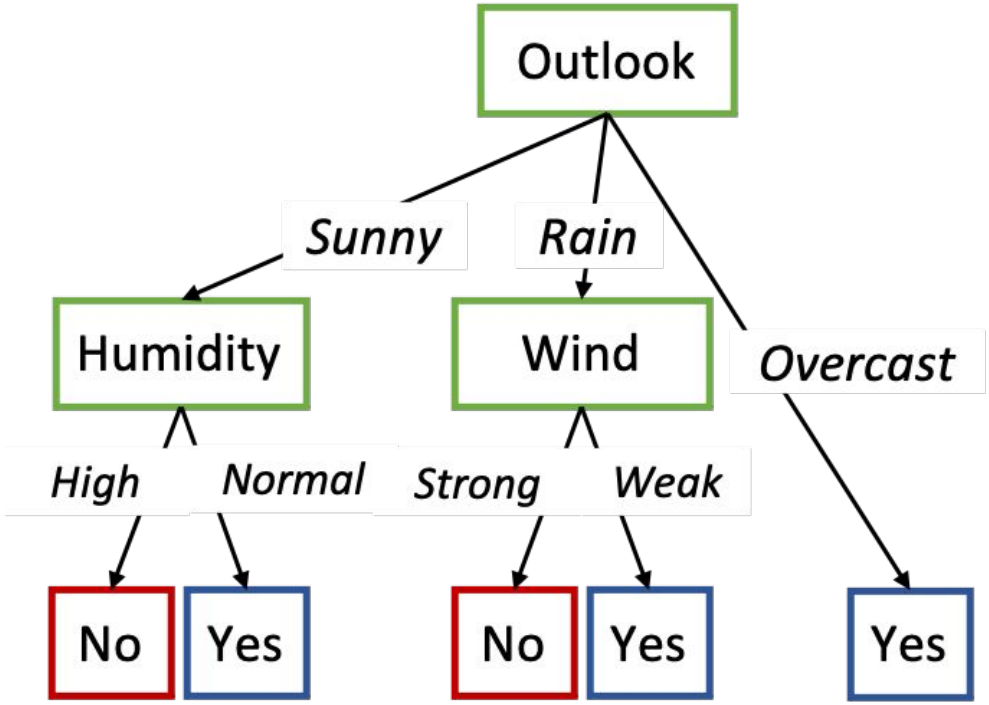
[Tomsett et al. 2018, Weld and Bansal 2018, Poursabzi-Sangdeh 2018, Mittelstadt et al. 2019]

A way to think about Explainability

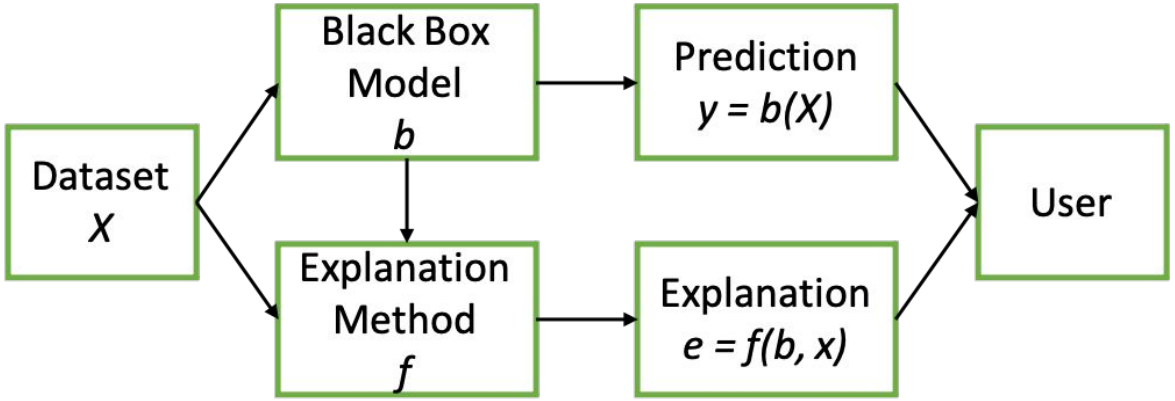


Explainable by Design Method





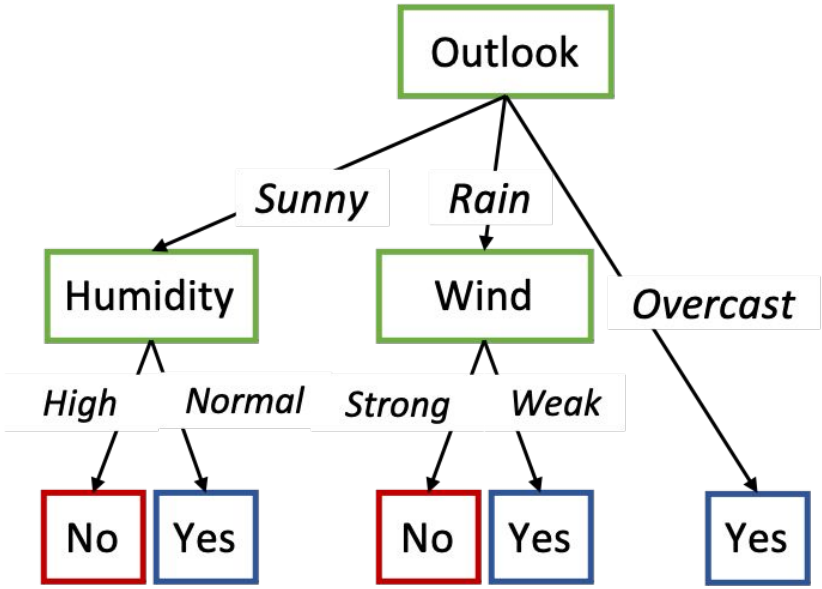
Black Box Explanations: Global vs Local



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

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

Local Explanations



Global Explanation

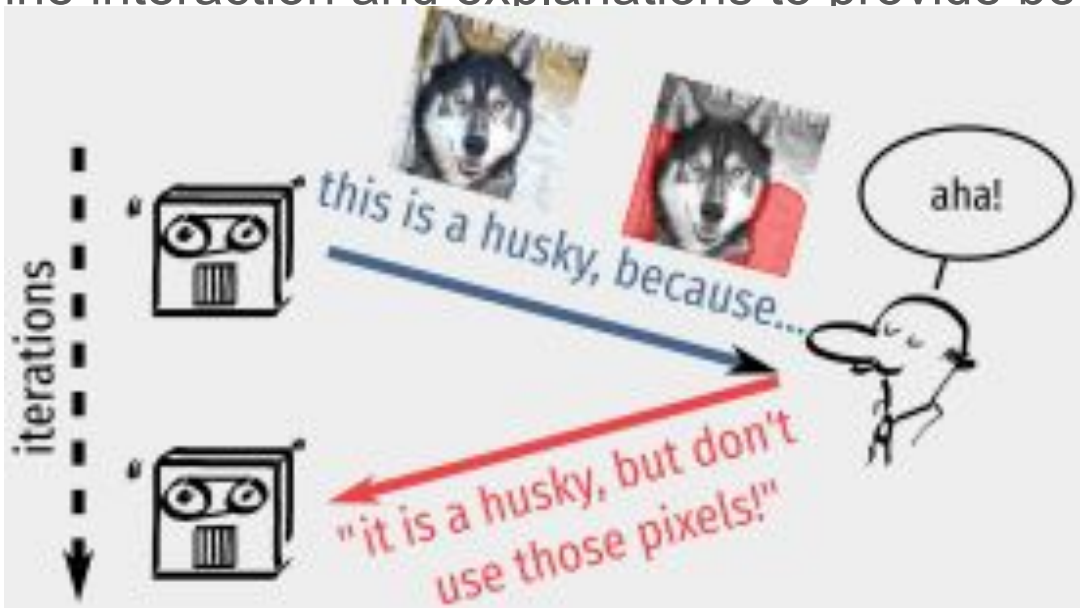
TAILOR

SOBIGDATA RESEARCH INFRASTRUCTURE

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

Research challenges on Explainable AI

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

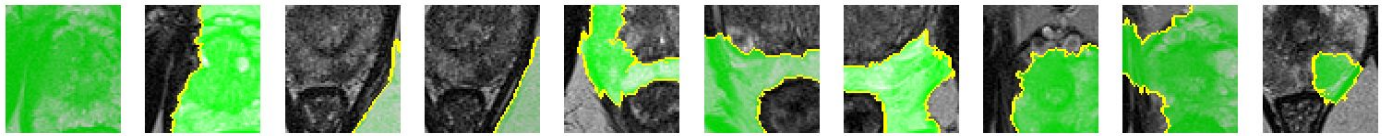


https://github.com/stefanoteso/awesome-explanatory-supervision

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

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 differents outputs, such as:
  - saliency maps (possibly with integrated gradients to be more precise):



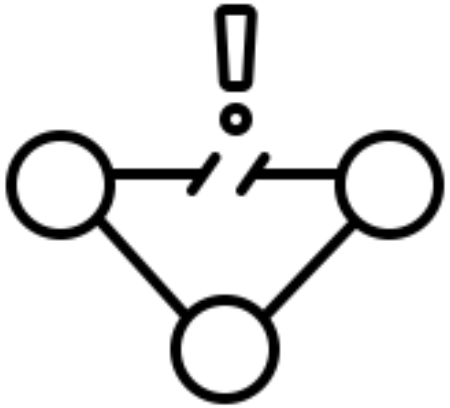
  - factual and counterfactual rules

```
r = { 29 > 0.50, 46 > 0.50, 47 > 0.50, 25 > 0.50, 58 > 0.50, 56 > 0.50, 40 > 0.50 } --> { class: 1 }
c = { { 46 <= 0.50 } --> { class: 0 }, { 40 <= 0.50 } --> { class: 0 }, { 58 <= 0.50 } --> { 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!

Reasoning over Uncertainty

collision: false

Pr(collision) = 0.1

Pr(collision) = 0.4

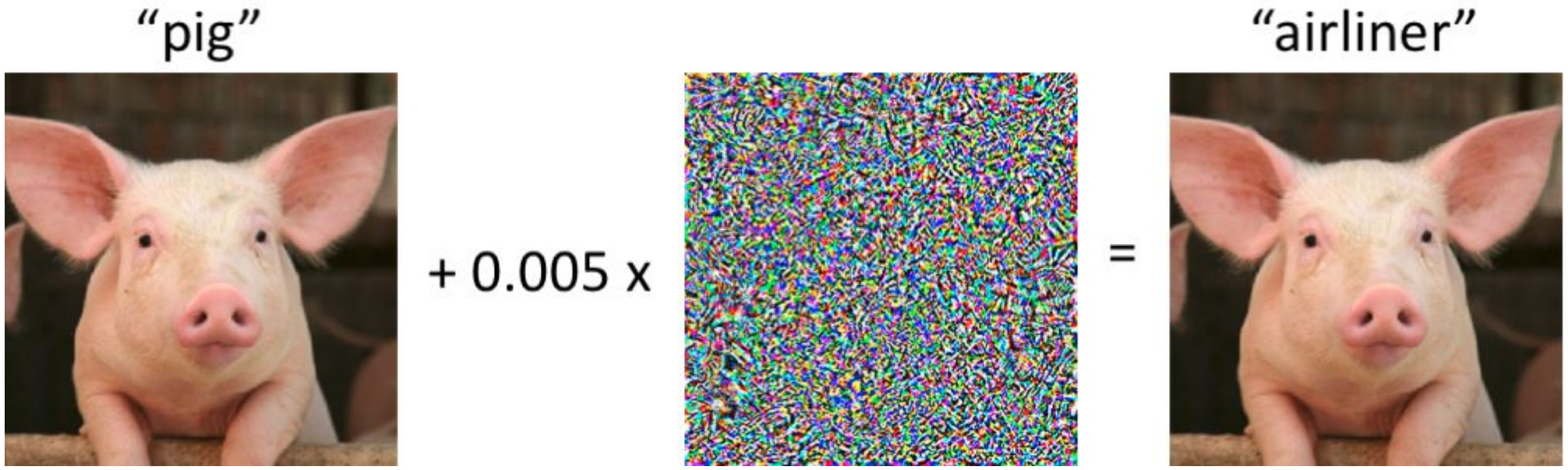
Reasoning over Predictions

Pr(collision now) = 0.0...

Pr(collision soon) = 0.5

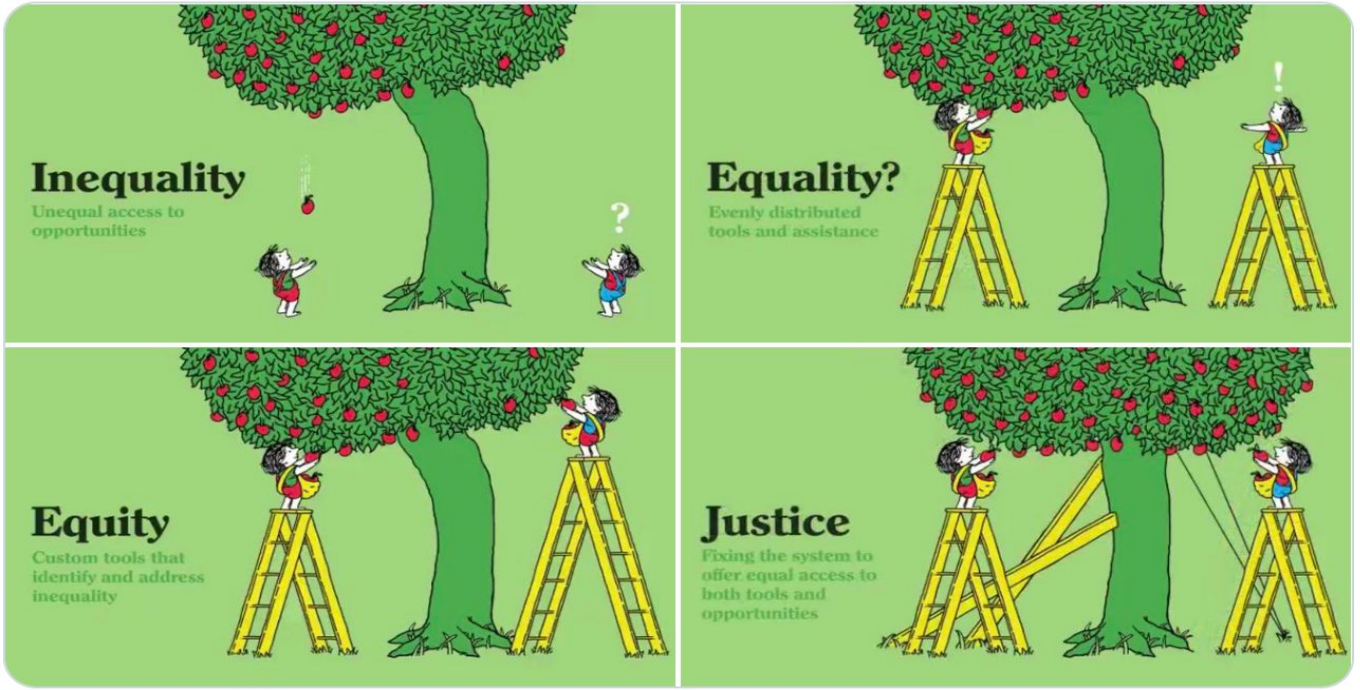
Probabilistic Predictive Stream Reasoning [Tiger and Heintz TIME 2016, IJAR 2020]

Dealing with truly Adversarial Examples



What is Fairness and Fairness-Aware ML?

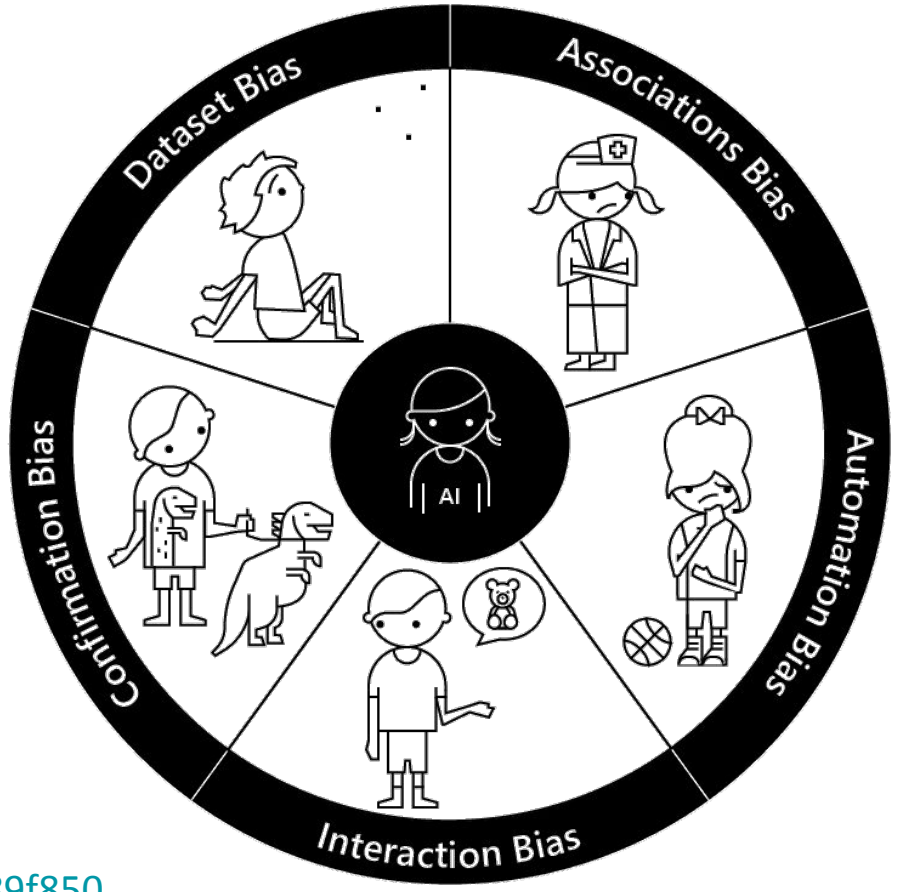
- Impartial and just treatment or behavior without favoritism or discrimination.



- *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

Artificial intelligence: Who is responsible for the diagnosis?

Emanuele Neri[1] · Francesca Coppola[2] · Vittorio Miele[3] · Corrado Bibbolino[4] · Roberto Grassi[5]

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.

TAILOR

SUBIG RESEARCH INFRASTRUCTURE

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

TAILOR

SOBIGDATA RESEARCH INFRASTRUCTURE

KDD Lab

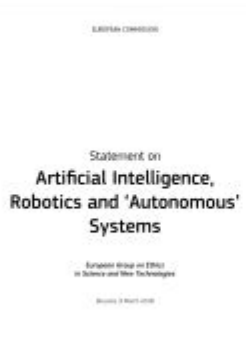
Istituto di Scienza e Tecnologie dell'Informazione "A. Faedo" Consiglio Nazionale delle Ricerche

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 se**
  (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

TAILOR
SOBIGDATA RESEARCH INFRASTRUCTURE
KDD Lab
Istituto di Scienza e Tecnologie dell'Informazione "A. Faedo" Consiglio Nazionale delle Ricerche

Respect for Privacy

Respect for Privacy

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

PbD
www.privacybydesign.ca

TAILOR

SOBIGDATA
RESEARCH INFRASTRUCTURE

KDD Lab

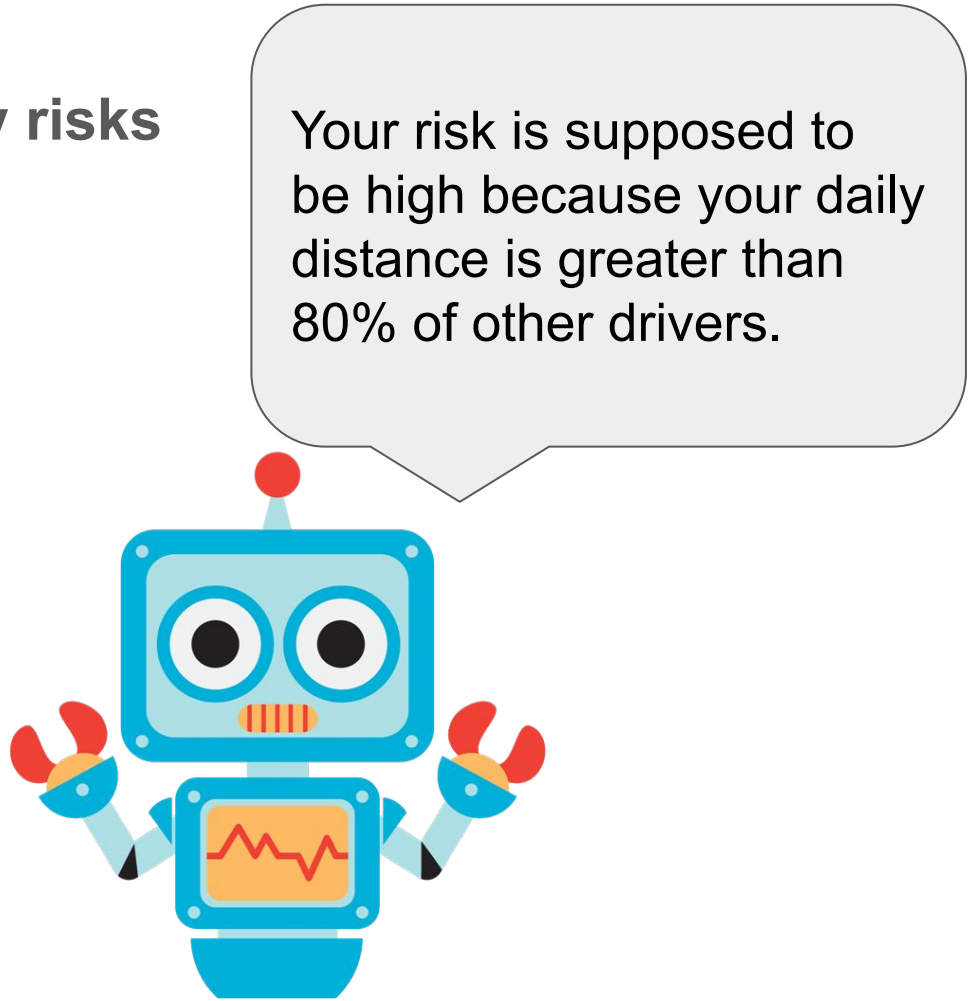
Istituto di Scienza e Tecnologie dell'Informazione "A. Faedo"
Consiglio Nazionale delle Ricerche

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**



TAILOR

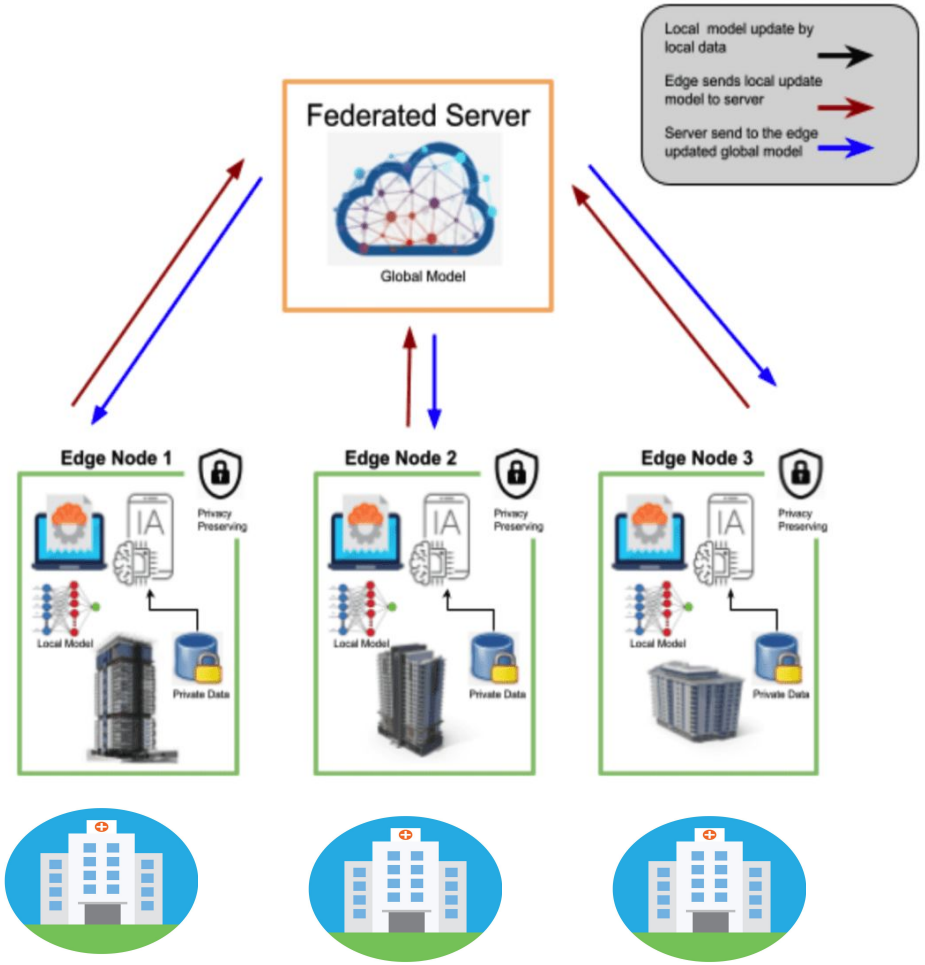
SOBIGDATA RESEARCH INFRASTRUCTURE

KDD Lab

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

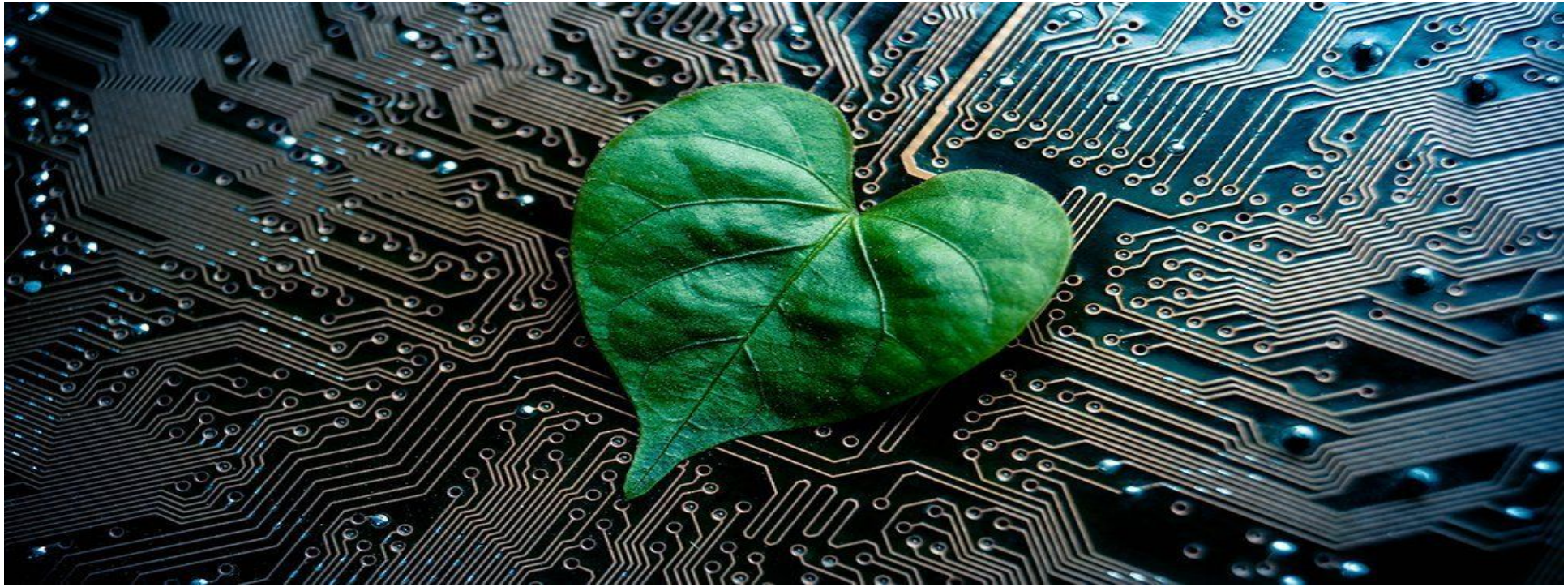
Federated Learning: Classical Setting

"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"



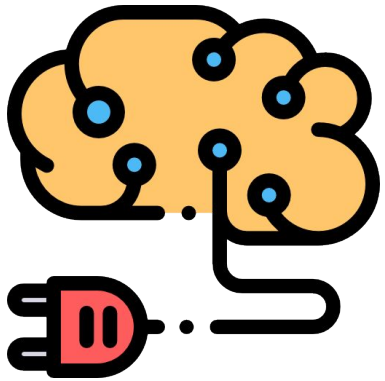
Sustainability

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



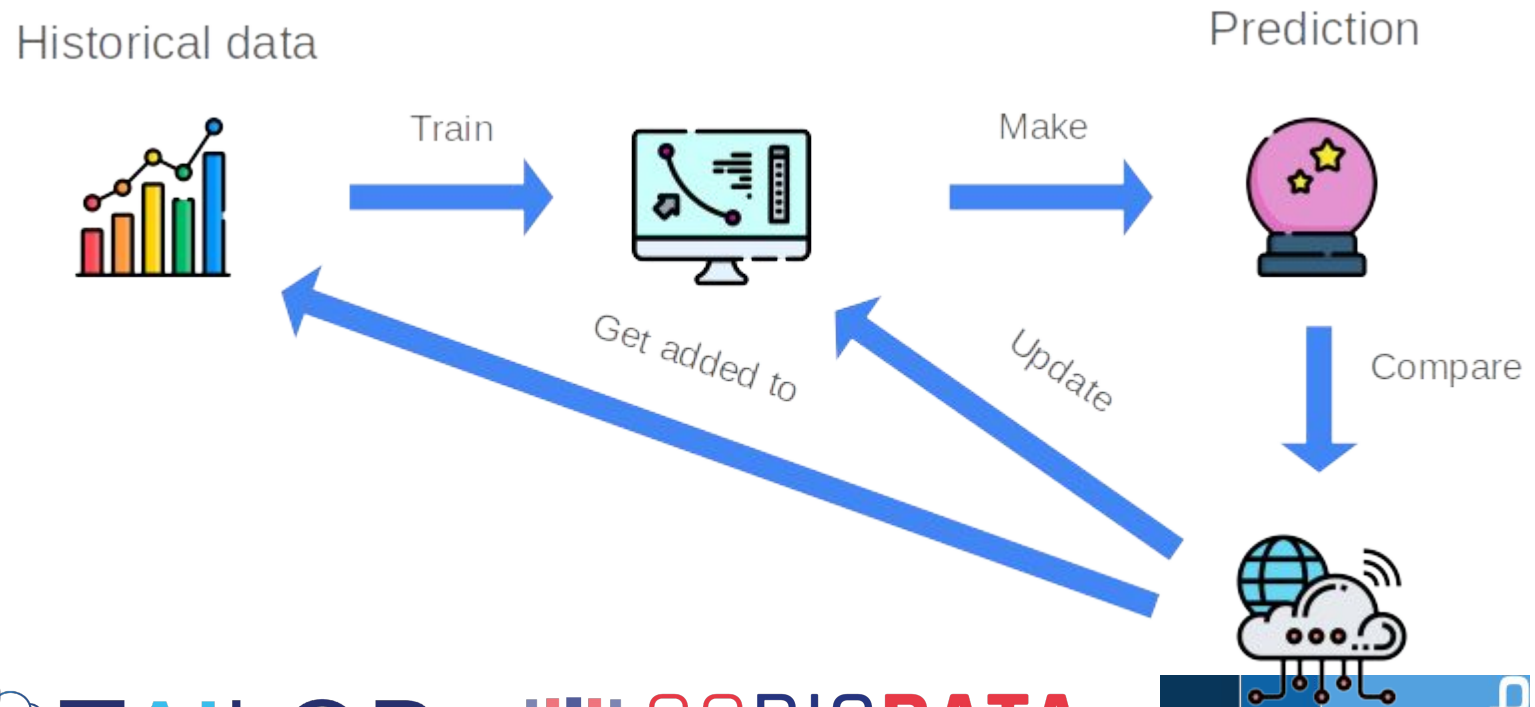
Research Challenge on Sustainability

- 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



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