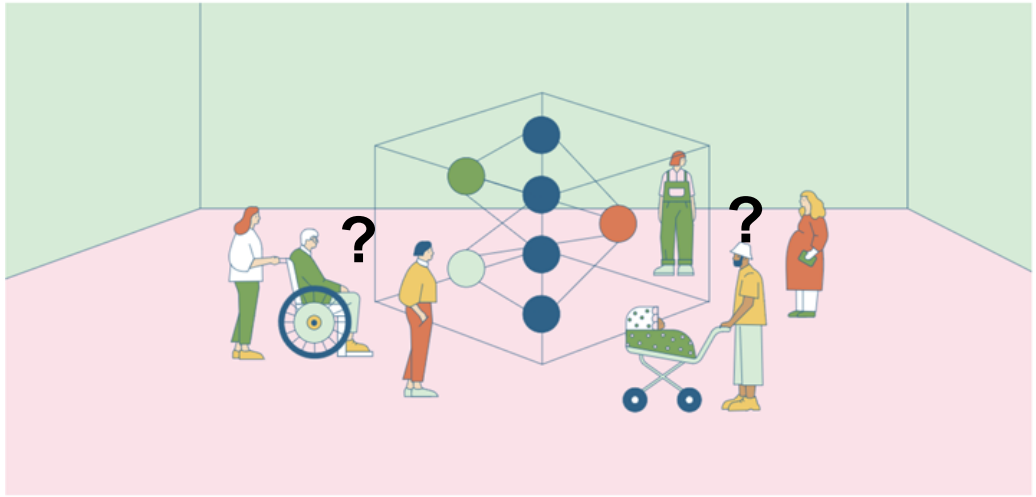
Équipes Traitement de l'Information et Systèmes

Explainability of Data Intensive AI Systems

Vassilis Christophides ENSEA, ETIS, IPAL
Evi Pitoura UOI, ATHENARC
Katerina Tzompanaki CYU, ETIS
Dimitris Kotzinos CYU, ETIS

Explain Automated Decisions



- The abundance of data offers many opportunities for technological innovations and for improved decision making in science, industry, daily life
- New challenges and dilemmas related to the way in which data-driven technologies are embedded in our society
- Critical decision making in high-stakes domains requires to ensure that relevant stakeholders are able to understand why a data-driven algorithm has come up to a certain result
  - Such an understanding helps determine if, when, and how much to rely on the automated decisions generated by data intensive systems

Classification of AI Systems Risk

Unacceptable-risk AI systems	High-risk AI systems	Limited- and minimal-risk AI systems
• Subliminal, manipulative, or exploitative techniques causing harm • Real-time, remote biometric identification systems used in public spaces for law enforcement • All forms of social scoring	• Systems that evaluate consumer creditworthiness • Recruiting or employee-management systems • Systems utilizing biometric identification in nonpublic spaces • Safety-critical systems (eg, systems that would put the health of citizens at risk due to failure) • Any systems used in the administration of justice	• AI chatbots • AI-enabled video and computer games • Spam filters • Inventory-management systems • Customer- and market-segmentation systems • Most other AI systems

https://gdprlocal.com/ai-risk-classification/ EU AI ACT Regulation Legislation

Individual's Rights under EU GDPR

The Right to Be Informed

Individuals have a right to know who is processing their personal data

The Right to Access

Individuals have the right to access any personal data that has been collected about them

The Right to Rectifications

Individuals have the right to require organizations to correct inaccurate personal data

The Right to Be Forgotten

Individuals have the right to have their personal data deleted and to prevent further collection

The Right to Restrict Processing

Individuals have the right to require organizations to restrict the processing of specific categories of personal data

The Right to Data Portability

Individuals have the right to require organizations to transfer personal data to a recipient of their choice

The Right to Object

Individuals have the right to consent, or withdraw consent, to the processing of their personal data

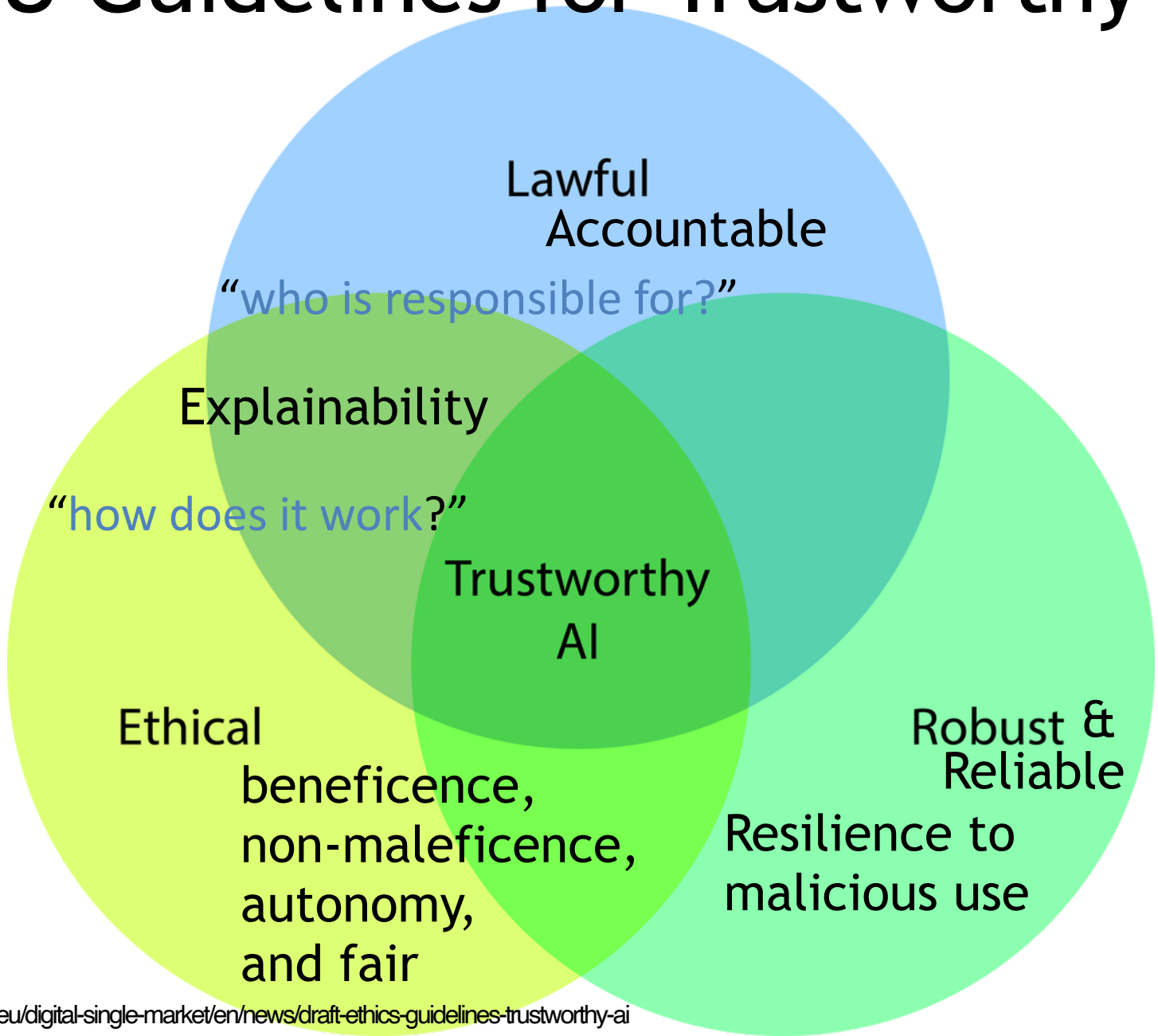
Rights in Relation to Automated Decision Making and Profiling

Individuals have the right to opt out of the use of their personal data by automated systems, such as artificial intelligence

https://www.universityofgalway.ie/itsecurity/gdpr/

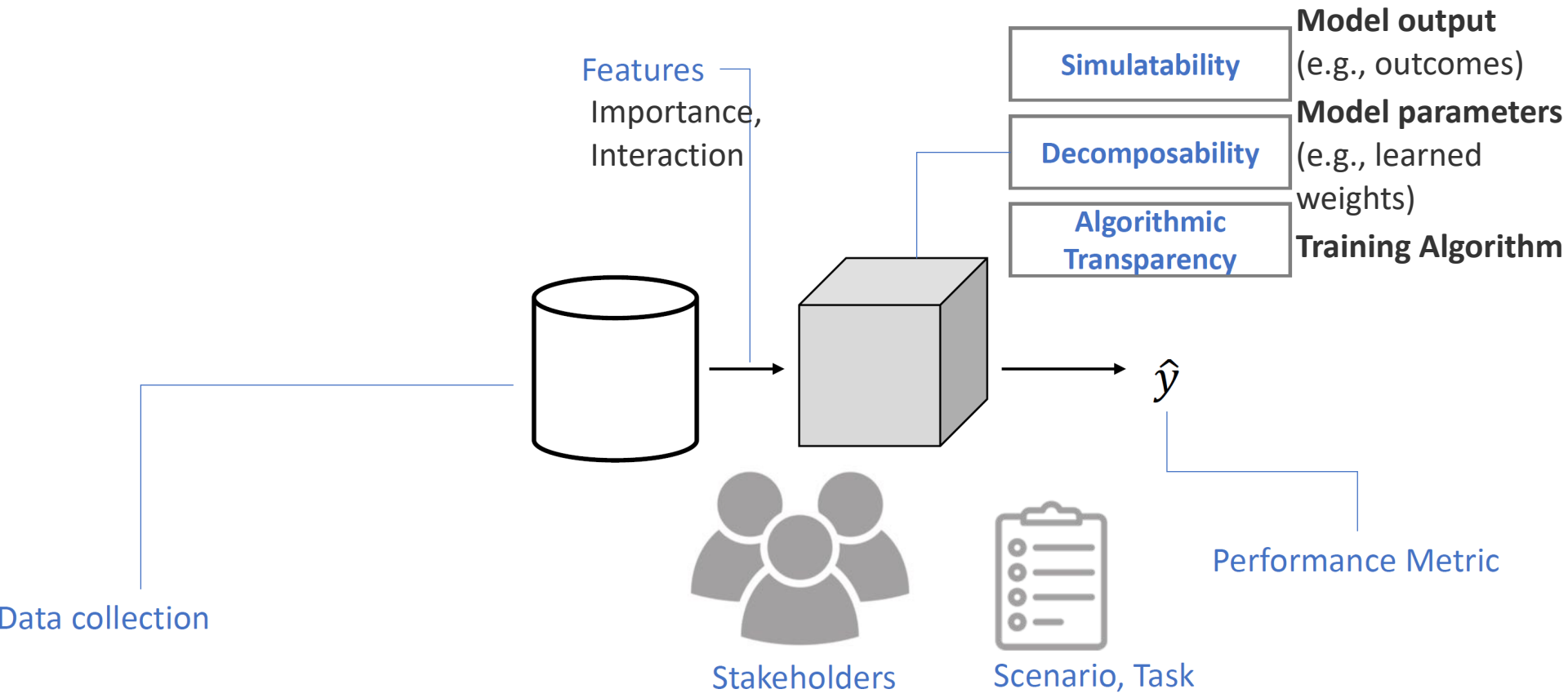
ETIS
Équipes Traitement de l'Information et Systèmes

EU Guidelines for Trustworthy AI



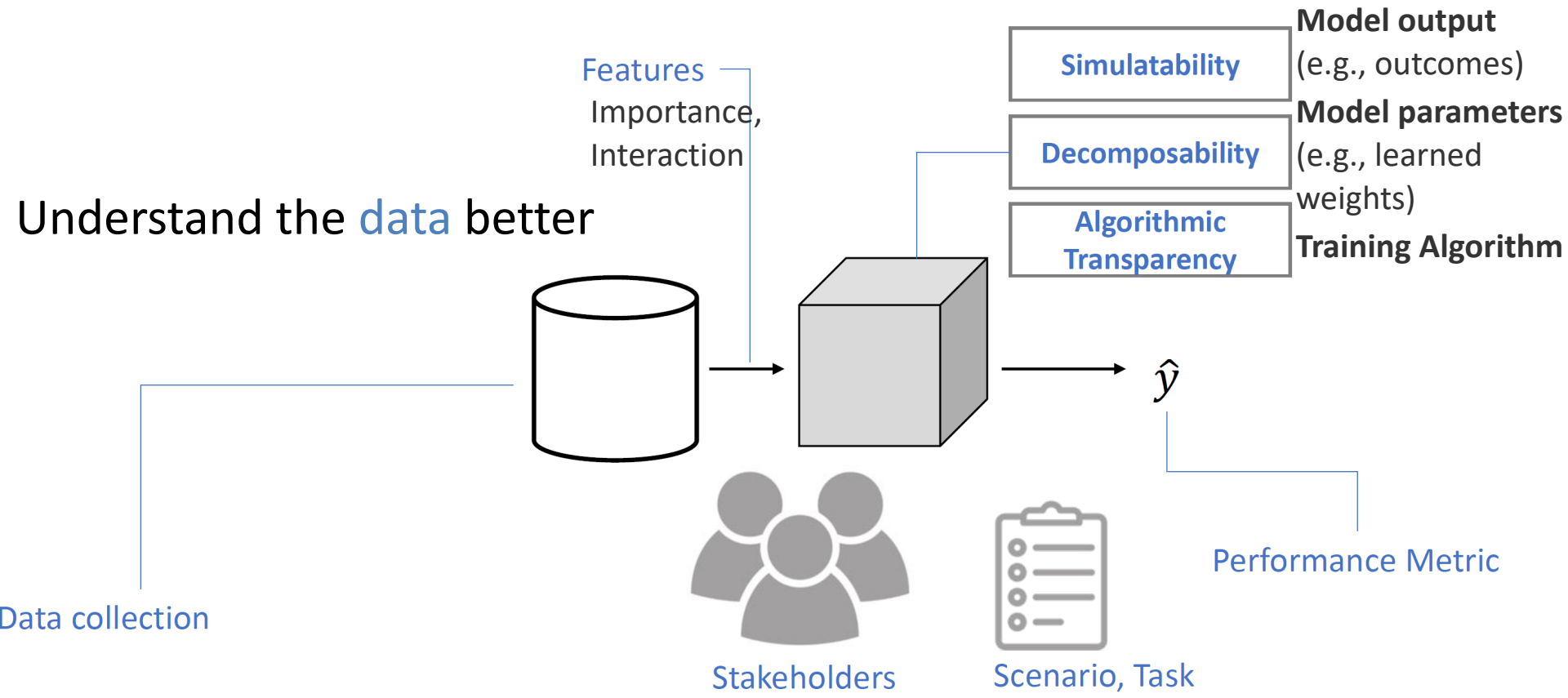
https://ec.europa.eu/digital-single-market/en/news/draft-ethics-guidelines-trustworthy-ai

Alternative Explanation Perspectives



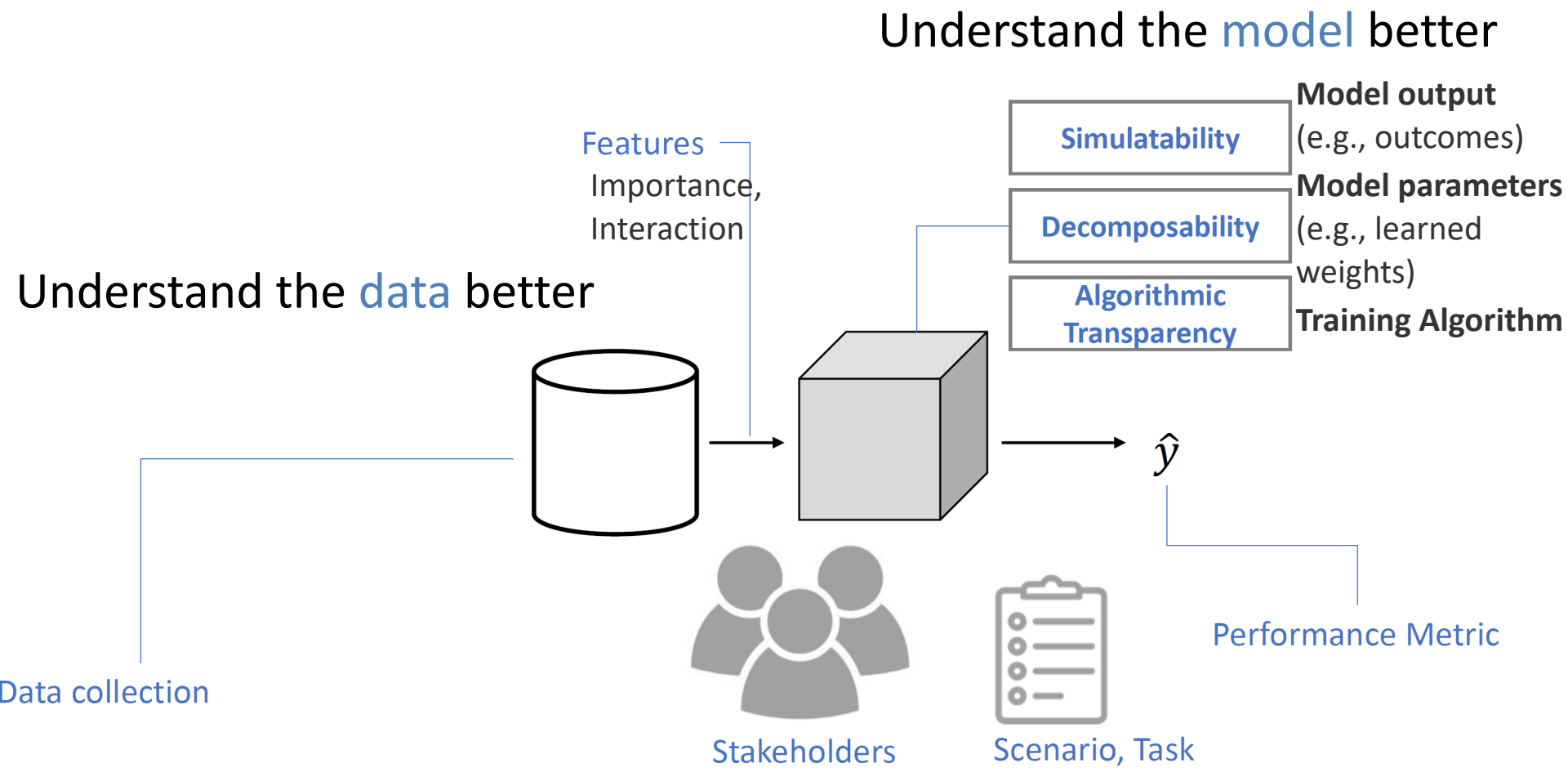
[C. Lipton 2016, Lepri et al. 2017, B. Mittelstadt et al. 2019, D. Weld, G. Bansal 2019]

Alternative Explanation Perspectives



[C. Lipton 2016, Lepri et al. 2017, B. Mittelstadt et al. 2019, D. Weld, G. Bansal 2019]

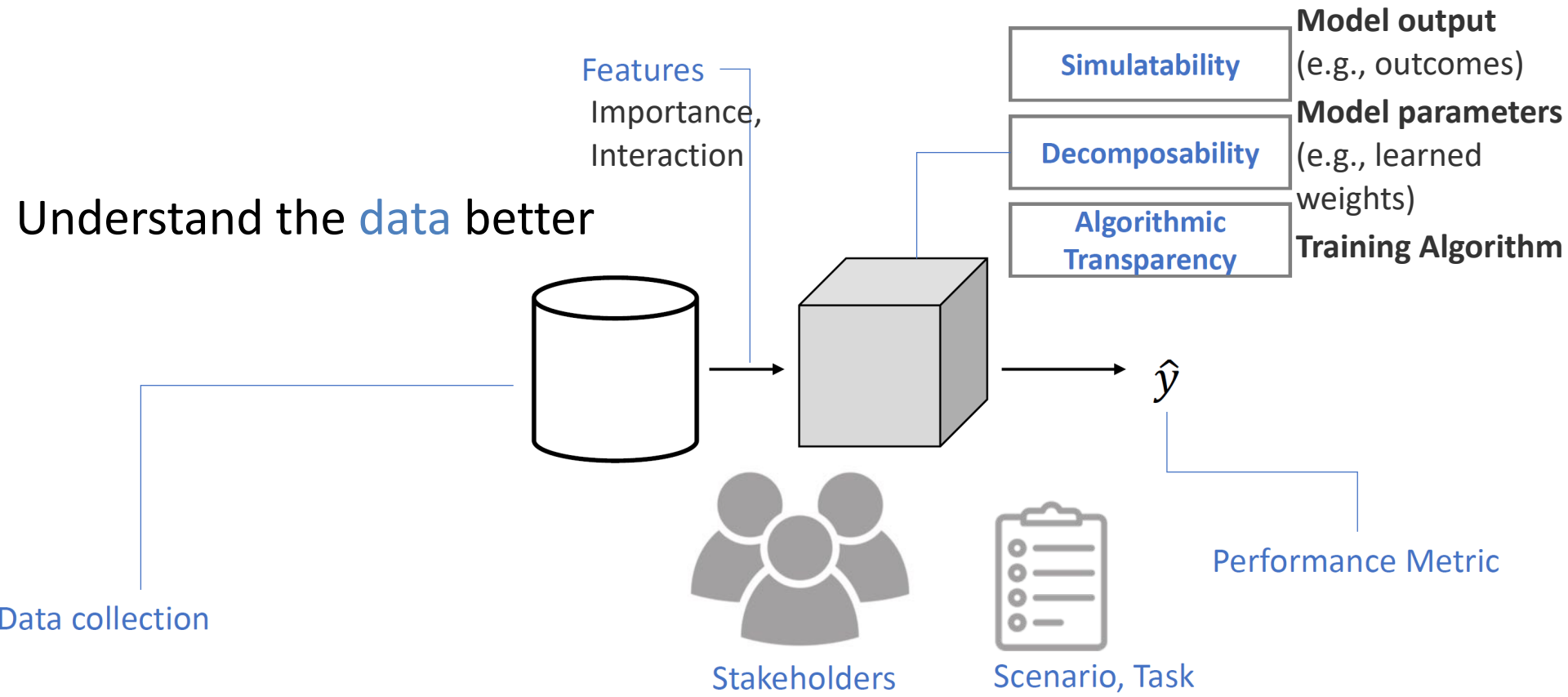
Alternative Explanation Perspectives



[C. Lipton 2016, Lepri et al. 2017, B. Mittelstadt et al. 2019, D. Weld, G. Bansal 2019]

Alternative Explanation Perspectives

Understand the model better



Understand user cognitive capacity & domain knowledge better

[C. Lipton 2016, Lepri et al. 2017, B. Mittelstadt et al. 2019, D. Weld, G. Bansal 2019]

ETIS
Équipes Traitement de l'Information et Systèmes

XAIDATA School Objectives

- The XAIDA Spring School aims to familiarize participants of different scientific backgrounds with the core concepts and methods in the emerging field of eXplainable Artificial Intelligence (XAI)
- We will understand explainability from alternative perspectives:
  - Different Families of Methods
    - ✓ Feature vs Data Attribution, Local vs Global, Associative vs Causal
  - Different Processing/Modeling Tasks
    - ✓ Classification, Regression/Forecasting, Query Answering, Retrieval Augmented and Concept-based Generation
  - Different Data Modalities
    - ✓ Tabular Data, Images, Time Series, Graphs
  - Different Evaluation Techniques
    - ✓ Quantitative, Qualitative
  - Different Domains
    - ✓ Healthcare, Predictive Maintenance, Material Science, etc.

Équipes Traitement de l'Information et Systèmes

XAIDATA Program

Thursday 28/5/2026 (Day 1)

Time	Title	Speakers/Notes	Room
9:00-9:15	**School Opening**	Organisers	Auditorium
9:15-10:00	**Explainability Basics**	V. Christophides, E.Pitoura, K.Tzompanaki	Auditorium
10:00-11:00	**Counterfactuals for Fairness and Explainability**	D. Sacharidis	Auditorium
11:00 -11:30	COFFEE BREAK	COFFEE BREAK	Salle 111
11:30-12:30	**Causal Feature Selection for Time Series Forecasting**	E. Vareille	Auditorium
12:30 -13:45	LUNCH BREAK	LUNCH BREAK	Salle 111
14:00- 14:45	**Concept-Based Explainability**	I. Falih	Auditorium
14:45-15:15	**Explainability for Time-to-Event Predictions**	A. Giannoulidis	Auditorium
15:15 -15:30	COFFEE BREAK	COFFEE BREAK	Salle 111
15:30-17:00	**Hands on Workshop: Experimenting with different explainability methods for Predictive Maintenance**	A. Giannoulidis, E. Vareille	M03

Équipes Traitement de l'Information et Systèmes

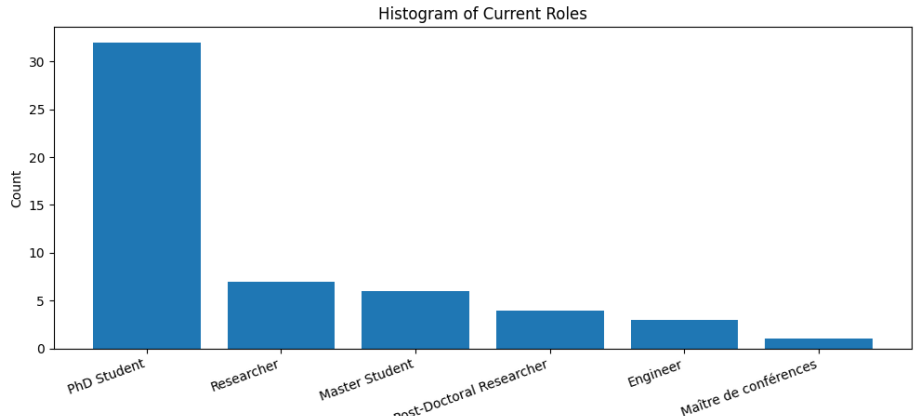
XAIDATA Program

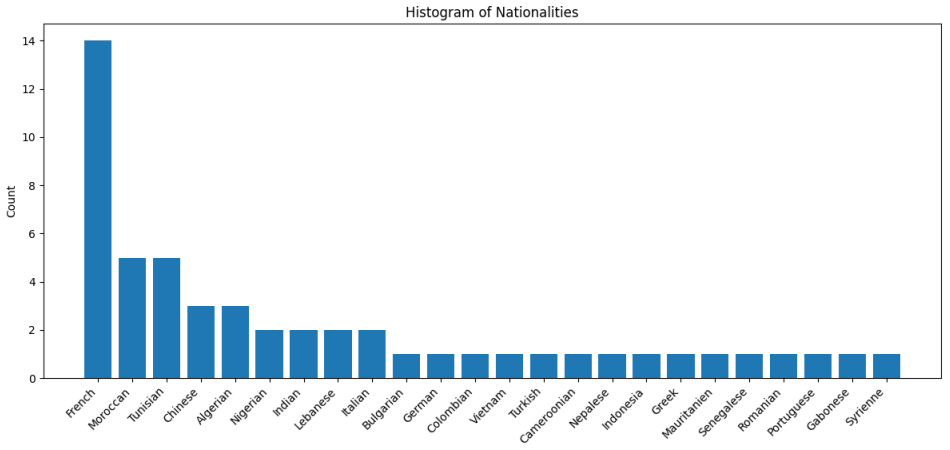
Friday 29/5/2026 (Day 2)

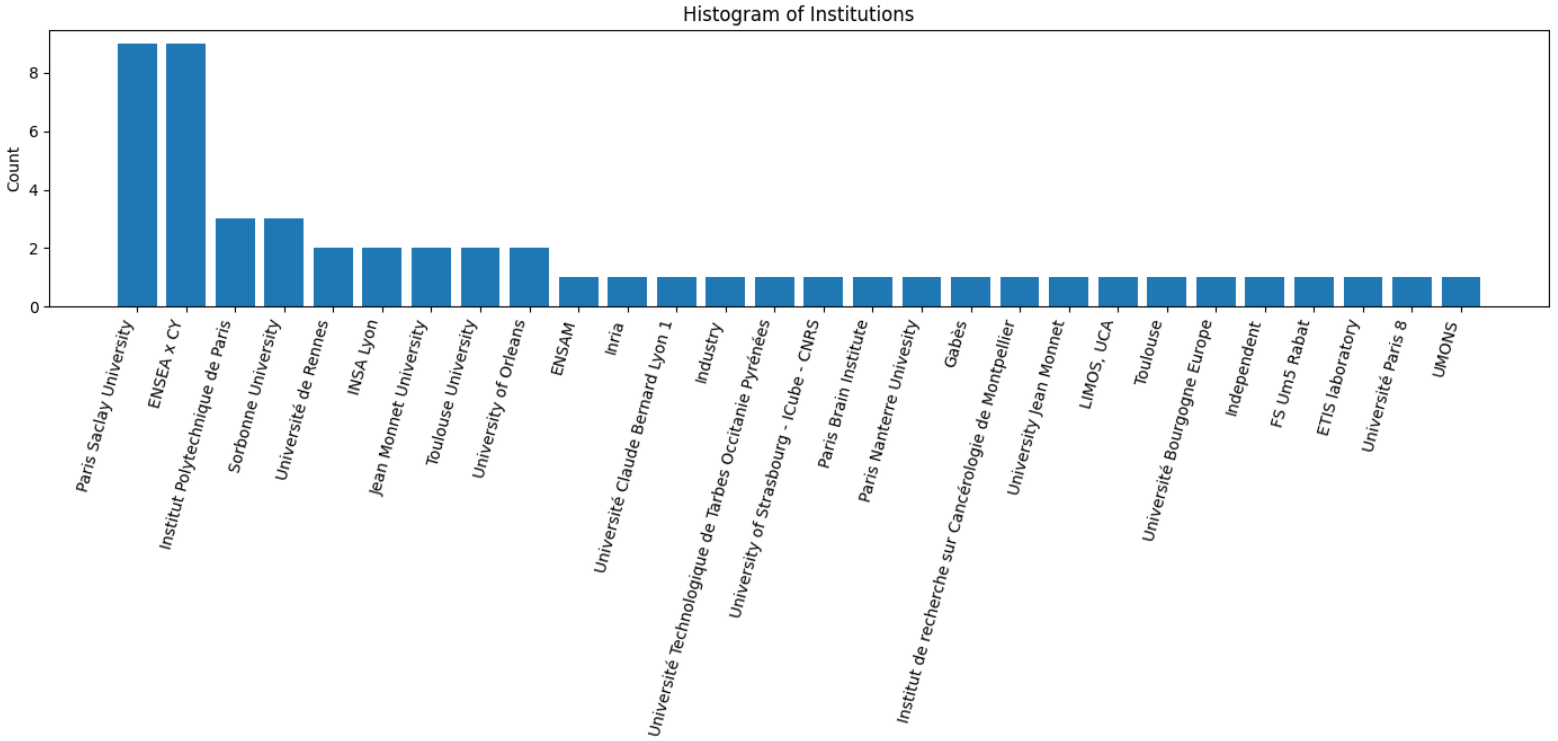
Time	Title	Speakers / Notes	Room
09:00-10:00	**Explainability for Retrieval Augmented Generation**	E.Pitoura	Auditorium
10:00-11:00	**Explainability for Graph Tasks**	G. Renton	Auditorium
11:00 -11:30	COFFEE BREAK	COFFEE BREAK	Salle 111
11:30-12:30	**Explaining Queries on Inconsistent Databases**	B. Raddaoui, K. Tzompanaki, Y. Gu, Y. Ma	Auditorium
12:30 -13:45	LUNCH BREAK	LUNCH BREAK	Salle 111
14:00-15:00	**Hands on Workshop: Evaluating Explainability through user studies.**	L. Galarraga	Auditorium
15:00 -15:20	COFFEE BREAK	COFFEE BREAK	Salle 111
15:20-17:00	**Hands on Workshop: Evaluating Explainability through user studies.**	L. Galarraga	M03

XAIDATA Participant Statistics

- 53 applications
- 32 accepted







ETIS

Équipes Traitement
de l'Information
et Systèmes