Privacy by Design in Big Data Analytics & AI

Anna Monreale

Dipartimento di Informatica

Università di Pisa

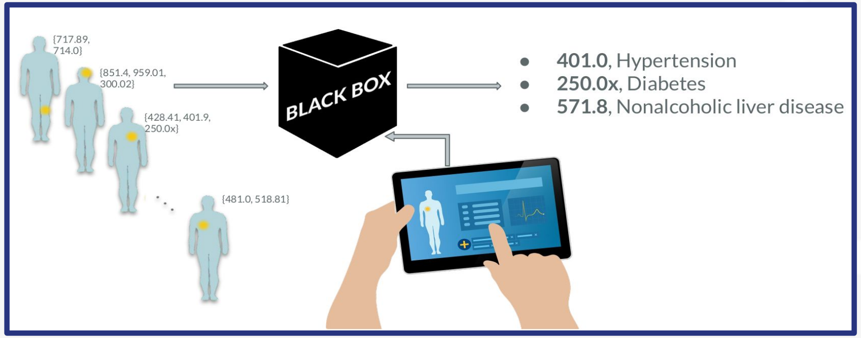
email: anna.monreale@unipi.it

UNIVERSITÀ DI PISA

SOBIGDATA
RESEARCH INFRASTRUCTURE

KDD Lab

AI & Big Data



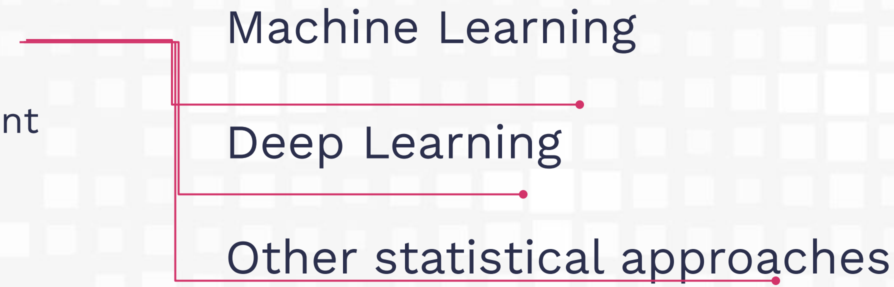
- **401.0**, Hypertension
- **250.0x**, Diabetes
- **571.8**, Nonalcoholic liver disease

WHAT IS A.I.?

A practical definition of AI

‘**Artificial intelligence system**’ (AI system) means a system that

1. receives machine and/or **human-based data** and inputs
2. infers how to achieve a given set of human-defined objectives using learning, reasoning or modelling implemented with the techniques and approaches listed in **Annex I**
3. generates outputs in the form of content (generative AI systems), predictions, recommendations or decisions, which influence the environments it interacts with.



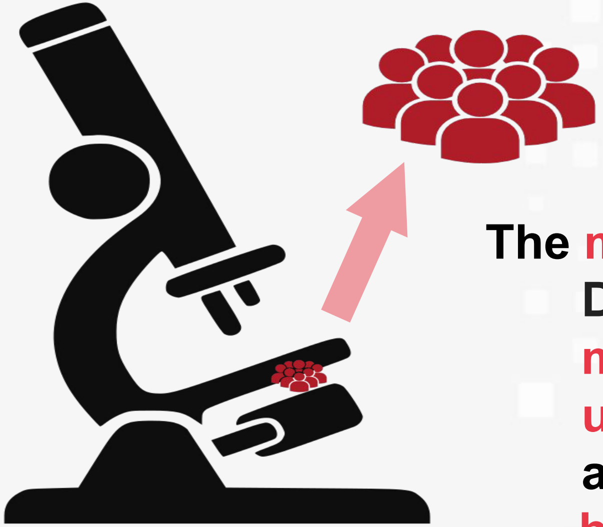
Machine Learning

Deep Learning

Other statistical approaches

AI Act, TITLE I, Article 3

AI, Big Data Analytics & Social Mining



The **main tool** for a Data Scientist to **measure, understand,** and possibly **predict human behavior**

Data Scientist needs to take into account ethical and legal aspects and social impact of data science & AI

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

Trustworthy AI as our foundational ambition, with three components

Lawful AI

complying with all applicable laws and regulations

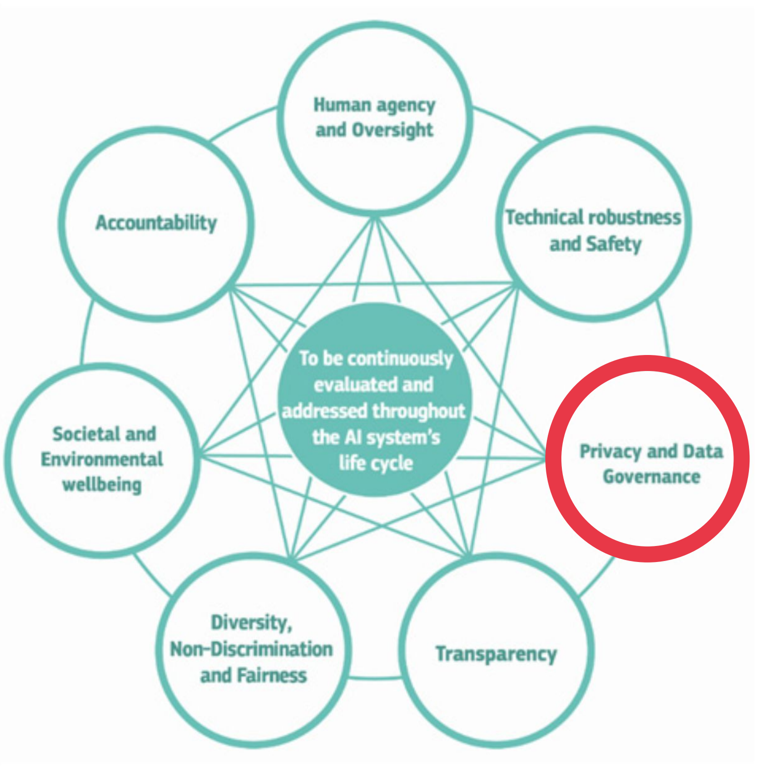
Ethical AI

ensuring adherence to ethical principles and values

Robust AI

perform in a safe, secure and reliable manner, both from technical and a social perspective, with safeguards to foresee and prevent unintentional harm

Requirements


Human agency and Oversight
Technical robustness and Safety
Privacy and Data Governance
Transparency
Diversity, Non-Discrimination and Fairness
Societal and Environmental wellbeing
Accountability
To be continuously evaluated and addressed throughout the AI system's life cycle

General Data Protection Regulation

PRIVACY

Personal Data

Personal data is defined as **any information** relating to an identity or identifiable natural person.

- Your name
- Home address
- Photo
- Email address
- Bank details
- .....

An **identifiable person** is one who can be identified, **directly or indirectly,** in particular by reference to an identification number or to one or more factors specific to his physical, physiological, mental, economic, cultural or social identity.

Sensitive Data

Sensitive personal data is a specific set of “**special categories**” that must be treated with extra security

- Racial or ethnic origin
- Political opinions
- Religious or philosophical beliefs
- Trade union membership
- Genetic data
- Biometric data

Anonymity according to 1995/46/EC

- The principles of protection must apply to any information concerning an identified or identifiable person

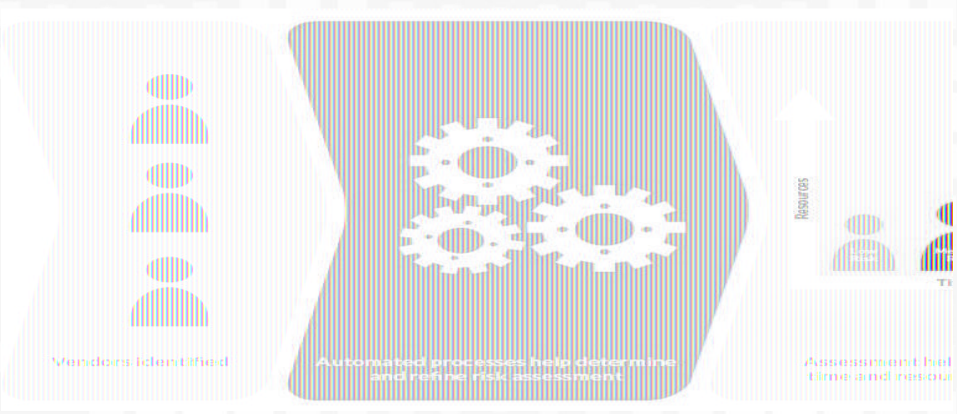
- To determine whether a person is identifiable, account should be taken of all the means likely reasonably to be used either by the controller or by any other person to identify the said person

- The principles of protection shall not apply to data rendered anonymous in such a way that the data subject is no longer identifiable

Privacy by Design Principle

- Privacy by design is an approach to protect privacy by inscribing it into the design specifications of information technologies, accountable business practices, and networked infrastructures, from the very start

- Developed by Ontario's Information and Privacy Commissioner, Dr. Ann Cavoukian, in the 1990s


Vendors identified
Automated processes help determine and refine risk assessment
Resources
Assessment hel
time and resou

Privacy Risk Assessment

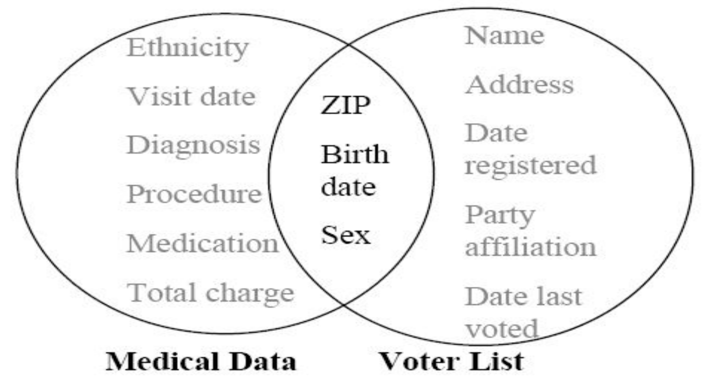
Privacy by Design

Why accessing data we can jeopardize individual privacy?

Privacy risk as Re-identification risk

- Sweeney managed to re-identify the medical record of the governor of Massachusetts
- MA collects and publishes sanitized medical data for state employees (microdata) **left circle**
- voter registration list of MA (publicly available data) **right circle**

- looking for governor's record & joining the tables:
  - 6 people had his birth date
  - 3 were men
  - 1 in his zipcode



Latanya Sweeney: k-Anonymity: A Model for Protecting Privacy. International Journal of Uncertainty, Fuzziness and Knowledge-Based Systems 10(5): 557-570 (2002)

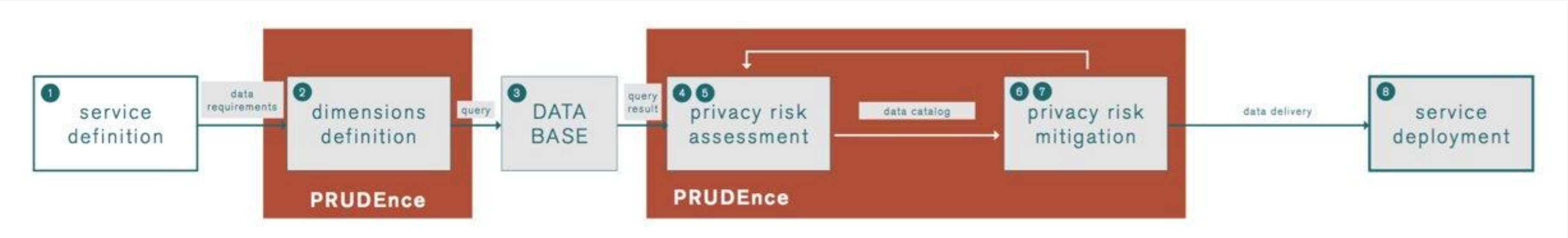
Linking Attack

Governor: birth date = **1950**, ZIP = **300111**

ID	Gender	YoB	ZIP	DIAGNOSIS
1	F	1962	300122	Cancer
2	F	1960	300133	Gastritis
3	M	1950	300111	Heart Attack
4	M	1955	300112	Headache
5	F	1965	300200	Dislocation
6	M	1953	300115	Fracture

Which is the disease of the Governor?

Data-Driven Privacy Risk Assessment



[1] Pratesi, F., Monreale, A., Trasarti, R., Giannotti, F., Pedreschi, D., Yanagihara, T.: Prudence: a system for assessing privacy risk vs utility in data sharing ecosystems. TDP 2018

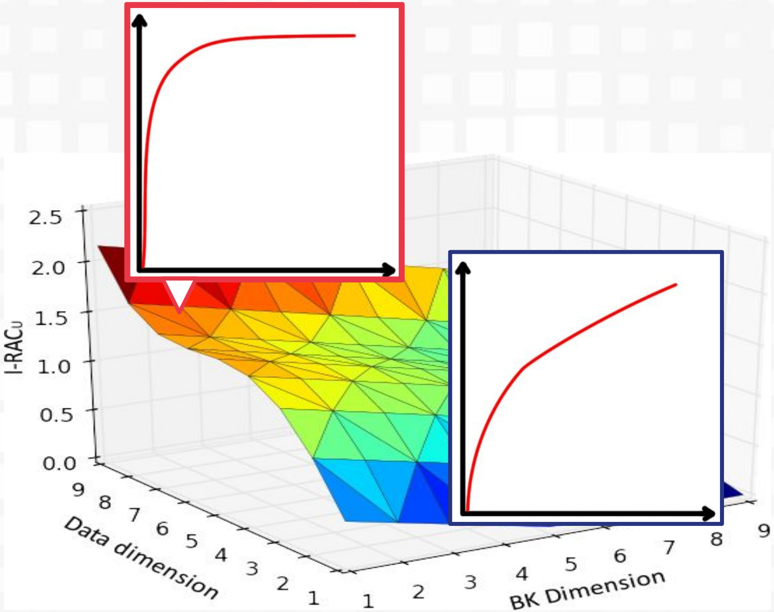
Privacy Risk Assessment Framework for Data Sharing

For each:

- **Data Format**, i.e., the data needed for the service
- **Risk Assessment Setting**, i.e., the set of pre-processing and privacy attacks

The Data Catalog provides:

- **Quantification of Privacy Risk**, i.e., the evaluation of the real risk of re-identification
- **Quantification of Data Quality**, i.e., the quality level we can achieve with private data, compared with the data quality of original data.



Linking Attack

Governor: birth date = **1950**, ZIP = **300111**

ID	Gender	YoB	ZIP	DIAGNOSIS
1	F	1962	300122	Cancer
2	F	1960	300133	Gastritis
3	M	1950	300111	Heart Attack
4	M	1955	300112	Headache
5	F	1965	300200	Dislocation
6	M	1953	300115	Fracture

Which is the disease of the Governor?

Making data anonymous

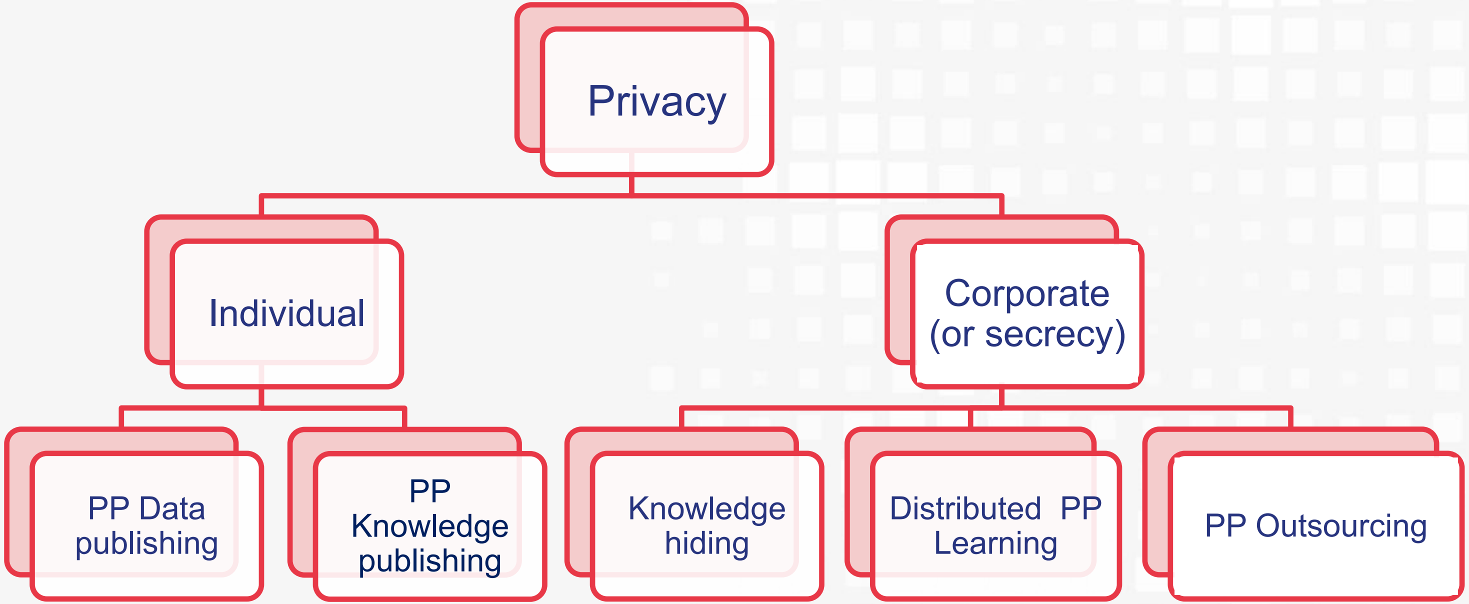
K-anonymity

Governor: Birth Date **= 1950,** ZIP = **300111**

ID	Gender	YoB	ZIP	DIAGNOSIS
1	F	[1960-1956]	300***	Cancer
2	F	[1960-1956]	300***	Gastritis
3	M	[1950-1955]	30011*	Heart Attack
4	M	[1950-1955]	30011*	Headache
5	F	[1960-1956]	300***	Dislocation
6	M	[1950-1955]	30011*	Fracture

Which is the disease of the Governor?

Ontology of Privacy Mitigation



Privacy

Individual

Corporate (or secrecy)

PP Data publishing

PP Knowledge publishing

Knowledge hiding

Distributed PP Learning

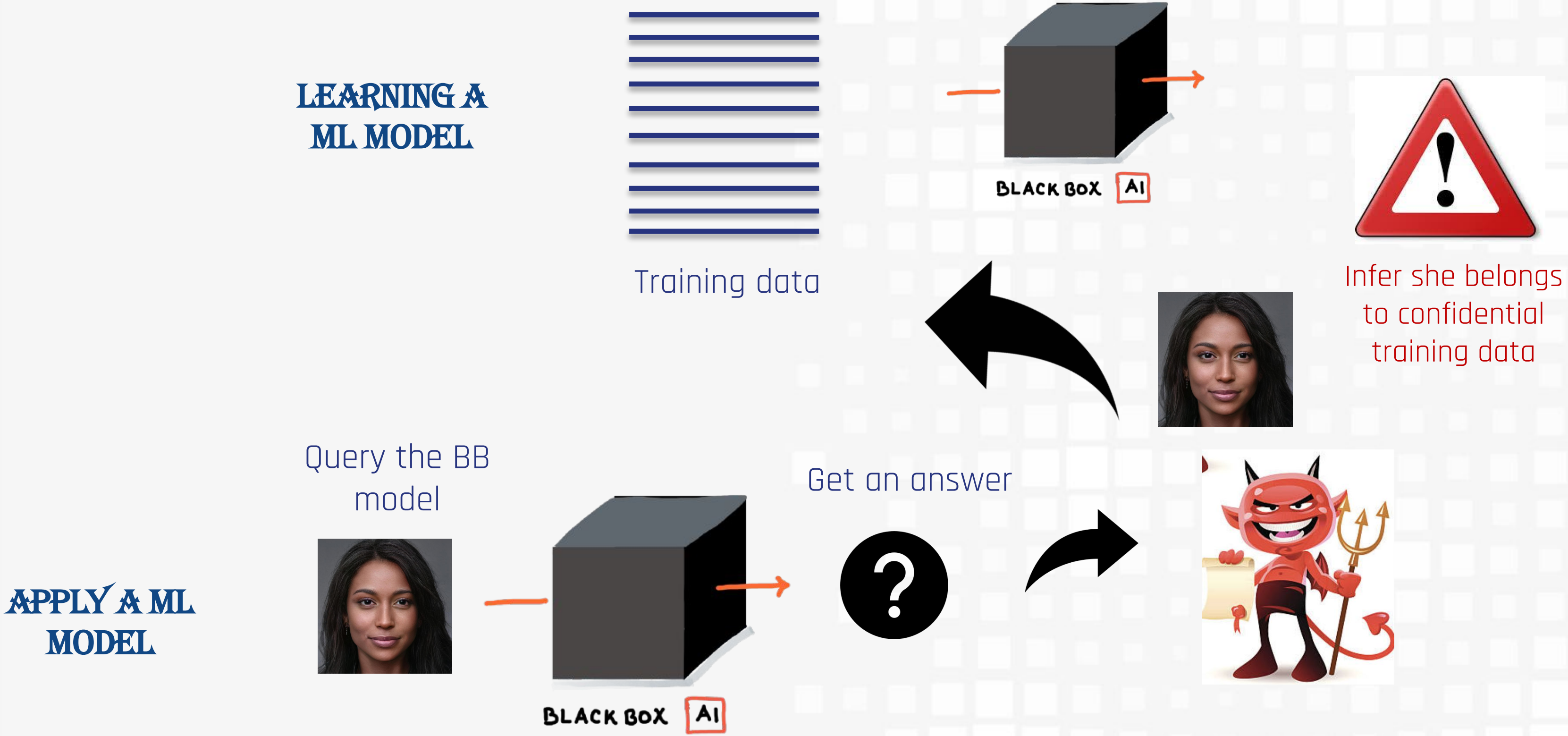
PP Outsourcing

Anonymity based Approaches

Randomization based Approaches

Can we jeopardize individual privacy without accessing data?

Privacy risk of ML models



LEARNING A ML MODEL

Training data

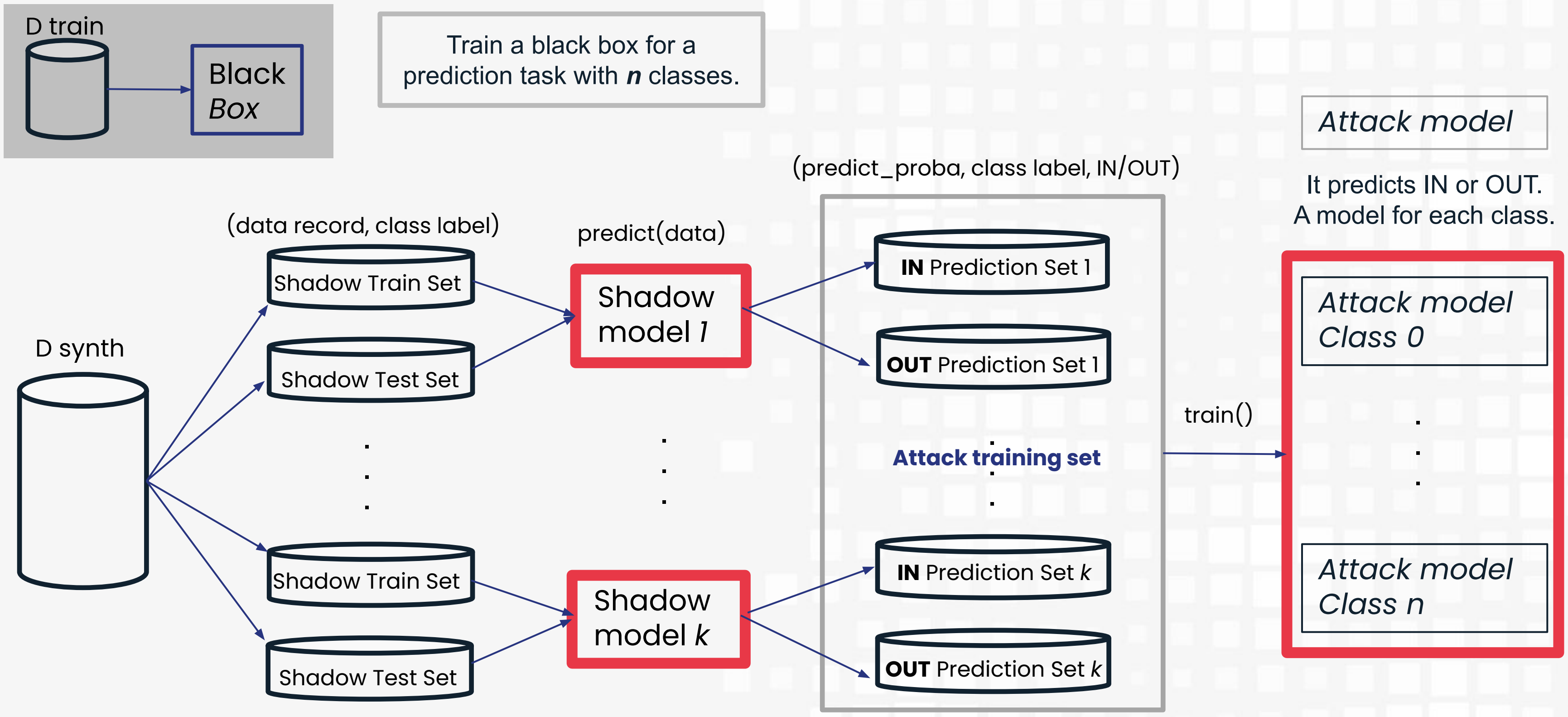
Infer she belongs to confidential training data

Query the BB model

Get an answer

APPLY A ML MODEL

The privacy attack: MIA

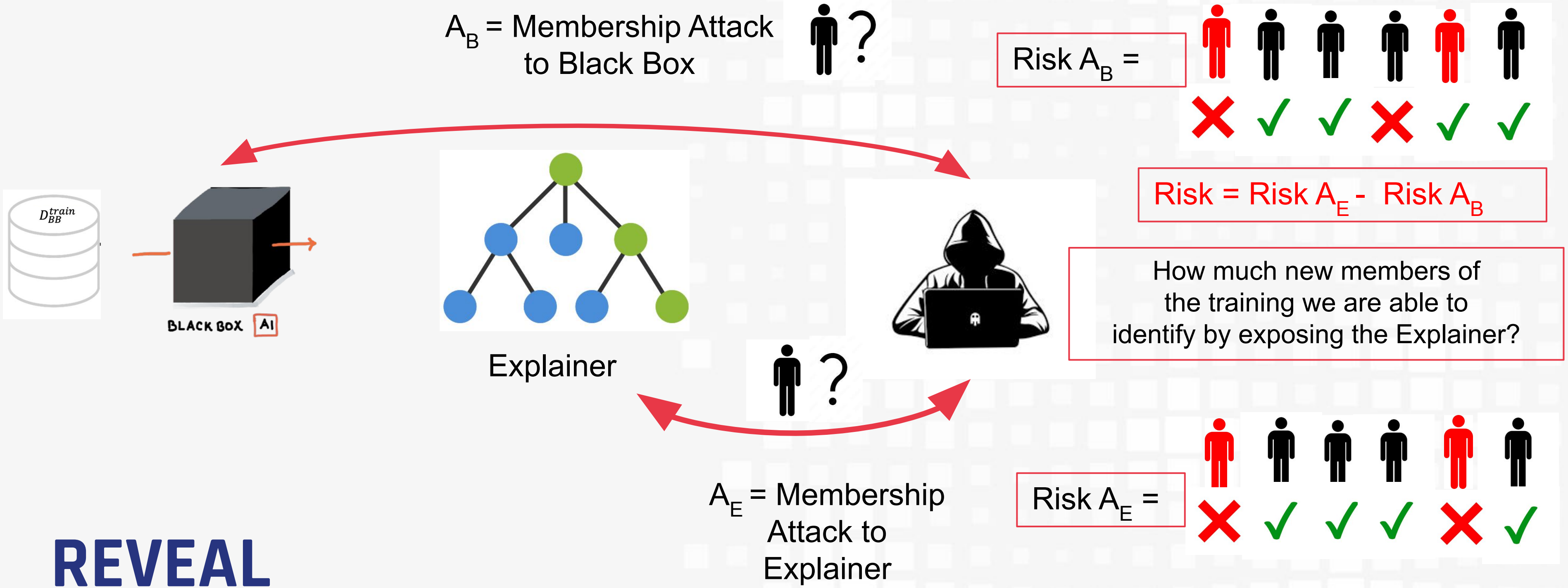


Train a black box for a prediction task with ***n*** classes.

It predicts IN or OUT.
A model for each class.

Reza Shokri, Marco Stronati, Congzheng Song, and Vitaly Shmatikov. Membership inference attacks against machine learning models. In 2017 IEEE Symposium on Security and Privacy

What about Explainers and privacy risks?



A_B = Membership Attack to Black Box

Risk A_B = ✗ ✓ ✓ ✗ ✓ ✓

Risk = Risk A_E - Risk A_B

How much new members of the training we are able to identify by exposing the Explainer?

A_E = Membership Attack to Explainer

Risk A_E = ✗ ✓ ✓ ✓ ✗ ✓

REVEAL

Conclusion: how to address privacy issues?

- Privacy-by-design: a proactive approach to privacy protection
- Assessing the privacy risk in training data
- Assessing the privacy risk of AI and potential XAI models
- Mitigating privacy risks by balancing protection and data utility

THANK YOU!

QUESTIONS?

SOBIGDATA
RESEARCH INFRASTRUCTURE