Data Privacy-Preservation Mechanisms

Current Theory and Limitations

Osonde Osoba, Ph.D.
October 2019

RAND

Finding our Bearings

Basic Premise:
Exponential growth of our data ecosystem. Parts of that ecosystem is relevant to health.

Goal:
To use available data to improve healthcare services (utility)

Normative Constraint :
Maintain “Privacy” of health subjects

Confidentiality Preserving authorized restrictions on information access and disclosure, including means for protecting personal privacy and proprietary information. [44 U.S.C., SEC. 3542]
Integrity Guarding against improper information modification or destruction, and includes ensuring information non-repudiation and authenticity. [44 U.S.C., SEC. 3542]
Availability Ensuring timely and reliable access to and use of information. [44 U.S.C., SEC. 3542]

Modes of Use

Access
- ➢ Basic Individual Record Access
- ➢ Batch Record Access

(e.g. for population monitoring)

Modeling
- ➢ Individual Prediction Models

(e.g. diagnostics)
- ➢ Population-level Models

(e.g. epidemic forecasts)

Davis, John S., and Osonde Osoba. "Improving privacy preservation policy in the modern information age." *Health and Technology* 9, no. 1 (2019): 65-75.

Typical Problem Statement

Age	Zip	Diagnosis
28	90145	**Measles**
21	90141	**Flu**
21	92238	**Flu**
55	92256	**Cancer**
53	92124	**Obesity**
48	92204	**Obesity**

Quasi-identifiers: Age, Zip
Sensitive Attribute: Diagnosis

The goal is to prevent “disclosures” i.e. the exposure of sensitive information *linked* to specific subjects. Subject specified by identifiers + quasi-identifiers. Think of it as a game between data owner and a curious adversary.

Privacy-Preservation Techniques (1/2)

Quasi-identifiers | Sensitive Attribute

Age	Zip	Diagnosis
28	90145	**Measles**
21	90141	**Flu**
21	92238	**Flu**
55	92256	**Cancer**
53	92124	**Obesity**
48	92204	**Obesity**

Age	Zip	Diagnosis
[21-28]	9****	**Measles**
[21-28]	9****	**Flu**
[21-28]	9****	**Flu**
[48-55]	92***	**Cancer**
[48-55]	92***	**Obesity**
[48-55]	92***	**Obesity**

K-Anonymity ensures that subjects cannot be identified and linked to sensitive attributes with certainty.

k-Anonymity:

- basic intuitive privacy mechanism
- Still imperfect. Homogeneity attacks

L-Diversity:

Updates k-anonymity. Requires >L diversity in sensitive attributes per sub-class

Skewness Attack:

A form of disclosure or privacy breach e.g. anonymized table discloses that anyone under 30 in the database more likely has flu.

Privacy-Preservation Techniques (2/2)

	Zip	Age	Disease
1	47677	29	gastric ulcer
2	47602	22	gastritis
3	47678	27	stomach cancer
4	47905	43	gastritis
5	47909	52	flu
6	47906	47	bronchitis
7	47605	30	bronchitis
8	47673	36	pneumonia
9	47607	32	stomach cancer

	Zip	Age	Disease
1	4767*	<40	gastric ulcer
3	4767*	<40	stomach cancer
8	4767*	<40	pneumonia
4	4790*	>40	gastritis
5	4790*	>40	flu
6	4790*	>40	bronchitis
2	4760*	<40	gastritis
7	4760*	<40	bronchitis
9	4760*	<40	stomach cancer

“Syntactic” vs. “Semantic” Privacy
No syntactic operations will prevent all info disclosure.
Focus on limiting what info can be learnt from a subject being in a database

t-Closeness:
t-closeness tries to make sub-groups indistinguishable in the sensitive distribution from the full table.

Differential Privacy

Table D_1	
Row	Income
1	50,000
2	58,000
3	72,000
4	59,000
5	68,000

vs.

Table D_2	
Row	Income
1	50,000
2	58,000
3	72,000
4	59,000
5	68,000
6	350,000

$$\forall B, \qquad \frac{P(M(D_1) \in B)}{P(M(D_2) \in B)} < e^{\varepsilon}$$

- Differential Privacy puts a bound on information disclosure resulting from inclusion in a database.
- Limitations:
  - Subjective choice for epsilon
  - Generally better tailored to an online data access model
  - Repeated queries increases disclosure risk. Typically track with a privacy budget.

“The Myth of PII…”

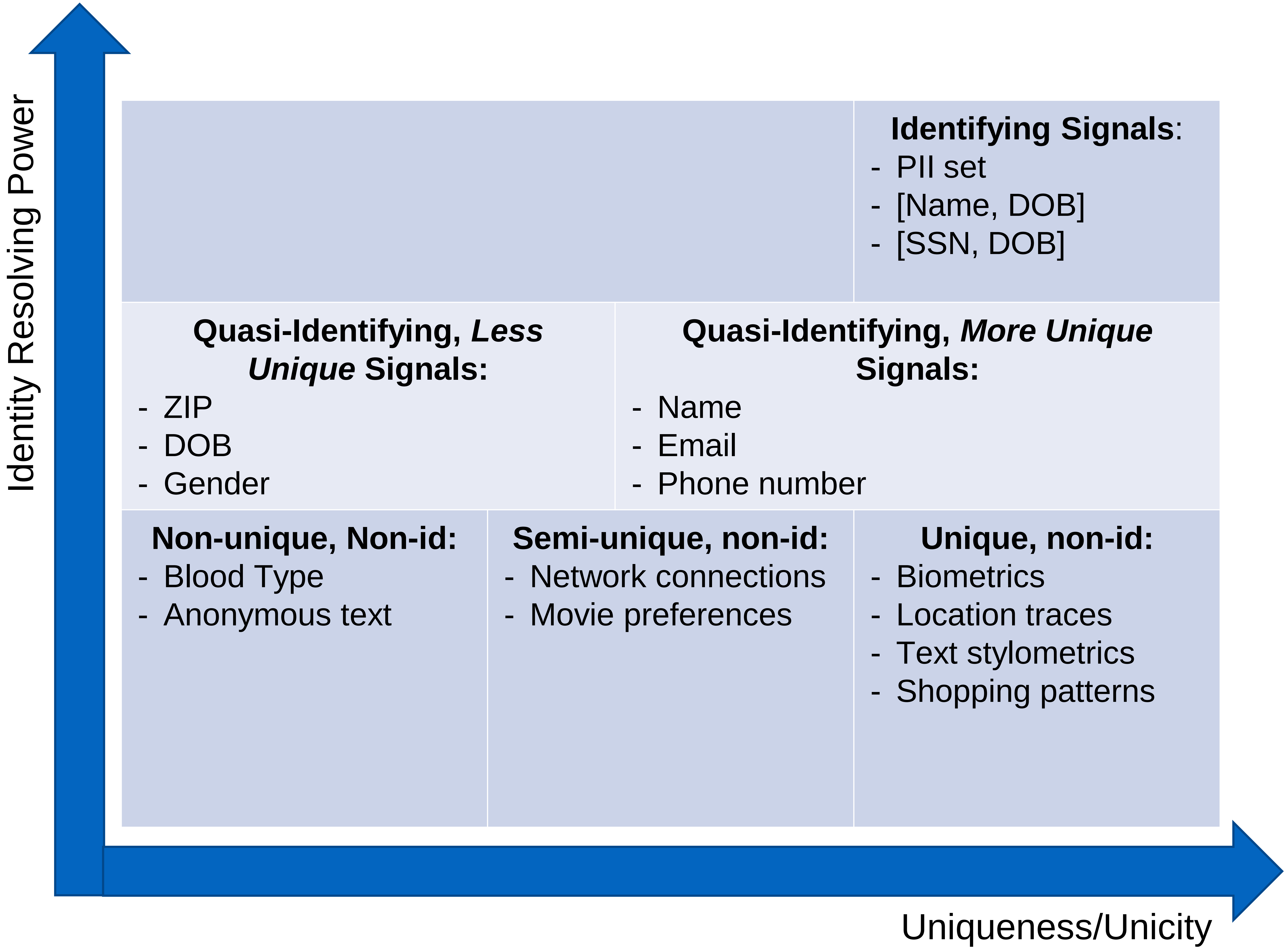
Privacy Regulations often founded on the assumption that specific descriptive signals are especially revelatory of subjects’ identities e.g. HIPAA’s PII designation.

With large-scale secondary data & powerful compute,
this turns out to be untrue

Any information that distinguishes one person from another can be used for re-identifying data.

Narayanan, Arvind, and Vitaly Shmatikov. "Myths and fallacies of personally identifiable information." *Communications of the ACM* 53, no. 6 (2010): 24-26.

Caveat: Assumes existence of & access to a “population register”
Distinguish between:
Uniqueness & **Identifiability**

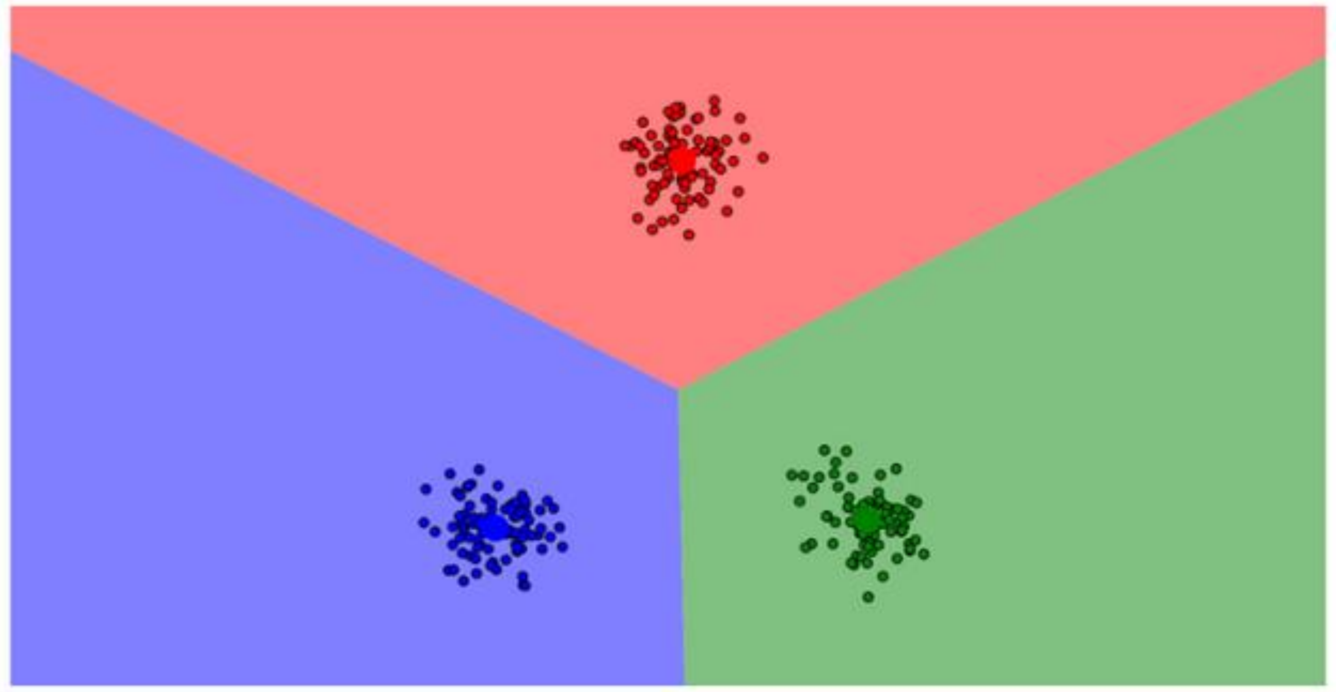


Notional Breakdown of Uniqueness vs. Identifiability

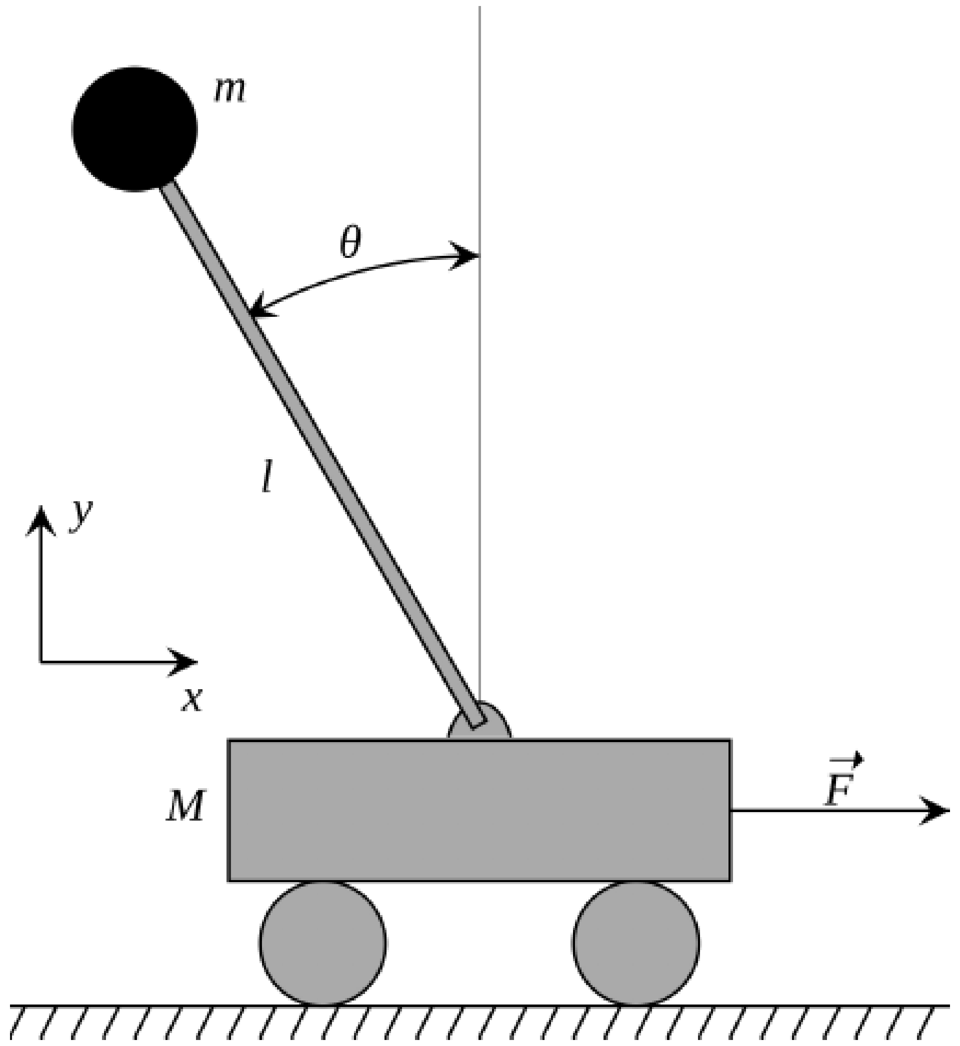
Machine Learning Modes

Machine Learning

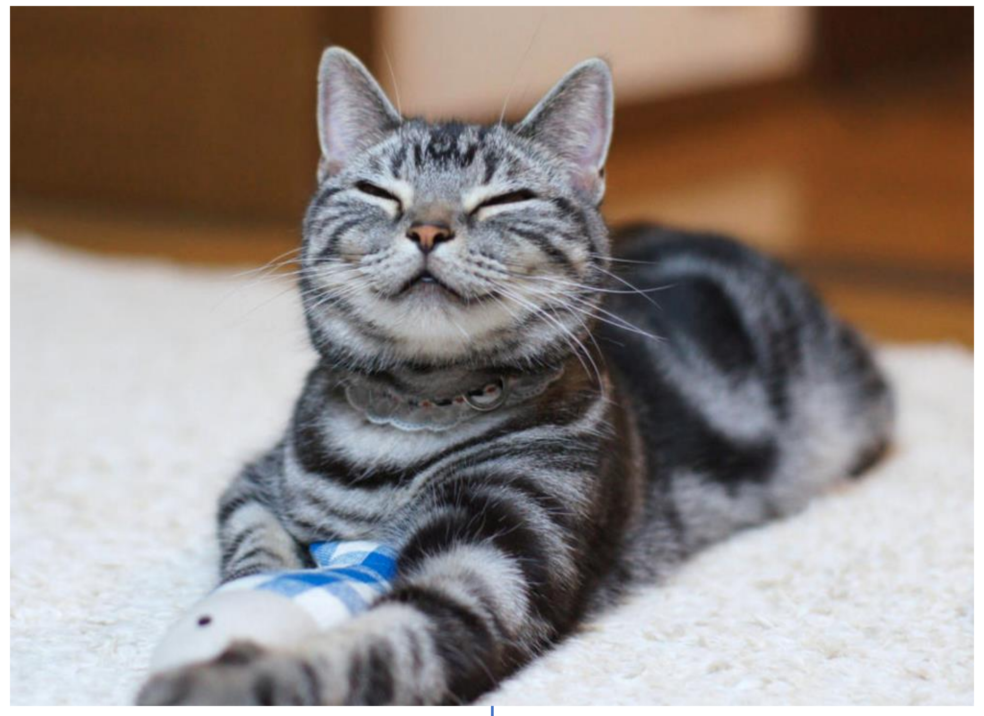
- Unsupervised Learning
- Reinforcement Learning
- Supervised Learning



- Clustering
- Density Estimation



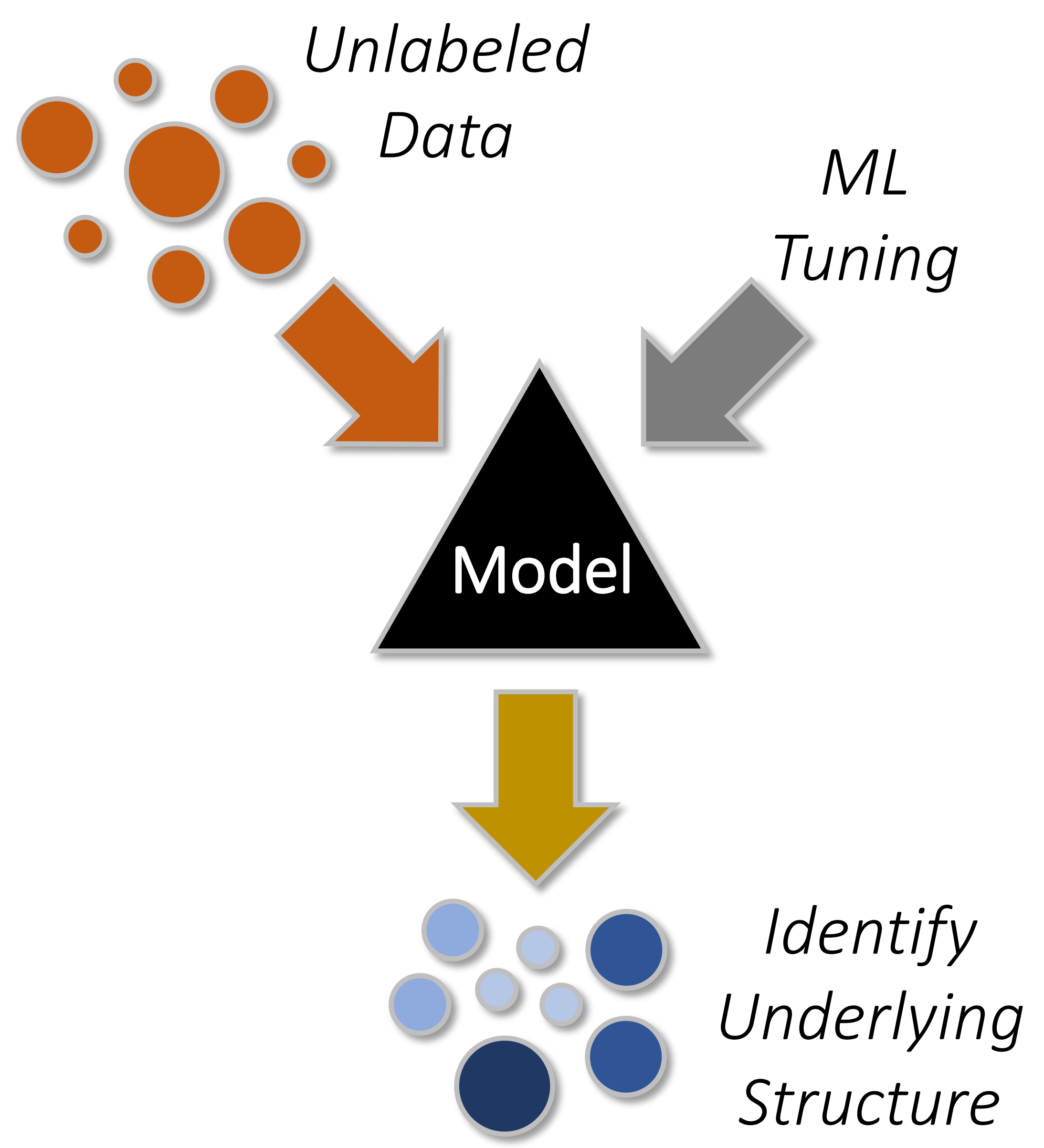
Control



Y=$f(x)$

Prediction

Varying Levels of Supervision…



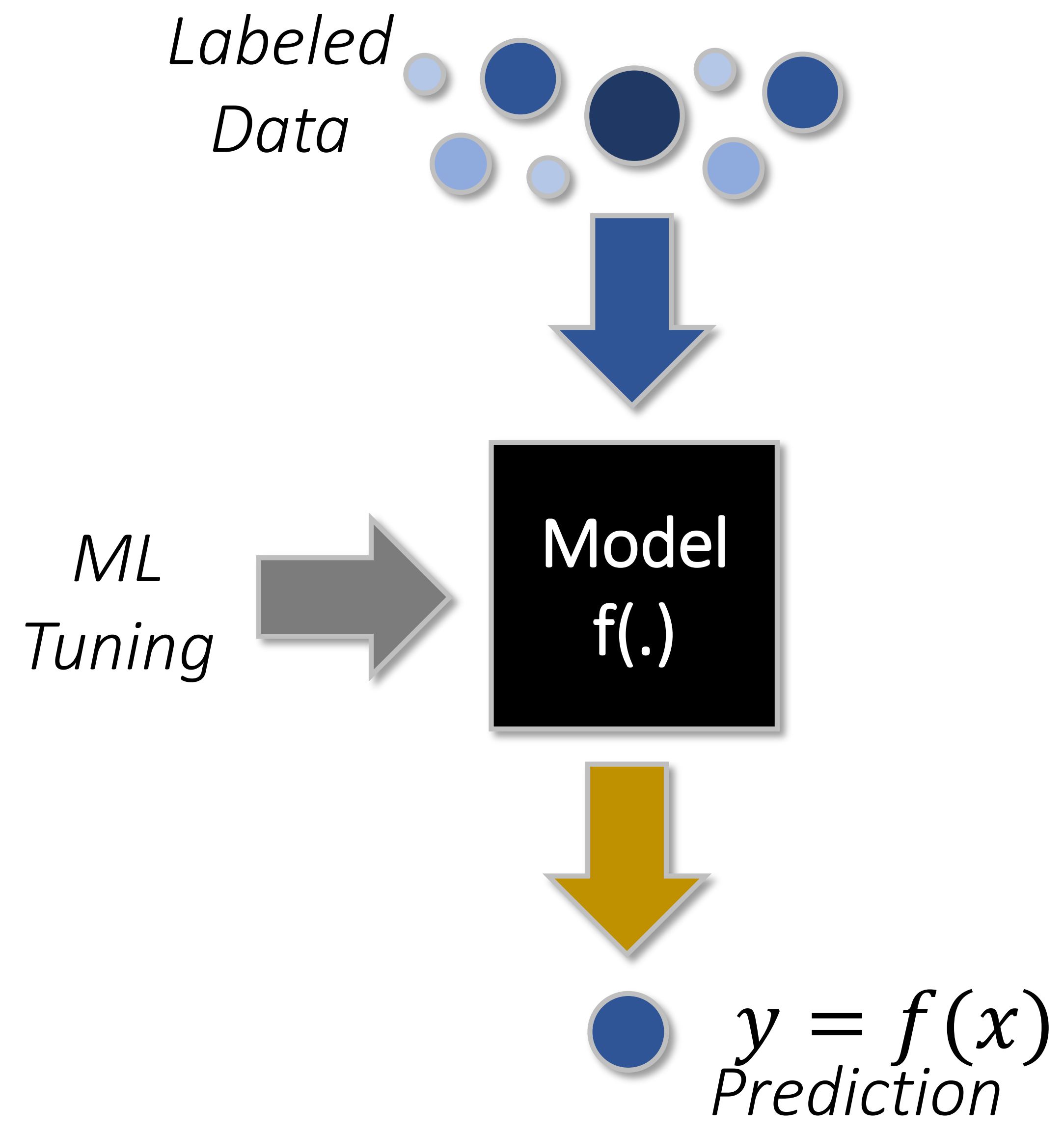
No supervisory signals

Input (s): samples (x)
Decision (d): cluster assign (z)
Evaluation (r): Goodness-of-Fit (G)

$$\mathrm{G}(\theta) = \ln p(x;\theta)$$
Or sum-of-square distances

- Unsupervised Learning:
  - Maximize the log-likelihood, G(θ) (for density estimation tasks)
- Supervised Learning:
  - Maximize predictive accuracy Acc(θ)
- All use some version of *empirical risk minimization* to update parameters

$$\theta_{t+1} = \theta_t - \eta_t \nabla_\theta J(\theta)$$



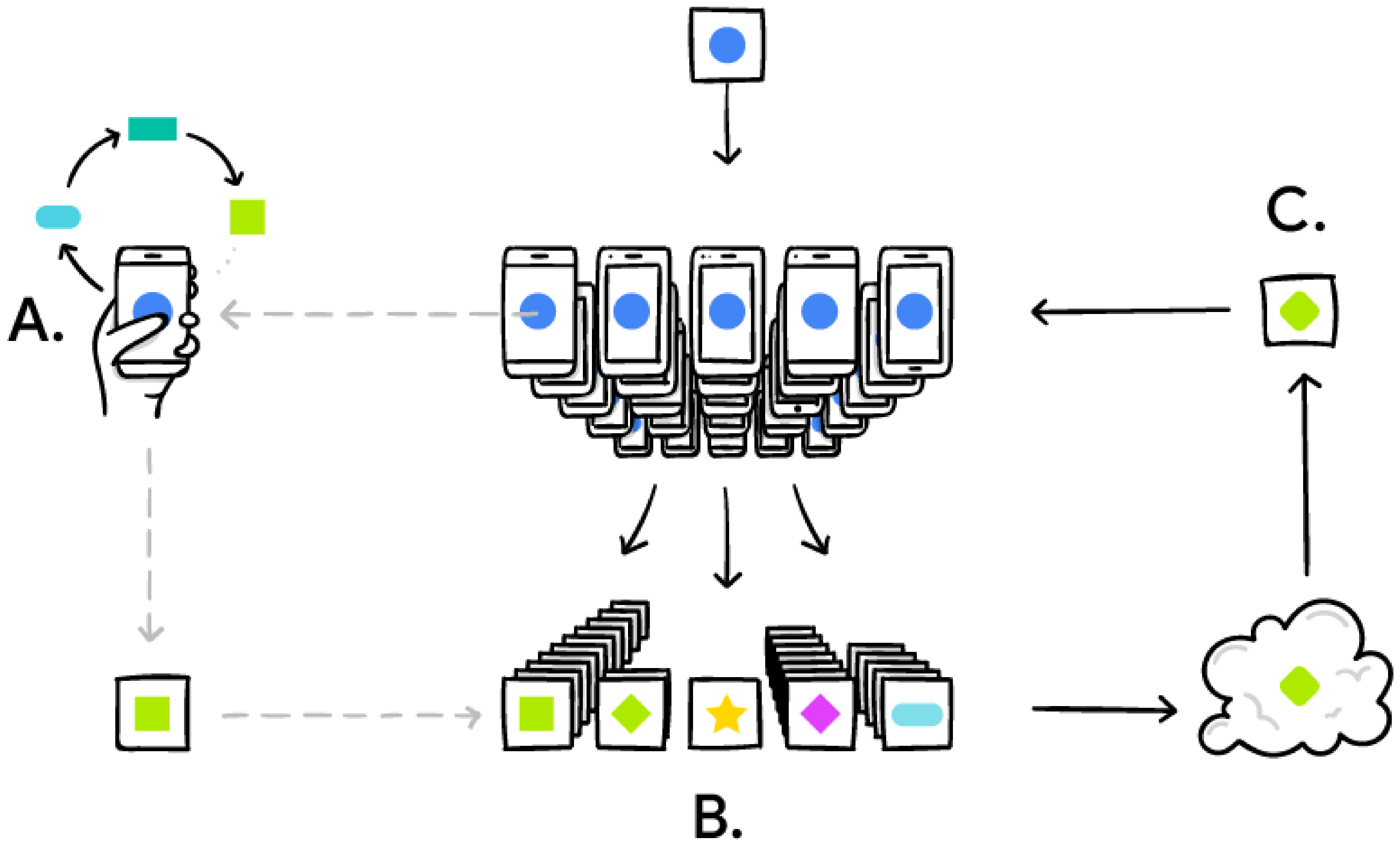
Full supervisory signals

Input (s): samples (x)
Decision (d): predictions (y)
Evaluation (r): accuracy (Acc)

$$\mathrm{Acc}(\theta) = J(y, \hat{y}(x,\theta))$$
e.g. J = MSE or Cross-Entropy

New Tools in the Arsenal: Federated Learning

- Data is useful for more than just access.
- AI/ML models require data for tuning
- Curating such data for modeling elevates privacy risk
- Federated Learning provides a means for training advanced models without access to subject data



Source: https://ai.googleblog.com/2017/04/federated-learning-collaborative.html

New Threats: Model Inversion

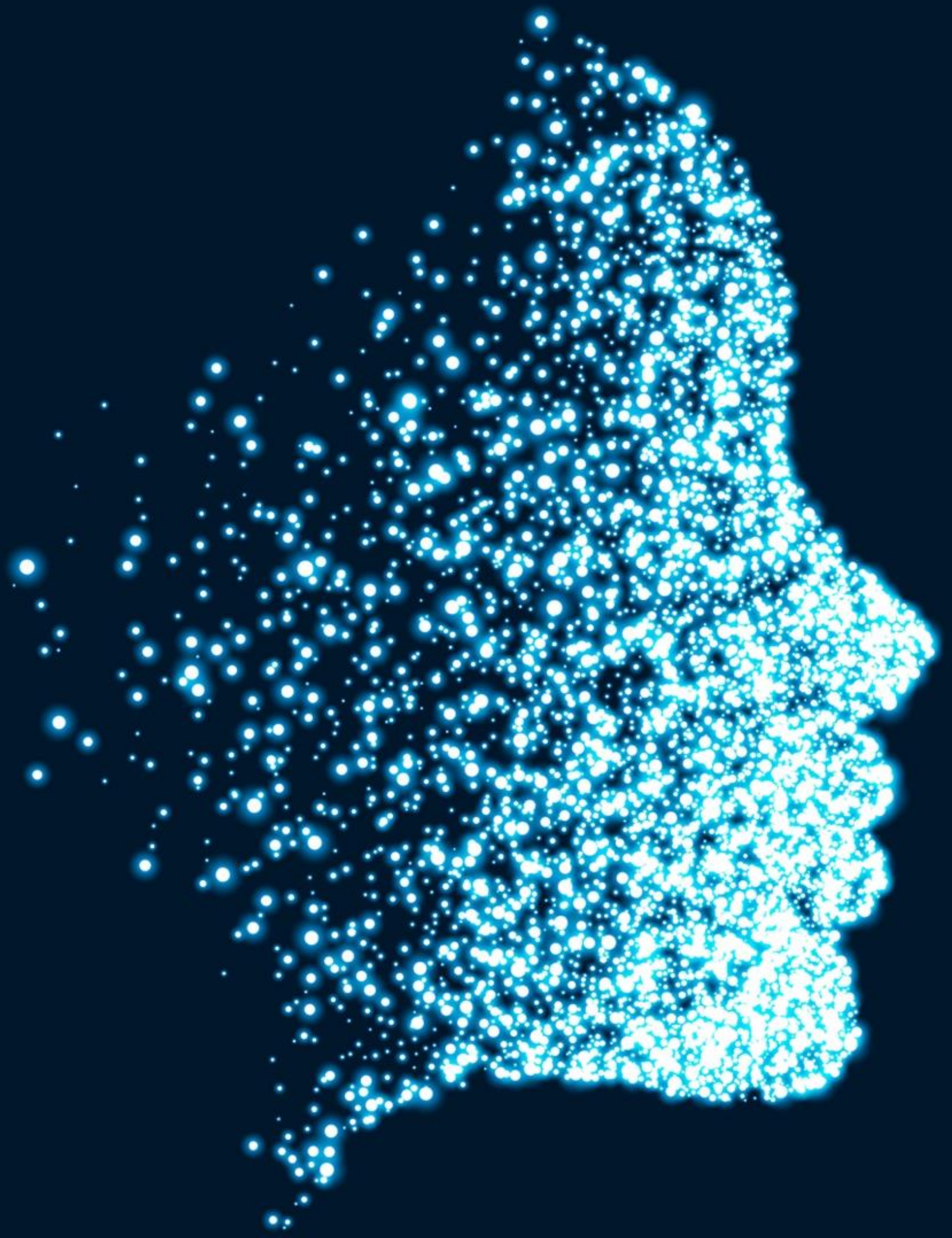
- Advanced models necessarily contain information about the data they are trained on
- Recent work demonstrates inversion attacks on such models
- i.e. design algorithms to reconstruct training data using just access to the model

Figure 1: An image recovered using a new model inversion attack (left) and a training set image of the victim (right). The attacker is given only the person's name and access to a facial recognition system that returns a class confidence score.

Source: Fredrikson, Jha, Ristenpart 2016

Relevant Old & Emerging Threats: Cybersecurity

- Ransomware/Wannacry
- Advanced Persistent Threat
- Internet of Things/Bodies
- Is Privacy Fair?



RAND

StarLineArts via Getty Images