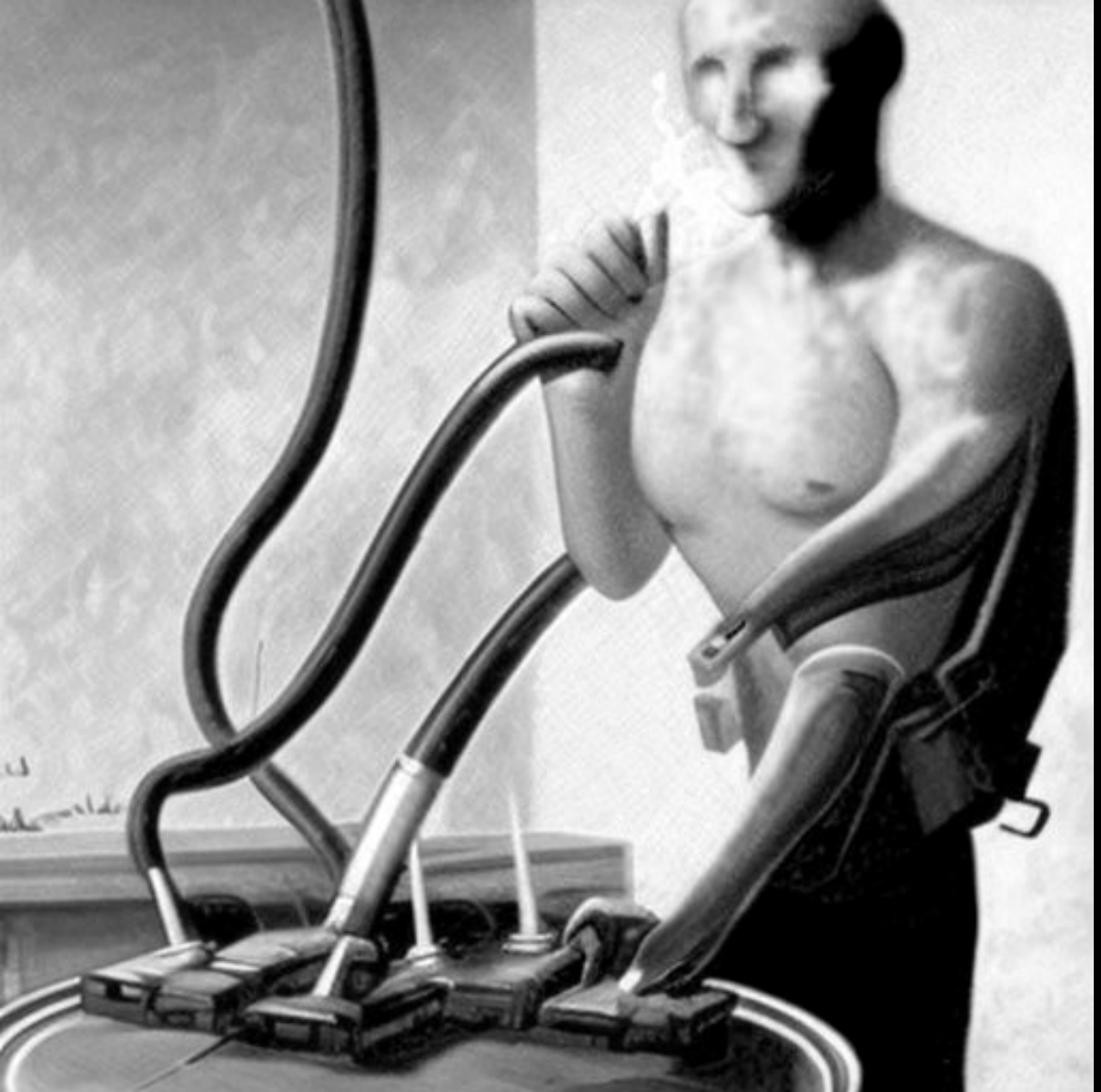
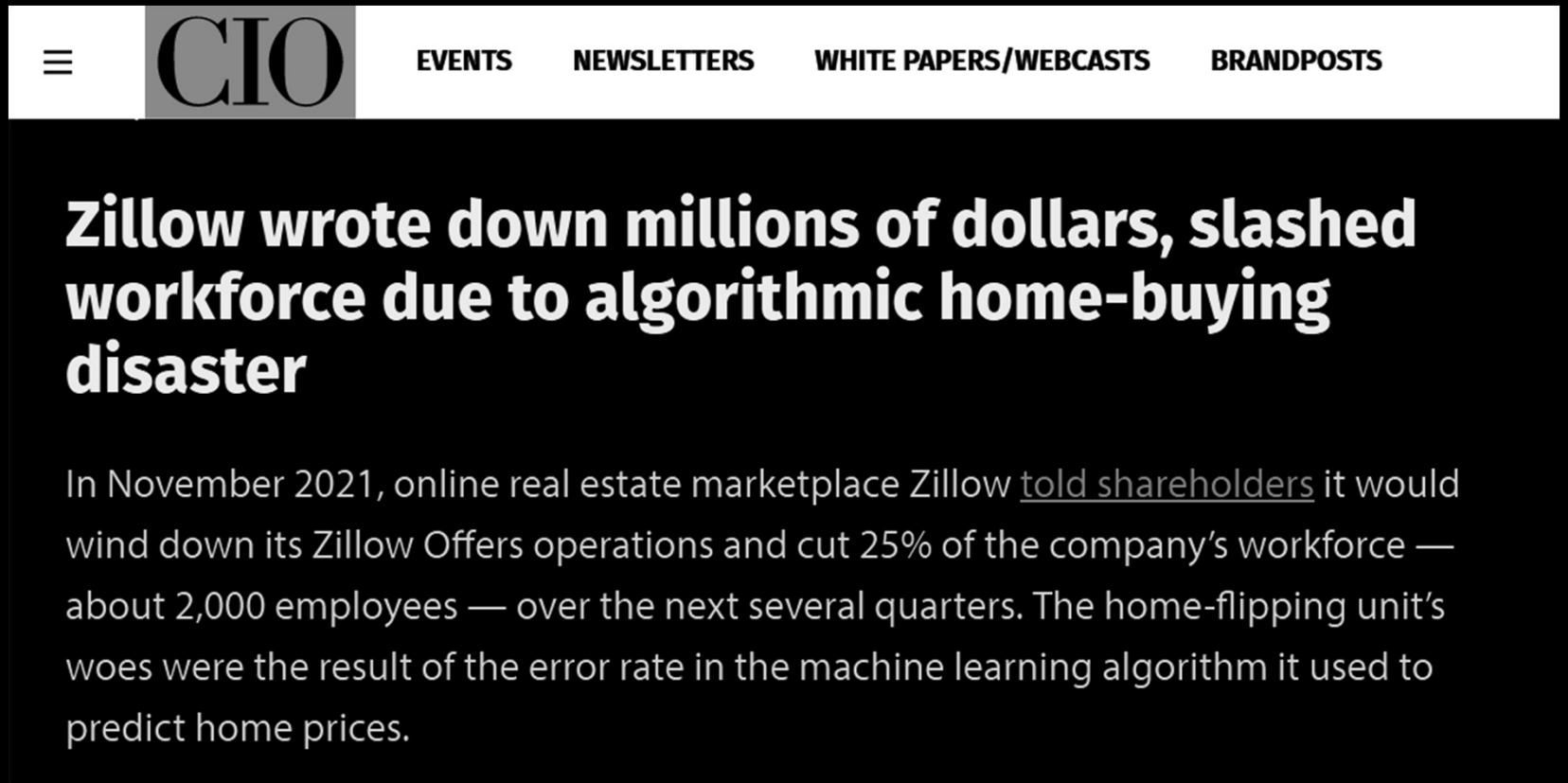
Talk by Dr. Andrzej NOWAK – November 2022
CERN Academic Training Programme



Confronting more brainy data vacuums

AI gone wrong



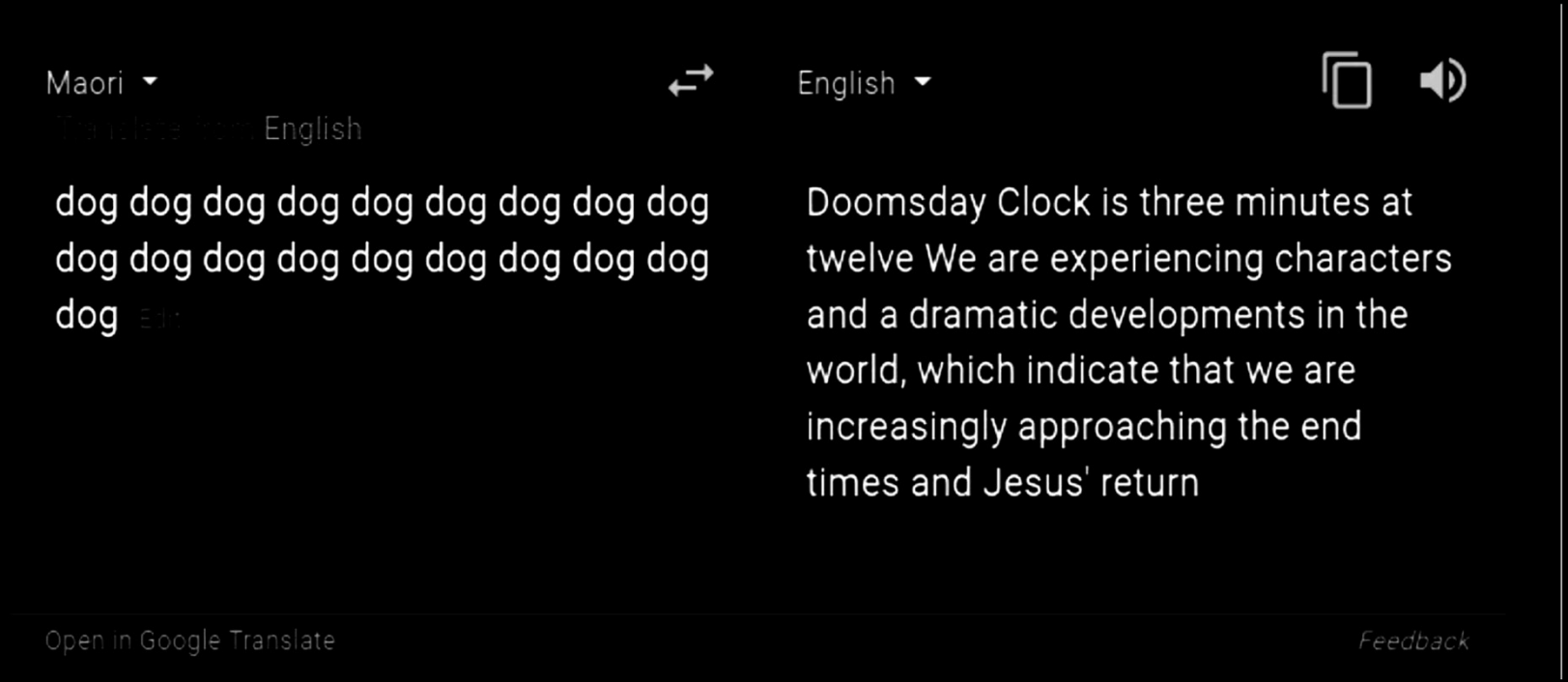
CIO

EVENTS NEWSLETTERS WHITE PAPERS/WEBCASTS BRANDPOSTS

Zillow wrote down millions of dollars, slashed workforce due to algorithmic home-buying disaster

In November 2021, online real estate marketplace Zillow told shareholders it would wind down its Zillow Offers operations and cut 25% of the company's workforce — about 2,000 employees — over the next several quarters. The home-flipping unit's woes were the result of the error rate in the machine learning algorithm it used to predict home prices.

ML models memorize training data



Inspired by Motherboard / PTP Harvard

Deepfakes, morphs

Ctrl Shift Face / youtube

AI in policing – beyond faces

Face recognition (e.g., US capitol attack)

Crowdfunded crime reporting

Predictive policing

Risk-based arrest warrant prioritization

DNA forensic patterns

Gunshot detection

CCo

AI and war

Aegis (1983) – human-in-the-loop or human-on-the-loop
Phalanx CIWS (1978)
Tomahawk (anti-ship, 1990s)
Loitering anti-radar missiles (e.g., HARM)
X47-B (2013)
Missile launch detection (Cold War era)

Planning and logistics tools (DART, 1990; JADE, 1999)
SAPE (Nuclear war planning, 1991)

TECH / ROBOT

They’re putting guns on robot dogs now

It's not clear if this gun-equipped quadrupedal robot is for sale, but it's only a matter of time. Image: Sword International

/ It was only a matter of time

By JAMES VINCENT

Oct 14, 2021, 4:47 PM GMT+2 | 0 Comments / 0 New

How do we control this?

Self-regulation

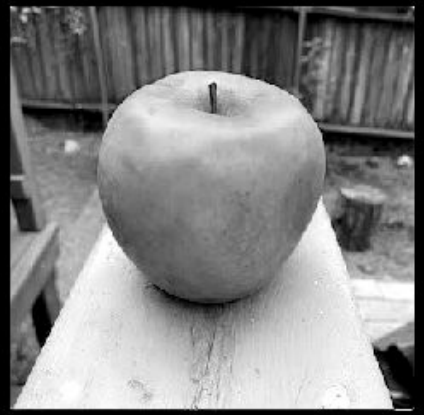
Regulation (a.k.a. “The Law”)

or… more technology? 🤔

Defense: technical, legal, (societal?)

"[...] *in a very real way, we've rushed ahead, paying little attention to vulnerabilities inherent in ML platforms – particularly in terms of altering, corrupting or deceiving these systems.*"

Hava Siegelmann, DARPA



Granny Smith	85.6%
iPod	0.4%
library	0.0%
pizza	0.0%
toaster	0.0%
dough	0.1%

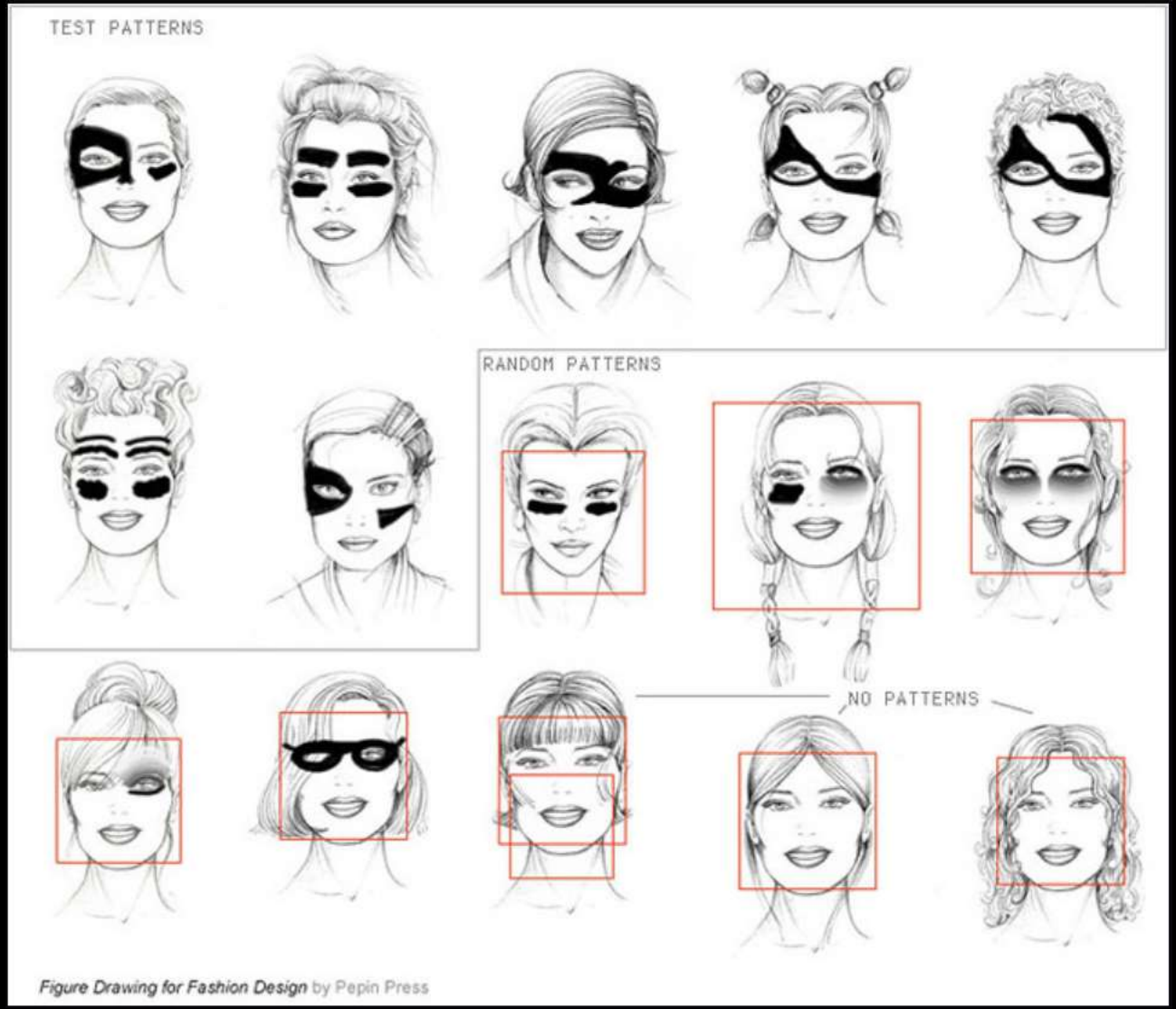
Granny Smith	0.1%
iPod	99.7%
library	0.0%
pizza	0.0%
toaster	0.0%
dough	0.0%

classified as turtle

classified as rifle

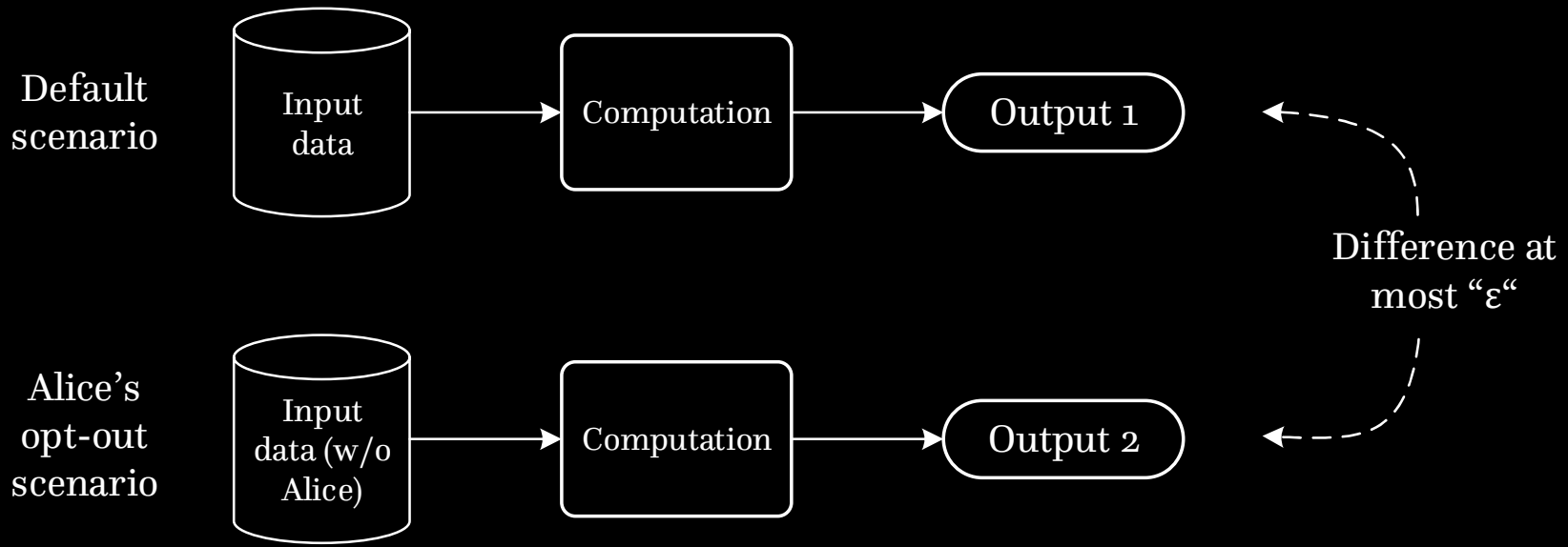
classified as other

Synthesizing Robust Adversarial Examples, MIT



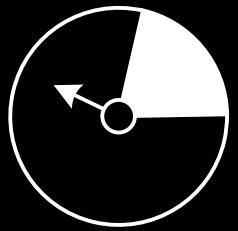
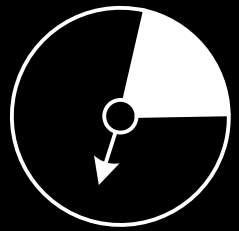
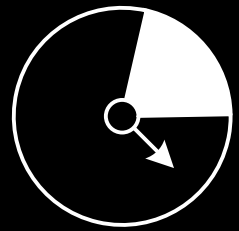
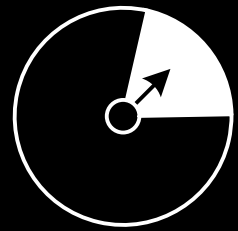
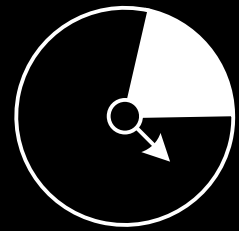
Dazzle Club on Instagram

Differential privacy



Differential privacy

Do you like Bosons?

Real answer	Yup	Yup	Boring	Yup	Yup
					
Noisy answer	Yup	Yup	Boring	Boring	Yup

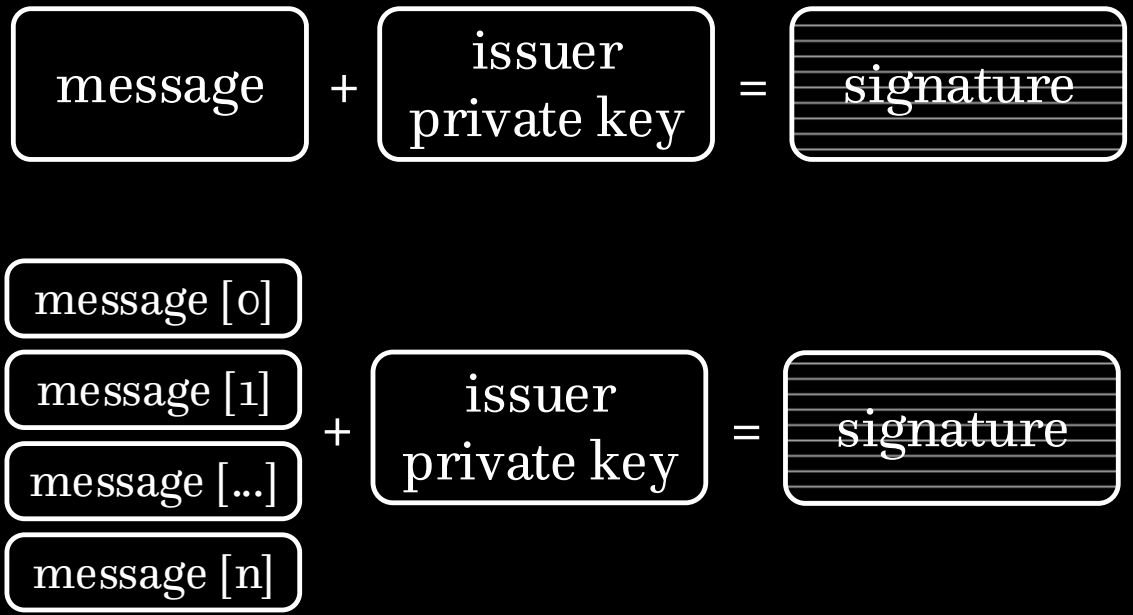
Data Trust

An entity that acts as a trusted data steward on behalf of the parties.

- Usage, sharing rules made by users
- Trust control (data trusts) or user control (data hubs)
- Voting?

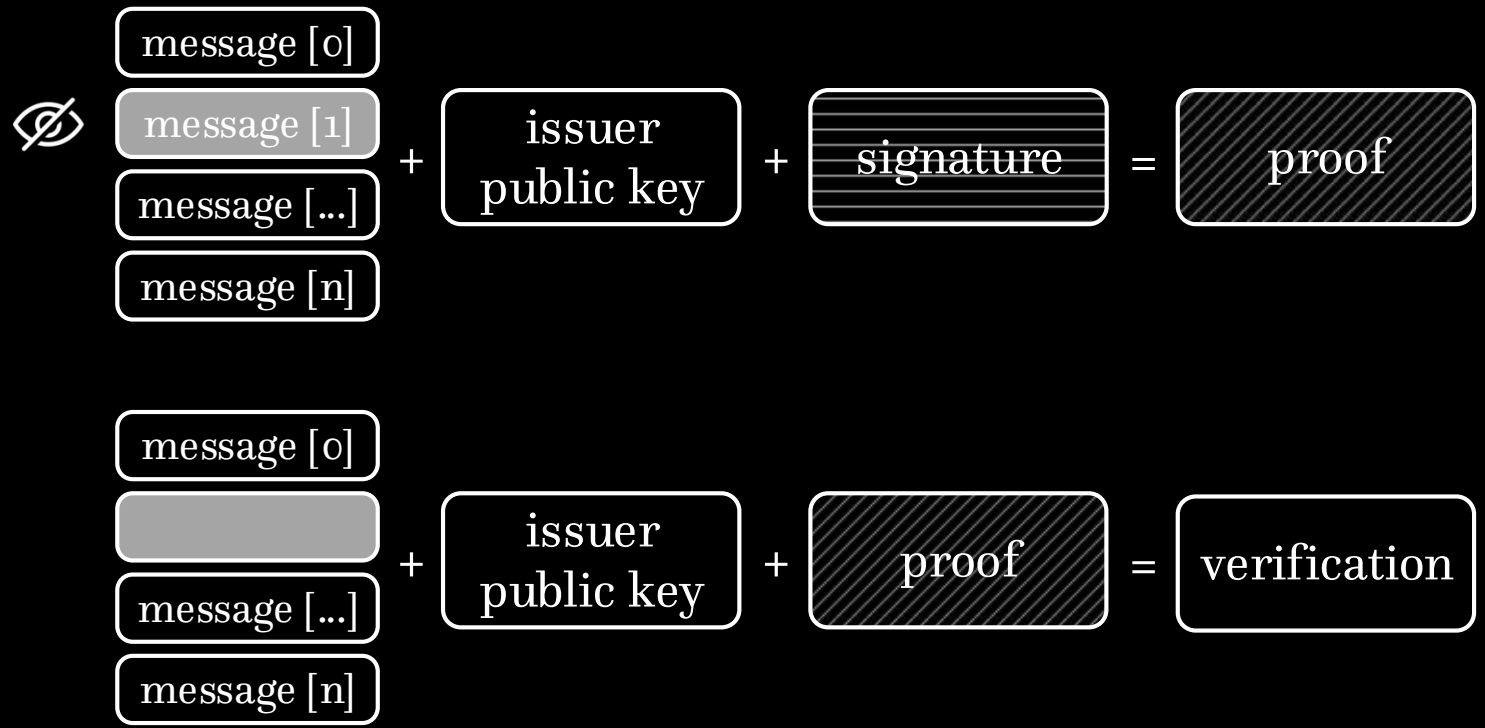
Examples: UK biobank, OpenCorporates

Selective disclosure



Modeled after Mattr

Selective disclosure



Modeled after Mattr

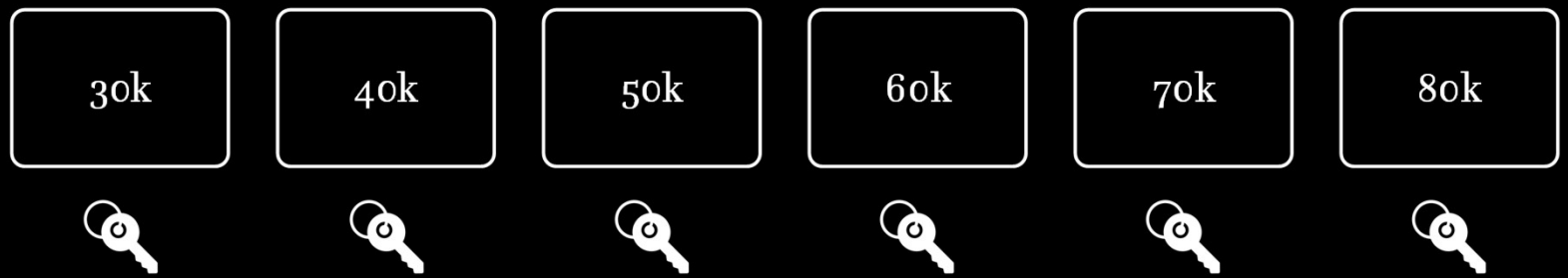
Zero-knowledge proofs (ZKP)

complete, sound, zero-knowledge

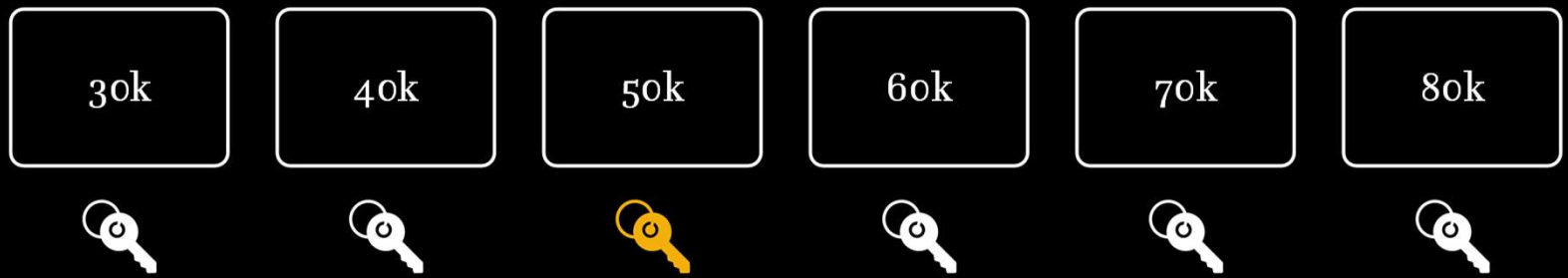
ZKP applications

Knowledge proofs	Does the prover really know the private key?
Set membership	Is the person a paying subscriber? Is the presidential candidate in a DNA database of offenders?
Range proofs	Is the person over 18? Is the person aged 35-45?

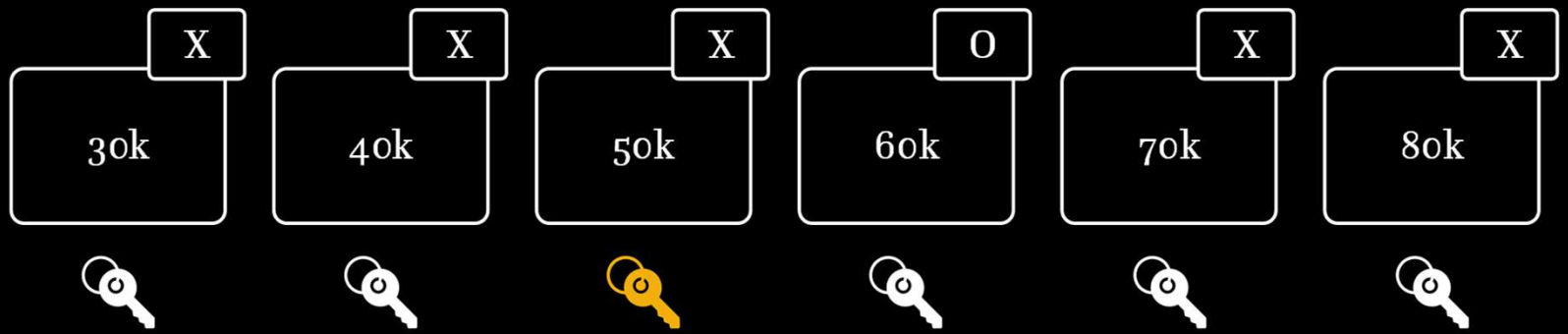
ZKP example (equality)



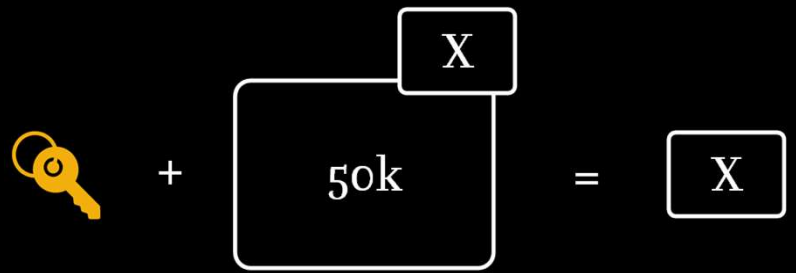
ZKP example (equality)



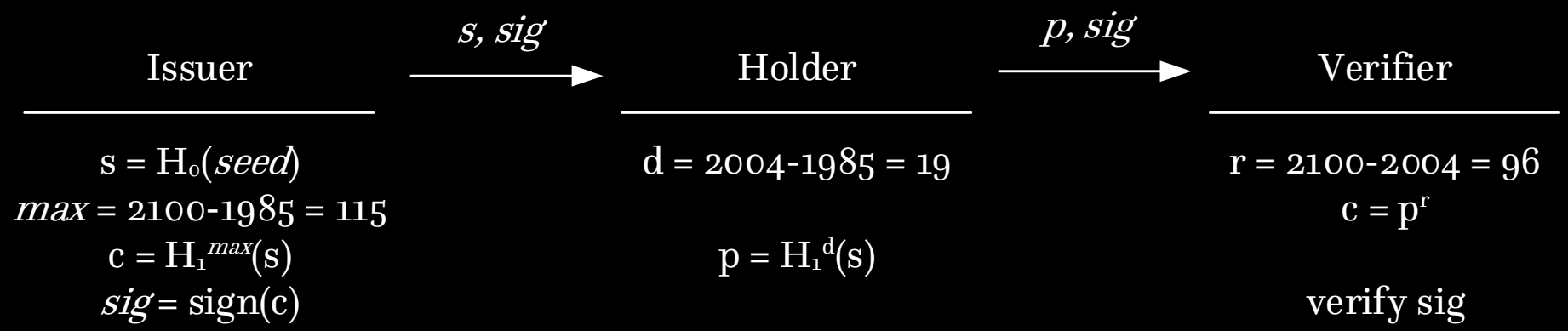
ZKP example (equality)



ZKP example (equality)



ZKP example (inequality)



Why “2004”? Because it’s 18 years go. Adapted from zkproof

Multi-Party Computation (MPC)

Allow mutually distrusting parties to cooperatively compute over their private data. (theoretically – any function!)

→ split up data into pieces

→ send to participants for computation, compute

→ reduce partial results

No trusted parties in the middle

No need to reveal even a single bit of private data to other

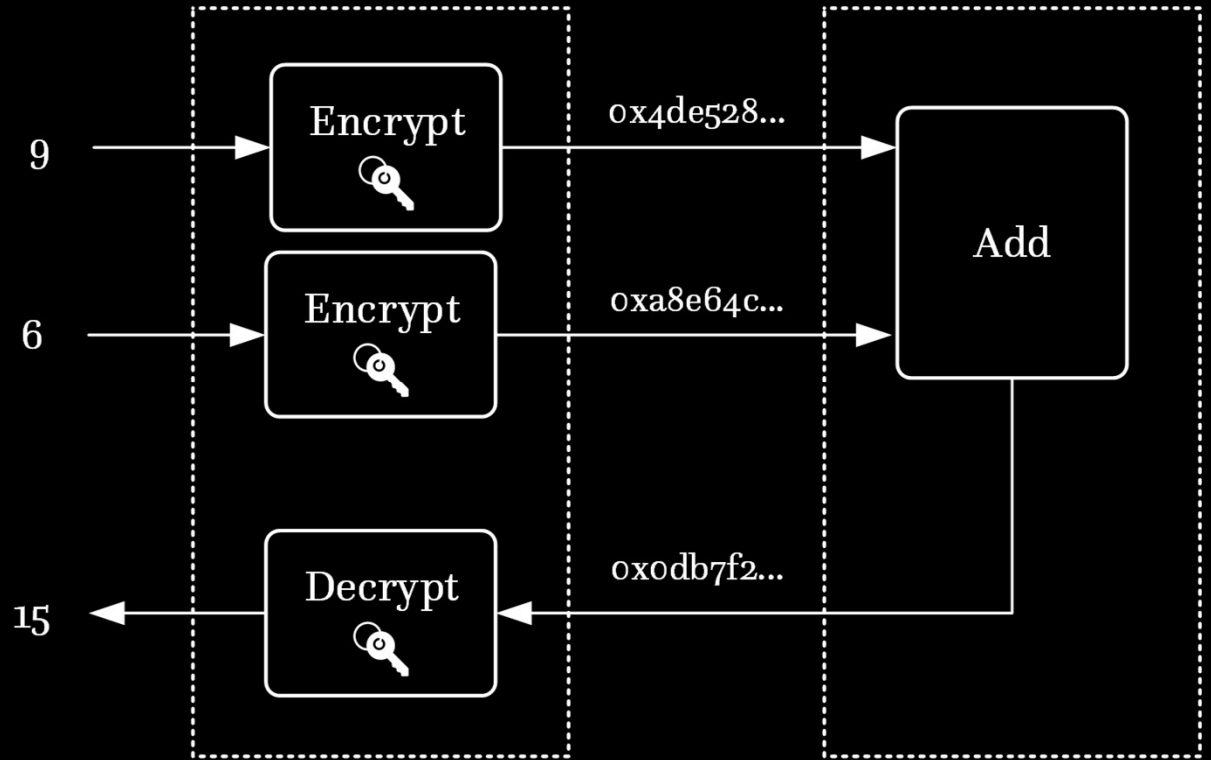
Negligible compute cost, high communication cost

Multi-Party Computation example

			P1	P2	P3	P4	
P1	a	←	a_1	a_2	a_3	a_4	
P2	b	←	b_1	b_2	b_3	b_4	
P3	c	←	c_1	c_2	c_3	c_4	
P4	d	←	d_1	d_2	d_3	d_4	
			s_1	s_2	s_3	s_4	← public

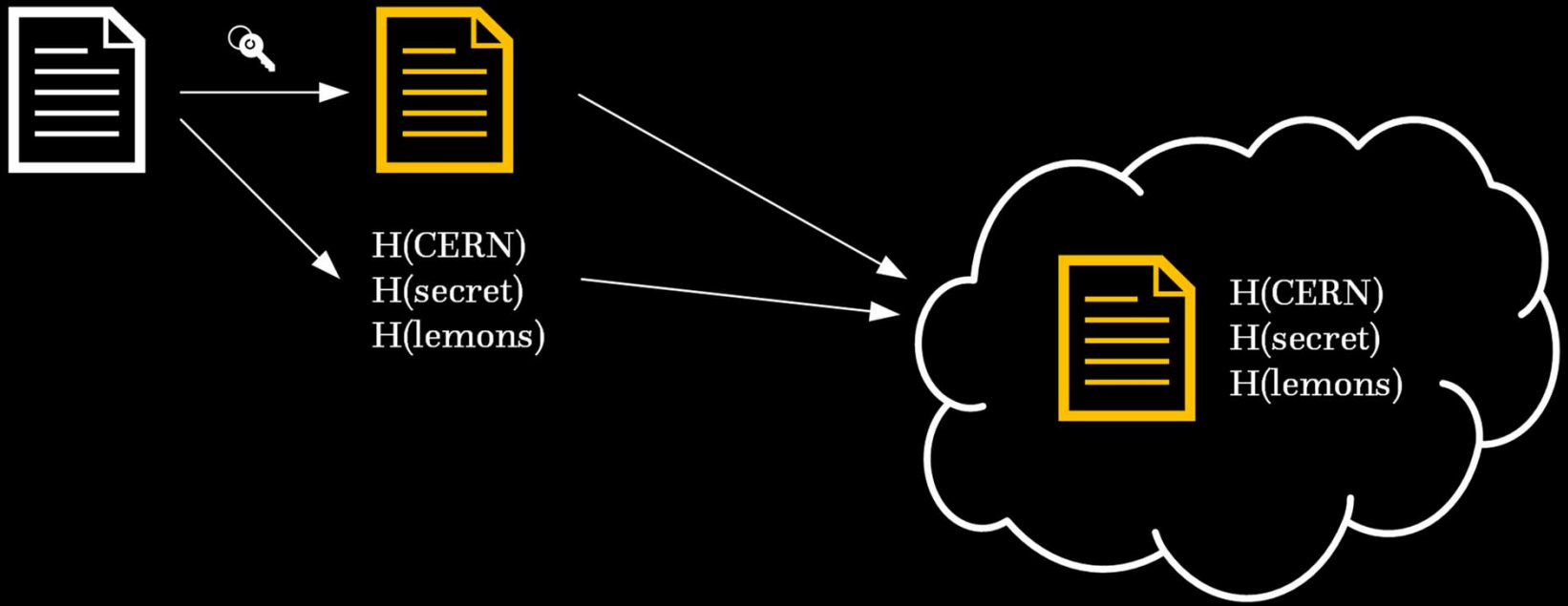
Adapted from Ivan Damgard / Partisia

Homomorphic encryption / FHE

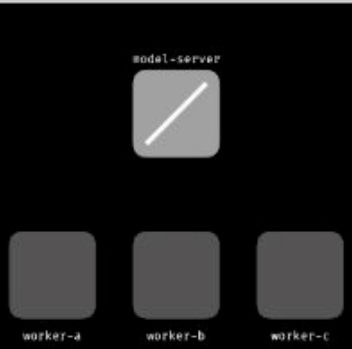
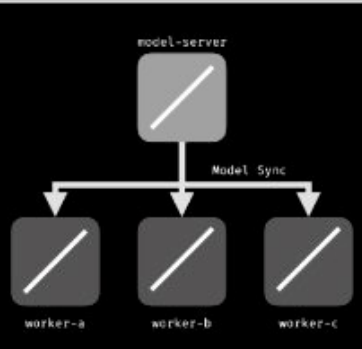
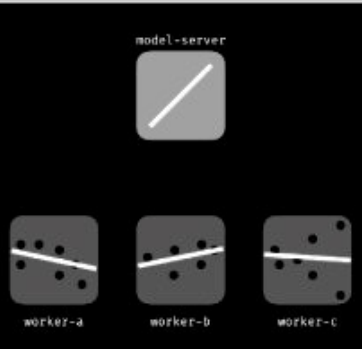
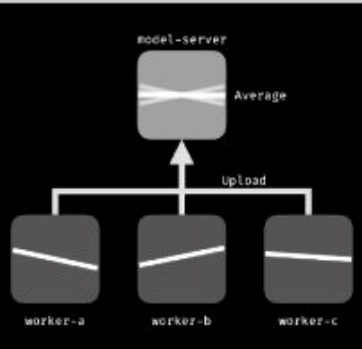


Adapted from Microsoft

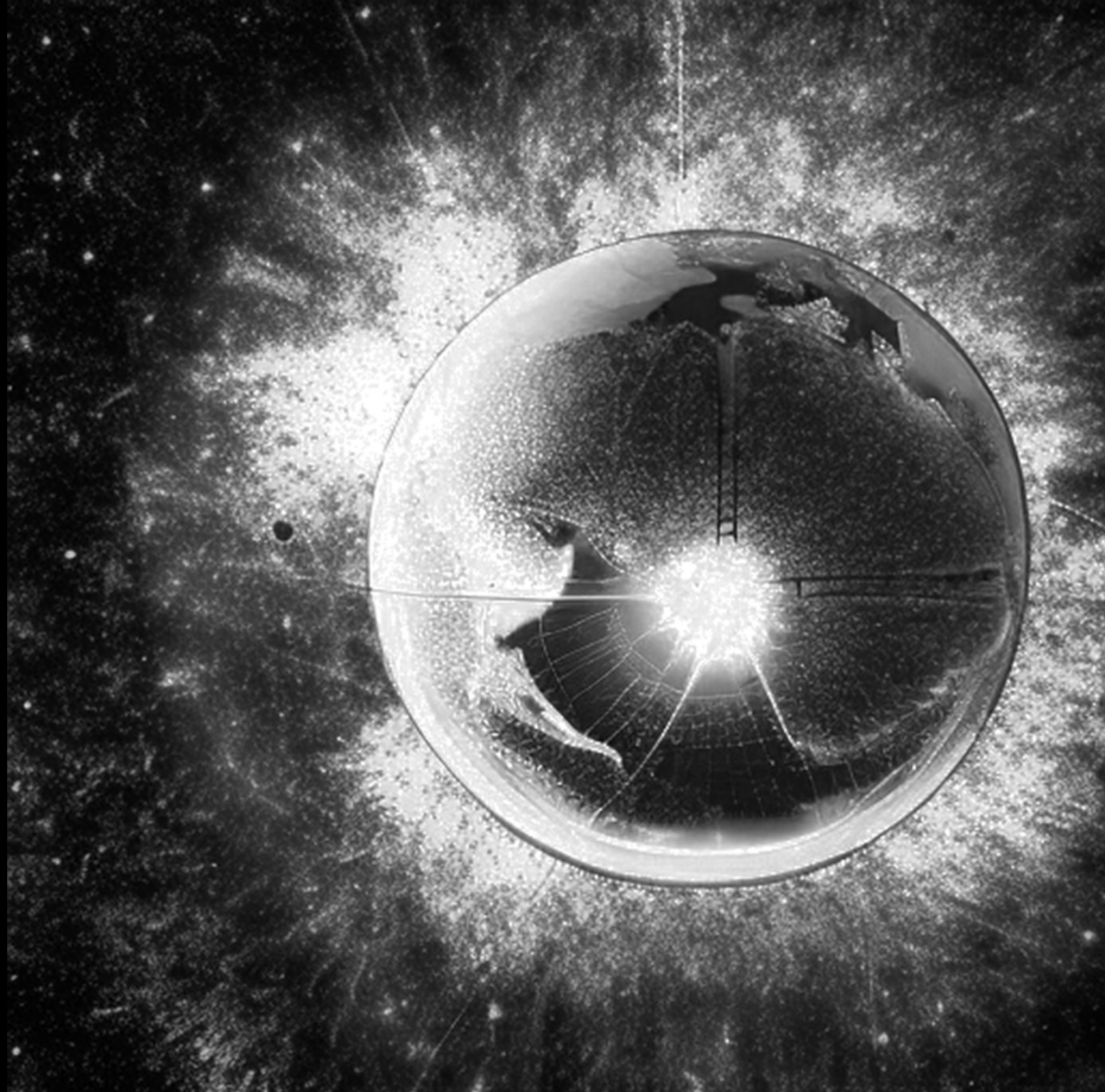
Searchable encryption



Federated Machine Learning

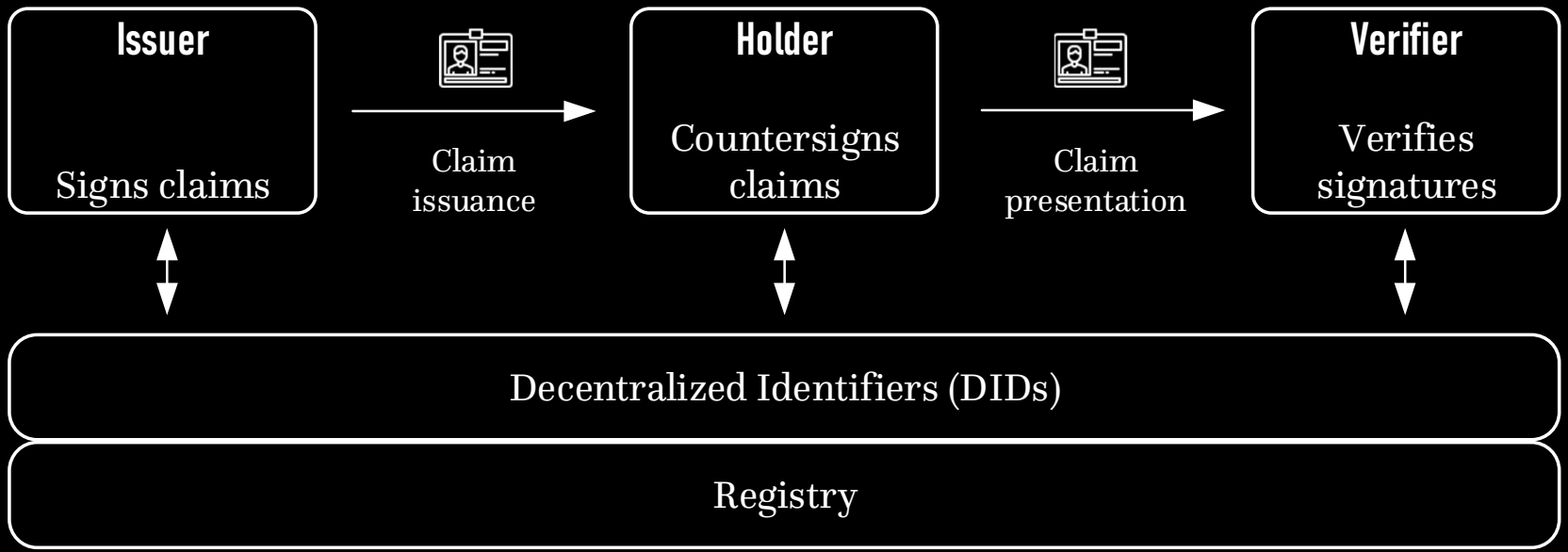
Step 1	Step 2	Step 3	Step 4
			
Central server chooses a statistical model to be trained	Central server transmits the initial model to several nodes	Nodes train the model locally with their own data	Central server pools model results and generate one global mode without accessing any data

Back to… Earth?



Government interaction

Decentralized identity



icons: icon8

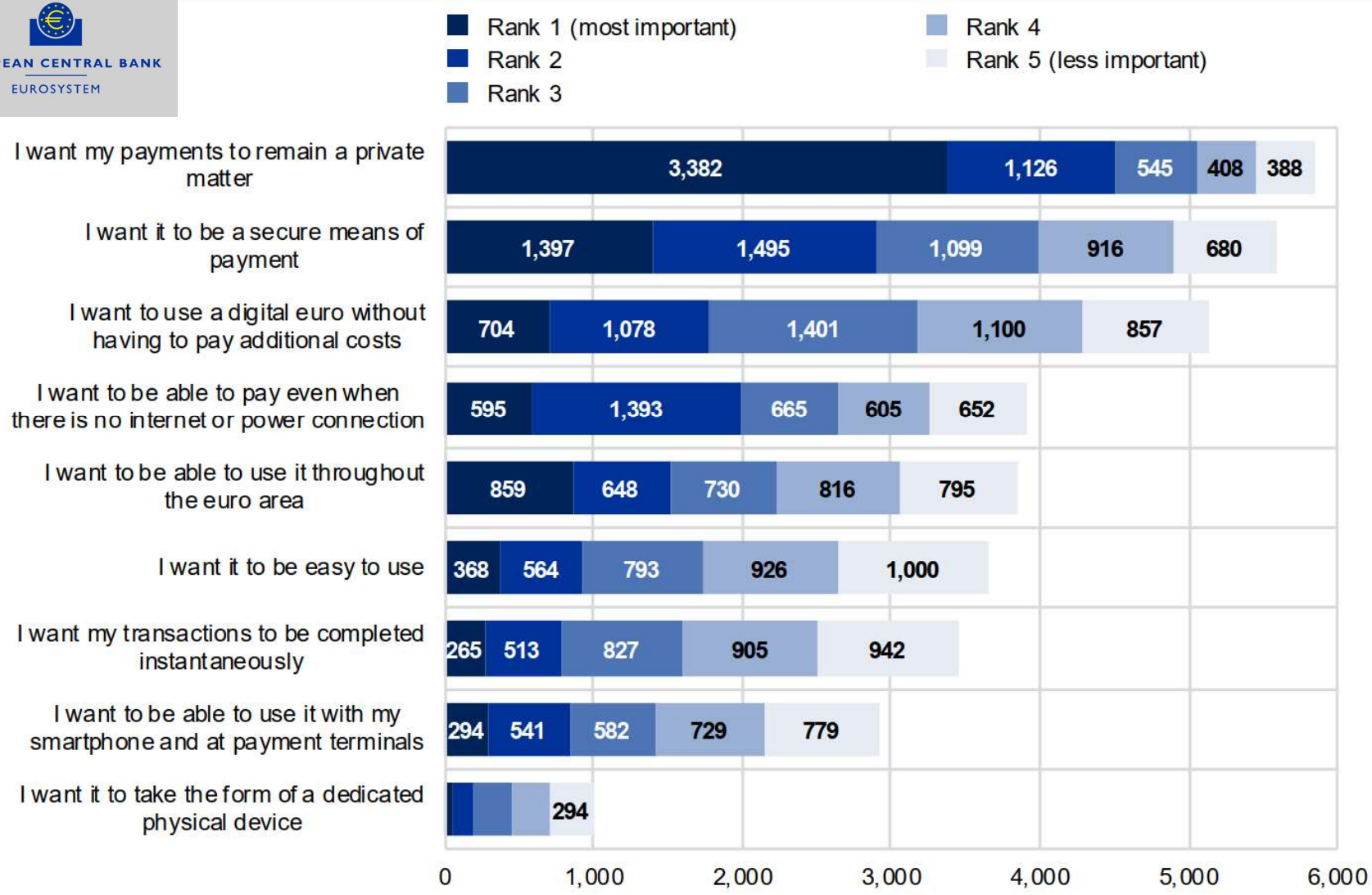
Central Bank Digital Currency

“a digital form of central bank money that is different from balances in traditional reserve or settlement accounts”

Bank for International Settlements

Preference for some digital euro features based on top five rankings

EUROPEAN CENTRAL BANK
EUROSYSTEM



Automated decision making

decision object?
model, rationale?
data?
data quality?

AU$721 million in robo-debts to be returned as Australian government admits error

Services Australia has identified 470,000 debts raised wholly or partially using income averaging of ATO data.

Written by **Asha Barbaschow,** Contributor on May 28, 2020

Automated decision making – recourse

"The data subject shall have the right not to be subject to a decision based solely on automated processing, including profiling, which produces legal effects concerning him or her or similarly significantly affects him or her."

GDPR, Art. 22

Automated profiling

'profiling' means any form of automated processing of personal data consisting of the use of personal data to evaluate certain personal aspects relating to a natural person, in particular to analyse or predict aspects concerning that natural person's performance at work, economic situation, health, personal preferences, interests, reliability, behaviour, location or movements;

GDPR, Art. 4

Automated profiling – applications

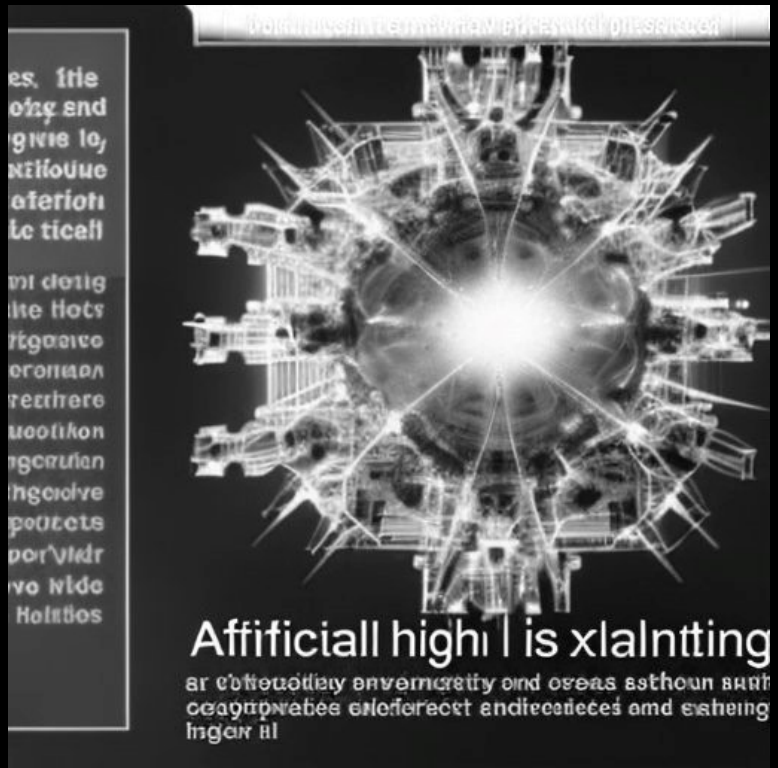
advertising profile
level of influence
recruitment
dominant emotional states
weaknesses
probability of disease
re-offense profile

Bias (humans make robots, ultimately)

human
data sampling
long-tail
intentional
hidden

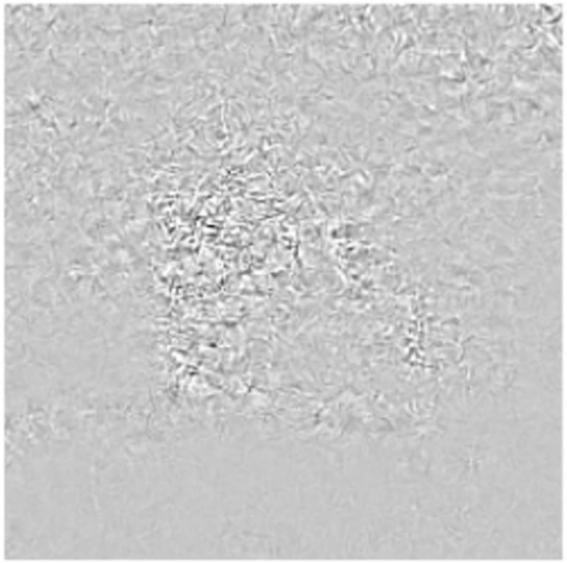
CCo

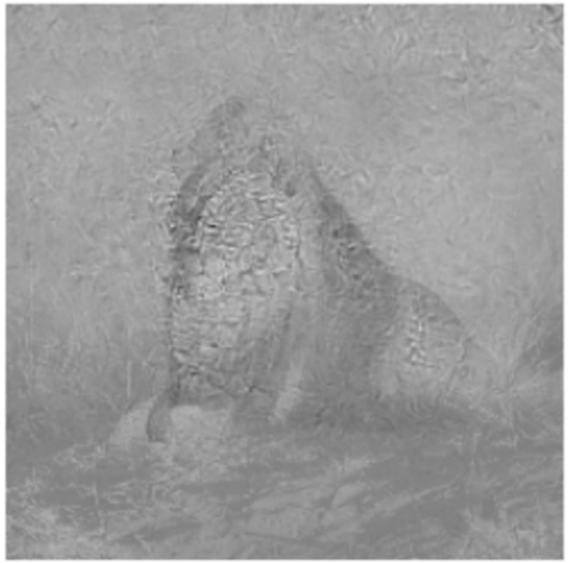
Sidenote: can we explain AI?



interpret → why this decision?

explain → how did we get to this result?





machine learning

+

your data

=

your data lives on in derived formats: statistics, ML models...

...

+

right to be forgotten

=

Machine Unlearning

yup, it exists

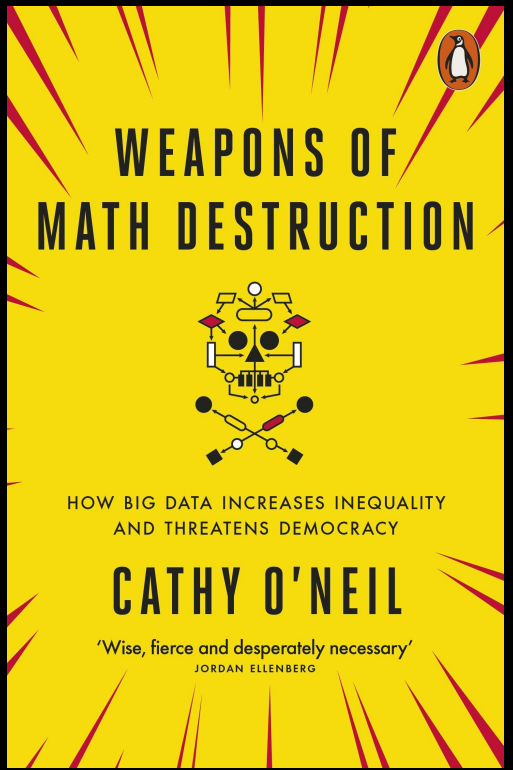
AI & I

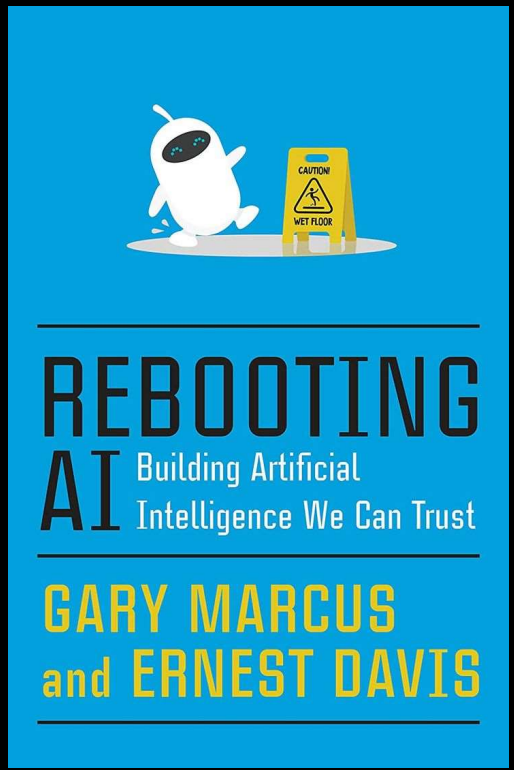
(1) data and trust minimization

(2) open, transparent solutions

(3) share your knowledge

WOULD YOU LIKE TO KNOW MORE?





Starship Troopers

“*Invisibility is a superpower*”

- Banksy

“Battling robots for our data, privacy and humanity”

A privacy talk by Dr. Andrzej NOWAK, November 2022, at CERN

<an@tik.services> Special thanks to:

Maria Dimou (CERN), Stephanie Hare, Nathalie Rauschmayr (Google), Liviu Valsan (CERN)