

# Data protection in the age of AI

## GDPR and Copyrights

ITWG

Virtual coffee break with a colleague  
January 27<sup>th</sup>, 2026

Alenka Blas and Ruti Rous



REPUBLIC OF SLOVENIA  
COURT OF AUDIT

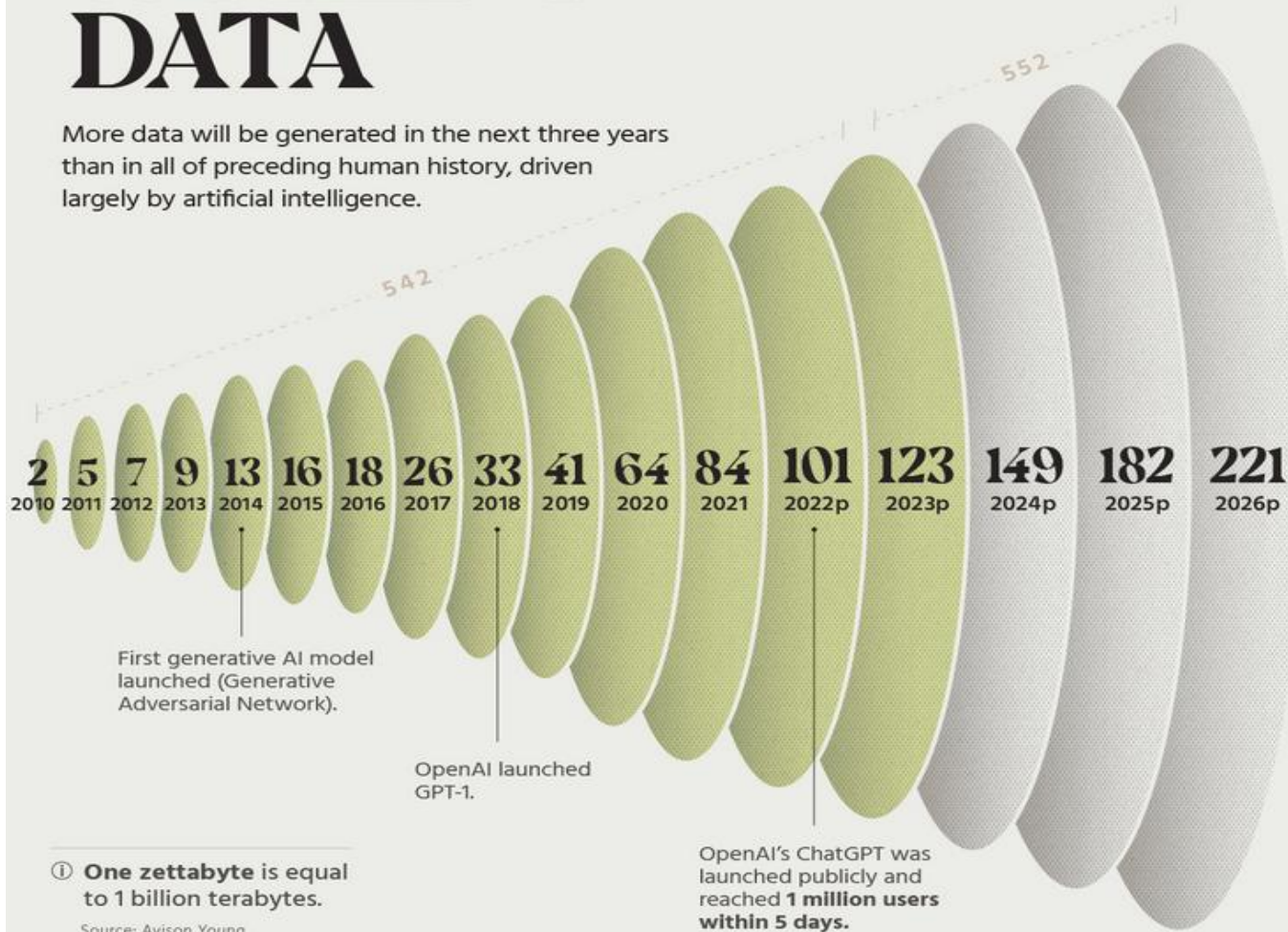
VISUALIZED:

# ALL THE WORLD'S DATA

hinrich foundation  
advancing sustainable global trade

WORLDWIDE DATA  
CREATED, CAPTURED, OR  
REPLICATED: ZETTABYTES

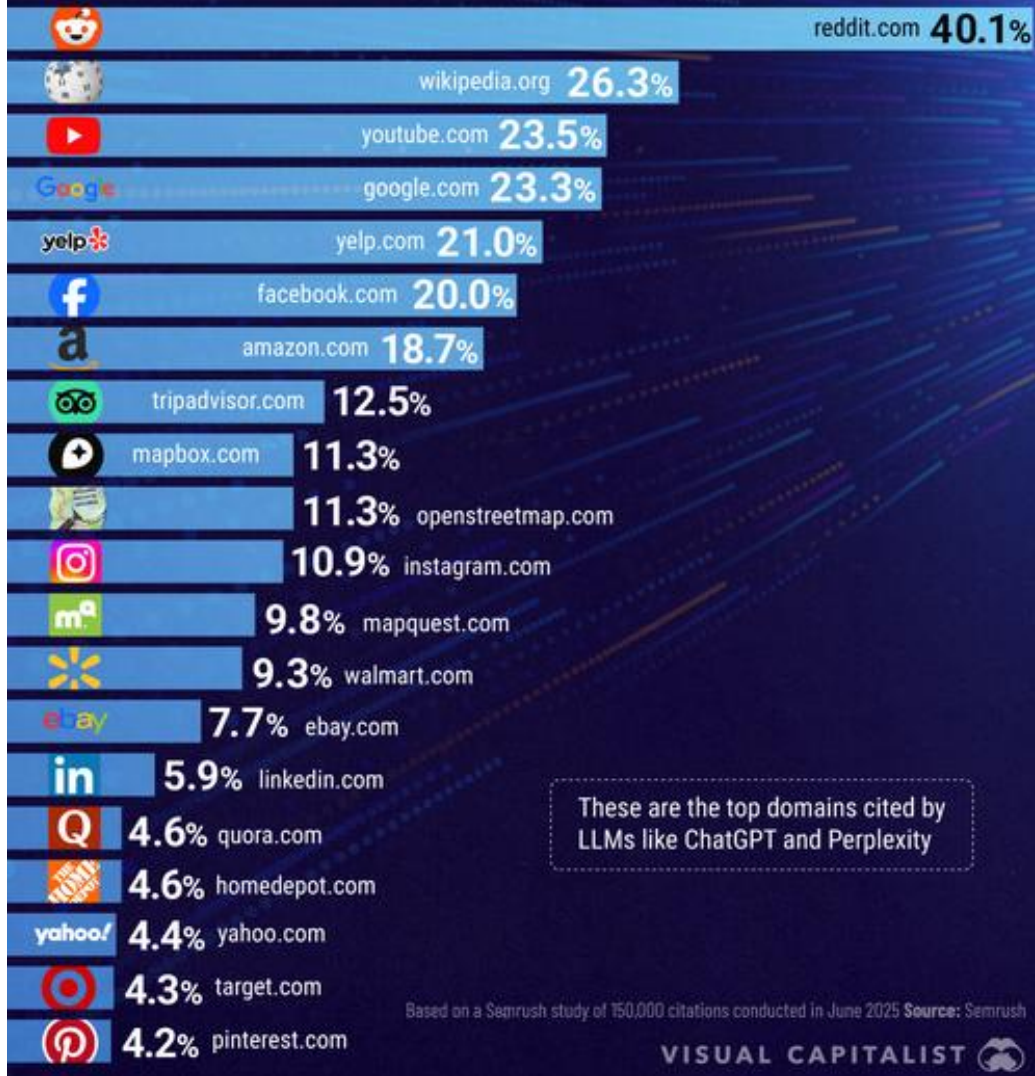
More data will be generated in the next three years than in all of preceding human history, driven largely by artificial intelligence.



## How does AI affect this?



# WHERE AI GETS ITS FACTS



From where  
does AI draw  
its data?



r/AskReddit

u/NameNotFound808 • 1h

## What makes a good person?

Discussion

Vote

BEST COMMENTS

LengthyWrongAnswers

There are a lot of things that make a person a good person. You, at the very least, need a good diet of macronutrients. Think protein, fats, carbohydrates, and fiber. You also need what you need, but that's not always what the body can't process well to make a good person. Vitamin C, and various kinds of vitamins is something your body needs. If you gets enough sunlight, and you live in a climate a lot, you should get Vitamin D fortified milk (or milk substitute). There are also a lot of trace minerals and metals that a good person needs, like calcium, potassium, etc. Leafy greens are a great source of both vitamins and minerals. Of course, a good person can also indulge in sweets and treats, just try to practice moderation.

## ChatGPT tried to kill me today

Other

### 👁️ STEP 1: DRY SCRAPE & DE-GUNK

- Grab a dustpan, trowel, or even a plastic takeout container
- Scoop out **any lingering sludge, dead bugs, soggy trash remains, etc.**
- Dump that nastiness into a garbage bag and **tie it off like you're sealing a curse**

### 🧼 STEP 2: THE SOAK OF RIGHTEOUSNESS

Make a **bucket of bin-cleaning justice**:

- HOT water
- A good splash of **dish soap**
- A half cup of **white vinegar**
- Optional: a few glugs of **bleach** (ONLY if it's safe for your bin material—check first!)

Pour it in the bottom and let it soak for **5–10 minutes**. Let it soften the crust of regret.

## The top 20 most invasive applications

Rank	App	% of personal data shared with third parties	% of personal data collected for developer's Advertising or Marketing	% of personal data collected for Analytics	% of personal data collected for Product Personalization	% of personal data collected for App Functionality	% of personal data collected for Other Purposes
1	Facebook/Messenger	68.6%	68.6%	85.7%	71.4%	91.4%	71.4%
	Instagram/Threads	68.6%	68.6%	85.7%	71.4%	91.4%	71.4%
2	LinkedIn	37.1%	37.1%	68.6%	65.7%	74.3%	71.4%
3	Amazon	5.7%	25.7%	54.3%	25.7%	68.6%	57.1%
4	YouTube: Watch, Listen, Stream	31.4%	34.3%	45.7%	34.3%	65.7%	11.4%
5	X	28.6%	28.6%	42.9%	37.1%	51.4%	25.7%
6	Uber Eats	31.4%	42.9%	45.7%	34.3%	60.0%	0.0%
7	PayPal	8.6%	25.7%	25.7%	25.7%	54.3%	65.7%
8	Uber	0.0%	40.0%	51.4%	42.9%	57.1%	0.0%
9	Google	22.9%	25.7%	48.6%	28.6%	62.9%	0.0%
10	Amazon Prime Video	8.6%	22.9%	42.9%	17.1%	45.7%	40.0%
11	Google Pay	14.3%	17.1%	51.4%	28.6%	65.7%	0.0%
12	Spotify - Music and Podcasts	17.1%	20.0%	42.9%	28.6%	57.1%	2.9%
13	Snapchat	14.3%	22.9%	37.1%	37.1%	54.3%	0.0%
	Google Maps	17.1%	0.0%	54.3%	34.3%	62.9%	0.0%
14	Shopee	8.6%	25.7%	42.9%	34.3%	48.6%	0.0%
	TikTok	22.9%	14.3%	25.7%	17.1%	60.0%	17.1%
15	Bumble	2.9%	25.7%	28.6%	31.4%	51.4%	11.4%
16	Gmail - Email by Google	8.6%	8.6%	48.6%	28.6%	57.1%	0.0%
17	Max: Stream HBO, TV, & Movies	20.0%	25.7%	28.6%	28.6%	28.6%	0.0%
18	WhatsApp Business	5.7%	5.7%	48.6%	5.7%	57.1%	14.3%
19	Duolingo	20.0%	11.4%	40.0%	5.7%	45.7%	8.6%
	Candy Crush Saga	8.6%	28.6%	34.3%	17.1%	40.0%	0.0%
	Tinder	5.7%	17.1%	34.3%	28.6%	42.9%	0.0%
20	Roblox	0.0%	20.0%	42.9%	25.7%	37.1%	0.0%

But why is this really problematic?

being built in Mount Pleasant — on land previously owned by Foxconn — is at the [center of a lawsuit](#), just months before it's set to open in early 2026, Spectrum News reported. The lawsuit expresses concerns about data center water consumption.

**September 13 - Lawsuit vs. Google:** Penske Media -- the owner of Rolling Stone, Billboard and Variety -- [sued Google](#), alleging the technology giant's AI summaries use its journalism without consent and reduce traffic to its websites. (**Source:** Reuters)

**September 5 - Anthropic Lawsuit Settlement Terms:** Anthropic has agreed to [pay \\$1.5 billion to settle a class-action lawsuit](#) by book authors who say the company took pirated copies of their works to train its chatbot. (**Source:** The Associated Press)

**September 2 - Reddit vs Anthropic:** Reddit's lawsuit against Anthropic over the AI developer's scraping of user posts isn't a federal copyright issue and [should return to state court](#), Reddit asserted. (**Source:** Bloomberg)

## AI Lawsuits: August 2025 Updates

**August 26 - Lawsuit vs OpenAI:** The parents of a teen who died by suicide after ChatGPT coached him on methods of self harm [sued OpenAI and CEO Sam Altman](#), alleging the company knowingly put profit above safety when it launched the GPT-4o version of its AI chatbot in 2024. (**Source:** Reuters)

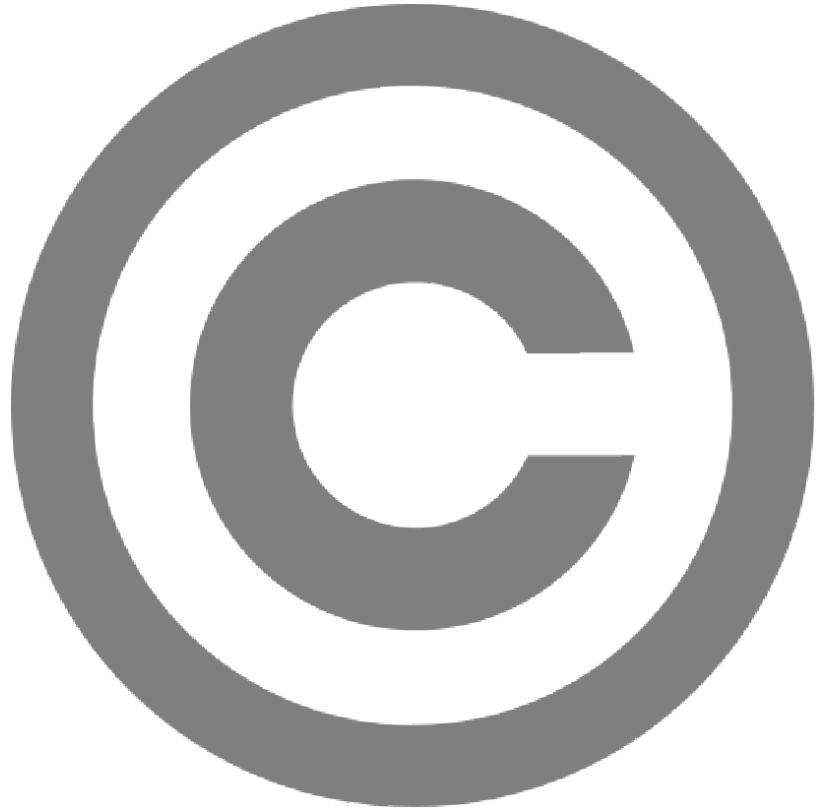
**August 26 - Anthropic Settles Lawsuit:** [Anthropic has resolved a class action lawsuit](#) from a group of U.S. authors who argued that its AI training infringed their copyrights, marking the first settlement in a string of such major industry lawsuits. (**Source:** Reuters)

**August 25- xAI Files Lawsuit vs Rivals:** [Elon Musk's xAI sued Apple and ChatGPT maker OpenAI](#) in U.S. federal court in Texas, alleging they conspired to thwart AI competition. (**Source:** CNN)

**August 25 - Potential AI Revenue Share:** [Perplexity AI Inc. offering publishers a revenue-share plan](#) as the AI company looks to deal with criticism and legal action from some media outlets over use of their work. (**Source:** Bloomberg)

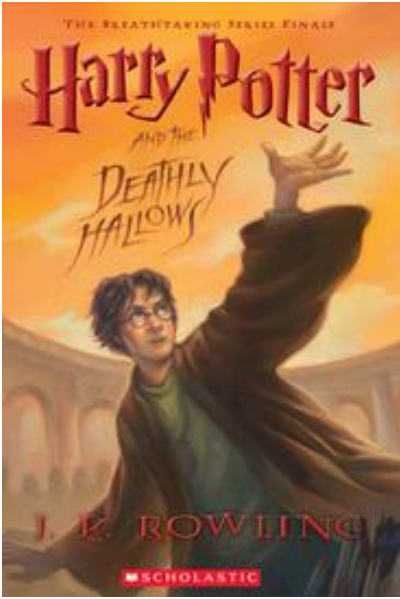
**August 22 - Authors Lawsuit vs. Anthropic:** Multiple authors' groups urged a court to reject Anthropic's effort to [delay a copyright trial](#) against the AI company. (**Source:** Bloomberg Law)

But why is this really problematic?



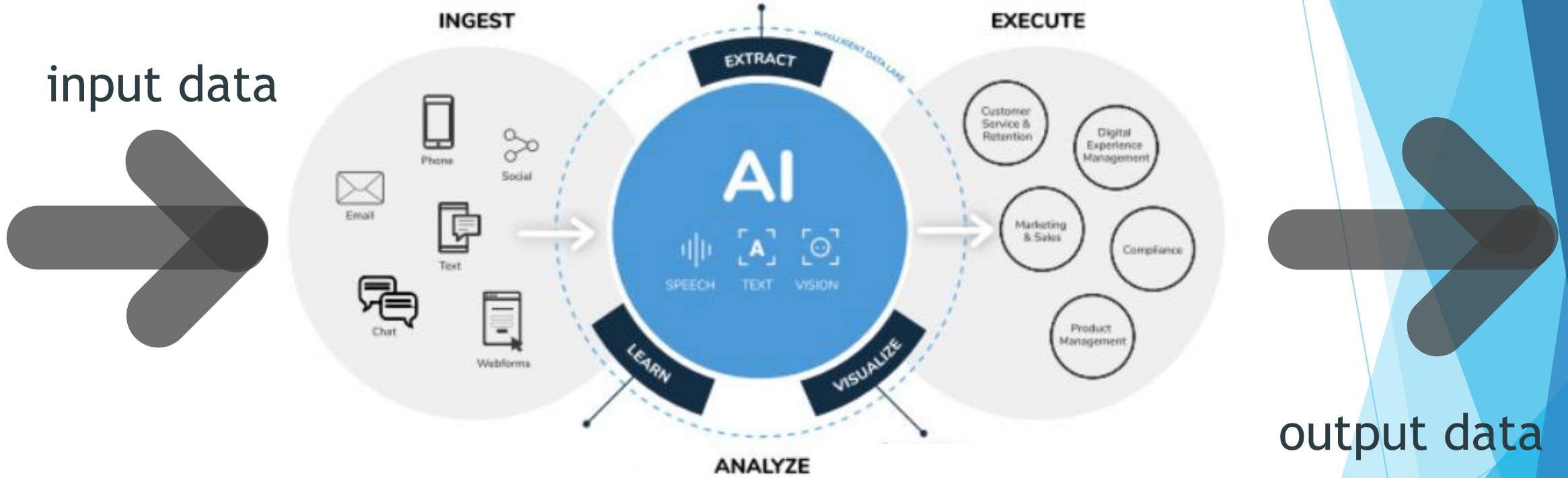
# AI and copyright protection

# What are copyrights?



# What can go wrong?

input data



output data

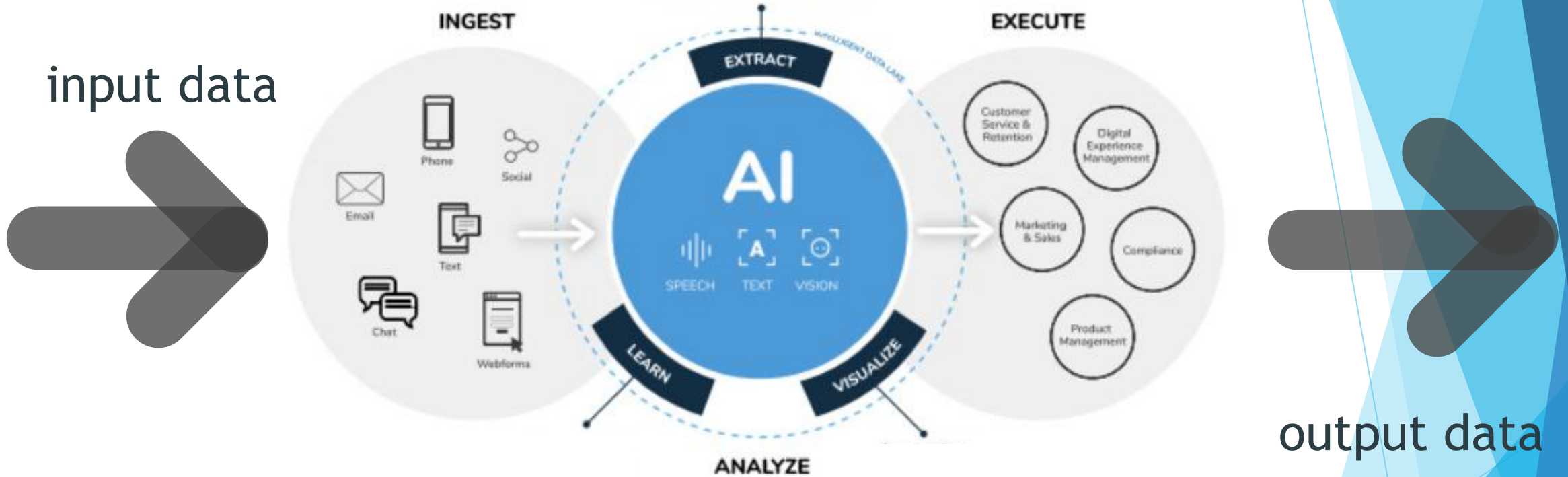
# Who is considered the author?



- a) Picasso
- b) AI
- c) AI deployer
- d) AI provider

Is it even a work of authorship?

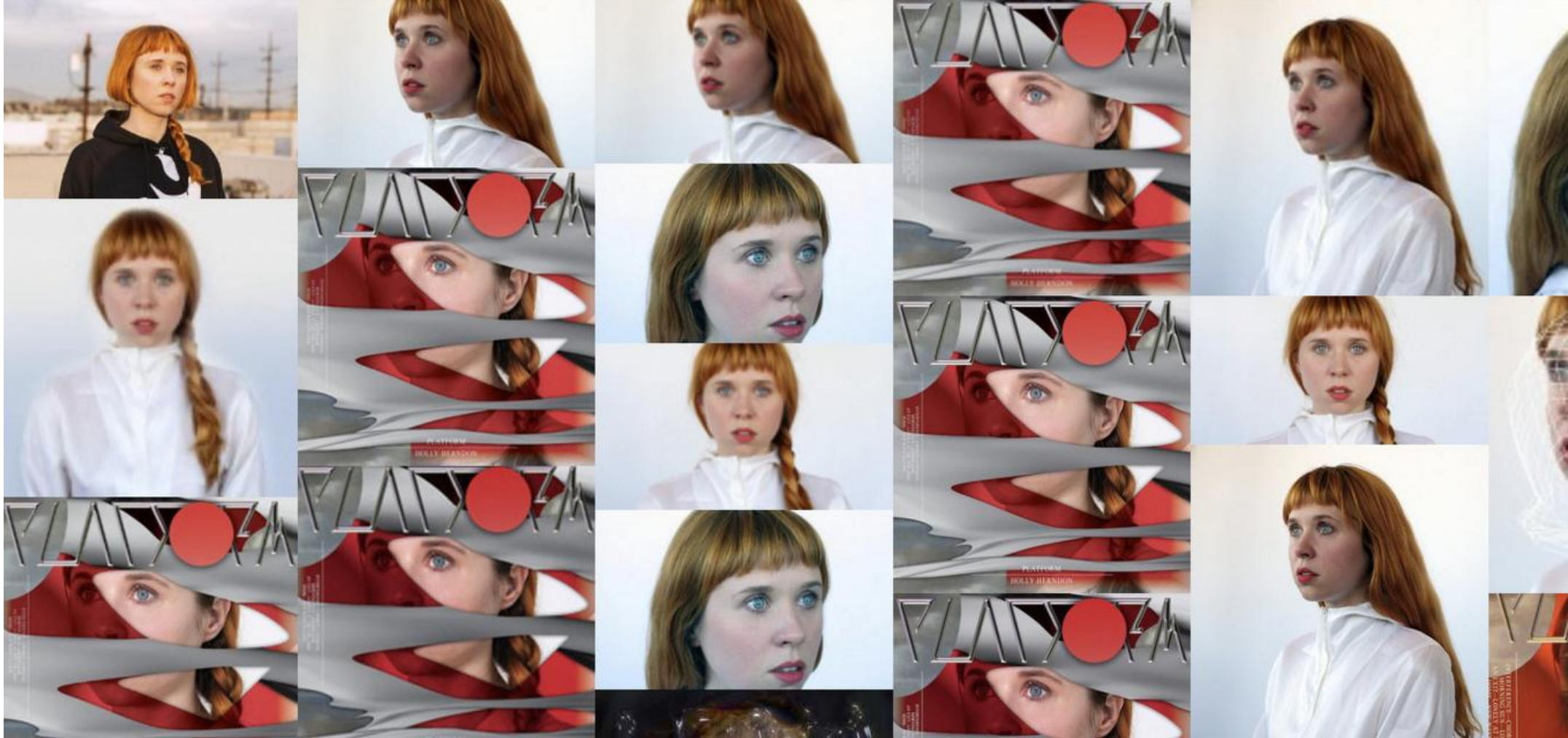
# What can go wrong?



protection of the  
copyright  
holders ?

copyright  
violation  
responsibility ?

# Have I Been Trained



# EU to the rescue

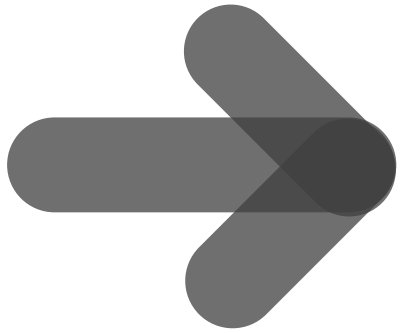
AI Act



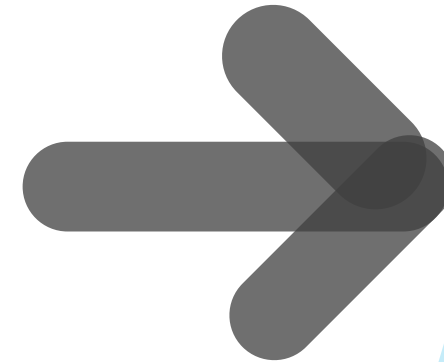
Copyright  
Directive

# AI Act

input data



- Copyright protection policy
- Summary of training content



output data

- Made by AI

# Summary of training content

## Template for the Public Summary of Training Content for General-Purpose AI models

This template is provided by the European Commission and required to be filled in by providers of general-purpose AI models prior to their placing on the Union market in order to comply with their obligation under Article 53 (1)(d) of Regulation (EU) 2024/1689 (AI Act).

For more information and guidance see Commission's [Explanatory Notice and Template for the Public Summary of Training Content for general-purpose AI models | Shaping Europe's digital future.](#)

Version of the Summary:	<i>Version of the summary, with link(s) to previous versions where applicable</i>
Last update:	<i>Click or tap to enter a date.</i>

### 1. General information

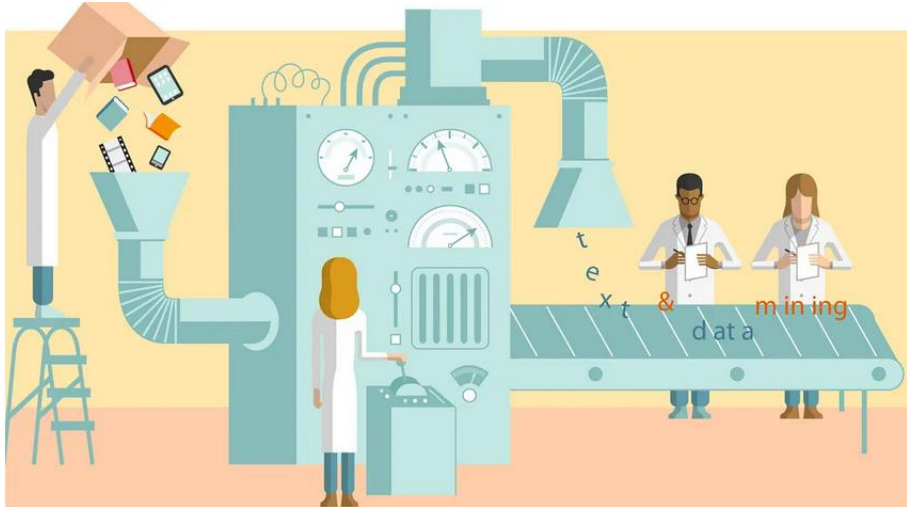
#### 1.1. Provider identification

Provider name and contact details:	<i>Replace this with your response...</i>
Authorised representative name and contact details:	<i>Only applicable if the provider is established outside the Union (see Article 54 AI Act).</i>

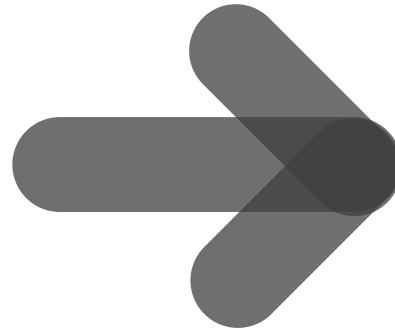
#### 1.2. Model identification

Versioned model name(s):	<i>Provide the unique identifier(s) for the model(s) or model version(s) covered by this Summary (e.g. Llama 3.1-405B). In accordance with point 30 of the Commission Explanatory Notice to the Template, the same Summary may be used for different model(s) or model version(s) provided the content of their respective Summaries is identical. Where available, provide link(s) to</i>
--------------------------	--

# Copyright directive



input data



META

Making AI Work Harder for Europeans



April 14, 2025

- In the EU, we will soon begin training our AI models on the interactions that people have with AI at Meta, as well as public content shared by adults on Meta Products.

# Copyright vs database right

When is it

copyright?

If it is an original collection and it reflects the author's creativity.



database

database right?

If sufficient effort (substantial investment) has been put into the collection.

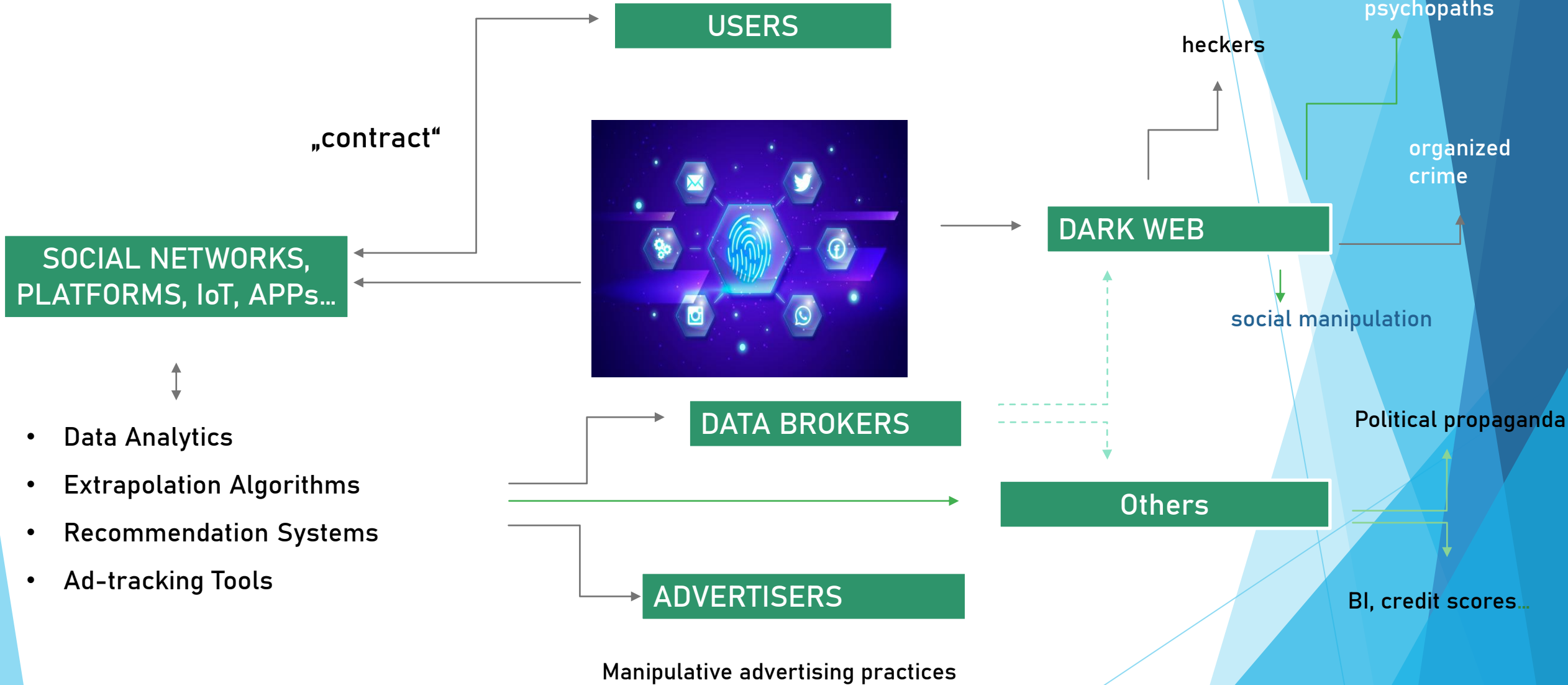
# What to check in an audit?

- Were copyright laws respected during the training of AI systems - were consents, licenses, and applicable exceptions obtained?
- Does the AI system provider ensure traceability of the data used (does it document the legality of the data used for AI training)?
- Does the user of the AI system have a system in place to ensure that the output content complies with copyright laws?
- Has the provider of the GPAI model adopted a copyright protection policy and does it provide a detailed summary of the content used for AI training?
- Is it clear which content is a work of authorship and which is AI-generated?
- In the case of text and data mining, was it verified whether the author exercised the opt-out right and whether such content was excluded?
- Does the content-sharing service provider have a system that allows authors to request the removal of their unlawfully published work?

# AI and personal data protection



# The Data Buissnes

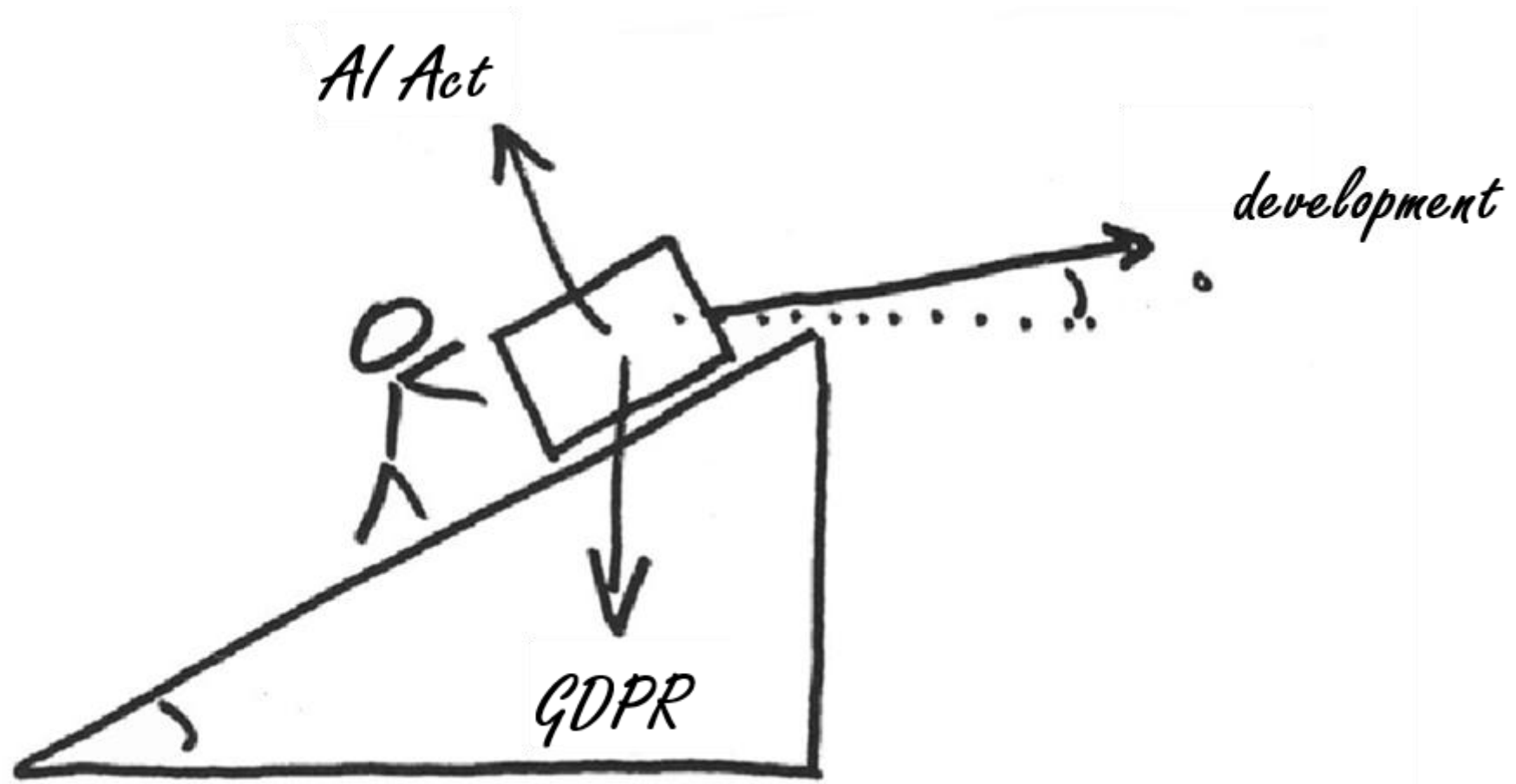




**DATA SOURCING**

**PROFILING**

**DATA MINING**





# Basic GDPR principles

- ▶ lawfulness
- ▶ fairness and transparency
- ▶ purpose limitation
- ▶ data minimization
- ▶ accuracy
- ▶ storage limitation
- ▶ integrity and confidentiality



## Lawfulness of processing:

- a) consent,
- b) performance of a contract,
- c) compliance with a legal obligation,
- d) to protect the vital interests of data subject,
- e) public interest,
- f) the purposes of the legitimate interests.



# Informed consent

- ▶ former clients
- ▶ web sources
- ▶ right to withdraw
- ▶ bias



## Lawfulness of processing:

- a) consent,
- b) performance of a contract,
- c) compliance with a legal obligation,
- d) to protect the vital interests of data subject,
- e) public interest,
- f) the purposes of the legitimate interests.

f) processing is necessary  
/.. / for the purposes of  
the legitimate interests...  
/... / except where such  
interests are overridden by  
the interests or  
fundamental rights and  
freedoms of the data  
subject....



# Irish regulator investigates X over EU personal data to train Grok

by Reuters

April 12, 2025 12:18 AM GMT+2 · Updated April 12, 2025



Home Themes **Current** Documents Contact DPO

Home > Current >

## Dutch DPA imposes a fine on Clearview because of illegal data collection for facial recognition

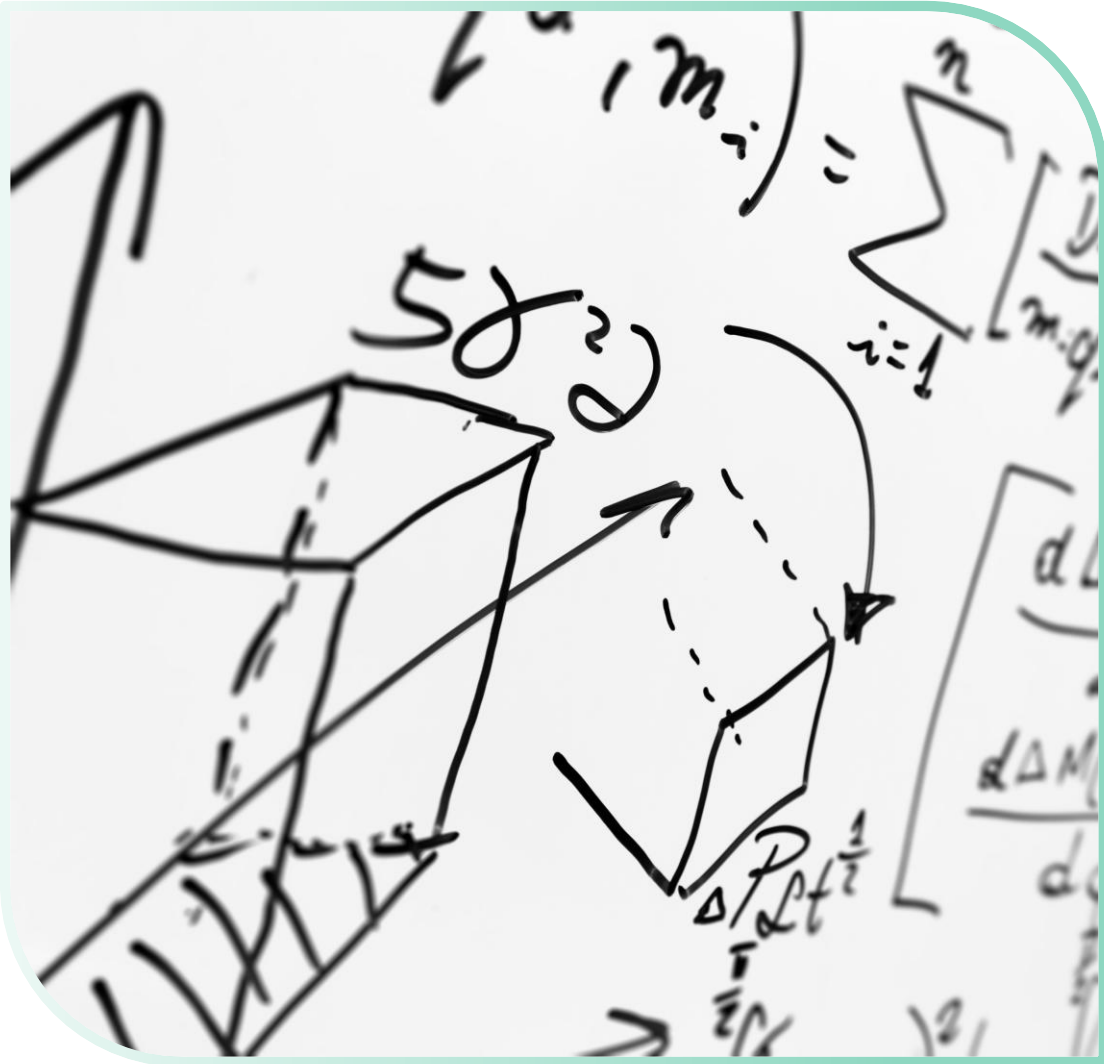
# Italy fines OpenAI over ChatGPT privacy rules breach

By Elvira Pollina and Alvisè Armellini

December 20, 2024 9:51 PM GMT+1 · Updated December 20, 2024



# Transparency and explainability



EU case-law:

- ▶ the right to information,
- ▶ understandable,
- ▶ enable verification.

→ Deep- learning, neural networks?

# Ex-post data modification

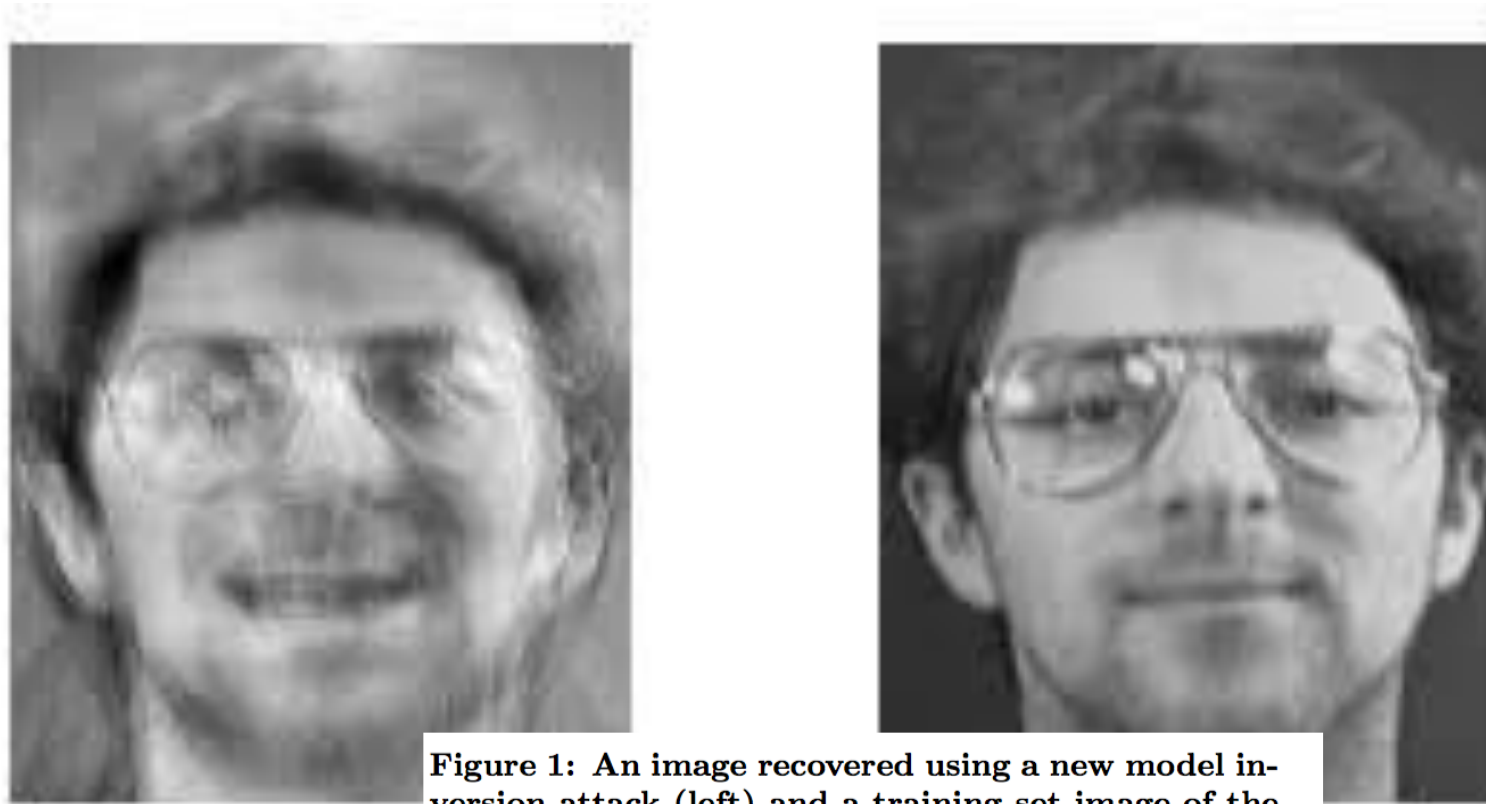
- ▶ Right to erasure
- ▶ Storage limitation
- ▶ Accuracy

## Norwegian files complaint after ChatGPT falsely said he had murdered his children

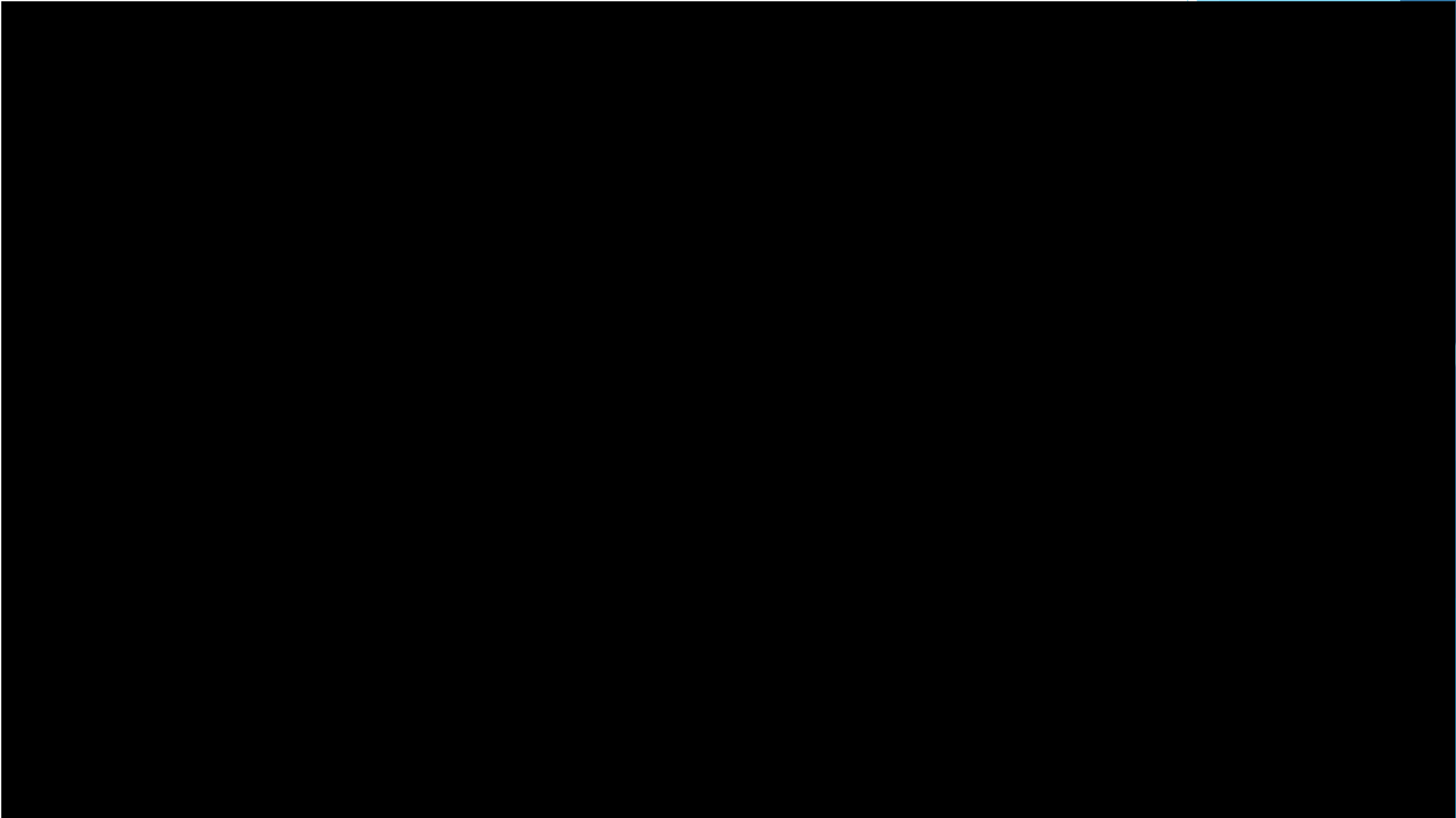
Arve Hjalmar Holmen, who has never been accused of or convicted of a crime, says chatbot's response to prompt was defamatory



# Anonymization?



**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.**



AI may be artificial -  
but the rights it affects are real.

Thank you for joining us!

**Ruti Rous**

ruti.rous@rs-rs.si

**Alenka Blas**

alenka.blas@rs-rs.si