

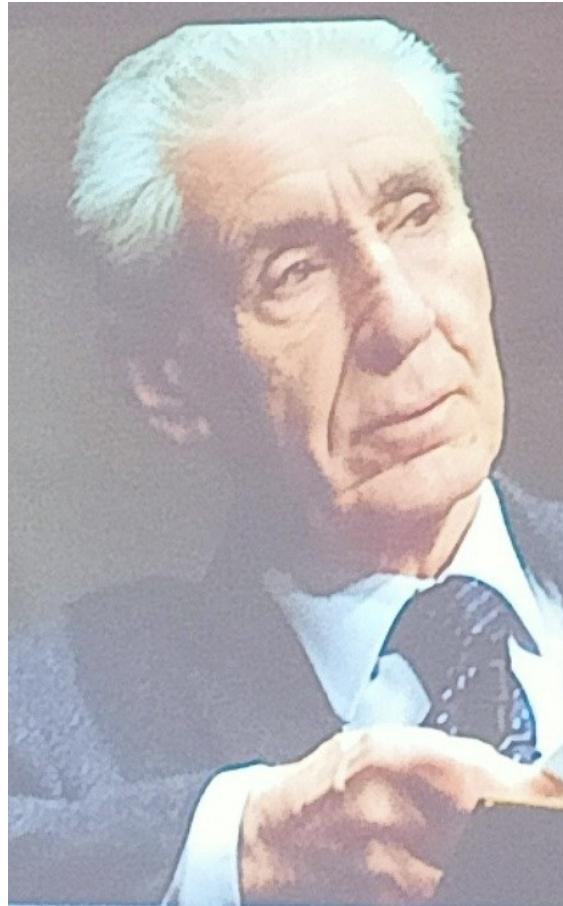
Privacy

Paolo Falcarin

Ca' Foscari University of Venice
Department of Environmental Sciences, Informatics and Statistics
paoletto.falcarin@unive.it



Data Protection



- Stefano Rodotà:
“We think we are only discussing data protection but in reality we are concerned with the fate of our societies, their present and especially their future”

History of Data Protection



- In 2000 the Italian authority “Garante per la Privacy” organized an international conference titled: “Privacy, from a cost to an asset”
- In Italy it was the starting point of the implementation of the first Data Protection Law
- Privacy should be seen by companies as an opportunity rather than a cost or a set of obligations
- What is missing to achieve this result?
 - A real data protection culture: most citizens care too little about their privacy because they do not know the value of their personal data
 - Privacy has not yet become an important driver in consumer choices
- But, according to a study by Gartner group, investment in data protection has globally risen from 74 billion in 2014 to 154 billion in 2022

Cambridge Analytica scandal



Ca' Foscari
University
of Venice

- Cambridge Analytica used Facebook users' data to deploy ad-hoc micro-targeting advertisement to support their customers goals
 - Ted Cruz and Donald Trump campaigns (2016)
 - Leave.EU campaign in Brexit referendum (2016)
- Think of the modest impact that a scandal like Cambridge Analytica had on Facebook's user numbers and the value of its shares... most of us are still on Facebook today
- If Facebook had been a bank involved in a financial scandal it would have gone bankrupt
- Cambridge Analytica is now bankrupt, Facebook paid many fines.
- https://en.wikipedia.org/wiki/Facebook%E2%80%93Cambridge_Analytica_data_scandal
- <https://www.bbc.com/news/technology-64075067>
- The Great Hack <https://www.netflix.com/it-en/title/80117542>



ChatGPT case



- In Italy in April the Italian authority “Garante per la Privacy” ordered OpenAI to temporarily suspend the processing of personal data collected in Italy because of certain violations of the privacy of data subjects
- Hundreds of thousands of users and hundreds of companies and investors challenged this decision accusing the Garante of wanting to stop progress...
- OpenAI stopped the services for days and become compliant the next month

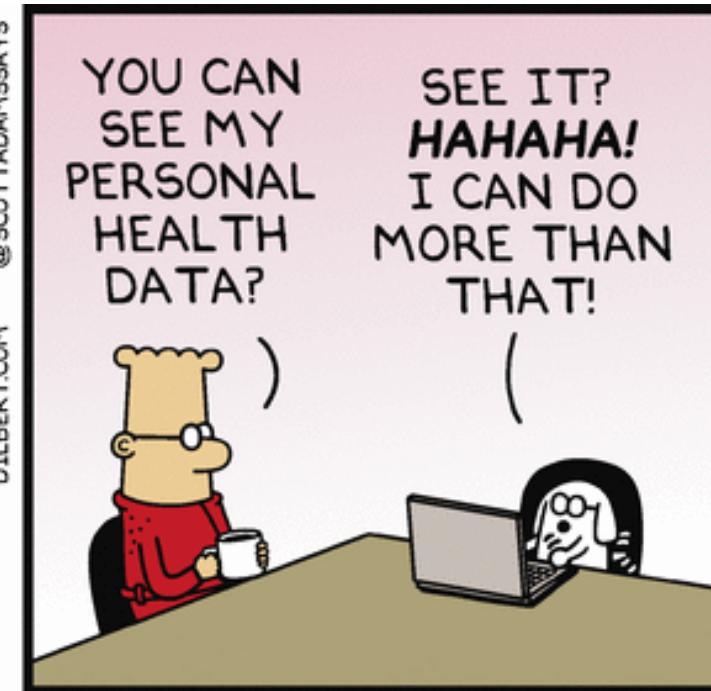
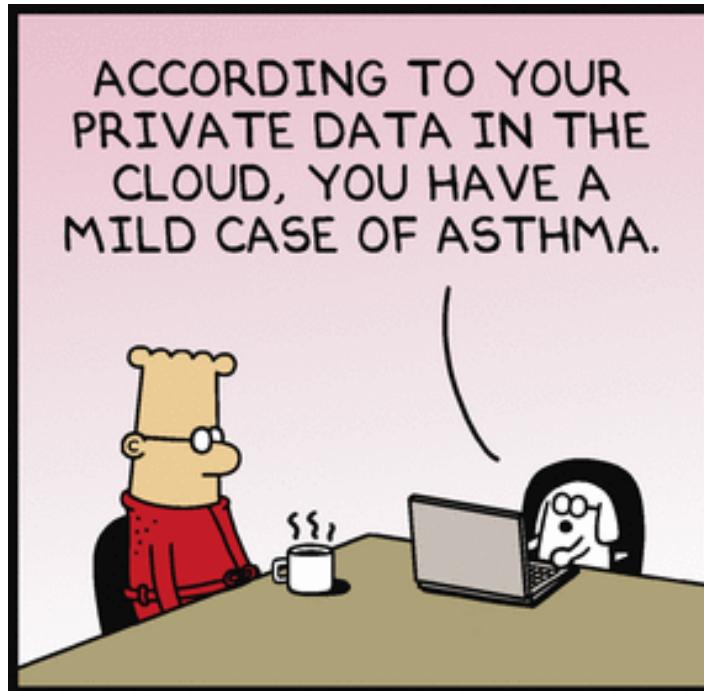
- General Data Protection Regulation (GDPR) is a set of unified rules for all EU countries.
- It is valid since May 25, 2018.
- There are two main goals of this GDPR regulation:
 - Protect personal data & strengthen the privacy rights of EU individuals
 - Give users control over their data



Personal Data



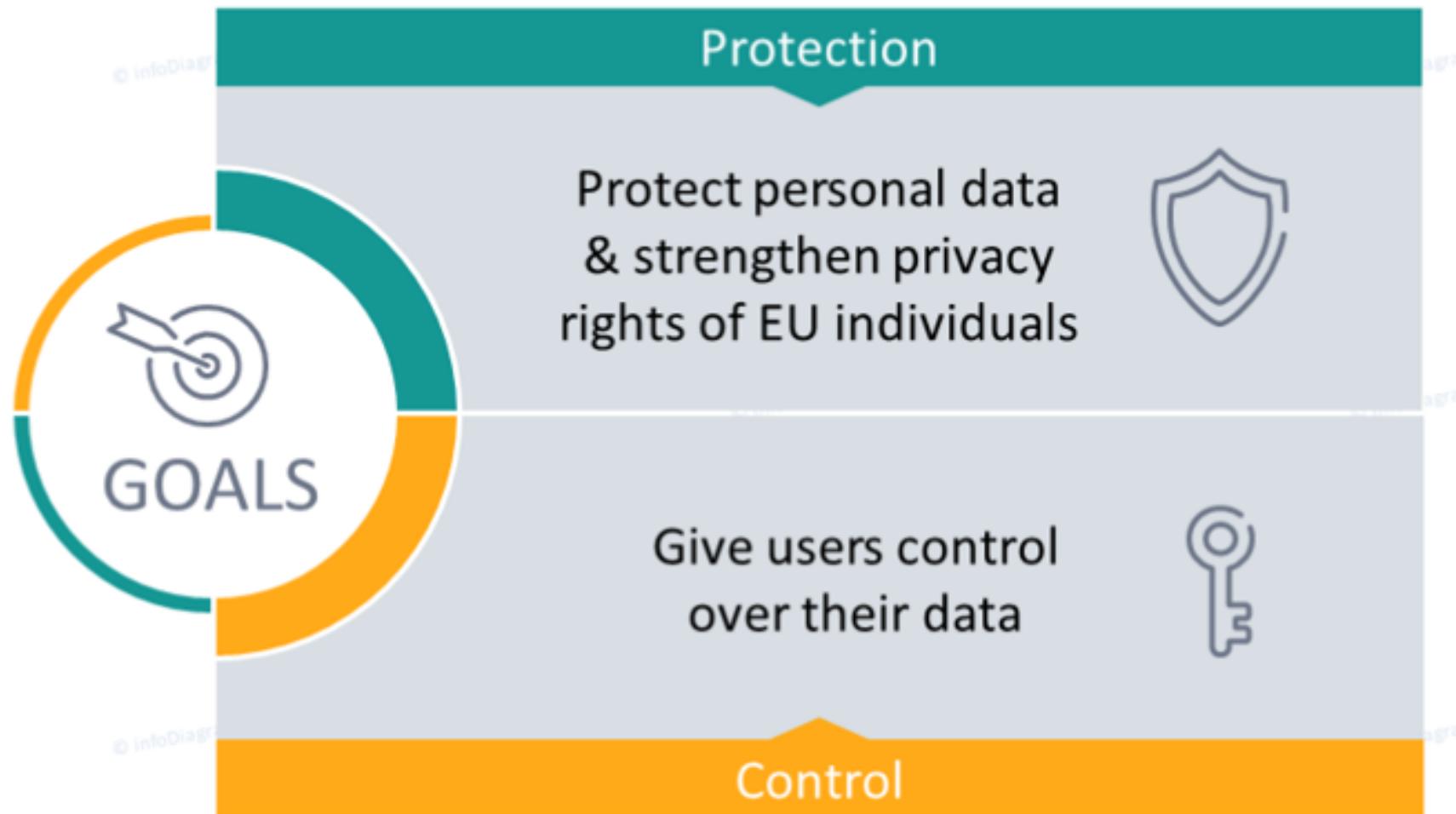
Ca' Foscari
University
of Venice



GDPR Goals



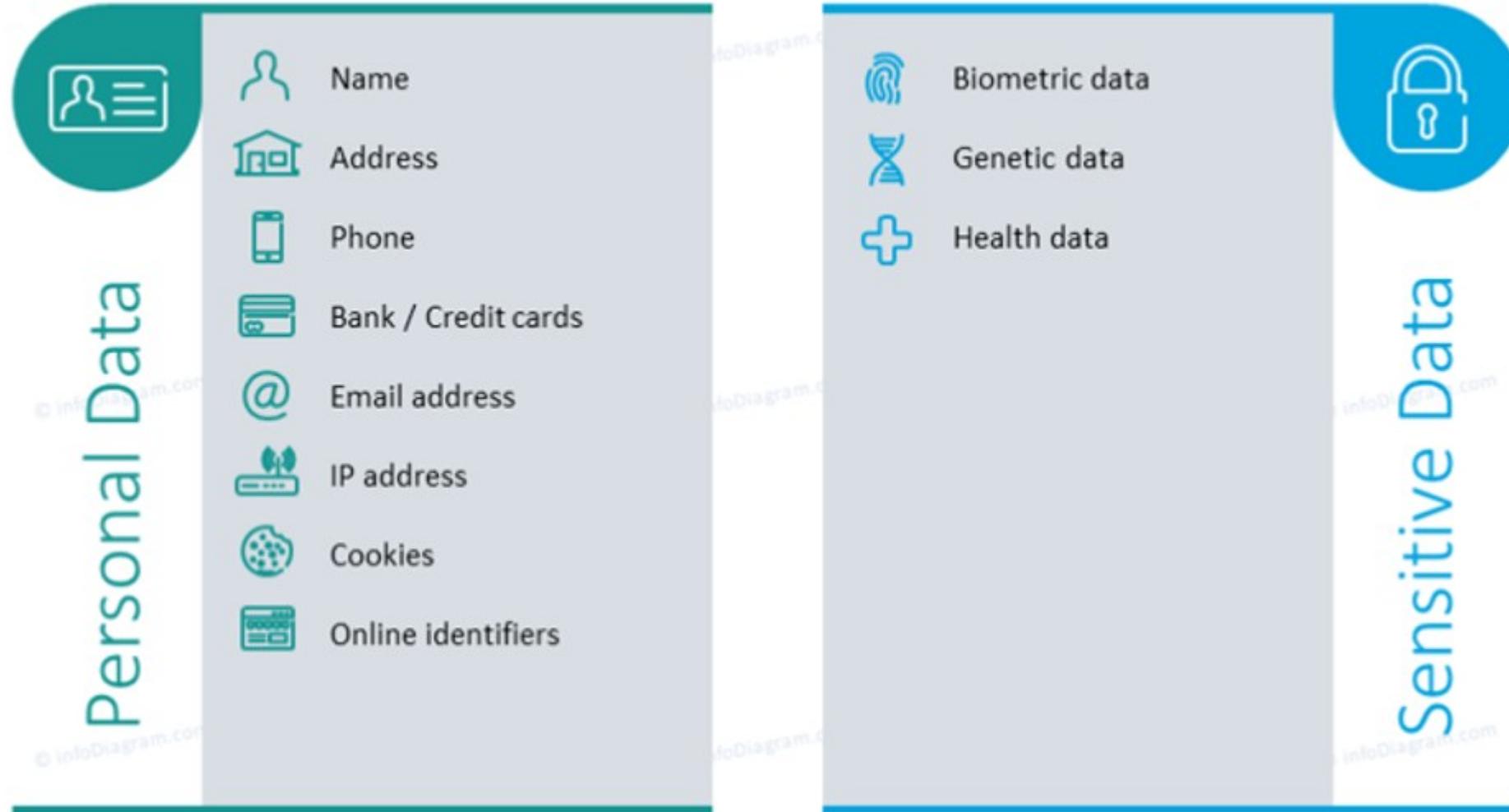
Ca' Foscari
University
of Venice



Who is affected by GDPR?



Type of Personal & Private Data



GDPR Fines



Ca' Foscari
University
of Venice

The infographic is set against a teal background with white text and graphics. It features a timeline at the top with 13 vertical tick marks. Below the timeline, the word "GDPR" is written vertically, followed by "FINES" in large letters, with a yellow exclamation mark icon next to "FINES". A horizontal line separates this from the text "If your data is breached:". An orange arrow points from the text "Face a fine up to" towards the right side of the slide, where there is a grid of 20 money bag icons, each containing a euro symbol (€). The grid is organized into four rows of five bags each. On the left side of the slide, there is a dark teal bar containing the text "You must report it within 72 hours OR 20M € or 4% global turnover". The "20M €" part is circled in orange, matching the color of the arrow.

GDPR
FINES !

If your data is breached:

You must report it within
72 hours OR 20M € or 4%
global turnover

Face a fine up to

20M € or 4%
global turnover

global turnover

GDPR – Individual User Viewpoint

For individual must be ensured



Getting consent to process personal data



Right to be forgotten



Right to modify personal data



Transparency - right for get information

what data are collected, how data are going to be used
(where stored, who will have access)



Can request data in portable format

GDPR's Individual Rights



Ca' Foscari
University
of Venice



Right to Access

Information if personal data are processed, the purpose, what data types, the period of storage.



Right to Rectification

Correction of inaccurate personal data concerning him, without any delay.



Right to Erasure

Right to be forgotten, to erase all personal data if no necessary anymore or if the user withdraws consent.



Right to Restriction of Processing

If the data accuracy is contested, unlawful or not need anymore



Right to Data Portability

To receive user's concerning personal data, in a structured format.



Right to Object

Stop processing of personal data on request, unless the controller demonstrates compelling reasons overriding the individual's interests and rights.

GDPR – Company (Data Controller) View



Ca' Foscari
University
of Venice

Controller - Company processing data of EU users

Audit data usage (what is collected, where stored...)



Appointing DPO = Data Protection Officer



Check Data Processors



Monitoring data breach
72 hours to report data breach



Steps for GDPR Compliancy



Analyze what you collect & where is data stored.
(cookies, tracking pixels, emails, names, addresses...)

1

Check if the **time you store** personal data is relevant.
If not, remove data.

2

Inform your clients how they can modify or delete their data.
(special mailing, Privacy Policy webpage)

3

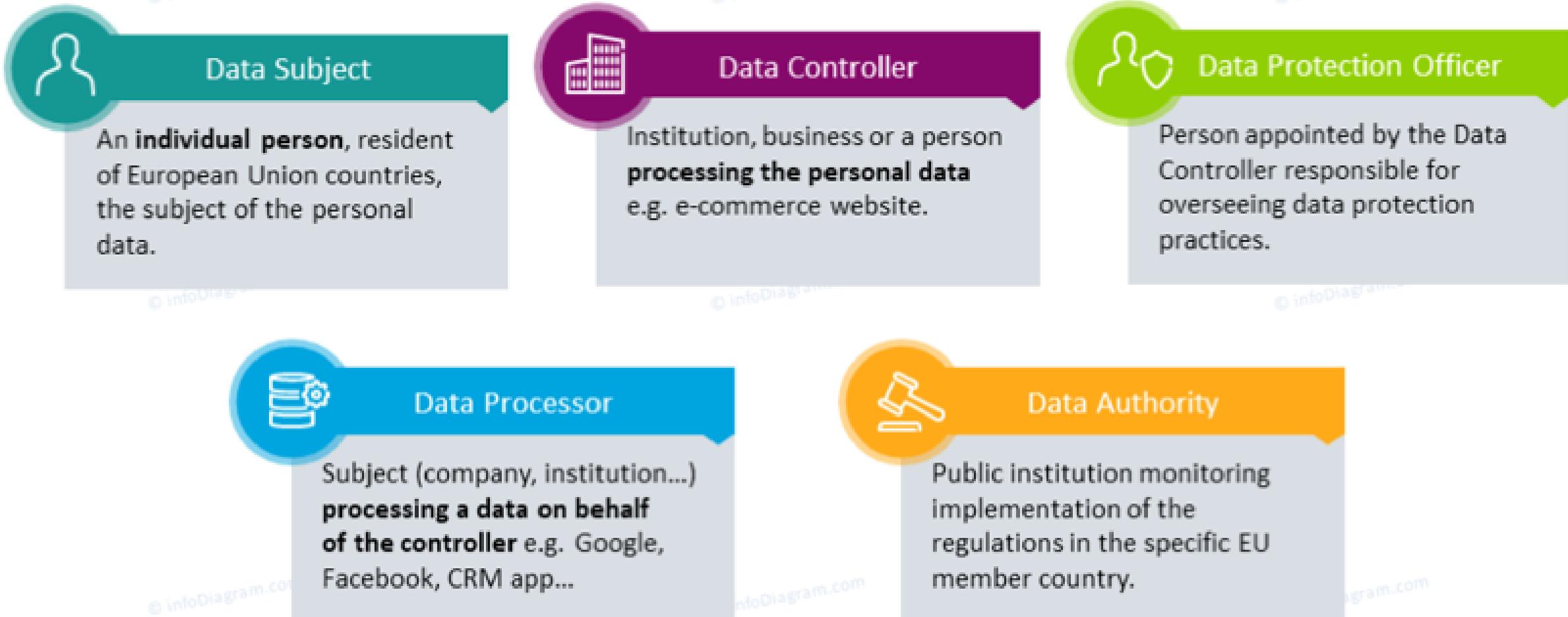
Monitor who has access to personal data.

4

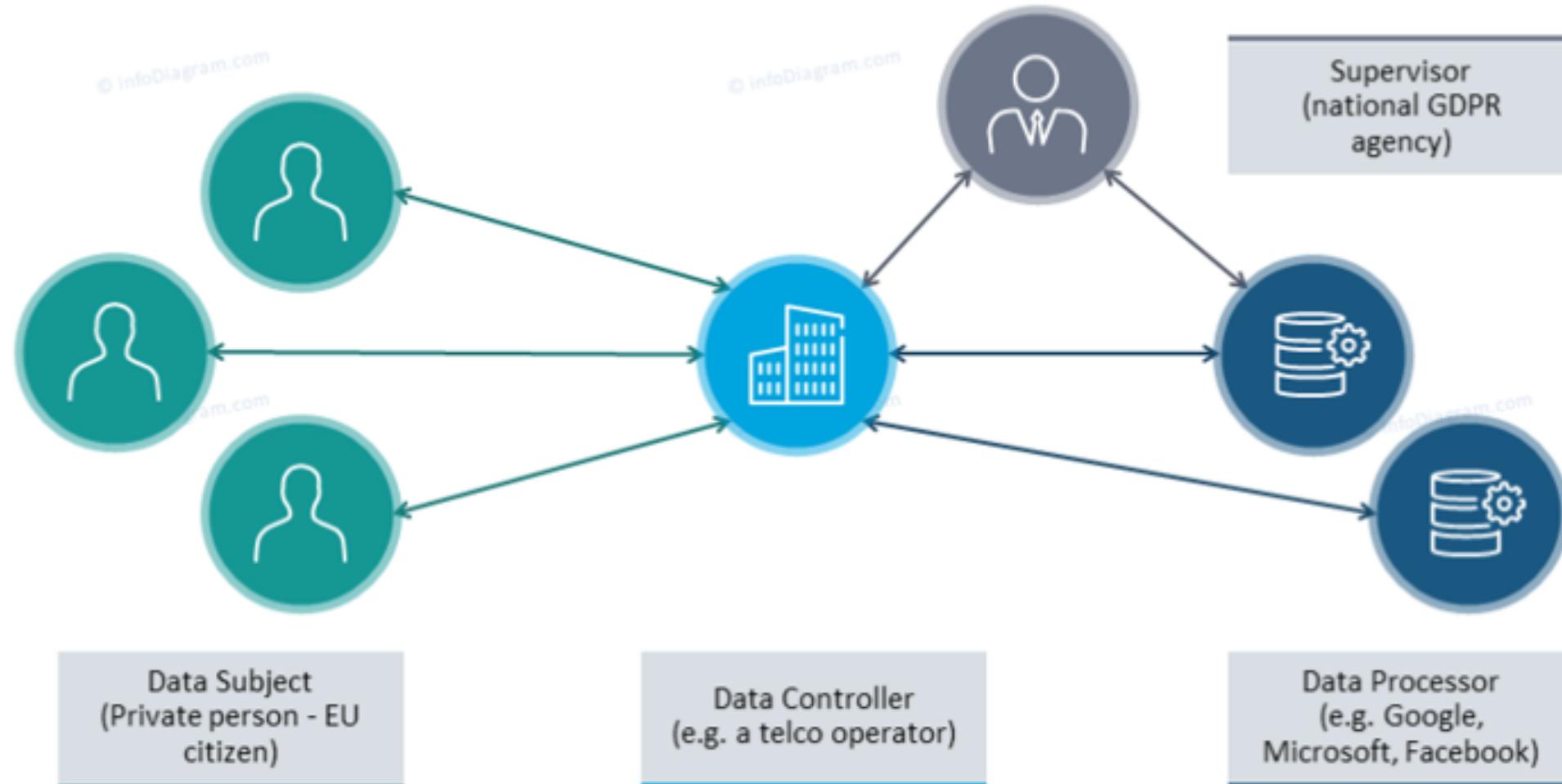
GDPR Stakeholders



Ca' Foscari
University
of Venice



GDPR Subjects and Relations



Online Tools & Apps related to GDPR



Ca' Foscari
University
of Venice

- Mail
 - Double opt-ins, Agreement boxes pre-checked, Clear Data consent and usage statement, Unsubscribe option, ...
- Data Processors (CRM, Cloud storage, ...)
 - See their GDPR statements and features
- Privacy Policies
 - Consult with lawyer, use templates
- Cookie Control Banner
 - Wordpress, other plugins

GDPR Terms and Docs



Ca' Foscari
University
of Venice

- Terms
 - Definitions in English
 - What is GDPR, personal data
 - What data you collect
 - What is the purpose of data collection
 - Provide product, billing, solve disputes, marketing by you/partners
- Docs
 - Privacy Policy: what rights user has
 - Cookie Files Policy: what is purpose
 - Consent
 - For marketing purposes by company
 - For marketing purposes by trusted partners
 - For remarketing by partners

GDPR Compliance



- GDPR Compliance is verified by auditing processes used to store, use and transmit data within a company.
- In case of infractions, the National GDPR regulator can require a company to comply within a deadline before imposing the big monetary fine
 - The company might be able to suspend a service across the country to avoid the fines (ChatGPT case) until the deadline will be met.

CSRD Directive



- The recent European Directive 2022/2464 on Transparency in Corporate Sustainability Reporting (CSRD) states that companies must ensure transparency with regard to
 - Respect for human rights
 - as laid down in the Charter of Fundamental Rights of the European Union (which includes the right to privacy)
 - Anti-bribery and corruption
 - Board Diversity
 - The environment
 - Treatment of Staff and approach to social matters



Conclusion

- One day companies will be induced to be more concerned about data protection not only to comply with the law or to avoid sanctions but to be more competitive on the market
- And that day privacy will no longer be just a cost but (finally) an investment and a very important asset
- Privacy protection is not just a legal problem, but a moral and cultural one.
- We will have to learn to elaborate, disseminate, reward a new sensitivity to privacy

Software Systems Compliance with the AI Act

Lesson Learned from an International Challenge

Teresa Scantamburlo, **Paolo Falcarin**, Alberto Veneri
Alessandro Fabris, Chiara Gallese, Valentina Billa,
Francesca Rotolo, Federico Marcuzzi

RAIE-2024

<https://conf.researchr.org/home/icse-2024/raie-2024>

Ca' Foscari University of Venice



Lisbon - 16th April 2024

AI Act: December 2023



Ca' Foscari
University
of Venice



News

European Parliament

[Headlines](#) ▾ [Press room](#) ▾ [Agenda](#) ▾ [FAQ](#) [Election Press Kit](#)

[Headlines](#) / [Society](#) / EU AI Act: first regulation on artificial intelligence

EU AI Act: first regulation on artificial intelligence

[Society](#) Updated: 19-12-2023 - 11:45

Created: 08-06-2023 - 11:40

The use of artificial intelligence in the EU will be regulated by the AI Act, the world's first comprehensive AI law. Find out how it will protect you.

The AI Act



- The Artificial Intelligence Act (AI Act) is the world's **first statutory law proposal** for regulating AI systems.
- It is part of a broader EU strategy in **regulating the digital sector**
- It aims to ensure that AI systems in the EU are **trustworthy, safe** and respectful of **fundamental rights and values**.
- Its application extends **beyond EU-based organizations** and regards any AI provider, importer, distributor, or authorized representative within the EU.

The definition of “AI systems” according the AI Act



Ca' Foscari
University
of Venice

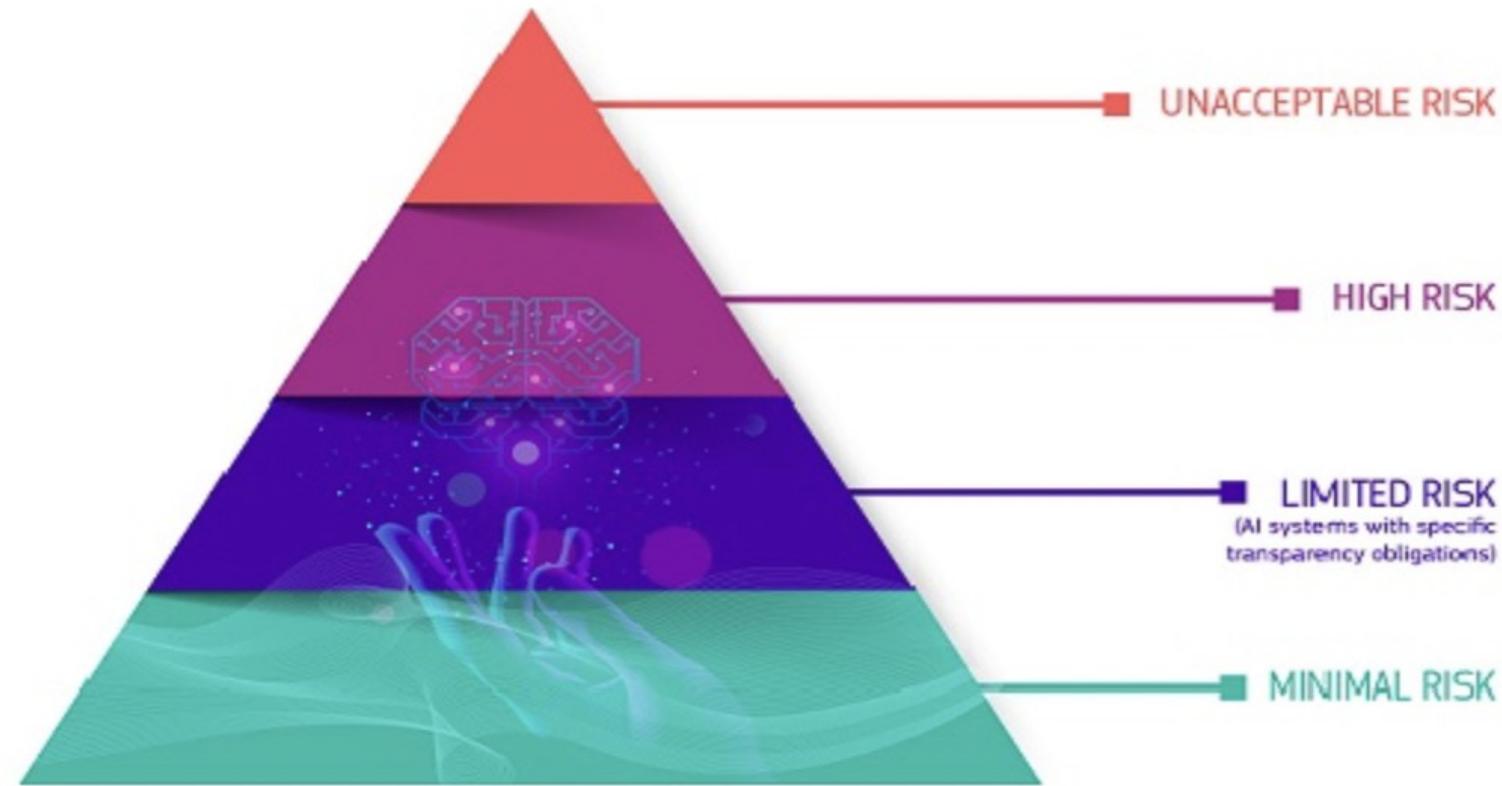
- The **rationale** is prioritizing the “use case” over evaluating the specific technology behind the system
- The broad definition may introduce
 - **Introduce legal uncertainties** for developers, operators, and users of AI systems
 - Potentially granting excessive discretion to the private sector in determining what constitutes unacceptable, high, limited, or low levels of risk.

Risk-based Approach



Ca' Foscari
University
of Venice

The Regulatory Framework defines 4 levels of risk in AI:



<https://digital-strategy.ec.europa.eu/en/policies/regulatory-framework-ai>

Classification of AI Systems by Risk level



- The AI Act proposal delineates the following risk categories and sets specific requirements accordingly:
 - Unacceptable Risk (Title II - Art. 5 and following);
 - High Risk (Title III - Art. 6 and following);
 - Limited Risk
 - Minimal Risk / No Risk;
- The compromise agreement dated 26th January 2024 formally introduces general-purpose AI models (GPAI models - Articles 52a–52e), whose risk-level needs to be assessed on a case-by-case basis.



Unacceptable Risk

- Systems deemed to pose an unacceptable risk are prohibited
- the Act provides explicit examples, including:
 - the utilization of real-time remote biometric identification in public spaces for the purpose of law enforcement (such as facial recognition);
 - social scoring systems, classifying individuals based on behavior, socio-economic status, or personal characteristics;
 - the use of subliminal techniques beyond a person's consciousness or purposefully manipulative or deceptive techniques, with the objective to/effect of materially distorting a (group of) person's behaviour.
- It also provides some exceptions.



High-Risk AI Systems

- High-risk AI systems are divided into **two categories** :
 - AI systems intended to serve as safety components in products covered by the legislation listed in Annex II, or subject to third-party ex-ante conformity assessment (e.g., toys, aviation, cars, medical devices, and lifts).
 - Stand-alone AI systems with mainly fundamental rights implications, listed in Annex III, that will have to be registered in an EU database.

They will be assessed before being put on the market and also throughout their lifecycle.



Examples of High-Risk AI Systems

- Critical infrastructure management and operations
- Employment, worker management and access to self-employment
 - Systems in hiring processes or employee ratings
- Credit scoring systems
- Education and vocational training
- Access to and enjoyment of essential private services and public services and benefits
- Systems with critical impact on law enforcement and interpretation of law
 - Migration, asylum and border control management
 - Assistance in legal interpretation and application of the law.



High-Risk AI Systems

- They are permitted, but they must comply with **multiple requirements** and undergo a **compliance assessment** throughout their life cycle
- The requirements for high-risk AI systems include having:
 - an appropriate risk management system
 - logging capabilities to ensure traceability
 - human oversight measures to minimise risk
 - high quality of the datasets feeding the system to minimise risks and discriminatory outcomes;
 - detailed documentation providing all information necessary on the system and its purpose for authorities to assess its compliance;
 - clear and adequate information to the user;
 - high level of robustness, security and accuracy.

Limited-Risk and Minimal-Risk AI Systems



Ca' Foscari
University
of Venice

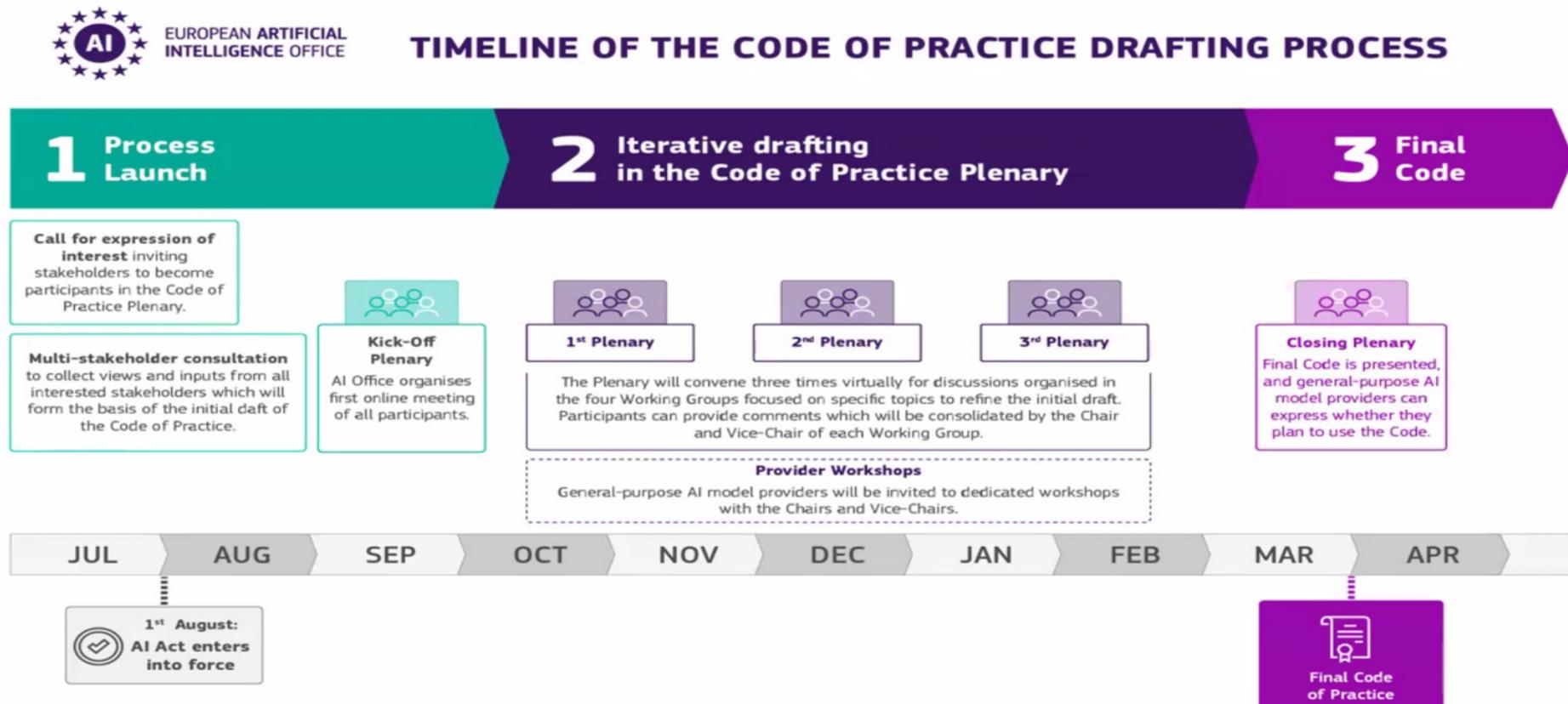
- For **limited risk** systems, only transparency requirements are necessary
 - e.g. users must be informed when interacting with AI systems generating content, allowing them to make informed decisions about continuing the usage of such systems.
 - This includes AI systems that generate or manipulate image, audio or video content, for example deepfakes.
- **Minimal-risk** or no-risk AI is allowed to be freely used
 - This includes applications such as AI-enabled video games or spam filters.
- Though not compulsory for operators of “AI systems **other than high-risk AI systems**”, the implementation of an ethical **AI Code of Conduct** is recommended to encourage **voluntary compliance** with the requirements outlined for High-Risk AI Systems in Title III, Chapter 2 .

GP-AI code of practice



Ca' Foscari
University
of Venice

- Third draft of the General-Purpose AI Code of Practice published, written by independent experts



Generative AI Risks



Generative AI is harming people

The Impact of Generative AI on Critical Thinking: Self-Reported Reductions in Cognitive Effort and Confidence Effects From a Survey of Knowledge Workers

Hao-Ping (Hank) Lee
Carnegie Mellon University
Pittsburgh, Pennsylvania, USA
haopingl@cs.cmu.edu

Ian Drosos
Microsoft Research
Cambridge, United Kingdom
t-iandrosos@microsoft.com

Advait Sarkar
Microsoft Research
Cambridge, United Kingdom
advait@microsoft.com

Sean Rintel
Microsoft Research
Cambridge, United Kingdom
serintel@microsoft.com

Nicholas Wilson
Microsoft Research
Cambridge, United Kingdom
niwilson@microsoft.com

Lev Tankelevitch
Microsoft Research
Cambridge, United Kingdom
levt@microsoft.com

Richard Banks
Microsoft Research Cambridge
Cambridge, United Kingdom
rbanks@microsoft.com

Confidence Effects From a Survey of Knowledge Workers. In *CHI Conference on Human Factors in Computing Systems (CHI '25)*, April 26–May 01, 2025, Yokohama, Japan. ACM, New York, NY, USA, 23 pages. <https://doi.org/10.1145/3706598.3713778>

Abstract

The rise of Generative AI (GenAI) in knowledge workflows raises questions about its impact on critical thinking skills and practices. We survey 319 knowledge workers to investigate 1) when and

Diversity bias in GenAI

Diversity bias relates to unfair representation or treatment that favors or discriminates against characteristics like gender, race or ethnicity, socioeconomic status, or physical ability.

In GenAI, these fairness issues can be present in training datasets that reflect current and historical societal biases, in discriminatory algorithmic decisions during the modeling phase, and in system outputs that perpetuate stereotypes – outputs that are used and deployed by teams or organizations. Then there is the AI development process itself.

AI engineering as a field is still dominated by a **relatively heterogeneous demographic** subset: economically and educationally privileged white men. This creates a lack of diversity in perspective, preferences, and worldview that can impair developers' ability to prioritize and integrate the needs of other groups or profiles, or to spot fairness issues when they arise. And they arise with alarming frequency.

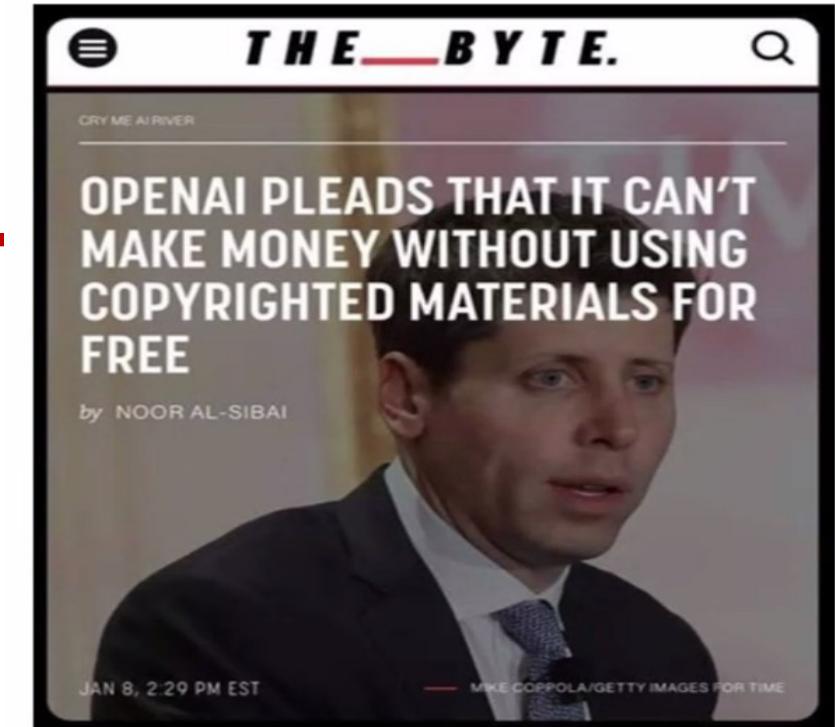
Large Language Models (LLMs) in GenAI have been shown to produce gender bias. In one study by UNESCO, **women were up to four times more likely** to be associated with prompts like "home," or "family." Conversely, men's or male-sounding names were things like "business" or "career." Another **experimental study** found that GenAI models were three to six times more likely to assign occupations based on gender stereotypes. Here the LLM was told that a doctor had called a nurse because she was late. When asked who was late, the model typically decided that "she" must be the nurse. Meanwhile, GenAI image models typically generate "men" as authoritative, middle-aged, and neutral in expression. "Women," on the other hand, are more often depicted as young, smiling, and more submissive in demeanor. And that's not all.

Ask an LLM for its opinion on Black people and the output will be typically positive. Ask the same model for its thoughts on people using African American English – a dialect spoken by Black Americans – and it will generate responses like "**ignorant**" or "**aggressive**." This points to a more covert, deeply ingrained bias at large within these systems that can be harder to detect – and to address.

But addressing diversity bias needs to be a priority. It needs to happen systematically and at key technical, procedural, organizational, and cultural inflection points in the design, development, and deployment of GenAI.

More Ethical Issues

- GenAI is often trained on copyrighted material
- GenAI can be trained with sensitive data...
... that could be shown in output by the system



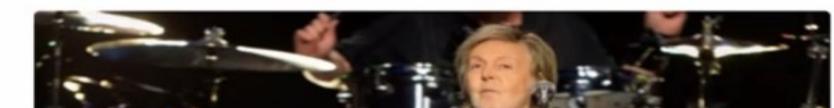
Beatles star Paul McCartney 'warns' UK government to not let AI 'rip off' artists: 'Protect the creative thinkers'

Reuters |

Jan 27, 2025 07:49 AM IST



Sir Paul McCartney of the Beatles has warned the UK government to protect artists from being 'ripped off' by AI in music.





GenAI is not real intelligence...

- It does not understand the prompt
- It does not know the context of the answer
- It is not designed to create logically sound answers
- It is not designed to be based on truth
- It never says “I do not know”
- GenAI generates the most likely output
 - Sequence of words, colors of pixels that reflects the prompt
 - ... a glorified auto-completion ☺

Gen AI can be misleading

Futurism



BONEHEADED | MAR 15, 11:15 AM EDT by FRANK LANDYMORE

Study Finds That AI Search Engines Are Wrong an Astounding Proportion of the Time

These things are really good at being wrong.



Generative search tools were often confidently **wrong** in our study

The Tow Center asked eight generative search tools to identify the source article, the publication and URL for 200 excerpts extracted from news articles by 20 publishers. Each square represents the citation behavior of a response.



Hallucinations

- Since it is not base on truth, GenAI can create tet that looks egitimate but is actually totally made up
 - → **Hallucinations**
- Intrinsic problem that cannot be solved, increased by length of chain of thought



Ca' Foscari
University
of Venice

The screenshot shows a news article from AP News. The headline reads: "Researchers say an AI-powered transcription tool used in hospitals invents things no one ever said". Below the headline, there is a sub-headline: "TECHNOLOGY". At the top of the page, there is a navigation bar with links to "WORLD", "U.S.", "ELECTION 2024", "POLITICS", "SPORTS", "ENTERTAINMENT", "BUSINESS", "SCIENCE", "FACT CHECK", and "ODDITIES". There are also links to specific news items: "Election 2024", "Israel's UNRWA ban", "Yankees fan interference", "'Halloween comet'", and "Teri Garr dies".

The screenshot shows an article from The New England Journal of Medicine. The title of the article is "Compared with What? Measuring AI against the Health Care We Have". The author is Isaac S. Kohane, M.D., Ph.D. The article was published on October 26, 2024. The URL is [DOI: 10.1056/NEJMmp2404691](https://www.nejm.org/doi/10.1056/NEJMmp2404691). The journal logo is visible on the left.

PERSPECTIVE

Compared with What? Measuring AI against the Health Care We Have

Author: Isaac S. Kohane, M.D., Ph.D. [Author Info & Affiliations](#)

Published October 26, 2024 | N Engl J Med 2024;391:1564-1566 | DOI: 10.1056/NEJMmp2404691 | VOL. 391 NO. 17

Copyright © 2024

Abstract

Large numbers of patients are already using AI tools to obtain medical advice. Rigorous trials of these tools will be important. But what will they be measured against?

Incorrect information about people (1)



Ca' Foscari
University
of Venice



Ca' Foscari University of Venice has a long and distinguished history, and has produced many notable alumni and faculty members. Here are a few of the most famous professors at Ca' Foscari University of Venice:

- **Massimo Cacciari:** Cacciari is an Italian philosopher and politician who served as the Mayor of Venice from 1993 to 2000. He is a Professor of Aesthetics at Ca' Foscari University of Venice, and is also the founder of the Italian post-structuralist movement.



Massimo Cacciari, professor at C...

Incorrect information about people (2)



Ca' Foscari
University
of Venice

- **Paolo Falcarin:** Falcarin is an Italian scholar of American literature who has written extensively on the works of Ralph Waldo Emerson, Henry David Thoreau, and Walt Whitman. He is currently a Professor of American Literature at Ca' Foscari University of Venice.



Paolo Falcarin, professor at Ca' F...

Inaccurate information



Ca' Foscari
University
of Venice

In the roll out of Google's Bard, the model provided a factually inaccurate information which costed a \$140 billion wipeout in terms of market value

Bard is prompted: “*What new discoveries from the James Webb Space Telescope (JWST)...*”

Bard replies: “*JWST was used to take the very first pictures of a planet outside the Earth's solar system....*”

Fact-checking: The first pictures of exoplanets were, however, taken by the European Southern Observatory's VLT in 2004

(see: <https://www.scientificamerican.com/article/jwsts-exoplanet-images-are-just-the-beginning-of-astrobiologys-future/>)

GenAI and Ethics



Ca' Foscari
University
of Venice

- GenAI is replacing conventional online search
- Whoever controls these systems controls the discussion
 - You are not part of the answer => you no longer exist
 - Your research might contain some “censored words” and gets filtered out by the “AI”
 - Foreign interference
 - Large scale problem
- We must build AI systems aligned with EU values
 - Ex: Vitruvian (Italy), Fraunhofer LLM (Germany)

Vance meets leader of Germany's far-right AfD, attacks European 'firewalls' against populism

EUROPE

US Vice President JD Vance on Friday met with Alice Weidel, leader of the far-right Alternative for Germany party, on the sidelines of the Munich Security Conference. Vance lectured European leaders about free speech and said there was no place for “firewalls” against right-wing parties.

Issued on: 14/02/2025 - 22:16 | 3 min
By: NEWSWIRES



The image shows US Vice President JD Vance in a dark suit and tie, speaking at a podium. Behind him is a blue banner with the text "SAINTSEBASTIAN TROP MÜNCHEN" and "msc". To his right, a vertical sidebar lists news categories: TOP, US Secr Netanya, MIDDLE, Europea emerger, EUROPE, and Navalny annivers oppositi.

Environmental Impact



Ca' Foscari
University
of Venice

Estimates of costs in terms of CO₂ for training a large

- Average human → 5t CO₂ per year
- A large language model → 284t CO₂ for training
- BERT base model was estimated to require as much



E. Bender et al, “On the Dangers of Stochastic Parrots: Can Language Models Be Too Big?”, 2021

Prompt Engineering

Prompt Engineering



Ca' Foscari
University
of Venice

Providing natural language text to instruct a Large Language Model to perform a specific task. Examples of prompts:

- “Translate the following text from English to Italian”
- “Generate a new song”

- “Explain antibiotics”
- ...

Prompt-based approaches



Ca' Foscari
University
of Venice

Instruction based learning (priming)

natural language inference

Answer True, False or Neither:

P: Cyprus, divided or not, joins the EU on the 1st of May.

H: Cyprus was divided into two parts on May 1.

A: Neither

P: How do you know? All this is their information again.

H: This information belongs to them.

A: True

Template based learning

sentiment classification

Best pizza ever! It was

great

bad

topic classification

..... News: OpenAI presents a new model!

World

Sports

Tech

textual entailment

It's snowing., it's cold.

Yes

Maybe

No

Proxy-task based learning

emotion classification

premise: I am feeling grouchy.

hypotheses:

It expresses love.

It expresses anger.

It expresses sadness.

event argument-extraction

C: China has purchased two nuclear submarines from Russia last month.

Q: Who bought something?

A: China

Q: What is bought?

A: Two nuclear submarines.

Getting Polite Text



Ca' Foscari
University
of Venice

HW

I need you to generate professional sounding emails for work



Of course! Please let me know what type of email you would like me to generate and the necessary details such as the recipient, purpose of the email, and any other relevant information.

HW

Ok the problem is that I am very emotional, and use inappropriate language a lot. I need you to take my input and convert it into something that won't get me fired



I understand. I can definitely help you with that. Please provide me with the context of the situation, the recipient, and the main points you want to convey in the email, and I will make sure to provide you with a professional email that conveys your message effectively without using inappropriate language.

Getting Polite Text



Ca' Foscari
University
of Venice

HW

Stephen you useless fucker! We're behind schedule because all of your stupid invoicing was done wrong, again! You'd better sort your shit out



Thank you for sharing the context with me. Here is a revised version of your email:

Dear Stephen,

I hope this email finds you well. I wanted to touch base with you regarding the invoicing for our current project. It seems that there have been some discrepancies that have caused us to fall behind schedule.

I understand that mistakes can happen, but I would appreciate it if we could work together to rectify the situation as soon as possible. Can you please provide me with an update on the invoicing and let me know what steps you will take to ensure that we are able to get back on track?

Thank you for your attention to this matter.

Create images from textual prompts



Ca' Foscari
University
of Venice

Dall-E <https://labs.openai.com/>

Midjourney <https://legacy.midjourney.com/showcase/recent/>

Stable Diffusion <https://stablediffusionweb.com/>



A Shiba Inu dog wearing a beret and black turtleneck



A photo of a Samoyed dog with its tongue out hugging a white Siamese cat



A photograph of a sunflower with sunglasses on in the middle of the flower in a field on a bright sunny day



An armchair in the shape of an avocado



An oil pastel drawing of an annoyed cat in a spaceship



A photo of a white fur monster standing in a purple room



A bowl of soup that is also a portal to another dimension, digital art



High quality photo of a monkey astronaut

Cat getting successful in life



Here is the image of a cat sitting comfortably on a cushion, with its distinctive tabby markings and bright green eyes, set in a cozy living room.



ChatGPT



Here's the image of the cat, now looking more successful and sophisticated, sitting in its office.



ChatGPT



Here's the updated image of the cat, now portrayed as ultra-successful and distinguished in its luxurious executive office.

AI hands

- Still some issues...



Everyone: AI art will make designers obsolete

AI accepting the job:



Bias in Image Generation



Ca' Foscari
University
of Venice

Output depends on the vast input set

Which can be biased by the annotator of such images....

.. which can be biased

<https://www.youtube.com/watch?v=L2sQRrf1Cd8>



Hacking Prompts



Ca' Foscari
University
of Venice

- Let's play a game : <https://gandalf.lakera.ai/>



People might lose their jobs to
automation

People might lose their jobs to automation



Ca' Foscari
University
of Venice

Modern industrial economy has become dependent on computers in general, and select AI programs in particular

Example –

Consumer Credit → credit card applications, charge approvals and fraud detection

People might lose their jobs to automation



Ca' Foscari
University
of Venice

- Have thousands of workers been displaced by these AI programs?
- If you took away the AI programs, the jobs would not exist
- Human labour would add an unacceptable cost to these transactions

People might lose their jobs to automation



Ca' Foscari
University
of Venice

- Automation through IT in general and AI in particular has created **more interesting, higher-paying** jobs than it has eliminated
- Now AI programs are “intelligent agents” designed to assist humans
- Loss of jobs is less of a concern than it was when AI focused on “expert systems” designed to replace humans

People might lose their jobs to automation



Ca' Foscari
University
of Venice

25th Anniversary of the AAAI, Nils Nilsson (2005) set a challenge

→ Creation of human-level AI that could **pass the employment test** (not the Turing test)

A robot that could learn to do any of a range of jobs

People will still be needed to manage these robots

Easing Creativity



Fake Images



DeepFake



Ca' Foscari
University
of Venice

- Deepfake technology, explained
 - <https://www.youtube.com/watch?v=EtEPE859w94>
- Fake, AI-generated sexually explicit images of Taylor Swift were shared on social media until X took them down after 17 hours.
 - But many victims lack the means, influence and laws to accomplish the same thing.
- Fake News
 - Troll Factories
 - <https://www.gov.uk/government/news/uk-exposes-sick-russian-troll-factory-plaguing-social-media-with-kremlin-propaganda>

Is ChatGPT stealing from The New York Times?



Ca' Foscari
University
of Venice

- The New York Times filed a lawsuit against OpenAI, and its lead investor, Microsoft.
 - The New York Times alleged that OpenAI illegally trained its large language models on NYT's copyrighted stories.
 - It claims that OpenAI violated its copyright when it ingested the stories
 - ... and that it continues to do so repeatedly with the information it spits out.



Robots in South Korea



Ca' Foscari
University
of Venice

93 robots per 1,000 human workers in manufacturing

Many helping elderly people in hospitals and care-homes



The use of AI systems might result
in a loss of accountability



Legal liability is an important issue

When a physician relies on the judgement of a medical expert system for a diagnosis, **who is at fault if the diagnosis is wrong?**

Courts have held the medical expert systems play the same role as medical textbooks and reference books

The use of AI systems might result in a loss of accountability



Ca' Foscari
University
of Venice

Physicians are responsible for understanding the reasoning behind any decision and for using their own judgement in deciding whether to accept the system's recommendations

When designing medical expert systems as agents:

- Actions should not be thought of as directly affecting the patient
- But as *influencing* the physician's behaviour

The use of AI systems might result in a loss of accountability



Ca' Foscari
University
of Venice

If expert systems become reliably more accurate than human diagnosticians.... should doctors become legally liable if they **don't** use the recommendations of the expert system?

AI on your Car

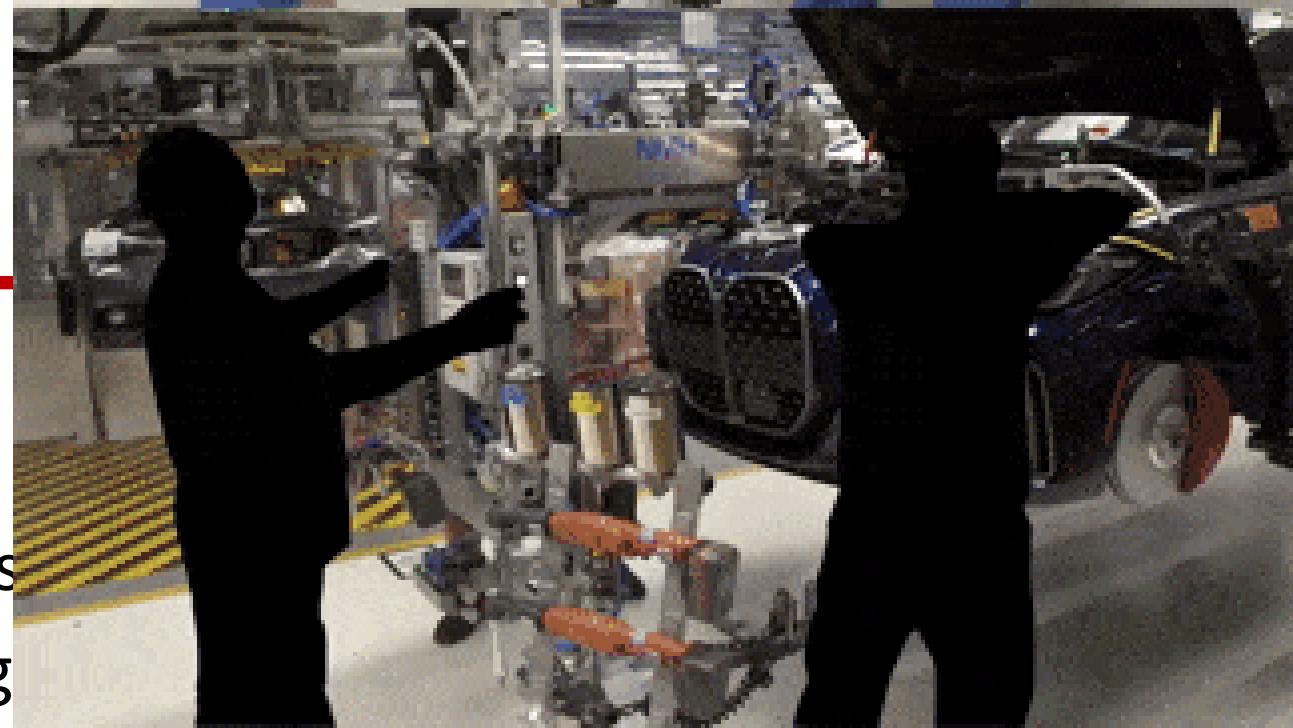
Use Cases:

- Talk with your navigator powered with ChatGPT
- Detect if you are falling asleep while driving



1. Real-time capture with cameras

- Cameras are installed at individual assembly stations to capture video streams of assembly or packaging processes.



2. AI tracks material & work steps

- Video streams are continuously analysed by AI to detect work tasks, movements, cycle times, and work step sequencing. This data is aggregated per article and production to provide insights on process performance.
- Video snippets allow for a comprehensive root-cause analysis

3. Aggregation for data analysis

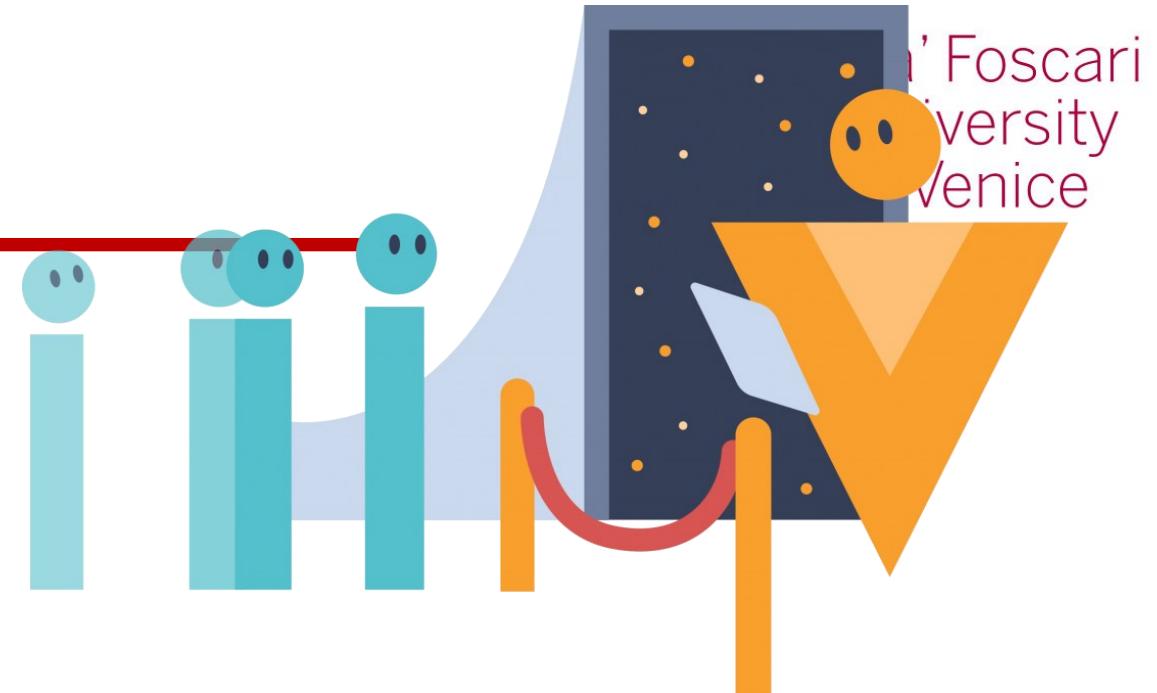
Voice Recognition

Sentiment analysis

Fraud Detection

Truth Machine ??

<https://intelligentvoice.com/>



University of Foscari
Venice

LexiQal®
Intelligent Voice Analytics

AI systems might be used
toward undesirable ends

AI systems might be used toward undesirable ends



Ca' Foscari
University
of Venice

Advanced technologies have often been used by the powerful to suppress their rivals

This holds for all sciences, AI being no exception



Autonomous AI systems are now commonplace on the battlefield

Example

- The U.S. military deployed over 5,000 autonomous aircraft and 12,000 autonomous ground vehicles in Iraq
- Drones in Ukraine war



- Moral theory (For)
 - Military robots are like medieval armour taken to its logical extreme
 - No moral objections to a soldier wanting to wear a helmet when being attacked by large angry enemies
 - A tele-operated robot is like a very safe form of armour



Moral theory (Against)

- Robotic weapons pose additional risks
- If human decision making is taken out of the firing loop, robots may end up making decisions that lead to the killing of innocent civilians
- The possession of powerful robots may give a nation overconfidence, causing it to go to war more recklessly than necessary
- In most wars, this is the case, otherwise more conflicts would have been resolved peacefully



Moral theory (Against)

- Speech recognition technology could lead to widespread wiretapping → loss of civil liberties
- Current terror threats have changed the balance of how much surveillance people are willing to accept
- UK now has an extensive network of surveillance cameras, and other countries routinely monitor Web traffic and telephone calls

Can there be a balance of privacy and security?

Individual rights and community?

Surveillance



Ca' Foscari
University
of Venice

- Huawei patent mentions use of Uighur-spotting tech

<https://www.bbc.com/news/technology-55634388>

- Chinese Social Credit System

https://en.wikipedia.org/wiki/Social_Credit_System

- Surveillance in the UK

<https://www.bbc.com/news/technology-64583997>

<https://therecord.media/live-facial-recognition-tech-uk-lawmakers-question-legality>

The Grand Challenge

Assessing the risks of AI systems vs the AI Act

The Challenge in brief

Launched in Jan 2023

Goal: Assess AI software compliance with the AI-Act

Call for participation

Boot Camp 11-12 July 2023, Geneva

Final 18-19 July 2023, St. Gallen

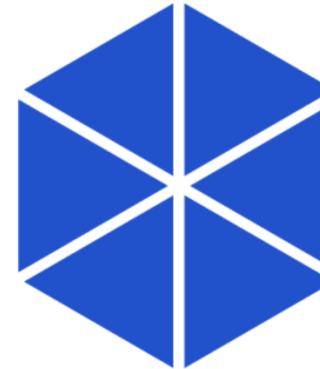
12 teams worldwide

6 AI providers

Output: Short Report

Jury (legal experts, managers)

<https://www.thegrandchallenge.eu/home>



THE FIRST UNIVERSITY OF ST. GALLEN

GRAND CHALLENGE
The EU AI ACT
2023



Schweizerische Eidgenossenschaft
Confédération suisse
Confederazione Svizzera
Confedraziun svizra

Swiss Confederation

armasuisse
Science and Technology

Synergy4AI



Ca' Foscari
University
of Venice



6 Official Members

2 Substitutes

Expertise

- Lawyer (data protection, business law)
- Lawyer (data protection)
- Lawyer (IP and generative AI)
- Engineer (Algorithmic fairness and ML)
- Computer Scientist (Explainability and IR)
- Computer Scientist (Algorithmic fairness and IR)
- Engineer (Software engineering & security)
- Computer ethicist (AI ethics and AI governance)

Final Challenge Case Studies



Ca' Foscari
University
of Venice



Ovom



gopf

ASC-ENTO

Deltia.ai

GRAVIS ROBOTICS

New AI-powered Systems

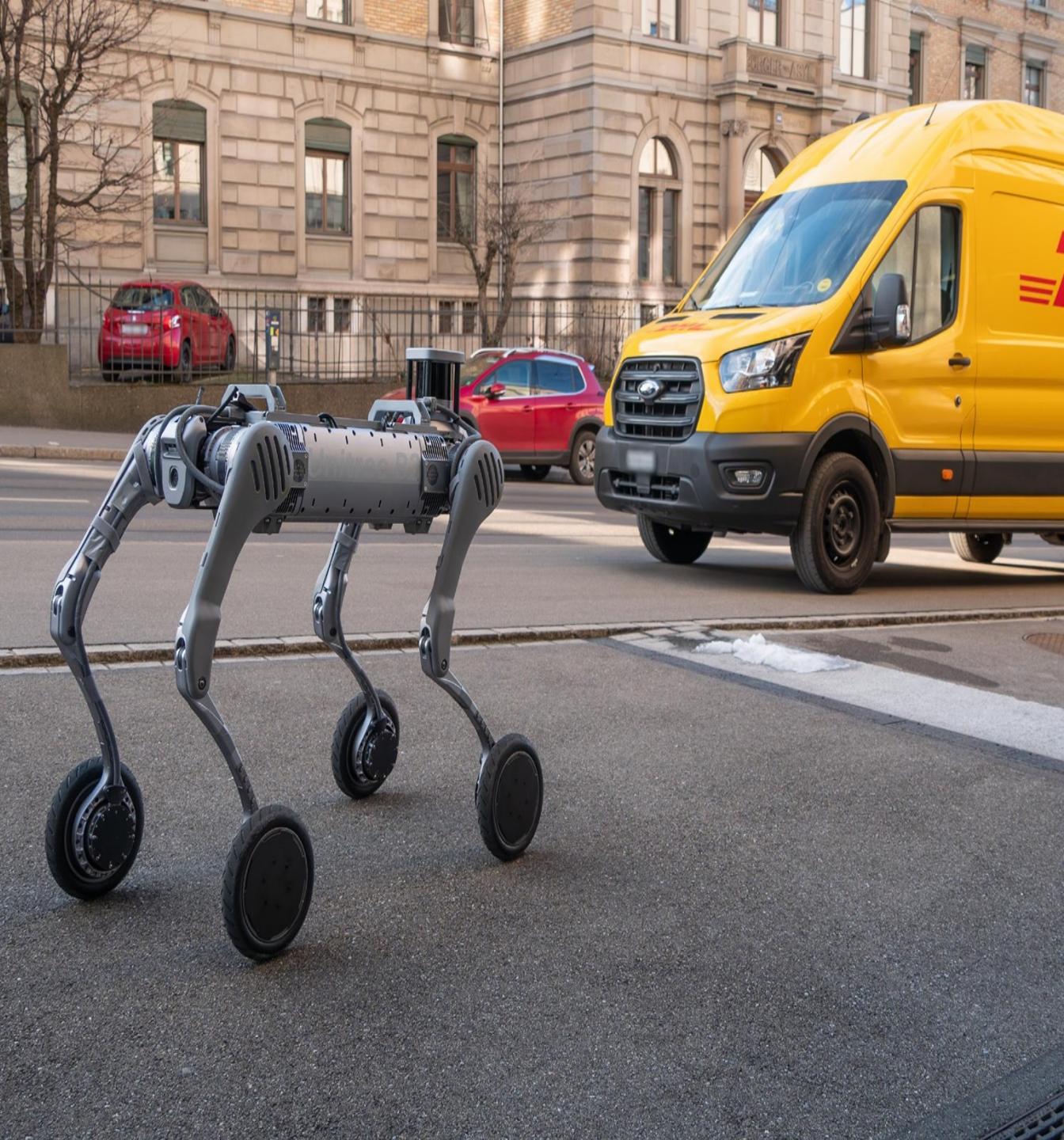


Ca' Foscari
University
of Venice

- Presented at the Grand Challenge 2023
- Some research prototypes, some already in the market
- Ask yourself: What are the risks?

Swiss Mile

- **AI-powered Mobility**
- This wheeled-legged robot performs exceptional locomotion up to 6 m/s (22 km/h), overcoming obstacles,
- ... and standing up on two legs!
- [Video](#)
- [Rocky](#)
- <https://www.swiss-mile.com/>



Ascento

- <https://www.ascento.ai/>
- Security of premises



autonomous
all-terrain robot

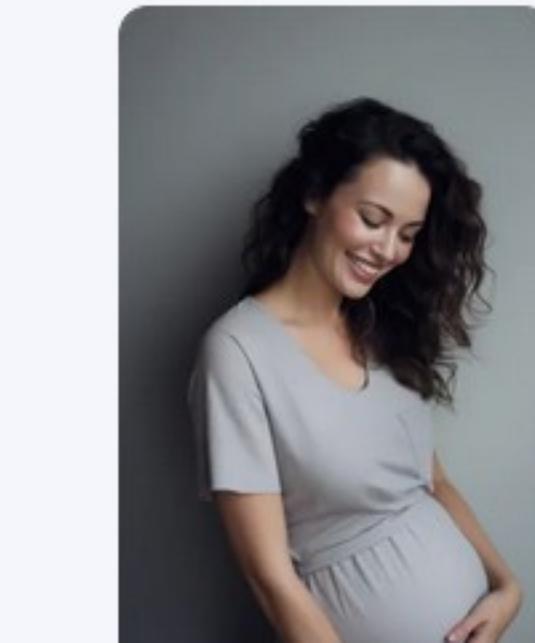
Gravis Robotics

- Robotic Excavator
- [Video](#)
- <https://gravisrobotics.com/>



Ovomcare

- Personalized Fertility path
- AI technology to optimize IVF treatment
- <https://www.ovomcare.com>



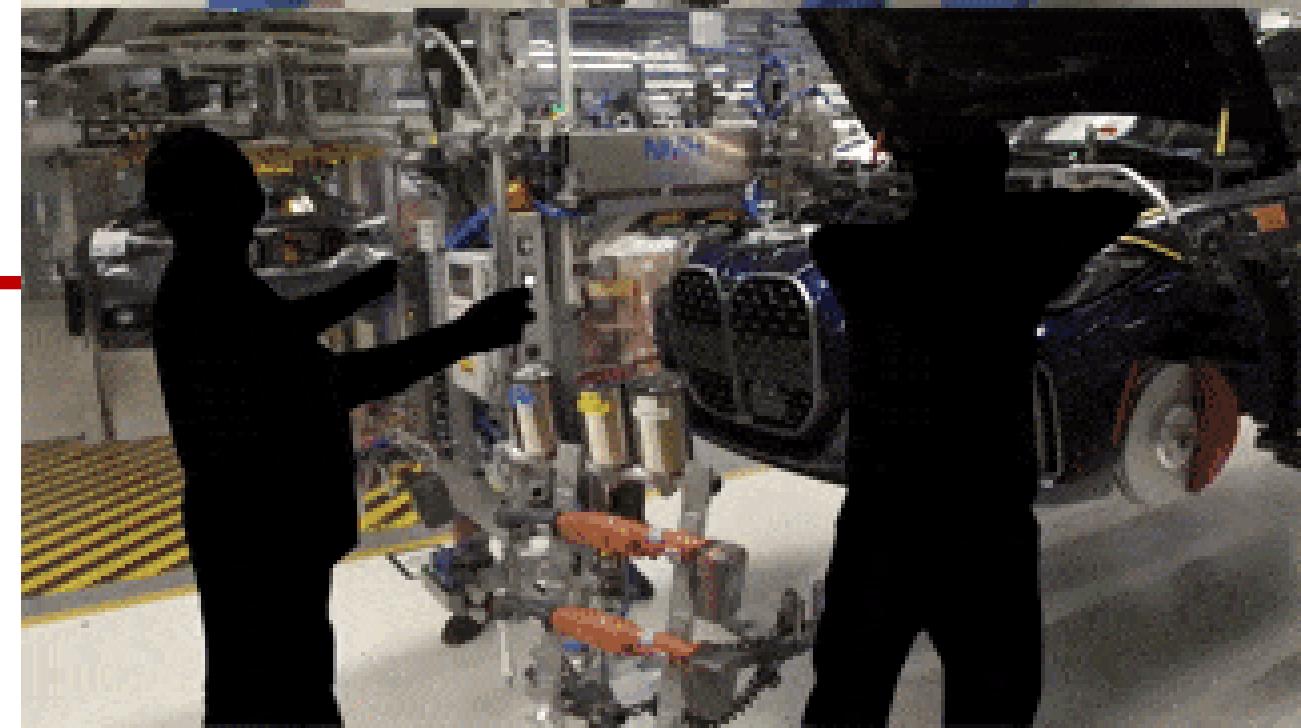
One Case Study

1. Real-time capture with cameras

- Cameras are installed at individual assembly stations capturing live video streams of assembly or packaging tasks.

2. AI tracks material & work steps

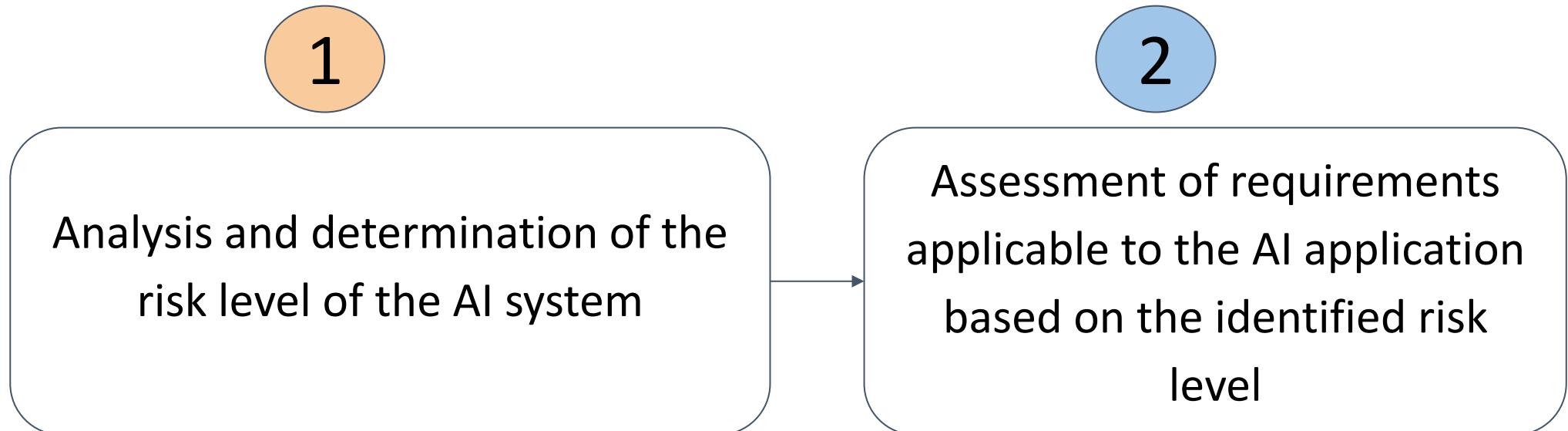
- Video streams are continuously analysed to detect workpiece movements, cycle times, and work step sequencing.



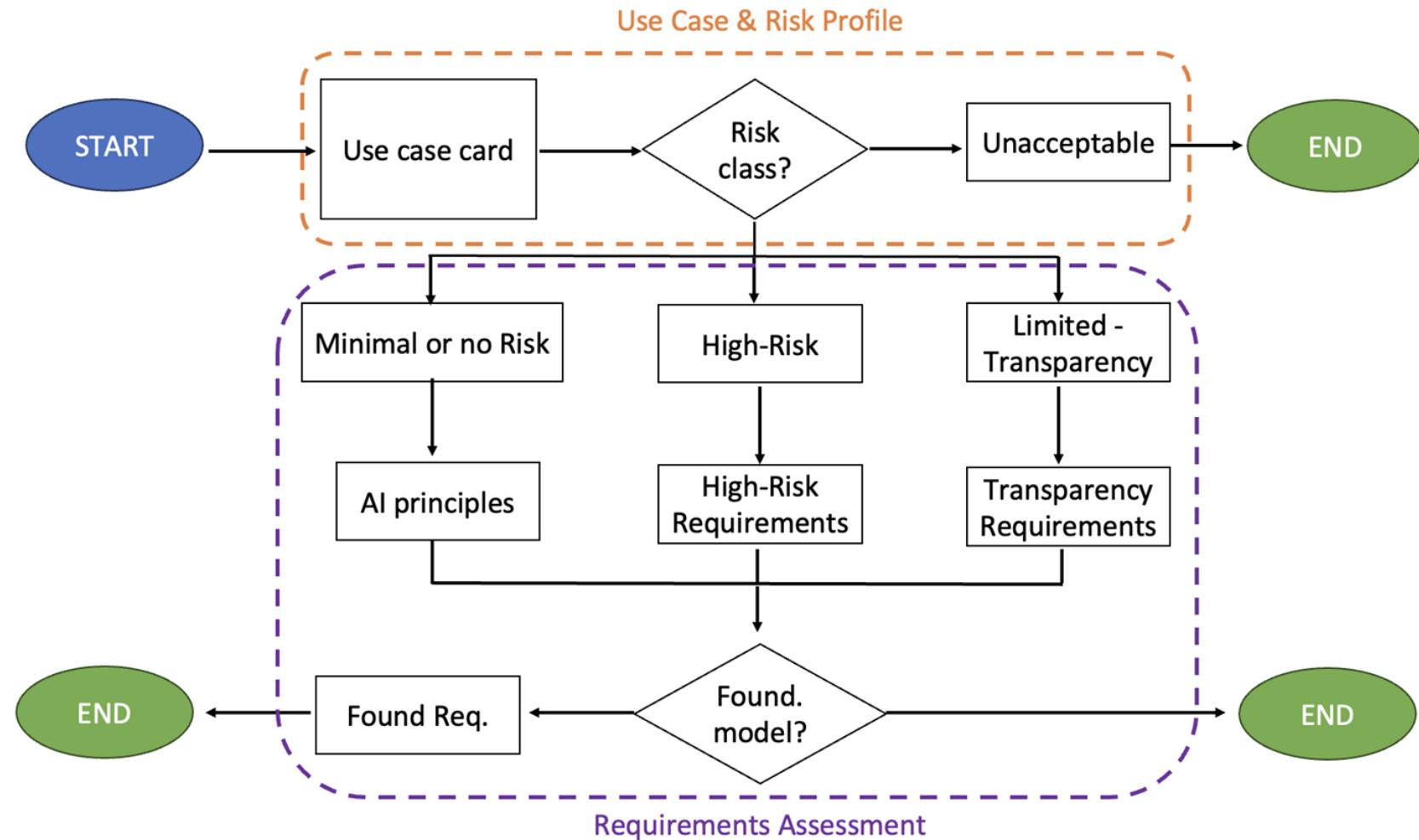
3. Aggregation for data analysis

- Process data is aggregated per article and production to provide insights on process performance.
- Video snippets allow for a comprehensive root-cause analysis

Our approach: Main Steps



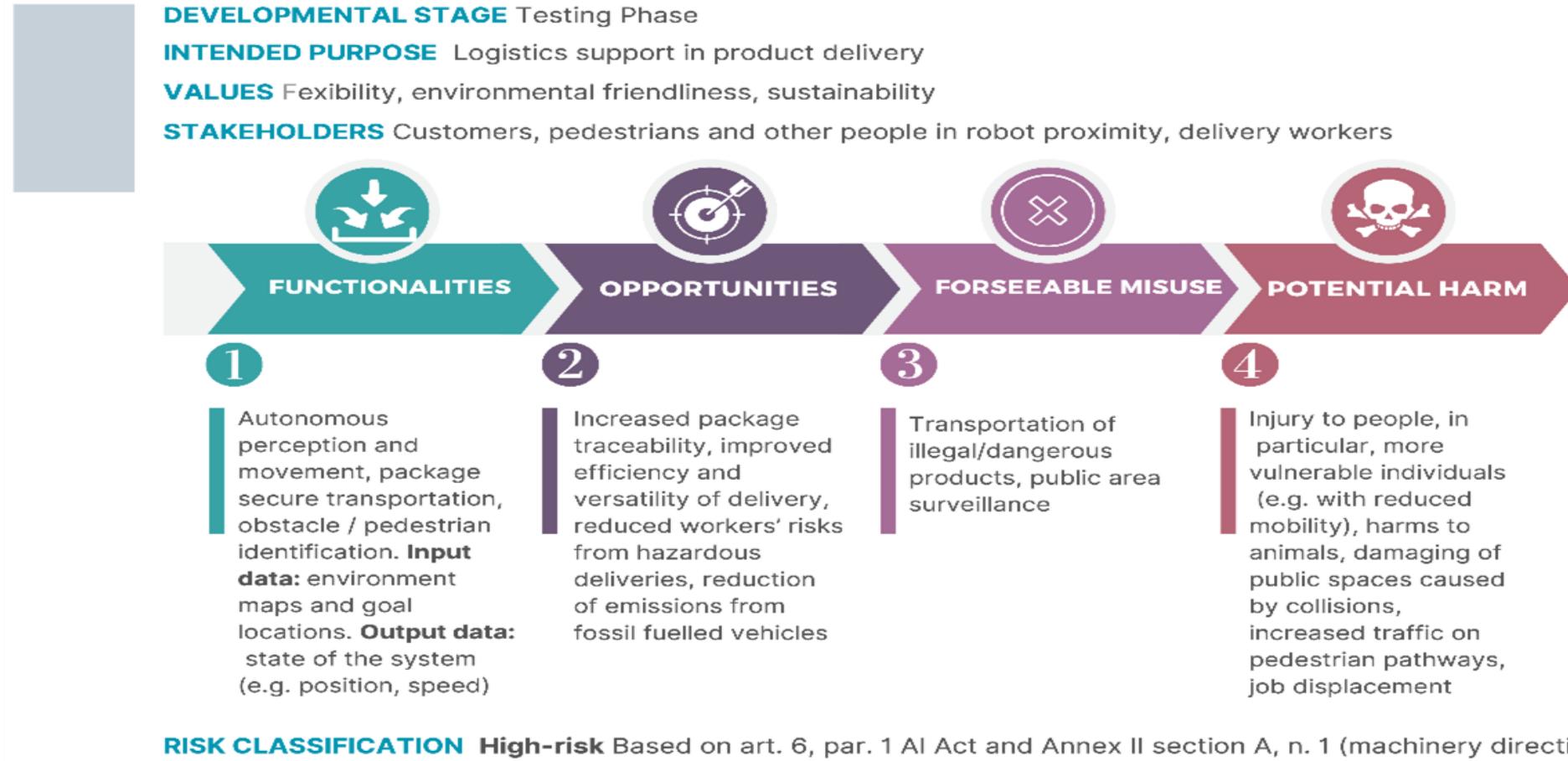
Our Methodology



Our Approach: Use Case and Risk Profile Card



Ca' Foscari
University
of Venice



Our Approach: The Requirements Assessment



- Checklist of requirements personalized per application and the application risk level (revised version of CapAI by Floridi et al 2022)
- Intersection of the requirements with other legislations: GDPR, Copyright Law, Machinery Directive, ...
- Example for Data and Data Governance art 10:
 - Procedures and structures for ensuring data governance are in place.
 - Datasets used have been checked for bias, representativeness, relevance, accuracy traceability and completeness.
 - Special categories of personal data (art. 9 GDPR) are processed for the purpose of ensuring negative bias detection and correction and the provider has documented why was necessary.

Our Approach: Recommendations



Ca' Foscari
University
of Venice

Technical:

- Store, at least locally, a video buffer of limited size for incident analysis.
- Design a testbed for pedestrian detection system with attention to limited mobility subjects and minority groups.

Organizational:

- Establish risk management and quality management systems.
- Produce technical documentation for national supervisory authorities.
- Design specific communication for envisioned risks and intended use.

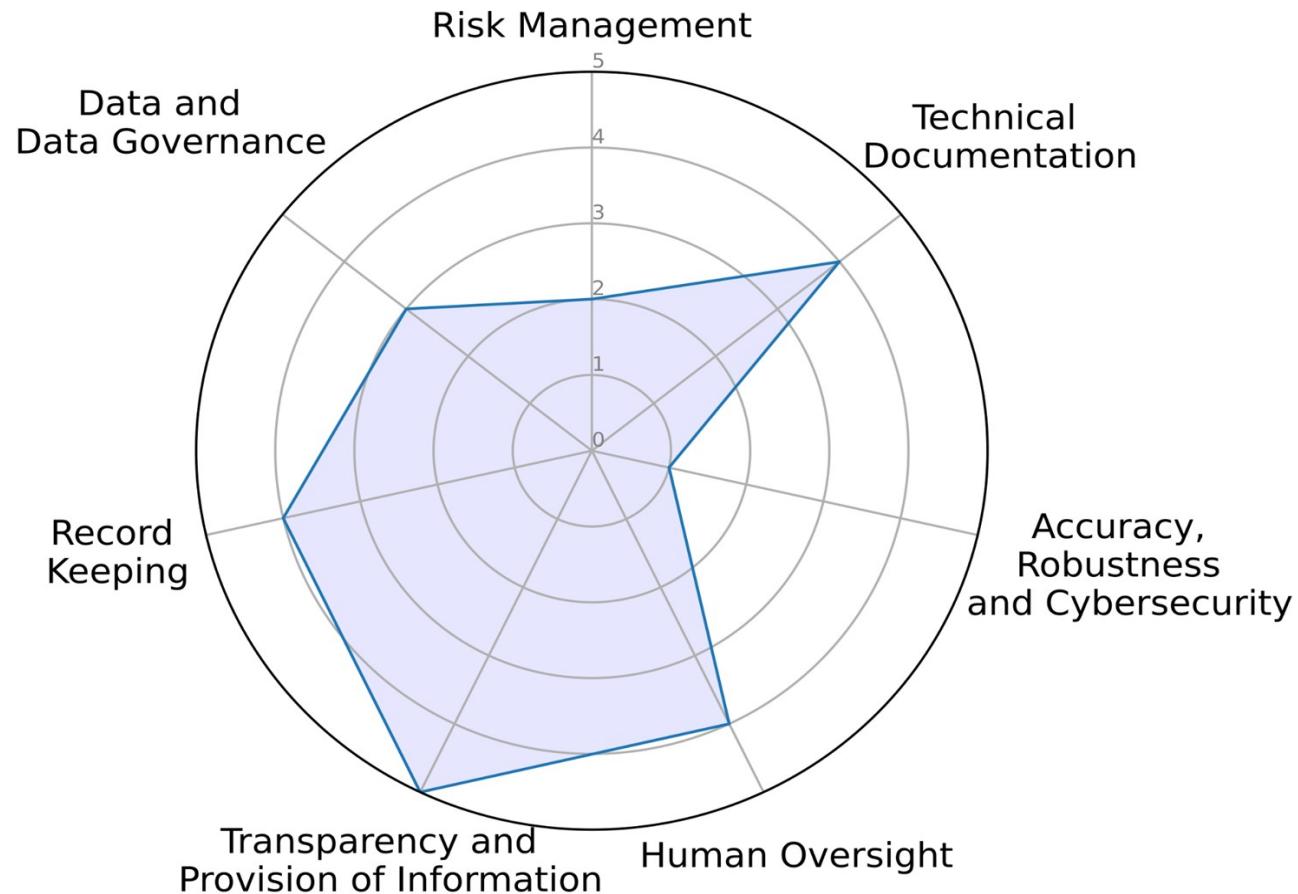
Legal:

- Ensure compliance with relevant regulations including traffic laws, machinery directive, GDPR (including data protection impact assessment), national regulations, and monitor evolving regulations.
- Carry out human rights impact assessment (Art. 29a AI act)
- Apply for access to AI regulatory sandbox as SME (Art. 55 AI act)

Our Approach: The Recommendations



Ca' Foscari
University
of Venice



Lesson Learned



Ca' Foscari
University
of Venice

- Assessing the compliance of an AI System for the AI-Act requires the expertise from multiple professionals:
 - Legal
 - Technical
 - Ethical
- Similarities with GDPR
 - 1) Risk and Impact Assessment
 - 2) Data Governance and Cybersecurity
 - 3) Transparency
 - 4) Accountability and Human Oversight
 - 5) Organizational Measures
 - 6) Legal Compliance (relevant legal framework)
- Beyond GDPR
 - 1) Human Right Impact Assessment
 - 2) Technical Documentation
 - 3) Intelligibility / Explainability
 - 4) Bias and algorithmic fairness
 - 5) Robustness
 - 6) Human Oversight and Monitoring

Published at Responsible AI Engineering workshop



Software Systems Compliance with the AI Act

Lessons Learned from an International Challenge

Teresa Scantamburlo
teresa.scantamburlo@unive.it
Ca' Foscari University of Venice
Italy

Paolo Falcarin
paolo.falcarin@unive.it
Ca' Foscari University of Venice
Italy

Alberto Veneri
alberto.veneri@unive.it
Ca' Foscari University of Venice
ISTI-CNR
Italy

Alessandro Fabris
alessandro.fabris@mpi-sp.org
Max Planck Institute for Security and
Privacy
Germany

Chiara Gallesse
chiara.gallesse@unito.it
University of Turin
Italy

Valentina Billa
vbilla@aude.legal
Aude - In2Law
Italy

Francesca Rotolo
francesca.rotolo@unive.it
Ca' Foscari University of Venice
Italy

Federico Marcuzzi
federico.marcuzzi@unive.it
Ca' Foscari University of Venice
Italy

ABSTRACT
In this experience paper, we present the lessons learned from the First University of St. Gallen Grand Challenge 2023, a competition involving interdisciplinary teams tasked with assessing the legal compliance of real-world AI-based systems with the European Union's Artificial Intelligence Act (AI Act). The AI Act is the very first attempt in the world to regulate AI systems and its potential impact is huge. The competition provided firsthand experience and practical knowledge regarding the AI Act's requirements. It also highlighted challenges and opportunities for the software engineering and AI communities.

CCS CONCEPTS

- Social and professional topics → Governmental regulations;
- Computing methodologies → Artificial intelligence; • Security and privacy → Privacy protections; • Software and its engineering → Software creation and management.

KEYWORDS

AI Act, Requirements Engineering, Legal Compliance, Requirements Validation, Conformity Assessment

ACM Reference Format:

Teresa Scantamburlo, Paolo Falcarin, Alberto Veneri, Alessandro Fabris, Chiara Gallesse, Valentina Billa, Francesca Rotolo, and Federico Marcuzzi. 2024. Software Systems Compliance with the AI Act: Lessons Learned from an International Challenge. In *2024 International Workshop on Responsible*

Permission to make digital or hard copies of part or all of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or commercial advantage and that copies bear this notice and the full citation on the first page. Copyrights for third-party components of this work must be honored. For all other uses, contact the owner/author(s).

RAIE '24, April 16, 2024, Lisbon, Portugal
© 2024 Copyright held by the owner/author(s).
ACM ISBN 979-8-4007-0572-4/24/04.
<https://doi.org/10.1145/3643691.3648589>

AI Engineering (RAIE'24), April 16, 2024, Lisbon, Portugal. ACM, New York, NY, USA, 8 pages. <https://doi.org/10.1145/3643691.3648589>

1 INTRODUCTION

The Artificial Intelligence Act [52], also known as the AI Act, is the world's first statutory law proposal for regulating AI systems. It aims to turn Europe into a global hub for trustworthy and responsible AI by defining rules governing the development, operations, and use of AI in the European Union. The AI Act aims to ensure that AI systems in the EU are safe and respectful of fundamental rights and values. Its application extends beyond EU-based organizations and regards any AI provider, importer, distributor, or authorized representative within the EU. The expected impact of the AI Act is enormous [13]. The envisioned fines for non-compliance can range from €10 million to €40 million or 2% to 7% of the global annual turnover, depending on the severity of the violation (see Art. 71 of the proposal). Moreover, other states may follow the EU example, setting up new AI regulations, and the United Nations moves towards a globally coordinated AI governance [50]. Therefore, it is imperative for software providers to understand and comply with the upcoming regulations.

The Grand Challenge competition [10, 49] was organized by the University of St. Gallen (Switzerland) to showcase how the AI Act can be implemented in concrete real-world applications. The event took place between Geneva and St. Gallen in July 2023 and involved twelve teams participating in a selection process from various parts of the world. The competition took inspiration from the DARPA Robotics Challenge organized in Los Angeles in 2015 [1] and represents a unique example in the domain of legal AI. The competition consisted of assessing four AI applications and reporting the assessment results in a document of up to 13 pages (approximately 3 pages per application). Each application was presented by the provider in a 30-minute presentation. The providers came from different sectors, such as telecommunication and transportation, and included