deberti jacchia

EU AI Act legislative tools and developments in university education

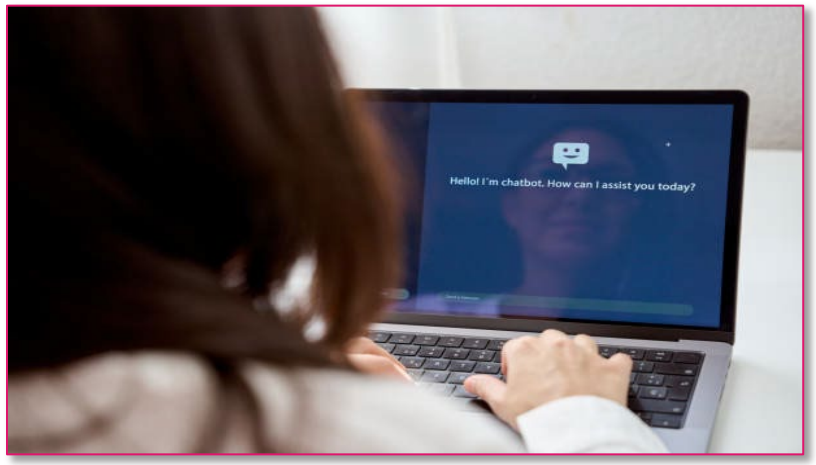
Jacopo Piemonte / Federico Aluigi

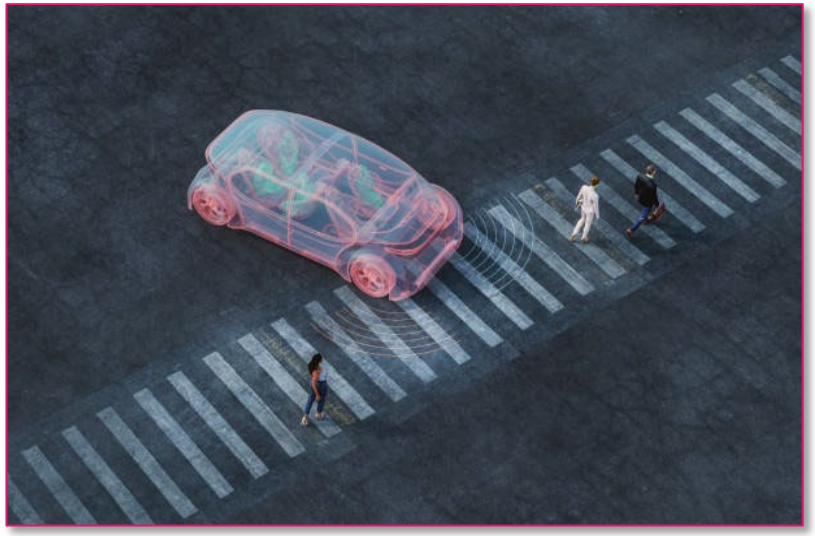
15 May 2024

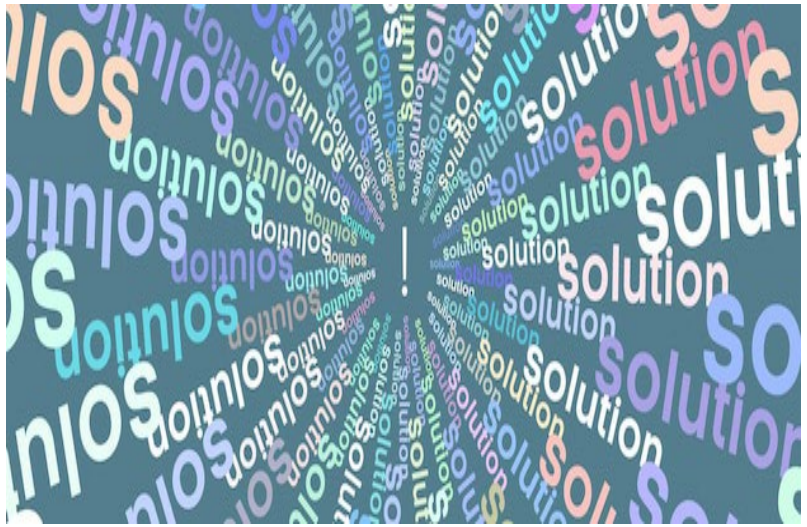
The EU AI Act Explained

1. Introduction

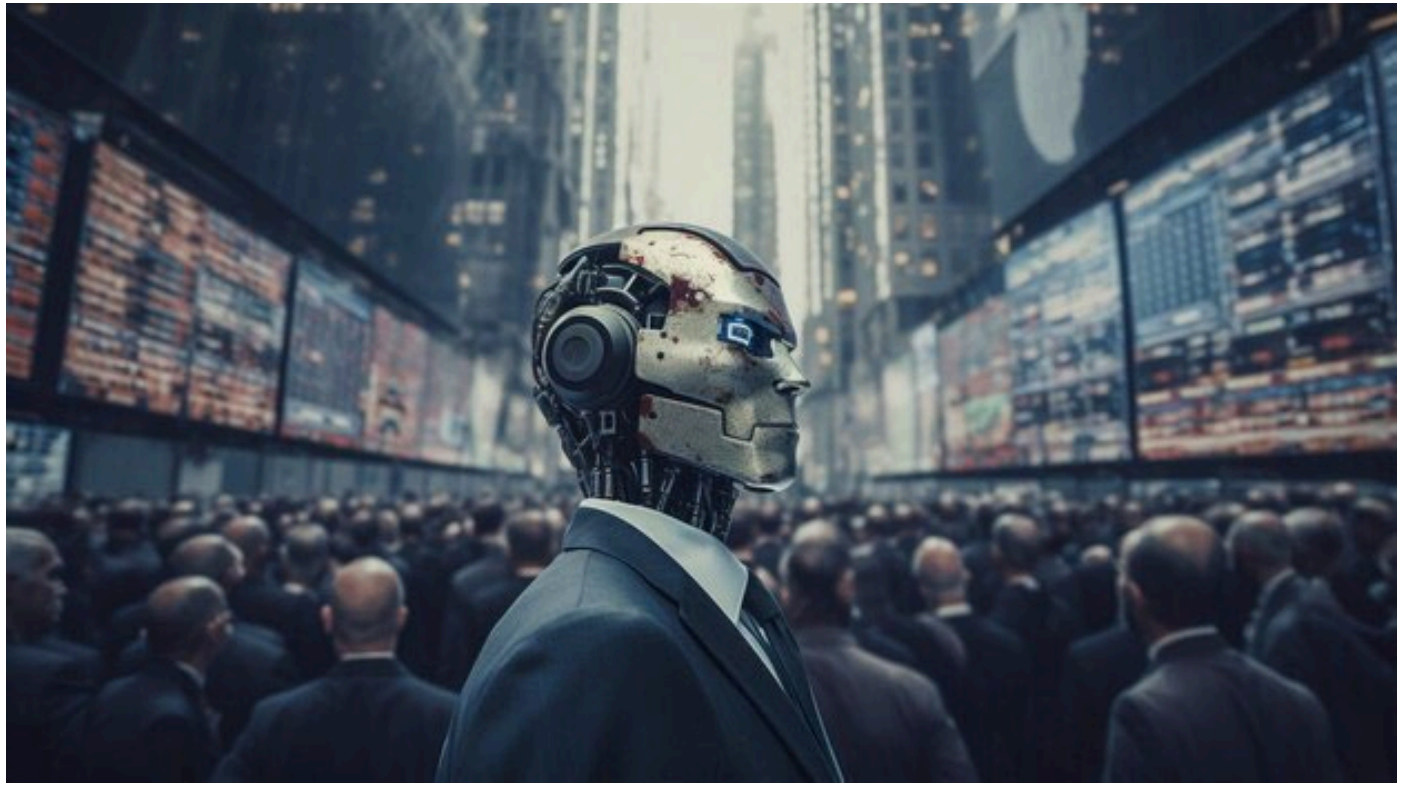
2024: THE YEAR OF ARTIFICIAL INTELLIGENCE (“AI”)?



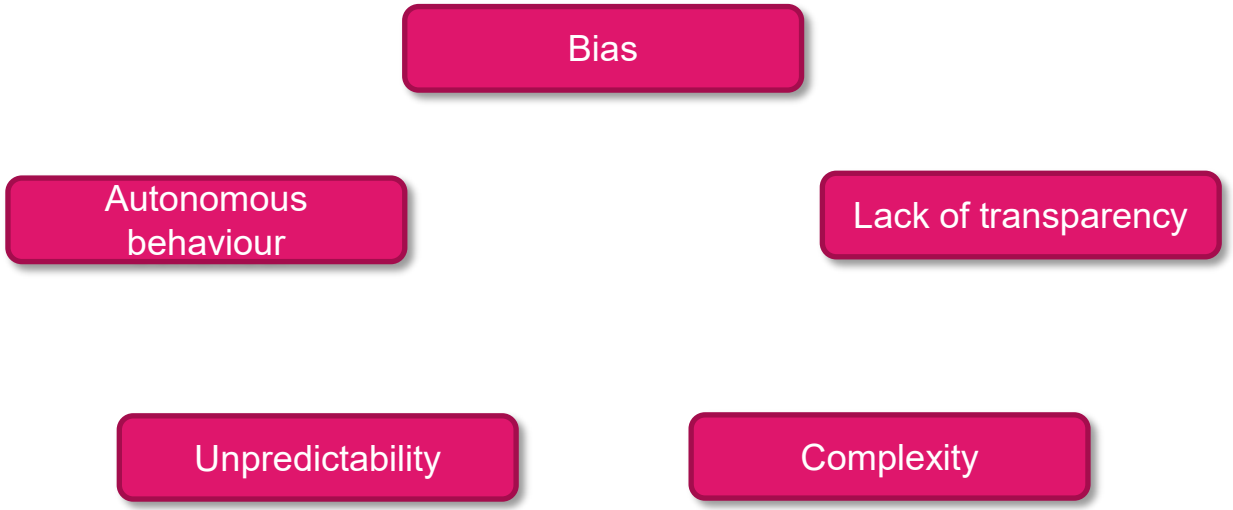




CHALLENGES OF AI (1)



CHALLENGES OF AI (2)



CHALLENGES OF AI (3)

Images generated by Midjourney in accordance to the following prompt:

- «Typical day in Italy»

Source: Horizon Security S.r.l.

CHALLENGES OF AI (4)
Deep Fake

Images generated by Midjourney

Source: Horizon Security S.r.l.

CHALLENGES OF AI (5)
Deep Fake

Images generated by Midjourney

Source: Horizon Security S.r.l.

2. The fundamentals of the EU AI ACT

EUROPEAN UNION, A PIONEER IN THE AI REGULATION

April 2021: the European Commission launches the first proposal for a regulation on AI.

March 2024: final text receives green light from the European Parliament.

NEXT STEPS:

After formal approval by the Eu Council of Ministers, the final text should be published in the Official Journal of the European Union **by the first half of 2024**.

The text will come into effect 20 days after the publication of the AI Act in the Official Journal of the European Union.

THE BROAD DEFINITION OF AI SYSTEM CONTAINED IN THE AI ACT

"… *'AI system' means a machine-based system designed to operate with varying levels of autonomy, that may exhibit adaptiveness after deployment and that, for explicit or implicit objectives, infers, from the input it receives, how to generate outputs such as predictions, content, recommendations, or decisions that can influence physical or virtual environments* …" (**AI Act**, Article 3, para. 1).

Exemptions:

- AI systems used exclusively for military and scientific research and development purposes;
- Natural persons using AI systems for purely non-professional purposes.

THE RISK BASED APPROACH

3. The Unacceptable Risk AI Systems

UNACCEPTABLE RISK (1)
The list

- AI to exploit vulnerabilities.
- AI to manipulate individuals.
- Social scoring.
- Predictive policing.
- Biometric identification systems in publicly accessible spaces for law enforcement.
- AI used for the emotion recognition within the workplace and educational institutions.
- Biometric categorisation.
- Scraping of facial images from the internet or CCTV footage to create facial recognition databases.

UNACCEPTABLE RISK (2)
Examples

"… *Social Scoring*…".

UNACCEPTABLE RISK (3)
Examples

"… AI used for making risk assessments of natural persons in order to assess or predict the risk of a natural person committing a criminal offence, based solely on the profiling of a natural person or on assessing their personality traits and characteristics…".

UNACCEPTABLE RISK (4)
Examples

"… *Biometric categorization of natural persons …*".

UNACCEPTABLE RISK (5)
Effective day and fines

The bans will come into effect six months after the publication of the AI Act (*i.e.,* presumably by the end of 2024).

Fines up to 35 million Euro or 7% of total worldwide annual turnover.

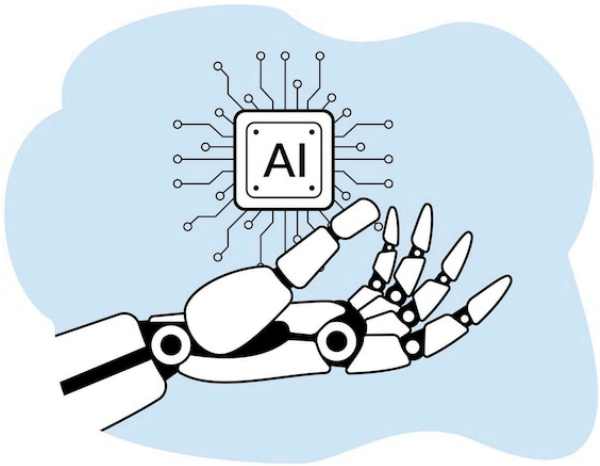
4. The High-risk AI Systems

HIGH RISK (1)
The Areas

- Remote biometric identification and categorisation.
- Critical infrastructure.
- Education and vocational training.
- Employment, workers management and access to self-employment.
- Access to and enjoyment of essential private services and essential public services and benefits.
- Law enforcement.
- Migration, asylum and border control management.
- Administration of justice and democratic processes.

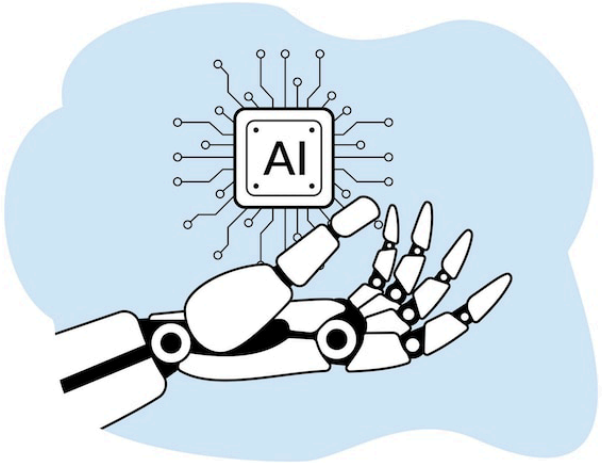
HIGH RISK (2)
Examples

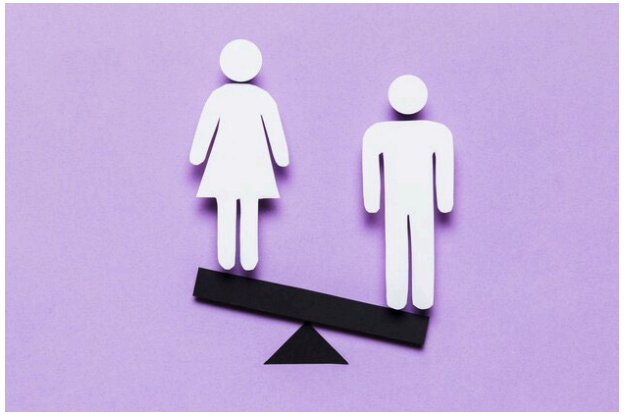
"… Access to and enjoyment of essential private services …".



HIGH RISK (3)
Examples

"… Employment, workers management and access to self-employment …".

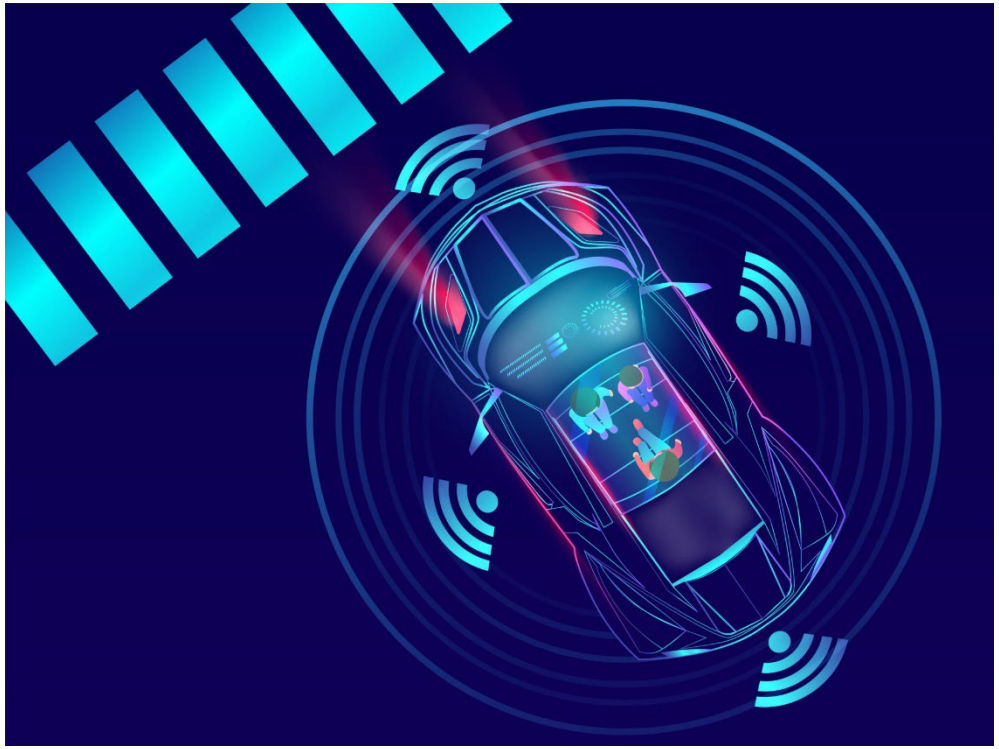


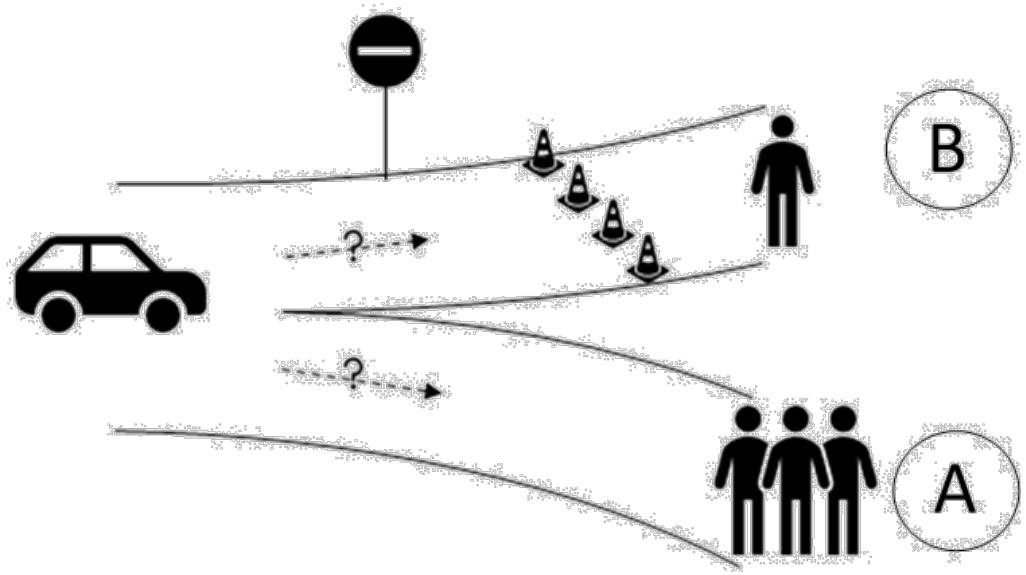


HIGH RISK (4)
The Compliance Activity To Be Carried Out
Ex Ante

- Risk Management System.
- Data and Data Governance.
- Technical Documentation.
- Traceability.
- Transparency obligations.
- Human Supervision.
- Accuracy, Robustness and Security.

HIGH RISK (5)
Focus on the human intervention





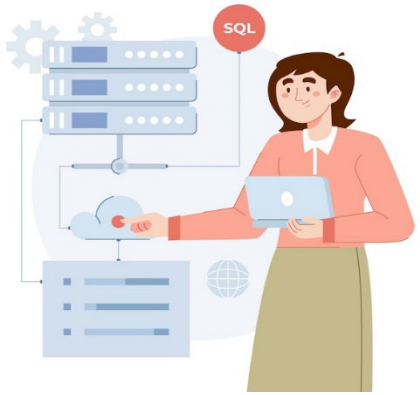
Source: Horizon Security S.r.l.

HIGH RISK (5)
The EU Declaration of Conformity

In the document, it will be necessary to document how each requirement of the AI Act mentioned in the previous slide is met (*e.g.*, Risk Management System, Data Governance, *etc.*).

HIGH RISK (6)
Further steps to be taken before the AI system can be placed on the market

The CE marking shall be affixed indicating compliance with the general principles and applicable Union Laws.



Registration in the EU Database.

HIGH RISK (7)
Effective day and fines

The rules will come into effect 24 months after the publication of the AI Act (with some exceptions).

Fines up to 15 million Euro or 3% of total worldwide annual turnover.

5. Transparency-risk AI Systems

TRANSPARENCY-RISK (1)
What they are

- AI systems that interact with individuals, such as those that can carry risks of manipulation (but which cannot be considered unacceptable or high-risk AI systems).
- Possible examples are artistic deepfakes or chatbots.
- There are specific obligations of transparency towards the users.

TRANSPARENCY-RISK (2)
Example

A company creates a deepfake video in which Maria Montessori is giving a lesson.

The user shall be warned that the deepfake was created using AI.

TRANSPARENCY-RISK (3)
Effective day and fines

The rules will come into effect 24 months after the publication of the AI Act.

Fines up to 7.5 million Euro or 1% of total worldwide annual turnover.

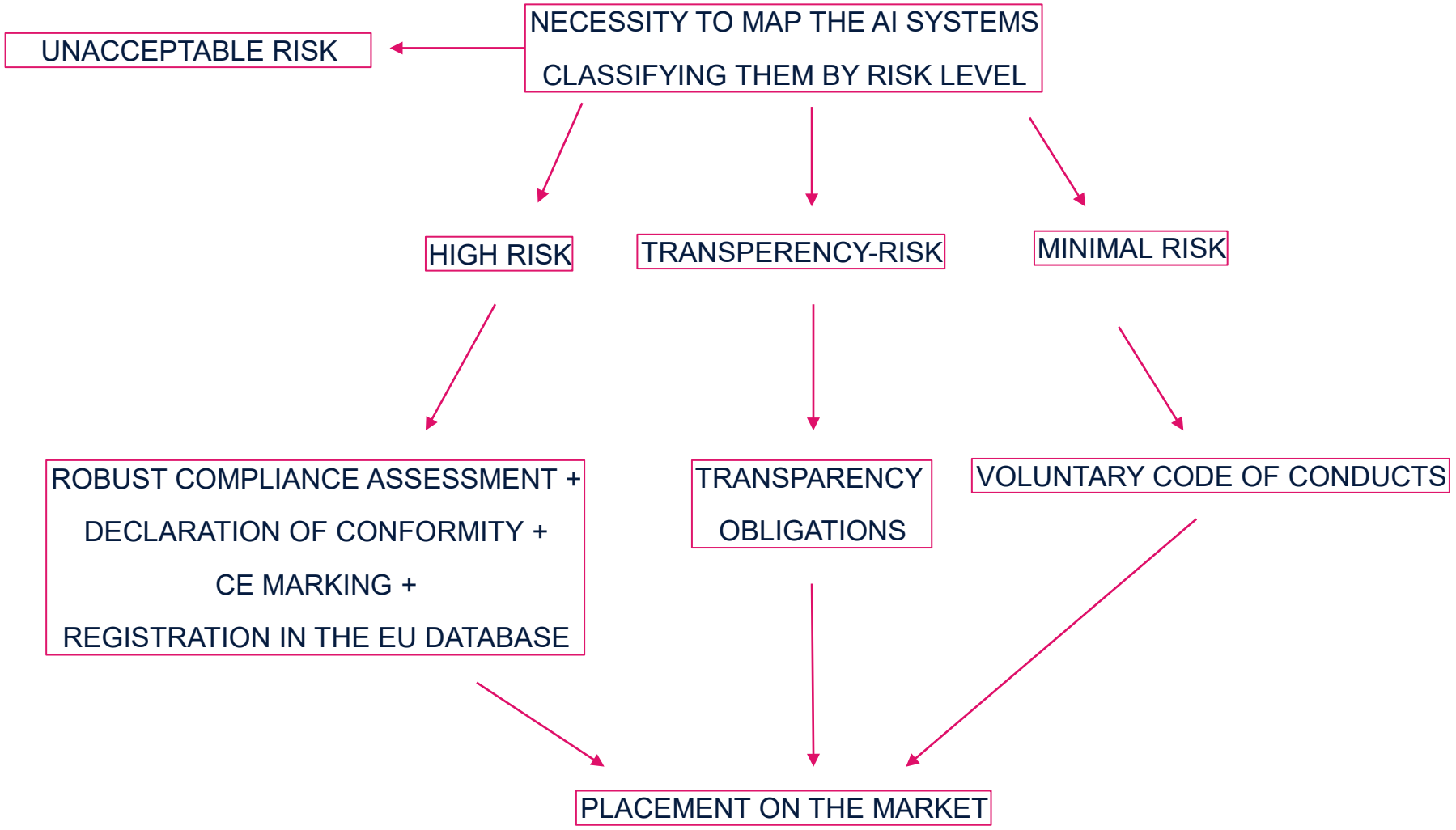
6. The minimal risk AI Systems

MINIMAL RISK

- They include categories such as anti-spam filters.
- In this case, it is established that such AI systems may be submitted to the voluntary application of codes of conduct.
- The rules will start to apply after 24 months from the publication of the AI ACT.

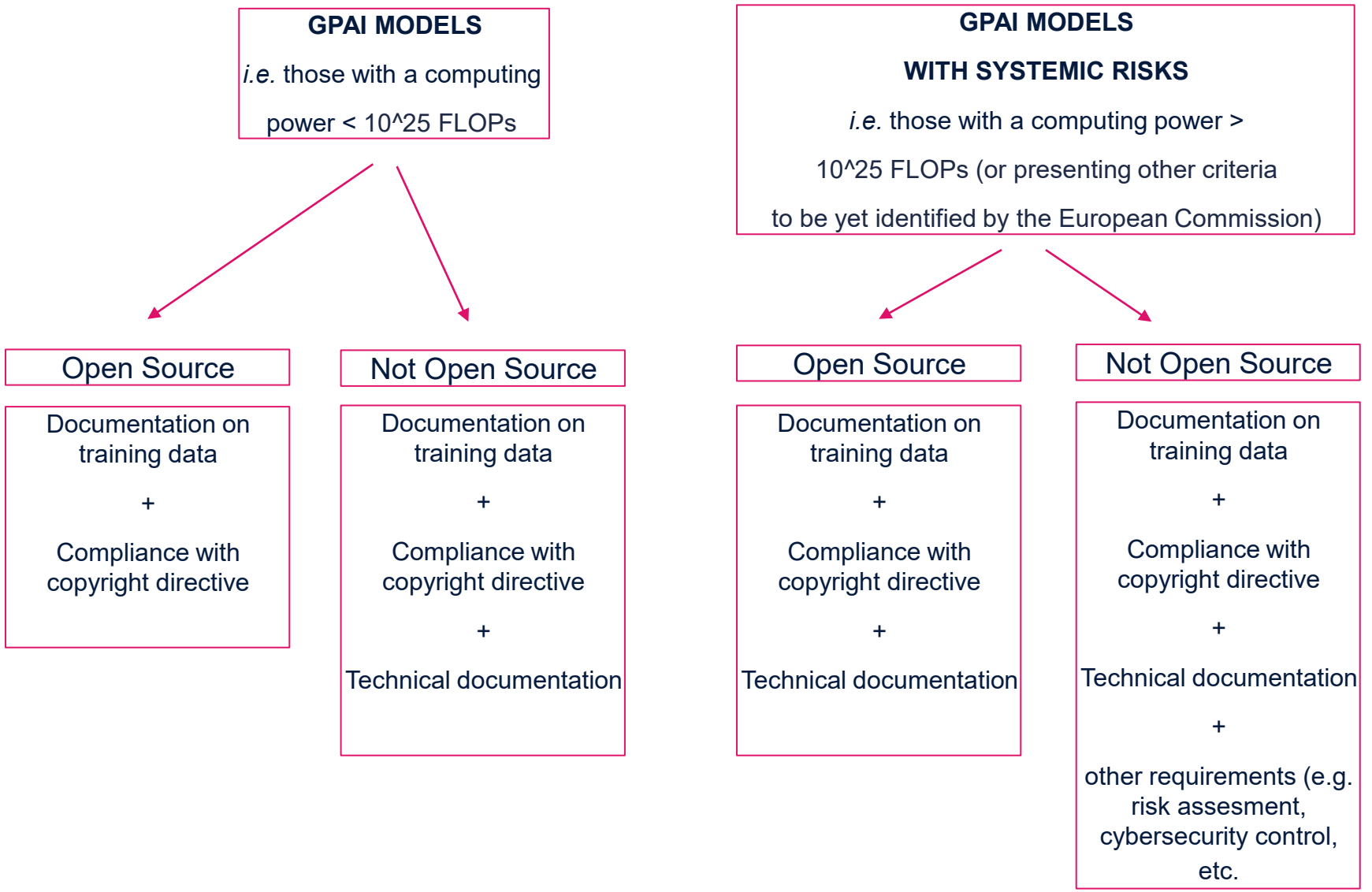
7. A final recap

A FINAL RECAP



8. The General-Purpose AI Models (GPAI)

GPAI


GPAI MODELS
i.e. those with a computing power < 10^25 FLOPs
Open Source
Documentation on training data
+
Compliance with copyright directive
Not Open Source
Documentation on training data
+
Compliance with copyright directive
+
Technical documentation
GPAI MODELS
WITH SYSTEMIC RISKS
i.e. those with a computing power > 10^25 FLOPs (or presenting other criteria to be yet identified by the European Commission)
Open Source
Documentation on training data
+
Compliance with copyright directive
+
Technical documentation
Not Open Source
Documentation on training data
+
Compliance with copyright directive
+
Technical documentation
+
other requirements (e.g. risk assesment, cybersecurity control, etc.

9. The AI Governance

THE AI GOVERNANCE (1)

AI Office

Composition

- Office set up within the Commission composed by independent experts.

Functions

- Enforce the rules on the GPAI Models.
- Developing methods for assessing AI Models.

AI Board

Composition

- One representative for any EU Member State.

Functions

- Coordination of the application of the text liaising with the single national Authorities.
- Publication of written opinions.

THE AI GOVERNANCE (2)

Authority/Authorities yet to be decided

- Power to impose fines.
- Coordination with the other national autorithies for ensuring the implementation of the AI ACT.

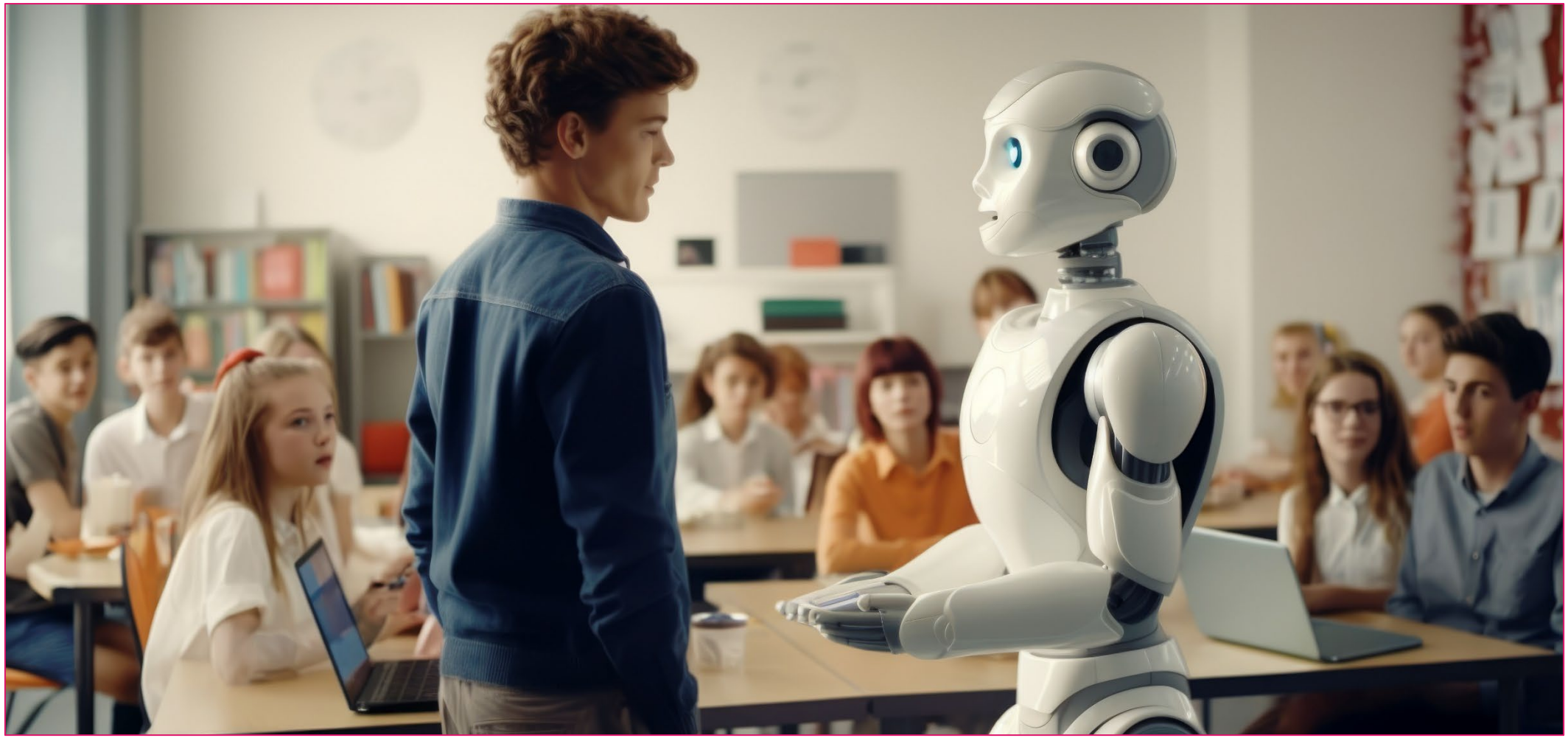
10. Critics moved to the AI ACT

CRITICS MOVED TO THE AI ACT

- Could this hinder technology and leave European companies behind?
- Does this introduce a new layer of cumbersome activities for companies, adding to the already significant compliance duties imposed, for example, under the General Data Protection Regulation?
- Is it feasible for legislation to anticipate every new development in constantly evolving technologies like artificial intelligence? (*e.g.*, the original proposal text did not include the phenomenon of generative AI).

11. The AI Act and education: handing over to the regulations

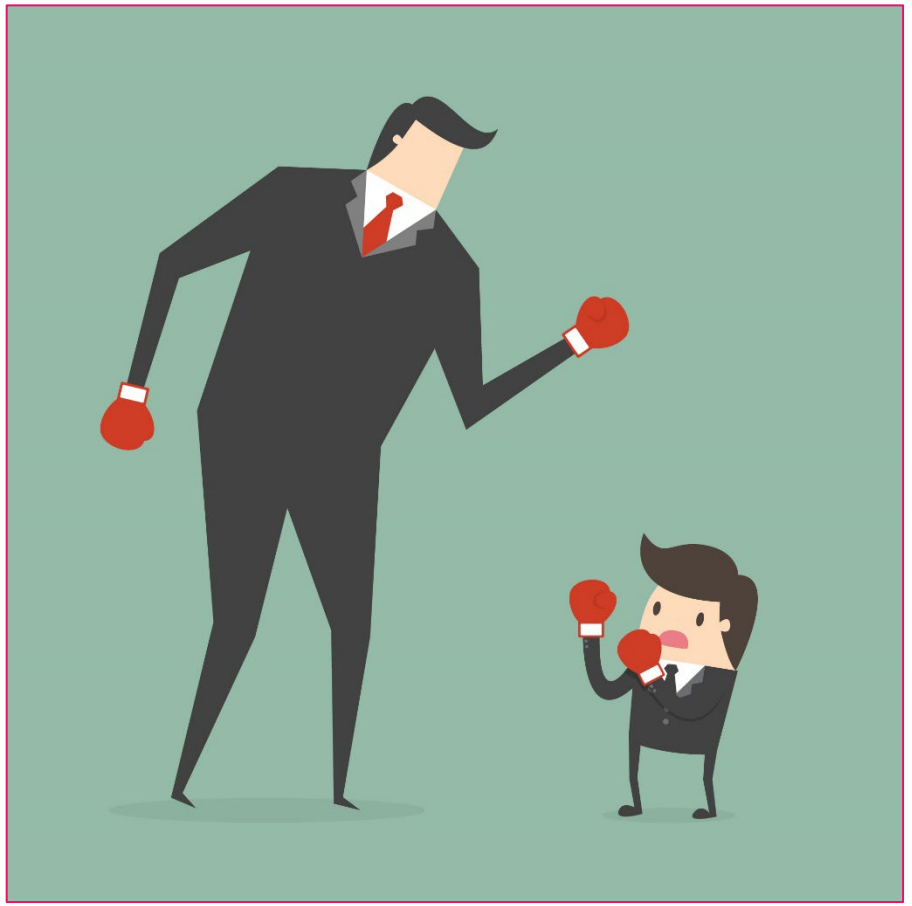
In which way does European regulation of artificial intelligence impact the field of education?



Recitals: the concept of imbalance of power (1)

AI ACT, RECITAL 44:

*"… **Considering the imbalance of power in the context of work or education, combined with the intrusive nature of these systems** (**AI systems**), **such systems could lead to detrimental or unfavourable treatment of certain natural persons or whole groups thereof**. Therefore, the placing on the market, the putting into service, or the use of AI systems intended to be used to detect the emotional state of individuals in situations related to the workplace and education should be prohibited …".*



Recitals: impact on a person's ability to secure a livelihood (2)

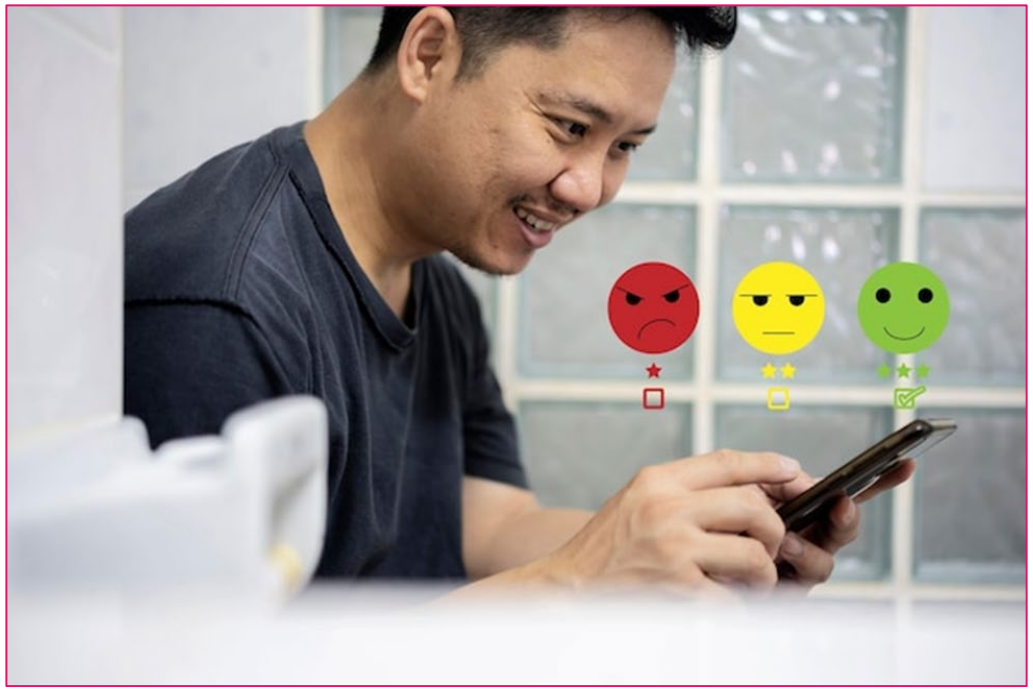
AI ACT, RECITAL 56:

"… AI systems used in education or vocational training, in particular for determining access or admission, for assigning persons to educational and vocational training institutions or programs at all levels, for evaluating learning outcomes of persons, for assessing the appropriate level of education for an individual and materially influencing the level of education and training that individuals will receive or will be able to access or for monitoring and detecting prohibited behavior of students during tests ***should be classified as high-risk AI systems, since they may determine the educational and professional course of a person's life and therefore may affect that person's ability to secure a livelihood*** *…".*

AI systems prohibited in education (1)

AI ACT, Article 5 (*f*):

The placing on the market, putting into service for this specific purpose, or use of AI systems ***to infer emotions of a natural person in workplace and educational institutions is prohibited, except where the AI system is intended for medical or safety reasons****.*

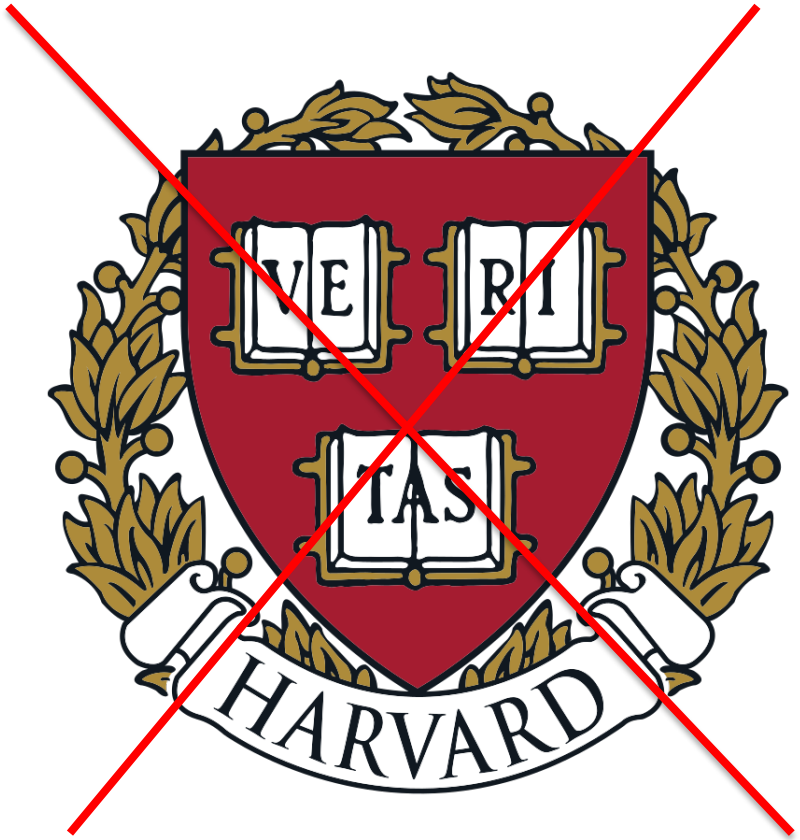


High-risk systems in education (2)

AI ACT, Annex 3.3:

Fall into the high-risk classification:

(a) *AI systems intended to be used to **determine access or admission** or to assign natural persons to educational and vocational training institutions at all levels;*



High-risk systems in education (3)

*(b) AI systems intended to be used to **evaluate learning outcomes**, including when those outcomes are used to steer the learning process of natural persons in educational and vocational training institutions at all levels;*

High-risk systems in education (4)

(c) *AI systems intended to be used for the purpose of **assessing the appropriate level of education that an individual will receive or will be able to access**, in the context of or within educational and vocational training institutions at all levels;*

(d) *AI systems intended to be used for **monitoring and detecting prohibited behavior of students during tests** in the context of or within educational and vocational training institutions at all levels.*

12. Living Guidelines on the responsible use of Generative AI in Research

The new "*Living Guidelines on the responsible use of Generative AI in Research*" by the European Commission (March 2024)

What is Generative Artificial Intelligence?

Generative Artificial Intelligence is a branch of artificial intelligence that focuses on creating systems capable of **autonomously generating original content, such as images, text, sounds, or even videos**. Unlike other types of AI that can only process existing data or perform actions based on predefined rules, generative AI systems can produce new data that has not been previously inputted into the algorithm.

Positive examples of the use of these tools by researchers

- Supporting non-native speakers in producing **texts in multiple languages**;
- Producing **text summaries** from different sources across extremely large corpuses quickly;
- Automatically retrieving and contextualising a **wide body of knowledge**.

On the other hand... risks and potential abuses (1)

- Risks related to **intellectual property and privacy**;
- **Delegation** of significant tasks that require human involvement to artificial intelligence;
- **Technological access disparities**: students lacking access to the necessary technologies for utilizing generative AI may be disadvantaged compared to their peers, exacerbating inequalities in education;

On the other hand... risks and potential abuses (2)

- **Overlap with Human Skills**: Automating educational tasks with AI could reduce opportunities for students to develop essential skills such as communication, collaboration, and problem-solving;

- **Technological Dependence**: Overreliance on generative AI may limit students' cognitive and decision-making skills, rendering them less capable of navigating complex real-world situations.

Why these Guidelines?

- The scientific community should use the Generative AI in a **responsible** manner;

- The development of a robust framework for generative AI in scientific research cannot be the sole responsibility of policymakers: **universities, research organisations, funding bodies, research libraries, learned societies, publishers and researchers at all stages of their careers are essential in shaping the discussion on AI** and how it can serve the public interest in research;

- The Guidelines have to be considered as a **supporting tool** for research funding bodies, research organisations and researchers; they are not binding and they complement and build on the EU AI policy, including the Artificial Intelligence Act.

Recommendations for researchers (1)

1. **Remain ultimately responsible for scientific output:**

- Researchers maintain a **critical approach** to using the output produced by generative AI and are aware of the tools' limitations, such as **bias**, hallucinations and inaccuracies;

- AI systems are neither authors nor co-authors. Authorship implies agency and responsibility, so it lies with human researchers.

Recommendations for researchers (2)

What is a bias in artificial intelligence?

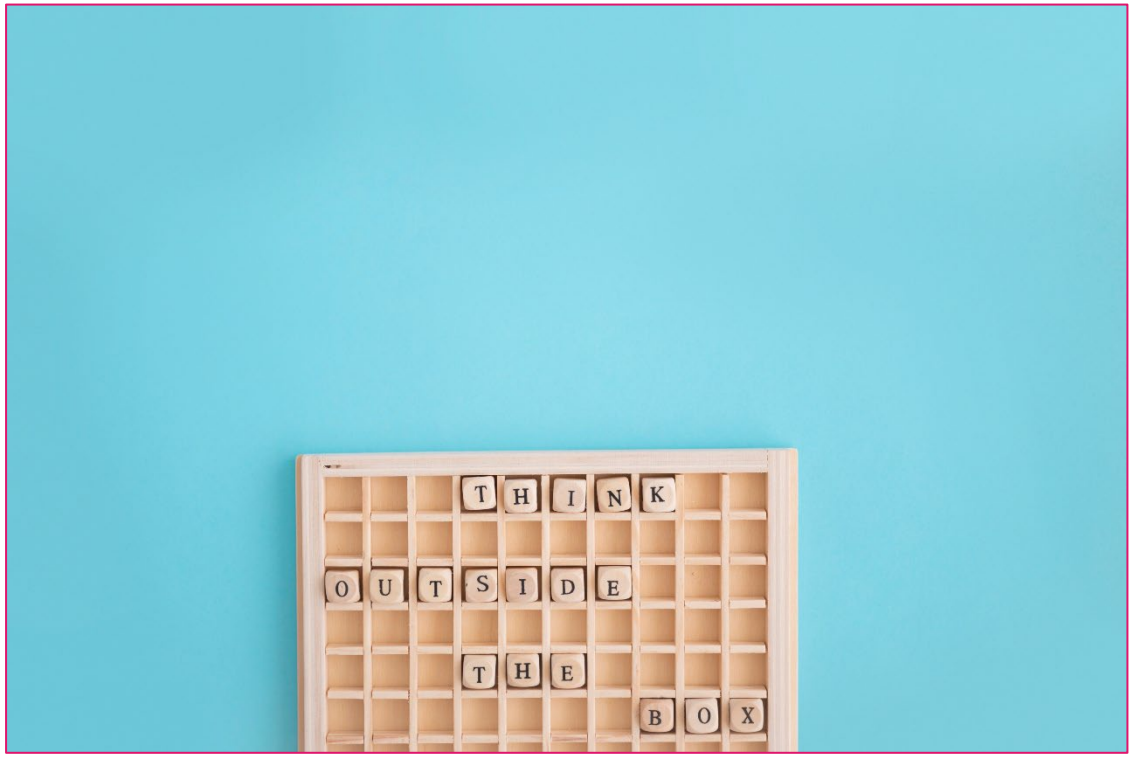
A bias refers to systematic errors or distortions in the data or algorithms used in AI systems, resulting in **unfair or discriminatory outcomes**, particularly against certain groups or individuals. This bias can arise from various sources, such as biased training data, algorithmic biases, or even human biases embedded in the design process.

Recommendations for researchers (3)

2. Use generative AI transparently:

- detailing **which generative AI tools has been used substantially** in the research process. Reference to the tool could include the name, version, date, etc. and how it was used and affected the research process. If relevant, researchers make the input (**prompts**) and output available, in line with open science principles;

- disclosing or discussing the limitations of generative AI tools used, including possible biases in the generated content, as well as possible mitigation measures.



Recommendations for researchers (4)

What is a prompt in artificial intelligence?

A prompt refers to a specific **input or instruction provided to an AI model** to generate a desired output. It can be a text-based query, command, or context given to the AI system to initiate a response.

Recommendations for researchers (5)

Example: prompt for writing a fantasy novel

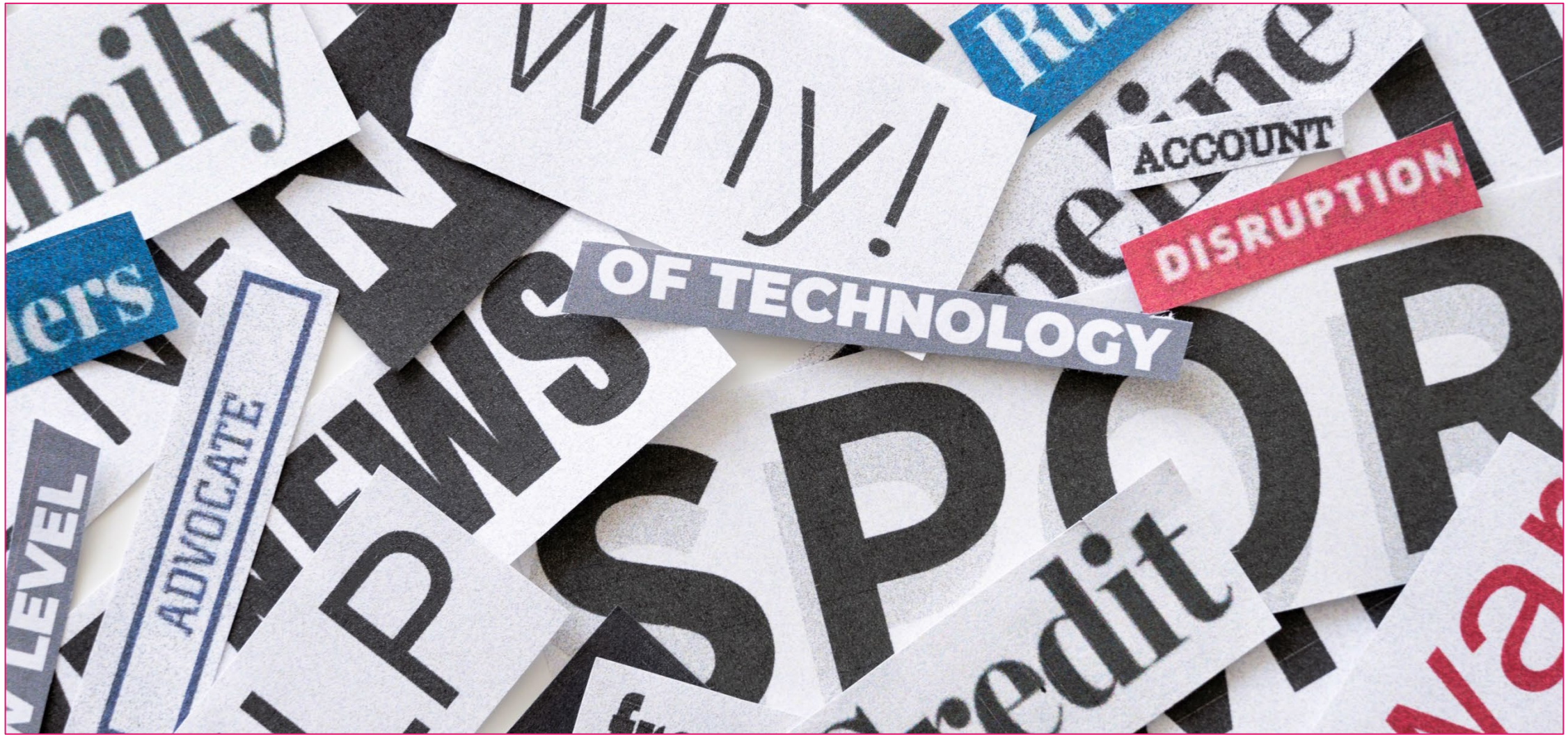
"The main character is a half-elf and half-human, seeking redemption by avenging his father. He will walk for a month and defeat a great dragon".

Recommendations for researchers (6)

3. Pay particular attention to issues related to privacy and intellectual property rights when sharing sensitive or protected information with AI tools:

Wait a moment... what are we talking about?

Intellectual property rights and Artificial Intelligence: An unhappy marriage



New York Times vs. OpenAI (1)

- The intensification of the debate on artificial intelligence underscores the growing need for global regulation, as demonstrated by The New York Times' legal action against OpenAI and Microsoft;

- The critical challenge facing the AI industry concerns **the use of copyrighted works in model training.**

New York Times vs. OpenAI (2)

- The New York Times' legal action raises fundamental ethical and legal questions about the **use of journalistic works in training artificial intelligence algorithms**;

- This is a true problem: OpenAI responds that it would be impossible to create services like ChatGPT if they were prevented from relying on all copyrighted works. The company claimed that "***copyright law does not prohibit training***".

New York Times vs. OpenAI (3)

The legal dilemma is as follows: **does training AI algorithms using copyrighted works constitute a violation of the copyright?**

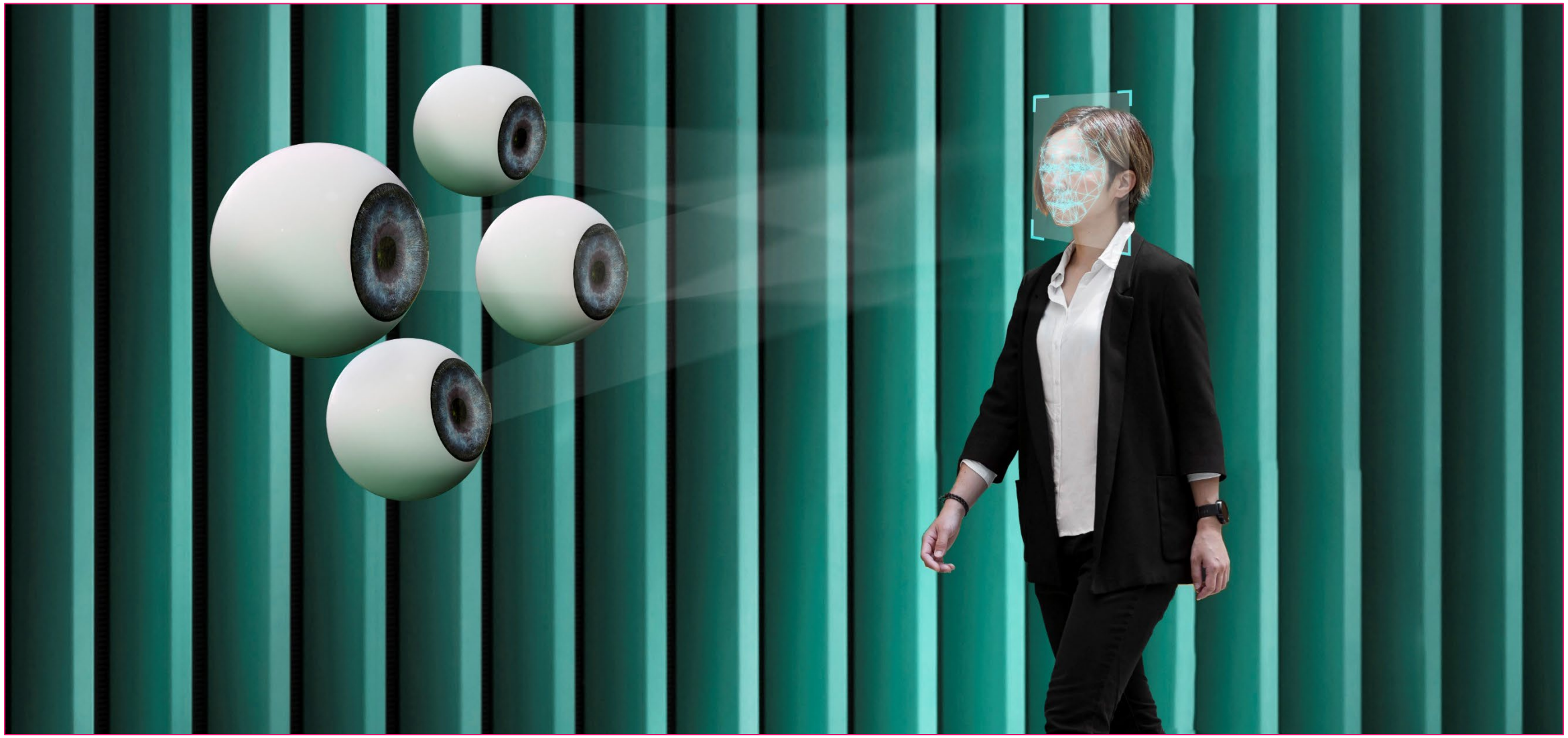
Recommendations for researchers (7)

3. Pay particular attention to issues related to intellectual property rights when sharing sensitive or protected information with AI tools.

… *Therefore?*

You should protect unpublished or sensitive work by taking care **not to upload it into an online AI system** unless there are assurances that the data will not be re-used (*e.g.*, to train future language models).

Privacy and Artificial Intelligence: Living Together Under One Roof



The Italian Data Protection Authority vs. Bocconi University (1)

- The Italian Data Protection Authority imposed a fine of 200,000 euros on Bocconi University for violating privacy regulations due to the use of AI systems;

- Specifically, the sanction was issued due to the university's use of proctoring software to **monitor students during written exams**;

- The software, developed by the American company "Respondus Inc," includes "Lockdown Browser" to prevent online searches during exams and "Respondus Monitor" to monitor students via webcam.

The Italian Data Protection Authority vs. Bocconi University (2)

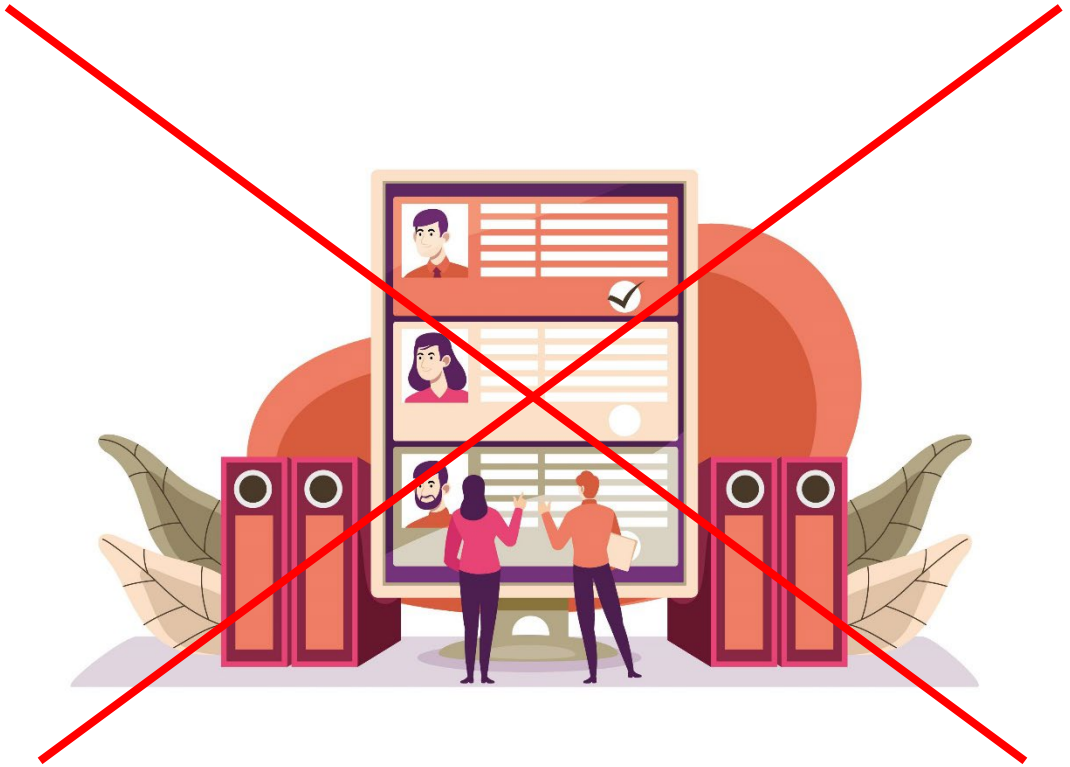
- Bocconi University was sanctioned due to the **lack of consent from students** and the inadequate information provided about the personal data processing;
- The Authority highlighted **violations of fundamental privacy principles**, including transparency and limitation of data processing;
- The sanction was also based on the **lack of a sufficient legal basis for biometric data processing**.

Recommendations for researchers (8)

3. Pay particular attention to issues related to privacy when sharing sensitive or protected information with AI tools.

… *Therefore?*

You should take care **not to provide third parties' personal data to online generative AI systems** unless the data subjects have given their consent and researchers have a clear goal for the intended use of the personal data.



Recommendations for researchers (9)

4. Continuously learn how to use generative AI tools properly to maximise their benefits, including by undertaking training.

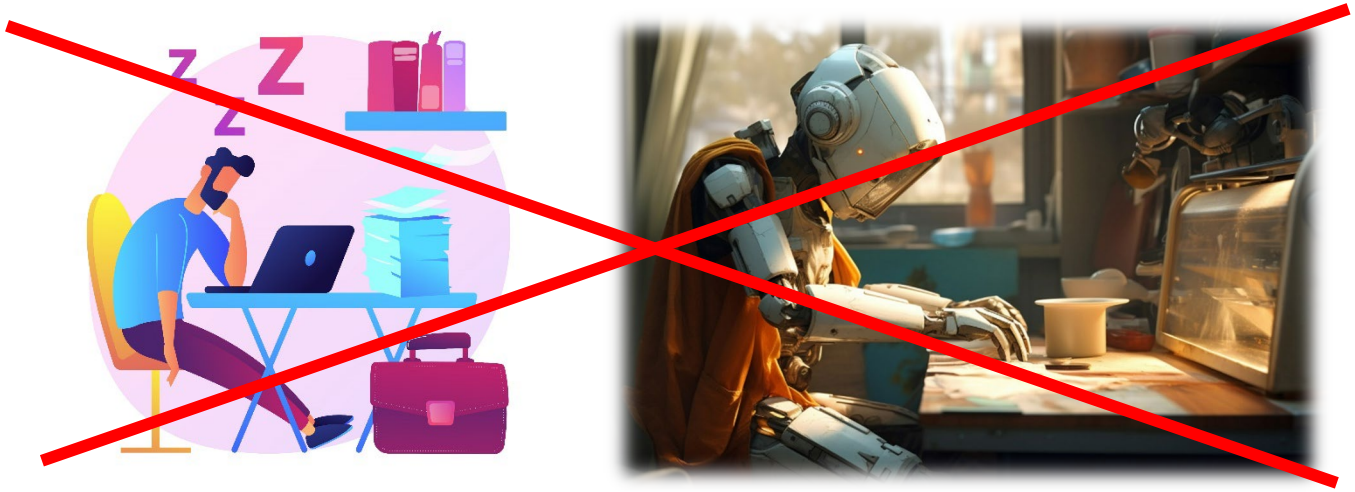
Generative AI tools are evolving quickly, and new ways to use them are regularly discovered. **Researchers stay up to date on the best practices** and share them with colleagues and other stakeholders.

Recommendations for researchers (10)

5. Refrain from using generative AI tools substantially in sensitive activities that could impact other researchers or organisations (for example peer review, evaluation of research proposals, etc).

What is a substantial use?

Example: using Generative AI to search background info for a review is not a substantial use, while delegating the evaluation or the assessment of a paper is a substantial use.



13.Case Study: KU Leuven’s Guidelines for Teaching Staff

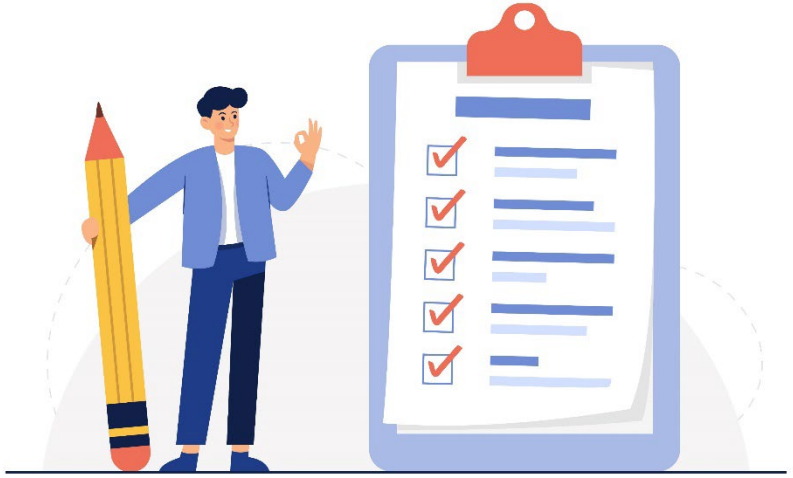
Responsible use of Generative Artificial Intelligence, Guidelines for teaching staff

Recommendations for teaching staff (1)

- Teaching staff are expected to clearly **inform students** about **whether or not they are allowed to use GenAI** for assignments such as visual, writing and programming assignments;

- Students are expected to be **transparent about the use of GenAI** so that their knowledge, understanding and skills can be assessed fairly and correctly;

- If a student uses GenAI and is not transparent about its use, this may be considered an **irregularity.**



Recommendations for teaching staff (2)

You should discuss the specifics of your course:

- When a student gets an assignment, you **clearly indicate what is and what isn't allowed**;
- Clearly communicate the limitations of GenAI to your students.
- Discuss how you expect students to be transparent about the use of GenAI in their assignments, and how they have to report and monitor it;
- Give students information on how they have to cite and reference GenAI.

Recommendations for teaching staff (3)

- Students at KU Leuven are encouraged to be transparent about their use of GenAI in assignments.

- To facilitate transparency, students are advised to include additional information regarding their GenAI usage in assignments. This may involve providing **screenshots of interactions with GenAI, highlighting relevant details, or explaining the purpose and method of GenAI utilization**.

Recommendations for teaching staff (4)

And what about… detecting GenAI?

"*Turnitin's **AI detector** is no longer available since the start of the academic year 2023-2024. Numerous investigations in 2023 resulted in its **unreliability***" (Guidelines, KU Leuven).

At the moment, there are no detectors with full reliability: "*Consider the **AI scores** of AI detectors as **purely indicative** (…),*
Look for clues other than AI scores. The result of an AI detector is not enough to establish an anomaly. Thus, even now, a similarity score of Turnitin is not a sufficiently conclusive argument to impose a sanction" (Guidelines, KU Leuven).

Recommendations for teaching staff (5)

So, what tools remain available?

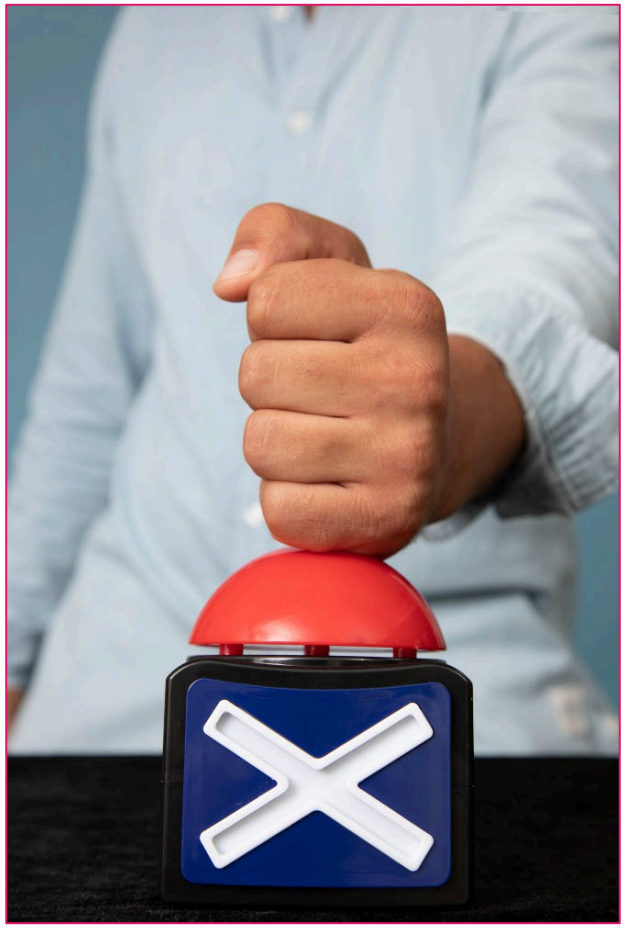
There are only clues, indications, traces...

- a clear break in writing style;
- deviations in content;
- indications via the references;
- the student not being able to explain or defend the paper.

Recommendations for teaching staff (6)

- If a combination of circumstances or a convergence of clues is identified, which cannot be attributed to chance, this may justify **imposing a sanction;**

- Teachers or supervisors who suspect based on multiple indications are encouraged to **seek guidance from the faculty's plagiarism expert** to ensure appropriate actions are taken.



THANK YOU!

SHARING VALUE WITH YOU.

MILANO
Via San Paolo, 7 · 20121 Milano, Italia
T. +39 02 72554.1 · F. +39 02 72554.400
milan@dejalex.com

ROMA
Via Vincenzo Bellini, 24 · 00198 Roma, Italia
T. +39 06 809154.1 · F. +39 06 809154.44
rome@dejalex.com

BRUXELLES
Chaussée de La Hulpe 187 · 1170 Bruxelles, Belgique
T. +32 (0)26455670 · F. +32 (0)27420138
brussels@dejalex.com

MOSCOW
Ulitsa Bolshaya Ordynka 37/4 · 119017, Moscow, Russia
T. +7 495 792 54 92 · F. +7 495 792 54 93
moscow@dejalex.com

www.dejalex.com