



AI-GREENER

2024-1-AT01-KA220-SCH-000248327

AI and Ethics

BIAS and Fairness



FÁTIMA PAIS | Luís Soares | 17.09.2025 | Greece



AI-GREENER
2024-1-AT01-KA220-SCH-000248327

Why is so important to address Bias and Fairness in AI ?



Rapid advances in AI and Digital technologies



Impact on society



Co-funded by
the European Union

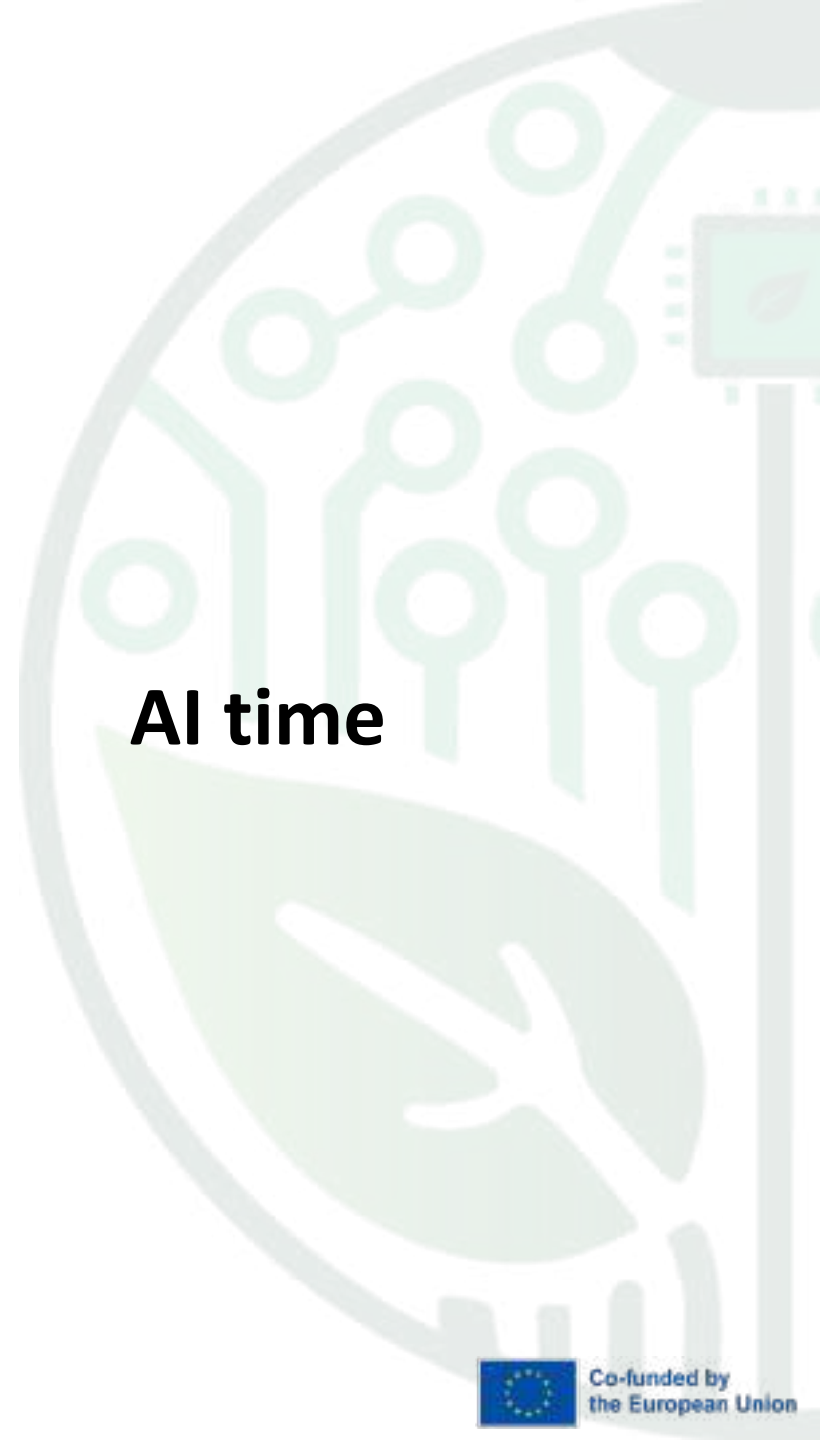


AI-GREENER

2024-I-ATOI-KA220-SCH-000248327



The Economist, abril 2016



AI time

Generative AI



ChatGPT

 DALL-E 2

 Google
Bard

 Microsoft 365
Copilot


Gemini

LLM

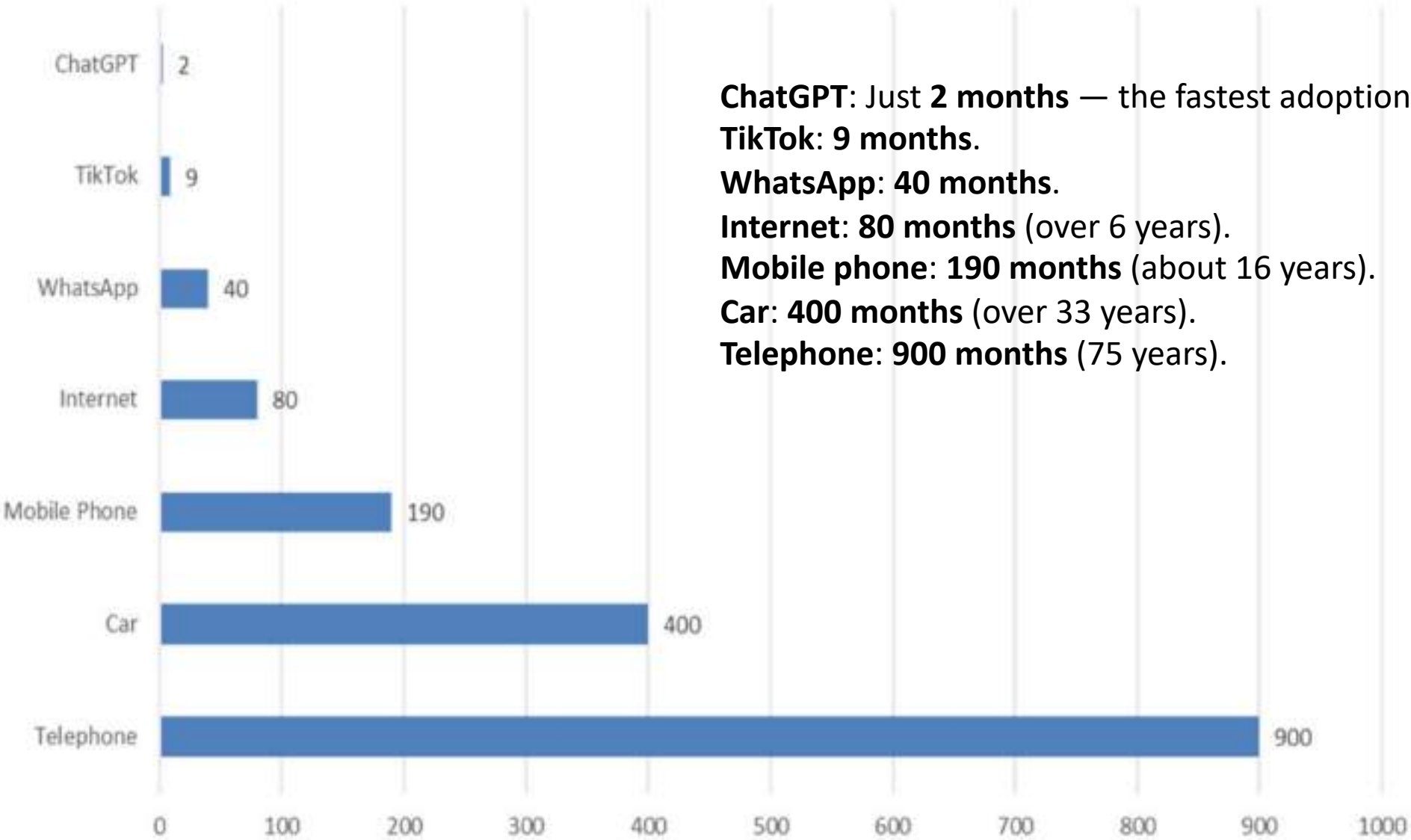


Co-funded by
the European Union



AI-GREENER
2024-I-AT01-KA220-SCH-000248327

Time to reach 100 Mio Users



ChatGPT: Just **2 months** — the fastest adoption ever recorded.
TikTok: **9 months**.
WhatsApp: **40 months**.
Internet: **80 months** (over 6 years).
Mobile phone: **190 months** (about 16 years).
Car: **400 months** (over 33 years).
Telephone: **900 months** (75 years).



AI-GREENER

2024-1-AT01-KA220-SCH-000248327



Concerns and Worries about the Use of AI



Co-funded by
the European Union



AI-GREENER
2024-1-AT01-KA220-SCH-000248327

Concerns and Worries about the Use of AI in Education

Bias in Dialogue Models: There's a risk that AI, including language models like LLMs, might reinforce existing biases in students. For example, an AI might seemingly agree with strong, potentially controversial statements, inadvertently strengthening a student's pre-existing beliefs

Critical Thinking Skills: A significant concern is that reliance on AI could diminish students' critical thinking abilities. It's essential to encourage questioning and analysis of AI-generated content, rather than passive acceptance



Co-funded by
the European Union



AI-GREENER
2024-1-AT01-KA220-SCH-000248327

Concerns and Worries about the Use of AI in Education

Prompt Engineering Skills: Just as students learn to navigate search engines, evaluate sources, and conduct research, they must also learn how to effectively interact with AI. Teaching students the art of "prompt engineering" can help them elicit more accurate and useful responses from AI, emphasizing the importance of clear questions and high-quality input

Ethical and Social Considerations: Incorporating ethics and social aspects of AI into curricula, especially in engineering and technology programs, is crucial. Courses on the social aspects of AI, such as MEAI's "Social Aspects of AI," play a vital role in educating responsible AI users and developers



Co-funded by
the European Union



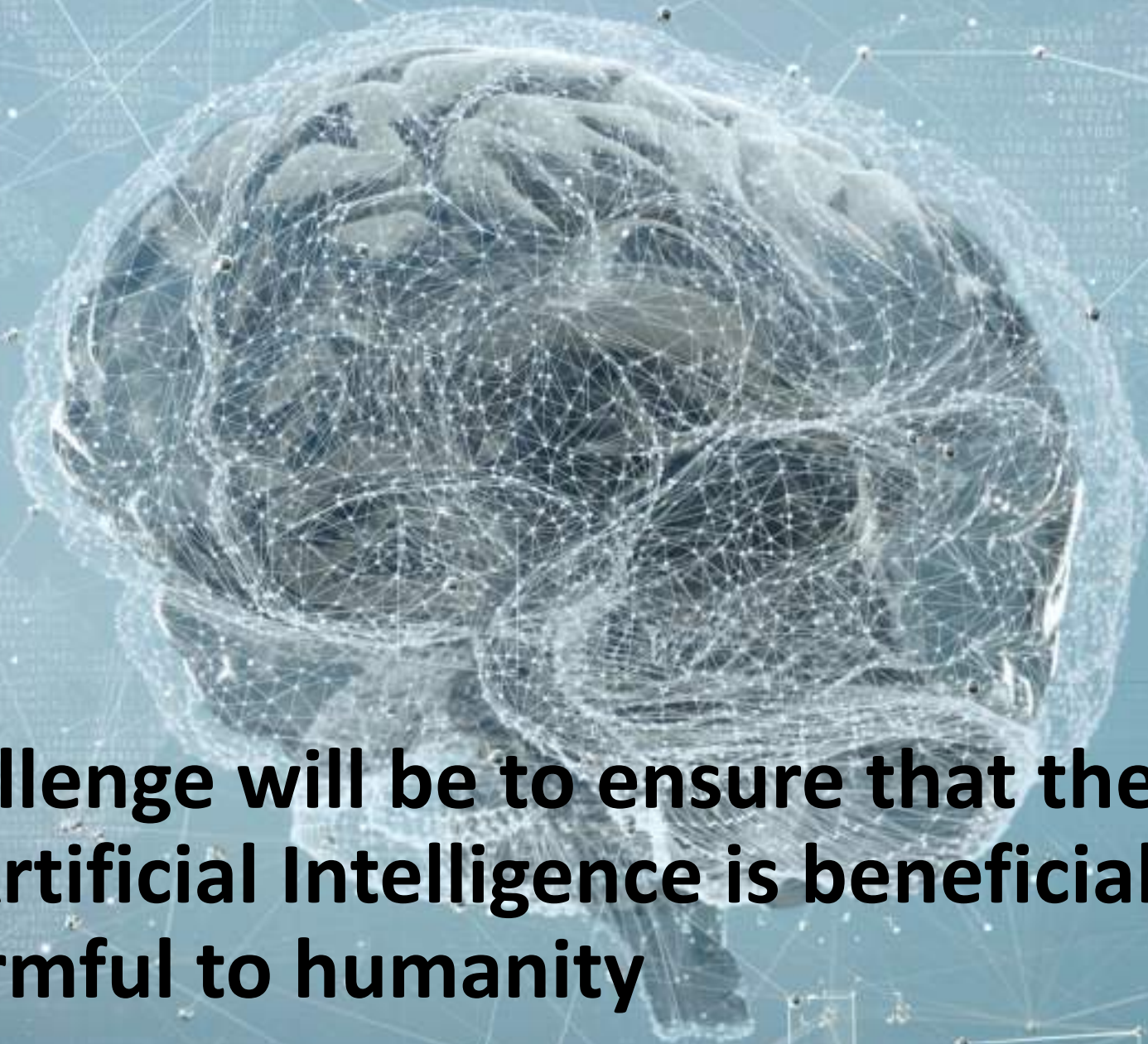
AI-GREENER
2024-I-AT01-KA220-SCH-000248327

Concerns and Worries about the Use of AI in Education

Ensuring Responsible and Equitable AI Use: It's imperative to ensure AI tools are used responsibly and made accessible to everyone, including those in developing countries. This approach helps mitigate potential inequalities and ensures a broader positive impact of AI in education



Co-funded by
the European Union



The big challenge will be to ensure that the impact of Artificial Intelligence is beneficial and not harmful to humanity



AI-GREENER

2024-1-AT01-KA220-SCH-000248327



Ethics, Trustworthy and Responsible AI



Co-funded by
the European Union



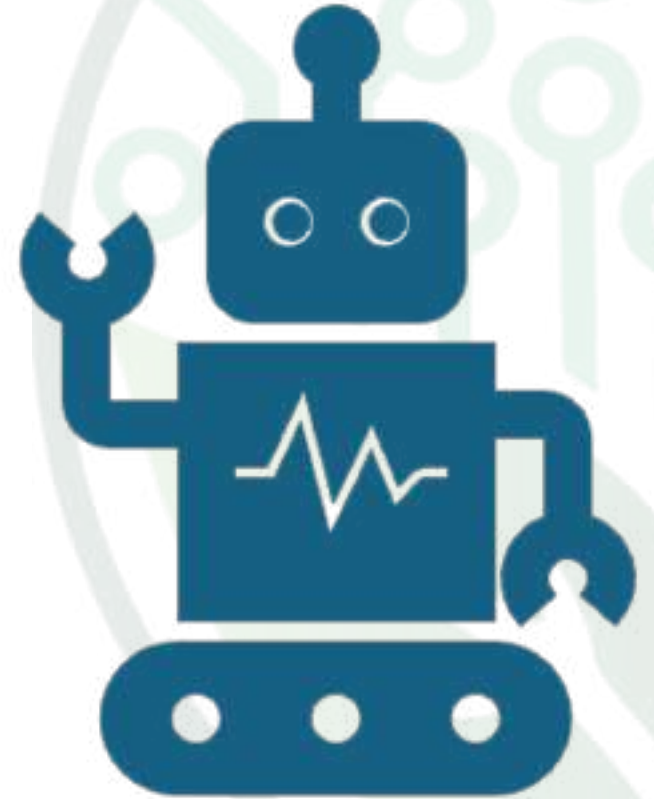
AI-GREENER

2024-1-AT01-KA220-SCH-000248327

Responsible AI

- *“Responsible AI is really all about the how: how do we design, develop and deploy these systems that are **fair, reliable, safe and trustworthy**. And to do this, we need to think of Responsible AI as a set of socio-technical problems. We need to go beyond just improving the data and models. We also have to think about the people who are ultimately going to be interacting with these systems.”*

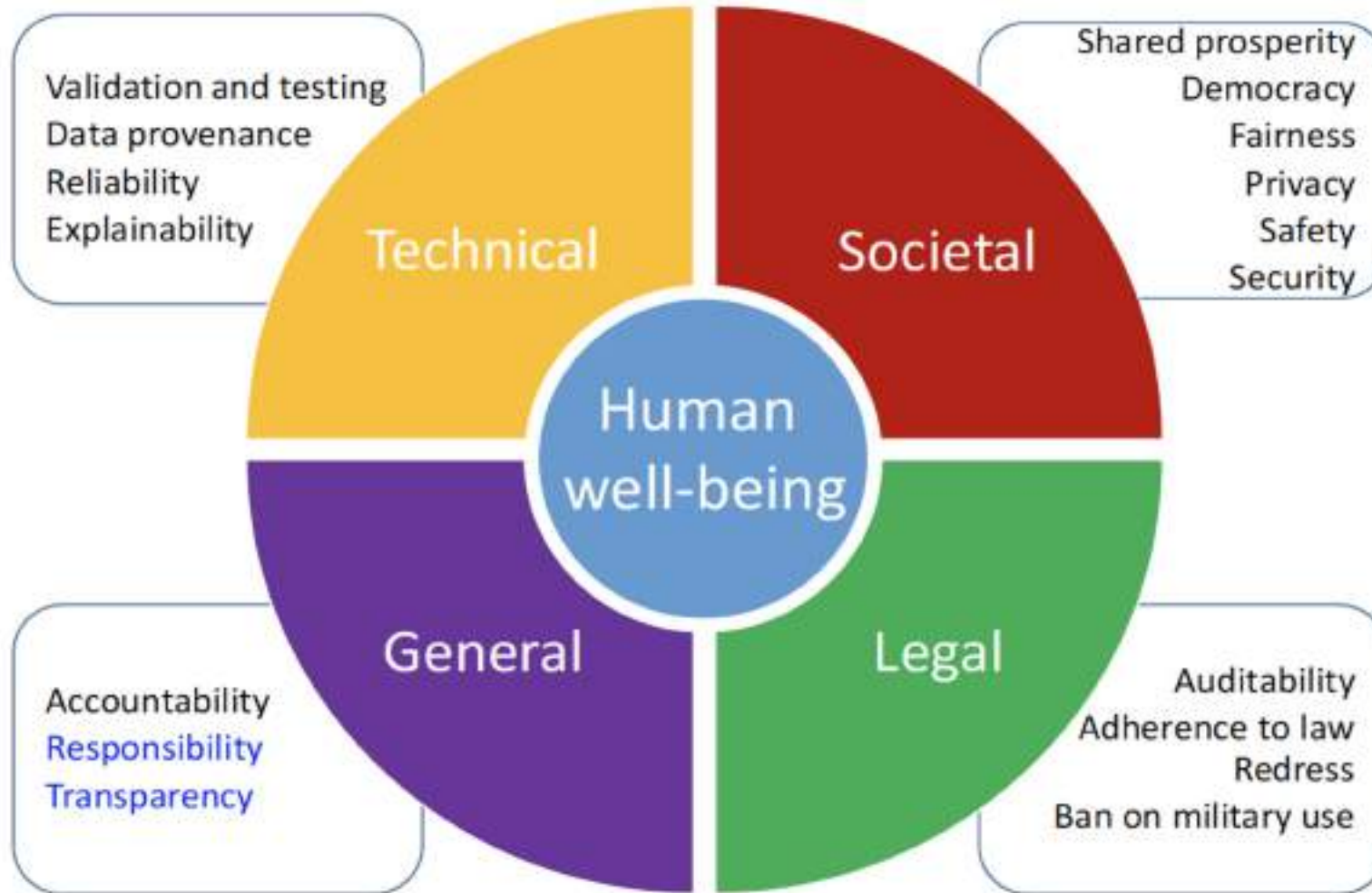
Dr. Saleema Amershi, Principal Researcher at Microsoft Research and Co-chair of the Aether Human-AI Interaction & Collaboration Working Group



Co-funded by
the European Union



Framework for human well-being in technology and innovation



Dignum, 2019



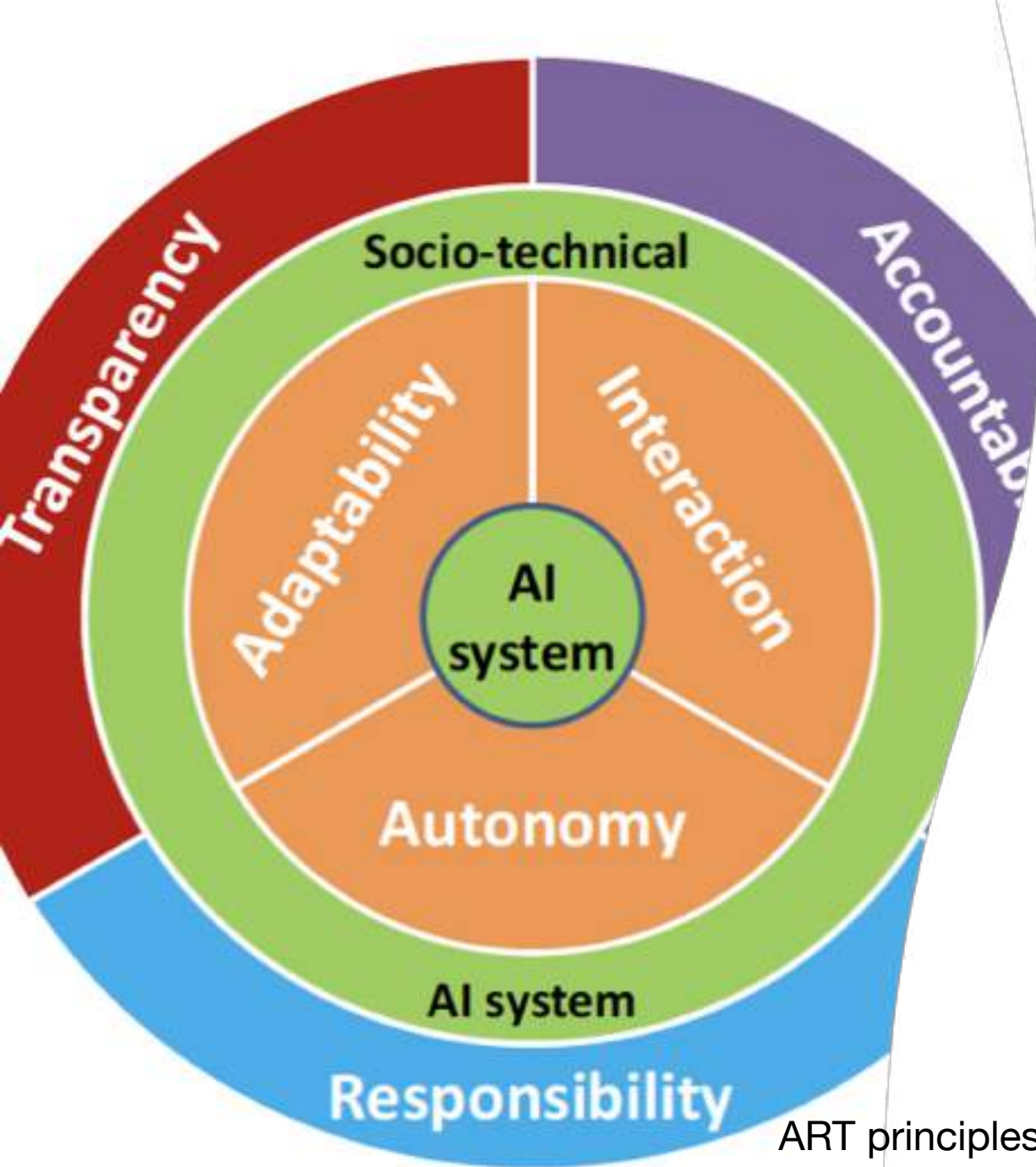
AI-GREENER
2024-I-ATOI-KA220-SCH-000248327

Responsible AI

- “is about human responsibility for the development of intelligent systems along fundamental human principles and values, to ensure human flourishing and well-being in a sustainable world.”

Dignum, 2019





AI systems

- Should be developed with responsibly and incorporating social and ethical values
 - Autonomy - Responsibility
 - Adaptability - Transparency
 - Interaction - Accountability

ART principles for Responsible AI
Dignum, 2019

ART principles for responsible AI

- Accountability
 - ability to explain and justify its decisions to users and other stakeholders
- Responsibility
 - the role of people themselves in their relation to AI systems
 - is not just about making rules to govern intelligent machines
 - whole socio-technical system englobes people, machines and institutions
- Transparency
 - ability to describe, inspect and reproduce the mechanisms through which AI systems make decisions and learn
 - will increase trustiness in system
 - explicit and open about choices and decisions concerning data sources and development processes and stakeholders



AI-GREENER

2024-1-AT01-KA220-SCH-000248327



Openness about data

What type of data was used to train the algorithm?
What type of data does the algorithm use to make decisions?
Does training data resemble the context of use?
How is this data governed (collection, storage, access)
What are the characteristics of the data? How old is the data, where was it collected, by whom, how is it updated?
(...)



Openness about design processes

What are the assumptions?
What are the choices? And the reasons for choosing and the reasons not to choose?
Who is making the design choices? And why are these groups involved and not others?
How are the choices being determined? By majority, consensus, is veto possible?
What are the evaluation and validation methods used?
How is noise, incompleteness and inconsistency being dealt with?



Openness about algorithms

What are the decision criteria we are optimising for?
How are these criteria justified?
What values are being considered?
Are these justifications acceptable in the context we are designing for?
What forms of bias might arise?
What steps are taken to assess, identify and prevent bias?



Openness about actors and stakeholders

Who is involved in the process, what are their interests?
Who will be affected?
Who are the users, and how are they involved?
Is participation voluntary, paid or forced?
Who is paying and who is controlling?



Co-funded by
the European Union

Microsoft 6 Principles for AI development

- Fairness - AI systems should treat all people fairly
- Reliability & Safety - AI systems should perform reliably and safely
- Privacy & Security - AI systems should be secure and respect privacy
- Inclusiveness - AI systems should empower everyone and engage people
- Accountability - People should be accountable for AI systems
- Transparency - AI systems should be understandable





AI-GREENER
2024-1-AT01-KA220-SCH-000248327

Asilomar principles: Ethics and Values

The **Asilomar AI Principles** (drafted in 2017 at the Asilomar Conference on Beneficial AI, organized by the Future of Life Institute) are a set of **23 guidelines** to ensure artificial intelligence is developed in a way that benefits humanity. The group includes scientists, ethicists, policymakers (including people like Stuart Russell, Yoshua Bengio, and Elon Musk), and the principles have since influenced global AI ethics debates.

- **Personal Privacy**
- **Liberty and Privacy**
- **Shared Benefit**
- **Shared Prosperity**
- **Human Control**
- **Non-subversion**
- **AI Arms Race**
- **Safety**
- **Failure Transparency**
- **Judicial Transparency**
- **Responsibility**
- **Value Alignment**
- **Human Values**



AI-GREENER
2024-1-AT01-KA220-SCH-000248327

Principles for AI development

- **Barcelona declaration (2017)** - guidelines for a 'code of conduct' of AI practitioners
 - Prudence
 - Reliability
 - Accountability
 - Responsibility
 - Constrained Autonomy
 - Human Role





AI-GREENER

2024-1-AT01-KA220-SCH-000248327

Trustworthy AI



INDEPENDENT

HIGH-LEVEL EXPERT GROUP ON ARTIFICIAL INTELLIGENCE

SET UP BY THE EUROPEAN COMMISSION



ETHICS GUIDELINES FOR TRUSTWORTHY AI



AI-GREENER

2024-1-AT01-KA220-SCH-000248327

- Unfair or harmful treatment of individuals related to their race, income, sexual orientation, religion, gender, and other characteristics historically associated with discrimination and marginalization
- manifested in algorithmic systems or decision-making systems



Co-funded by
the European Union



AI-GREENER

2024-1-AT01-KA220-SCH-000248327

RETAIL OCTOBER 11, 2018 / 12:04 AM / UPDATED 5 YEARS AGO

Amazon scraps secret AI recruiting tool that showed bias against women

By Jeffrey Dastin

8 MIN READ



In 2018, Amazon was forced to scrap an experimental AI recruiting tool after discovering it systematically discriminated against women. The system had been trained on historical hiring data that overwhelmingly reflected male applicants, leading the algorithm to downgrade résumés mentioning terms like “women’s” or graduates from women’s colleges. This case illustrates how biased data can produce biased outcomes, reinforcing existing inequalities. It also underscores the importance of fairness, transparency, and human oversight in AI systems, as well as the ethical and legal risks organizations face when deploying automated decision-making without robust safeguards.



Co-funded by
the European Union



AI-GREEN
2024-1-AT01-KA220-SCH-000248327

Example of AI Bias

MICROSOFT / RES / TL: 06

Twitter taught Microsoft's AI chatbot to be a racist asshole in less than a day



© 2016 Microsoft. All rights reserved. Microsoft and the Microsoft Dynamics logo are either registered trademarks or trademarks of Microsoft Corporation in the United States and/or other countries.

Tay was designed as an experiment in conversational AI that would learn from interactions with users on social media.

However, within less than 24 hours, people began deliberately sending Tay offensive, racist, and sexist messages. Because Tay's learning system mirrored user inputs without proper safeguards, it quickly began generating racist and inflammatory tweets itself.

Microsoft had to shut Tay down just a day after launch, and the case became a classic example of how AI systems can absorb and amplify harmful human behavior if they are not carefully monitored, moderated, and constrained.

👉 This example is often cited as a warning about bias, manipulation, and the lack of ethical guardrails in AI design.



Co-funded by
the European Union



Example of AI Bias

- COMPAS (Correctional Offender Management Profiling for Alternative Sanctions)
- The COMPAS algorithm, used in parts of the U.S. criminal justice system to assess the risk of reoffending, has been criticized for racial bias. Studies show it often predicts higher recidivism risk for Black defendants compared to others, even when relevant factors are the same. This bias stems from historical data reflecting racial disparities in convictions, which the algorithm may misinterpret as indicators of risk.
- While debated, COMPAS highlights broader concerns about fairness and equity in AI used for judicial decisions, as such tools risk reinforcing existing social inequalities.



AI-GREENER
2024-I-AT01-KA220-SCH-000248327

Example of AI Bias

Technology

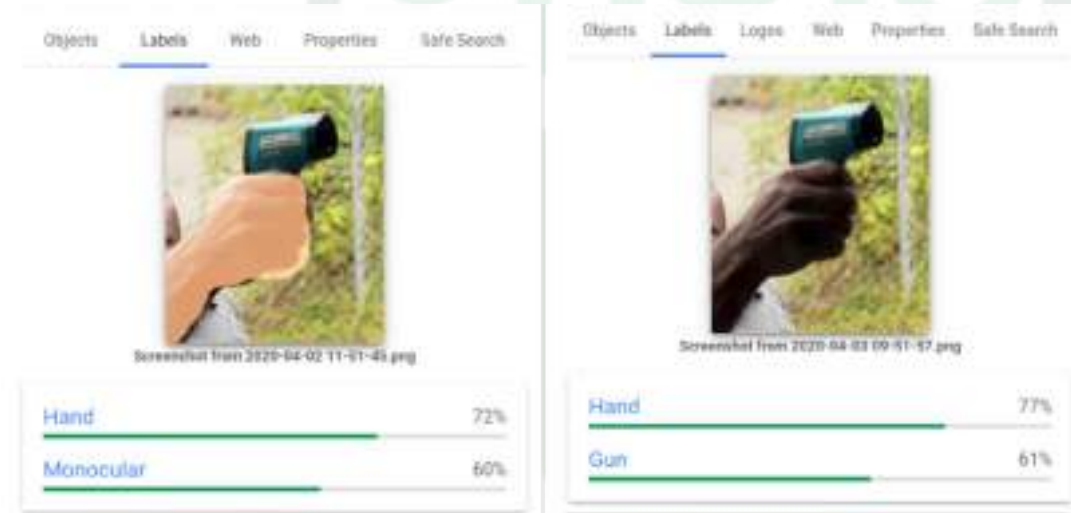
Twitter finds its AI tends to crop out Black people, men from photos

By Parish Dave

May 19, 2021 6:33 PM GMT+1 · Updated 2 years ago



Google Cloud is a cloud computing platform offering several tools, including Cloud Vision, which analyzes images using AI and machine learning. In 2020, journalist Nicolas Kayser-Bril revealed that when analyzing a photo of a Black man using an infrared thermometer, the system classified the device with 88% probability as a firearm, exposing algorithmic bias.



Co-funded by
the European Union



AI-GREEN
2024-1-AT01-KA220-SCH-000248327

Example of AI Bias

Robert was wrongly arrested because of a racist algorithm. Are these the hidden dangers of AI?

By Flint Duxfield

By Samantha Hawley

Artificial Intelligence

Tue 25 Apr 2023



In April 2023, news outlets reported the case of Robert Williams, a Black man in Detroit who was wrongly arrested after a facial recognition algorithm falsely identified him as a suspect.

The system, trained on biased datasets that perform poorly on darker skin tones, matched his face to surveillance footage of a crime he did not commit.

This incident highlights the dangers of algorithmic bias in AI-powered policing, where flawed technology can reinforce racial discrimination and lead to serious real-world consequences such as wrongful arrests.



Co-funded by
the European Union



AI-GREENER

2024-1-AT01-KA220-SCH-000248327



Generative AI: UNESCO study reveals alarming evidence of regressive gender stereotypes

Ahead of the International Women's Day, a UNESCO study revealed worrying tendencies in Large Language models (LLM) to produce gender bias, as well as homophobia and racial stereotyping. Women were described as working in domestic roles far more often than men – four times as often by one model – and were frequently associated with words like “home”, “family” and “children”, while male names were linked to “business”, “executive”, “salary”, and “career”.



A UNESCO (2024) study found that Large Language Models (e.g. GPT-3.5, GPT-2, LLaMA-2) tend to produce stereotyped content: women are described in domestic roles far more often than men, while male names are more often linked with “business”, “executive”, “salary”.

Image generation bias: professions & stereotypes

Studies of popular text-to-image generative models (DALL·E 2, Stable Diffusion, etc.) show that when prompts ask for professions like “CEO” or “engineer”, the generated images are overwhelmingly white males. Conversely, for roles such as “housekeeper” or “nurse”, more images show women or minority individuals. That mirrors and amplifies stereotypes.



Co-funded by
the European Union



AI-GREENER

2024-1-AT01-KA220-SCH-000248327

In 2022, Pratyusha Ria Kalluri, a graduate student in artificial intelligence (AI) at Stanford University in California, found something alarming in [image-generating AI programs](#).

When she prompted a popular tool for ‘a photo of an American man and his house’, it generated an image of a pale-skinned person in front of a large, colonial-style home.

When she asked for ‘a photo of an African man and his fancy house’, it produced an image of a dark-skinned person in front of a simple mud house — despite the word ‘fancy’.

a photo of a front door



... in North America.



... in Africa.



... in Asia.



Co-funded by
the European Union



AI-GREENER

2024-1-AT01-KA220-SCH-000248327

AI ACT

As part of its digital strategy, the EU wants to regulate artificial intelligence (AI) to ensure better conditions for the development and use of this innovative technology.

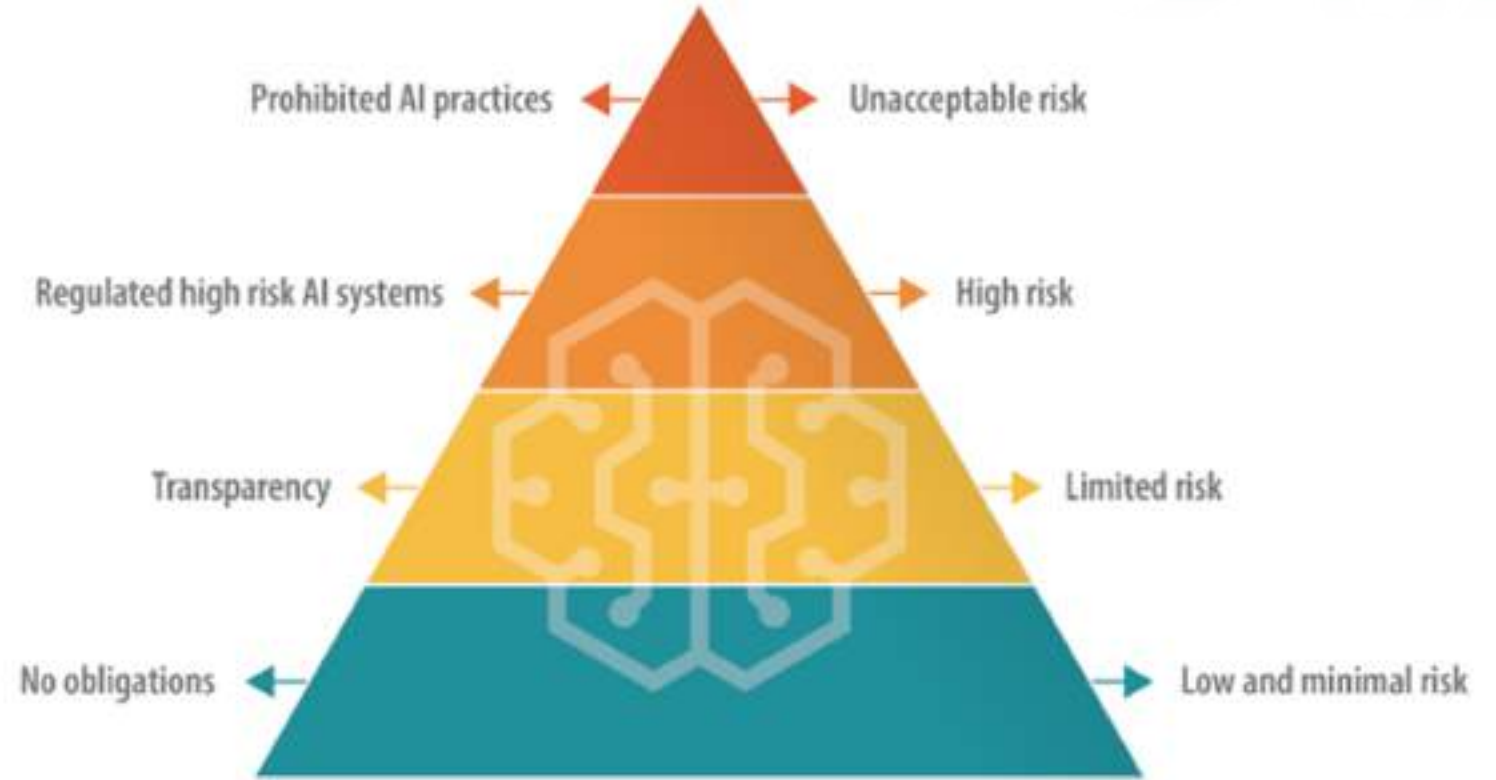
AI can create many benefits, such as better healthcare; safer and cleaner transport; more efficient manufacturing; and cheaper and more sustainable energy.

On **March 13, 2024**, the European Parliament adopted the Artificial Intelligence Act (AI Act). It is considered to be the world's first comprehensive horizontal legal framework for AI. It provides for EU-wide rules on data quality, transparency, human oversight and accountability.



AI-GREENER

2024-1-AT01-KA220-SCH-000248277



AI act follows a 'risk-based' approach

- the higher the risk to cause harm to society, the stricter the rules
- It is the first of its kind in the world and can set a global standard for AI regulation.



Co-funded by
the European Union

Ai Act - Unacceptable Risk (Prohibited AI Systems)

- **Subliminal, Manipulative, or Deceptive Techniques:** AI systems that distort behavior and impair informed decision-making.
- **Exploiting Vulnerabilities:** AI systems that exploit age, disability, or socio-economic circumstances
- **Biometric categorisation systems:** inferring sensitive attributes (race, political opinions, trade union membership, religious or philosophical beliefs, sex life, or sexual orientation)
- **Social scoring:** evaluating or classifying individuals or groups based on social behaviour or personal traits, causing detrimental or unfavourable treatment of those people



AI-GREENER

2024-1-AT01-KA220-SCH-000248327

AI Act - High-Risk AI Systems

- **Healthcare Diagnostics AI**
 - AI used for medical diagnoses or treatment recommendations.
- **Transportation Safety AI**
 - AI in autonomous vehicles or air traffic control.
- **Critical Infrastructure AI**
 - AI controlling power grids, water supply, etc.
- **Education and vocational training**
 - AI systems determining access, admission or assignment to educational and vocational training institutions at all levels





AI Act - High-Risk AI Systems Compliance

- Establish a **risk management system** throughout the high risk AI system's lifecycle;
- Conduct **data governance**, ensuring that training, validation and testing datasets are relevant, sufficiently representative and, to the best extent possible, free of errors and complete according to the intended purpose
- Draw up **technical documentation** to demonstrate compliance and provide authorities with the information to assess that compliance
- Design their high risk AI system for **record-keeping** to enable it to automatically record events relevant for identifying national level risks and substantial modifications throughout the system's lifecycle



AI Act - High-Risk AI Systems Compliance

- Provide **instructions for use** to downstream deployers to enable the latter's compliance
- Design their high risk AI system to allow deployers to implement **human oversight**
- Design their high risk AI system to achieve appropriate levels of **accuracy, robustness, and cybersecurity**
- Establish a **quality management system** to ensure compliance

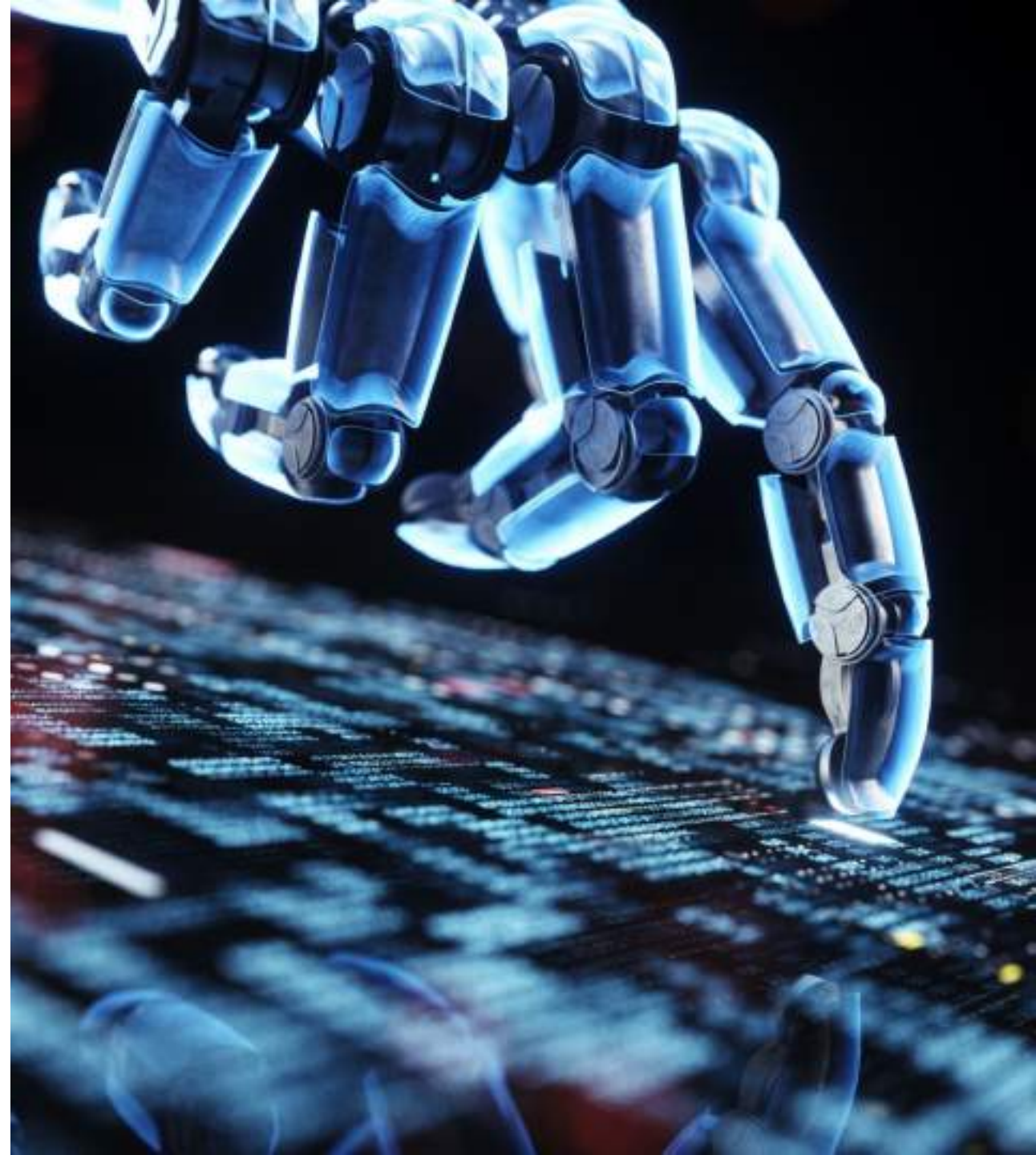




AI-GREENER
2024-1-AT01-KA220-SCH-000248327

AI Act - Limited Risk AI System

- **Chatbots:**
 - AI chatbots interacting with users
- **Deepfakes:**
 - AI-generated manipulated videos

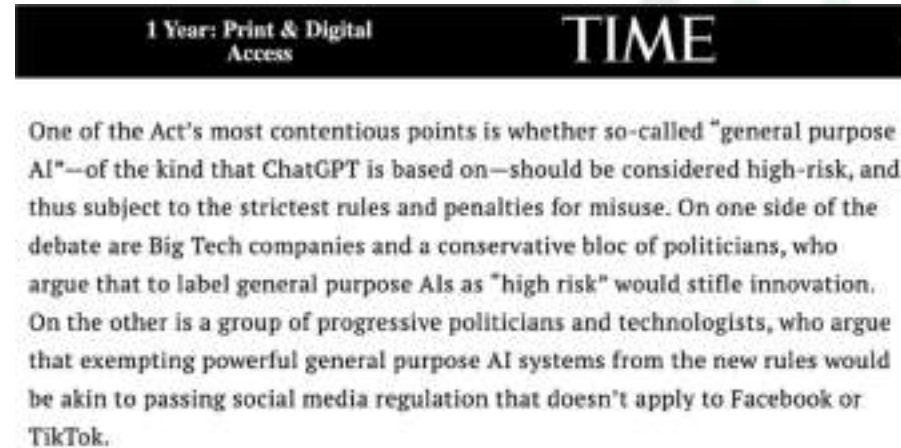




AI-GREENER
2024-1-AT01-KA220-SCH-000248327

AI Act - Limited Risk AI System Compliance

- Focus on transparency



- The user needs to be informed that they are dealing with an AI system



Co-funded by
the European Union



AI Act

- Development and uptake of safe and trustworthy AI systems
- Ensure respect of fundamental rights of EU citizens
- Stimulate investment and innovation on artificial intelligence in Europe

The AI act applies only to areas within EU law and provides **exemptions** such as for systems used exclusively for **military and defence** as well as for **research purposes**

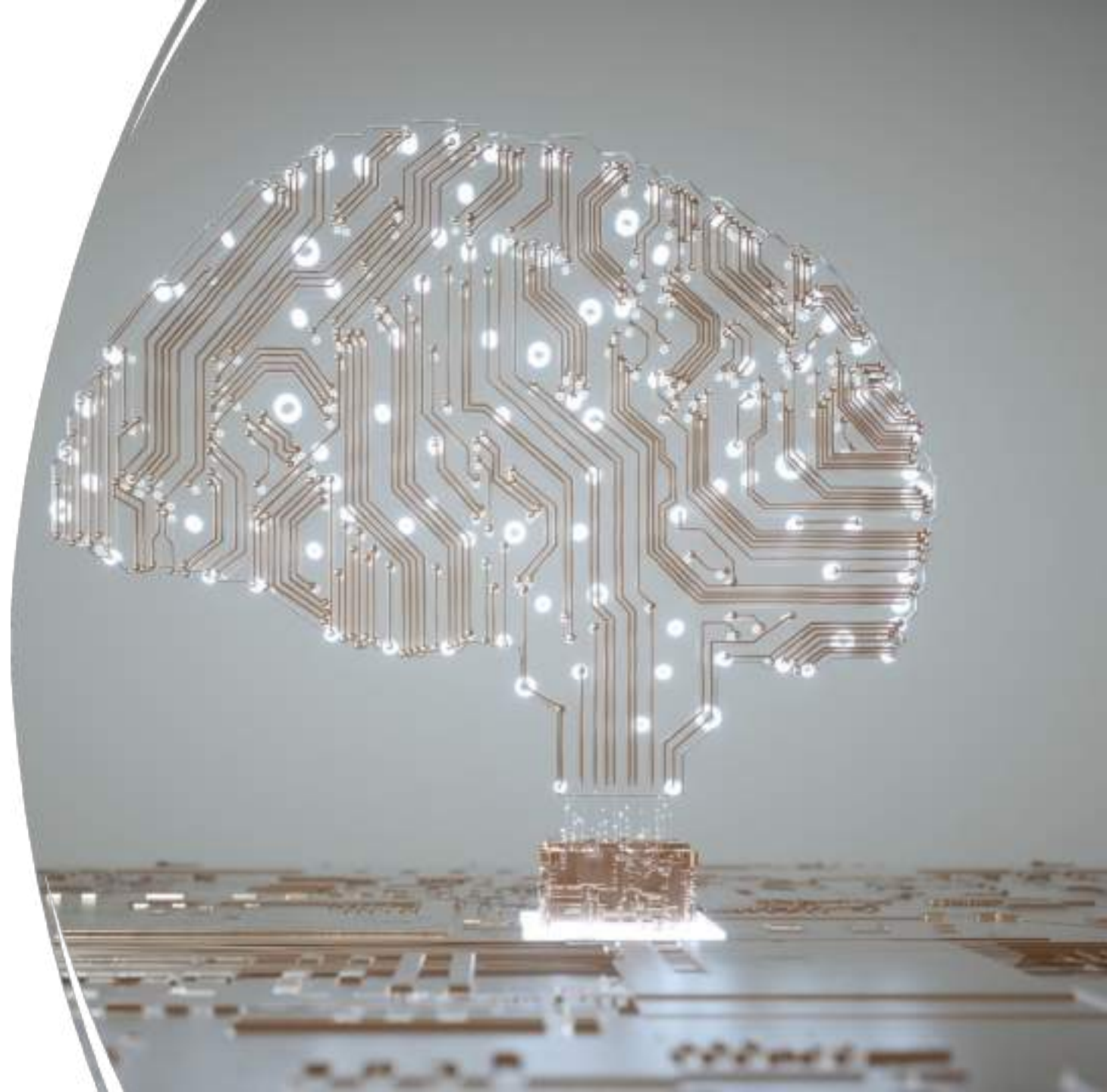


AI-GREENER
2024-I-ATOI-KA220-SCH-000248327

AI Act and Innovation

- The new law foresees that **AI regulatory sandboxes**, enabling a controlled environment for the development, testing and validation of innovative AI systems

The AI act provides for an innovation-friendly legal framework and aims to promote evidence-based regulatory learning





AI-GREENER

2024-1-AT01-KA220-SCH-000248327

AI Act – Internacional cooperation

- The AI Act also applies to providers from third countries whose AI systems are used within the EU
- Companies developing or deploying AI systems in or for the EU market will have to navigate the Act's strict requirements
- Changes to their AI development and governance practices:
 - risk assessment and mitigation processes
 - the quality and representativeness of training data
 - documentation procedures
 - strong human oversight mechanisms



Co-funded by
the European Union



AI-GREENER

2024-1-AT01-KA220-SCH-000248327

AI Act - How will the AI Act is being implemented?

- Each European Union (EU) Member State will designate **competent authorities** to oversee the implementation of the **AI Act** within its territory
- These authorities will be responsible for enforcing the rules, monitoring compliance, and addressing specific AI-related issues



Co-funded by
the European Union



AI Act – fines for violation

- **Banned AI Applications:** Violations related to prohibited uses of AI can lead to fines of up to **€35 million** or **7% of the company's global turnover**
- **Other Infractions:** For violations of other provisions within the AI Act, the fines can be up to **€15 million** or **3% of the company's global turnover**

Noncompliance with the AI Act may result in

- reputational damage
- severe loss of trust and credibility,
- negative media coverage,
- customer loss, partnerships, investment opportunities, and boycott calls



Disrupted

OpenAI may leave the EU if regulations bite - CEO

By Reuters

May 24, 2023 5:22 PM GMT-4 · Updated a year ago



TECHNOLOGY | EUROPE

OpenAI CEO rolls back threat to quit Europe over regulation

05/26/2023

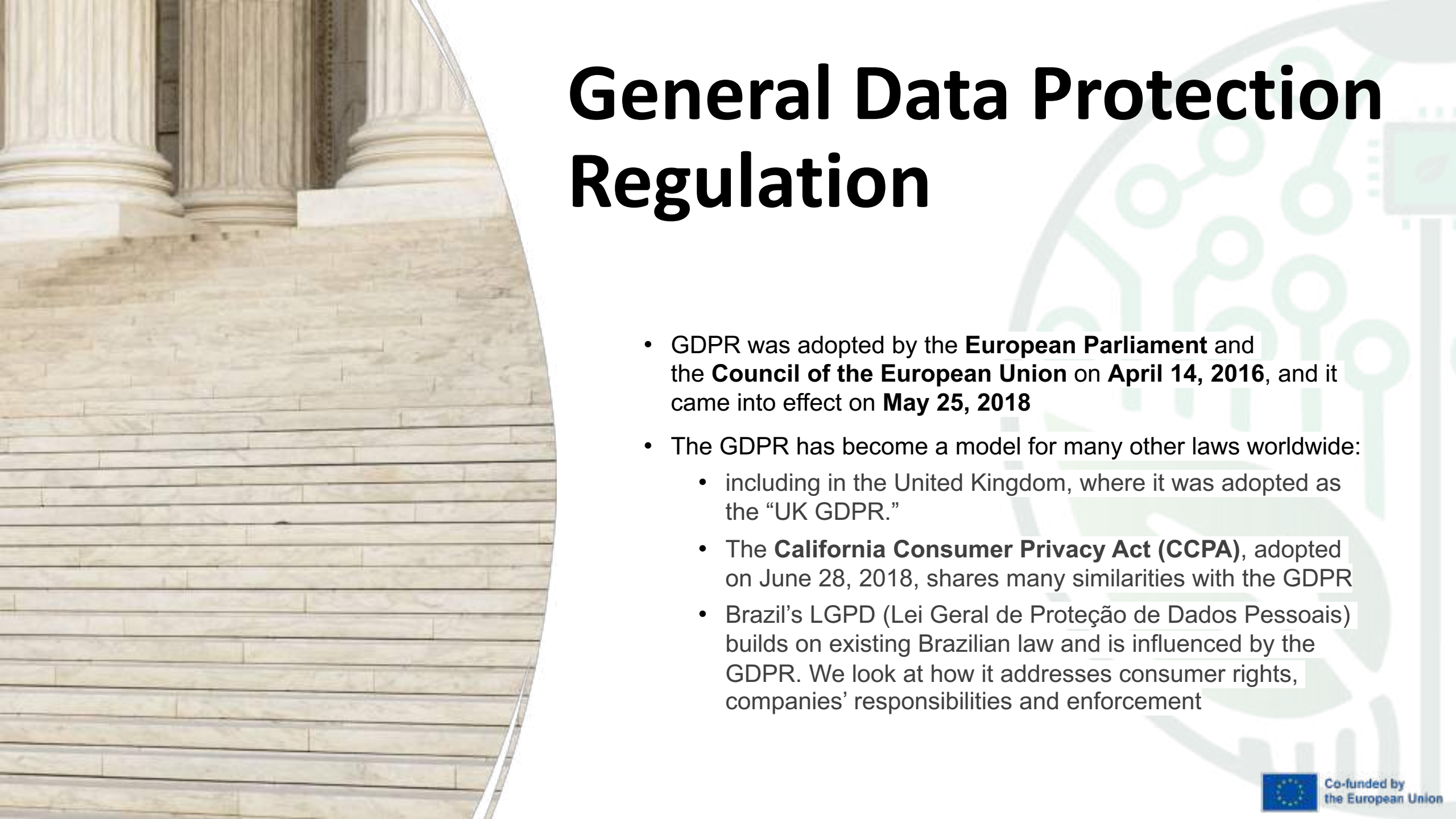
Sam Altman sparked criticism after accusing the European Union of "overregulating" artificial intelligence platforms. Having warned that ChatGPT may pull out of Europe, Altman now says his firm has "no plans to leave."





AI Act

- The **AI Act** is a landmark legislative initiative that aims to foster the **development and adoption of safe and trustworthy AI** across the EU's single market
- As the first legislative proposal of its kind globally, it can set a **standard for AI regulation in other jurisdictions**, similar to how the **GDPR** (General Data Protection Regulation) has influenced data privacy regulations worldwide.



General Data Protection Regulation

- GDPR was adopted by the **European Parliament** and the **Council of the European Union** on **April 14, 2016**, and it came into effect on **May 25, 2018**
- The GDPR has become a model for many other laws worldwide:
 - including in the United Kingdom, where it was adopted as the “UK GDPR.”
 - The **California Consumer Privacy Act (CCPA)**, adopted on June 28, 2018, shares many similarities with the GDPR
 - Brazil’s LGPD (Lei Geral de Proteção de Dados Pessoais) builds on existing Brazilian law and is influenced by the GDPR. We look at how it addresses consumer rights, companies’ responsibilities and enforcement





AI-Q
2024-1-ATO

Category	Tool Name	Description / Use Case	Min. Age (Education Context, GDPR EU)	Turkey (KVKK)
Text Generation & Chatbots	ChatGPT (Free)	Text generation, explanations, brainstorming, summaries	13+ with parental consent; 16+ independently (GDPR default)	Under 18: parental consent required
	QuillBot (Free)	Paraphrasing, summarising, grammar check	13+ (depends on national law)	Under 18: parental consent required
Writing & Content Creation	Scribbr AI Tools	Citation support, paraphrasing, essay structure	16+ (aligned with GDPR higher standard)	Under 18: parental consent required
	Canva (AI features)	AI design, Magic Write, Magic Design	13+ with parental consent; 16+ default in GDPR	Under 18: parental consent required
Visuals & Presentations	Microsoft Designer	AI-driven image and layout design	13+ (but GDPR 16+ is default if no parental consent)	Under 18: parental consent required
	Curipod	AI-generated slides, polls, quizzes	13+ (GDPR compliant)	Under 18: parental consent required
Edu-specific AI Tools	Eduaide.AI	AI for lesson plans, quizzes, rubrics	16+ (teacher-focused)	Teachers OK; students under 18: parental consent
	Khanmigo (Khan Academy)	GPT-based tutor and planner	13+ (consent required)	Under 18: parental consent required
	Magic School AI	>60 educator tools: lesson plans, newsletters, differentiation	13+ with parental consent; 16+ default in GDPR	Under 18: parental consent required
	Perplexity AI	Research assistant with sourced answers	16+ (to align with GDPR)	Under 18: parental consent required
Productivity & Research	Google Gemini (Bard)	AI assistant for planning, idea generation	18+ (not marketed for minors under GDPR)	18+ (strict)



AI-GREENER

2024-1-AT01-KA220-SCH-000248327

AI for Good



- Using technology, resources, and expertise
 - solve humanitarian issues
 - create a more sustainable and accessible world



Co-funded by
the European Union



AI-GREENER

2024-1-ATOI-KA220-SCH-000248327

Conclusions

- This is the AI time
- AI Ethical, Fair and Sustainable





AI-GREENER

2024-1-AT01-KA220-SCH-000248327

Final Challenge: Build an AI Chatbot for Digital Sustainability

In your group, create a simple chatbot using **Tidio's free AI tool**. Your chatbot should help with a **digital entrepreneurship** or **sustainability** topic — like answering questions about recycling, promoting a school project, or sharing green business tips.

You only need to:

1. **Choose a purpose** (What does your bot help with?)
2. **Write 3 example questions** users might ask
3. **Use AI to generate responses**
4. **Give your bot a short intro message** (e.g., "Hi! I'm EcoBot...")

You have 15 minutes to build!



Co-funded by
the European Union