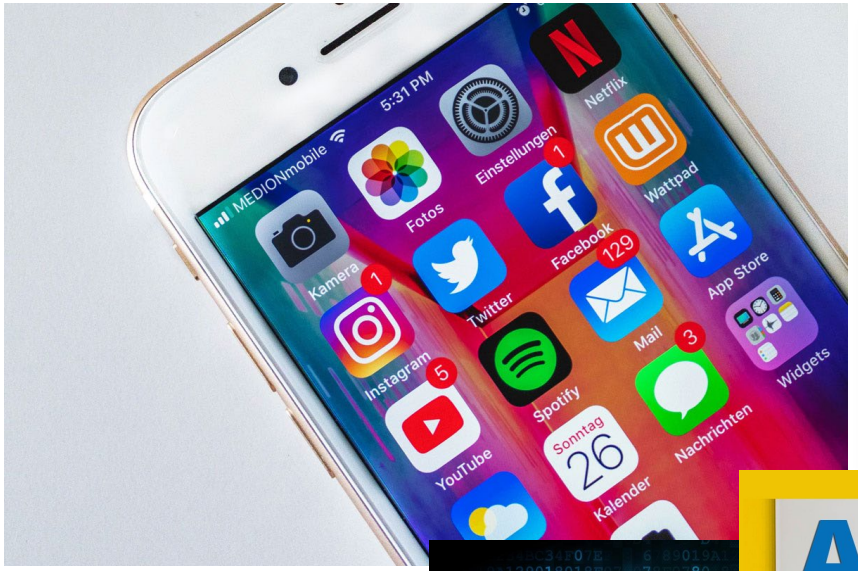
Evaluación de proyectos con IA en los CEI

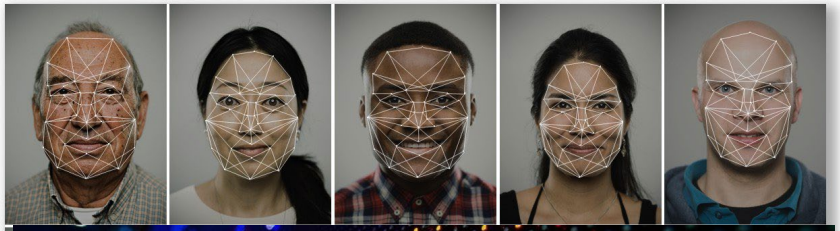
Observatori d'Ètica en Intel·ligència Artificial de Catalunya
Universitat de Girona

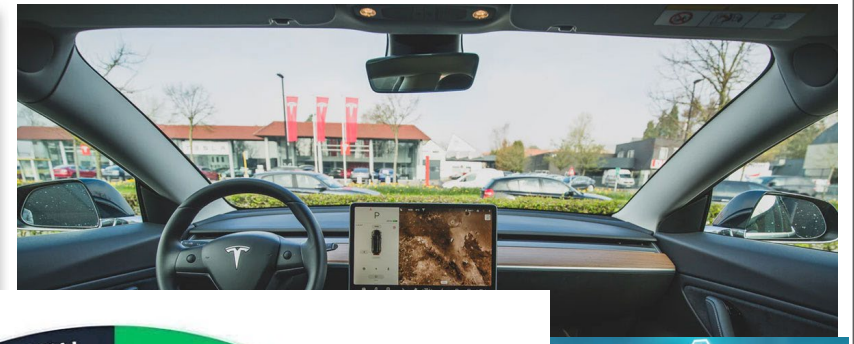
Albert Sabater Coll

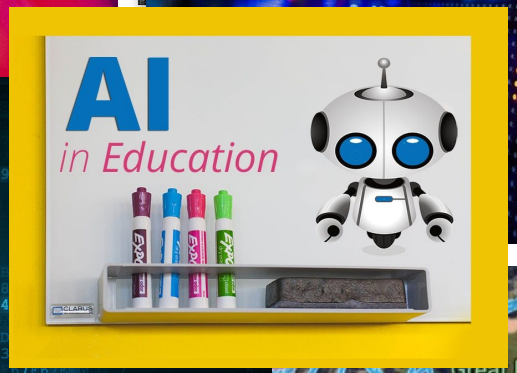
OEIAC | Universitat de Girona

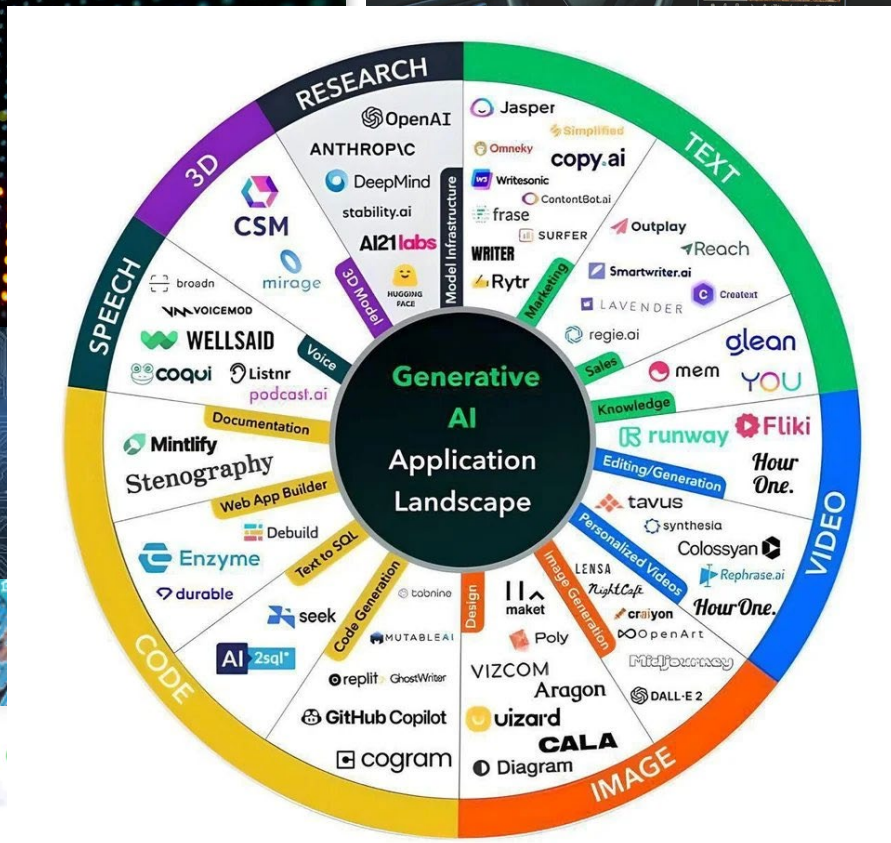
Generalitat de Catalunya

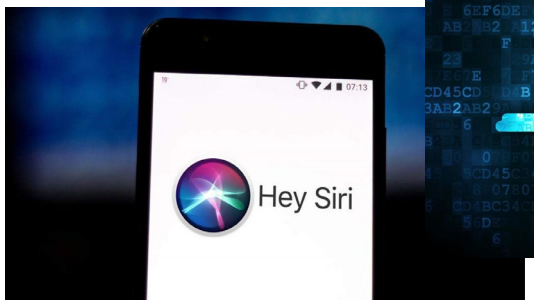

MEDIONmobile
5:31 PM
Netflix
Einstellungen
Fotos
Wattpad
Facebook
Kamera
App Store
Twitter
Mail
Instagram
Spotify
Widgets
YouTube
Sonntag
26
Kalender
Nachrichten

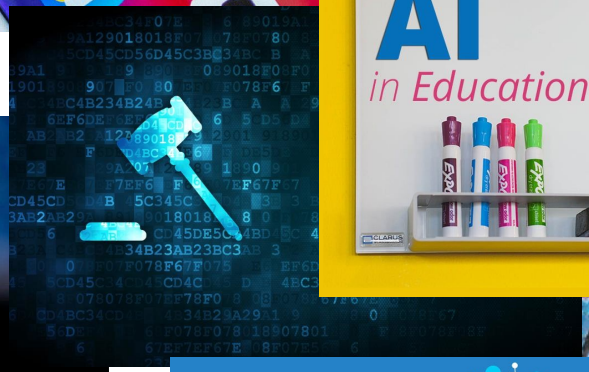





AI
in Education


Generative AI Application Landscape
RESEARCH
TEXT
VIDEO
IMAGE
CODE
SPEECH
3D
OpenAI
ANTHROPIC
DeepMind
stability.ai
AI21labs
Jasper
copy.ai
Writesonic
ContentBot.ai
frase
SURFER
WRITER
Rytr
Outplay
Reach
Smartwriter.ai
LAVENDER
Creatext
regie.ai
glean
mem
YOU
runway
Fliki
Hour One.
tavus
synthesia
Colossyan
Rephrase.ai
LENSA
NightCafe
craiyon
OpenArt
Midjourney
DALL·E 2
VIZCOM
Aragon
uizard
CALA
Diagram
maket
Poly
tabnine
MUTABLE AI
replit
GhostWriter
GitHub Copilot
cogram
seek
AI 2sql
Enzyme
durable
Debuild
Mintlify
Stenography
Documentation
Web App Builder
Text to SQL
Code Generation
Design
Image Generation
Personalized Videos
Editing/Generation
Knowledge
Sales
Marketing
Model Infrastructure
3D Model
Voice
CSM
mirage
broadn
VOICEMOD
WELLSAID
coqui
Listnr
podcast.ai
HUGGING FACE

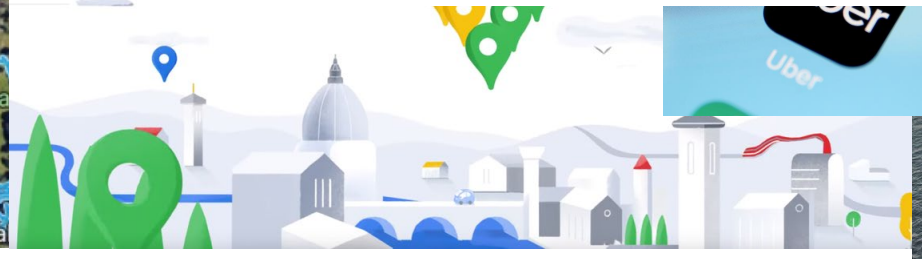

Hey Siri

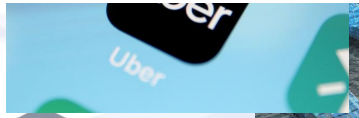


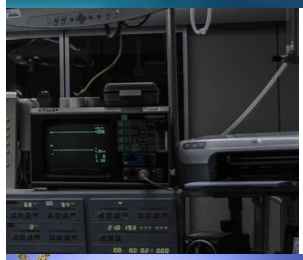

amazon

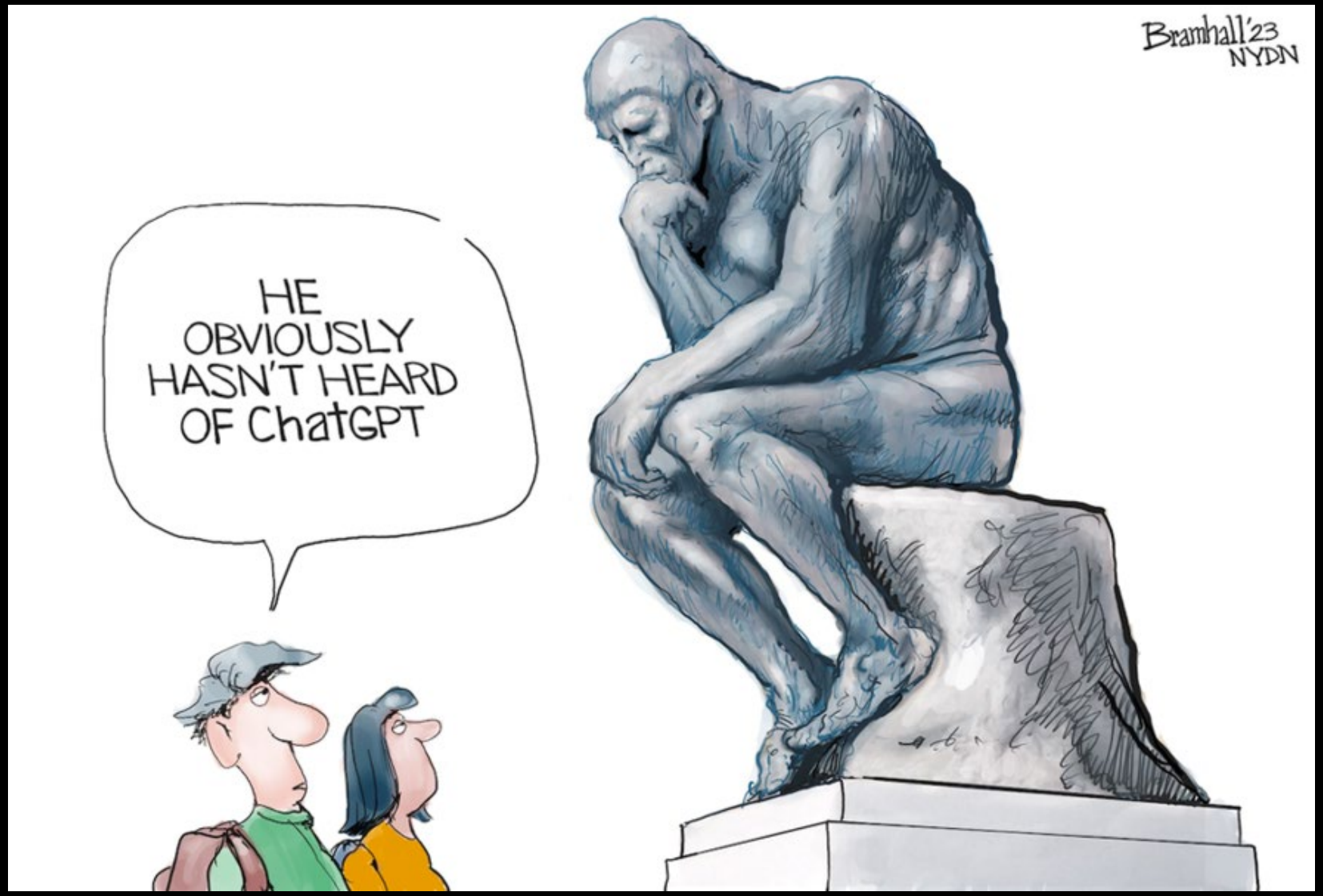

JUST ASK
amazon alexa
amazon


Sheep Mea
Centra




Uber





Fuente: https://www.dallasnews.com/opinion/commentary/2023/02/23/editorial-cartoon-chatgpt/

HIGHER EDUCATION

Everyone Is Cheating Their Way Through College ChatGPT has unraveled the entire academic project.

By James D. Walsh, Intelligencer staff writer

Journal of Marketing

AMA AMERICAN MARKETING ASSOCIATION

Impact Factor: **11.5** / 5-Year Impact Factor: **15.0**

Restricted access | Research article | First published online January 13, 2025

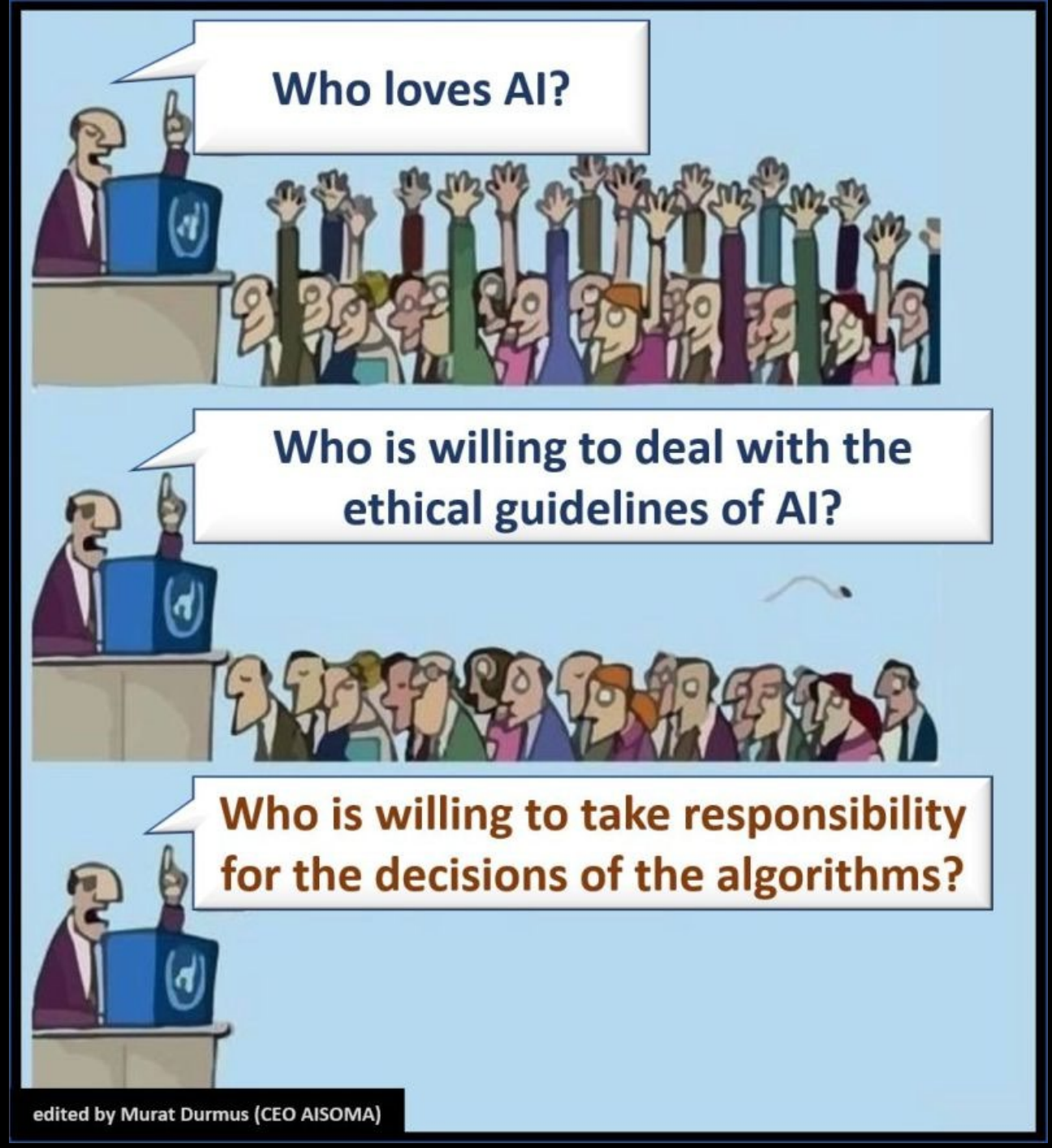
EXPRESS: Lower Artificial Intelligence Literacy Predicts Greater AI Receptivity

Stephanie Tully, Chiara Longoni, and Gil Appel View all authors and affiliations

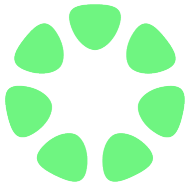
Accepted Manuscripts | https://doi.org/10.1177/00222429251314491

Abstract

As artificial intelligence (AI) transforms society, understanding factors that influence AI receptivity is increasingly important. The current research investigates which types of consumers have greater AI receptivity. Contrary to expectations revealed in four surveys, cross country data and six additional studies find that people with lower AI literacy are typically more receptive to AI. This lower literacy-greater receptivity link is not explained by differences in perceptions of AI's capability, ethicality, or feared impact on humanity. Instead, this link occurs because people with lower AI literacy are more likely to perceive AI as magical and experience feelings of awe in the face of AI's execution of tasks that seem to require uniquely human attributes. In line with this theorizing, the lower literacy-higher receptivity link is mediated by perceptions of AI as magical and is moderated among tasks not assumed to require distinctly human attributes. These findings suggest that companies may benefit from shifting their marketing efforts and product development towards consumers with lower AI literacy. Additionally, efforts to demystify AI may inadvertently reduce its appeal, indicating that maintaining an aura of magic around AI could be beneficial for adoption.


Who loves AI?
Who is willing to deal with the ethical guidelines of AI?
Who is willing to take responsibility for the decisions of the algorithms?
edited by Murat Durmus (CEO AISOMA)

IA responsable: 3 ingredientes

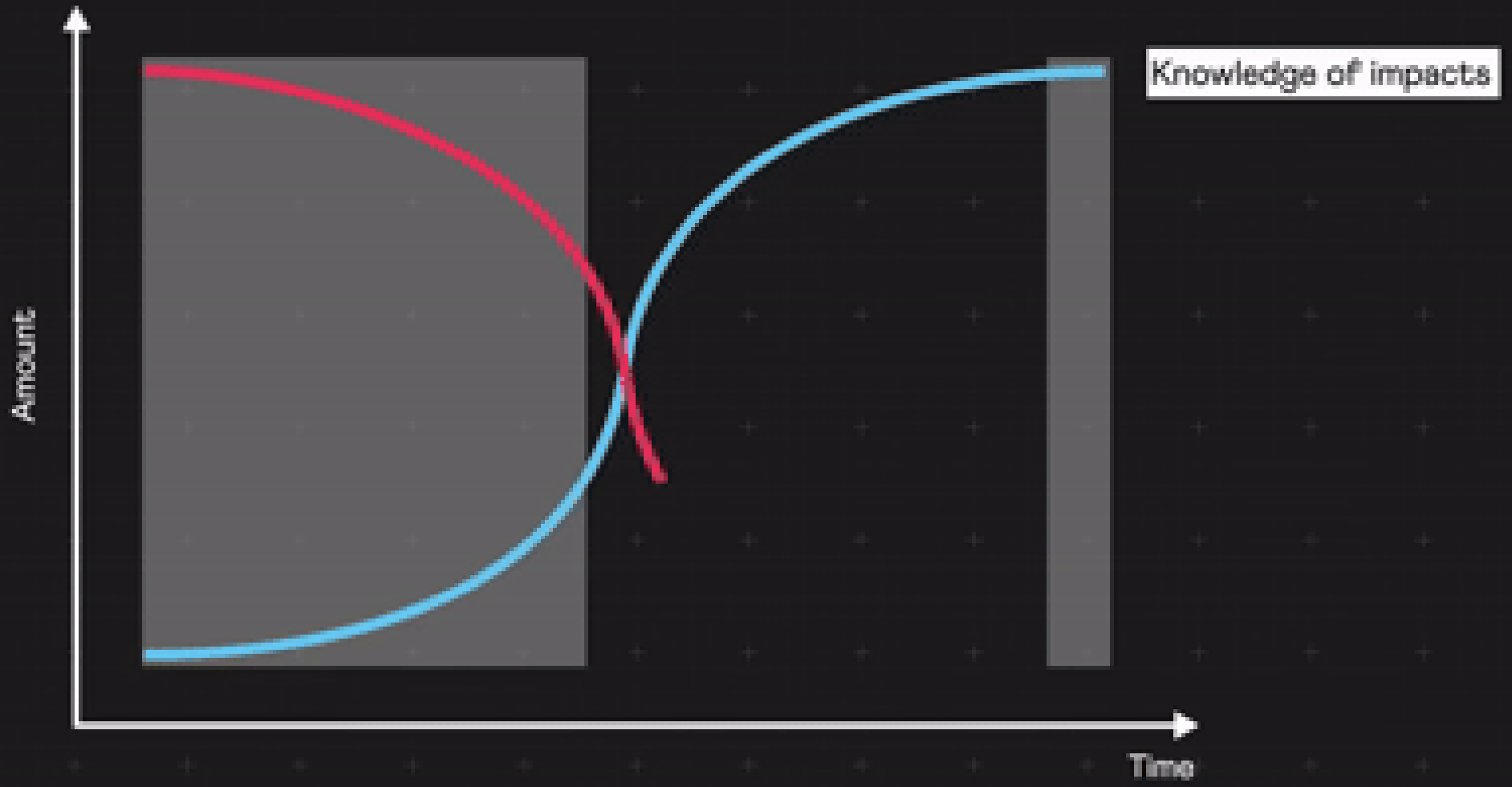


1. Es necesario conocer los aspectos legales y éticos (¡no son un complemento!).

2. Es necesario realizar un seguimiento y una validación interna y externa.

3. Es necesario preguntar (visión crítica) y documentar las elecciones y sus impactos.

The Collingridge Dilemma - When Do You Regulate Technology?



Artificial Intelligence Act: MEPs adopt landmark law

Press Releases PLENARY SESSION IMCO LIBE 13-03-2024 - 12:25

AI Act/RIA e investigación (I)

- La investigación queda excluida de la aplicación del RIA.

Art. 2.6: *This Regulation does not apply to AI systems or AI models, including their output, specifically developed and put into service for the sole purpose of scientific research and development.*

- No sólo se refiere al desarrollo sistemas de IA sin restricciones concretas, también a la utilización de sistemas de IA en principio no permitidos si se realiza con fines científicos.

AI Act/RIA e investigación (II)

- La exclusión finaliza en el momento en que el sistema desarrollado sea introducido en el mercado o puesto en servicio.

Art. 3.8: *This Regulation does not apply to any research, testing or development activity regarding AI systems or AI models prior to their being placed on the market or put into service. Such activities shall be conducted in accordance with applicable Union law. Testing in real world conditions shall not be covered by that exclusion.*

AI Act/RIA e investigación (III)

- Aun cuando la investigación queda excluida del RIA, si se recomienda seguir las directrices establecidas en los códigos de conducta voluntaria.

Art. 95: *1. The AI Office and the Member States shall encourage and facilitate the drawing up of codes of conduct, including related governance mechanisms, intended to foster the voluntary application to AI systems, other than high-risk AI systems, of some or all of the requirements set out in Chapter III, Section 2.*

Innovación vs regulación

- Es necesario salir del falso debate sobre innovación vs regulación.
- La regulación son trampolines para la innovación tecnológica (no es necesario utilizar la tecnología existente "tal y como es").
- La regulación siempre impulsa la transformación tecnológica, garantizando la aceptación social y empujando a la diferenciación competitiva entre organizaciones.

Principales temas a tratar

Uso responsable y ético

- Entender las limitaciones
- Tratamiento ético de los datos
- Mantener la integridad
- Preservar la confianza

Aprendizaje y accesibilidad

- Utilitzar como complemento
- Accessibilidad condicionada
- Promover el feedback
- Proporcionar cursos

MIT News
ON CAMPUS AND AROUND THE WORLD

Explained: Generative AI's environmental impact

Rapid development and deployment of powerful generative AI models comes with environmental consequences, including increased electricity demand and water consumption.

Adam Zewe | MIT News
January 17, 2025

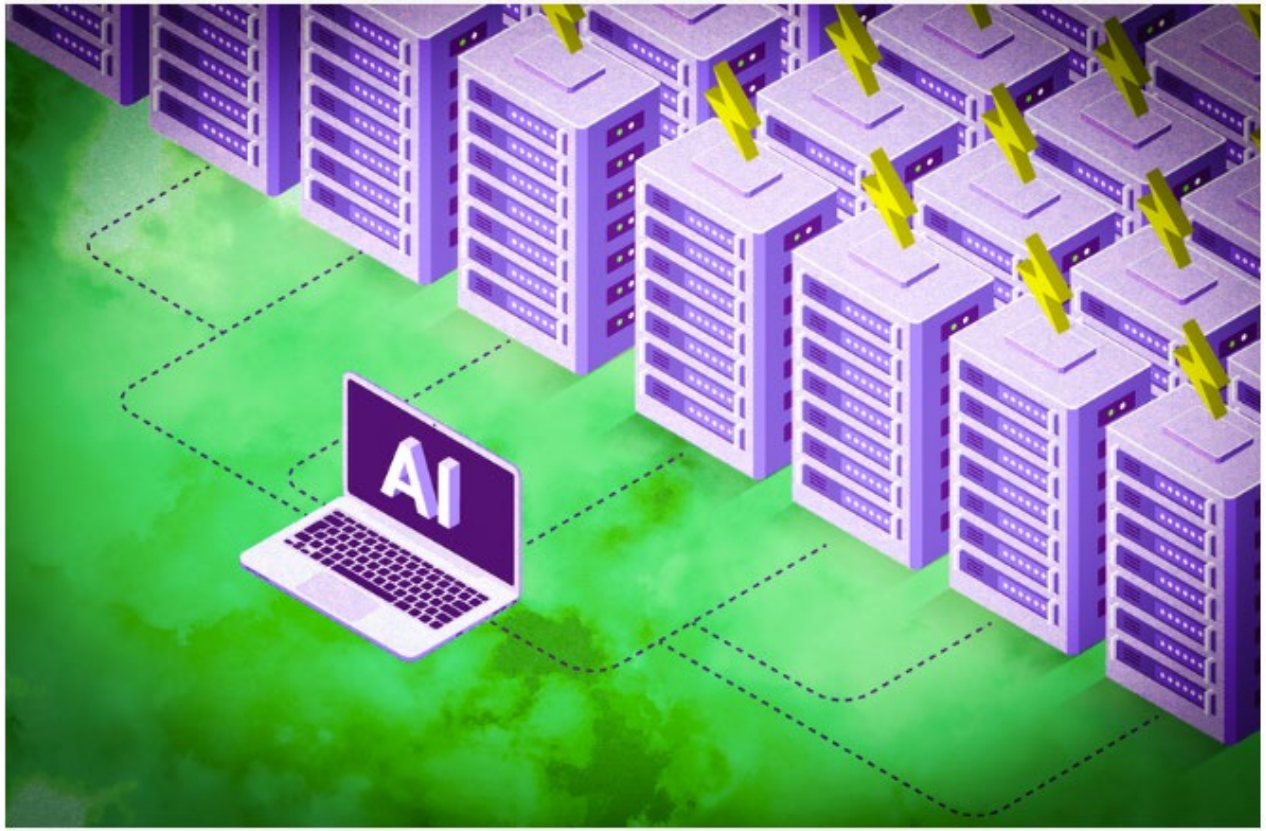


Image credit to MIT News

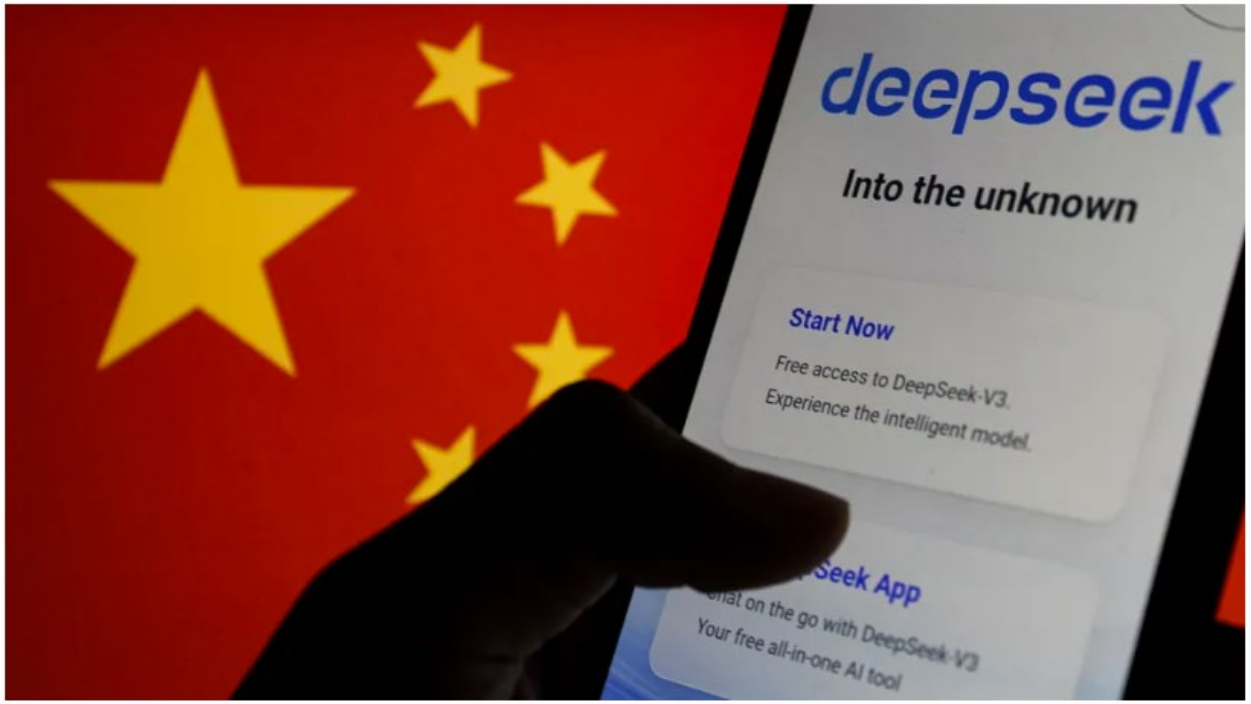
China will enforce clear flagging of all AI generated content starting from September

News By Jowi Morales published 3 days ago

AI text, audio, video, images, and even virtual scenes will all need to be labeled.

Comments (16)

When you purchase through links on our site, we may earn an affiliate commission. Here's how it works.



(Image credit: Shutterstock)

Image credit to Tom's Hardware

Preocupaciones y progreso

- Les preocupaciones se centran en 3 tipos de progresso de la IA:

1. **Progreso genuino y rápido en el ámbito de la percepción / imágenes**

   Preocupaciones a causa de la precisión (distinción entre lo real / falso)

2. **Progreso y mejora en las áreas de automatización de criterios**

   Preocupaciones a causa de errores automatizados (algunos inevitables)

3. **Progreso dudoso en la predicción de fenómenos sociales**

   Preocupaciones por inexactitud y subjetividad (disfrazada de objetividad)

Preocupaciones específicas (I)

- **Puntos ciegos y posibles sesgos**. Todos los modelos de IA actuales se definen completamente por sus datos de entrenamiento y, por lo tanto, perpetúan las omisiones y sesgos de dichos datos.

- **Validación y responsabilidad**. Existe el riesgo de que los sistemas sean lo suficientemente buenos como para que los usuarios se vuelvan confiados y complacientes, pero lo suficientemente malos como para que surjan problemas graves con consecuencias profundas.

Preocupaciones específicas (II)

- **Transparencia y documentación**. Mantener la transparencia en la documentación y los informes seguirá siendo fundamental para garantizar la reproducibilidad y replicabilidad de los hallazgos de investigación generados con IA.

- **Privacidad y protección de datos**. No se debe asumir a priori que las herramientas de IA sean privadas o seguras. Los usuarios deben comprender los riesgos potenciales asociados a la introducción de datos sensibles, privados, confidenciales o de propiedad exclusiva en estas herramientas.

Parece obvio pero no lo es

- Los sistemas de IA no pueden responder a preguntas cuyas respuestas no estén de alguna forma en su conjunto de datos de formación.
- Los sistemas de IA no pueden resolver problemas sobre los que no han sido entrenados. Esto no sólo incluye los datos sino la arquitectura.
- Los sistemas de IA no pueden adquirir nuevas habilidades sin nuestra ayuda. Esto incluye desde la recogida de información al etiquetado y análisis.
- Los sistemas de IA no pueden inventar cosas nuevas.

Google

images of ai

All | Images | Videos | Books | News | More | Tools | Saved | SafeSearch

artificial intelligence | future | technology | machine learning | logo | human | creative | animated | design | computer | deep learning | clip art

iStock
321,500+ Ai Stock Photos, Pictures ...


Analytics Insight
Artificial Intelligence – The His...

Simplilearn.com
Artificial Intelligence [AI ...


European Parliament
EU AI Act: first regulation on ...


TechRepublic
7 Best AI Art Generators of 2023


Simplilearn.com
7 Types of Artificial Intelligence That ...

ITChronicles
AI Applications Today: Where Arti...


HubSpot Blog
What Jobs Will AI Replace & W...

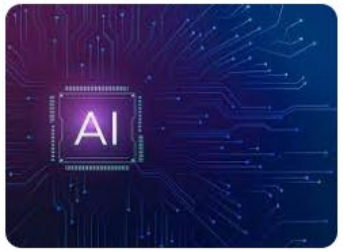

Forbes
Risks Posed By AI

Elegant Themes
Artificial Intelligence ...


IBM
What is generative AI, what are ...


The Economic Times
Artificial Intelligence: Four ski...

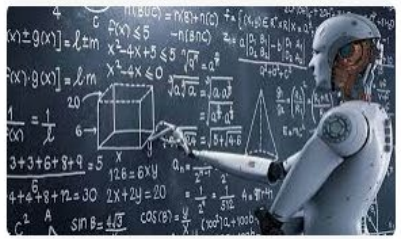

Forbes
The Future Of Artificial Intelligen...



Freepik
Ai Images - Free Download on Free...


ET Telecom
AI: Artificial intelligence ma...


Simplilearn.com
Artificial Intelligence? Types, ...


Columbia News - Columbia University
Will ChatGPT and AI Help or Harm Us ...


Chitkara University
What Is Artificial Intelligence And ...

Analytics Vidhya
Advantages and Disadvantages o...


BBC
BBC Newsround


The Conversation
AI and the future of work: 5 experts on ...

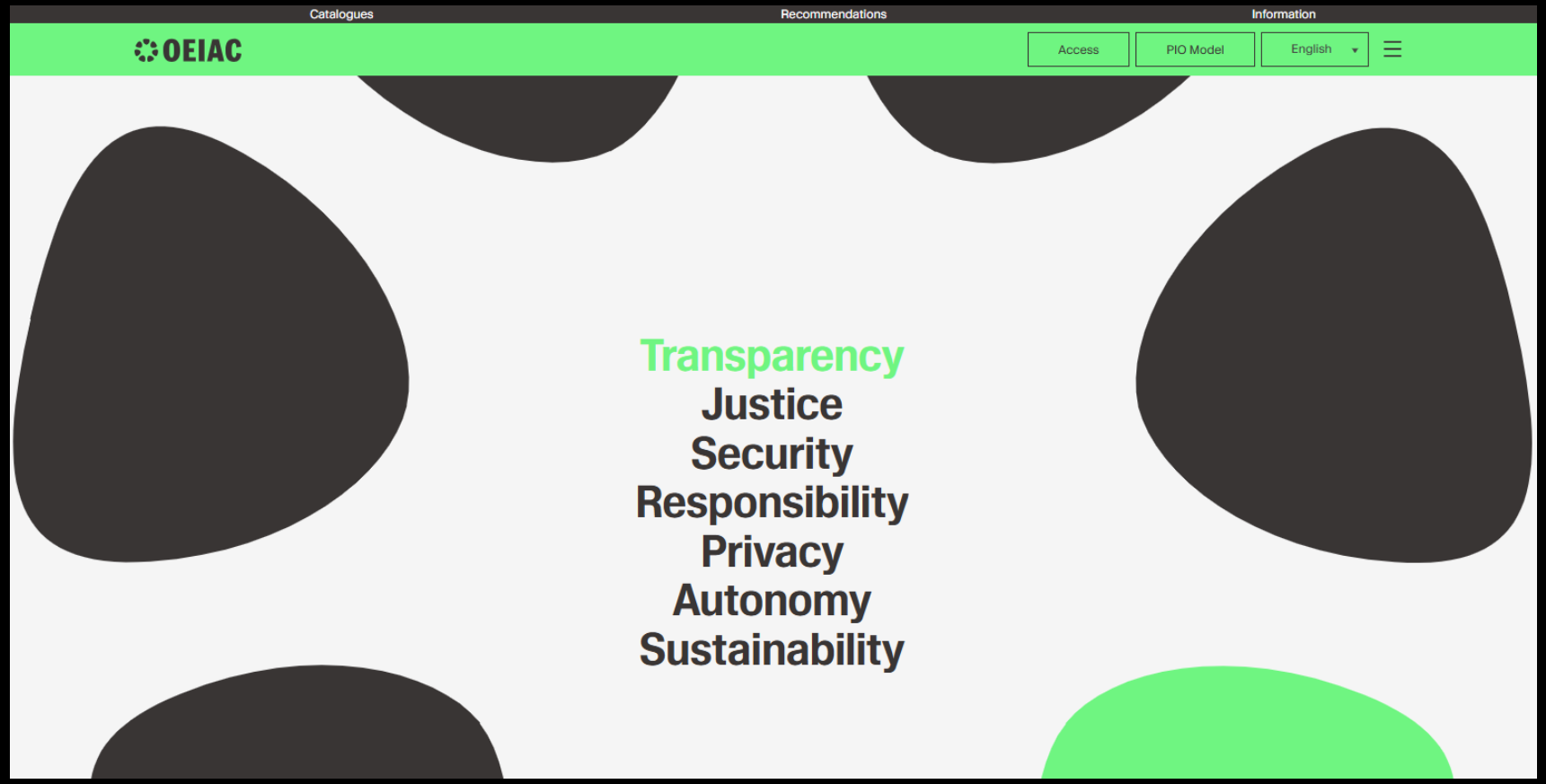
La imagen de la IA

- Describir los sistemas de IA con agencia independiente de la supervisión humana o dando a entender que pronto podrían hacerlo no es ético.
- Asumir que los algoritmos de IA aprenden de la misma forma que lo hacen los humanos es problemático ya que dan credibilidad a las afirmaciones de que la IA es realmente “inteligente”.
- Por ejemplo, son habituales y erróneas las comparaciones de algoritmos de aprendizaje profundo con la forma en la que funciona el cerebro humano.

Artificial Intelligence Act: MEPs adopt landmark law

Press Releases PLENARY SESSION IMCO LIBE 13-03-2024 - 12:25

PIO Model



www.oeiac.cat

1	**Pre-evaluation 1** To determine if the AI system is considered an unacceptable risk based on the AI Act regulations.
2	**Pre-evaluation 2** To classify the AI system according to whether its risk is high, limited, or minimal in accordance with the AI Act regulations.
3	**Pre-evaluation Results** Allows users of the PIO Model Rights and Obligations to position themselves on some of the main legal requirements of the AI Act regulations.
4	**Start of Evaluation** Through a checklist organized based on ethical principles, legal provisions, ethical recommendations, and examples.

4	**Start of Evaluation** Through a checklist organized based on ethical principles, legal provisions, ethical recommendations, and examples.
5	**Data Collection** Through the information provided by users of the PIO Model via the checklist based on ethical and legal grounds.
6	**Obtaining Metrics** Compliance with legal provisions and ethical recommendations of the PIO Model, taking into account the risk categories of the AI Act regulations.
7	**Results Visualization** Through a badge that reflects the degree of achievement of the ethical principles and legal provisions contained in the PIO Model.

Pre-assessment

Unacceptable risk:

This refers to a very limited set of uses of AI that are particularly harmful and contravene EU values, as they infringe fundamental rights and are therefore prohibited.

High risk

This refers to a limited number of AI systems that may have an adverse impact on the security of individuals or their fundamental rights under the EU Charter of Fundamental Rights and are therefore considered high risk.

Limited risk

This refers to the risks associated with a lack of transparency in the use of AI and therefore the need for specific transparency obligations to ensure that humans are informed when necessary, building trust.

Minimal risk

This refers to all other AI systems that can be developed and used in accordance with existing legislation without additional legal obligations. On a voluntary basis, providers of such systems may choose to apply ethical principles such as those of the PIO Model.

1.1 Preliminary analysis of system use

1.1.1 Are the results generation process of your AI system well documented and reproducible in case problems need to be discovered in the future?

See reference

See example

YES	NO	DOES NOT APPLY

1.1.2 Do you know precisely what your AI system is used for?

See reference

See example

YES	NO	DOES NOT APPLY

1.2.1 Is the development process of your system and the people involved documented and public?

See reference

See example

YES	NO	DOES NOT APPLY

1.1 Preliminary analysis of system use

1.1.1 Are the results generation process of your AI system well documented and reproducible in case problems need to be discovered in the future?

See reference

This question relates to Articles 11, 12, 53.1 and 55.1 of Regulation 2024/1689 of the Parliament and of the Council on Artificial Intelligence (AI Act), concerning technical documentation, records and obligations of persons supplying general purpose models.

It is also related to Article 7 of the Product Liability Directive, which deals with the defective nature of products.

It also refers to the fairness section of the ICO toolkit and the Reliability, Fall-back plans and Reproducibility section of ALTAI.

See example

YES	NO	DOES NOT APPLY

Global results

1 Transparency and explainability

Answered questions: 4 of 4 — 75% YES / 25% NO

2 Justice and equity

Answered questions: 2 of 2 — 50% YES / 50% NO

3 Safety and non-maleficence

Answered questions: 2 of 2 — 50% YES / 50% NO

4 Responsibility and accountability

Answered questions: 2 of 2 — 50% YES / 50% NO

5 Privacy

Answered questions: 2 of 2 — 50% YES / 50% NO

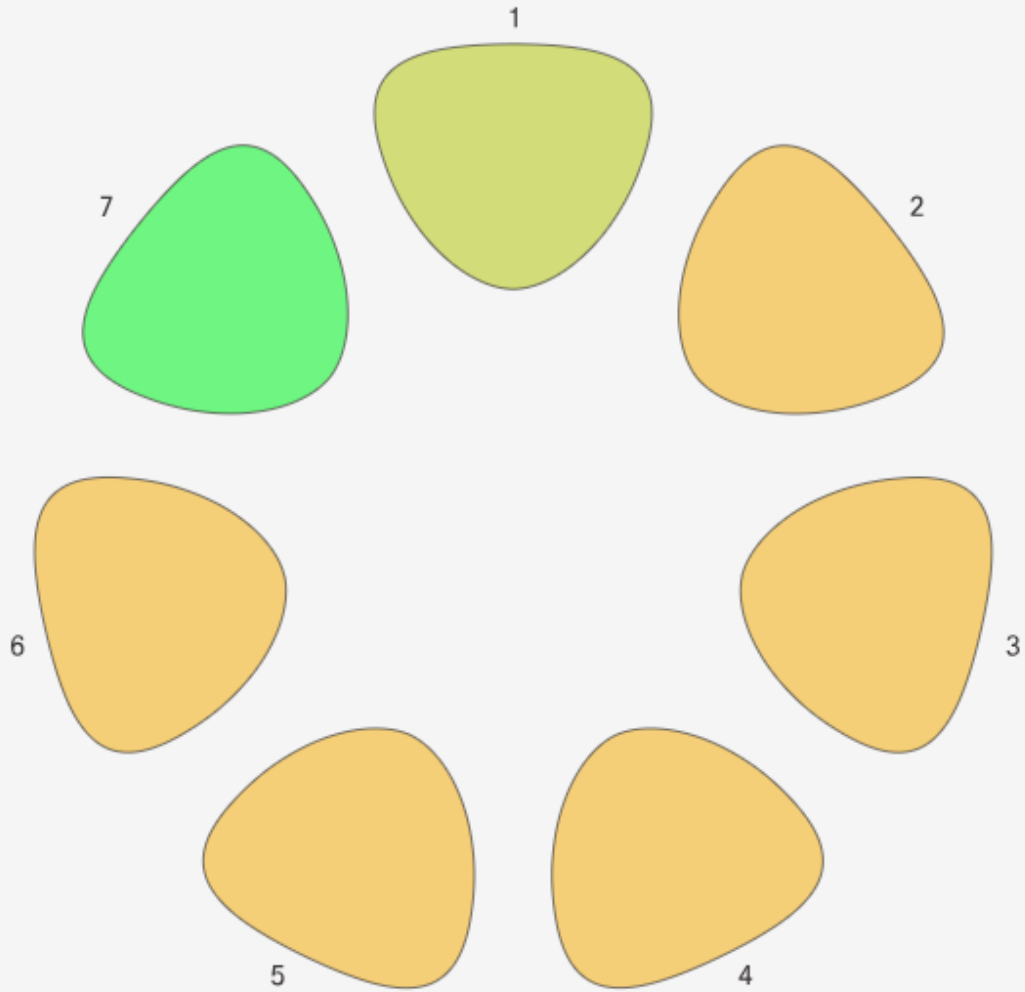
6 Autonomy

Answered questions: 2 of 2 — 50% YES / 50% NO

7 Sustainability

Answered questions: 2 of 2 — 100% YES / 0% NO

0% 0-19% 20-39% 40-59% 60-79% 80-99% 100%



Download badge | Download results

Risk matrix

The following risk matrix is a tool to get a snapshot of the level of risk by severity (regulatory or AI Act) and the likelihood of acting on it. The values in each box represent the negative responses provided according to the severity and probability mentioned. This information is also used for improvement points.

	SEVERITY ACCORDING TO THE REGULATION OR AI ACT		
PROBABILITY OF TAKING AN ACTION	Minimum	Limited	High
Very high	0	0	2
High	0	0	0
Neither high nor low	0	0	1
Low	0	0	0
Very low	0	0	2

Download matrix

Points for improvement

Below you will find those questions that you answered with a “NO” and their degree of severity. The colors used in this section refer to the given combination of severity levels (as per AI Act regulations) and the likelihood that you will take some action soon. These appear in the same table above.

Very low | Low | Low/Medium | Medium | Medium/High | High

2 Justice and equity

2.1.2 Have you ensured that your AI system does not use variables or "proxies" that may be unfairly discriminatory?	This question refers to Articles 9, 15.3 27.1, 55.1 and 95 of Regulation 2024/1689 of the Parliament and of the Council on Artificial Intelligence (AI Act), concerning risk management measures, resilience and robustness of AI systems, fundamental rights impact assessment for high-risk AI systems, obligations of providers of general-purpose systemic risk systems and promotion of voluntary codes of conduct for vulnerable persons or groups. It is also related to Article 7 of the Product Liability Directive, which deals with the defective nature of products. Furthermore, the question is based on Article 10 of the Council of Europe Framework Convention on Artificial Intelligence and Human Rights, Democracy and the Rule of Law, which refers to Equality and non-discrimination, as well as the human right to non-discrimination, established in Article 14 of the European Convention on Human Rights, Article 21 of the EU Charter of Fundamental Rights, Article 26 of the International Covenant on Civil and Political Rights and Article 7 of the Universal Declaration of Human Rights. In this sense, the question is also related to the section *Specific safeguards to ensure non-discrimination when using* of the FRA’s report *Getting The Future Right, Artificial Intelligence And Fundamental Rights*. Finally, we highlight the Fairness and non-discrimination section of UNESCO’s recommendations on ethics in Artificial Intelligence and is related to the Fairness section of the GSMA’s AI Ethics Assessment.

Catalogues

A-B

Examples

Real and hypothetical examples to the questions of the PIO Model.

100+

Legal and recommendations

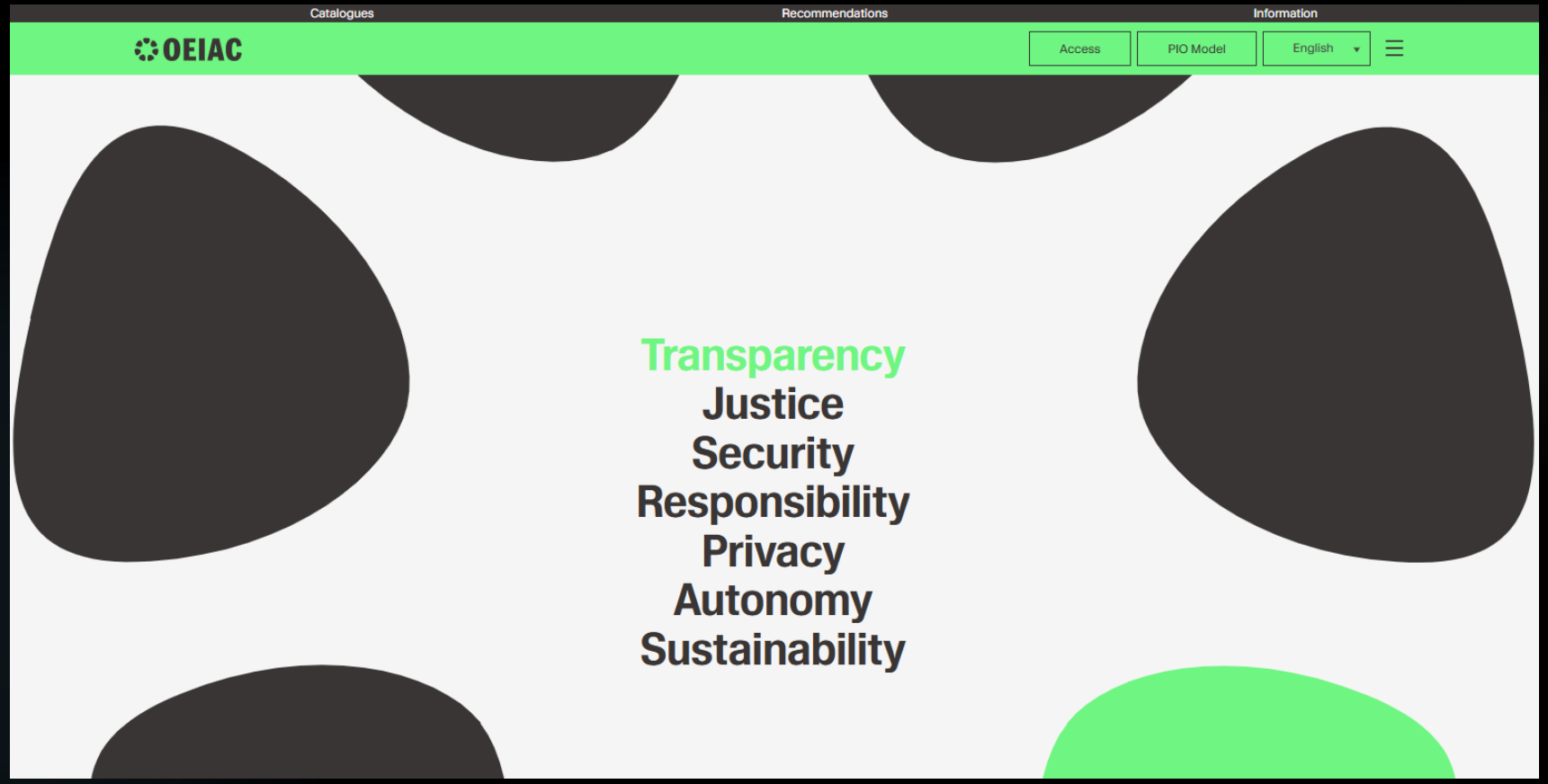
Legal standards on which the questions in the PIO Model have been based.

We offer recommended content for the drafting of contract clauses relating to AI systems.

Transparency | Justice | Security | Responsibility

Privacy | Autonomy | Sustainability | All recommendations

PIO Model



www.oeiac.cat

Evaluación de proyectos con IA en los CEI

Observatori d'Ètica en Intel·ligència Artificial de Catalunya
Universitat de Girona

Albert Sabater Coll

OEIAC | Universitat de Girona

Generalitat de Catalunya