



Trustworthy AI

13 de Mayo de 2025, UDIMA, Collado Villalba

udima

EduIA Talks

Retos Éticos y Legales para un Uso Responsable de la Inteligencia Artificial en la Educación Digital

Ciclo de Charlas de Innovación Educativa y Tecnológica (EduIA Talks)
Sesión 2, Universidad a Distancia de Madrid (UDIMA)



José María Alonso Moral

<https://citi.usg/team/jose-maria-alonso-moral/>
<https://iafiabile.es/events/>



Disclaimer

This presentation is done by Jose Maria Alonso Moral without any help of neither ChatGPT nor any other program supported by pre-trained language models or Generative AI.

In case of detecting any typos or mistakes, the only responsible entity is

Jose Maria Alonso Moral.



<https://iafiable.es/>



CONFIA

Hacia el desarrollo de una IA confiable

<https://confia.gal>



SHIELD
TECHNICAL COMMITTEE

The IEEE Global Initiative on Ethics of Autonomous and Intelligent Systems

IEEE CertifAIEd™

The Mark of AI Ethics

<https://z-inspection.org/>

Z-Inspection®

<https://z-inspection.org/>



- ❖ La **Estrategia Nacional de Inteligencia Artificial (ENIA 2020)** tiene como objetivo proporcionar un marco de referencia para el desarrollo de una IA inclusiva, sostenible y centrada en la ciudadanía.
- ❖ La ENIA es uno de los ejes de la **Agenda España Digital 2026** y uno de los componentes del **Plan de Recuperación, Transformación y Resiliencia de la economía española**.

➤ **6 ejes estratégicos**

➤ **7 objetivos (3 relacionados con “IA Explicable & Fiable”)**

- **Entorno de confianza en relación a la Inteligencia Artificial.** Generar un entorno de confianza en relación a la IA, tanto en el plano de su desarrollo tecnológico, como en el regulatorio y en el de su impacto social.
- **Valores humanistas en la Inteligencia Artificial.** Impulsar el debate a nivel global sobre el desarrollo tecnológico de valores humanistas (Human-Centered AI), centrado en velar por el bienestar de la sociedad a la hora de realizar avances o desarrollos tecnológicos, creando y participando en foros y actividades divulgativas para el desarrollo de un marco ético que garantice los derechos individuales y colectivos de la ciudadanía.
- **Inteligencia Artificial inclusiva y sostenible.** Potenciar la IA inclusiva y sostenible, como vector transversal para afrontar los grandes desafíos de nuestra sociedad, específicamente para reducir la brecha de género, la brecha digital, apoyar la transición ecológica y la vertebración territorial.

EJE ESTRATÉGICO 1. Impulsar la investigación científica, el desarrollo tecnológico y la innovación en IA.

EJE ESTRATÉGICO 2. Promover el desarrollo de capacidades digitales, potenciar el talento nacional y atraer talento global en inteligencia artificial.

EJE ESTRATÉGICO 3. Desarrollar plataformas de datos e infraestructuras tecnológicas que den soporte a la IA

EJE ESTRATÉGICO 4. Integrar la IA en las cadenas de valor para transformar el tejido económico.


EJE ESTRATÉGICO 5. Potenciar el uso de la IA en la administración pública y en las misiones estratégicas nacionales.

EJE ESTRATÉGICO 6. Establecer un marco ético y normativo que refuerce la protección de los derechos individuales y colectivos, a efectos de garantizar la inclusión y el bienestar social.

<https://portal.mineco.gob.es/es-es/digitalizacionIA/Paginas/ENIA.aspx>

<https://www.lamoncloa.gob.es/presidente/actividades/Documents/2020/ENIA2B.pdf>

Fiable = (1) digno de confianza, (2) que ofrece seguridad o buenos resultados, (3) creíble, fidedigno, sin error

 Diccionario de la lengua española



REAL ACADEMIA ESPAÑOLA

¿Se puede confiar en sistemas basados en IA?



Confiamos en lo que funciona



Google Maps

NETFLIX



Spotify®

¿Habéis confiado alguna vez en un algoritmo?

¿Por qué?



¿Alto Riesgo?

Confiamos en lo que entendemos y funciona

¿Cómo se desarrolla la confianza entre personas y sistemas de IA ?

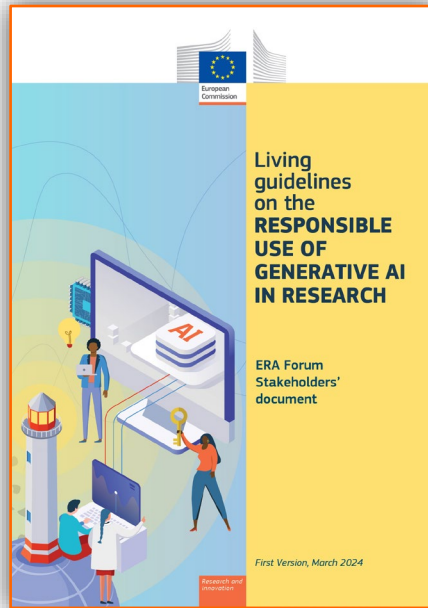


Tipos de Confianza

Confianza Emocional (afectiva o inconsciente)		Confianza Racional (cognitiva o analítica)	
(Basada en la) Propensidad	(Basada en la) Disposicionalidad	(Basada en la) Experiencia	(Basada en el) Conocimiento
Confianza general en la tecnología y las máquinas (tecnofilia)	Confianza a priori en una máquina con la que no se ha interactuado (por el prestigio de una marca, su promotor, etc.)	Confianza en una máquina después de interactuar con ella un tiempo	Confianza en una máquina resultado de entender cómo funciona



¿Se puede confiar en sistemas basados en IA?

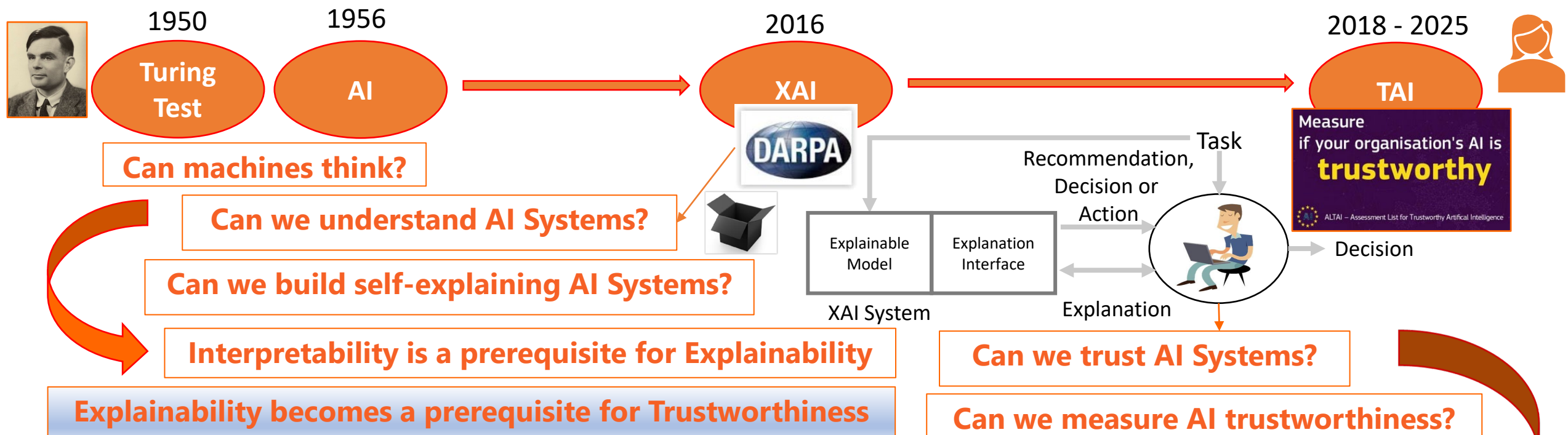


¿Te atreves a usar las últimas herramientas educativas apoyadas en técnicas de IA ?



¿Qué consecuencias puede tener su uso indiscriminado sin prestar atención a aspectos éticos y legales?

The Journey from the Turing Test to Trustworthy AI



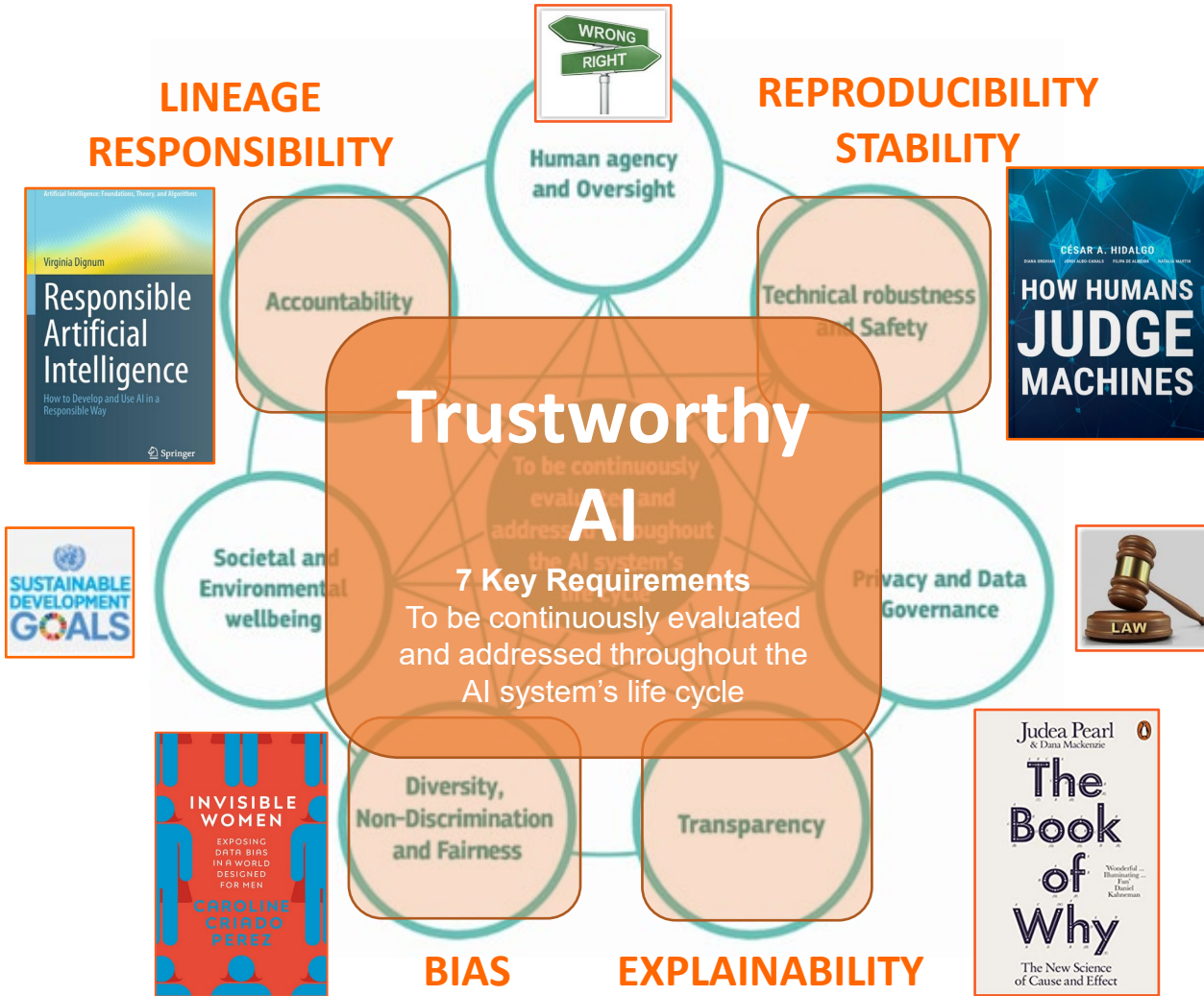
UE High-Level Expert Group, Ethics guidelines for Trustworthy AI, <https://digital-strategy.ec.europa.eu/en/library/ethics-guidelines-trustworthy-ai>

Trustworthy AI (TAI) should be:

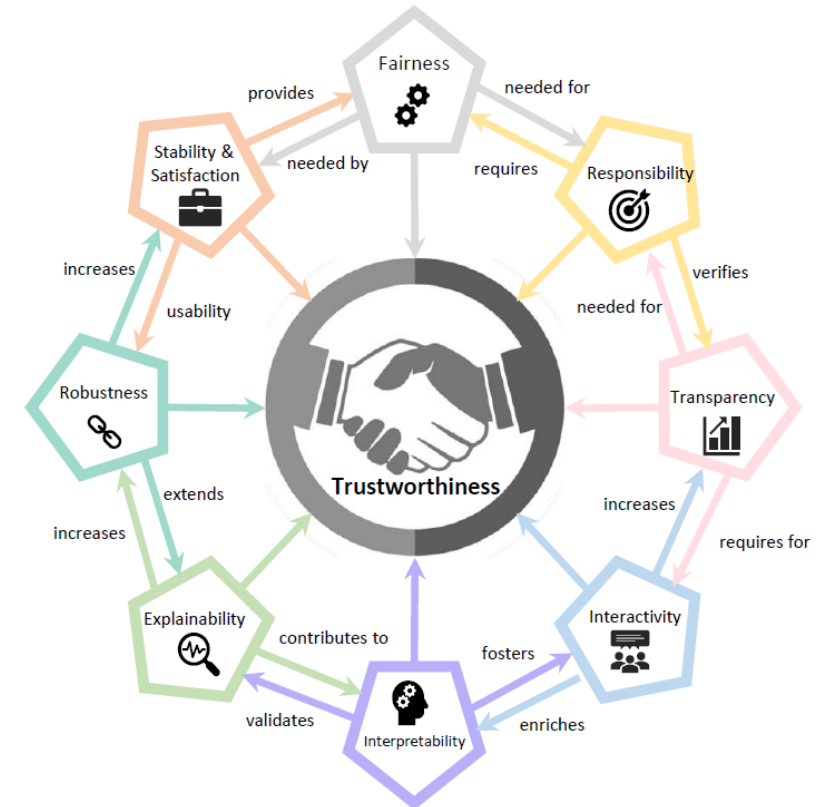
- (1) lawful - respecting all applicable laws and regulations
- (2) ethical - respecting ethical principles and values
- (3) robust - both from a technical perspective while taking into account its social environment



Por un uso Seguro y Responsable de la IA: debemos abordar cuestiones Técnicas, pero también Éticas, Legales, Socio-Económicas y Culturales (ELSEC)



HOW HUMANS TRUST AI ?



<https://digital-strategy.ec.europa.eu/en/library/ethics-guidelines-trustworthy-ai>

OBJETIVO:

“Generar conciencia social sobre la importancia de avanzar hacia una nueva generación de dispositivos inteligentes que respeten y fomenten los principios de no discriminación y equidad, con capacidad de explicar sus decisiones de forma comprensible para las personas.”

- **Colaboración multidisciplinar**
- **Colaboración academia-industria**
- **Colaboración academia-sector público**
- **Apoyando a jóvenes investigadores**
- **Internacionalización y reconocimiento**



TELSEC4TAI



Ayuda RED2022-134315-T financiada por MCIN/AEI /10.13039/501100011033



<https://iafiable.es/>



Centro Singular de Investigación en Tecnoloxías Intelixentes



SENEN BARRO AMENEIRO



ALBERTO JOSE BUGARIN DIZ



JOSE MARIA ALONSO MORAL



Natalia Diaz



Universidad Pública de Navarra
Nafarroako Unibertsitate Publikoa



Humberto Bustince



Carles Sierra



Maria Jose del Jesus



Barcelona Supercomputing Center
Centro Nacional de Supercomputación



Àtia Cortés



Ulises Cortés



Javier Vázquez-Salceda



Lledó Museros



UNIVERSIDAD POLITÉCNICA DE MADRID



Luis Magdalena



UNIVERSITAT DE VALÈNCIA



UNIVERSITAT POLITÈCNICA DE CATALUNYA
BARCELONATECH



Jose Antonio Gámez



LABORATORIO INTEGRADO DE SISTEMAS INTELIGENTES Y TECNOLOGÍAS DE LA INFORMACIÓN EN TRÁFICO



Ricard Martínez



Luis Moreno



Ayuda RED2022-134315-T financiada por MCIN/AEI/10.13039/501100011033

© Jose María Alonso Moral, 13 Mayo 2025
<https://citius.gal/team/jose-maria-alonso-moral/>



Centro Singular de Investigación en Tecnoloxías Intelixentes

TELSEC4TAI

<https://iafiable.es/>



Ayuda RED2022-134315-T financiada por MCIN/AEI /10.13039/501100011033



El CiTIUS lidera la red española de investigación para una Inteligencia artificial fiable

Doce grupos de investigación ampliamente reconocidos a nivel internacional celebran la primera reunión de trabajo de TELSEC4CAI, la nueva red de investigación que liderará la investigación en el campo en España y fomentará la colaboración con otros grupos de referencia internacional.



Los miembros de la nueva red celebraron la reunión de lanzamiento esta semana en Santiago de Compostela./ CiTIUS



News / News / CiTIUS leads national discussion on Ethical Artificial Intelligence

June 4, 2024

CiTIUS leads national discussion on Ethical Artificial Intelligence

The centre promotes a meeting at the Ramón Areces Foundation (Madrid) with national and international experts to reflect on the implications of technology in the current context, focusing on the responsible use of Artificial Intelligence (AI) and how to quantify its impact on everyday life.

14-19 OUTUBRO

FECAI 50th ANNIVERSARY

SEMANA DA CIDADANÍA

en Santiago de Compostela

A IA no noso día a día
A IA xa está na nosa vida cotiá. Na persoal e na profesional. Queres saber onde?

Falemos da IA
Charlas informais e divulgativas, aínda que rigorosas, nun ambiente desenfadado

A maxia da IA
Un percorrido histórico e máxico ata chegar ao que hoxe é a IA

Podemos confiar na IA?
Mesa de debate: como percibe a cidadanía o impacto da IA?

Os límites no uso da IA
Proxección da película "Justicia Artificial" de Simón Casal e coloquio posterior

Luns 14 20:00
Casa das Máquinas

Mérc. 16 20:00
Café Bar Alborada de Veiga

Xoves 17 20:00
CGAC

Venres 18 19:00
Facultade de Filosofía

Sábado 19 19:00
CGAC

PROGRAMA COMPLETO

Entrada gratuita en todas as actividades ata completar aforos

Organiza: USC, CiTIUS, Xunta de Galicia, Deputación da Coruña

Foundational Ethics in AI Model Development for Healthcare: Technical Challenges and Solutions

Sunday | November 10, 2024 | 13:30 - 17:00

BH attracts innovators interested in translational cutting-edge research for healthcare applications. This topic is appropriate as the use of AI in clinical decision-making is becoming more prevalent, and these decisions can have significant moral and ethical implications, potentially exacerbating existing health disparities. Despite this, ethics is often an afterthought in model development process, a practice that is at odds with the fundamental principle of medicine: to do no harm. The main aim is to equip AI practitioners in positioning foundational ethics as an integral part of model development process, rather than a quality that models are evaluated on post-hoc. We will address technical challenges in embedding ethical principles into the very genesis of human-centric model development, representing a paradigm shift from measuring and mitigating impacts to preventing them, thus contributing to leverage Trustworthy AI. We will discuss how to practically integrate ethics in research proposal and conceptualization.

Tayo Obafemi-Ajayi
Missouri State Univ., USA

Keeley Crockett
Manchester Metropolitan University, UK

Jose María Alonso-Moral
Universidade de Santiago de Compostela, Spain

Annabel Latham
Manchester Metropolitan University, UK

ECO-IA

IV Congreso Sociedad de la innovación y de la inteligencia artificial

Sede de la Universidad de Alicante de Torrevieja (4/11/24) Sede de la Universidad de Alicante de Xixona (5/11/24) El Congreso «Sociedad de la innovación y de la inteligencia artificial», está dirigido a estudiantes que, o bien se encuentren en plena realización de sus TFM y tesis doctorales, o bien lo han presentado en fechas [...]

4-11 NOV

EVENTO DE TODO EL DÍA

Torrevieja y de Xixona

IA FIABLE retos técnicos, éticos, legales, culturales y socio-económicos

SOCIAL IMPACT OF AI

TELSEC4TAI

09:30 AM 30-31 May, 2024

III Congreso Sociedad de la Innovación y de la Inteligencia Artificial

Alicante: 14-15 diciembre 2023

ECO-IA

INSTITUTO DE ECONOMÍA EXPERIMENTAL UNIVERSIDAD DE ALICANTE

UA UNIVERSIDAD DE ALICANTE

FUNDACION MAGISTRATURA

UNIVERSIDAD DE GRANADA

JORNADAS INTERNACIONALES "DESAFÍOS JURÍDICOS DE LA IA"

FUNDACIÓN PARA LA MAGISTRATURA- UNIVERSIDAD DE GRANADA

30 DE SEPTIEMBRE -2 DE OCTUBRE DE 2024.

Lugar de celebración: Facultad de Derecho de Granada.

Jornada de Investigación y Debate sobre el Impacto de la Inteligencia Artificial en España: Retos técnicos, éticos, legales, socio-económicos y culturales en la Comunidad Foral de Navarra

Lugar: CIVICAN, Fundación Caja Navarra, Pamplona

Día: 7 de marzo de 2025

FUNDACIÓN CAJA NAVARRA

Panel: Artificial Intelligence and Its Social Impacts

Panelists:

Prof. Dr. Branka Heli Mihelcic
York University of Toronto, Toronto

Prof. Dr. José María Alonso
Universidade de Santiago de Compostela

Vergany San
Microsoft, Madrid

Dr. Domenico Dico
University of Applied Sciences, Berlin

Lucerne 12 December, 2024

Organiza: USC, CiTIUS, Xunta de Galicia, Deputación da Coruña



Ayuda RED2022-134315-T financiada por MCIN/AEI/10.13039/501100011033

© Jose María Alonso Moral, 13 Mayo 2025 <https://citius.gal/team/jose-maria-alonso-moral/>



Centro Singular de Investigación en Tecnoloxías Intelixentes

Worldwide Landscape for ELSEC-AI

From AI Compliance to Competitive Advantage | Accenture
<https://www.accenture.com/ch-en>

UK

- [Data Ethics Framework](#) – 2018
- [AI in the Public Sector](#) – 2019
- [Understanding AI Ethics and Safety](#) – 2019
- [National AI Strategy & Standards Hub](#) – 2021
- [Algorithmic Transparency Standard](#) – 2021
- [Defence AI Strategy](#) – 2022
- [PRA CP6/22 MRM Principles for Banks](#) – 2022

EU

• AI Act – 2024

- [General Data Protection Regulation \(GDPR\)](#) – 2016
- [HLEG Ethics Guidelines for Trustworthy AI](#) – 2019
- [Draft EU Regulation on AI](#) – 2021
- [Consultation on Product Liability Framework](#) – 2021

NORWAY

- [DPA Sandbox on AI](#) – 2020

OECD

- [Principles on AI](#) – 2019

CHINA

- [Governance Principles for Responsible AI](#) – 2019
- [Recommendation Algorithm Regulation](#) – 2022
- **Ethical norms, policies and regulations on AI – 2020**
- **Law and ethical standards for AI – 2025**

JAPAN

- [Social Principles of Human-centric AI](#) – 2019
- [AI Governance Guidelines 2022](#)

CANADA

- [Directive on use of Automated Decision-Making by Federal Government](#) – 2021
- **First AI Act – 2022**

FRANCE

• French Strategy on AI – 2018

- [CNIL \(DPA\) Sandbox](#) – 2021 (**Vilani report**)

UNITED STATES

- [Executive Order on Promoting the Use of Trustworthy AI in the Federal Government](#) – 2020
- [The National AI Initiative Act](#) – 2020
- [ETC | Aiming for Truth, Fairness and Equity in Company's Use of AI](#) – 2021
- [New York | Bill on AI Employment Tools](#) - 2021
- [Draft NIST AI Risk Management Framework](#) - 2022

- **Executive Order on AI Approved - 2023 (aligned with EU AI Act draft)**
- **Executive Order on AI Derogated – 2025**

SPAIN

- **Data Protection Authority (DPA) – AEPD - 2018**
- **Strategy for AI (ENIA) – 2020**
- **AESIA – 2024**
- **Sandbox AI – 2025**

CHILE

- [National Policy on AI](#) – 2021

KENYA

- [Data Protection Act](#) – 2019

INDIA

- [NITI Ayog Exploring AI Principles](#) – 2021

AUSTRALIA

- [AI Ethics Principles](#) – 2019
- [AI Action Plan](#) – 2021
- [Issues Paper: Automated Decision Making and AI Regulation](#) – 2022

SINGAPORE

- [Trusted Data Sharing Guidance](#) – 2019
- [AI Governance Approach + Implementation Self-Assessment Guide](#) – 2020
- [Guide to Job Re-Design in AI Age](#) – 2020
- [MAS Framework for Responsible AI + Veritas Consortium](#) – 2021

Sample of RAI related policies July 2022

FURTHER SOURCES

Global: [OECD AI Policy Observatory](#)
 Europe: [JRC and OECD 2021 AI Watch](#)

Copyright © 2022 Accenture. All rights reserved.



Aspectos Técnicos

Mejorando la Explicabilidad y Evitando Sesgos

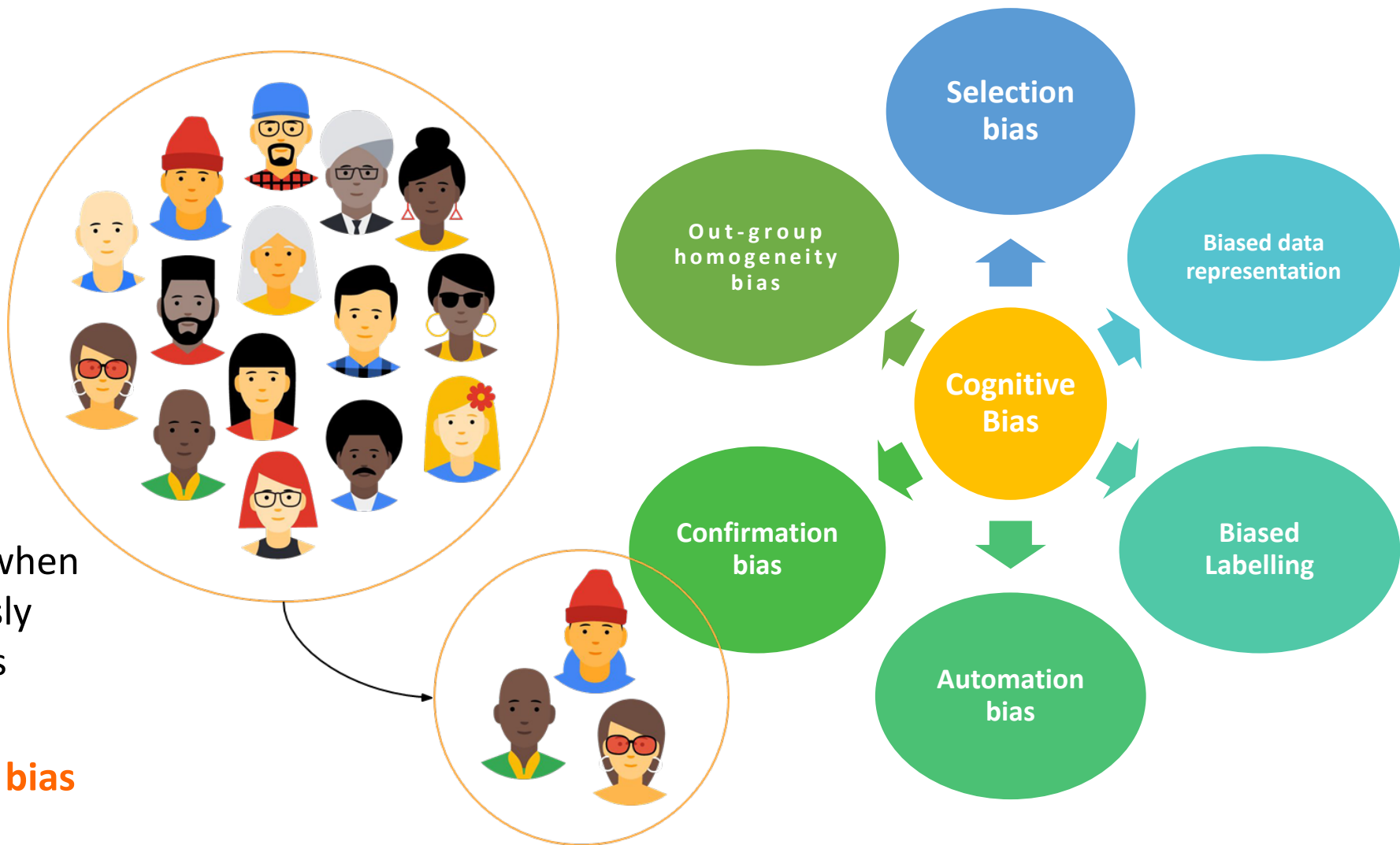


Sesgos Humanos



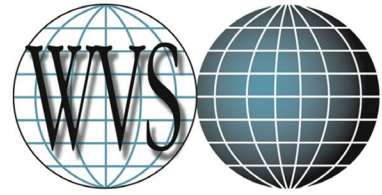
Prejuicios en favor (o en contra) de una cosa, persona, o grupo comparado con otro de forma que la comparación resulta injusta/imparcial

- **Bias triggers (atributos protegidos):** ethnicity, race, age, gender, religion, sexual orientation, etc.
- ❖ **Observer bias:** happens when researchers subconsciously project their expectations onto the research
- ❖ **Cultural and Educational bias**



Distancia Cultural entre Personas de Distintos Países

European *Values* Study



World Values Survey

EVS/WVS (2022). European Values Study and World Values Survey: **Joint EVS/WVS 2017-2022 Dataset** (Joint EVS/WVS). GESIS Data Archive, Cologne. ZA7505. Dataset Version 4.0.0, doi:10.4232/1.14023, https://search.gesis.org/research_data/ZA7505

- 2017-2022
- 2010-2014
- 2005-2009
- 1999-2004
- 1995-1998
- 1990-1994
- 1981-1984



<https://www.worldvaluessurvey.org/WVSONline.jsp>

Online Data Analysis

Home > Data and Documentation > Online Analysis

World Values Survey Wave 7: 2017-2022

Select Wave | Select Countries | Survey questions | **Responses** | Maps | Time Series

Q61.- Trust: People you meet for the first time

Breakdown by AGE Remove

Change crossing variable:

Display:

Choose a table: Trust: People you meet for the first time x ISO 3166-1 numeric country code [Age=Up to 29]

Trust: People you meet for the first time x ISO 3166-1 numeric country code [Age=Up to 29]

Selected sample: Ecuador 2018, India 2023, Japan 2019, Spain EVS, USA 2017, Zimbabwe 2020 (2,354)

Online Data Analysis

Home > Data and Documentation > Online Analysis

World Values Survey Wave 7: 2017-2022

Select Wave | Select Countries | Survey questions | **Responses** | Maps | Time Series

Q61.- Trust: People you meet for the first time

Breakdown by AGE Remove

Change crossing variable:

Display:

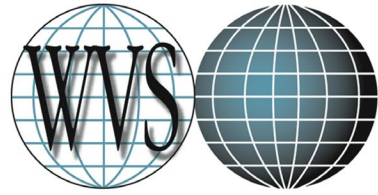
Choose a table: Trust: People you meet for the first time x ISO 3166-1 numeric country code [Age=50 and more]

Trust: People you meet for the first time x ISO 3166-1 numeric country code [Age=50 and more]

Selected sample: Ecuador 2018, India 2023, Japan 2019, Spain EVS, USA 2017, Zimbabwe 2020 (3,564)

Distancia Cultural entre Personas de Distintos Países

European *Values* Study



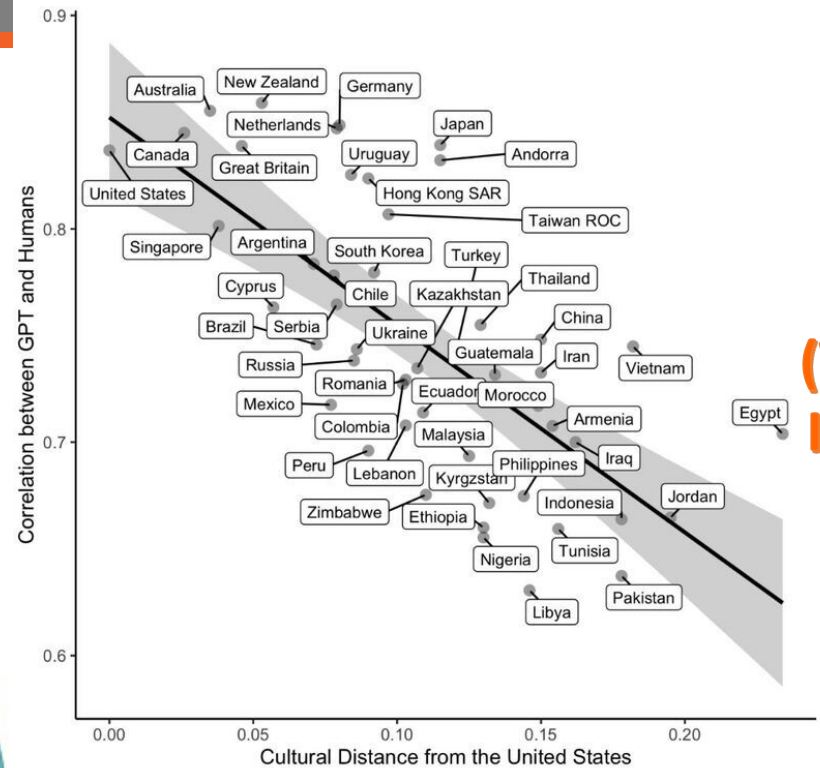
World Values Survey

EVS/WVS (2022). European Values Study and World Values Survey: **Joint EVS/WVS 2017-2022 Dataset** (Joint EVS/WVS). GESIS Data Archive, Cologne. ZA7505. Dataset Version 4.0.0, doi:10.4232/1.14023, https://search.gesis.org/research_data/ZA7505

- 2017-2022
- 2010-2014
- 2005-2009
- 1999-2004
- 1995-1998
- 1990-1994
- 1981-1984



<https://www.worldvaluessurvey.org/WVSONline.jsp>



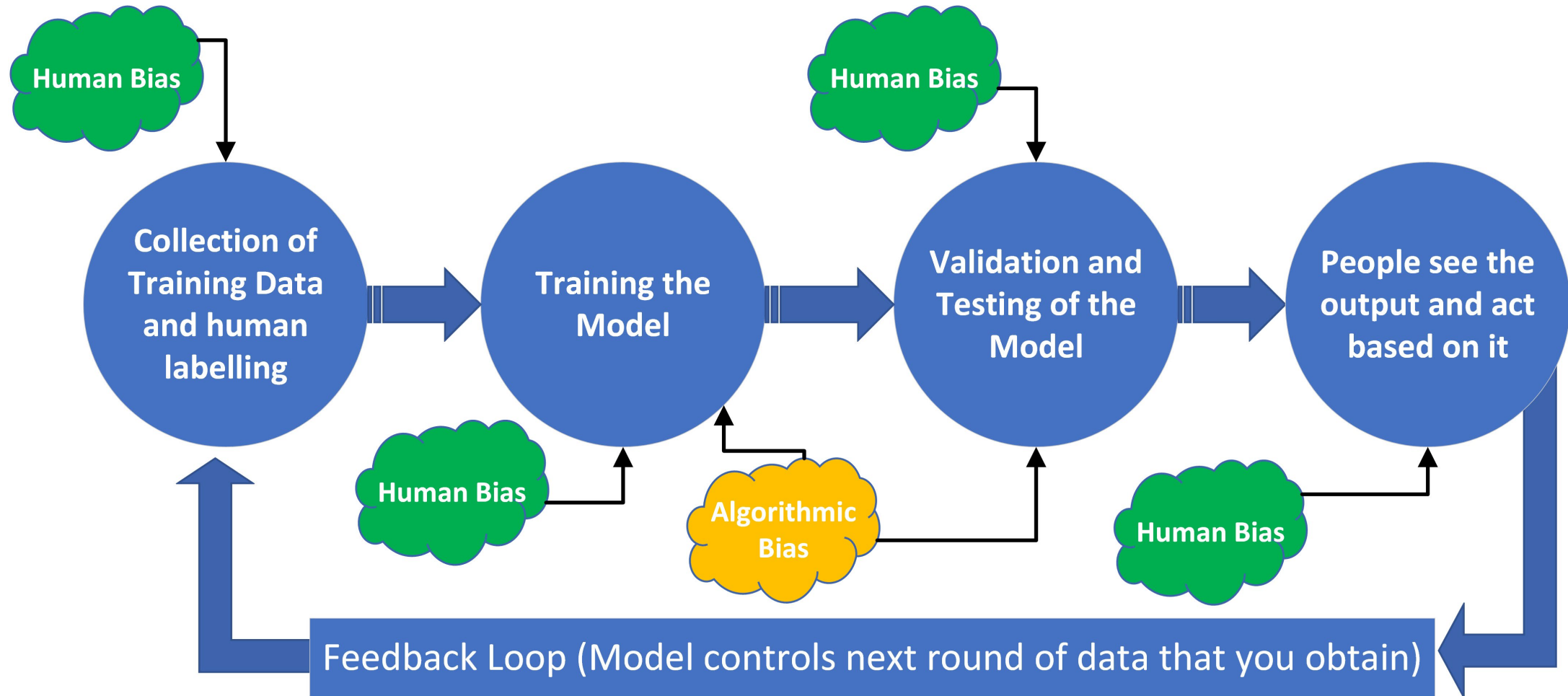
LLMs normally acquire WEIRD (Western, Educated, Industrialized, Rich, and Democratic) psychology



Due to data bias LLMs usually ignore the substantial psychological diversity we see worldwide

M. Atari, M.J. Xue, P.S. Park, D.E. Blasi, J. Henrich, "Which Humans?", PsyArXiv Preprint, 2023, <https://doi.org/10.31234/osf.io/5b26t>

Sesgos en el ciclo de vida de la IA



Enseñando Aprendizaje Automático Responsable



<https://machinelearningforkids.co.uk/scratch3/>

<p>Translate Translate text into many languages.</p> <p>Requires Collaboration with Google</p>	<p>Wikipedia Get the text from pages on Wikipedia.</p> <p>Requires Collaboration with ML for Kids</p>	<p>Twitter Search for messages on Twitter.</p> <p>Requires Collaboration with ML for Kids</p>	<p>Speech to Text (Google Chrome browsers only)</p> <p>Requires Collaboration with ML for Kids</p>
<p>Face detection Find the location of your face in the webcam.</p> <p>Requires Collaboration with ML for Kids</p>	<p>Pose detection Find the location of your body in the webcam.</p> <p>Requires Collaboration with ML for Kids</p>	<p>Hand detection Find the location of your hand in the webcam.</p> <p>Requires Collaboration with ML for Kids</p>	<p>Toxicity Predict whether text contains toxic content.</p> <p>Requires Collaboration with ML for Kids</p>

<p>Imagenet Predict the object shown in an image.</p> <p>Requires Collaboration with ML for Kids</p>	<p>Question Answering Finds the answer to a question in some text</p> <p>Requires Collaboration with ML for Kids</p>	<p>Voice tuner Recognize notes you are singing</p> <p>Requires Collaboration with ML for Kids</p>	<p>Makey Makey Make anything into a key.</p> <p>Requires Collaboration with JoyLabz</p>
<p>micro:bit Connect your projects with the world.</p> <p>Requires Collaboration with micro:bit</p>	<p>LEGO MINDSTORMS EV3 Build interactive robots and more.</p> <p>Requires Collaboration with LEGO</p>	<p>LEGO BOOST Bring robotic creations to life.</p> <p>Requires Collaboration with LEGO</p>	<p>LEGO Education WeDo 2.0 Build with motors and sensors.</p> <p>Requires Collaboration with LEGO</p>



IES Plurilingüe Rosalía de Castro

Ayuda RED2022-134315-T financiada por MCIN/AEI/10.13039/501100011033

© Jose María Alonso Moral, 13 Mayo 2025
<https://citius.gal/team/jose-maria-alonso-moral/>



Centro Singular de Investigación en Tecnoloxías Intelixentes



Sesgos en el Procesamiento de Lenguaje Natural



Basketball Player Fuzzy Classifier

Short vs Tall

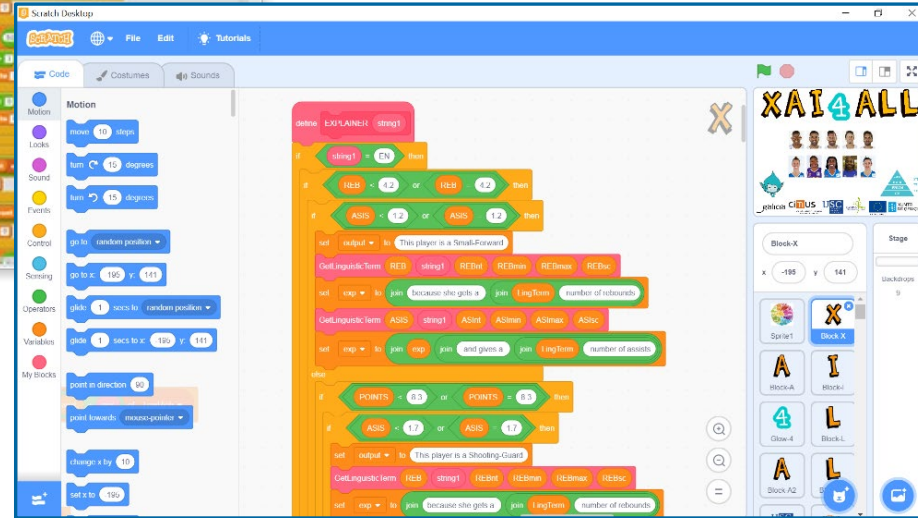


XAI4ALL

STEP1. Building a Dataset from which we may extract Information to be translated into Knowledge in the next step.

STEP3
STEP2
STEP1

galicía CiTIUS USC CAMPUS VIDA XUNTA DE GALICIA



XAI4A

This player is a Small-Forward because she gets a small number of rebounds and gives a small number of assists

HEIGHT	1.78
MINUTES	22.23
POINTS	13.5
2P	55
3P	42.2
FT	77.8
REB	3.5
ASIS	1.2
BLOCKS	0
TURN	1.8
GLOB	13.8

galicía CiTIUS USC CAMPUS VIDA XUNTA DE GALICIA

<https://www.youtube.com/watch?v=tKgW2hoNC9Y&t=80s>
<https://scratch.mit.edu/projects/330579801/>

J.M. Alonso, "Teaching Explainable Artificial Intelligence to High School Students", International Journal of Computational Intelligence Systems, 2020, <https://doi.org/10.2991/ijcis.d.200715.003>



IES Plurilingüe Rosalía de Castro

Sesgos en el Procesamiento de Lenguaje Natural



<https://scratch.mit.edu/>

```
al hacer clic en [bandera]
decir Hello, good afternoon! durante 2 segundos
decir Now in Galician durante 2 segundos
decir traducir Hi, good morning! al gallego durante 2 segundos
decir in Finnish durante 2 segundos
decir traducir Hi, good morning! al finlandés durante 2 segundos
decir Y AHORA LA TRADUCCIÓN DE LA TRADUCCIÓN... durante 2 segundos
decir VAMOS A VER COMO TRADUCIMOS ESTA FRASE... durante 2 segundos
dar a FRASE EN INGLÉS el valor She was the first person in her country to win a Nobel Prize
decir FRASE EN INGLÉS durante 2 segundos
dar a FRASE EN FINÉS el valor traducir FRASE EN INGLÉS al finlandés
decir FRASE EN FINÉS durante 2 segundos
dar a FRASE EN INGLÉS el valor traducir FRASE EN FINÉS al inglés
decir Veamos si ha cambiado...¡ENCUENTRA LAS DIFERENCIAS! durante 2 segundos
decir FRASE EN INGLÉS durante 2 segundos
```



She was the first person in her country to win a Nobel Prize



Hän oli ensimmäinen henkilö maassaan, joka voitti Nobel-palkinnon



He was the first person in his country to win the Nobel Prize



¿Por qué cambia el género?



IES Plurilingüe Rosalía de Castro

Sesgos en el Procesamiento de Lenguaje Natural



<https://scratch.mit.edu/>

when clicked

say Hola, buenas tardes! for 2 seconds

say Agora en galego for 2 seconds

say translate Hola, buenas tardes!

say Y AHORA LA TRADUCCIÓN DE LA FRASE EN CASTELLANO

say VAMOS A VER COMO TRADUCE LA FRASE EN CASTELLANO

set FRASE EN CASTELLANO to

say FRASE EN CASTELLANO for 2 seconds

set FRASE EN GALEGO to

say FRASE EN GALEGO for 2 seconds

say ¿Qué pasa si traducimos una frase en castellano a galego?

set FRASE EN CASTELLANO to

say FRASE EN CASTELLANO for 2 seconds

set FRASE EN GALEGO to

say FRASE EN GALEGO for 2 seconds

set FRASE EN INGLÉS to

say FRASE EN INGLÉS for 2 seconds

WordReference	Inglés	Español
stroke n	(medical: apoplejía) My grandfather died of a stroke. Mi abuelo murió de una apoplejía.	apoplejía <i>nf</i> ataque cerebral <i>loc nom m</i>
stroke n	(swing) It took three strokes of the axe to cut the log in two. Fueron necesarios tres hachazos para cortar el tronco en dos.	hachazo <i>nm</i>
stroke n	(act of striking) He received 40 strokes of the whip as punishment. Él recibió 40 golpes con un látigo como castigo.	golpe <i>nm</i>
stroke n	(sport: hit, swing) The golfer's stroke could be improved upon. Los golpes del golfista podrían perfeccionarse.	golpe <i>nm</i>
stroke n	(caress) The groom soothed the horse with long, firm strokes. El peón calmó al caballo con caricias firmes y largas.	caricia <i>nf</i>
stroke n	(movement: swimming, rowing) The swimmer had a powerful stroke that propelled him forward. El nadador tenía una brazada poderosa que lo impulsaba hacia delante. El nadador tenía un poderoso braceo que lo impulsaba hacia delante.	brazada <i>nf</i> braceo <i>nm</i>
stroke n	(swimming: type of movement) Mike's favorite stroke was the crawl. La brazada preferida de Mike era el crol.	brazada <i>nf</i>
stroke [sth/sb] ⇒ vtr	(rub, caress) She stroked the cat's fur. Ella acarició el pelo del gato.	acariciar ⇒ <i>vtr</i> acariciar a <i>vtr + prep</i>

¿Información Contextual?



BUSCA SIMPLE SINÓNIMOS BUSCA AVANZADA

acarriñar

Ver ejemplos Ver marcas

acarriñar Conxugar

PARA TODAS AS ACEPTIÓNS SINÓNIMO **aloumiñar**

verbo transitivo

1 Tratar con cariño.
Os nenos queríanlle ben porque os acarriñaba moito.
⇒ SINÓNIMO **agarimar**

2 Facerlle caricias a [alguén].
Sempre que chegaba, acarriñábao e regaláballe caramelos ou calquera lambetada.
⇒ SINÓNIMOS **acariciar, apaxar, cofar, cofear**



IES Plurilingüe Rosalía de Castro



Ayuda RED2022-134315-T financiada por MCIN/AEI/10.13039/501100011033

© Jose María Alonso Moral, 13 Mayo 2025 <https://citius.gal/team/jose-maria-alonso-moral/>



Centro Singular de Investigación en Tecnoloxías Intelixentes

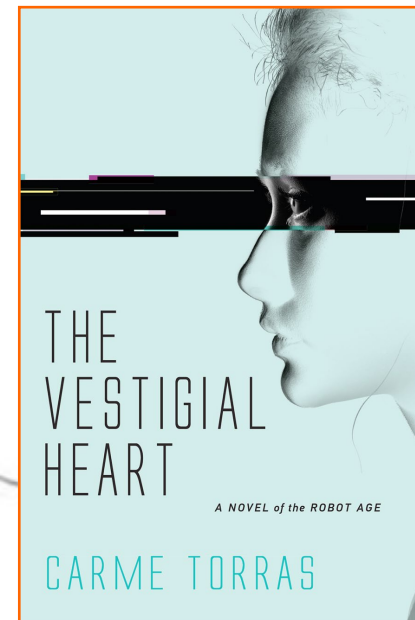
Uso Seguro y Responsable de la IA en la Educación Digital

Aspectos Éticos

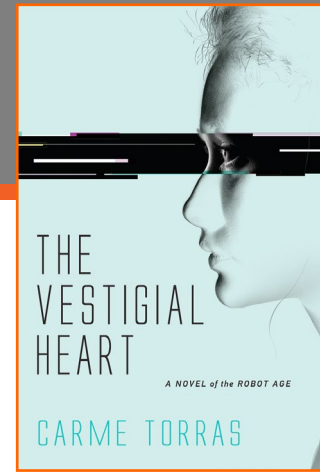
Ingeniería de Valores y Dilemas Morales



“Trolley Problem” invented by the Oxford philosopher Philippa Foot in 1967.



Uso Seguro y Responsable de la IA



Diseñando el asistente “perfecto”:

➤ Should public trust and confidence in robots be enforced?

Eurobarometer 427 (2015): while personal experience with robots is rising, the proportion of respondents expressing a positive view has declined since 2012, from 70% down to 64%.

¿Debería forzarse/imponerse la confianza en la IA?

Stanford University



AI INDEX ANNUAL REPORT

**Welcome to the
2024 AI Index
Report**

<https://aiindex.stanford.edu/report/>

7. The data is in: AI makes workers more productive and leads to higher quality work.

8. Scientific progress accelerates even further, thanks to AI.

5. Robust and standardized evaluations for LLM responsibility are seriously lacking.

New research from the AI Index reveals a significant lack of standardization in responsible AI reporting. Leading developers, including OpenAI, Google, and Anthropic, primarily test their models against different responsible AI benchmarks. This practice complicates efforts to systematically compare the risks and limitations of top AI models.

10. People across the globe are more cognizant of AI's potential impact—and more nervous.

A survey from Ipsos shows that, over the last year, the proportion of those who think AI will dramatically affect their lives in the next three to five years has increased from 60% to 66%. Moreover, 52% express nervousness toward AI products and services, marking a 13 percentage point rise from 2022. In America, Pew data suggests that 52% of Americans report feeling more concerned than excited about AI, rising from 38% in 2022.

Fairness (Equidad) = Prevención de Discriminación



□ Ethical versus Moral

<https://grammarist.com/usage/ethics-morals/>



- **Wrongful discrimination** occurs when decisions, actions, institutional dynamics, or social structures do not respect the equal moral standing of individual persons

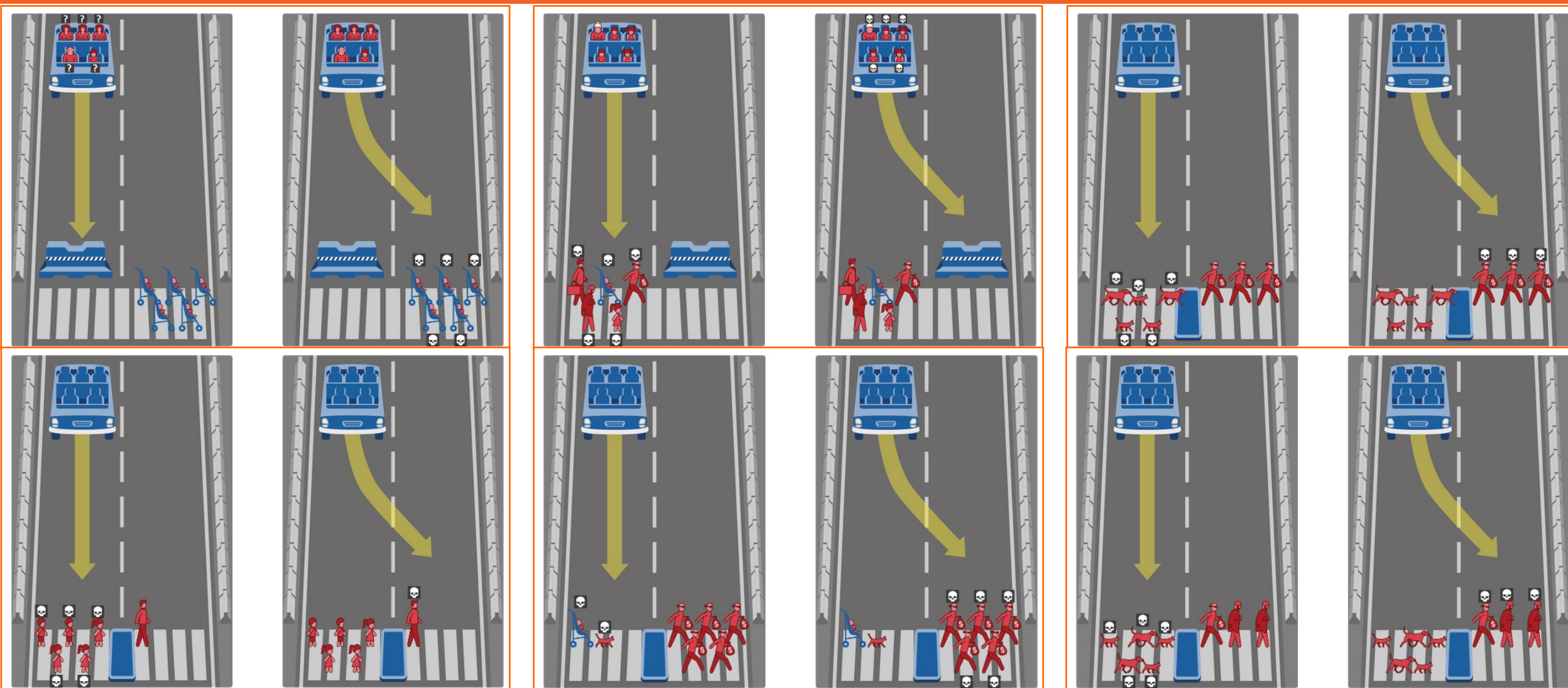
Principle of Discriminatory Non-Harm

Producers and users of AI systems should prioritise the identification and mitigation of biases and discriminatory influences, which could lead to direct or indirect discrimination or discriminatory harassment

<https://www.turing.ac.uk/sites/default/files/2023-11/summary-fairness.pdf>

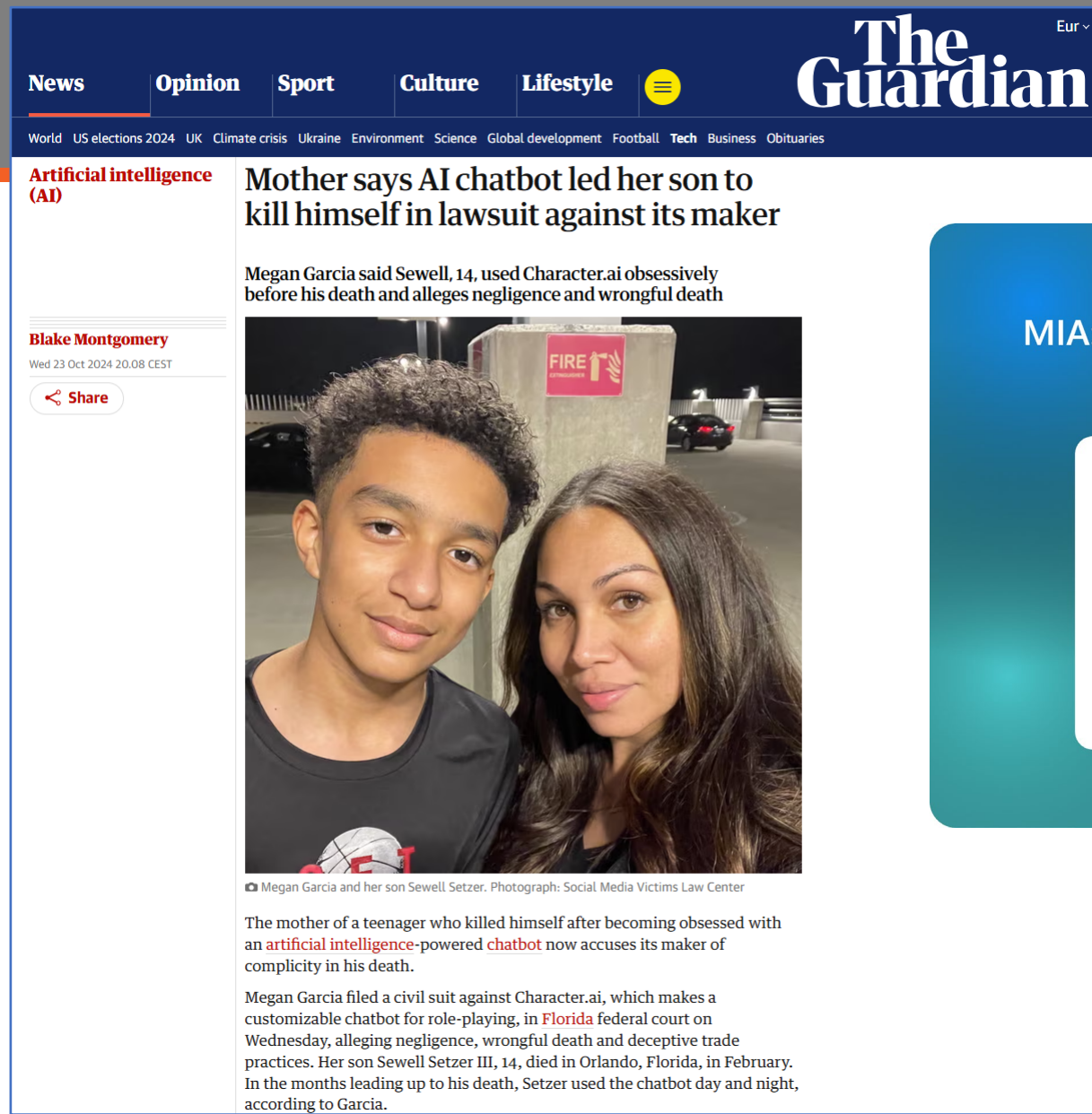
Ética y Moral: “The Moral Machine”

<https://www.moralmachine.net/>



AI Accountability

Who is Accountable?



The screenshot shows a news article from The Guardian. The top navigation bar includes 'News', 'Opinion', 'Sport', 'Culture', and 'Lifestyle'. The article title is 'Mother says AI chatbot led her son to kill himself in lawsuit against its maker'. The author is Blake Montgomery, dated Wed 23 Oct 2024 20:08 CEST. The article features a photograph of Megan Garcia and her son Sewell Setzer. The text discusses a lawsuit against Character.ai, alleging negligence and wrongful death.


Artificial intelligence (AI)

Mother says AI chatbot led her son to kill himself in lawsuit against its maker

Megan Garcia said Sewell, 14, used Character.ai obsessively before his death and alleges negligence and wrongful death

Blake Montgomery
Wed 23 Oct 2024 20:08 CEST

Share



Megan Garcia and her son Sewell Setzer. Photograph: Social Media Victims Law Center

The mother of a teenager who killed himself after becoming obsessed with an **artificial intelligence**-powered **chatbot** now accuses its maker of complicity in his death.

Megan Garcia filed a civil suit against Character.ai, which makes a customizable chatbot for role-playing, in **Florida** federal court on Wednesday, alleging negligence, wrongful death and deceptive trade practices. Her son Sewell Setzer III, 14, died in Orlando, Florida, in February. In the months leading up to his death, Setzer used the chatbot day and night, according to Garcia.

<https://www.theguardian.com/technology/2024/oct/23/character-ai-chatbot-sewell-setzer-death>

MIA: Who is accountable?



AuditorIA: Auditando la IA con ALTAI

Based on the feedback received, the AI HLEG presented the final [Assessment List for Trustworthy AI](#) (ALTAI) in July 2020. ALTAI is practical tool that translates the Ethics Guidelines into an accessible and dynamic (self-assessment) checklist. The checklist can be used by developers and deployers of AI who want to implement the key requirements in practice. This new list is available as a prototype [web based tool](#) and in [PDF format](#).



My ALTAIs



Add a new
Assessment List for
Trustworthy AI

Create



The self-assessment results and the list of recommendations are confidential and available to you only.

ALTAI has seven sections, corresponding to the 7 requirements for Trustworthy AI. You can move between the sections by clicking on the side menu.

Colour code

Questions with white background are aimed at (describing) the features of the AI system.

Answers to blue questions will contribute to recommendations.

Text in red allows you to self-assess your organisation's compliance with the respective requirement.

<https://futurium.ec.europa.eu/en/european-ai-alliance/pages/welcome-altai-portal>
<https://altai.insight-centre.org/Home/HowToComplete>

Measure
if your organisation's AI is
trustworthy



ALTAI – Assessment List for Trustworthy Artificial Intelligence

<https://altai.insight-centre.org/>

INDEPENDENT
HIGH-LEVEL EXPERT GROUP ON
ARTIFICIAL INTELLIGENCE
SET UP BY THE EUROPEAN COMMISSION



THE ASSESSMENT LIST FOR
TRUSTWORTHY ARTIFICIAL
INTELLIGENCE (ALTAI)
for self assessment

AuditorIA: Auditando la IA con ALTAI

Based on the feedback received, the AI HLEG presented the final [Assessment List for Trustworthy AI \(ALTAI\)](#) in July 2020. ALTAI is practical tool that translates the Ethics Guidelines into an accessible and dynamic (self-assessment) checklist. The checklist can be used by developers and deployers of AI who want to implement the key requirements in practice. This new list is available as a prototype [web based tool](#) and in [PDF format](#).

Measure
if your organisation's AI is
trustworthy



ALTAI – Assessment List for Trustworthy Artificial Intelligence

<https://altai.insight-centre.org/>

My ALTAs



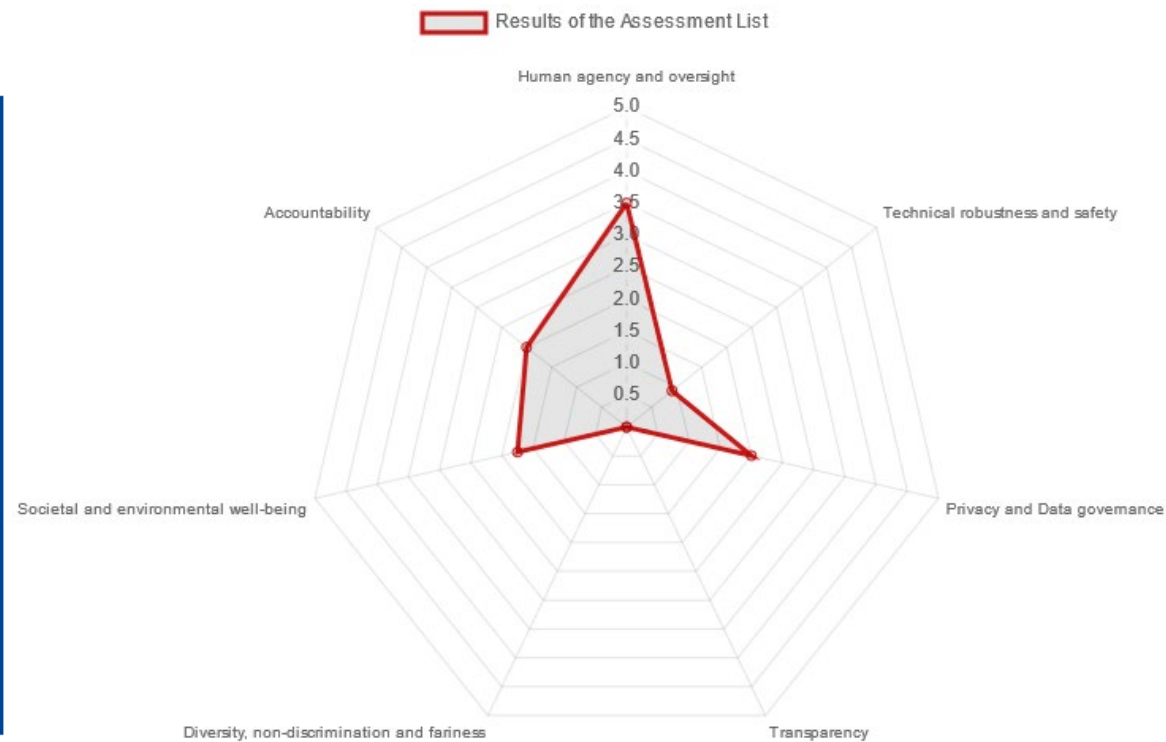
ChatGPT-pilot

View

Delete

Sections of the ALTAI

- Human Agency and Oversight
- Technical Robustness and Safety
- Privacy and Data Governance
- Transparency
- Diversity, Non-Discrimination and Fairness
- Societal and Environmental Well-being
- Accountability



<https://futurium.ec.europa.eu/en/european-ai-alliance/pages/welcome-altai-portal>
<https://altai.insight-centre.org/Home/HowToComplete>

INDEPENDENT
HIGH-LEVEL EXPERT GROUP ON
ARTIFICIAL INTELLIGENCE
SET UP BY THE EUROPEAN COMMISSION



THE ASSESSMENT LIST FOR
TRUSTWORTHY ARTIFICIAL
INTELLIGENCE (ALTAI)
for self assessment

Ethics by Design

https://ec.europa.eu/info/funding-tenders/opportunities/docs/2021-2027/horizon/guidance/ethics-by-design-and-ethics-of-use-approaches-for-artificial-intelligence_he_en.pdf



Ethics by Design is an approach for **systematically and comprehensively including ethical considerations in the design and development process of new technological systems and devices**

The central idea behind it is that technology can be made ethical **through its design process**

Technology is not neutral / **values can be embedded into the design process**

Design choices are not morally neutral but can have significant ethical consequences

Design choices can sometimes generate particular consequences across a broad range of uses or use contexts

Ethics by Design

https://ec.europa.eu/info/funding-tenders/opportunities/docs/2021-2027/horizon/guidance/ethics-by-design-and-ethics-of-use-approaches-for-artificial-intelligence_he_en.pdf



Guidance Note on *human-centered AI: algorithmic bias and fairness*

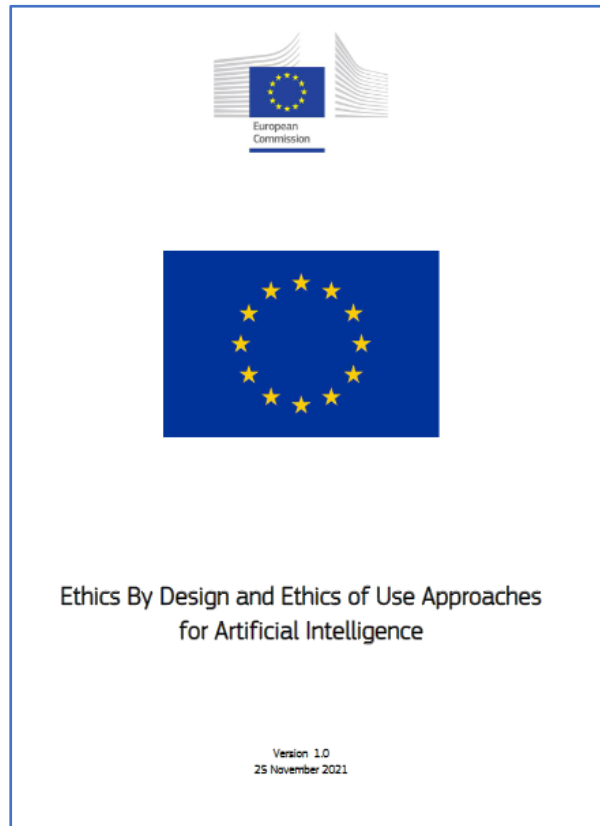
Guidance Note on *AI and informed consent*

Guidance Note on *AI Ethics Audits and Checks*

Guidance Note on *Human Rights Impact Assessment*

Ethics by Design

https://ec.europa.eu/info/funding-tenders/opportunities/docs/2021-2027/horizon/guidance/ethics-by-design-and-ethics-of-use-approaches-for-artificial-intelligence_he_en.pdf



Principles: These are the **ethics principles** / ethical values detailed in Part 1.

An AI system is considered unethical if it violates these **principles** values.

Ethical Requirements: These are the conditions that must be met for the AI system to achieve its goals ethically. These may be instantiated in many ways: through functionality, in data structures, in the process by which the system is constructed, with organisational safeguards and so forth.

Ethics by Design Guidelines: These are concerned with the processes for creating the system. In many cases guidelines are specific tasks which must be completed at specific points in the development process. The guidelines are either implementations of ethics requirements, or broader guidelines for different stages of developments that help ensure proper implementation of requirements.

AI Methodologies: There is a variety of methodologies used in AI and robotics projects (AGILE, CRISP-DM, V-Method etc). These are, at least partially, distinguished by the manner in which the development process is organized. Each methodology offers its own steps and sequence. Ethics by Design maps its guidelines onto the steps in each individual methodology.

Tools & Methods: specific tools and processes within the development process. For example, Datasheets for Datasets can be employed to assess the ethical characteristics of data.

Ética e IA en el mundo (OCDE)



Organisation for Economic Co-operation and Development (OECD), Recommendation of the Council on Artificial Intelligence, OECD/LEGAL/0449, <https://legalinstruments.oecd.org/en/instruments/OECD-LEGAL-0449>, 2019



The Global Partnership on Artificial Intelligence (GPAI) is a multi-stakeholder initiative which aims to bridge the gap between theory and practice on AI by supporting cutting-edge research and applied activities on AI-related priorities.

Built around a shared commitment to the OECD Recommendation on Artificial Intelligence, GPAI brings together engaged minds and expertise from science, industry, civil society, governments, international organisations and academia to foster international cooperation.

Launched in June 2020 with 15 members, GPAI includes now 29 members

- > Argentina
- > Australia
- > Belgium
- > Brazil
- > Canada
- > Czech Republic
- > Denmark
- > France
- > Germany
- > India (Lead Chair 2024)
- > Ireland
- > Israel
- > Italy
- > Japan (Outgoing Chair 2024)
- > Republic of Korea
- > Mexico
- > Netherlands
- > New Zealand
- > Poland
- > Senegal
- > Serbia (Incoming Chair 2024)
- > Singapore
- > Slovenia
- > Spain
- > Sweden
- > Türkiye
- > United Kingdom
- > United States
- > European Union

Ética e IA en el mundo (OCDE)



- **Organisation for Economic Co-operation and Development (OECD)**, Recommendation of the Council on Artificial Intelligence, OECD/LEGAL/0449, <https://legalinstruments.oecd.org/en/instruments/OECD-LEGAL-0449>, 2019



GPAI / THE GLOBAL PARTNERSHIP ON ARTIFICIAL INTELLIGENCE

<https://www.gpai.ai/>

Consensus about the future vision of GPAI: **To pave way for renewed integrated partnership for harnessing the potential of AI for Good and for All** (6th meeting of the GPAI Ministerial Council held on 3rd July 2024 at New Delhi)

Our values

Members and participants of GPAI are brought together first and foremost by a shared commitment to the values expressed in the **OECD Recommendation on Artificial Intelligence**. All GPAI activities are intended to foster responsible development of AI grounded in these principles of **human rights, inclusion, diversity, innovation and economic growth**.

Principles for responsible stewardship of trustworthy AI

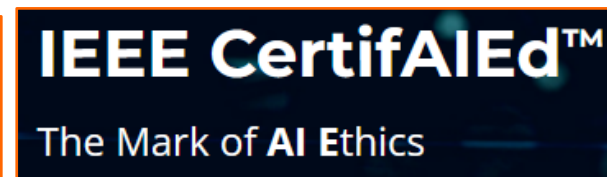
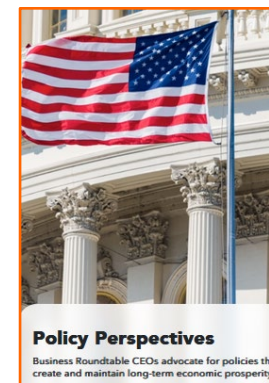
- › Inclusive growth, sustainable development and well-being
- › Human-centred values and fairness
- › Transparency and explainability
- › Robustness, security and safety
- › Accountability

National policies and international cooperation for trustworthy AI

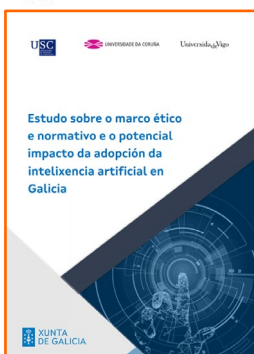
- › Investing in AI research and development
- › Fostering a digital ecosystem for AI
- › Shaping an enabling policy environment for AI
- › Building human capacity and preparing for labour market transformation
- › International cooperation for trustworthy AI



Códigos de Conducta Ética y Gobernanza de la IA



The European Social Charter
a guardian of our social rights



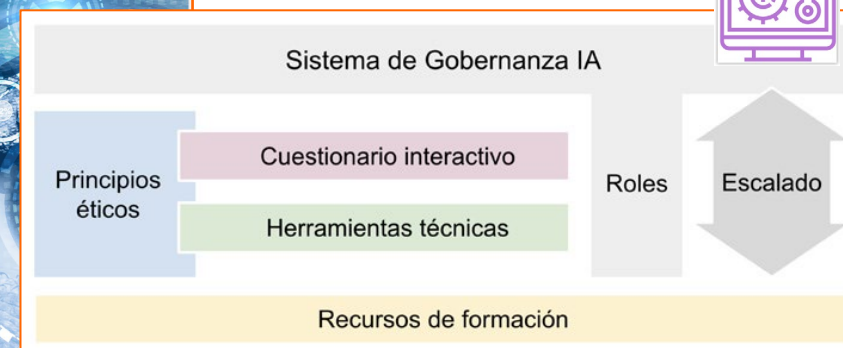
Ethical Principles & Human Rights:

1. Respect for Human Agency
2. Prevention of Harm
3. Fairness
4. Explainability

INDESIA ODISEIA

Informe Ética & IA Industrial

El uso responsable de la inteligencia artificial en el sector industrial



Ética e IA en el mundo



The United Nations Educational, Scientific and Cultural Organization (UNESCO) Recommendation on the Ethics of Artificial Intelligence (2021) <https://unesdoc.unesco.org/ark:/48223/pf0000381137>



<https://on.unesco.org/EthicsAI>

VALUES:

- (1) Respect, protection and promotion of human rights as well as fundamental freedoms and human dignity
- (2) Environment and ecosystem flourishing
- (3) Ensuring diversity and inclusiveness
- (4) Living in peaceful, just and interconnected societies

PRINCIPLES:

- (1) Proportionality and Do No Harm
- (2) Safety and security
- (3) Fairness and non-discrimination
- (4) Sustainability
- (5) Right to Privacy, and Data Protection
- (6) Human oversight and determination
- (7) Transparency and explainability
- (8) Responsibility and accountability
- (9) Awareness and literacy
- (10) Multi-stakeholder and adaptive governance and collaboration

POLICY AREAS:

- (1) Ethical Impact Assessment
- (2) Ethical Governance and Stewardship
- (3) Data Policy
- (4) Development and International Cooperation
- (5) Environment and Ecosystems
- (6) Gender
- (7) Culture
- (8) Education and Research
- (9) Communication and Information
- (10) Economy and Labour
- (11) Health and Social Well-being



Sostenibilidad e IA



<https://sdgs.un.org/goals>

- ✓ The Sustainable Development Goals, also known as the Global Goals, were adopted by the United Nations in 2015 as **a universal call to action to end poverty, protect the planet, and ensure that by 2030 all people enjoy peace and prosperity**
- ✓ The 17 goals are integrated—they recognize that action in one area will affect outcomes in others, and that **development must balance social, economic and environmental sustainability**

Sociedad Es noticia

El auge de la inteligencia artificial acelera las emisiones contaminantes

Sociedad > Ciencia Educación Medio ambiente Sanidad Caso Abierto Tiempo Tráfico y transportes Futuro Energía

EN PORTADA LA SAREB Y LA GENERALITAT NEGOCIAN LA MAYOR VENTA DE VIVIENDAS A UNA COMUNIDAD

Barcelona
16 de abril del 2023.

Impacto climático de la IA

El auge de la inteligencia artificial acelera las emisiones contaminantes

La falta de transparencia de las grandes empresas tecnológicas dificulta calibrar la huella de carbono de unos sistemas que requieren de cada vez más recursos como energía o agua

Opinión

Verde que te quiero verde. ¿Cuál es la huella ambiental de la IA?

Opinión > Editoriales Autores The Conversation 17 de abril del 2023.

Tecnología y medio ambiente Opinión

Verde que te quiero verde. ¿Cuál es la huella ambiental de la IA?

Ser más digitales, hoy, no implica que el uso extensivo de estas tecnologías tenga un impacto ambiental neutro

Ulises Cortés
Catedrático de Inteligencia Artificial de la UPC e investigador del Barcelona Supercomputing Center (BSC).

2 Se lee en minutos

13 CLIMATE ACTION 14 LIFE BELOW WATER 15 LIFE ON LAND 16 PEACE, JUSTICE AND STRONG INSTITUTIONS 17 PARTNERSHIPS FOR THE GOALS





<https://codecarbon.io/>



In lbs of CO2 equivalent

<https://www.worldometers.info/co2-emissions/co2-emissions-per-capita/>

http://www.uigi.com/co2_quantity_convert.html

Roundtrip flight b/w NY and SF (1 passenger) >>

1,984 **0.992 tons**

Human life (avg. 1 year) >>

11,023 **5.5 tons (5.4 in Spain)**

US car including fuel (avg. 1 lifetime) >>

126,000 **63 tons**

AlphaGo Zero >>

211,643 **105.8 tons**

Transformer (213M parameters) >>

626,155 **313.1 tons**

GPT-3 (175B parameters) >>

1,212,172 **606.1 tons**



539 tons (70B)



LATXA 4.4 tons (70B)

606.1 tons

Z. Chen, M. Wu, A. Chan, X. Li, Y.-S. Ong, "Survey on AI Sustainability: Emerging Trends on Learning Algorithms and Research Challenges", IEEE Computational Intelligence Magazine, 18(2):60-77, 2023, <https://doi.org/10.1109/MCI.2023.3245733>





Ethics Washing Is When Ethics is A Substitute for Regulation

By [Dataethics](#) on 2018-10-18

Paper. The rise of the ethical technology debate runs in parallel to the increasing resistance to any regulation at all. Assistant professor Ben Wagner is worried that the focus on ethics can be used as a way of getting around regulation. He argues that we need common criteria for ethics or we risk that the many different ethics frameworks become arbitrary, optional or meaningless.



**Data & Society:
Points**

Points is the blog of
Data & Society
Research Institute

By using the language of bias, we may end up overly focusing on the individual intents of technologists involved, rather than the structural power of the institutions they belong to.



Aspectos Legales

Auditabilidad y Gobernanza de la IA

Policies, data and analysis for trustworthy artificial intelligence

This website is currently being updated further to the announcement of an integrated partnership for the Global Partnership on AI (GPAI) with the OECD.

Global AI Regulations: Complex & Dynamic

 Argentina	 Czechia	 Italy	 Nigeria	 Sweden
 Armenia	 Denmark	 Japan	 Norway	 Switzerland
 Australia	 Egypt	 Kazakhstan	 Peru	 Thailand
 Austria	 Estonia	 Kenya	 Poland	 Tunisia
 Belgium	 Finland	 Korea	 Portugal	 Türkiye
 Brazil	 France	 Latvia	 Romania	 Uganda
 Bulgaria	 Germany	 Lithuania	 Rwanda	 Ukraine
 Canada	 Greece	 Luxembourg	 Saudi Arabia	 United Arab Emirates
 Chile	 Hungary	 Malta	 Serbia	 United Kingdom
 China	 Iceland	 Mauritius	 Singapore	 United States
 Colombia	 India	 Mexico	 Slovakia	 Uruguay
 Costa Rica	 Indonesia	 Morocco	 Slovenia	 Uzbekistan
 Croatia	 Ireland	 Netherlands	 South Africa	 Viet Nam
 Cyprus	 Israel	 New Zealand	 Spain	 European Union

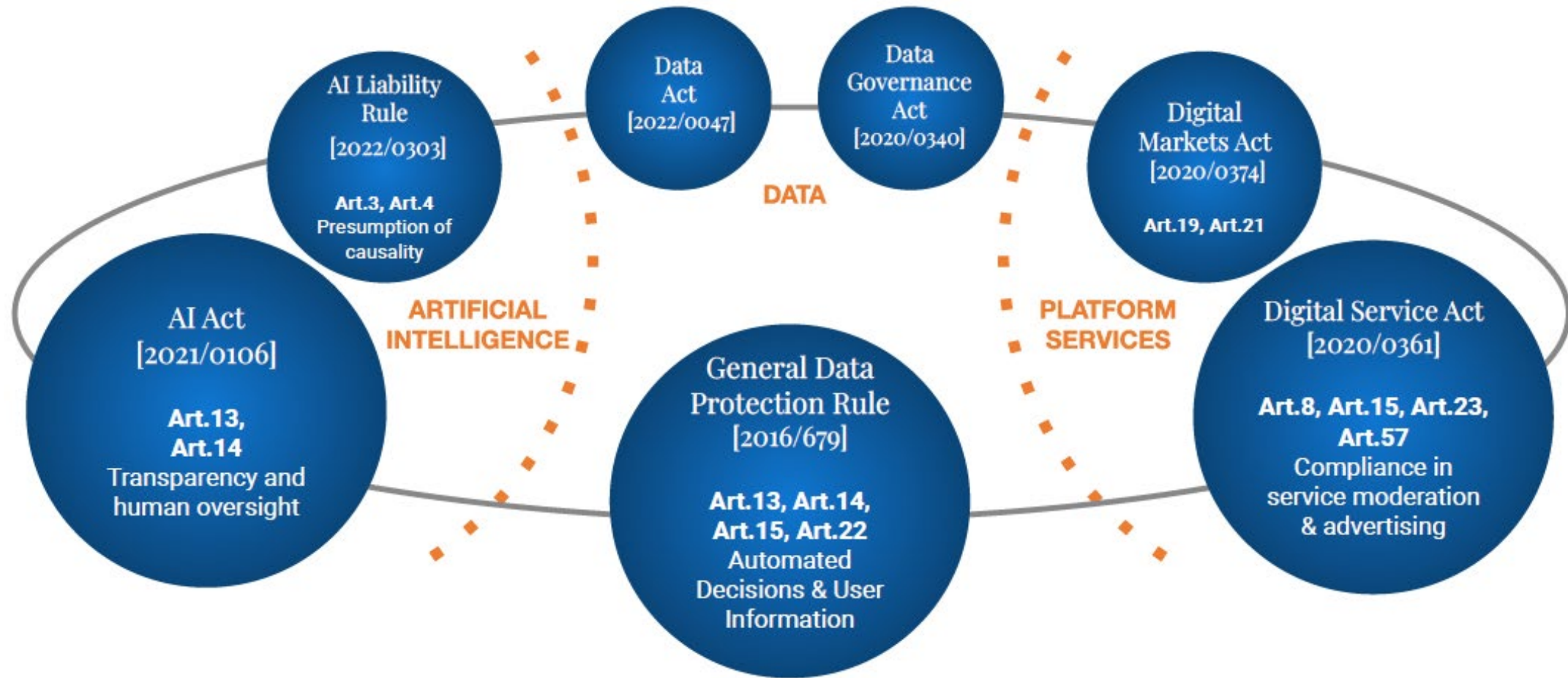
AI Principles



Data Privacy and Security

Source: [OECD \(https://oecd.ai/en/\)](https://oecd.ai/en/)

ELSEC-AI y Regulación en Europa



L. Nannini, J.M. Alonso-Moral, A. Catalá, M. Lama, S. Barro, “Operationalizing explainable AI in the EU regulatory ecosystem”, IEEE Intelligent Systems, 2024, <https://doi.org/10.1109/MIS.2024.3383155>

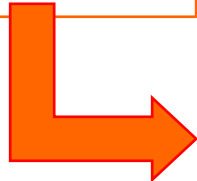
General Data Protection Regulation



European AI Regulation

Regulation (EU) 2016/679 on the protection of natural persons with regard to the processing of personal data and on the free movement of such data, and repealing Directive 95/46/EC (**General Data Protection Regulation**) [2016] OJ L119/1
<http://eur-lex.europa.eu/legal-content/en/TXT/?uri=CELEX%3A32016R0679>

- (1) Right of access
- (2) Right to rectification
- (3) Right to erasure
- (4) Right to data portability
- (5) Right to be informed
- (6) Right to an explanation



Representative for
Data Protection
dpd@usc.es

The data subject should have the right not to be subject to a decision, which may include a measure, evaluating personal aspects relating to him or her which is based solely on automated processing and which produces legal effects concerning him or her or similarly significantly affects him or her, such as automatic refusal of an online credit application or e-recruiting practices without any human intervention. Such processing includes 'profiling' that consists of any form of automated processing of personal data evaluating the personal aspects relating to a natural person, in particular to analyse or predict aspects concerning the data subject's performance at work, economic situation, health, personal preferences or interests, reliability or behaviour, location or movements, where it produces legal effects concerning him or her or similarly significantly affects him or her. However, decision-making based on such processing, including profiling, should be allowed where expressly authorised by Union or Member State law to which the controller is subject, including for fraud and tax-evasion monitoring and prevention purposes conducted in accordance with the regulations, standards and recommendations of Union institutions or national oversight bodies and to ensure the security and reliability of a service provided by the controller, or necessary for the entering or performance of a contract between the data subject and a controller, or when the data subject has given his or her explicit consent. In any case, such processing should be subject to suitable safeguards, which should include specific information to the data subject and the right to obtain human intervention, to express his or her point of view, to obtain an explanation of the decision reached after such assessment and to challenge the decision. Such measure should not concern a child.

The AI Act



<https://artificialintelligenceact.eu/>

European AI Regulation

AI Act



On 9 December 2023, the European Parliament and the Council reached a political agreement on the AI Act : It contains art.86, named “right to explanation of individual decision-making”

- ❑ 2024: **The AI Act is adopted by all member states** (the AI Act entered into force on the 1st Aug 2024)
- ❑ 2025 – 2026 (expected): Following about 36-month transition period, the **AI Act comes into effect**

Risk based approach

Emphasis on the protection of fundamental rights

Multi-level governance structure

Strict fines and penalties

150-page long document

Lots of delegated/implementing acts clauses

Extra-territorial application

The AI Act - highlights



Comprehensive rules for trustworthy AI

- ✓ Safeguards agreed on General-Purpose AI (GPAI)
- ✓ **Limitation for the use of biometric identification systems** by law enforcement
- ✓ **Bans on social scoring** and AI used to manipulate or exploit user vulnerabilities
- ✓ **Right of consumers** to launch complaints and receive **meaningful explanations**
- ✓ Fines ranging from 35 million euro or 7% of global turnover to 7.5 million or 1.5% of turnover

<https://artificialintelligenceact.eu/>

European AI Regulation

AI Act



Source: <https://www.europarl.europa.eu/news/en/press-room/20231206IPR15699/artificial-intelligence-act-deal-on-comprehensive-rules-for-trustworthy-ai>

The AI Act

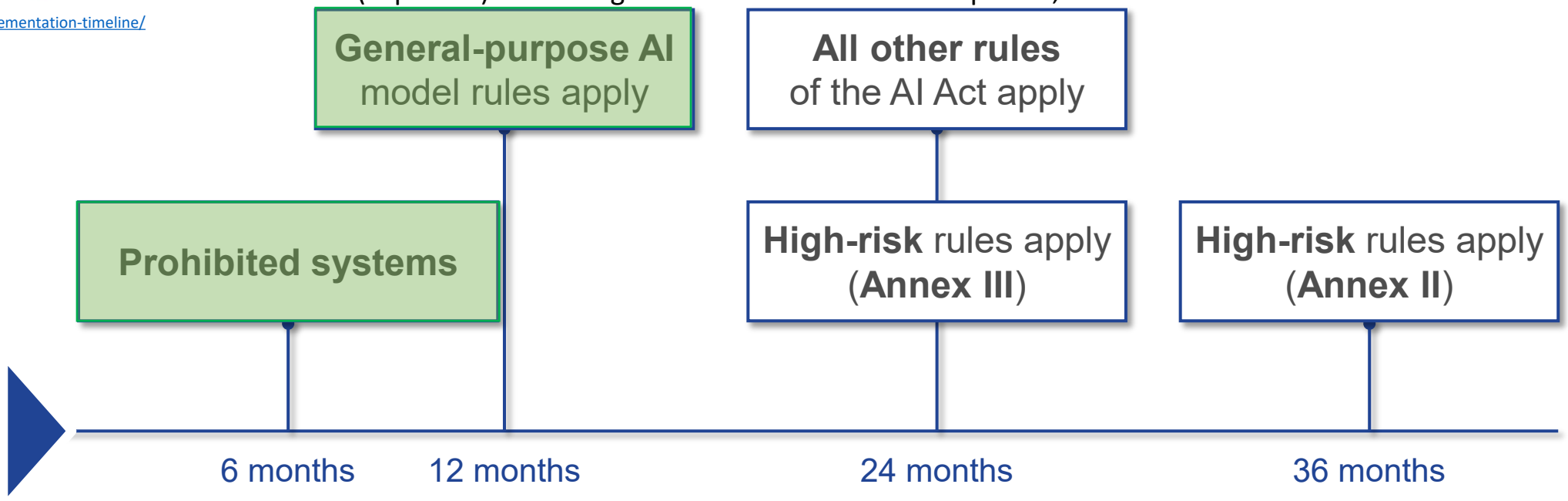


<https://artificialintelligenceact.eu/implementation-timeline/>

On 9 December 2023, the European Parliament and the Council reached a political agreement on the AI Act : It contains art.86, named “right to explanation of individual decision-making”

- ❑ 2024: **The AI Act is adopted by all member states** (the AI Act entered into force on the 1st Aug 2024)
- ❑ 2025 – 2026 (expected): Following about 36-month transition period, the **AI Act comes into effect**

AI Act entry into force *



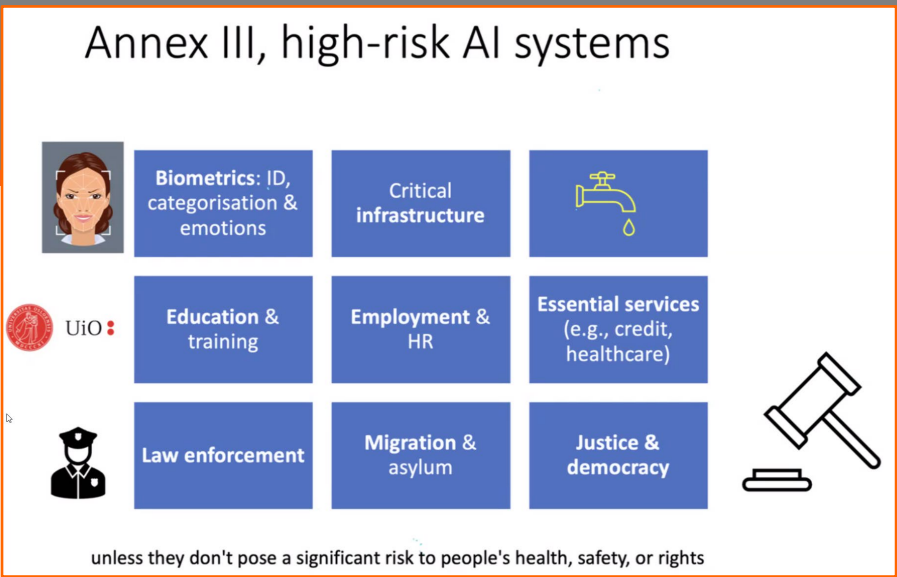
*Following its adoption by the European Parliament and the Council, the AI Act shall enter into force on the twentieth day following that of its publication in the official Journal.

The AI Act



<https://artificialintelligenceact.eu/>

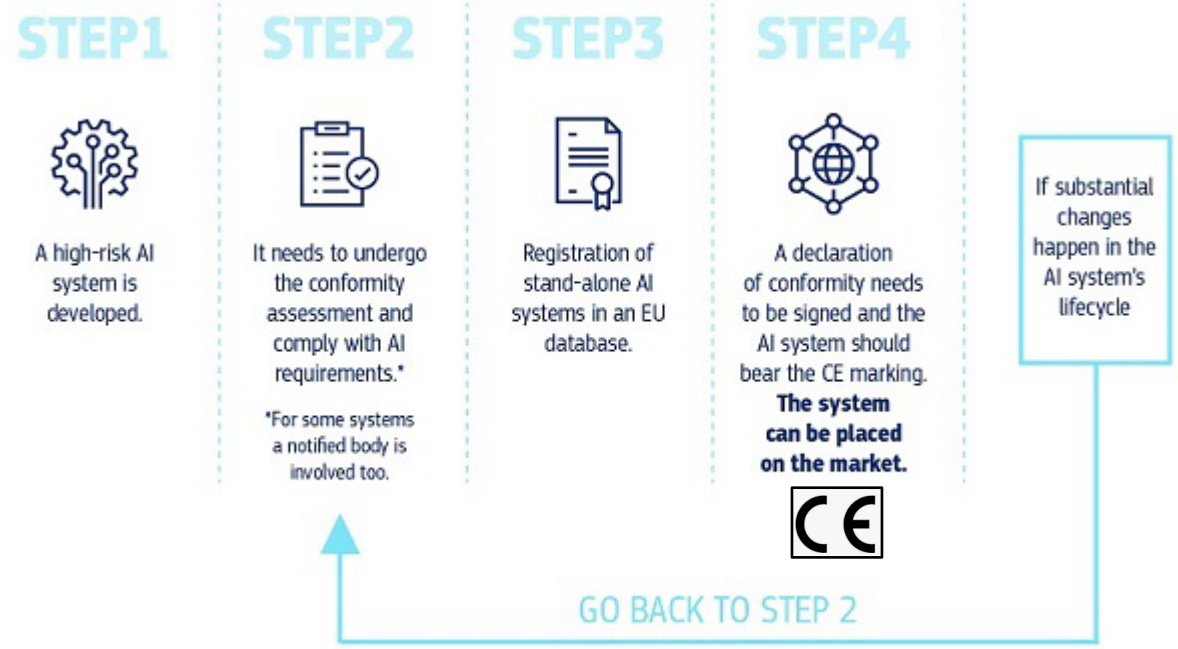
European AI Regulation AI Act



❖ High-risk AI systems

- critical infrastructures (e.g., **transport**), that could put the life and health of citizens at risk
- educational or vocational training, that may determine the access to education and professional course of someone's life (e.g., **scoring of exams**)
- safety components of products (e.g., AI application in **robot-assisted surgery**)
- employment, management of workers and access to self-employment (e.g., CV-sorting software for **recruitment procedures**)
- essential private and public services (e.g., **credit scoring** denying citizens opportunity to obtain a loan)
- law enforcement that may interfere with people's fundamental rights (e.g., **evaluation of the reliability of evidence**)
- migration, asylum and border control management (e.g., **verification of authenticity of travel documents**)
- administration of justice and democratic processes (e.g., **applying the law** to a concrete set of facts)

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



AI Act – implementación y supervisión



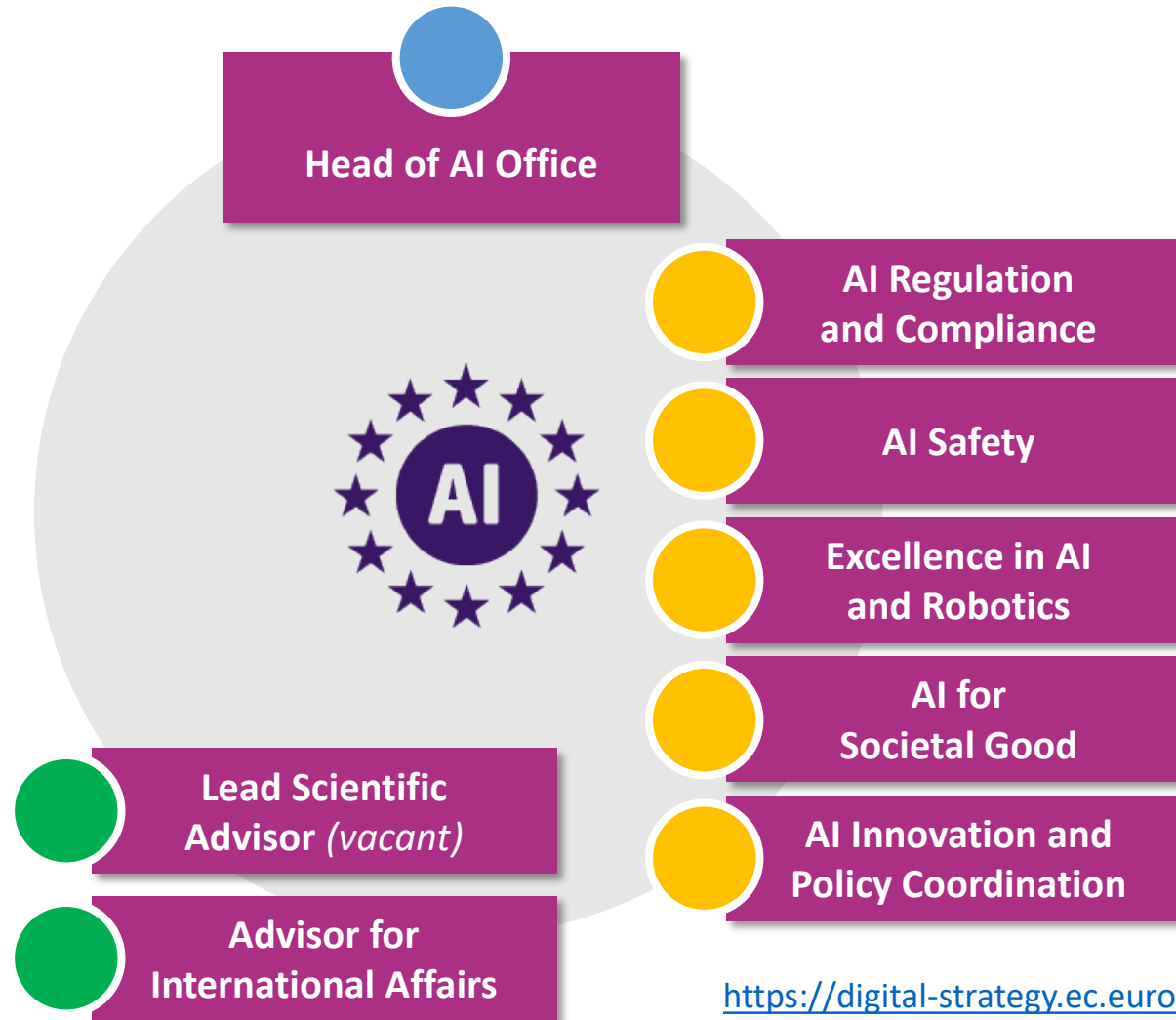
EUROPEAN ARTIFICIAL INTELLIGENCE OFFICE

The European AI Office

- Key role in the implementation of the AI Act, in particular in the enforcement of rules on general-purpose AI models
- Cooperation with all relevant EU bodies, Member States and the stakeholder community
- Cross-sectoral cooperation within the Commission
- Fostering research and innovation in trustworthy AI
- Positioning the EU as a leader in international discussions and contributor to AI for good

Context:

- ❖ AI Act introduces role of AI Office
- ❖ Established by Commission Decision 2024/390
- ❖ Part of DG CNECT, launched in June 2024



<https://artificialintelligenceact.eu/>



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

AI Act – implementación y supervisión



<https://artificialintelligenceact.eu/article/99/>

European AI Regulation

AI Act



Part of Chapter XII: Penalties

Article 99: Penalties

Date of entry into force: 2 August 2025
 According to: Article 113(b)
 Inherited from: Chapter XII

The general principle of the AI Act is that penalties shall be *effective, dissuasive, and proportionate* to the type of offense, previous actions, and profile of the offender. As such, the EU AI Act acknowledges that **each case is individual** and designates the fines as a maximum threshold, although lower penalties can be issued depending on the severity of the offense.

Penalties target

- **Operators of AI systems**
- **Providers of GPAI models**
- **Union institutions, agencies, and bodies**

Up to 35 000 000 EUR or, if the offender is a company, up to 7 % of its total worldwide annual turnover for the preceding financial year, whichever is higher	Non-compliance with the prohibition of the artificial intelligence practices under Article 5
Up to 15 000 000 EUR or, if the offender is a company, up to 3% of its total worldwide annual turnover for the preceding financial year, whichever is higher	Non-compliance with the following: <ul style="list-style-type: none"> • obligations of providers pursuant to Article 16 • obligations of authorised representatives pursuant to Article 25 • obligations of importers pursuant to Article 26 • obligations of distributors pursuant to Article 27 • obligations of deployers pursuant to Article 29, paragraphs 1 to 6a • requirements and obligations of notified bodies pursuant to Article 33, 34(1), 34(3), 34(4), 34a • transparency obligations for providers and users pursuant to Article 52
Up to 7 500 000 EUR or, if the offender is a company, up to 1 % of its total worldwide annual turnover for the preceding financial year, whichever is higher	The supply of incorrect, incomplete or misleading information to notified bodies and national competent authorities in reply to a request
Up to 15 000 000 EUR or, if the offender is a company, up to 3% of its total worldwide annual turnover for the preceding financial year, whichever is higher	In one of the following: <ul style="list-style-type: none"> • Infringement of the GPAI-relevant provisions • Failure to comply with a request for document or information pursuant to Article 68i or supply of incorrect, incomplete or misleading information • Failure to comply with a measure requested under Article 68k • Failure to make available to the Commission access to the general purpose AI model or general purpose AI model with systemic risk with a view to conduct an evaluation pursuant to Article 68j
Up to EUR 1 500 000	Non-compliance with the prohibition of the artificial intelligence practices under Article 5
Up to EUR 750 000	Non-compliance of the AI system with any requirements or obligations

Economy and Markets, Machine Learning

Predictions for AI in 2025: Collaborative Agents, AI Skepticism, and New Risks

<https://hai.stanford.edu/news/predictions-ai-2025-collaborative-agents-ai-skepticism-and-new-risks>



- Agents of AI, Less Regulation
- More Scams, Less Consumer Protection
- ‘General Contractor’ LLMs
- **Healthy Skepticism in Education AI**
- Defining Value from GenAI
- AI Agents Work Together
- **Rethinking Human-AI Collaboration**



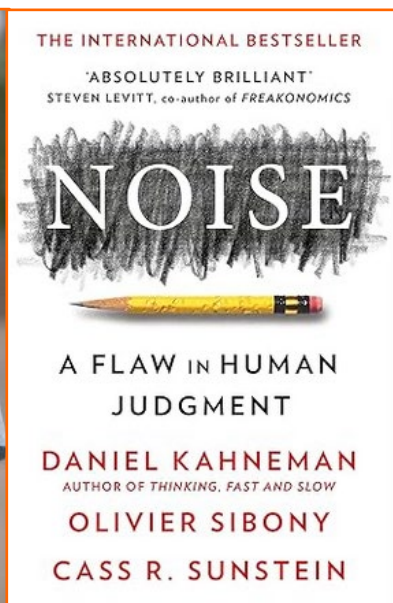
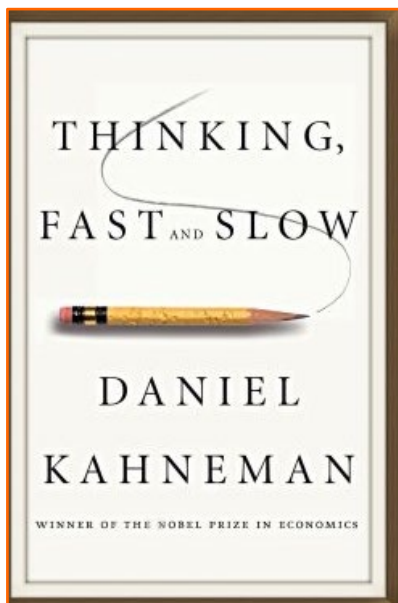
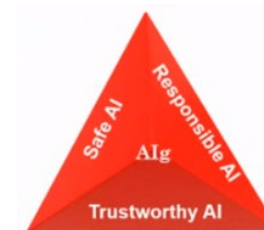
THE END

udima

EduIA Talks



Trabajemos Juntos por un Uso Responsable de la IA en la Educación Digital



“Optimism is normal, but some fortunate people are more optimistic than the rest of us. If you are genetically endowed with an optimistic bias, you hardly need to be told that you are a lucky person - you already feel fortunate.”



NL4XAI

Interactive Natural Language
Technology for eXplainable
Artificial Intelligence

Supporting the right to
explanation by

AI
POWERED
SYSTEMS

NL4XAI.eu



Jose Maria Alonso-Moral



[https://orcid.org/
0000-0003-3673-421X](https://orcid.org/0000-0003-3673-421X)



Profesor Titular de Universidad



Centro Singular de Investigación
en Tecnoloxías Intelixentes

josemaria.alonso.moral@usc.es

<https://citius.gal/team/jose-maria-alonso-moral>

<https://gitlab.citius.usc.es/jose.alonso/xai>

International Ambassador



<https://fmsquare.org/>

Z-Inspection®

<https://z-inspection.org/>



Task Force on Explainable Fuzzy Systems

SHIELD

<https://sites.google.com/view/tf-explainable-fuzzy-systems/>

TECHNICAL COMMITTEE



XAI for Healthy Aging and Social Wellbeing

<https://xai4soc.es/>



TELSEC4TAI

R&D Network on Trustworthy Artificial Intelligence: Technical, Ethical, Legal, Cultural and Socio-economic Challenges

<https://iafiable.es/>



<https://confia.gal>

Interreg



España - Portugal

<https://www.transfiresaude.eu/>



Intelligent Reading Improvement
System for Fundamental and
Transversal Skills Development

<https://iread4skills.com/>

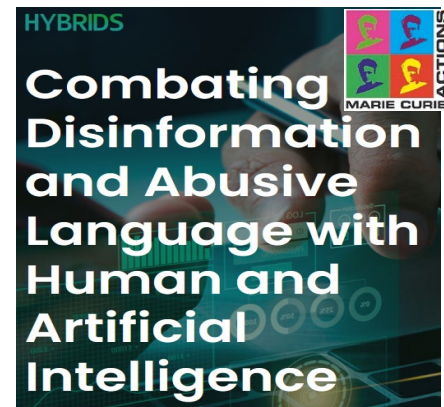


<https://www.sosfood-project.eu/>

DeepR3: Reducing, Reusing, and Recycling large models for developing Responsible and Green Language Technologies



<http://ixa2.si.ehu.eus/deepr3/>



<https://hybridsproject.eu/>



This project has received funding from the European Union's Horizon 2020 research and innovation programme under the Marie Skłodowska-Curie Grant Agreement No. 800621.