

## VIII JORNADA DE PATRIMONI CULTURAL

*LA INTEL·LIGÈNCIA  
ARTIFICIAL APLICADA  
AL PATRIMONI  
CULTURAL*



Generalitat de Catalunya  
Agència Catalana  
del Patrimoni Cultural



## Patrimoni cultural i IA

*Com no perdre el nord en un moment de canvis accelerats*

**Karma Peiró**

17 d'octubre de 2024. Auditori La Pedrera . Barcelona

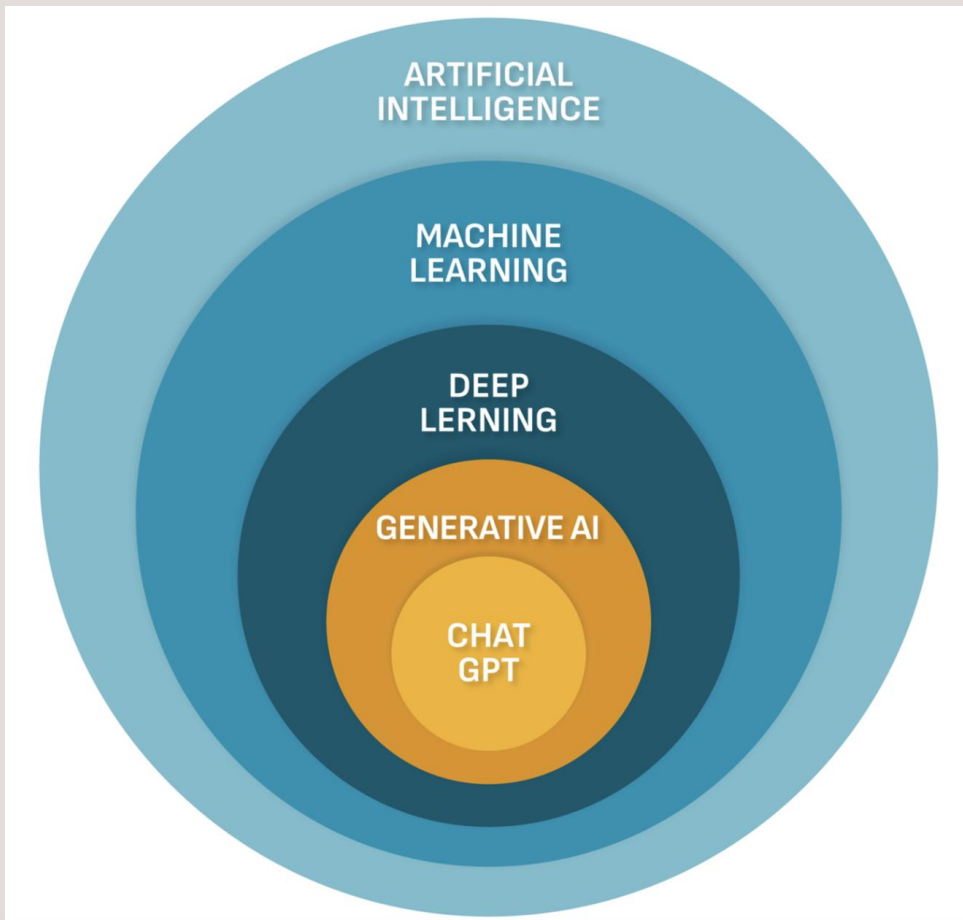
# Què és la IA... avui?

La intel·ligència artificial potencia que les màquines funcionin i reaccionin com els humans.

Per aconseguir aquest objectiu, es fan servir algorismes sofisticats que prediuen i prenen decisions de manera automatitzada.

Algunes de les aplicacions d'IA en la nostra vida quotidiana més conegudes són el reconeixement facial o d'imatges, assistents virtuals, recomanadors, robots socials, etc.

La IA no és nova, fa 50 anys que existeix:  
dades massives + capacitat de computació



Els algoritmes més sofisticats fan servir **aprenentatge automàtic** (*machine learning*) i **aprenentatge profund** (*deep learning*).

Dins d'aquest, es troba la **IA Generativa**, tan popular per l'aplicació *ChatGPT*, entre d'altres

# ¿Qué NO es la IA... en 2024?

Acabar con la visión antropomórfica

TECNOLOGÍA

¿Se cumple la profecía? Se revela una inteligencia artificial contra los humanos y parece comenzar una batalla



## La Inteligencia Artificial y el futuro de la humanidad: ¿Un nuevo amanecer o el ocaso de nuestra era?

¿Cuándo la inteligencia artificial hará todas las tareas humanas? La predicción se acorta cada vez más

# Errors, Biases and Overconfidence in Artificial Emotional Modeling

October 2019

DOI: [10.1145/3358695.3361749](https://doi.org/10.1145/3358695.3361749)

Conference: IEEE/WIC/ACM International Conference on Web Intelligence

 Valentina Franzoni ·  Jordi Vallverdu ·  Alfredo Milani

Overview

Stats

Comments (1)

Citations (23)

References (5)

## Abstract

With the diffusion and successful new implementation of several machine learning techniques, together with the substantial cost decrease of sensors, also included in mobile devices, the field of emotional analysis and modeling has boosted. Apps, web apps, brain-scanning devices, and Artificial Intelligence assistants often include emotion recognition features or emotional behaviors, but new researches contain, maintain, or create several design errors, which analysis is the main aim of this paper.

# Les emocions de les màquines

“**Intel·ligència Artificial emocional**”, “**Robots emocionals**”, o “**computació afectiva**” són termes que cada vegada escoltem més sovint en els diversos mitjans de comunicació. (S’està) projectant un escenari que a molta gent preocupa: el de les màquines que entenguin completament les emocions humanes o, encara més, tinguin emocions per si mateixes”.

(Franzoni, Vallverdu, and Milani 2019).



*"La inteligencia artificial que tenemos ahora es de habilidades sin comprensión; puede evaluar la probabilidad y predecir la siguiente palabra en un texto, pero para que haya una inteligencia artificial general tiene que haber comprensión".*

Ramón López de Mántaras, fundador del Instituto de Investigación de Inteligencia Artificial del CSIC

[480. Bárbara Villuendas, 2024](#)

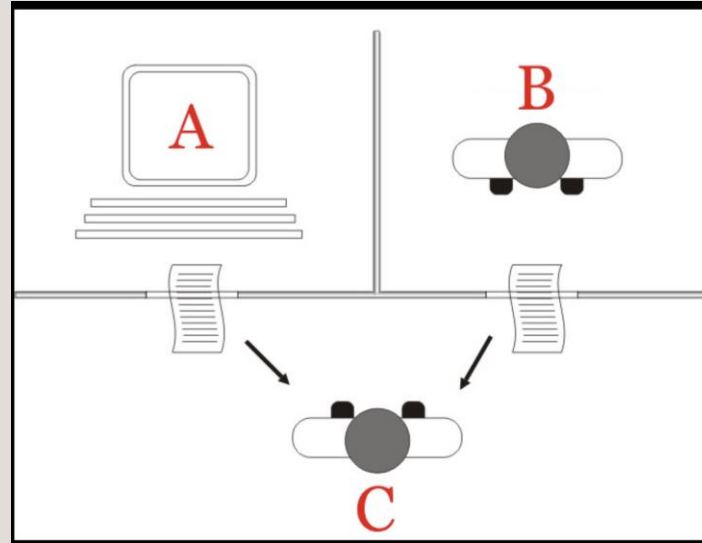
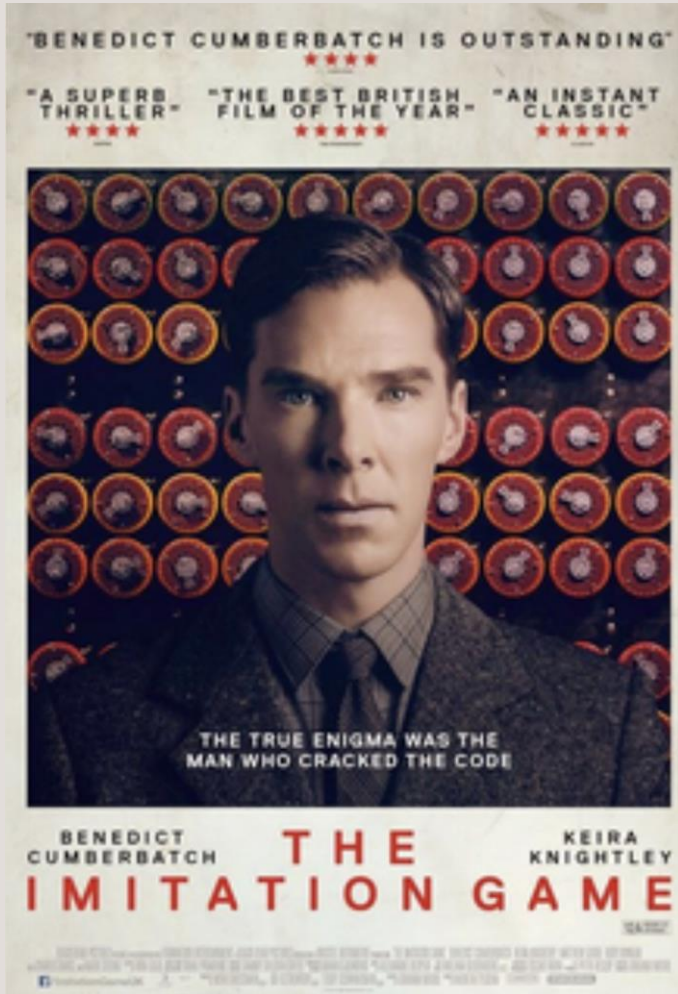
Ramon López de Mántaras (2024)

*"El problema no es la IA, sino el Dr. Frankenstein"*



[Alan Turing, "Computing Machinery and Intelligence", 1950](#)

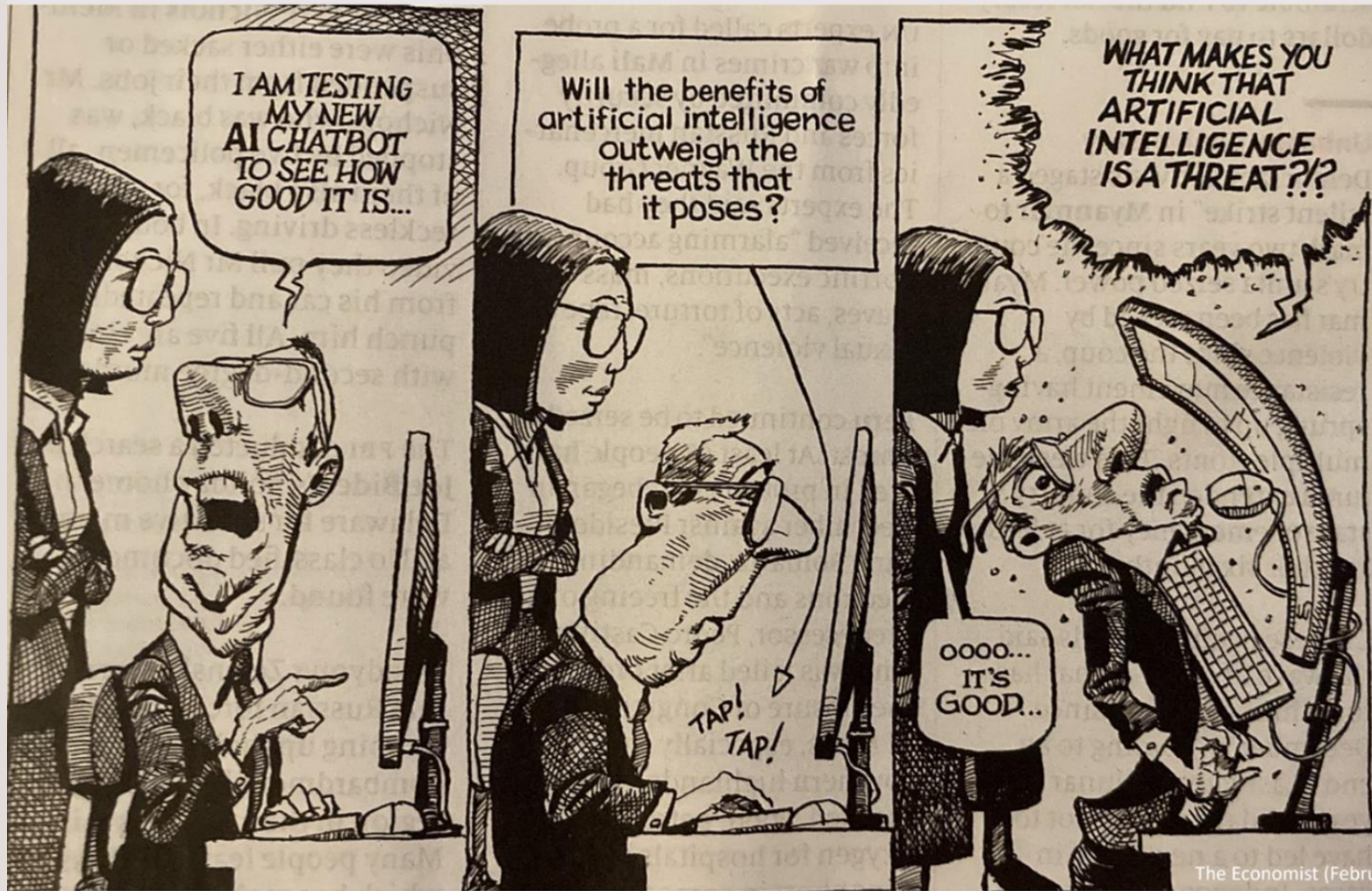
# “Can machines think?”



*“We can only see a short distance ahead, but we can see plenty there that needs to be done”, A.Turing, 1950*

1. Què és i què no és la IA?
2. **E**rrors per avançar
3. Oportunitats: IA +Patrimoni Cultural
4. Reptes per als propers anys
5. Conclusions





I AM TESTING  
MY NEW  
AI CHATBOT  
TO SEE HOW  
GOOD IT IS...

Will the benefits of  
artificial intelligence  
outweigh the  
threats that  
it poses?

WHAT MAKES YOU  
THINK THAT  
ARTIFICIAL  
INTELLIGENCE  
IS A THREAT?!?

TAP!  
TAP!

OOOO...  
IT'S  
GOOD...

# ChatGPT



## Examples

"Explain quantum computing in simple terms" →

"Got any creative ideas for a 10 year old's birthday?" →

"How do I make an HTTP request in Javascript?" →



## Capabilities

Remembers what user said earlier in the conversation

Allows user to provide follow-up corrections

Trained to decline inappropriate requests



## Limitations

May occasionally generate incorrect information

May occasionally produce harmful instructions or biased content

Limited knowledge of world and events after 2021

Send a message.



Free Research Preview. ChatGPT may produce inaccurate information about people, places, or facts. [ChatGPT May 12 Version](#)

## Faire attention!



Message ChatGPT



ChatGPT can make mistakes. Check important info.



OpenAI

[Research](#)

[Products](#)

[Safety](#)

[Company](#)



# Safety at every step

We believe in AI's potential to make life better for everyone, which means making it safe for everyone.

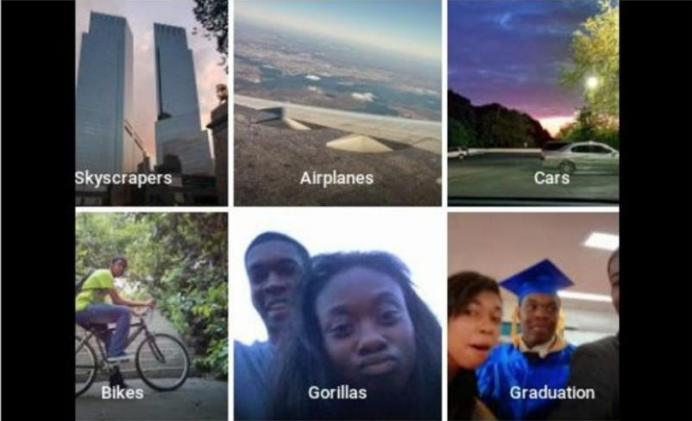

BBC Sign in Home News Sport Reel Worklife Travel

# NEWS

Tech

## Google apologises for Photos app's racist blunder

1 July 2015





The image shows a 2x3 grid of photos. The top row contains: 'Skyscrapers' (tall buildings), 'Airplanes' (airplane on a runway), and 'Cars' (a car on a road). The bottom row contains: 'Bikes' (a person on a bicycle), 'Gorillas' (two gorillas), and 'Graduation' (a person in a graduation cap).

## Errors de la IA en el passat...

REUTERS World Business Markets Breakingviews Video More

RETAIL OCTOBER 11, 2018 / 1:04 AM / UPDATED 3 YEARS AGO

### Amazon scraps secret AI recruiting tool that showed bias against women

By Jeffrey Dastin 8 MIN READ  

SAN FRANCISCO (Reuters) - Amazon.com Inc's [AMZN.O](#) machine-learning specialists uncovered a big problem: their new recruiting engine did not like women.

- Algorisme reconeixement imatges amb etiquetatge automàtic erroni
- Entrenament amb dades del passat esbiaixades

## *Wrongfully Accused by an Algorithm*

In what may be the first known case of its kind, a faulty facial recognition match led to a Michigan man's arrest for a crime he did not commit.

El primer arrest injustificat per una concordància defectuosa d'un algoritme de reconeixement facial



"This is not me," Robert Julian-Borchak Williams told investigators. "You think all Black men look alike?" Sylvia Jarrus for The New York Times

NATE

**RACIAL JUSTICE REQUIRES ALGORITHMIC JUSTICE. SUPPORT THE MOVEMENT.**

**TECHNOLOGY SHOULD  
SERVE ALL OF US. NOT JUST  
THE PRIVILEGED FEW.**

Join the Algorithmic Justice League in the movement towards equitable and accountable AI.

Enter your email

JOIN THE MOVEMENT

**CODED BIAS**



A SHALINI KANTAYYA FILM



Errors de la IA en el passat...

# Errors de la IA en el passat...



EL PAÍS

## Internacional

EUROPA · EE.UU · MÉXICO · AMÉRICA LATINA · ORIENTE PRÓXIMO · ASIA · ÁFRICA · FOTOS · OPINIÓN · ÚLTIMAS NOTICIAS

BORIS JOHNSON >

### Los estudiantes británicos derrotan al algoritmo de Johnson

Las manifestaciones fuerzan al Gobierno a rectificar el modelo matemático alternativo a los exámenes de selectividad, adoptado a causa de la pandemia y considerado injusto y elitista



Un grupo de estudiantes, en una protesta contra el nuevo método de evaluación propuesto por el Gobierno de Boris Johnson, en Godshall, este lunes. PAUL ELLIS (AFP)



R3D  
Red en Defensa  
de los Derechos Digitales

QUIÉNES SOMOS

QUÉ HACEMOS

BLOG

PUBLICACIONES



## UN ALGORITMO DISCRIMINATORIO EN PAÍSES BAJOS ACUSÓ A MILES DE FAMILIAS DE FRAUDE CONTRA EL SISTEMA DE BENEFICIOS SOCIALES

- Algorisme mal programat, amb més al centre que a l'alumne
- Algorisme discriminatori que tenia en compte variables no pertinents per concedir les ajudes

## El escándalo en un pequeño pueblo de España por las imágenes de decenas de niñas y jóvenes desnudas generadas por IA



## Errors de la IA en el passat...



Imagen viral retocada por IA del Papa Francisco El Androide libre

Fins a on es pot falsejar la realitat?  
*AI Undress App*

AI INCIDENT DATABASE

English | X | S | F | L | I | N | Sign Up

Discover + Submit

Welcome to the AIID

- Discover Incidents
- Spatial View
- Table View
- Entities
- Taxonomies
- Word Counts
- Submit Incident Reports
- Submission Leaderboard
- Blog
- AI News Digest
- Risk Checklists
- Random Incident

## Welcome to the AI Incident Database

Search over 3000 reports of AI harms

Search Discover

### Incident 817: AI-Generated Images Spread Misinformation During Hurricane Helene Response

**"The Fog of Disaster is Getting Worse"** Latest Incident Report  
theatlantic.com 2024-10-12

Keeping track of events during a natural disaster was hard enough in the past, before people with dubious motives started flooding social media with sensational images generated by artificial intelligence. In a crisis, public officials, fir...

Read More →

OECD AI Policy Observatory

Blog Live data Countries Priority issues Tools Resources About

Home AI Incidents Monitor

With the support of Patrick J. McGovern

## OECD AI Incidents Monitor (AIM)

Automated monitor of incidents and hazards from public sources (beta).

AI-related legislation is gaining traction, and effective policymaking needs evidence, foresight and international cooperation. The OECD AI Incidents Monitor (AIM) documents AI incidents and hazards to help policymakers, AI practitioners, and all stakeholders worldwide gain valuable insights into the risks and harms of AI systems. Over time, AIM will help to show risk patterns and establish a collective understanding of AI incidents and hazards and their multifaceted nature, serving as an important tool for trustworthiness.

Please see methodology and disclosures here.

Enter a concept or a keyword and press enter. Use advanced options to further narrow the search. Reset

ADVANCED SEARCH OPTIONS

### Summary visualisations of AI incidents & hazards

Evolution of incidents and hazards

### Summary statistics of AI incidents & hazards

	Incidents & hazards	Articles
All time total	13543	73369
Current month's total	170	692
Last month's total	524	2390
Peak month	2024-02	2024-01
Peak amount	842	4482
% change (month-over-month)	-1.87	-11.97
% change (quarter-over-quarter)	-20.96	-26.01
% change (year-over-year)	143.73	148.75

\*Note: Percent change is calculated based on preceding full months (i.e. the current

AI Risk Repository OVERVIEW DATABASE CAUSAL TAXONOMY DOMAIN TAXONOMY HOW TO USE FAQ NEWS

A Comprehensive Database of Risks from AI Systems

## What are the risks from Artificial Intelligence?

A comprehensive living database of over 700 AI risks categorized by their cause and risk domain.

Read preprint Explore database

Home Database About Context Harms Glossary Reclaiming AI Futures

# AI Observatory

'Artificial Intelligence', 'Machine Learning', 'Big Data' and 'Automation' are routinely used to invoke utopian vistas of objective, efficient and incorruptible systems with the potential to revolutionise society. We are told, repeatedly, that with enough data and computational power, applying the unbiased logic of machines and computer software, long-standing social problems – from crime to poverty – can be alleviated.

The language of 'AI' and the 'digital revolution', however, hides the crossed wires, the missing parts and the poisonous exhaust underneath the gleaming hoods of computers and data machines.

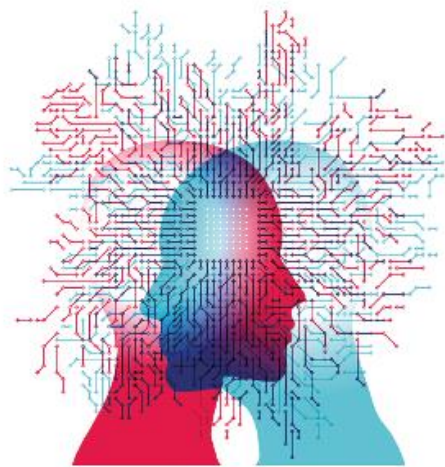
The potential of data, computers and networked technologies for social and scientific progress is undoubtable – they can possess the capacity for processing information and performing cognitive tasks which far surpass human ability.

The language of 'AI' and the 'digital revolution', however, hides the crossed wires, the missing parts and the poisonous exhaust underneath the gleaming hoods of computers and data machines.



# INTEL·LIGÈNCIA ARTIFICIAL

Decisions Automatitzades a Catalunya



apdcat

Generalitat  
de Catalunya



## 2. ADA EN ACCIÓ: TREBALL DE RECERCA A CATALUNYA

*Per Karma Peiró*

# AJUDA A LA DECISIÓ O AUTOMATITZATS?



*En l'àmbit judicial....*

*Predir la reincidència criminal RisCanvi*

*Reincidència criminal juvenil SAVRY*

*Predir la violència de gènere (VIOGEN)*

*Detectar les denúncies falses (VERIPOL)*

*IA per a detectar frau*

*Orientació per a l'extradició (o no) de migrants*

ARTICLES

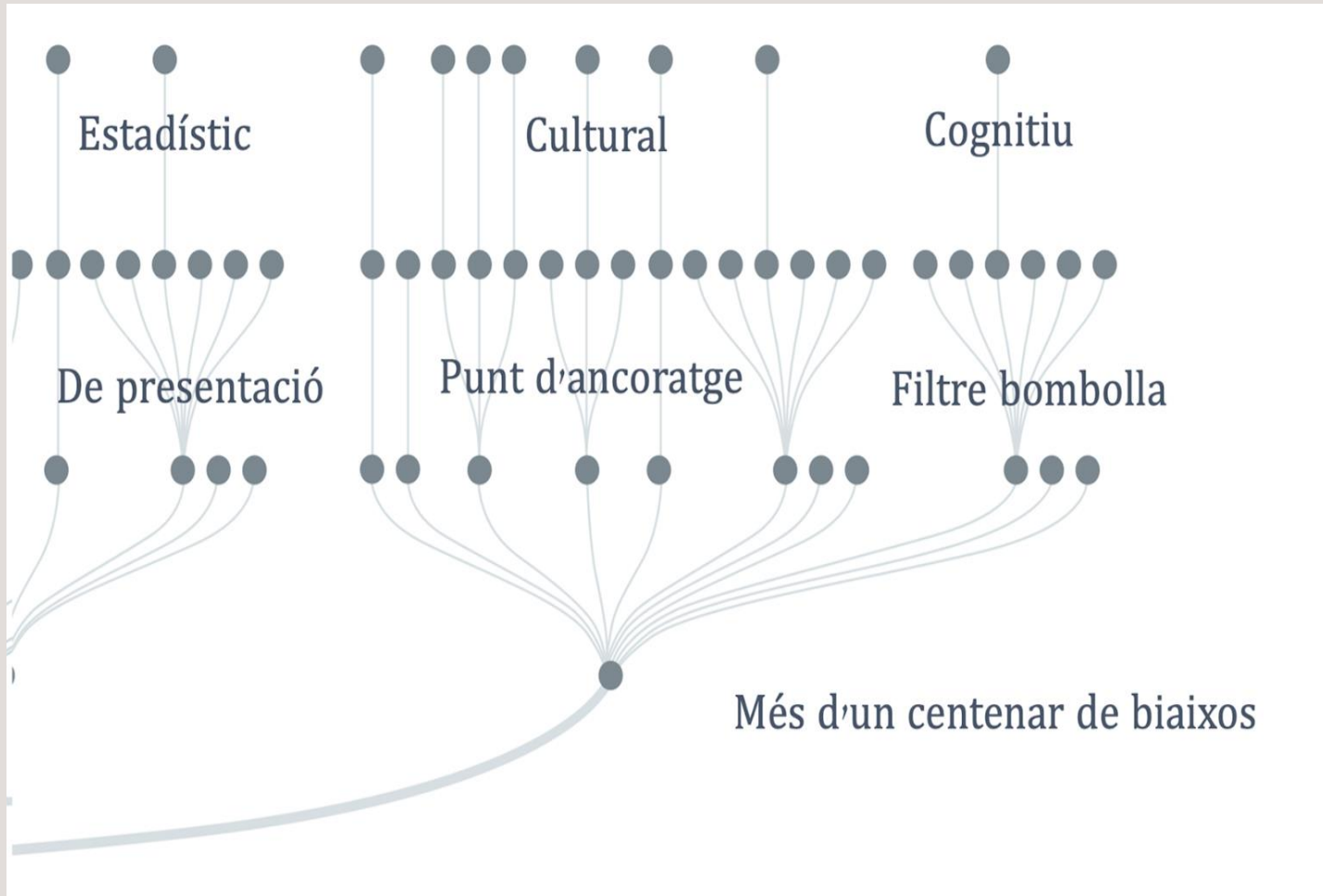
### «¿Es posible acabar con los sesgos de los algoritmos?»

Karma Peiró · 14 de noviembre de 2022 · 2 · algoritmos, big data, datos, Digital Mind, privacidad, Ricardo Baeza-Yates, sesgos por Ricardo Baeza-Yates y Karma Peiró



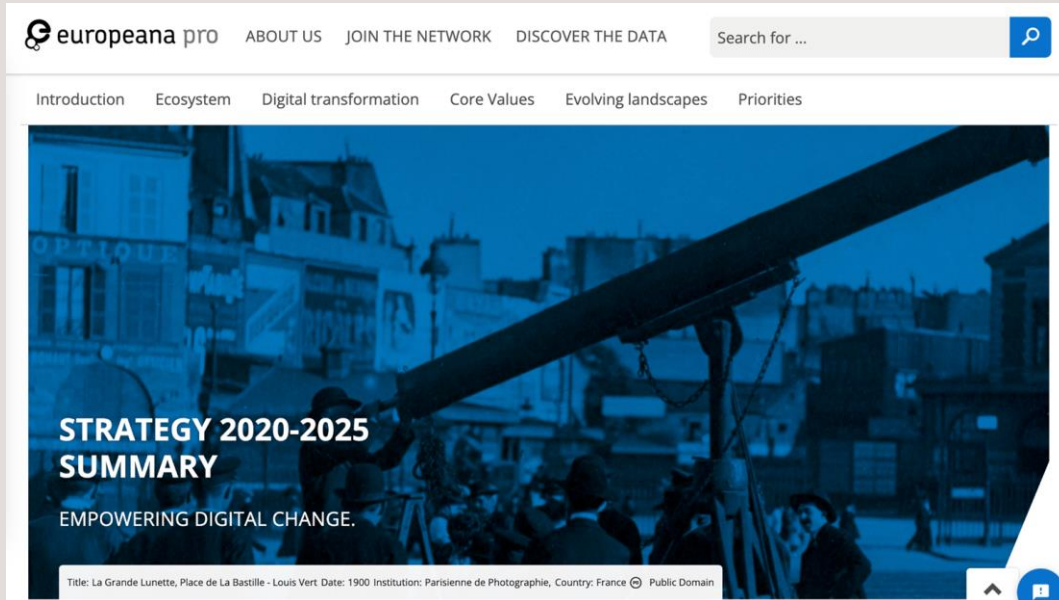
El futuro se está creando a partir de la inteligencia artificial, ya que buena parte de las decisiones que tomamos en nuestras vidas están influenciadas por la acción de algoritmos de aprendizaje automático, que nos ahorran tiempo y facilitan la vida. Sin embargo, los sistemas automatizados tienen sesgos que pueden convertir nuestra cotidianidad en una burbuja de felicidad, mostrándonos sólo lo que nos gusta. Si no hacemos algo para cambiar esta situación, quizás un día descubriremos...

Ricardo Baeza-Yates & Karma Peiró (2022)



1. Què és i què no és la IA?
2. Errors per avançar
3. **Oportunitats: IA + Patrimoni Cultural**
4. Reptes per als propers anys
5. Conclusions

# Oportunitats: IA + Patrimoni Cultural



## Estratègia d'IA per a projectes de Patrimoni cultural (2020-25)

- **Anàlisi de dades massives:** Identificar patrons i tendències que poden indicar riscos i necessitats de protecció del Patrimoni cultural.
- **Detecció de canvis:** Monitorar canvis a l'entorn i en la condició dels béns culturals, alertar sobre possibles amenaces.
- **Simulació i modelatge:** Escenaris futurs i modelar l'impacte de polítiques i estratègies de protecció al patrimoni cultural.
- **Suport a la gestió:** Prendre decisions informades i eficaces, per estalviar recursos.

# ANNA, NOUVELLE GUIDE VIRTUELLE : L'INTELLIGENCE ARTIFICIELLE AU SERVICE DU PATRIMOINE

12 juin 2024

Fort de ses 132 000 visiteurs uniques annuels, notre musée numérique est notre musée le plus visité (devant le musée Saint-Remi : 45 000 visiteurs en 2023) ! Afin de permettre au plus grand nombre de mieux s'approprier ce riche outil, les musées de Reims ont engagé un nouvel agent : Anna, un agent conversationnel autrement appelé chatbot. L'agent conversationnel s'appuie sur des technologies d'intelligence artificielle (IA) développées par la société **DAVI The**



Envie d'un voyage  
dans nos collections ?

En guides virtuels o  
sistemas de  
traducció...

Biaix en l'accessibilitat  
o  
en les col·leccions  
històriques

# Development of a Decision Support System for Improved Resilience & Sustainable Reconstruction of historic areas to cope with Climate Change & Extreme Events based on Novel Sensors and Modelling Tools

[Ficha informativa](#)[Resultados resumidos](#)[Informe](#)[Resultados](#)

## El uso de la IA en la conservación del patrimonio cultural europeo

Con ayuda de la tecnología, investigadores financiados por la UE mitigan los riesgos que amenazan a monumentos antiguos y yacimientos culturales.



© saiko3p/stock.adobe.com

## Projecte HYPERION, finançat per la UE (2019- 23)

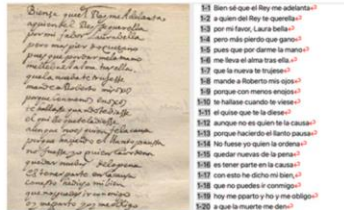
La IA ajuda els investigadors a detectar i analitzar amenaces climàtiques, terratrèmols, erupcions volcàniques, deteriorament pel canvi climàtic per a monuments i enclavaments històrics.

# Artificial intelligence helps discover an unknown work of Lope de Vega

01 FEB 2023

The *Anuario Lope de Vega. Texto, literatura, cultura* edited by the UAB's CER PROLOPE research group, has now published the discovery of a new comedy by Lope de Vega, entitled *La francesa Laura*, discovered in the manuscript collections of the Biblioteca Nacional de España by researchers from the universities of Valladolid and Vienna. The use of computer tools has been essential for its detection, the results of which have been fully supported by philological analysis. The researchers and PROLOPE will soon publish the critical edition of the work together with Editorial Gredos.

In the past few years, artificial intelligence (AI) has become a highly valued tool in several research tasks and projects. The more rigorous historical-philological research has also benefited from this new type of resource. And proof of that is the case of the discovery of *La francesa Laura*, an unknown play



## Unlock the past with Transkribus

Transkribus enables you to automatically recognise text easily, edit seamlessly, collaborate effortlessly, and even train your custom AI for digitizing and interpreting historical documents of any form.

See features

Try for free

100 credits for free. Every month.





# Automated Modelling of Masonry Wall Facades Using Digital Documentation

*Research Undertaken by Post-doctoral Researcher Dr Enrique Valero, Dr Frédéric Bosché and Dr Alan Forster*

Historic Environment  
Scotland

Atenció al **biaix** de les dades  
pel **lloc** de procedència



Linlithgow Palace courtyard west wall, 3D-coloured point cloud.

June 2020



## HERITAGE QUEST: volunteers search for prehistoric structures



Universiteit  
Leiden



Nationaal Park  
Utrechtse Heuvelrug



PROVINCIE UTRECHT



Landschap+  
Erfgoed Utrecht

INNOVATION

# How Artificial Intelligence Completed Beethoven's Unfinished Tenth Symphony

On October 9, the work will be performed in Bonn, Germany, and a recording will be released

Ahmed Elgammal, *The Conversation*

September 24, 2021

---



## MOST POPULAR

-  See 11 of the Best Wildlife Photographs From Years Past
-  Archaeologists Discover Intricately Decorated Coffins Belonging to the Only Daughter of an Ancient Egyptian Governor
-  Montana Rancher Who Created Giant, Hybrid Sheep Sentenced to Six Months in Prison
-  You Could See Dazzling Auroras Tonight as 'Strong' Solar Storm Hits. Here's What to Know



WATCH TO LEARN MORE

# A COMPUTER LEARNS TO PAINT LIKE REMBRANDT



0:00

## How will AI coexist with human music? The future is now

Article by Karma Peiró, a result of the project "Challenges and Opportunities in Music Tech"

24.04.2023



**T**he processes of creating, disseminating, learning, and listening to music are being completely

La IA com a **complement** per a la creació cultural

Incerteses:

- **Drets d'autor** quan l'autoria és d'una tecnologia
- L'acceptació dels **públics**, les audiències per l'art procedent d'uns programaris

NIGHTSHADE

Glaze and Nightshade Q & A Download User's Guide Email Papers/Media

### What Is Nightshade?

Why Does it Work, and Limitations

Since their arrival, generative AI models and their trainers have demonstrated their ability to download any online content for model training. For content owners and creators, few tools can prevent their content from being fed into a generative AI model against their will. Opt-out lists have been disregarded by model trainers in the past, and can be easily ignored with zero consequences. They are unverifiable and unenforceable, and those who violate opt-out lists and do-not-scrape directives can not be identified with high confidence.

In an effort to address this power asymmetry, we have designed and implemented Nightshade, a tool that turns any image into a data sample that is unsuitable for model training. More precisely, Nightshade transforms images into "poison"

The Nightshade Team

@TheGlazeProject  
@theglazeproject  
Email Us

# Project AI4Culture Platform

“An online building hub for the application of **artificial intelligence** technologies in the cultural heritage sector” (2024)

- *Reconeixement de **text multilingüe** en documents escanejats;*
- *Generar i validar **subtítols multilingües**;*
- *Enriquiment amb informació extreta d'**imatges i enllaços semàntics**;*
- ***Traducció automàtica** de metadades del patrimoni cultural.*



1. Què és i què no és la IA?
2. Errors per avançar
3. Oportunitats: IA +Patrimoni Cultural
4. **Reptes per als propers anys**
5. Conclusions

# CONTEXT

- [Artificial Intelligence Strategy of Catalonia](#) and the [Charter for Digital Rights and Responsibilities](#) (Catalan, Government, 2020)
- [Algorithm and data strategy for the ethical impulse of artificial intelligence](#) (Barcelona City Council, 2021)
- [Artificial Intelligence Strategy of Spain](#) (2020)
- [Charter for Digital Rights](#) (Spanish Government, 2021)
- [AI ACT \(2024\)](#)

## A European approach to artificial intelligence

The EU's approach to artificial intelligence centres on excellence and trust, aiming to boost research and industrial capacity and ensure fundamental rights.

The European approach to artificial intelligence (AI) will help build a resilient Europe for the Digital Decade where people and businesses can enjoy the benefits of AI. It focuses on 2 areas: excellence in AI and trustworthy AI. The European approach to AI will ensure that any AI improvements are based on rules that safeguard the functioning of markets and the public sector, and people's safety and fundamental rights.







[Societat digital](#) > [Drets i responsabilitats digitals](#) > [Carta catalana per als drets i les responsabilitats digitals](#)

# Carta catalana per als drets i les responsabilitats digitals

Com seguim defensant els drets humans i les llibertats fonamentals en les noves circumstàncies de l'era digital? La societat catalana contribueix a aquest debat global amb la **Carta catalana per als drets i les responsabilitats digitals**.

El món digital no és un espai a part, aliè al món que habitem; és una dimensió que aporta elements nous i determinants a les nostres vides. La **Carta catalana per als drets i les responsabilitats digitals** recull drets humans acceptats fins ara que requereixen una implementació actualitzada amb la seva versió digital —drets *digitalitzats*. Així mateix, s'hi té en compte una perspectiva d'ampliació de llibertats fonamentals per contribuir a la construcció de drets emergents.

El punt 17 dels [Objectius de Desenvolupament Sostenible \(ODS\) de les Nacions Unides](#) defensa establir «aliances per als objectius mundials: enfortir els mitjans per implementar i revitalitzar les associacions mundials per a un desenvolupament sostenible». La **Carta catalana per als drets i les responsabilitats digitals** neix amb l'objectiu de contribuir al

**Un projecte coordinat pel Govern que té per objectiu promoure un marc legislatiu i democràtic per garantir els drets humans en l'era digital**



Open Access

Review

# Artificial Intelligence for Digital Heritage Innovation: Setting up a R&D Agenda for Europe

by Sander Münster<sup>1,\*</sup>  , Ferdinand Maiwald<sup>2</sup>  , Isabella di Lenardo<sup>3</sup> , Juha Henriksson<sup>4</sup>  , Antoine Isaac<sup>5</sup>  ,  
Manuela Milica Graf<sup>1,6</sup>  , Clemens Beck<sup>1</sup>   and Johan Oomen<sup>7</sup> 

<sup>1</sup> Junior Professorship for Digital Humanities (Object/Image), Friedrich Schiller University Jena, 07737 Jena, Germany

<sup>2</sup> Institute of Photogrammetry and Remote Sensing, Technische Universität Dresden, 01062 Dresden, Germany

<sup>3</sup> Digital Humanities Institute, École Polytechnique Fédérale de Lausanne (EPFL-DHI), CH-1015 Lausanne, Switzerland

<sup>4</sup> Musical Archive Finland, Sörnäisten rantatie 25, 00500 Helsinki, Finland

<sup>5</sup> Europeana Foundation, 2595 BE The Hague, The Netherlands

<sup>6</sup> [manuela.milica.graf@uni-jena.de](mailto:manuela.milica.graf@uni-jena.de)

## Abstract

Artificial intelligence (AI) is a game changer in many fields, including cultural heritage. It supports the planning and preservation of heritage sites and cities, enables the creation of virtual experiences to enrich cultural tourism and engagement, supports research, and increases access and understanding of heritage objects. Despite some impressive examples, the full potential of AI for economic, social, and cultural change is not yet fully visible. Against this background, this article aims to (a) highlight the scope of AI in the field of cultural heritage and innovation, (b) highlight the state of the art of AI technologies for cultural heritage, (c) highlight challenges and opportunities, and (d) outline an agenda for AI, cultural heritage, and innovation.

**Keywords:** cultural heritage; AI; agenda; overview



February 2022



Opportunities and Challenges  
of **Artificial Intelligence** Technologies  
for the **Cultural and Creative Sectors**



*“Ten cultural and creative sectors were analysed including architecture, book publishing, fashion, film, museums, music, news media, performing arts, visual art and video games” (2022)*

Necessitat que la IA aporti una oferta per a reflectir la diversitat cultural europea.

# Reptes IA + Patrimoni Cultural

- **Pèrdua d'autenticitat/amençar diversitat:** L'ús excessiu de l'IA podria portar a una homogeneïtzació de la creació cultural i reduir l'originalitat.
- **Biaixos:** Afectar la interpretació del patrimoni cultural.
- **Qüestions ètiques i legals:** Privacitat de dades i els drets de propietat intel·lectual.
- **Falta de formació:** Entre professionals del sector sobre coneixements d'IA, podria dificultar aplicar-la adequadament.
- **Impacte llocs treball:** L'automatització podria afectar tasques.
- **Preservació de la diversitat cultural:** La IA podria amenaçar la diversitat, si no es gestiona adequadament.

1. Què és i què no és la IA?
2. Errors per avançar
3. Oportunitats IA +Patrimoni Cultural
4. Reptes per als propers anys

## 5. **C**onclusions



## ARTIFICIAL INTELLIGENCE

Automated Decision-making in Catalonia



apdcat

Generalitat de Catalunya

Professora del Departament de Medicina de la Universitat de Barcelona (UB).  
Subdirectora de l'Observatori de Bioètica i Dret de la UB

### ITZIAR DE LECUONA:<sup>94</sup> “Cal moderar l’entusiasme tecnològic que vivim”

L'experta en els aspectes ètics i legals de la biomedicina alerta que, en l'àmbit de la recerca biomèdica, bona part dels projectes de recerca que s'avaluen estan basats en la convergència de tecnologies, entre les quals es troben la intel·ligència artificial i l'aplicació de l'anàlisi de dades massives.

“La intel·ligència artificial ja és aquí, i tots estem contribuint amb els nostres conjunts de dades personals al seu desenvolupament”, explica. “Avui és possible explotar dades a escala massiva, per millorar la presa de decisions mitjançant el desenvolupament d'algorismes, que permetin –correlacionant les dades– permetin determinar patrons de comportament i predir conductes. I això ho volen fer tots: des de la iniciativa privada fins al sistema públic de salut. Per tenir sistemes de salut més eficients o perquè la ciutadania es beneficiï d'una medicina personalitzada. No fer-ho seria gravíssim des del punt de vista ètic”, afegeix De Lecuona.

### El paper del ciutadà

Portem la intel·ligència artificial integrada en els aparells, com és el cas dels mòbils. Emetem dades contínuament des dels diferents dispositius digitals que utilitzem. També, hi ha sistemes més complexos en els quals s'exploten les nostres dades personals amb diversos objectius.



Cap de Tecnologia de NCENT. Catedràtic d'Informàtica de la Universitat Pompeu Fabra i de la Northeastern University

## RICARDO BAEZA-YATES:<sup>107</sup> “La revolució tecnològica actual necessita ètica”

### El Pep Consciències dels prejudicis

Si fem servir els algorismes sabent que tenen biaixos; per què deixem, doncs, que prenguin decisions? I Ricardo Baeza-Yates respon: “Per què tenim humans que prenen decisions si també s’equivoquen? Tots tenim prejudicis. Els sistemes automatitzats són de gran ajuda en situacions en què els biaixos no influeixen gaire, com, per exemple, en el control aeri: tenir persones moltes hores, en tensió, és més perillós que entrenar màquines per a aquesta tasca. No es cansen, estan programades, són més eficients.”

“De biaixos, n’hi ha un munt, el problema és que no ens adonem que existeixen fins que cometem un error. El mateix passa amb els nostres prejudicis, que no en som conscients la major part de les vegades. Les màquines poden ajudar-nos a adonar-nos-en i a crear

això passi, la gent no se’n adona que hi ha biaixos en les dades que s’introdueixen en els algorismes”, explica el catedràtic d’Informàtica de la UPF. Baeza-Yates diu que la tecnologia sempre va més ràpida que la part social, i que només ens preocupem de l’ètica quan sorgeixen problemes, com ara mateix.

### ARTIFICIAL INTELLIGENCE

Automated Decision-making in Catalonia



adocat

Generalitat de Catalunya



### ARTIFICIAL INTELLIGENCE

Automated Decision-making in Catalonia



gub.cat

Generalitat de Catalunya

Investigadora del grup d'Enginyeria del Coneixement i Aprenentatge Automàtic del Centre d'Investigació en Ciència Intel·ligent de Dades i Intel·ligència Artificial de la UPC. Vicedegana de Big Data, Ciència de Dades i Intel·ligència Artificial del Col·legi Oficial d'Enginyeria Informàtica de Catalunya

## KARINA GIBERT:<sup>126</sup> “El cost energètic de la IA és enorme”

“La tecnologia, en principi, és neutra i el que té usos més o menys ètics és el que fem amb ella”, explica la investigadora Karina Gibert. “Els problemes ètics que cal tenir en compte avui són els biaixos, sí, però també l'explicabilitat, que és una qüestió molt més profunda i que ja s'ha comentat en aquest capítol.”

Gibert vol puntualitzar que només hi ha una part de la IA que no és explicable, la que es coneix com a *subsimbòlica*. “La part simbòlica, que imita la forma com els humans resolen els problemes, és totalment explicable. Però és terriblement cara, des del punt de vista computacional, i mai no ha fet un salt a problemes reals d'una certa complexitat perquè col·lapsa.”

“La part subsimbòlica intenta aconseguir que els ordinadors resolguin problemes que requereixen intel·ligència amb més qualitat que els humans, des del punt de vista dels resultats que aporta”, continua explicant la investigadora, “sense importar gaire si la manera com resol el problema imita o no la manera com ho fa l'humà. I aquesta és la IA no explicable, la que, entre altres coses, ha donat lloc al *deep learning*. Com a contrapartida, el *deep learning* ha aconseguit donar resultats respecte a problemes grossos que fins ara no s'havien resolt. Però està clar que, perquè el resultat d'una intel·ligència artificial tingui impacte en un context real, cal argumentar com sigui la decisió que s'ha pres. I això ho sabem els que –com jo– fa trenta anys que treballem en aplicacions reals.”





## ARTIFICIAL INTELLIGENCE

Automated Decision-making in Catalonia



apdcat

Generalitat  
de Catalunya

**Professor investigador del Consell Superior d'Investigacions Científiques (CSIC). Exdirector de l'Institut d'Investigació d'Intel·ligència Artificial (IIIA)**

### **RAMÓN LÓPEZ DE MÁNTARAS:<sup>99</sup> “S’hauria de poder certificar que els algorismes estan nets de biaixos”**

Ramon López de Mántaras va coordinar l'elaboració de l'Estratègia Espanyola per a la Intel·ligència Artificial<sup>100</sup> –feta pública el març del 2019–, un pas importantíssim per definir les normes ètiques que corresponen a la seva implementació. Igualment, va promoure la Declaració de Barcelona,<sup>101</sup> redactada fa dos anys per la comunitat científica i que marca molt bé les directrius ètiques que s'haurien de seguir.

Un dels punts prioritaris que la Declaració de Barcelona menciona és el del retiment de comptes (*accountability*). És a dir, que quan un algorisme prengui decisions, les persones afectades puguin rebre –en termes entenedors– una explicació de per què les ha pres i se'ls permeti qüestionar-les amb arguments raonats. Però dos anys més tard, aquest és un aspecte pendent de resoldre que preocupa molt els investigadors.

“Com a consumidor, ciutadà, tens dret a demanar explicacions,<sup>102</sup> amb la Llei de protecció de dades europea. Pots preguntar per què t'han denegat un ajut, o no t'han donat un crèdit. El codi de l'algorisme no te'l donaran, però s'hauria de poder exigir que una entitat neutral avalués si l'algorisme és just. No existeix una entitat semblant a Catalunya. Algorithm Watch<sup>103</sup> o Ethical Tech Society<sup>104</sup> poden fer aquesta funció en l'àmbit internacional.



## ARTIFICIAL INTELLIGENCE

Automated Decision-making in Catalonia



apdcat

Generalitat de Catalunya

Doctora en informàtica i professora d'investigació a l'Institut de Robòtica i Informàtica Industrial (CSIC-UPC)

### CARME TORRAS:<sup>105</sup> “Cal formar en criteris ètics a qui desenvolupa la tecnologia”

“Hi ha actualment una gran tendència a treballar amb aprenentatge profund (*deep learning*), un tipus de xarxa neuronal amb moltes capes que el que fa és associar entrades amb sortides. És un tipus d'aprenentatge anomenat de *caixa negra*, perquè no se sap amb exactitud què passa en aquest procés d'associació. Si les dades tenen biaixos, l'algorisme aprèn malament. Però qui el posa en marxa, sigui un banc per donar-te un crèdit o un organisme públic per avaluar un currículum o assignar una subvenció, no et podrà donar una explicació de com s'ha arribat a aquella decisió. Hi ha un projecte molt ambiciós de la Unió Europea<sup>106</sup> d'introduir el concepte d'*explicabilitat* (*explainability*) per obtenir explicacions de com ha processat la màquina les dades per arribar a una conclusió determinada. La comunitat informàtica està molt sensibilitzada per trobar una manera general d'argumentar les decisions que pren la màquina. A més, cal expressar els arguments en termes comprensibles per persones no-expertes en informàtica.”

Carme Torras –autora de novel·les de ciència-ficció des de les quals enfoca cap a on evolucionarem els humans i les màquines– reclama dos aspectes: “1. Regulació per assegurar els criteris mínims en l'aplicació de la intel·ligència artificial en la societat, en el moment actual. 2. Educació, començant per formar en criteris ètics als qui desenvolupen la tecnologia.”



## ARTIFICIAL INTELLIGENCE

Automated Decision-making in Catalonia



apdcat

Generalitat de Catalunya

Director científic del grup d'Intel·ligència Artificial d'Alt Rendiment en el Centre de Supercomputació de Barcelona. Catedràtic d'Intel·ligència Artificial de la Universitat Politècnica de Catalunya

## ULISES CORTÉS:<sup>110</sup> “Cal ensenyar els límits de la tecnologia i pensament crític”

“Abans qui decidia donar-te un préstec o fer-te una operació de cor eren els algorismes que estaven al cap del banquer o del metge. Tu no sabies amb quins criteris ho decidien. Ara està passant el mateix, però amb màquines. Seria bo que una empresa pogués assegurar que la decisió és explicable si hi estàs en desacord. Però avui això és impossible, perquè ningú no els està dissenyant perquè siguin explicables.”

“Hi ha una petita porció de la població que té por dels automatismes pels errors comesos fins ara i, en concret, de l'ús de l'aprenentatge automàtic (*machine learning*). Però a la majoria de les persones, els importa ben poc el que es faci amb els algorismes. I els hauria d'importar, perquè les seves vides canvien constantment en funció del que decideixen un munt de màquines. S'han de produir casos molt flagrants, com el de Cambridge Analytica, per començar a prendre'n consciència. No és dolent fer servir Facebook, si se saben les conseqüències. El dolent és deixar a un menor que hi entri sense advertir-li els riscos que comporta”.

“Tinc clar que en una societat occidental, de passat cristià, hi ha un conjunt de valors comuns. Dels quals ja parlaven Plató i Aristòtil. Als EUA és d'utilitat plantejar-se: quin és el mal mínim que puc causar? És molt evident que el més just per a nosaltres no ho és per als de Malawi. Però ens ajuntem i decidim quin és el consens.”

# Conclusió...



La investigadora Carissa Véliz y la portada de su libro 'Privacidad es poder'. | Legada la cabeza

Professora de filosofia i ètica a la Universitat d'Oxford



Carissa Véliz  
@CarissaVeliz

...

Perfect analogy for the age of [#algorithms](#). The dad is [#AI](#). The kids are us. And we are not doing much better than them. [#AIethics](#)

[Tradueix el tuit](#)



Moltes gràcies  
per la vostra atenció!



**Karma Peiró**

+ info: <http://karmapeiro.com>

Correu-e: [hello@karmapeiro.com](mailto:hello@karmapeiro.com)

LinkedIn i Twitter: [@kpeiro](#)