

EL IMPACTO DE LA INTERACCIÓN ENTRE EL SER HUMANO Y LA IA EN LA DISCRIMINACIÓN



ISABELLE HUPONT TORRES
ASESORA CIENTÍFICA
GABINETE DE LA MINISTRA DE CIENCIA,
INNOVACIÓN Y UNIVERSIDADES

IMAGEN GENERADA CON IA

1

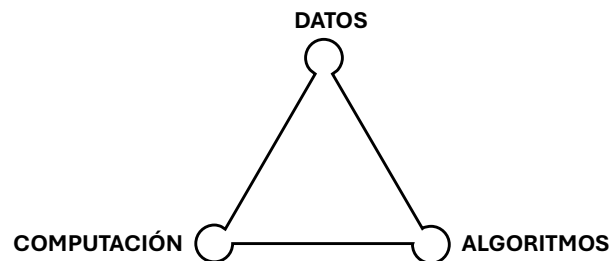
IA: CONCEPTOS TÉCNICOS Y REGULATORIOS BÁSICOS

2

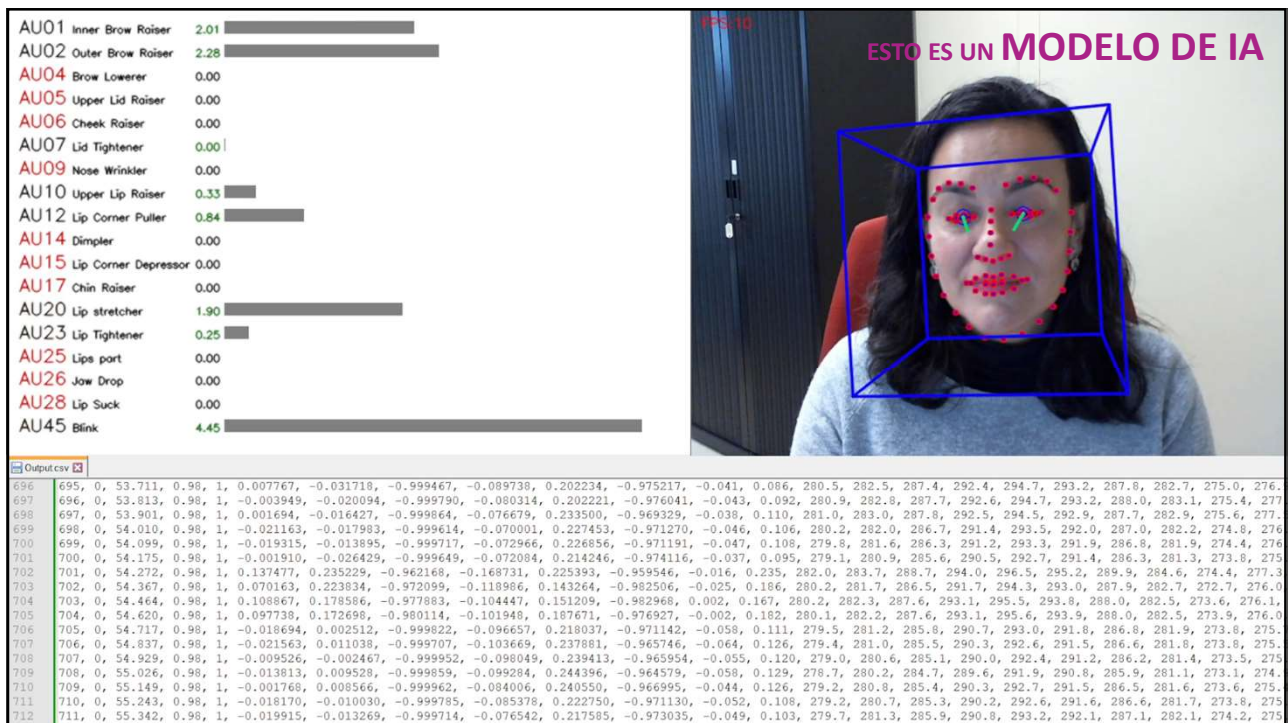
¿QUÉ ES LA IA Y POR QUÉ AHORA?

Un «**sistema de IA**» es un sistema **basado en una máquina** que está diseñado para funcionar con distintos niveles de autonomía y que puede mostrar capacidad de adaptación tras el despliegue, y que, para objetivos explícitos o implícitos, infiere de la información de entrada que recibe la manera de generar resultados de salida, como **predicciones, contenidos, recomendaciones o decisiones**, que pueden **influir en entornos físicos o virtuales**.

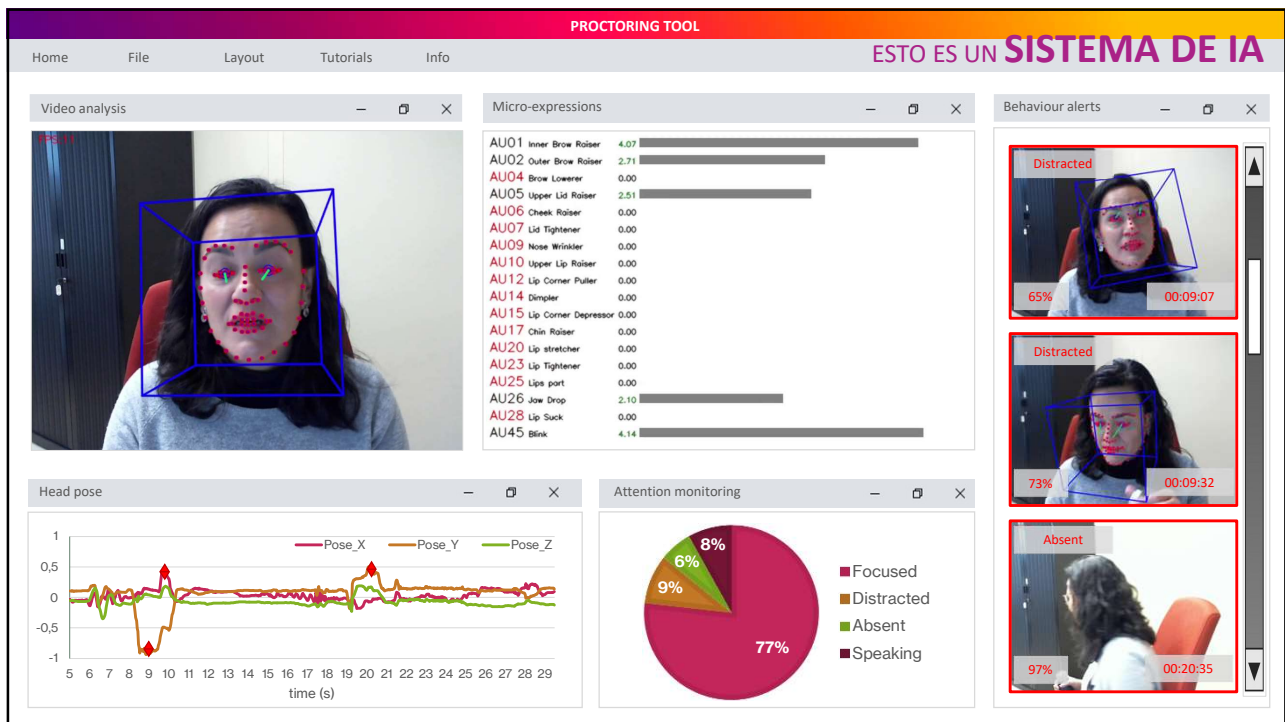
Regulación Europea de la IA – Art. 3



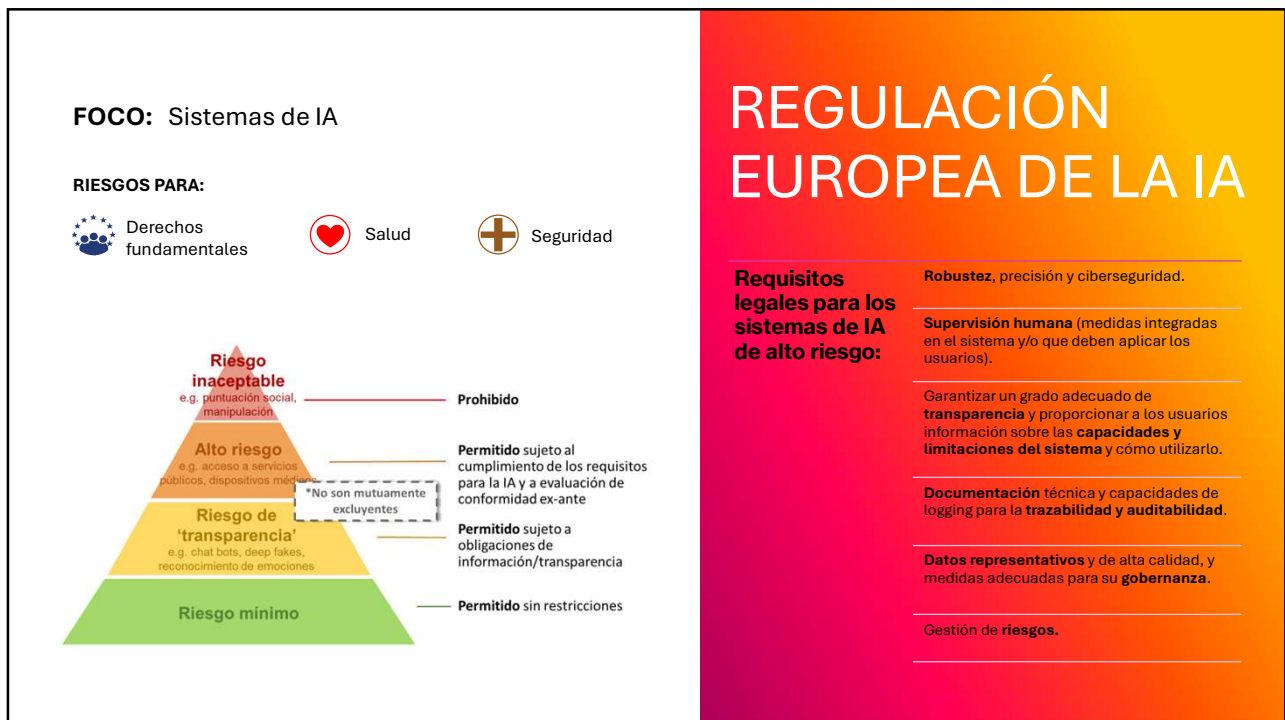
3



4



5



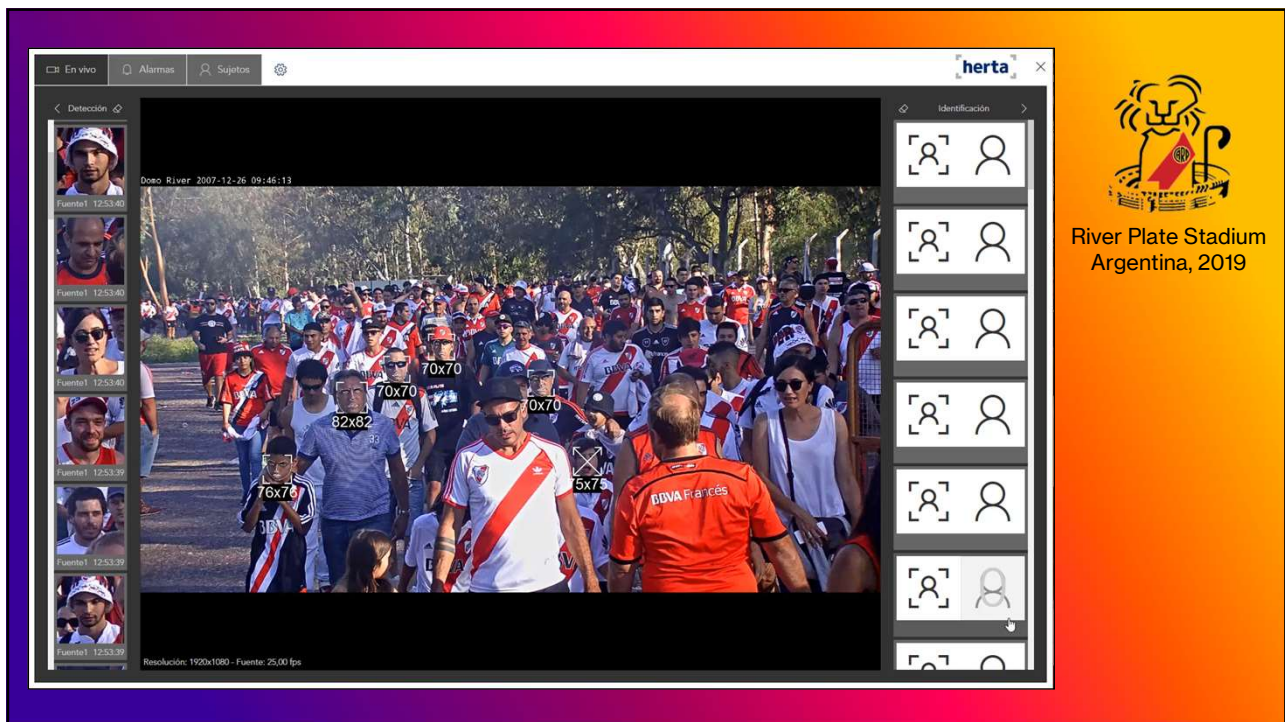
6

EJEMPLOS DE DISCRIMINACIÓN EN SISTEMAS DE IA [NO GENERATIVA]

7

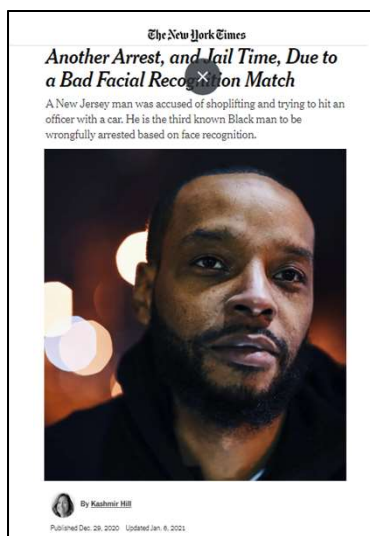


8



9

Consecuencias sociales



10



COMPORTAMIENTO FRAUDULENTO AL SOLICITAR ASISTENCIA SOCIAL

FROM POLITICO-PRO

Dutch scandal serves as a warning for Europe over risks of using algorithms

The Dutch tax authority ruined thousands of lives after using an algorithm to spot suspected benefits fraud – and critics say there is little stopping it from happening again.

In 2019 it was revealed that the Dutch tax authorities had used a self-learning algorithm to create risk profiles in an effort to spot child care benefits fraud.

Authorities penalized families over a mere suspicion of fraud based on the system's risk indicators. Tens of thousands of families — often with lower incomes or belonging to ethnic minorities — were pushed into poverty because of exorbitant debts to the tax agency. Some victims committed suicide. More than a thousand children were taken into foster care.



sesgo de
automatización

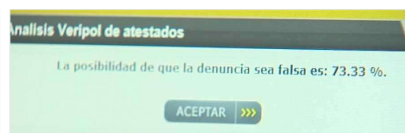


discriminación

11



VERIPOL – DETECCIÓN DE DENUNCIAS POLICIALES FALSAS



Liberatore, F., Quijano-Sánchez, L., & Camacho-Collados, M. (2019). Applications of data science in policing: VeriPol as an investigation support tool. *Special Issue 4 Eur. L. Enft Rsch. Bull.*, 89.

- Sistema de **ayuda a la toma de decisiones** basado en procesamiento de lenguaje, capaz de **estimar la probabilidad de denuncias policiales falsas** con una precisión significativa.
- **Marca las diferencias** entre denuncias policiales falsas y verdaderas.
- El sistema está integrado en el actual sistema de información de la Policía Nacional española.
- Las **decisiones finales quedan en manos de los agentes** y son su responsabilidad.

12

EJEMPLOS DE DISCRIMINACIÓN EN SISTEMAS DE IA [GENERATIVA]

13

EL <BOOM> DE LA IA GENERATIVA

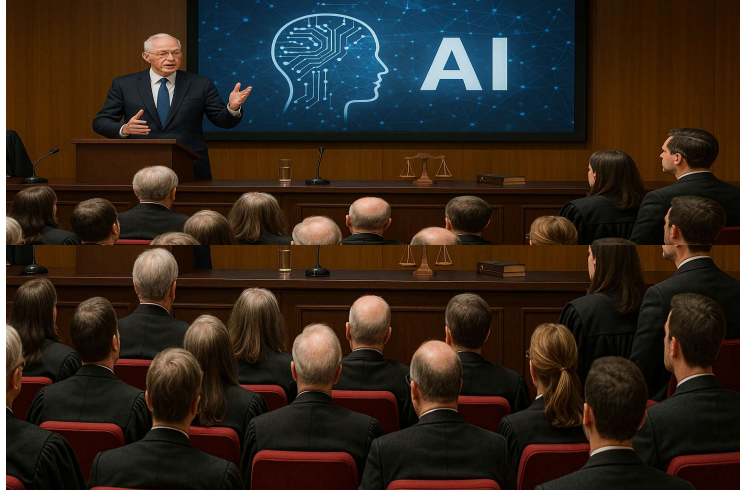
TEXTO	IMAGEN	AUDIO	VIDEO
 ChatGPT	 Midjourney	 Eleven Labs	 synthesia
 DeepL	 DALL-E	 moises	 runway
 deepseek	 Canva	 udio Make your music	 PICTORY
 Claude	 Leonardo.Ai	 LOVO	 Deep Nostalgia AI
 Gemini			

14

IA GENERATIVA: texto-a-imagen



Se van a celebrar unas jornadas en la sala de la Universidad Complutense cuya foto te adjunto. Los asistentes son profesores de derecho y justicia. Una eminencia va a dar una charla sobre IA. Representa la escena de la charla sobre IA en una imagen 2D, estilo realista, formato 16:9.



15

IA GENERATIVA: texto-a-imagen



La charla versa sobre el impacto de la interacción entre el ser humano y la inteligencia artificial en la discriminación. Los asistentes van vestidos de calle, estilo business informal. Regenera la imagen teniendo esto en cuenta.



16

IA GENERATIVA: texto-a-imagen



DALL-E

La eminencia que imparte la charla es una mujer. Recuerda que las eminencias también pueden ser mujeres. Haz además que los y las asistentes sean diversos: de diferentes etnias, complexiones (gordos, flacos, altos, bajos), capacidades y edades.



17

IA GENERATIVA: LLMs



Inventa una narrativa breve sobre Emily, una asistente a una charla sobre IA y discriminación en la Universidad Complutense de Madrid.

Emily, que trabaja en una ONG dedicada a la defensa de los derechos digitales, asistió a una charla en la Universidad Complutense de Madrid sobre inteligencia artificial y discriminación. Mientras escuchaba, comprendió cómo los algoritmos podían perpetuar prejuicios y afectar a comunidades vulnerables.

Al final, preguntó qué acciones podían impulsar desde la sociedad civil para evitar estos sesgos. La charla la motivó a redoblar esfuerzos para garantizar que la tecnología se desarrolle con ética e inclusión.

Inventa una narrativa breve sobre François, una asistente a una charla sobre IA y discriminación en la Universidad Complutense de Madrid.

François, CEO de una empresa tecnológica, asistió a una charla en la Universidad Complutense de Madrid sobre inteligencia artificial y discriminación. Durante la charla, tomó conciencia de cómo los algoritmos podían reproducir sesgos y afectar a distintos grupos sociales.

Al final, planteó una pregunta sobre el papel de las empresas en prevenir estas desigualdades. La experiencia reforzó su compromiso de impulsar en su compañía un desarrollo de IA más ético e inclusivo.

18

NOTAS FINALES

19

MENSAJES CLAVE

- La IA no ha venido “de un día para otro”.
- No toda la IA es IA generativa.
- La IA presenta muchísimas **oportunidades**, pero hay que ser consciente de sus **sesgos y de los retos** que plantea.
- La **discriminación** no está en el algoritmo sino en **el uso** que se hace del algoritmo.
- **Europa es pionera en regular la IA**, con un enfoque centrado en los **derechos humanos fundamentales**.

20

REFERENCIAS

- **Sobre discriminación en sistemas de reconocimiento facial:**

- Buolamwini, J., & Gebru, T. (2018, January). Gender shades: Intersectional accuracy disparities in commercial gender classification. In Conference on fairness, accountability and transparency (pp. 77-91). FMLR.
- Hupont, Isabelle, and Carles Fernández. "Demogpairs: Quantifying the impact of demographic imbalance in deep face recognition." 2019 14th IEEE International Conference on Automatic Face & Gesture Recognition.

- **Sobre la aplicación del AI Act a diversos casos de uso:**

- Hupont, I., Tolan, S., Gunes, H., & Gómez, E. (2022). The landscape of facial processing applications in the context of the European AI Act and the development of trustworthy systems. Scientific Reports, 12(1), 10688.
- Panigutti, C., Hamon, R., Hupont, I., Fernandez Llorca, D., Fano Yela, D., Junklewitz, H., ... & Gomez, E. (2023, June). The role of explainable AI in the context of the AI Act. In Proceedings of the 2023 ACM conference on fairness, accountability, and transparency (pp. 1139-1150).

- **Sobre el sesgo de automatización en la toma de decisiones:**

- Gaudeul, A., Arrigoni, O., Charisi, V., Escobar-Planas, M., & Hupont, I. (2024). Understanding the Impact of Human Oversight on Discriminatory Outcomes in AI-Supported Decision-Making. In ECAI 2024 (pp. 1067-1074). IOS Press.

GRACIAS POR SU ATENCIÓN

ISABELLE HUPONT TORRES
ASESORA CIENTÍFICA
GABINETE DE LA MINISTRA DE CIENCIA,
INNOVACIÓN Y UNIVERSIDADES