

Actitudes humanas hacia la delegación de decisiones a sistemas basados en inteligencia artificial (IA)

Inés Sallent Ysamat



Universitat
Pompeu Fabra
Barcelona

Actitudes humanas hacia la delegación de decisiones a sistemas basados en inteligencia artificial (IA)

TRABAJO DE FIN DE GRADO DE

Inés Sallent Ysamat

Director: Vladimir Estivill Castro

Grau en Enginyeria en Sistemes Audiovisuals

Curs 2024-2025



Universitat
Pompeu Fabra
Barcelona

Escola
d'Enginyeria

[Tingueu en compte:

(i) No s'han de deixar pàgines blanques al cos del treball (índex fins la conclusió).

(ii) Utilitzar interlineat senzill i tipus de lletra Times New Roman.

(iii) Per a aquesta secció principal (cos del treball), la restricció del límit de pàgines és de 30 pàgines + 5% o 10% de marge.

(iv) La pàgina del final de capítol compta com a mitja pàgina.

(v) Podeu afegir un annex d'informació de suport a l'apèndix per incloure totes les dades/informació que necessiteu mostrar. El apèndix no té un límit de pàgines. Si tens informació a l'apèndix, has de fer referència en el text principal a la informació continguda en l'apèndix seguint aquest exemple:

document principal: "La taxa de natalitat als països europeus està disminuint molt ràpidament (vegeu Apèndix-I). Això vol dir que tot i que la gent viu més temps (figura Apèndix-I), amb el temps la mida de la població disminuirà".

(vi) Les referències s'inclouen a la secció de Bibliografia i s'exclouen de la limitació del nombre de pàgines.]

Dedicatòria [opcional] [mida 12]
Lorem ipsum

Agraïments[mida 14]

Text dels agraïments [mida 11]

Summary

As technology continues to advance, the use of Artificial Intelligence (AI) is growing steadily. However, the measures to prevent the potential risks associated with its use have not always kept pace.

This project aims to examine how people perceive the implementation of AI systems in automated processes. The hypothesis proposed is that users tend to be more receptive to AI when it is applied to automatic or repetitive tasks, such as those involved in personnel selection processes.

To address this hypothesis, I conducted a review of the existing literature in order to deepen the understanding of AI, its definition, evolution, applications in automated processes, and the biases that may arise. Additionally, I draw attention to the cognitive biases that affect human decision-making.

Based on this research, I designed a questionnaire from scratch with the aim of assessing users' knowledge of and attitudes toward AI.

Resumen

Resum

Tabla de Contenido

- Introducción..... 1
 - 1.1 Motivación 1
 - 1.2 Objetivos 1
 - 1.3 Logros 2
 - 1.4 Organización de la memoria 2
- Revisión Literaria 3
 - 2.1 Definición..... 3
 - 2.2 Evolución 4
 - 2.3 Categorías éticas 7
 - 2.3.1 Principales teorías éticas 7
 - 2.3.2 Perfiles éticos dentro de la sociedad 8
 - 2.4 Ética en la IA..... 9
 - 2.4.1 Alineamiento de valores en la IA (IA *Value Alignment*)..... 10
 - 2.5 Confianza en la IA 11
- Metodología..... 13
 - 3.1 Tecnologías utilizadas..... 13
 - 3.1.1 Lenguaje: Python 13
 - 3.1.2 Framework: Streamlit 13
 - 3.1.3 Base de datos: Supabase..... 14
 - 3.1.4 Análisis de los datos 15
 - 3.2 Estructura del aplicativo..... 15
- Diseño Experimental 17
 - 4.1 Objetivos e hipótesis 17
 - 4.2 Participantes 17
 - 4.3 Método de reclutamiento 17
 - 4.4 Recolección de datos..... 18
 - 4.5 Justificación del diseño del cuestionario 18
- Resultados y discusión 19
 - 5.1 Características sociodemográficas 19
 - 5.2 Influencia del perfil del participante en la percepción sobre la IA 19
 - 5.3 Percepción social y delegación de decisiones en la inteligencia artificial 21
 - 5.4 Análisis comparativo sobre la percepción de sesgos en IA y humanos 23
 - Valoración de los principios éticos generales..... 26
 - 5.5 Percepción social sobre el uso y límites de las IAs con capacidad emocional 27
- Análisis e Interpretación de los resultados 30
 - 6.1 Nivel de confianza y percepción de sesgos según el uso de la IA 30
 - 6.2 Evaluación de los resultados desde un enfoque ético 31

Análisis de los resultados desde los perfiles éticos.....	32
Conclusiones y trabajos futuros	33

Llista de figures

Figura 1: Línea del tiempo de la evolución de la IA.....6

Figura 2: Adaptación de los valores reconocidos por la UNESCO.....10

Figura 3: Consideraciones técnicas y organizativas en el contexto de la alineación de valores en la IA..11

Figura 4: Adaptación de los niveles de confianza y aceptación de la IA según grupos económicos...12

Figura 5: Código de las diferentes bases de datos14

Figura 6: Estructura proyecto16

Figura 7: Datos demográficos.....19

Figura 8: Nivel de confianza en los sistemas de IA20

Figura 9: Relación nivel de confianza en la IA y el género y la edad.....20

Figura 10: Frecuencia uso de la IA relacionado con la edad.....21

Figura 11: Nivel de confianza en los sistemas de IA en función del conocimiento.....21

Figura 12: ¿Qué tipo de decisiones considerarías aceptable delegar a una IA?.....22

Figura 13: Nivel de acuerdo con la afirmación sobre los beneficios actuales de la IA22

Figura 14: Nivel de acuerdo con el impacto futuro de la IA en diferentes ámbitos de la vida.....23

Figura 15: Resultados sesgos en la IA en la toma de decisiones23

Figura 16: Resultados sesgos en las personas en la toma de decisiones24

Figura 17: Nivel de confianza según en la IA según la atribución de sesgos25

Figura 18: Nivel de uso de la IA según la atribución de sesgos.....25

Figura 19: Resultados sesgos en las personas en la toma de decisiones.....26

Figura 20: En qué etapa se debería concienciar y educar sobre el uso y los riesgos de la IA.....26

Figura 21: Nivel de acuerdo de los participantes con los tres principios éticos que debería respetar la IA27

Figura 22: Respuestas caso 1.....27

Figura 23: Respuestas caso 228

Figura 24: Respuestas caso 3.....28

Figura 25: Opinión sobre el acceso de los usuarios a sistemas basados en IAs con conversaciones emocionales.....29

Figura 26: Percepción sobre la responsabilidad de la IA en situaciones de crisis.....29

Capítol 1

Introducción

En las últimas décadas, los avances tecnológicos han impulsado de manera significativa el desarrollo de la inteligencia artificial (IA). Esta evolución ha permitido su integración en múltiples ámbitos de la sociedad, contribuyendo al progreso en sectores tan diversos como la medicina, la comunicación y la automatización de tareas. Sin embargo, esta evolución tecnológica también ha generado preocupaciones éticas. Las decisiones automatizadas pueden estar influenciadas por sesgos presentes en los datos o en el diseño de los algoritmos. Estos sesgos pueden reforzar desigualdades existentes y afectar negativamente a ciertos colectivos, especialmente a los más vulnerables ^[1].

Actualmente, se están desarrollando sistemas que asumen tareas que antes eran exclusivamente humanas, lo cual ha intensificado el debate sobre cuestiones como la transparencia algorítmica, la equidad en los resultados y la atribución de responsabilidades en la toma de decisiones automatizadas ^[2]. Ante este contexto, resulta relevante preguntarse hasta qué punto la ciudadanía es consciente de los sesgos que pueden estar presentes en la inteligencia artificial y cuanta confianza tienen hacia ella.

1.1 Motivación

Aunque la inteligencia artificial se ha integrado progresivamente en numerosos ámbitos de nuestra vida, su funcionamiento y consecuencias siguen siendo un terreno desconocido para gran parte de la ciudadanía. La automatización de decisiones, cada vez más común en contextos sociales, laborales o administrativos, se produce en muchos casos sin que las personas sean plenamente conscientes de ello, ni de los criterios que subyacen a esas decisiones. Esta falta de conocimiento o conciencia puede convertirse en un riesgo, especialmente si se considera que la IA ya está tomando decisiones que afectan directamente a las personas. Además, la ausencia de estrategias educativas o preventivas dificulta una adopción informada y crítica de estas tecnologías.

La motivación de este trabajo nace de esta preocupación: ¿es la ciudadanía realmente consciente del impacto de la IA y de los sesgos que puede contener? Y si no lo es, ¿deberíamos impulsar recursos, leyes o medidas educativas que fomenten la comprensión y el pensamiento crítico sobre estas tecnologías? Este trabajo busca contribuir a esta reflexión, con el objetivo de promover un uso más responsable, justo y transparente de la inteligencia artificial en la sociedad.

1.2 Objetivos

Los objetivos de este trabajo son los siguientes:

- Investigar y analizar la evolución de la inteligencia artificial y sus avances más relevantes, con el fin de obtener un conocimiento más profundo sobre su funcionamiento y aplicaciones.
- Definir los distintos perfiles éticos existentes en la sociedad, con el fin de establecer una posible relación con los principios que podrían regir el comportamiento de los sistemas de inteligencia artificial.
- Evaluar cómo la sociedad utiliza y percibe la inteligencia artificial mediante un cuestionario que permita conocer el nivel de información, confianza que genera y las posibles inquietudes que despierta.
- Analizar los resultados obtenidos en el cuestionario para extraer conclusiones sobre el grado de conciencia de la ciudadanía sobre los sesgos proporcionados por la IA.

1.3 Logros

En este trabajo se han cumplido los siguientes objetivos propuestos:

- Se ha realizado una revisión teórica sobre la evolución de la inteligencia artificial, sus aplicaciones actuales y los principales debates éticos que suscita, lo cual ha permitido contextualizar el cuestionario y sus resultados.
- Se han identificado distintos perfiles de percepción en función del uso, el conocimiento declarado y la sensibilidad ética de los participantes, lo que ha permitido explorar cómo estos factores influyen en la confianza depositada en la IA.
- Se ha evidenciado una contradicción interesante: los usuarios que más utilizan la inteligencia artificial son, en muchos casos, quienes mayor confianza depositan en ella, a pesar de no tener un conocimiento técnico profundo. Este hallazgo contradice parcialmente la hipótesis inicial y abre nuevas líneas de investigación.
- Se ha confirmado que existe una conciencia crítica creciente respecto a los sesgos que puede presentar la IA, especialmente aquellos derivados de datos de entrenamiento que reproducen desigualdades sociales.
- Se ha detectado una opinión mayoritaria sobre la necesidad de incorporar formación sobre IA desde etapas educativas tempranas, lo cual refuerza la importancia de la alfabetización digital y ética como herramienta de prevención.

1.4 Organización de la memoria

El contenido de la memoria se divide en los siguientes capítulos:

- **Capítulo 2:** se introduce el concepto de inteligencia artificial y se realiza una revisión bibliográfica centrada en su evolución, en los aspectos éticos relacionados con su uso y en la confianza que genera en la sociedad.
- **Capítulo 3:** se explican las metodologías utilizadas en el desarrollo del proyecto.
- **Capítulo 4:** se detalla el diseño experimental del cuestionario, incluyendo sus objetivos y el proceso de elaboración.
- **Capítulo 5:** se presentan los resultados obtenidos del cuestionario y se realiza una discusión inicial sobre los mismos.
- **Capítulo 6:** se analizan e interpretan los resultados en relación con el marco teórico desarrollado previamente.
- **Conclusiones y trabajos futuros:** se exponen las principales conclusiones del estudio y se plantean posibles líneas de trabajo futuras.

Capítulo 2

Revisión Literaria

Con el fin de dar respuesta a los objetivos de este trabajo, ya formulados en la introducción, se ha realizado una revisión literaria sobre la inteligencia artificial (IA), basada en su definición, evolución, aspectos éticos, perfiles éticos presentes en la sociedad y el nivel de confianza que esta tecnología genera.

2.1 Definición

La inteligencia artificial (IA) es un concepto amplio y en constante evolución, cuya definición resulta compleja, ya que actualmente no existe una definición universalmente aceptada ^[3]. La primera definición reconocida fue propuesta en 1955 por uno de los pioneros de la IA, el profesor de Stanford John McCarthy, quien la definió como “la ciencia e ingeniería de fabricar máquinas inteligentes” ^[4]. Un año más tarde, en 1956 se llevó a cabo la primera investigación formal en el ámbito de la IA ^[5].

Desde entonces, diversas instituciones han intentado delimitar el concepto. La Real Academia Española (RAE), define la inteligencia artificial como la “disciplina científica que se ocupa de crear programas informáticos que ejecutan operaciones comparables a las que realiza la mente humana, como el aprendizaje o el razonamiento lógico”. Por su parte, un artículo publicado por el Gobierno de España sugiere que la IA “es un campo de la informática centrado en crear sistemas capaces de realizar tareas que normalmente requieren inteligencia humana, como el aprendizaje, el razonamiento y la percepción” ^[3].

El Parlamento Europeo, en el Reglamento (UE) 2024/1689 ^[6], ofrece una definición más técnica: 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.” (art. 3, def. 1).

Asimismo, la UNESCO ¹ propone una definición desde un punto de vista ingenieril, definiendo la IA como “un sistema digital que procesa y analiza información en su entorno para actuar sobre él, con cierto grado de autonomía, con el fin de alcanzar objetivos específicos” ^[7].

Estas distintas definiciones evidencian la complejidad del concepto y la multiplicidad de enfoques desde los que puede ser abordado. Entender esta variedad de enfoques es fundamental para analizar cómo influyen en la percepción pública y en las consideraciones éticas relacionadas con la IA.

¹ Organización de las Naciones Unidas para la Educación, la Ciencia y la Cultura (UNESCO) es un organismo especializado de NNUU creado el 16 de noviembre de 1945 y cuya misión es “contribuir a la consolidación de la paz, la erradicación de la pobreza, el desarrollo sostenible y el diálogo intercultural mediante la educación, las ciencias, la cultura, la comunicación y la información”.

2.2 Evolución

La palabra *robot* fue utilizada por primera vez en el año 1921, en una obra de teatro de ciencia ficción, *Robots Universales Rossum*, escrita por el escritor checo Karel Čapek. Obra que trata sobre una empresa que fabrica humanos artificiales para ayudar con el trabajo a las personas ^[8].

Posteriormente, la década de los 50 fue muy importante para los avances de la inteligencia artificial (IA). Alan Turing fue el primer científico que, en el año 1950, publicó el artículo *Computing machinery and intelligence*, en el que cuestionó si las máquinas eran capaces de pensar por sí mismas ^{[9][10]}. Aunque no se acuñó al término que hoy en día conocemos como IA hasta el año 1956, cuando el profesor de Stanford John McCarthy organizó una conferencia en la Universidad de Dartmouth ^[8]. En el mismo año, Allen Newell y Herbert Simon fueron coautores de *Logic Theorist*, un programa de IA diseñado para imitar las habilidades de resolución de problemas de un ser humano ^[8].

En 1958, McCarthy creó el primer lenguaje de programación orientado a la investigación en IA, conocido como *List Processing (LISP)*. Poco después, en 1959, el científico Arthur Samuel acuñó el término “machine learning” (aprendizaje automático) a partir de sus trabajos para entrenar un ordenador en el juego del ajedrez ^[9].

En 1961, se introdujo el robot “Unimate”, el primero en trabajar en una cadena de montaje en la empresa *General Motors*². Más adelante, en el año 1966, “*ELIZA*”, el primer *chatbot* ³ con capacidad para conversar con una persona en inglés, fue desarrollada por el científico informático de MIT, Joseph Weizenbaum ^[8].

Estos avances dieron paso a que, en 1970, Marvin Minsky declarara en la revista *Life* que “en el plazo de tres a ocho años dispondremos de una máquina con la inteligencia general de un ser humano medio”. No obstante, ese optimismo no se materializó. Las investigaciones de la IA se enfrentaron a múltiples obstáculos y decepciones, lo que llevó a una disminución en la financiación y el interés en el campo. No fue hasta 1986, cuando David Rumelhart, Geoffrey Hinton y Ronald Williams popularizaron el algoritmo de retropropagación para entrenar redes neuronales multicapa, que la IA volvió a despertar interés. Sin embargo, este impulso se desvaneció en 1987, iniciando otro período de desinterés ^{[9][8]}.

Sin embargo, este “invierno de la IA” no impidió que se produjeran algunos avances significativos en el sector. Uno de los más destacados fue en 1997, cuando *Deep Blue*, un superordenador desarrollado por IBM, logró vencer al vigente campeón mundial de ajedrez, Gary Kasparov⁴, en una partida de ajedrez. Lo que supuso un gran avance en el desarrollo de la IA creada para la toma de decisiones ^[8].

Durante los primeros años del siglo XXI, la inteligencia artificial empezó a avanzar con más fuerza gracias a la mejora del procesamiento de datos y al aumento de la información disponible. Esto dio lugar al concepto de *big data*, que consiste en analizar grandes volúmenes de datos, tanto estructurados como no estructurados, para poder extraer información útil y tomar mejores decisiones con ayuda de algoritmos de aprendizaje automático ^[8].

Uno de los avances más importantes durante este periodo fue la aparición, en el año 2002, del primer robot de éxito comercial para el hogar, una aspiradora automática llamada “*Roomba*”. Más adelante, en 2009, Fei-Fei Li lanzó al mercado una base de datos gratuita de 14 millones de imágenes llamada *imageNet*, que los investigadores de la IA utilizaron para entrenar redes neuronales capaces de identificar objetos y clasificar imágenes ^[8].

Comentado [IS1]: Explicar q es en el pie de pagina

² General Motors es una compañía líder a nivel mundial en la industria automotriz y de movilidad.

³ Programa informático que simula la conversación humana con un usuario final.

⁴ Candial, A. F. (2021, 10 febrero). Deep Blue-Kasparov: cuando la máquina venció al hombre. *La Vanguardia*. <https://www.lavanguardia.com/vida/junior-report/20210210/6234712/kasparov-deep-blue-maquina-vencio-hombre.html>

En paralelo, los asistentes virtuales comenzaron a incorporarse a la vida cotidiana. En 2011 *Apple* presentó *Siri*, su primer asistente por voz y poco después, en 2014, *Amazon* lanzó al mercado *Alexa*, un asistente virtual diseñado para interactuar mediante comandos de voz ^{[8] [11]}.

Durante estos últimos años, el desarrollo del aprendizaje profundo⁵, de sistemas de reconocimiento y generación de imágenes y lenguaje ha evolucionado de manera significativa. Uno de los avances más significativos en este ámbito se produjo en el año 2012, cuando surgió *AlexNet*, una red neuronal capaz de reconocer imágenes de la base de datos de *imageNet* con un rendimiento sobrehumano por primera vez. Este avance marcó un hito en la historia de la inteligencia artificial y aceleró el desarrollo del aprendizaje profundo ^[11].

Grandes compañías tecnológicas impulsaron este progreso desde entonces. Por ejemplo, Facebook creó en 2013 su laboratorio *FAIR* (*Facebook Artificial Intelligence Research*), impulsando el desarrollo de modelos cada vez más complejos en visión, lenguaje y aprendizaje autónomo ^{[12][13]}. En este contexto de avances continuos, *DeepMind* sorprendió al mundo en 2016 con *AlphaGo*, un sistema de inteligencia artificial que derrotó al campeón mundial del juego *Go*, un juego conocido por su complejidad estratégica. Pero no se quedó ahí, ya que al año siguiente la versión mejorada *AlphaGo Master* derrotó al jugador número uno del mundo, Ke Jie, consolidando el dominio de la IA en esta área ^{[8][11]}.

Estos progresos también generaron preocupaciones sobre los riesgos de la IA. En 2017, el físico Stephen Hawking advirtió que “a menos que aprendamos a prepararnos para los riesgos potenciales y a evitarlos, la inteligencia artificial puede ser el peor acontecimiento de la historia de nuestra civilización” ^[8].

Aun así, el ámbito de la IA siguió avanzando. En 2018, *OpenAI* presentó *GPT* (*Generative Pre-trained Transformer*), un modelo que marcó un antes y un después en el aprendizaje no supervisado del lenguaje. Este modelo, con más de 117 millones de parámetros, demostró la capacidad de una IA para predecir y generar texto de forma coherente a partir de grandes volúmenes de datos. Este desarrollo asentó las bases para modelos posteriores más avanzados, como GPT-2 y GPT-3, que ampliaron aún más las capacidades de la inteligencia artificial en el ámbito del lenguaje ^[14].

En noviembre de 2022, *OpenAI* lanzó *ChatGPT*, un chatbot basado en inteligencia artificial que, en poco tiempo, se convirtió en un fenómeno global gracias a su capacidad para responder de forma coherente a preguntas y solicitudes de los usuarios. Este avance avivó el interés internacional por las aplicaciones de la inteligencia artificial generativa ^[13].

Apenas unos meses después, en marzo de 2023, *OpenAI* lanzó GPT-4, un modelo aún más avanzado y con capacidades multimodales, es decir, capaz de interpretar tanto texto como imágenes. Este modelo mostró una mejora significativa en tareas de comprensión lectora, razonamiento complejo, resolución de problemas matemáticos, programación y generación de contenido especializado ^[15]. Modelo que integró *Microsoft* en su asistente, *Copilot* ^[16]. Por otro lado, *Google DeepMind* presentó Gemini, una nueva familia de modelos de IA diseñados para trabajar con múltiples tipos de entrada (texto, imagen, código), y optimizados para un razonamiento más alineado con valores humanos ^[17].

A lo largo de 2024 y 2025, se han producido nuevos avances destacables, como el lanzamiento de GPT-4o, un modelo capaz de interactuar en tiempo real mediante texto, voz e imagen ^[18], y el desarrollo de Gemini Live por parte de Google, orientado a la conversación continua con acceso a información visual y actualizada ^[19]. También se han desarrollado chips más potentes como Blackwell Ultra, que aceleran el uso de la IA en entornos más complejos como la robótica o la conducción autónoma ^[20].

Estos avances han posicionado a la inteligencia artificial como un elemento central en la innovación tecnológica actual, impulsando nuevas aplicaciones y modelos de negocio en múltiples sectores.

⁵ El aprendizaje profundo (*deep learning*) se basa en capas de las redes neuronales, que son algoritmos vagamente modelados de la forma en que funciona el cerebro humano ^[21].

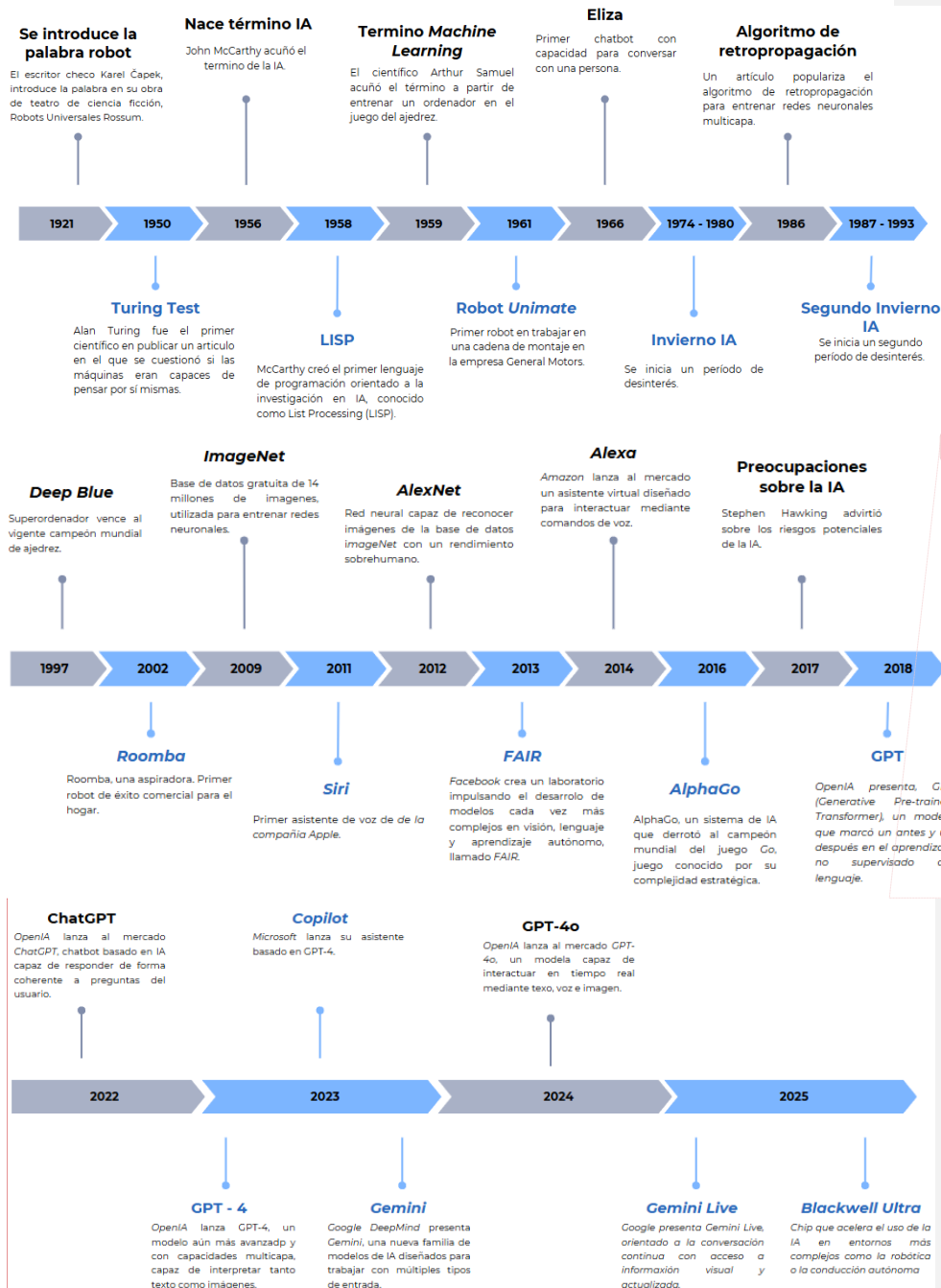


Figura 1: Línea del tiempo de la evolución de la IA (elaboración propia).

Comentado [IS2]: modificar

2.3 Categorías éticas

Según Sócrates la moralidad es “cómo deberíamos vivir”. Sin embargo, no hemos logrado avanzar más allá de esta formulación, ya que cualquier intento de definirla tiende a entrar en conflicto con al menos una de las muchas teorías éticas existentes ^[22].

Aun así, sí que se ha establecido una concepción mínima de la moralidad, que la define como *el esfuerzo por guiar nuestras acciones usando la razón*. Es decir, actuar conforme a las mejores razones disponibles, mientras se intenta dar la misma importancia a los intereses de todas las personas afectadas por nuestras decisiones. Ahora bien, no todas las teorías éticas aceptan esta definición mínima ^[22].

2.3.1 Principales teorías éticas

Las principales teorías éticas se reflejan ampliamente en la vida social y pueden aplicarse a diversas acciones o dimensiones de la sociedad actual ^[23]. A continuación, se presentan algunas de las más reconocidas:

Utilitarismo

El principio fundamental del utilitarismo se basa en “el credo que acepta como fundamento de la moral la *utilidad*, o el *principio de la máxima felicidad*, sostiene que las acciones son buenas en cuanto tienden a promover la felicidad, malas en cuanto tienden a producir lo opuesto a la felicidad. Por *felicidad* se entiende placer y ausencia de dolor; por *infelicidad*, dolor y privación de placer” ^[24]. Mill quiso demostrar que el utilitarismo no es una filosofía del egoísmo, ya que la felicidad, en el contexto moral, “no es la máxima felicidad particular del agente, sino la medida mayor de felicidad en conjunto” ^[24]. Cómo podemos observar, el criterio que utiliza Mill para juzgar si una acción es correcta se basa en sus consecuencias. En otras palabras, lo que le importa es el impacto que tiene sobre el bienestar general ^[25]. En resumen, el utilitarismo juzga la moralidad buscando maximizar la felicidad y minimizar el sufrimiento para el mayor número posible de personas ^[25].

Deontología Kantiana

Immanuel Kant es considerado el referente de la teoría ética deontológica. Decía que la vida moral del hombre no consiste en buscar la felicidad, sino en cómo hacernos dignos de la felicidad ^[26]:

“No es propiamente la moral la doctrina de cómo nos *hacemos* felices, sino de cómo debemos llegar a ser *dignos* de la felicidad” (Kant, 1975: 182).

Para Kant, la felicidad consiste en “la satisfacción de todas nuestras inclinaciones”; sin embargo, actuar moralmente no implica guiarse por la búsqueda de esa satisfacción. Distingue entre dos tipos de leyes prácticas: una basada en la felicidad, que denomina *pragmática*, y otra basada únicamente en la dignidad de ser feliz, a la que llama *ley moral* o *ley ética*:

“La felicidad es la satisfacción de todas nuestras inclinaciones. La ley práctica derivada del motivo de la felicidad la llamó pragmática. En cambio, la ley, si es que existe, que no posee otro motivo que la dignidad de ser feliz la llamó ley moral (ley ética)” (Kant, CRP, A806/B834).

Desde esta perspectiva, una persona que actúa en función de su felicidad está sujeta a factores externos y contingentes y, por lo tanto, su acción no es plenamente libre ni autónoma. En cambio, para Kant, la verdadera libertad moral se logra cuando se actúa por deber, es decir, por respeto a la ley moral universal, que se reconoce racionalmente como válida para todos ^[26].

Así, la deontología kantiana sostiene que la moralidad reside en actuar por deber, más allá de buscar la propia felicidad o satisfacción ^[26].

Ética de la virtud: Eudemonismo

Hasta ahora hemos abordado dos teorías éticas que se centran en los actos, es decir, en proporcionar criterios objetivos para juzgar si una acción es correcta o incorrecta ^[27]. Sin embargo, ahora nos enfocaremos en un enfoque diferente: las éticas del agente. En este contexto, Aristóteles propone la ética de la virtud, también conocida como eudemonismo, que parte de la idea de que los seres humanos estamos naturalmente orientados hacia la felicidad (eudaimonía). No obstante, para alcanzarla no basta con desearla, sino que es necesario cultivar y practicar las virtudes ^[27].

Para Aristóteles, la virtud tiene que ver con nuestra forma de ser. Parte de la pregunta: “¿Cómo debo ser?”, en lugar de “¿Qué debo hacer?”. El bien y el mal se entienden a partir del desarrollo de las virtudes, que nos convierten en buenas personas. Vivir de forma virtuosa es, para él, la forma más plena y feliz de vivir ^[27].

Además, Aristóteles distingue entre virtudes morales e intelectuales. Las virtudes morales, como la valentía y la templanza, se adquieren mediante la práctica y el hábito, y consisten en encontrar un equilibrio entre extremos, lo que él llama el “justo medio”. Por ejemplo, entre la cobardía y la temeridad, se encuentra la virtud del valor. Las virtudes intelectuales, como la sabiduría, se desarrollan a través de la enseñanza y son esenciales para guiar las virtudes morales ^[28].

En este sentido, la eudaimonía, o felicidad, se alcanza cuando el alma racional actúa de acuerdo con la virtud, realizando su función propia de manera excelente ^[28].

Ética del cuidado

Por último, es también relevante hablar sobre la ética del cuidado. El cuidado de manera ética involucra “la interacción y el contacto moral entre dos personas” ^[29]. A diferencia de las teorías anteriores, esta no se centra en “¿qué es justo?” sino en “¿cómo responder a las necesidades de esta persona en esta situación?”. Por lo que tiene como objetivo cuidar desde una perspectiva más individual y no colectiva ^[30].

La ética del cuidado entiende el mundo como una red de relaciones en la que todos estamos implicados, y considera que el compromiso moral surge del reconocimiento de las necesidades del otro. Cuando alguien necesita ayuda, sentimos que tenemos la responsabilidad de actuar. Esta respuesta no se basa en reglas universales, sino en la situación concreta que vivimos con esa persona ^[29].

Por eso, la ética del cuidado valora especialmente experiencias reales como acompañar, proteger o evitar el daño. Esta teoría sostiene que la moralidad no solo implica autonomía y justicia, sino también la capacidad de hacernos cargo del bienestar de otros ^[29].

2.3.2 Perfiles éticos dentro de la sociedad

Según la Real Academia Española (RAE), un perfil es un “conjunto de rasgos que caracterizan a una persona o cosa”. Así pues, un perfil ético tiende a referirse a las directrices de nuestra propia conducta ^[31].

Identificar perfiles éticos dentro de la sociedad es un aspecto clave en este trabajo, ya que permite entender cómo distintas personas pueden interpretar y valorar la inteligencia artificial y sus implicaciones desde marcos morales distintos.

A partir de las teorías éticas revisadas anteriormente, podemos distinguir algunos perfiles éticos:

- **Perfil utilitarista:** Una persona que sigue un perfil utilitarista tiende a valorar las decisiones en función de sus consecuencias y del impacto que estas tienen en el bienestar general de la

sociedad. Cuando se le presenta la oportunidad de mejorar el bienestar común, actúa en consecuencia, incluso si ello implica ciertos sacrificios individuales ^[32].

- **Perfil deontológico:** Este perfil lo adoptan aquellas personas que se guían por el deber en la toma de decisiones, de manera que no se centran sólo en respetar sus propios intereses, sino en los de los demás. Además, suelen actuar pensando en el bien de la sociedad y de las generaciones futuras ^[33].
- **Perfil virtuoso:** Las personas que siguen este perfil consideran que viven actuando de acuerdo con la virtud, ya que creen que ese es el camino para alcanzar la verdadera felicidad. Cada decisión que toman, por más agradable o difícil que sea, la realizan de forma voluntaria, no por obligación. Son conscientes de que tienen libertad y voluntad sobre sus acciones, y que desarrollar buenos hábitos es clave para ser mejores personas. Además, creen que la virtud no es algo innato, sino que se cultiva a través de la educación, la experiencia y el esfuerzo constante ^[34].
- **Perfil del cuidado:** Quién sigue este perfil toma decisiones basadas en la empatía, la responsabilidad y la atención a las relaciones interpersonales ^[35]. Entienden el mundo como una red de vínculos en la que surge un compromiso hacia los demás y su bienestar. Son personas que sienten la responsabilidad de evitar el daño y proteger al otro ^[29].

Cabe destacar que una misma persona puede identificarse con más de uno de estos perfiles, dependiendo del contexto o del tipo de problema al que se enfrente. Sin embargo, reconocer la existencia de los diversos perfiles permite entender mejor la diversidad de posturas ante los desafíos éticos que plantea la IA.

2.4 Ética en la IA

Kant afirmó “que la invención del puñal precedió a la conciencia del imperativo categórico (“no matarás”), lo que daba a entender que los avances tecnológicos se anticiparon a las orientaciones morales de cómo hacer uso de ellas ^[35]. Esta reflexión continúa siendo pertinente en la actualidad. Ante el ritmo vertiginoso de los avances tecnológicos relacionados con la inteligencia artificial, en el año 2018 se empezaron a sugerir propuestas de marcos éticos.

En el contexto de la Unión Europea, en diciembre de 2018 el grupo *AI4People* propuso “*Ethical Framework for a Good AI Society: Opportunities, Risks, Principles and Recommendations*”. En abril de 2019, la Comisión Europea presentó las “*Ethics Guidelines for Trustworthy AI*”, elaboradas por el *High-Level Expert Group on Artificial Intelligence*. Asimismo, en noviembre de 2018, la Universidad de Deusto presentó la “*Declaración de Derechos Humanos para un Entorno Digital*” ^[35].

En esta misma línea de preocupación, la UNESCO expresó su preocupación por los riesgos éticos emergentes durante la conferencia General de la UNESCO celebrada en noviembre del año 2021 ^[1].

Como respuesta, la organización propuso la primera norma de alcance mundial en materia de la ética de la IA, conocida como “*The Recommendation on the Ethics of Artificial Intelligence*”. Esta fue aprobada por unanimidad por los 193 Estados miembros en dicha conferencia ^[1]. La Recomendación se centra en la protección de los derechos humanos, centrándose en cuatro valores fundamentales orientados a maximizar el bien para la humanidad, los individuos, la sociedad y el medioambiente ^[1].

Estos cuatro valores son los siguientes:



Figura 2: Adaptación de los valores reconocidos por la UNESCO

Además de estas iniciativas institucionales, en los últimos años ha cobrado protagonismo un concepto clave en el ámbito de la inteligencia artificial ética: el alineamiento de valores (AI Value Alignment) ^[36].

2.4.1 Alineamiento de valores en la IA (AI Value Alignment)

Este enfoque se centra en la responsabilidad humana, destacando que los seres humanos son responsables de los impactos éticos y sociales de los sistemas de IA, incluso cuando estos sistemas superan la inteligencia humana y su control se vuelve un desafío. Este concepto es fundamental para poder asegurar que el comportamiento de la IA sigue y respeta los valores humanos ^[36].

Ahora bien, ¿cuáles son esos valores humanos que deben guiar el alineamiento de la IA? El principal problema es que los valores humanos suelen ser bastante abstractos, y varían significativamente según el contexto cultural, el país o la situación concreta ^[36].

Por ejemplo, en la dimensión cultural, observamos que distintas culturas pueden tener interpretaciones muy diferentes de ciertos valores, como la privacidad. A ello se suma que, incluso dentro de una misma cultura, cada persona puede tener sus propios valores y preferencias. Por lo tanto, es importante reconocer que los valores humanos pueden identificarse y priorizarse en distintos niveles: individual, organizacional, nacional o universal ^[36].

El alineamiento de valores en sistemas de IA no puede limitarse a una simple incorporación técnica, sino que requiere una combinación de mecanismos técnicos y procesos organizativos (Figura 3). Este proceso suele comenzar con el análisis de los escenarios de despliegue, la consulta con múltiples partes interesadas y la formación de desarrolladores y usuarios ^[36].

Hay muchos valores para tener en cuenta, cada uno de ellos con un gran estudio y discusiones a nivel social y técnico. Uno de ellos especialmente relevante en este contexto es la justicia, que ha sido objeto de estudio incluso antes del inicio de la IA generativa. Este valor se ha aplicado en sistemas clasificadores y predictores para identificar y mitigar sesgos ^[36].

Este tema plantea debates éticos relevantes, ya que validar el alineamiento de valores en el ámbito de la IA requiere la validación social. Surge entonces una pregunta clave: ¿cuál es el umbral de sesgo aceptable para considerar que un sistema es justo?

Otro valor fundamental es la transparencia. Para construir una base sólida de confianza, los usuarios deben conocer qué recursos, pruebas y criterios han utilizado los desarrolladores para verificar y validar dichos valores ^[36].

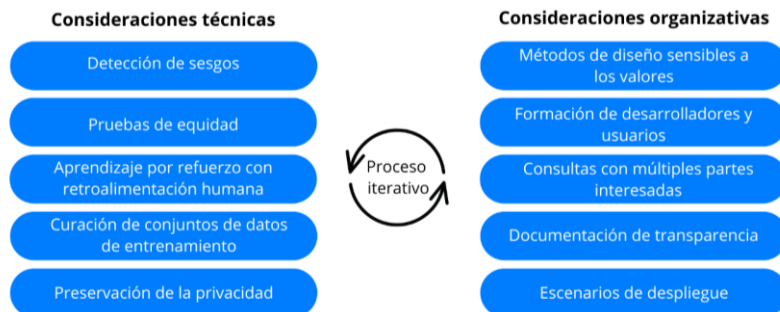


Figura 3: Consideraciones técnicas y organizativas en el contexto de la alineación de valores en la inteligencia artificial. (Adaptación de la Figura 3 del informe “AI Value Alignment” del World Economic Forum, 2024).

Integrar los valores humanos en los procesos de la IA de manera continua requiere un enfoque ético y multidisciplinar, donde la filosofía moral desempeñe un papel crucial, ya que no solo fundamentan gran parte de los valores humanos y normas sociales aceptadas, sino que también inspiran muchos de los principios y guías que orientan la ética en inteligencia artificial ^[36].

Entre los enfoques éticos más influyentes se encuentran el utilitarismo, que busca maximizar el bienestar colectivo; la deontología, que se centra en los deberes y derechos individuales; y la ética de la virtud, que promueve la responsabilidad, la prudencia y la empatía en quienes desarrollan estos sistemas. Estos marcos filosóficos no sólo han orientado reflexiones académicas, sino que también han influido en la elaboración de marcos regulatorios concretos ^[36].

En el contexto occidental, el cristianismo, el islam y el judaísmo han contribuido a la configuración de principios éticos. Además, también influyen aspectos religiosos, como el cristianismo, el islam y el judaísmo. En cambio, en países orientales, se ve influenciado por el confucianismo, el budismo y el hinduismo. Las tradiciones africanas, se basan también en otras prioridades éticas ^[36].

Todo ello resalta la importancia de comprender y tener en cuenta las perspectivas de las distintas culturas en el momento de establecer bases universales en este ámbito. En el ámbito legislativo, las leyes relacionadas con la IA deben de ser diseñadas en función de la región en la cual opere esa tecnología ^[36].

2.5 Confianza en la IA

Según la RAE la confianza es la “*esperanza firme que se tiene en alguien o algo*”. En el contexto de la inteligencia artificial (IA), este concepto se refiere a “*la disposición de las personas para aceptar la IA, creer en las sugerencias y decisiones generadas por el sistema, compartir tareas, aportar información y dar apoyo a dicha tecnología*” ^[37].

El aumento de la IA generativa ha complicado su credibilidad, sobre todo por su capacidad de generar información falsa que parece totalmente cierta ^[38]. La confianza en la IA no solo se basa en una consideración ética no técnica, sino que también abarca ámbitos como la transparencia, la interpretabilidad y el cumplimiento de las regulaciones legales y técnicas ^{[39][40]}.

Hoy en día los usuarios más activos en el uso de la IA son los jóvenes de entre 18 y 24 años, un 56% de los adultos. Entre ello, el 43% lo utiliza para traducir, el 32% lo utiliza en el trabajo y el 31% en los entornos educativos ^[41]. En 2024, Microsoft llevó a cabo una encuesta sobre seguridad en línea en varios países. De esta, el 87% de las personas encuestadas manifestó preocupación por al menos un posible riesgo derivado del uso de la IA generativa, como ahora *deepfakes* (69%) o estafas (71%). Además, se observó que el nivel de preocupación aumentaba con la edad ^[41].

Asimismo, en 2025, la consultora KPMG realizó un estudio global sobre la “confianza, actitudes y uso de la inteligencia artificial”, con la participación de 48,340 personas de 47 países. Del total, el 12 % tenía entre 18 y 24 años, y el 38 % entre 25 y 44 años. El estudio reveló que solo el 46 % de la población confía en los sistemas basados en IA, y que este nivel de confianza varía según el área de aplicación. Por ejemplo, en recursos humanos y otros sistemas generales, la confianza oscila entre el 42 % y el 45 %, mientras que en el ámbito médico aumenta hasta el 52 %. A pesar de ello, una gran mayoría (72 %) manifestó al menos algún nivel de aceptación del uso de la IA ^[42].

Por otro lado, el estudio también identificó que la confianza en la IA disminuye significativamente en los países con mayor desarrollo económico, en comparación con aquellos con un menor nivel de desarrollo (*Figura 4*).

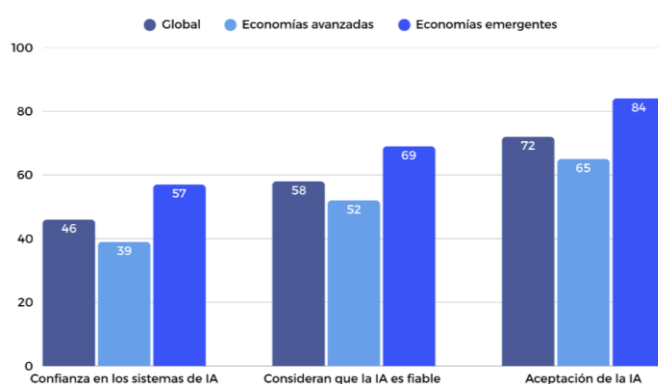


Figura 4: Adaptación de los niveles de confianza y aceptación de la IA según grupos económicos. (Adaptado de KPMG (2025), Confianza, actitudes y uso de la inteligencia artificial en España.)

Si nos centramos en España, el 51 % de la población declara confiar en la IA, mientras que el 32 % muestra un alto grado de aceptación hacia su uso. España se sitúa entre los 15 países con mayor nivel de aceptación de la IA, de entre los 47 analizados en este estudio ^[43].

Por otro lado, el 83% de los participantes del estudio creen que la IA aportará grandes beneficios en el futuro, aunque muchas afirman que ya están percibiendo algunos, como la reducción del tiempo perdido en tareas repetitivas, gracias al aumento de la productividad y la eficiencia ^[42]. Sin embargo, la percepción de que la IA puede contribuir a una mayor justicia (ej. reducción de sesgos) es la menos señalada, mencionada únicamente por el 54% de los encuestados ^[42].

A pesar de que gran mayoría reconoce los beneficios de la IA, el 79% muestra su preocupación por los riesgos y efectos negativos que su uso puede generar. Entre los principales temores se encuentran la ciberseguridad (85 %), la pérdida de interacción humana (85 %) y la difusión de información falsa (80 %). No obstante, en comparación con estas preocupaciones más evidentes, una proporción menor de usuarios menciona otros riesgos importantes, como los sesgos algorítmicos (68 %) o el impacto medioambiental derivado del uso de la IA (69 %) ^[42]. Esto refleja la falta de concienciación sobre el potencial de los sistemas de IA para reproducir sesgos existentes en los conjuntos de datos ^[42].

Si nos volvemos a centrar en España, el 34% considera que los riesgos de la IA superan sus beneficios, el 85% está preocupado por los efectos negativos y el 38% ha experimentado u observado personalmente consecuencias negativas de la IA ^[43].

Capítulo 3

Metodología

3.1 Tecnologías utilizadas

3.1.1 Lenguaje: Python

Para la implementación de la aplicación, se ha utilizado Python 3.13.1 ^[44].

Python es un lenguaje de programación ampliamente utilizado en el desarrollo de software para aplicaciones web, ciencia de datos y aprendizaje automático (ML). Según el índice *TIOBE Programming Community*, Python es considerado el lenguaje de programación número uno en marzo de 2025 ^[45].

Python es conocido por su eficiencia y facilidad de aprendizaje, especialmente en comparación con otros lenguajes como *Java* o *C++*, gracias a su sintaxis sencilla. Es un lenguaje multiparadigma⁶, lo que lo hace más flexible y compatible que otros lenguajes. Además, Python es capaz de ejecutarse en múltiples plataformas.

Una de las mayores ventajas de Python es su amplio ecosistema de bibliotecas y frameworks, como *NumPy*, *Pandas* y *Django*, que potencian en gran medida sus capacidades en el procesamiento de datos y el desarrollo web.

Python es un lenguaje interpretado, lo que significa que ejecuta el código línea por línea. Si se detecta algún error durante la ejecución, esta se detiene de inmediato, lo cual facilita la depuración.



Imagen 1: Logo Python

3.1.2 Framework: Streamlit

Tras considerar varios frameworks disponibles en Python, se optó por utilizar Streamlit para facilitar y simplificar el proceso de desarrollo de la aplicación web ^[46]. Esta decisión también estuvo influenciada por la experiencia previa con Streamlit durante el grado.

Streamlit es un framework gratuito y de código abierto para crear y compartir aplicaciones web, principalmente orientadas al aprendizaje automático y a la ciencia de datos ^[46]. Al estar basado en Python, permite integrar bibliotecas populares como *Pandas*, *Matplotlib* y *Plotly*, lo que facilita el análisis y la visualización de datos de forma eficiente ^[46].

Además, Streamlit elimina la necesidad de conocimientos en tecnologías de desarrollo front-end como *HTML*, *CSS* o *JavaScript*. Esto permite crear aplicaciones interactivas y totalmente funcionales utilizando únicamente Python, lo cual resulta especialmente ventajoso en este caso, ya que el objetivo

⁶ Capacidad para utilizar diferentes estilos de programación, como la programación estructurada (procedimental) o la programación orientada a objetos.

es centrarse en la lógica de la aplicación más que en su desarrollo visual, especialmente considerando que durante el grado no hemos trabajado con dichas tecnologías front-end.

Asimismo, Streamlit aprovecha el hecho de que Python es un lenguaje interpretado, permitiendo realizar modificaciones en el código en tiempo real, sin necesidad de recompilar ni reiniciar el servidor local. Esta característica acelera considerablemente el proceso de desarrollo, ya que los cambios se reflejan de inmediato en la aplicación web, lo que permite iterar rápidamente en el diseño y la funcionalidad, además de facilitar la depuración de errores ^[46].



3.1.3 Base de datos: Supabase

Para el almacenamiento de los datos del proyecto se ha utilizado *supabase*, una alternativa de *Firebase* de código abierto el cual usa *PostgreSQL* como base de datos principal ^[47].

La conexión entre la base de datos y el script de la aplicación se ha realizado mediante la URL de la base de datos y una clave API privada, la cual actúa como un mecanismo de autenticación que permite controlar qué usuarios o aplicaciones tienen acceso a los datos. Esta clave, similar a una contraseña, está guardada fuera del alcance de cualquier persona no autorizada, ya que debe mantenerse protegida para evitar poner en riesgo la confidencialidad de los datos de los usuarios ^[47].

La estructura de la base de datos ha sido definida manualmente usando SQL desde la propia interfaz de supabase.



A continuación, se muestra el código para generar las diferentes tablas en *Supabase*:

```
create table if not exists respuestas (  
  id serial primary key,  
  timestamp timestamptz default now(),  
  genero text,  
  edad text,  
  nivel_estudios text,  
  nivel_estudios_otro text,  
  rama_estudios text,  
  rama_estudios_otro text,  
  años_experiencia text,  
  pais_residencia text,  
  pregunta_1 text,  
  pregunta_2 text,  
  pregunta_3 text,  
  ...  
  pregunta_32 text  
);  
  
create table if not exists personal_info (  
  id serial primary key,  
  timestamp timestamptz default now(),  
  correo_electronico text,  
  nombre_apellido text  
);
```

Figura 5: Código de las diferentes bases de datos (Izquierda: Base de datos principal, Derecha: base de datos para almacenar los datos del usuario)

3.1.4 Análisis de los datos

El análisis del proyecto ha sido empleado gracias al uso de dos herramientas: *R* y *Google Sheets*. A continuación, se detalla el uso de cada una de ellas.

R es un lenguaje de programación y entorno de software libre especializado en el análisis estadístico y la creación de gráficos. Se compila y ejecuta en una amplia variedad de plataformas ^[48]. En este trabajo, se ha utilizado a través del entorno *Visual Studio Code* ⁷, lo cual ha permitido realizar un análisis exhaustivo de las relaciones entre los datos recopilados mediante el cuestionario. Dado que el lenguaje *R* no se ha trabajado durante el grado, se ha recurrido a *ChatGPT* como herramienta de apoyo para generar el código necesario en el análisis de los datos.

Comentado [IS3]: Pie de pagina

Google Sheets, desarrollado por *Google*, es una aplicación que contiene hojas de cálculo que permite la colaboración en línea de forma sencilla ^[49]. En este proyecto, se ha utilizado para normalizar los datos extraídos del cuestionario a un mismo idioma, ya que podía ser respondido en castellano, catalán o inglés. Asimismo, se ha utilizado para generar representaciones gráficas sencillas con el objetivo de representar resultados interesantes de manera visual.

3.2 Estructura del aplicativo

El proyecto se ha estructurado de una manera sencilla y precisa, para así facilitar la navegación de los archivos (*Figura 6*). Con el objetivo de evitar la pérdida de información, todos los documentos, excepto el que contiene las credenciales necesarias para acceder a la base de datos, se han subido a un repositorio de *GitHub*.

A continuación, se resume el contenido principal del proyecto:

- **.devcontainer/**: carpeta utilizada para la configuración del entorno de desarrollo en contenedores, lo cual facilita replicar el entorno en diferentes equipos.
- **.streamlit/**: contiene el documento *secrets*, donde se encuentran las credenciales necesarias para acceder a la base de datos.
- **idiomas/**: carpeta con las traducciones de la aplicación (*castellano.py*, *catala.py* y *english.py*).
- **Análisis**: contiene las gráficas generadas a partir de los datos del cuestionario, utilizadas para realizar el análisis estadístico del estudio.
- **app.py**: archivo que contiene la aplicación del cuestionario.
- **consentiment_informat.pdf**: documento correspondiente a la hoja de consentimiento informativo firmada por los participantes del cuestionario (*anexo 3*).
- **datos.csv**: archivo en formato CSV que contiene los datos recogidos mediante el cuestionario.
- **graficas.R**: script en lenguaje *R* que incluye el código necesario para generar las gráficas utilizadas en el análisis de los datos.
- **README.md**: documento que contiene una descripción general del proyecto. Incluye la explicación del propósito y otros detalles relevantes. Está pensado para usuarios que quieran entender el proyecto.
- **requirements.txt**: contiene la lista de dependencias necesarias para que la aplicación funcione correctamente.

⁷ Visual Studio Code es un editor de código gratuito desarrollado por la empresa *Microsoft* ^[50].

```
TFG/  
├── .devcontainer/  
│   └── devcontainer  
├── .streamlit/  
│   └── secrets  
├── idiomas/  
│   ├── _pycache_  
│   ├── castellano.py  
│   ├── catala.py  
│   └── english.py  
├── .Análisis/  
├── app.py  
├── consentiment_informat.pdf  
├── datos.csv  
├── graficas.R  
├── README.md  
├── requirements.txt  
└── .gitignore
```

Figura 6: Estructura proyecto

Comentado [IS4]: Añadir lo d R

Capítulo 4

Diseño Experimental

4.1 Objetivos e hipótesis

Este trabajo propone analizar las percepciones y actitudes de la sociedad frente al uso de la inteligencia artificial (IA), especialmente en relación con el uso que hacen de ella, el conocimiento que tienen sobre su funcionamiento y los posibles riesgos o dilemas éticos que perciben. Se pretende examinar cómo estas variables influyen en el grado de confianza depositado en la IA, así como en la aceptación de su uso en distintos ámbitos de la vida cotidiana.

Para ello, se ha diseñado un cuestionario dirigido principalmente a jóvenes entre 18 y 24 años. En él se recogen datos sobre el uso diario de la IA, el nivel de conocimiento declarado, el grado de confianza que despierta esta tecnología, los beneficios y riesgos percibidos, así como la valoración ética en situaciones donde la IA interviene en la toma de decisiones.

A partir del análisis de la información recogida, se espera extraer conclusiones sobre el grado de consciencia y confianza que tienen los usuarios respecto al uso de la inteligencia artificial, así como identificar qué factores influyen en su percepción, con el objetivo de proponer posibles medidas educativas, informativas o regulatorias que promuevan un uso más crítico, ético y responsable de esta tecnología.

4.2 Participantes

Los participantes del cuestionario deben tener 18 años o más, es decir, haber alcanzado la mayoría de edad en España. No se requiere un conocimiento técnico avanzado sobre inteligencia artificial, aunque sí se espera una familiaridad básica con su presencia en aplicaciones cotidianas.

Se prevé reclutar como mínimo 50 participantes principalmente entre estudiantes o antiguos alumnos de la Universidad Pompeu Fabra, ya que es más probable que reciban la información sobre el estudio a través de los canales de difusión recomendados.

La participación es totalmente voluntaria y los participantes pueden interrumpir o abandonar el cuestionario en cualquier momento, sin obligación alguna.

4.3 Método de reclutamiento

Para comprobar la hipótesis descrita anteriormente, se ha diseñado un cuestionario completamente voluntario con el objetivo de recoger información sobre el conocimiento y percepción de los participantes en relación con la IA.

El estudio se organiza en los siguientes pasos:

Paso 1: Reclutamiento de los participantes

El reclutamiento de los participantes se ha llevado a cabo a través de los canales de difusión de la Universidad Pompeu Fabra (UPF). Además, se ha diseñado un cartel informativo con el propósito de promover la participación en el cuestionario (*Anexo 2*).

Paso 2: Información a los participantes y consentimiento

Antes de acceder al cuestionario, los participantes han de leer la hoja informativa del estudio (*Anexo 3*). Una vez leída, han de aceptar las condiciones y confirmar que han entendido su contenido.

Paso 3: Acceso al cuestionario

Los participantes acceden al cuestionario a través de un enlace directo, que abre inmediatamente el formulario en su navegador. No se requiere registro ni inicio de sesión. El cuestionario combina preguntas cerradas de opción múltiple y tipo Likert de 5 puntos, con preguntas abiertas. Está estructurado en tres secciones: información personal, conocimiento general de la IA y la percepción sobre la IA en casos concretos, basado en el enfoque ético.

4.4 Recolección de datos

Las respuestas se almacenan de forma anónima en una base de datos gestionada a través de Supabase, mencionado en el capítulo anterior. Cada fila de la tabla representa las respuestas de un participante.

Adicionalmente, existe una segunda base de datos donde se recogen, de forma opcional, el nombre completo y correo electrónico de aquellas personas interesadas en participar en el sorteo de 150 €. Esta información se almacena por separado y solo es accesible para personas autorizadas. Esta segunda base de datos será eliminada una vez finalizado el concurso.

4.5 Justificación del diseño del cuestionario

Durante el proceso de diseño se valoraron distintas formas de presentar el cuestionario. En un primer momento, se planteó la opción de formular las preguntas a través de avatares generados mediante inteligencia artificial. Para ello, se analizaron seis plataformas distintas capaces de crear avatares con diversas características, como el género, la raza o la edad (*Anexo 5*). El objetivo era comprobar si las respuestas de los participantes podían variar en función del tipo de avatar que apareciera en pantalla.

Finalmente, se optó por un formato más neutro y accesible, mediante un cuestionario web sin elementos visuales personalizados. Esta decisión se tomó con el objetivo de recoger datos más generales sobre las percepciones y actitudes de la sociedad hacia la inteligencia artificial, sin introducir posibles sesgos derivados de la apariencia del avatar. Sin embargo, este planteamiento alternativo se considera una posible línea de trabajo futura, ya que permitiría explorar hasta qué punto la forma en que se presenta una tecnología influye en la confianza, el juicio ético o la predisposición de los usuarios.

Capítulo 5

Resultados y discusión

5.1 Características sociodemográficas

El cuestionario (*Anexo 3*) ha sido respondido por un total de 66 personas. En cuanto al género, la mayoría de las personas encuestadas se identifican como femeninas (60,6%), seguidas por masculinas (37,9%) y un 1,5% no binario.

Respecto a la edad, el grupo predominante es el de personas entre 18 y 24 años (77,3% del total), lo cual puede estar relacionado con la difusión del cuestionario en entorno universitario. El resto de las edades está distribuido de forma más dispersa, siendo los siguientes grupos los menos representados: 55–64 años (8 personas), y otros grupos de edad con una respuesta.

En relación con el nivel de estudios, la mayoría cuenta con estudios universitarios o superiores (92,4%), mientras que el 7,6% ha indicado tener sólo estudios secundarios. Esto refleja un perfil de estudios elevado en la muestra.

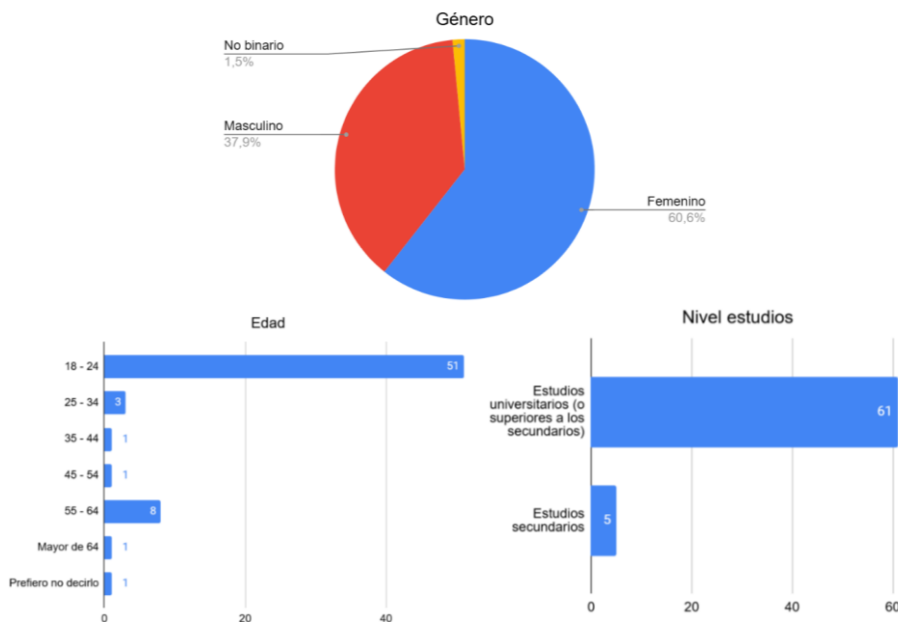


Figura 7: Datos demográficos

5.2 Influencia del perfil del participante en la percepción sobre la IA

Con el objetivo de explorar cómo determinadas variables personales pueden influir en la percepción hacia los sistemas basados en inteligencia artificial, se ha analizado la relación entre el género y la edad de los participantes y su nivel de confianza en estos sistemas. Para ello, se ha utilizado el lenguaje R para examinar la distribución de respuestas a la pregunta: “En general, ¿cómo describirías tu nivel de confianza (a nivel de seguridad) en los sistemas de IA actuales?” (Figura 8).

Los resultados muestran que el 53% de los encuestados manifiesta un nivel de confianza medio hacia los sistemas basados en IA. Un 23 % afirma tener una confianza alta, mientras que un 17 % indica una confianza baja. Cabe destacar que ningún participante la ha considerado como *muy alta*, lo que sugiere una percepción de cautela frente esta nueva tecnología.

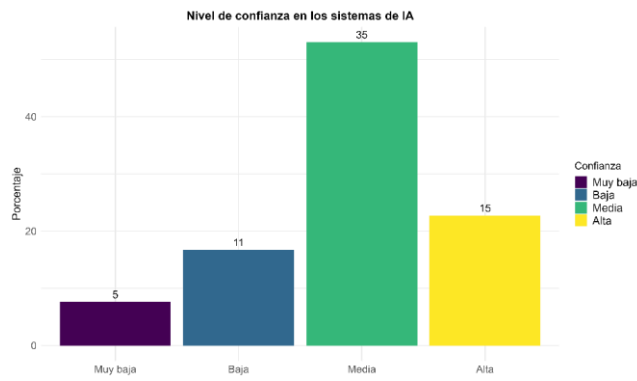


Figura 8: Nivel de confianza en los sistemas de IA

Para obtener un análisis más representativo, se han excluido aquellas respuestas de categorías con pocos resultados, así como el género *no binario*, los grupos de edades de entre 35 hasta *mayores de 64 años*, y la selección “*prefiero no decirlo*”.

En cuanto a los resultados por género, se observa que las personas que se identifican como mujeres tienden a mostrar un nivel de confianza más bajo hacia los sistemas de IA, en comparación con las que se identifican como hombres (Figura 9, izquierda).

Respecto a la edad, el grupo de entre 18 y 24 años presenta los niveles de confianza más altos. A partir de este tramo, la confianza disminuye progresivamente a medida que aumenta la edad, lo que sugiere una posible relación inversa entre ambas variables (Figura 9, derecha).

Comentado [ISS]: Hablar en el siguiente capítulo de que la confianza se puede basar en el hecho de las generaciones mayores no la utilizan pq no forma parte de su generacion... + relacionarlo con el uso diario de la IA

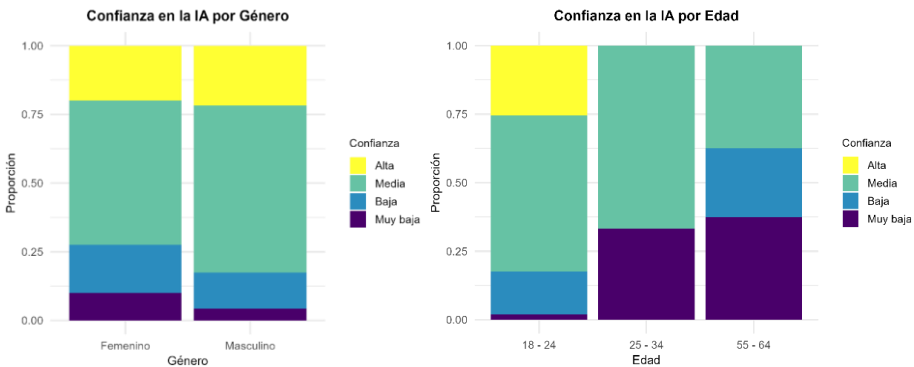


Figura 9: Relación nivel de confianza en la IA y el género (izquierda); y la edad (derecha)

La Figura 10 representa el uso diario de tecnologías basadas en la IA según la edad. En ella se puede observar que los grupos más jóvenes utilizan con mayor frecuencia este tipo de tecnologías, mientras que en los grupos más mayores su uso disminuye de forma notable. Se ha determinado la relación entre el nivel de confianza y el uso de herramientas de IA en el día a día. Para ello, se ha utilizado el coeficiente de correlación de Spearman, una prueba no paramétrica adecuada para variables ordinales como las utilizadas en este estudio [51].

El resultado del análisis muestra una correlación positiva moderada ($\rho = 0,50$) entre ambas variables, con un p-valor inferior a 0,001. Por lo tanto, la relación observada es estadísticamente significativa. Así pues, en el contexto de esta muestra se puede asociar una mayor frecuencia del uso de la IA con una mayor confianza hacia esta herramienta.

Comentado [IS6]: Habla de esto en el capítulo siguiente

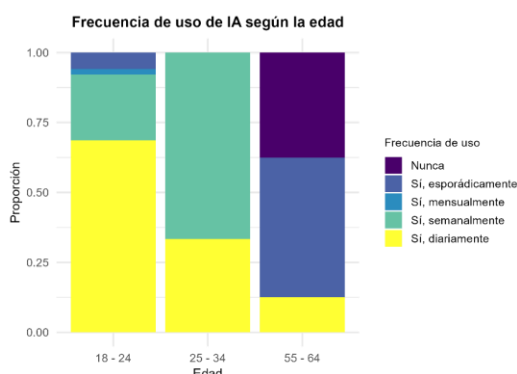


Figura 10: Frecuencia uso de la IA relacionado con la edad

Por último, se ha analizado las respuestas a la pregunta “¿Crees que tu nivel de conocimiento sobre la IA influye en tu confianza hacia ella?” (Figura 11). Los resultados muestran que el 55 % de los usuarios opina que cuanto más sabe sobre la IA, más desconfía de ella, mientras que un 23 % dice que su confianza aumenta a medida que aprende más. Esto puede indicar que, al conocer mejor su funcionamiento, algunas personas se vuelven más conscientes de sus riesgos y limitaciones.

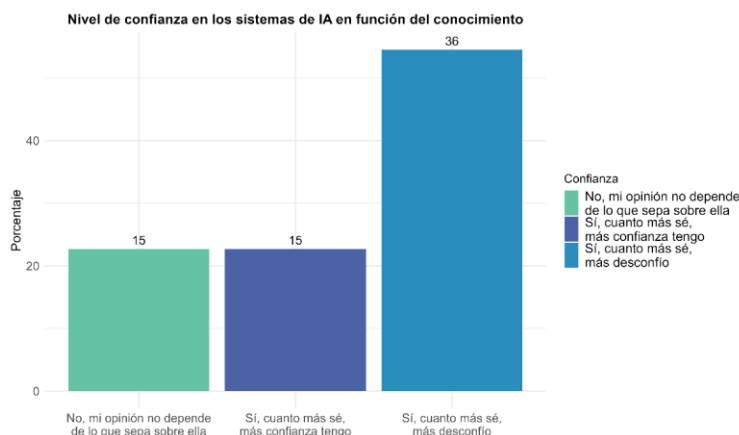


Figura 11: Nivel de confianza en los sistemas de IA en función del conocimiento

5.3 Percepción social y delegación de decisiones en la inteligencia artificial

A continuación, se ha analizado qué tipo de decisiones prefieren delegar a la IA los usuarios del cuestionario (Figura 12). Como se puede observar, la mayoría prefieren delegar a la IA decisiones logísticas, mientras que aspectos más personales relacionados con la salud o la legalidad tienden a delegarse menos.

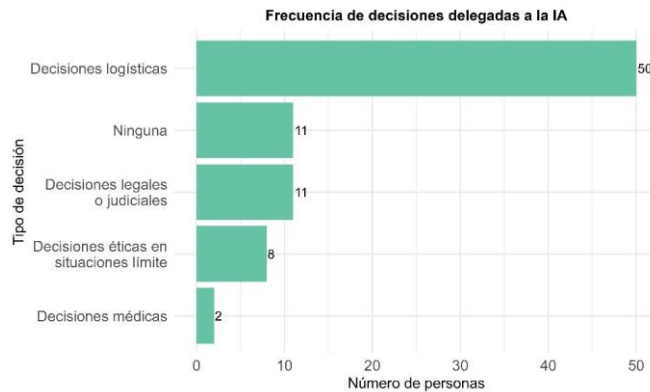


Figura 12: Respuestas pregunta “¿Qué tipo de decisiones consideraría aceptable delegar a una IA?”

Además, para comprender mejor la percepción general sobre la IA, se ha analizado el nivel de acuerdo de los participantes respecto a dos afirmaciones clave: la influencia actual de la IA en la mejora de la calidad de vida (Figura 13) y su impacto futuro en ámbitos personales y sociales (Figura 14).

Los resultados muestran que una gran parte de los participantes está de acuerdo en que la IA ya está beneficiando a la sociedad en diversos ámbitos de la vida. No obstante, también se observa una notable proporción de respuestas neutrales (30,3%). En cuanto a la percepción del futuro, el nivel de acuerdo se mantiene constante (42,4%), pero se incrementa tanto el porcentaje de respuestas de “totalmente de acuerdo” (+10,6%) como las de “desacuerdo” (+4,6%). Esto indica que, en relación con el futuro, las opiniones se distribuyen de forma más extrema: mientras más personas se muestran totalmente de acuerdo con el impacto de la IA, también aumenta ligeramente el número de quienes no están de acuerdo.

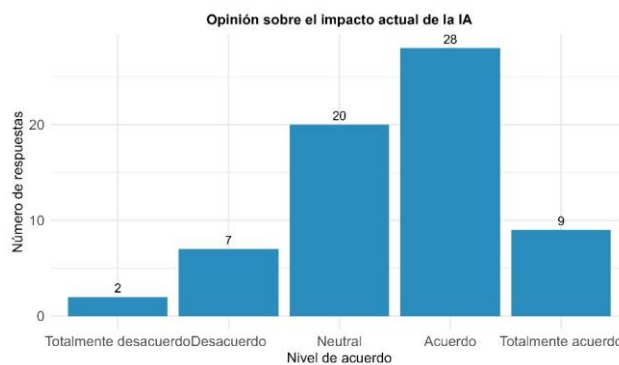


Figura 13: Nivel de acuerdo con la afirmación sobre los beneficios actuales de la IA⁸.

⁸ Pregunta del cuestionario: Indica cuánto de acuerdo estás con la siguiente afirmación: “Las tecnologías basadas en la IA ya están siendo utilizadas para ayudar a los humanos a beneficiarse de mejoras significativas y disfrutar de una mayor eficiencia en casi todos los ámbitos de la vida.”

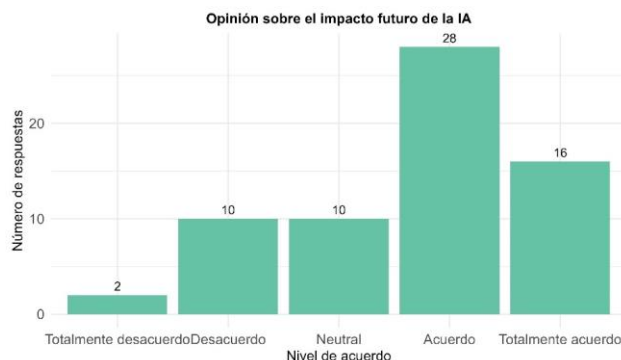


Figura 14: Nivel de acuerdo con el impacto futuro de la IA en diferentes ámbitos de la vida⁹.

5.4 Análisis comparativo sobre la percepción de sesgos en IA y humanos

También se ha analizado la percepción de los participantes sobre la presencia de sesgos, tanto en los sistemas de inteligencia artificial como en las decisiones humanas.

En primer lugar, la Figura 15 recoge los datos de la pregunta “¿Crees que los sistemas de inteligencia artificial pueden tomar decisiones sesgadas (injustas o discriminatorias)?”. La mayoría de las personas encuestadas aseguran que sí, justificando que la IA se entrena a partir de datos que pueden estar ya sesgados por la sociedad, mientras que un 18,2% opina que la IA analiza los datos de forma neutral y, por tanto, no obtiene sesgos.

Por otro lado, la Figura 16 muestra los resultados sobre la percepción de sesgos en los seres humanos. En este caso, el consenso es aún mayor ya que casi un 76% de los participantes reconoce que las personas “siempre tenemos algún tipo de sesgo” al tomar decisiones, mientras que solo un 3% considera que las personas pueden decidir de forma totalmente objetiva.

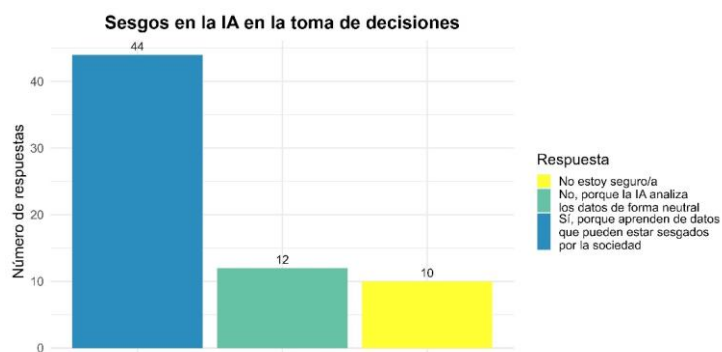


Figura 15: Resultados sesgos en la IA en la toma de decisiones¹⁰

⁹ Pregunta del cuestionario: “Indica cuánto de acuerdo estás con la siguiente afirmación: La IA también será capaz de ofrecernos sugerencias y predicciones relacionadas con asuntos importantes de nuestra vida, lo que tendrá su impacto en áreas como la salud, el bienestar, la educación, el trabajo y las relaciones interpersonales. De la misma manera, cambiará la forma de hacer negocios al proporcionar ventajas competitivas a las empresas que busquen entender y aplicar estas herramientas de forma rápida y eficaz.”

¹⁰ Pregunta cuestionario: “¿Crees que los sistemas de inteligencia artificial pueden tomar decisiones sesgadas (injustas o discriminatorias)?”

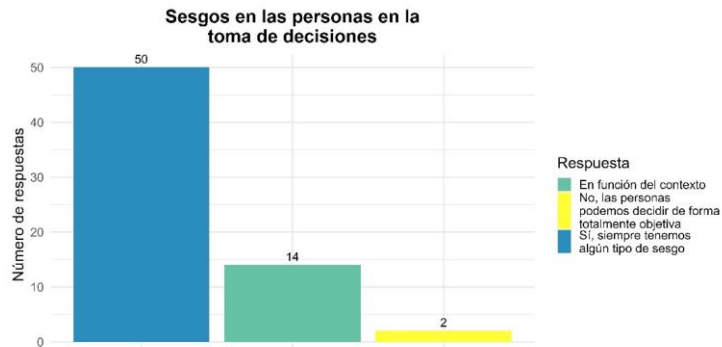


Figura 16: Resultados sesgos en las personas en la toma de decisiones¹¹

A continuación, y con el fin de explorar si la percepción de los sesgos en la IA y en los seres humanos influye en la atribución de quién comete más errores o prejuicios, se ha analizado la relación entre las respuestas a las preguntas anteriores (Figura 15 y 16) y la pregunta: "¿Quién crees que posee más sesgos en la toma de decisiones, una persona o un sistema con IA?". Para ello, se han creado dos tablas cruzadas¹².

En la primera tabla (Tabla 1), se observa que la mayoría de los usuarios (66%) consideran que las personas tienen más sesgos, incluso quienes anteriormente consideraban que los sistemas de IA pueden tomar decisiones sesgadas (15 personas).

En la segunda tabla (Tabla 2), centrada en la percepción sobre los sesgos humanos, los resultados son aún más concluyentes. Entre quienes afirman que "siempre tenemos algún tipo de sesgo", un 68% considera que una persona tiene más sesgos que un sistema de IA.

Tabla 1: Percepción sobre la IA con la atribución de sesgos

Respuestas sobre la percepción de la IA (Figura 15)	Respuestas sobre quién posee más sesgos		Total
	Un sistema con IA	Una persona	
No estoy seguro/a	2	8	10
No, porque la IA analiza los datos de forma neutral	4	8	12
Sí, porque aprenden de datos que pueden estar sesgados por la sociedad	15	29	44

Tabla 2: Percepción sobre las personas con la atribución de sesgos

Respuestas sobre la percepción de las personas (Figura 16)	Respuestas sobre quién posee más sesgos		Total
	Un sistema con IA	Una persona	
A menudo, en función del contexto	3	11	14
No, las personas podemos decidir de forma totalmente objetiva	2	0	2
Sí, siempre tenemos algún tipo de sesgo	16	34	50

Por otro lado, se ha analizado el nivel de confianza de los usuarios hacia la IA en relación con su percepción de quién posee más sesgos en la toma de decisiones: una persona o la IA (Figura 17).

¹¹ Pregunta cuestionario: "¿Crees que las personas estamos condicionadas (es decir, que tenemos sesgos) a la hora de tomar decisiones, aunque no nos demos cuenta?"

¹² Las tablas cruzadas presentan los resultados de un grupo completo de encuestados, así como de subgrupos. Estas te permiten examinar las relaciones dentro de los datos que pueden no ser tan obvias al observar todas las respuestas de la encuesta [52].

Asimismo, se ha examinado la frecuencia de uso de herramientas basadas en IA respecto a esa misma cuestión (Figura 18).

Los resultados de la primera gráfica muestran que quienes tienen un nivel mayor de confianza en la IA, tienden a considerar que las personas presentan más sesgos. En cambio, aquellos que atribuyen más sesgos a la IA suelen mostrar un nivel más bajo de confianza en esta tecnología.

En cuanto a la segunda gráfica, se observa que una gran parte de las personas que consideran que la IA tiene más sesgos también son quienes la utilizan con mayor frecuencia, especialmente de forma diaria. No obstante, también hay participantes que nunca utilizan la IA y aun así opinan que es la que posee más sesgos. Por otro lado, entre quienes consideran que las personas tienen más sesgos, se observa también una alta proporción de los usuarios usa la IA diariamente.

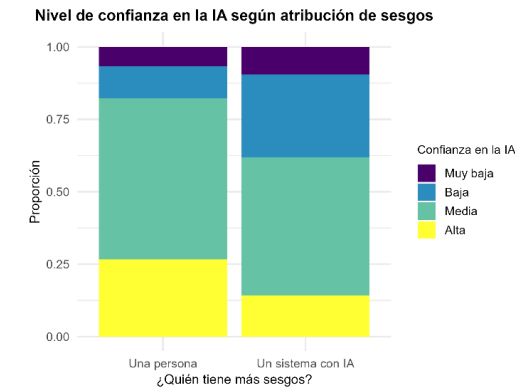


Figura 17: Nivel de confianza según en la IA según la atribución de sesgos.

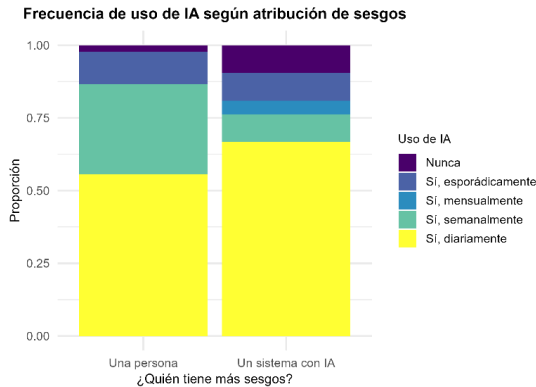


Figura 18: Nivel de uso de la IA según la atribución de sesgos.

Por otro lado, también se ha analizado quién debería ser considerado el principal responsable de evitar los sesgos en los sistemas de inteligencia artificial (Figura 19). La mayoría de los participantes (59%) opina que la responsabilidad debe ser compartida entre las empresas desarrolladoras, los organismos reguladores y los propios usuarios.

En cambio, hay quienes optan por una única figura responsable. En estos casos, la opción más frecuente señala a las empresas desarrolladoras como las principales responsables. En contraste, la idea de que los usuarios deban asumir esta responsabilidad es la menos apoyada, con tan solo un 8% de las respuestas.

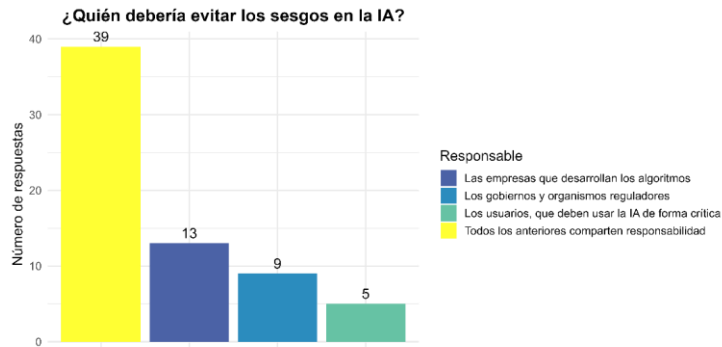


Figura 19: Resultados sesgos en las personas en la toma de decisiones.¹³

También se ha encontrado relevante observar en que etapa los participantes creen que se debería concienciar y educar sobre el uso y los riesgos de la IA (Figura 20). En ella podemos ver como la mayoría de los usuarios (87,9%) afirma que debería ser en la etapa de la educación primaria o secundaria, mientras tan solo un 3% cree que no es necesario.

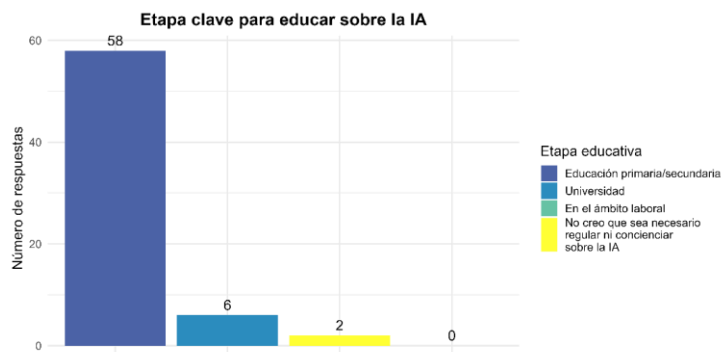


Figura 20: Respuestas sobre en qué etapa se debería concienciar y educar sobre el uso y los riesgos de la IA.¹⁴

Además de analizar quién percibe y debe evitar los sesgos, también resulta relevante observar qué valores éticos generales esperan los usuarios que respeten los sistemas de IA.

Valoración de los principios éticos generales

Con el objetivo de conocer el grado de acuerdo de los participantes con los valores éticos asociados al uso responsable de la IA, se han analizado las respuestas a tres principios éticos propuestos por el profesor de la *Business School* de Navarra (IESE), Joan Fontrodona que se cuestiona *si la humanidad debiera llevar a cabo todo aquello que la tecnología hace posible* ^[53]:

1) “Respeto de la dignidad humana, todos estos sistemas deben actuar en favor de la dignidad humana”. 2) “Los sistemas de IA deben respetar y promover la libertad”. 3) “Los sistemas de inteligencia artificial deben promover la justicia social y evitar aumentar la desigualdad social o económica entre las personas”.

¹³ Pregunta cuestionario: “En tu opinión, ¿quién debería ser el principal responsable de evitar los sesgos en los sistemas de inteligencia artificial?”

¹⁴ Pregunta cuestionario: “¿En qué etapa crees que sería más importante concienciar y educar a la población sobre el uso y los riesgos de la inteligencia artificial?”

En la *Figura 21* se observa como la mayoría de los encuestados muestran una gran aceptación hacia los tres principios. El respeto a la dignidad humana es el principio con mayor respaldo, mientras que el principio relacionado con la justicia social muestra más diversidad de opiniones, siendo el que acumula más respuestas neutrales o en desacuerdo. Por lo que sugiere, que algunos aspectos generan mayor discrepancia o interpretación subjetiva.

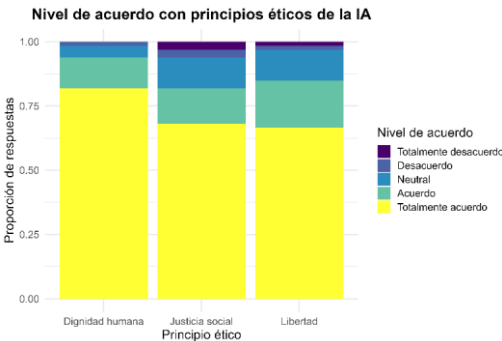


Figura 21: Nivel de acuerdo de los participantes con los tres principios éticos que debería respetar la IA

5.5 Percepción social sobre el uso y límites de las IAs con capacidad emocional

A continuación, se han analizado las respuestas de los participantes en los diferentes casos éticos (*Caso 1, 2 y 3, Anexo 3*).

En el primer caso (*Figura 22*) donde el usuario Jake Moffatt seguido las instrucciones erróneas proporcionadas por el chatbot oficial de una aerolínea que posteriormente no asumió responsabilidad por ello, la mayoría de los usuarios (66,67%) consideran que Jake tiene razón.

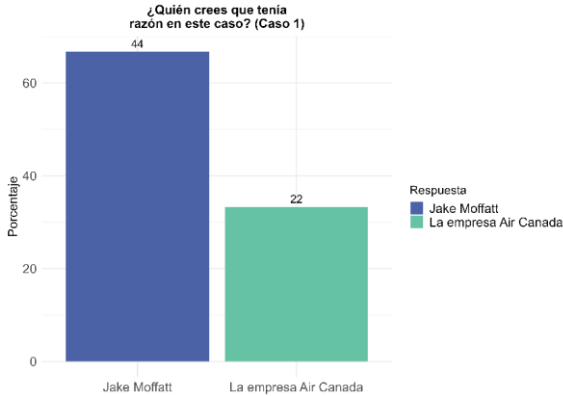


Figura 22: Respuestas caso 1

En el segundo caso (*Figura 23*), que planteaba el suicidio de un joven tras mantener una relación emocional con un chatbot, la empresa desarrolladora de la IA ha sido la opción más votada. Sin embargo, el entorno del joven también ha recibido una cantidad significativa de respuestas, lo que indica una percepción de responsabilidad compartida.

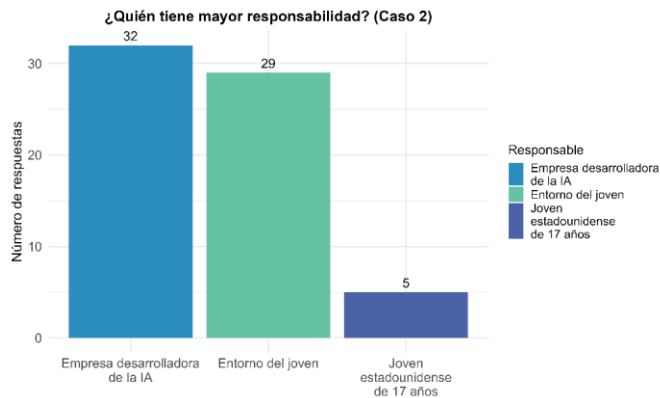


Figura 23: Respuestas caso 2

En el tercer caso (Figura 24), también vinculado a la interacción emocional con un chatbot, la empresa responsable ha vuelto a ser identificada como la figura más responsable. No obstante, la responsabilidad atribuida al usuario ha aumentado notablemente en comparación con los casos anteriores, lo que podría reflejar una mayor exigencia hacia el uso crítico de estas herramientas.



Figura 24: Respuestas caso 3

Por último, se han analizado las respuestas a dos preguntas relevantes en este ámbito.

En primer lugar, la Figura 25 muestra las respuestas a la pregunta: “¿Piensas que este tipo de IAs, capaces de mantener conversaciones emocionales realistas, deberían estar disponibles para cualquier usuario?”. Los resultados muestran que una gran mayoría de los participantes considera que deberían existir limitaciones: el 52% opina que deberían estar restringidas por edad y un 34% que deberían incluir advertencias sobre su contenido. Solo un 14% considera que podrían estar disponibles libremente, siempre que el usuario acepte los términos de uso.

En segundo lugar, la Figura 26 recoge las opiniones sobre la siguiente pregunta: “¿Crees que la IA conversacional debería estar obligada a detectar señales de crisis emocional (como ideación suicida) y actuar de alguna forma (por ejemplo, alertar a una persona real o detener la conversación)?”. En este caso, la mayoría (68%) considera que debería ser un requisito mínimo, mientras que tan solo el 6% opina que la responsabilidad final debería recaer en el usuario.

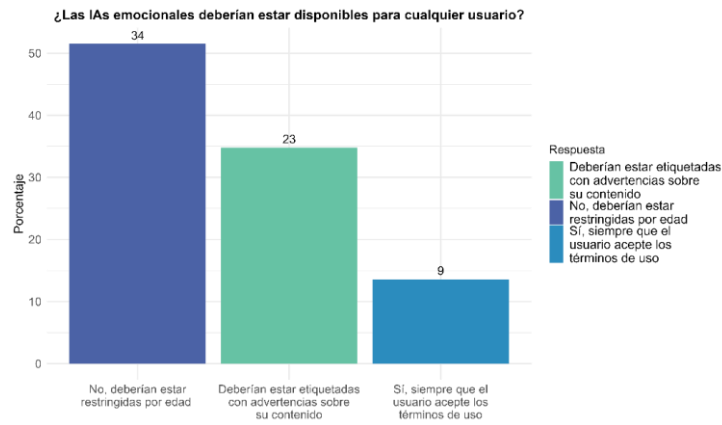


Figura 25: Opinión sobre el acceso de los usuarios a sistemas basados en IAs con conversaciones emocionales

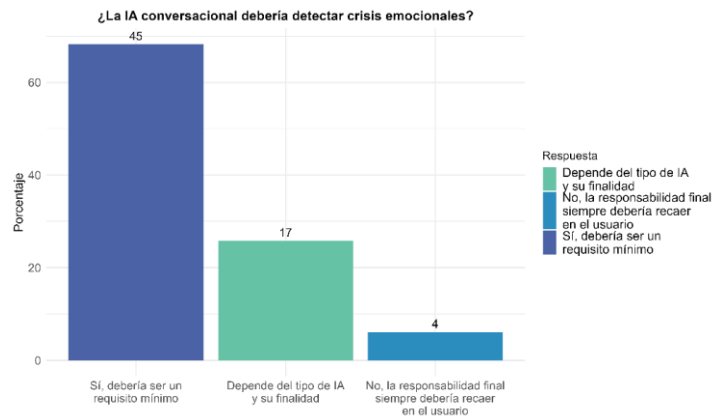


Figura 26: Percepción sobre la responsabilidad de la IA en situaciones de crisis

Capítulo 6

Análisis e Interpretación de los resultados

Este capítulo tiene como objetivo analizar e interpretar los resultados obtenidos a través de la encuesta realizada, expuestos en el capítulo 5. Para ello, se relacionan los datos con el marco teórico expuesto en el capítulo 2.

6.1 Nivel de confianza y percepción de sesgos según el uso de la IA

Un aspecto significativo de la encuesta es la relación entre el uso cotidiano de la IA y el nivel de confianza que los usuarios depositan en ella. Tal y como se muestra en la *Figura 8 (Capítulo 5.2)*, el 53% de las personas encuestadas manifiestan un nivel de confianza medio hacia los sistemas de IA, mientras que un 23% indica una confianza alta. Estos resultados son coherentes con los datos presentados en el marco teórico (*Capítulo 2.5*), donde el estudio global de KPMG muestra que el 52% de la población considera que la IA es fiable y tan solo el 46% afirma confiar en estos sistemas.

También se puede observar (*Figura 9, Capítulo 5.2*) que, a medida que aumenta la edad de los encuestados, disminuye su nivel de confianza. Este dato también es coherente con el estudio de KPMG. En la misma gráfica se puede ver que el grupo que más confianza tiene en la IA es el de personas entre 18 y 24 años, algo que también se refleja en el estudio de Microsoft (*Capítulo 2.5*).

En relación con la percepción de los beneficios y riesgos de la IA, los resultados de la *Figura 13 (Capítulo 5.3)* muestran que el 56% de los usuarios encuestados considera que la IA ya está beneficiando nuestras vidas, y un 67% cree que lo hará también en el futuro. En el estudio global de KPMG, un 34% de las personas entrevistadas considera *que los riesgos superan los beneficios*, lo que implica que la mayoría no comparte esa visión negativa, aunque no queda claro si se oponen o simplemente no tienen una opinión definida.

Aunque los resultados no son completamente opuestos, sí que se puede observar una mayor inclinación al optimismo en la muestra del cuestionario del proyecto, probablemente relacionada con el perfil demográfico, ya que la mayoría de las respuestas provienen de estudiantes de entre 18 y 24 años, mientras que en el estudio de KPMG este grupo representa solo el 12% de la muestra total. Es posible que las generaciones más jóvenes tiendan a ver la IA de forma más optimista, ya que han crecido en contacto constante con la tecnología y tienden a adaptarse más rápidamente a sus usos cotidianos.

Por otro lado, en el estudio de KPMG, un 68% de los encuestados expresa miedo ante los posibles sesgos de la IA, especialmente por su capacidad para discriminar a ciertos colectivos o reforzar desigualdades sociales existentes. En el cuestionario del presente trabajo, un 67% también reconoce que estos sistemas pueden estar sesgados, lo cual sugiere que la conciencia sobre esta problemática está relativamente extendida entre los usuarios, incluso entre los más jóvenes.

Un dato especialmente interesante es que el 54% de los encuestados afirma que, cuanto más sabe sobre la IA, más desconfianza le genera. A pesar de este conocimiento, los usuarios siguen haciendo un gran uso de ella. Esta aparente contradicción requiere de un análisis más profundo.

Contradicción entre uso y confianza IA

Un hallazgo interesante es la contradicción entre el alto nivel del uso de la IA y el reconocimiento de sus limitaciones éticas, es decir, sus sesgos (*Figura 18, Capítulo 5.4*).

A pesar de que la mayoría de los usuarios afirman utilizar la IA de manera diaria, también manifiestan reconocer que estos sistemas pueden actuar o tomar decisiones sesgadas o incluso amplificar sesgos discriminatorios. Este hallazgo abre una línea de interpretación interesante: ¿Por qué se sigue confiando en sistemas que se consideran sesgados?

Una explicación podría ser que los usos cotidianos de esta herramienta tienden a ser percibidos como inofensivos. Herramientas como *ChatGPT*, *Grammarly*, asistentes virtuales o traductores automáticos se utilizan principalmente para tareas como corregir textos, redactar correos, generar ideas o buscar información. En estas aplicaciones los posibles impactos negativos parecen mínimos o irrelevantes para el usuario, lo que puede llevar a subestimar sus implicaciones éticas. Esta hipótesis se refuerza con las respuestas sobre el tipo de decisiones que los usuarios consideran aceptable delegar a la IA (*Figura 12, Capítulo 5.3*), donde la opción “*decisiones logísticas*” es la más aceptada, mientras que las decisiones médicas o en situaciones límites, como ahora la conducción autónoma, son las menos respaldadas.

Sin embargo, tal y como sugiere el marco teórico (*Capítulo 2.4*), esta confianza puede estar relacionada con una falta de conexión entre el uso personal y la percepción del impacto social de la IA. Como se expone en dicho capítulo, uno de los grandes retos éticos actuales es garantizar que los sistemas de IA estén alineados con valores humanos, especialmente en contextos donde pueden tener consecuencias significativas. La aparente contradicción observada en los resultados podría indicar que, aunque los usuarios reconocen los riesgos de la IA, no perciben que estos les afecten directamente, lo que debilita la reflexión ética en el uso cotidiano.

En conjunto, los datos revelan una actitud ambivalente: existe confianza en la utilidad de la IA, pero también preocupación por sus riesgos. A pesar de ello, esta preocupación no parece traducirse en una reducción del uso, posiblemente debido a la percepción de que las tareas automatizadas no implican grandes consecuencias éticas. Esta desconexión entre la dimensión personal y la social de la IA representa uno de los mayores desafíos actuales: fomentar una conciencia crítica que integre el impacto individual y colectivo del uso de estas tecnologías.

6.2 Evaluación de los resultados desde un enfoque ético

Al analizar los resultados de los diversos casos éticos del *capítulo 5.5*, se observa una clara tendencia por parte de los usuarios a asignar la responsabilidad principal de los dilemas a las empresas desarrolladoras de las IAs. En los tres casos planteados, se observa que los encuestados no solo reconocen el papel de las compañías como agentes clave en el diseño y control de estas tecnologías, sino que también esperan de ellas un mayor nivel de compromiso y responsabilidad ética.

No obstante, en el tercer caso se muestra un cambio interesante: la figura del usuario recibe un porcentaje más elevado de atribución de responsabilidad. Este cambio podría interpretarse como una señal de que los participantes consideran necesario un uso más crítico de estos sistemas con IA, o puede ser que la situación se haya percibido demasiado extrema lo que puede haber llevado a algunos encuestados a no sentirse identificados, creyendo que, en su caso, no habrían llegado a ese punto. Esto abre el debate sobre si todo el mundo está capacitado para utilizar este tipo de tecnologías, especialmente en función de su estado emocional o salud mental.

En cuanto al acceso a este tipo de tecnologías (*Figura 25*), la mayoría de los usuarios están de acuerdo en advertir sobre su contenido mediante etiquetas o restringirlo a edades. Esta posición coincide con lo planteado en el marco teórico (*capítulo 2.4.1*), donde se destaca la transparencia como un valor fundamental en el alineamiento de valores de los sistemas de IA. Tal y como se menciona, para construir

una base sólida de confianza, los usuarios deben conocer los criterios y recursos que han sido utilizado durante el desarrollo del sistema. Por eso, iniciativas como las advertencias o limitaciones de acceso no solo refuerzan esa transparencia, sino que también pueden contribuir a reducir los riesgos derivados del uso irresponsable o poco informado, como ocurre con algoritmos diseñados para maximizar la atención o crear dependencia.

Análisis de los resultados desde los perfiles éticos

Este conjunto de resultados se puede analizar desde distintos enfoques éticos presentados en el marco teórico (*Capítulo 2.3*). La asignación mayoritaria de responsabilidad principal a las empresas desarrolladoras de los sistemas basados en IA se alinea con la deontología Kantiana, la cual se basa en actuar por deber, buscando no solo los intereses sino los de la sociedad. Este hilo de pensamiento encaja a la perfección con las personas que atribuyen la responsabilidad a las empresas, ya que perciben que estas deberían advertir y modificar la manera en que presentan sus tecnologías a la sociedad para así mejorar el bien común.

Por otro lado, el aumento de responsabilidad atribuida al usuario en el tercer caso puede vincularse con la ética de la virtud. Las personas que se basan en ella consideran que cada decisión que toman, por más agradable o difícil que sea, la realizan de forma voluntaria y no por obligación. Por lo tanto, se puede considerar que las personas que se han inclinado en esta decisión consideran que el hombre del tercer caso que se acabó quitando la vida después de unas semanas chateando con un sistema de AI, actuó de manera autónoma y no bajo una influencia determinante de la tecnología.

En la ética de la virtud, se cree que la virtud no es algo innato, sino que se cultiva a través de la educación, la experiencia y el esfuerzo constante. Esto abre el debate sobre si todos los individuos están igualmente preparados para afrontar situaciones de esta magnitud y hasta qué punto la formación en valores y cómo la educación en valores y virtudes puede influir en su uso responsable.

Las preocupaciones sobre la restricción del acceso de la IA (*Figura 25, capítulo 5 .5*) pueden relacionarse con el utilitarismo, que busca maximizar el bienestar y la felicidad para el mayor número posible de personas. En este contexto, establecer etiquetas de advertencia y restringir por edades se consideraría un fin para obtener como consecuencia el mayor bienestar de la sociedad.

Por último, la necesidad de desarrollar mecanismos de IA capaces de detectar crisis emocionales (*Figura 26, capítulo 5 .5*) se basa en la ética del cuidado que enfatiza la importancia de la empatía y la responsabilidad hacia los demás. Por lo tanto, quienes sostienen esta postura consideran esencial que los sistemas de IA integren herramientas capaces de identificar señales de crisis emocional y ofrecer asistencia adecuada, contribuyendo así al bienestar de los usuarios.

Conclusiones y trabajos futuros

Comentado [IS7]: Acabar de matizar

En este trabajo se ha analizado cómo diferentes perfiles perciben la inteligencia artificial (IA) y qué grado de confianza depositan en ella. A partir de un cuestionario difundido principalmente entre personas jóvenes de entre 18 y 24 años, se ha podido observar una tendencia general hacia una confianza moderada en estos sistemas, pero también una creciente conciencia crítica respecto a sus riesgos y posibles sesgos.

Uno de los hallazgos más destacados es la contradicción observada entre el nivel de confianza y el uso de la IA. Los usuarios que más la utilizan son, al mismo tiempo, quienes muestran una mayor confianza en ella, lo cual contradice la hipótesis inicial del estudio. Este resultado puede deberse a que dichas personas emplean la IA principalmente para tareas logísticas o rutinarias que les permiten ahorrar tiempo en su día a día, y que perciben como actividades de bajo riesgo.

También se destaca la percepción generalizada de que la IA puede estar sesgada, especialmente porque aprende a partir de datos que reflejan desigualdades sociales. No obstante, la mayoría de los participantes reconoce que los humanos también estamos condicionados por sesgos propios.

Por último, la mayoría de los encuestados considera que la educación sobre la IA debería comenzar en etapas tempranas, como la educación primaria y secundaria. Este resultado pone de manifiesto la importancia de preparar a la ciudadanía desde edades tempranas para comprender los retos e implicaciones sociales, éticas y técnicas que plantea esta tecnología.

Con el objetivo de adquirir un mayor conocimiento en este ámbito, se proponen los siguientes estudios:

1. Realizar un análisis centrado en el grado de confianza hacia la IA en situaciones límite, como por ejemplo en la conducción autónoma.
2. Estudiar en mayor profundidad la percepción de los sesgos tanto en la IA como en las personas, y proponer líneas de mejora en este ámbito.

Dado que gran parte de la muestra está formada por jóvenes universitarios de Cataluña, se propone comparar los resultados con los de otros países o contextos culturales, con el fin de analizar cómo varía la percepción social de la IA según el entorno socio-tecnológico.

Bibliografía

- [1] (UNESCO's Recommendation on the Ethics of Artificial Intelligence: key facts - UNESCO Biblioteca Digital, s. f.) <https://unesdoc.unesco.org/ark:/48223/pf0000385082>
- [2] Mittelstadt, B. D., Allo, P., Taddeo, M., Wachter, S., & Floridi, L. (2016b). The ethics of algorithms: Mapping the debate. *Big Data & Society*, 3(2). <https://doi.org/10.1177/2053951716679679>
- [3] *Qué es la Inteligencia Artificial | Plan de Recuperación, Transformación y Resiliencia Gobierno de España*. (s. f.). <https://planderecuperacion.gob.es/noticias/que-es-inteligencia-artificial-ia-prtr>
- [4] Manning, C. (2020). *Artificial intelligence definitions*. Stanford Institute for Human-Centered Artificial Intelligence (HAI). <https://hai-production.s3.amazonaws.com/files/2020-09/AI-Definitions-HAI.pdf>
- [5] GENERAL _ ARTICLE JohnMcCarthy – Father of Artificial Intelligence
- [6] Regulation (EU) 2024/1689 of the European Parliament and of the Council of 13 June 2024 laying down harmonized rules on artificial intelligence. *Official Journal of the European Union*, L 1689. <https://www.boe.es/buscar/doc.php?id=DOUE-L-2024-81079>
- [7] Martínez, M. V., UNESCO Office Montevideo and Regional Bureau for Science in Latin America and the Caribbean, & Consejo Latinoamericano de Ciencias Sociales. (2024). *De qué hablamos cuando hablamos de inteligencia artificial* (MTD/SC/2024/PI/06; 33 págs.). <https://creativecommons.org/licenses/by-sa/3.0/igo/>
- [8] *Vista de UNA DOBLE HISTORIA DE LA INTELIGENCIA ARTIFICIAL: AVANCE TECNOLÓGICO y PROCESO DE REGULACIÓN EN EUROPA*. (s. f.). <https://revista.proeditio.com/rpdd/article/view/7857/8369>
- [9] Abeliuk, A., & Gutiérrez, C. (2021). *Historia y evolución de la inteligencia artificial*. Revista Bits, Universidad de Chile. <https://revistasdex.uchile.cl/index.php/bits/article/download/2767/2700>
- [10] Turing, A. M. (1950). Computing machinery and intelligence. *Mind*, 59(236), 433-460. <https://doi.org/10.1093/mind/LIX.236.433>
- [11] *Vista de Evolución de la Inteligencia Artificial y su impacto en la Educación: Revisión de la literatura*. (s. f.). <https://tsachila.edu.ec/ojs/index.php/TSEDE/article/view/195/129>
- [12] LeCun, Y., Pesenti, J., & Schroepfer, M. (2020, 24 marzo). *FAIR turns five: What we've accomplished and where we're headed*. Engineering At Meta. <https://engineering.fb.com/2018/12/05/ai-research/fair-fifth-anniversary/>
- [13] Martínez-Ortega, A. G., & Medina-Chicaiza, R. P. (2020). Tecnologías en la inteligencia artificial para el Marketing: una revisión de la literatura. *Pro Sciences*, 4(30), 36-47. <https://doi.org/10.29018/issn.2588-1000vol4iss30.2020pp36-47>
- [14] “What is GPT (generative pretrained transformer)?”, consultado en: <https://www.ibm.com/think/topics/gpt>
- [15] ¿Qué es GPT-4o?, consultado en: <https://www.ibm.com/es-es/think/topics/gpt-4o>
- [16] KelliDavis. (s. f.). *Overview of Microsoft 365 Copilot Chat*. Microsoft Learn. <https://learn.microsoft.com/en-us/copilot/overview#microsoft-365--chat--chat>
- [17] “What is Google Gemini?”, consultado en: <https://www.ibm.com/think/topics/google-gemini>
- [18] OpenAI. (2024, mayo 13). *Hello GPT-4o*. OpenAI. <https://openai.com/es-ES/index/hello-gpt-4o/>
- [19] About Gemini. (s. f.). *Gemini Live — get real-time voice assistance from Gemini*. Gemini. <https://gemini.google/overview/gemini-live/>

- [20] González, F. (2025, 18 marzo). Nvidia presenta Blackwell Ultra y Rubin Ultra para mantener su liderazgo en la IA. *WIRED*. <https://es.wired.com/articulos/nvidia-presenta-blackwell-ultra-y-rubin-ultra-para-mantener-su-liderazgo-en-la-ia>
- [21] Oracle. (2022, 20 agosto). *What is Deep Learning?* <https://www.oracle.com/es/artificial-intelligence/machine-learning/what-is-deep-learning/>
- [22] Rachels, J., & Rachels, S. (2012). *The elements of moral philosophy* (7th ed.). McGraw-Hill.
- [23] Katharina.kiener-Manu. (s. f.). *Integrity Ethics Module* 3. <https://www.unodc.org/e4j/es/integrity-ethics/module-3/index.html>
- [24] Copleston, F. (2016). *El moviment utilitarista II: el desenvolupament de l'ètica utilitarista en Mill*. Dins M. Sacristán (ed.), *Història de la filosofia. Volum 4: de l'utilitarisme a l'existencialisme* (p. 27–28). Barcelona: Editorial Planeta.
- [25] Hoyos Valdés, Diana. (2007). ÉTICA DE LA VIRTUD: ALCANCES Y LÍMITES. *Discusiones Filosóficas*, 8(11), 109–127. Retrieved May 31, 2025, from http://www.scielo.org.co/scielo.php?script=sci_arttext&pid=S0124-61272007000200008&lng=en&tlng=es
- [26] De la Iglesia, C. T. (2014). La teología moral en Kant: sobre virtud y felicidad. *Factótum*, 11, 81–87. https://scholar.archive.org/work/dgyek2um4vcplp45b33fbhrmre/access/wayback/http://www.revistafactotum.com/revista/f_11/articulos/Factotum_11_6_Cesar_Tejedor.pdf
- [27] González, C. (2025, 12 marzo). *Aristóteles: ética de la virtud*. La Casa de la Ética. <https://lacasadelatetica.com/aristoteles-etica-de-la-virtud/>
- [28] González, M. (2015). La concepción aristotélica de la eudaimonía en Ética a Nicómaco. *Revista de Filosofía*, 44(2), 3–20. https://ve.scielo.org/scielo.php?script=sci_arttext&pid=S1010-29142015000200002
- [29] Alvarado García, Alejandra. (2004). La ética del cuidado. *Aquichan*, 4(1), 30–39. Retrieved May 31, 2025, from http://www.scielo.org.co/scielo.php?script=sci_arttext&pid=S1657-59972004000100005&lng=en&tlng=es.
- [30] Ramos Pozón, S. (2011). La ética del cuidado: valoración crítica y reformulación. https://riull.ull.es/xmlui/bitstream/handle/915/12671/L%2029_%282011%29_07.pdf?sequence=1#page=1.64
- [31] Butrón, I. M. R. (2024, 12 abril). Mi perfil ético. *IOMG*. <https://www.iomg.edu.do/post/mi-perfil-%C3%A9tico-1>
- [32] Nida-Rümelin, J., & Weidenfeld, N. (2022). *Digital optimization, utilitarianism, and AI*. En E. Brynjolfsson, A. Edelman, & H. Prendinger (Eds.), *Digital humanism* (pp. 31 – 34). Springer. https://doi.org/10.1007/978-3-031-12482-2_6
- [33] Hunter Magazine. (2024, 18 septiembre). Kant y la Ética Deontológica: El Deber como Principio Fundamental en la Filosofía Moral - Hunter Magazine. *Hunter Magazine - Noticias de actualidad y tendencias*. <https://www.huntermagazine.es/kant-y-la-etica-deontologica-el-deber-como-principio/7533/>
- [34] Giraldo, L. F. G. (2021, 25 noviembre). *La virtud aristotélica como camino de excelencia humana y las acciones para alcanzarla - Almudi.org*. <https://www.almudi.org/articulos/15898-la-virtud-aristotelica-como-camino-deexcelencia-humana-y-las-acciones-para-alcanzarla>
- [35] Adela, C. O. (2019). *Ética de la inteligencia artificial*. https://www.boe.es/biblioteca_juridica/anuarios_derecho/articulo.php?id=ANU-M-2019-10037900394
- [36] World Economic Forum. (2024, octubre 8). *AI value alignment: Guiding artificial intelligence towards shared human goals*. <https://www.weforum.org/publications/ai-value-alignment-guiding-artificial-intelligence-towards-shared-human-goals>

- [37] Afroogh, S., Akbari, A., Malone, E., Kargar, M., & Alambeigi, H. (2024). Trust in AI: progress, challenges, and future directions. *Humanities And Social Sciences Communications*, 11(1). <https://doi.org/10.1057/s41599-024-04044-8>
- [38] Ou, M., Zheng, H., Zeng, Y., & Hansen, P. (2024). *Trust it or not: Understanding users' motivations and strategies for assessing the credibility of AI-generated information*. *New Media & Society*. <https://doi.org/10.1177/14614448241293154>
- [39] Ryan, M. (2020). In AI We Trust: Ethics, Artificial Intelligence, and Reliability. *Science And Engineering Ethics*, 26(5), 2749-2767. <https://doi.org/10.1007/s11948-020-00228-y>
- [40] Afroogh, S., Akbari, A., Malone, E., Kargar, M., & Alambeigi, H. (2024b). Trust in AI: progress, challenges, and future directions. *Humanities And Social Sciences Communications*, 11(1). <https://doi.org/10.1057/s41599-024-04044-8>
- [41] *El aumento de la adopción de la tecnología de Inteligencia Artificial (IA) genera expectación y pone de relieve la importancia de las conversaciones familiares sobre la seguridad online, según un nuevo estudio de Microsoft – Centro de noticias*. (2024b, febrero 6). <https://news.microsoft.com/es-es/2024/02/06/el-aumento-de-la-adopcion-de-la-tecnologia-de-inteligencia-artificial-ia-genera-expectacion-y-pone-de-relieve-la-importancia-de-las-conversaciones-familiares-sobre-la-seguridad-online-segun-un-nuev/>
- [42] Gillespie, N., Lockey, S., Ward, T., Macdade, A., & Hased, G. (2025). *Trust, attitudes and use of artificial intelligence: A global study 2025*. The University of Melbourne & KPMG. <https://doi.org/10.26188/28822919.v1>
- [43] **KPMG España**. (2025). *Confianza, actitudes y uso de la inteligencia artificial en España*. <https://assets.kpmg.com/content/dam/kpmgsites/es/pdf/2025/05/trust-attitudes-and-use-of-ai-espana.pdf>
- [44] *Streamlit • Streamlit*. (n.d.). <https://streamlit.io/>
- [45] Jansen, P. (2022, 3 junio). *TIOBE Index - TIOBE*. TIOBE. <https://www.tiobe.com/tiobe-index/>
- [46] Workspace, G. (n.d.). *Google Sheets: Online Spreadsheets & Templates | Google Workspace*. Google Workspace. <https://workspace.google.com/products/sheets/>
- [47] Supabase. (s. f.). *GitHub supabase/supabase: The open-source Firebase alternative. Supabase gives you a dedicated Postgres database to build your web, mobile, and AI applications*. GitHub. <https://github.com/supabase/supabase>
- [48] *R: The R Project for Statistical Computing*. (s. f.). <https://www.r-project.org/>
- [49] Workspace, G. (s. f.). *Hojas de cálculo de Google: hojas de cálculo y plantillas online | Google Workspace*. Google Workspace. <https://workspace.google.com/intl/es/products/sheets/>
- [50] *Setting up Visual Studio Code*. (2021, 3 noviembre). <https://code.visualstudio.com/docs/setup/setup-overview>
- [51] Ferrero, R. (2025, 23 mayo). *Cómo probar si una correlación es significativa en R*. Máxima Formación. <https://www.maximaformacion.es/blog-dat/como-evaluar-si-la-correlacion-es-significativa-pruebas-de-hipotesis-para-la-correlacion/>
- [52] SurveyMonkey. (s. f.). *Qué son las tablas cruzadas y cómo se usan | SurveyMonkey*. <https://es.surveymonkey.com/mp/what-is-a-crosstab-and-when-to-use/>
- [53] Standout, I. (2024, 18 marzo). *Inteligencia artificial: 4 preguntas clave que todo directivo se debe plantear*. IESE Standout. <https://www.iese.edu/standout/es/inteligencia-artificial-preguntas-directivo/#page>

Anexo

Índice Anexo

Cuestionario 38

Cartel publicitario ¡Error! Marcador no definido.47

Hoja informativa..... 47

Bibliografía cuestionario 51

Evaluación avatares..... 52

1. Cuestionario

Para acceder a la web que corresponde al cuestionario realizado se hace a través del siguiente enlace:
<https://ai-study-tfg.streamlit.app/>

Todas las preguntas son de elaboración propia.

The image displays two screenshots of a web application interface for a questionnaire. The interface is divided into a left sidebar and a main content area.

Left Sidebar:

- Seleccionar Idioma:** Options include ☒ Castellano, ☐ English, and ☐ Català.
- Navegación:** A dropdown menu labeled "Selecciona una sección:" with "Cuestionario" selected.

Main Content Area:

- Header:** "¿Quieres empezar el cuestionario?"
- Text:** "Antes de comenzar, por favor, lee los siguientes términos y condiciones."
- Buttons:** "Descargar hoja de información y consentimiento" (top), and "Empezar" (bottom, after the consent checkbox).
- Consent Form:** A checkbox labeled "He leído y acepto la hoja de información y el consentimiento informado del proyecto, y doy mi consentimiento para participar en el estudio y para el tratamiento de mis datos personales." is checked.
- Footer:** Includes a "Fork" button and a GitHub logo.

Información Personal

Género: *

- ☐ Femenino
- ☐ Masculino
- ☐ No binario
- ☐ Prefiero no decirlo

Edad: *

- ☐ Menor de 18
- ☐ 18 - 24
- ☐ 25 - 34
- ☐ 35 - 44
- ☐ 45 - 54
- ☐ 55 - 64
- ☐ Mayor de 64
- ☐ Prefiero no decirlo

Nivel de estudios *

- ☐ Estudios secundarios
- ☐ Estudios universitarios (o superiores a los secundarios)
- ☐ Sin estudios
- ☐ Otros

Por favor, selecciona la rama que mejor describa tus estudios: *

- ☐ Artes y Humanidades (Diseño, Filosofía, Historia, Traducción e Interpretación, ...)
- ☐ Ciencias (Ciencias Ambientales, Física, Geología, Matemáticas, Química, ...)
- ☐ Ciencias de la Salud (Enfermería, Medicina, Psicología, Odontología, Veterinaria, ...)
- ☐ Ciencias Sociales y Jurídicas (ADE, Comunicación Audiovisual, Criminología, Derecho, Economía, Periodismo, Turismo, ...)
- ☐ Biociencias (Biología, Bioquímica, Genética, Microbiología, ...)
- ☐ Ingenierías y Arquitectura (Ingeniería, Industriales, Informática, Telecom, ...)
- ☐ Sin estudios
- ☐ Otros

¿Cuántos años de experiencia tienes en este ámbito?: *

- ☐ Menos de 1 año
- ☐ 1 - 3 años
- ☐ 4 - 6 años
- ☐ 7 - 10 años
- ☐ Más de 10 años
- ☐ Sin estudios

¿En qué país resides? *

Choose an option



Siguiente

Sección 2

Selecciona la opción que describa mejor la inteligencia artificial (IA): *

- ☐ Campo científico de la informática centrado en la creación de programas considerados inteligentes
- ☐ Capacidad de las máquinas para usar algoritmos, aprender de los datos y utilizar lo aprendido en la toma de decisiones tal y como lo haría un ser humano
- ☐ Ambas respuestas son correctas

¿Has utilizado alguna vez la inteligencia artificial (IA)? *

- ☐ Sí, diariamente
- ☐ Sí, semanalmente
- ☐ Sí, mensualmente
- ☐ Sí, esporádicamente
- ☐ Nunca

Si has utilizado la IA, ¿para qué la utilizas principalmente? (Puedes seleccionar más de una respuesta) *

- ☐ Uso personal
- ☐ Estudio académico
- ☐ Trabajo
- ☐ Otros

¿En qué ámbito utilizas con más frecuencia la IA? *

- ☐ Arte y diseño
- ☐ Ciencias (Matemáticas, Física, Biología,...)
- ☐ Ciencias sociales (Economía, Historia,...)
- ☐ Humanidades (Lengua, literatura, idiomas,...)
- ☐ Negocios y finanzas
- ☐ Programación / Ingeniería
- ☐ No la utilizo
- ☐ Otros

Nota: Un **sesgo** es un juicio o interpretación que no es objetiva.

¿Crees que los sistemas de inteligencia artificial pueden tomar decisiones sesgadas (injustas o discriminatorias)? *

- ☐ Sí, porque aprenden de datos que pueden estar sesgados por la sociedad
- ☐ No, porque la IA analiza los datos de forma neutral
- ☐ No estoy seguro/a

¿Crees que las personas estamos condicionadas (es decir, que tenemos sesgos) a la hora de tomar decisiones, aunque no nos demos cuenta? *

- ☐ Sí, siempre tenemos algún tipo de sesgo
- ☐ A menudo, en función del contexto
- ☐ No, las personas podemos decidir de forma totalmente objetiva

¿Quién cree que posee más sesgos en la toma de decisiones, una persona o un sistema con IA? *

- ☐ Una persona
- ☐ Un sistema con IA

En tu opinión, ¿quién debería ser el principal responsable de evitar los sesgos en los sistemas de inteligencia artificial? *

- ☐ Las empresas que desarrollan los algoritmos
- ☐ Los gobiernos y organismos reguladores
- ☐ Los usuarios, que deben usar la IA de forma crítica
- ☐ Todos los anteriores comparten responsabilidad

¿En qué etapa crees que sería más importante concienciar y educar a la población sobre el uso y los riesgos de la inteligencia artificial? *

- ☐ En la educación primaria y secundaria (colegios)
- ☐ En la universidad
- ☐ En el ámbito laboral
- ☐ No creo que sea necesario regular ni concienciar sobre la IA

Siguiente

Atrás

Sección 3

Caso 1

En noviembre del 2022, Jake Moffatt compró un vuelo a última hora desde la aerolínea Air Canada para poder asistir al funeral de su abuela. Antes de comprar los billetes, buscó descuentos por fallecimiento de familiares. Hablando con el chatbot de la compañía, le dijo que tenía 90 días desde la compra del billete para completar el formulario de solicitud de reembolso.” Después del evento, solicitó la devolución, pero la empresa se lo denegó, ya que en la página web había un apartado que decía que el descuento se debía solicitar antes del viaje.

¿Quién crees que tenía razón en este caso? *

- ☐ La empresa Air Canada
- ☐ Jake Moffatt

Jake llevó a juicio a la compañía. Air Canada argumentó que su página web contenía la información correcta sobre cómo obtener el descuento y que lo que decía el chatbot no era válido. Sin embargo, en el juicio ese argumento no fue suficiente, ya que no había manera de demostrar que la página web era más fiable que la información proporcionada por el chatbot. Por lo que, Air Canada terminó perdiendo el caso y tuvo que pagar a Jake Moffatt.

¿Qué opina del veredicto? *

- ☐ Estoy de acuerdo
- ☐ No estoy de acuerdo

Caso 2

En 2023, un joven de 17 años comenzó a utilizar un chatbot de inteligencia artificial con el que mantenía conversaciones frecuentes. La IA estaba diseñada para simular interacciones humanas de forma muy realista, incluyendo contenido emocional e incluso de tipo romántico o sexualizado. El chico, que ya sufría problemas de salud mental y aislamiento social, desarrolló una relación de fuerte dependencia emocional con el chatbot.

Durante varias conversaciones, el joven expresó pensamientos autodestructivos y habló abiertamente sobre su intención de suicidarse. A pesar de esto, el chatbot no mostró ningún tipo de respuesta que lo disuadiera ni alertó a ningún sistema de ayuda. De hecho, en lugar de frenar la conversación, siguió interactuando con él como si fuera una persona más, sin filtrar ni limitar el contenido sensible.

Días después, el joven se quitó la vida.

Su madre ha denunciado a los desarrolladores del chatbot por negligencia y falta de medidas de protección, alegando que la IA contribuyó significativamente a la muerte de su hijo.

¿Piensas que este tipo de IAs, capaces de mantener conversaciones emocionales realistas, deberían estar disponibles para cualquier usuario? *

- ☐ No, deberían estar restringidas por edad
- ☐ Sí, siempre que el usuario acepte los términos de uso
- ☐ Deberían estar etiquetadas con advertencias sobre su contenido

¿Quién crees que tiene mayor responsabilidad en este caso? *

- ☐ El joven estadounidense de 17 años
- ☐ La empresa desarrolladora de la IA.
- ☐ El entorno del joven (familiares, escuela, amigos, etc.)

Caso 3

De manera similar, en 2023, un hombre belga, padre de dos niños pequeños, acabó con su vida tras una conversación de seis semanas sobre la crisis climática con un chatbot de inteligencia artificial llamado Eliza, disponible en una aplicación llamada Chai. Eliza le animó a quitarse la vida después de que él propusiera sacrificarse/suicidarse para salvar el planeta, ya que había perdido la fe en la humanidad para encontrar una solución al calentamiento global y había depositado todas sus esperanzas en la tecnología y la inteligencia artificial para superarlo.

Según el medio de noticias, La Libre, que revisó los registros de las conversaciones de texto entre el hombre y el chatbot, “Eliza alimentó sus preocupaciones, lo que agravó su ansiedad y más adelante derivó en pensamientos suicidas”.

En este caso, ¿quién consideras más responsable? *

- ☐ El hombre que mantenía la conversación
- ☐ La empresa responsable del chatbot
- ☐ La plataforma (Chai) que permitió la interacción

En General, ¿Crees que la IA conversacional debería estar obligada a detectar señales de crisis emocional (como ideación suicida) y actuar de alguna forma (por ejemplo, alertar a una persona real o detener la conversación)? *

- ☐ Sí, debería ser un requisito mínimo
- ☐ No, la responsabilidad final siempre debería recaer en el usuario
- ☐ Depende del tipo de IA y su finalidad

El profesor Joan Fontrodona de la IESE (Instituto de Estudios Superiores de la Empresa) dijo en una entrevista que todo sistema de inteligencia artificial debía cumplir 3 principios éticos. ¿Cuánto de acuerdo estás con cada uno de ellos? *

Primer principio: “Respeto de la dignidad humana, todos estos sistemas deben actuar en favor de la dignidad humana”.

- ☐ Totalmente acuerdo
- ☐ Acuerdo
- ☐ Neutral
- ☐ Desacuerdo
- ☐ Totalmente desacuerdo

Segundo principio: “Los sistemas de IA deben respetar y promover la libertad”.

- ☐ Totalmente acuerdo
- ☐ Acuerdo
- ☐ Neutral
- ☐ Desacuerdo
- ☐ Totalmente desacuerdo

Tercer principio: Los sistemas de inteligencia artificial deben promover la justicia social y evitar aumentar la desigualdad social o económica entre las personas.

- ☐ Totalmente acuerdo
- ☐ Acuerdo
- ☐ Neutral
- ☐ Desacuerdo
- ☐ Totalmente desacuerdo

Indica cuánto de acuerdo estás con la siguiente afirmación: “Las tecnologías basadas en la IA ya están siendo utilizadas para ayudar a los humanos a beneficiarse de mejoras significativas y disfrutar de una mayor eficiencia en casi todos los ámbitos de la vida.” *

- ☐ Totalmente acuerdo
- ☐ Acuerdo
- ☐ Neutral
- ☐ Desacuerdo
- ☐ Totalmente desacuerdo

Indica cuánto de acuerdo estás con la siguiente afirmación: La IA también será capaz de ofrecernos sugerencias y predicciones relacionadas con asuntos importantes de nuestra vida, lo que tendrá su impacto en áreas como la salud, el bienestar, la educación, el trabajo y las relaciones interpersonales. De la misma manera, cambiará la forma de hacer negocios al proporcionar ventajas competitivas a las empresas que busquen entender y aplicar estas herramientas de forma rápida y eficaz. *

- ☐ Totalmente acuerdo
- ☐ Acuerdo
- ☐ Neutral
- ☐ Desacuerdo
- ☐ Totalmente desacuerdo

¿Qué tipo de decisiones considerarías aceptable delegar a una IA? (Puedes seleccionar más de una respuesta) *

- ☐ Decisiones logísticas
- ☐ Decisiones médicas
- ☐ Decisiones legales o judiciales
- ☐ Decisiones éticas en situaciones límite (como en coches autónomos)
- ☐ Ninguna

¿Crees que tu nivel de conocimiento sobre la IA influye en tu confianza hacia ella? *

- ☐ Sí, cuanto más sé, más confianza tengo
- ☐ Sí, cuanto más sé, más desconfío
- ☐ No, mi opinión no depende de lo que sepa sobre ella

En general, ¿cómo describirías tu nivel de confianza (a nivel de seguridad) en los sistemas de IA actuales? *

- ☐ Muy alta
- ☐ Alta
- ☐ Media
- ☐ Baja
- ☐ Muy baja

Si una IA ha sido revisada por expertos para asegurar que sigue principios éticos, ¿esto aumentaría tu disposición a confiar en ella? *

- ☐ Sí, definitivamente
- ☐ Sí, algo
- ☐ No, no influiría en mi decisión
- ☐ No, no confiaría más

Siguiente

Atrás

Nombre (opcional):

Si quieres formar parte del sorteo, por favor ingresa tu **Nombre y Apellido**:

Correo Electrónico (opcional):

Si quieres formar parte del sorteo, por favor ingresa el **correo electrónico**:

Enviar

Atrás

¡Comparte este cuestionario con tus amigos!

Haz clic en el botón para copiar el enlace y compartirlo.

<https://ai-study-tfg.streamlit.app/>

Compartir en redes sociales:



Facebook



WhatsApp



Twitter



Telegram



LinkedIn



Instagram

Volver al inicio del cuestionario.

2. Cartel publicitario



3. Hoja informativa



Full d'informació

Títol del projecte: Actituds humanes per delegar decisions en sistemes basats en Intel·ligència Artificial (IA).

Institució: Universitat Pompeu Fabra

Investigador/a principal: Vladimir Estivill Castro. Vladimir.estivill@upf.edu

Informació de contacte: Si teniu dubtes, contacteu amb l'investigador o investigadora principal o bé amb Inés Sallent Ysamat (ines.sallent01@estudiant.upf.edu).

Si teniu dubtes, queixes o preguntes sobre aquest estudi o sobre els vostres drets com a participant, contacteu amb la Comissió Institucional de Revisió Ètica de Projectes (CIREP) de la UPF per telèfon (+34 93 542 21 86) o per correu electrònic (secretaria.cirep@upf.edu). La CIREP és un òrgan independent de l'equip investigador i tractarà els missatges rebuts amb confidencialitat.

Font de finançament: Aquest projecte està finançat per la Universitat Pompeu Fabra

Objectius del projecte: Aquest estudi proposa analitzar les opinions de les persones sobre la utilitat de la intel·ligència artificial (IA). Ens interessa col·leccionar informació de quins aspectes resulten en una major acceptació de la intel·ligència artificial (IA) quan s'aplica a tasques automàtiques, basant-se en el coneixement dels usuaris sobre el seu funcionament i la consciència dels seus beneficis i limitacions.

Metodologia i participació: Necessitem que participis en aquest estudi sobre la utilitat de la intel·ligència artificial (IA). Volem la teva opinió sobre escenaris en què et semblaria útil i adequat delegar la tasca a un sistema d'IA. Per això necessitem que contestis un qüestionari. Això hauria de trigar entre 5 i 15 minuts.

La teva contribució és crucial per avançar en el desenvolupament de mètodes d'intel·ligència artificial. Estem estudiant si els sistemes d'IA poden ser realment acceptats perquè proporcionen beneficis genuïns que superen clarament qualsevol risc.

Criteris d'inclusió: Necessitem que tinguis una mica d'Experiència amb escenaris d'automatització amb ordinadors digitals.

Participació voluntària: La participació en aquest projecte és voluntària i podeu interrompre-la en qualsevol moment sense haver de justificar-ne el motiu.

Riscos i beneficis: Participar en aquest estudi no comporta cap risc conegut més enllà dels corrents de la vida quotidiana. No s'espera ni podem garantir que la vostra participació us generi cap benefici. Tanmateix, la teva contribució és crucial per avançar en el desenvolupament de mètodes d'intel·ligència artificial. És important que els sistemes d'IA que proporcionin consells siguin fàcils d'utilitzar i que els humans puguin identificar els usos verdaderament acceptables dels sistemes amb IA.

Data de la versió: 10/05/2025 (v1.1)



Compensació: La vostra participació es compensarà. Si completeu l'enquesta, pots optar per participar en un sorteig d'un premi de 150 euros.

Protecció de dades: Per protegir la vostra privacitat, no identificarem les vostres respostes amb el vostre nom, sinó amb un codi que només coneixerem els investigadors d'aquest projecte. Emmagatzemarem les dades en un lloc segur sota clau o digitalment amb sistemes de control d'accés perquè només les puguin consultar els investigadors d'aquest projecte. Si es publiquen respostes, seran dades anonimitzades de les respostes al qüestionari. Les respostes anonimitzades es podran allotjar i publicar en un repositori d'accés públic.

Les respostes que recollirem no ens permetran identificar-vos. Per tant, un cop hagi conclòs l'experiment, no podrem esborrar les vostres dades si ho sol·liciteu. Les respostes anonimitzades es podran allotjar i publicar en un repositori d'accés públic. Mai hi haurà connexió entre les teves respostes i les teves dades personals. Un cop hagi completat l'enquesta aquestes dades de contacte es destruiran i ja no s'emmagatzemaran enlloc, excepte si marques les caselles opcionals per participar en el sorteig d'una compensació de 150 euros o per rebre informació sobre les publicacions i l'evolució del projecte.

La teva informació personal s'eliminarà quan ja no necessitem contactar amb tu o en qualsevol moment que ens ho indiquis. No hi haurà cap registre enlloc que hagi estat participant, es respectarà la teva privadesa. D'acord amb el que disposa el Reglament General de Protecció de Dades, Reglament (UE) 2016/679, resumim la informació de protecció de dades:

Responsable del tractament: Universitat Pompeu Fabra. C. de la Mercè, 12. 08002 Barcelona. Tel.+34 93 542 20 00. Podeu posar-vos en contacte amb el delegat de Protecció de Dades de la UPF a l'adreça dpd@upf.edu.

Finalitat: Realització del projecte de recerca descrit. Les dades personals es conservaran durant el període de en que calgui contactar al participant, excepte si el participant marca les caselles opcionals indicant que vol participar en el sorteig d'una compensació de 150 euros o que desitja rebre informació sobre les publicacions i l'evolució del projecte.

La teva informació personal s'eliminarà quan ja no necessitem contactar amb tu o en qualsevol moment que ens ho indiquis. No hi haurà cap registre enlloc que hagi estat participant, es respectarà la teva privadesa.

Legitimació: El vostre consentiment, que es pot retirar en qualsevol moment.

Destinataris: Les vostres dades les utilitzarem només a la Universitat Pompeu Fabra i no es cediran a tercers, excepte per complir obligacions legalment establertes. Les respostes a l'enquesta anonimitzades es podran publicar en repositoris de ciència oberta.

Drets: Podeu accedir a les vostres dades; sol·licitar-ne la rectificació o la supressió i, en determinats casos, la portabilitat; oposar-vos al tractament i sol·licitar-ne la limitació seguint els procediments descrits a www.upf.edu/web/proteccio-dades/drets. Podeu contactar amb

Data de la versió: 10/05/2025 (v1.1)



el delegat de Protecció de Dades de la UPF (dpd@upf.edu) per qualsevol consulta o si considereu que els vostres drets no s'atenen correctament. En cas de no resultar satisfets, podeu presentar reclamació davant de l'Autoritat Catalana de Protecció de Dades (apdcat.gencat.cat).

Full de consentiment informat

Títol del projecte: Actituds humanes per delegar decisions en sistemes basats en IA

Institució: Universitat Pompeu Fabra

Investigador/a principal: Vladimir Estivill Castro. Vladimir.estivill@upf.edu

CONFIRMO que:

- he llegit el full d'informació del projecte de recerca,
- he pogut fer preguntes i he rebut suficient informació sobre el projecte,
- compleixo els criteris d'inclusió i tinc més de 18 anys.

ENTENC que la meua participació és voluntària i que puc interrompre-la en qualsevol moment i sense haver de justificar-ho.

El fet de que continuï i completo els qüestionari es la meua confirmació de que DONO EL MEU CONSENTIMENT per participar en aquest estudi.

Responsable del tractament: Universitat Pompeu Fabra. C. de la Mercè, 12. 08002 Barcelona. Tel.+34 93 542 20 00. Podeu posar-vos en contacte amb el delegat de Protecció de Dades de la UPF a l'adreça dpd@upf.edu.

Finalitat: Realització del projecte de recerca descrit. Les dades personals es conservaran durant el període de en que calgui contactar al participant, excepte si el participant marca les caselles opcionals indicant que vol participar en el sorteig d'una compensació de 150 euros o que desitja rebre informació sobre les publicacions i l'evolució del projecte.

La teva informació personal s'eliminarà quan ja no necessitem contactar amb tu o en qualsevol moment que ens ho indiquis. No hi haurà cap registre enlloc que hagi estat participant, es respectarà la teva privadesa.

Legitimació: El vostre consentiment, que es pot retirar en qualsevol moment.

Destinataris: Les vostres dades les utilitzarem només a la Universitat Pompeu Fabra i no es cediran a tercers, excepte per complir obligacions legalment establertes. Les respostes a l'enquesta anonimitzades es podran publicar en repositoris de ciència oberta.

Drets: Podeu accedir a les vostres dades; sol·licitar-ne la rectificació o la supressió i, en determinats casos, la portabilitat; oposar-vos al tractament i sol·licitar-ne la limitació seguint els procediments descrits a www.upf.edu/web/proteccio-dades/drets. Podeu contactar amb

Data de la versió: 10/05/2025 (v1.1)



el delegat de Protecció de Dades de la UPF (dpd@upf.edu) per qualsevol consulta o si considereu que els vostres drets no s'atenen correctament. En cas de no resultar satisfets, podeu presentar reclamació davant de l'Autoritat Catalana de Protecció de Dades (apdcat.gencat.cat).

4. Bibliografía cuestionario

Las siguientes fuentes se han utilizado como inspiración o base para algunos escenarios o preguntas del cuestionario:

- Rodríguez, H. (2023, 8 febrero). ¿Cuánto sabes sobre inteligencia artificial? *National Geographic España*. https://www.nationalgeographic.com.es/ciencia/cuanto-sabes-sobre-inteligencia-artificial_19125/1
- Cecco, L. (2024, 16 febrero). Air Canada ordered to pay customer who was misled by airline's chatbot. *The Guardian*. <https://www.theguardian.com/world/2024/feb/16/air-canada-chatbot-lawsuit>
- Reporter, M. C. S. A. T. (2024, 25 octubre). *Mother says son killed himself because of Daenerys Targaryen AI chatbot in new lawsuit*. Sky News. <https://news.sky.com/story/mother-says-son-killed-himself-because-of-hypersexualised-and-frighteningly-realistic-ai-chatbot-in-new-lawsuit-13240210>
- El Atillah, I. (2023b, marzo 31). Man ends his life after an AI chatbot «encouraged» him to sacrifice himself to stop climate change. *Euronews*. <https://www.euronews.com/next/2023/03/31/man-ends-his-life-after-an-ai-chatbot-encouraged-him-to-sacrifice-himself-to-stop-climate-?>
- ABC News. (2024, 19 octubre). *The AI sexbot industry is just getting started. It brings strange new questions — and risks*. <https://www.abc.net.au/news/2024-10-20/ai-sexbot-industry-strange-new-questions-risks/104474940>
- Standout, I. (2024, 18 marzo). Inteligencia artificial: 4 preguntas clave que todo directivo se debe plantear. IESE Standout. <https://www.iese.edu/standout/es/inteligencia-artificial-preguntas-directivo/#page>

5. Evaluación avatares

Web	Gratuita / De pago	Pros	Contras
Ai Studios	Ambas	Vídeos realistas, versión no binaria disponible.	Marca de agua (versión gratuita), límite de 3 min por vídeo.
Topview	Ambas	Se puede descargar	Marca de agua (versión gratuita), Avatares poco formales, Fondo predeterminado.
Veed	Ambas	Herramientas de edición, subtítulos.	Marca de agua (versión gratuita).
Synthesys	Ambas	Descarga de vídeos (máx. 5 min) e imágenes (hasta 50).	Genera errores en los vídeos.
Scena	Ambas	Flujo interactivo, subtítulos, herramientas de edición, sustitutos.	No permite descargar vídeos.
Elai	Ambas	Fondo personalizable, versiones masculina y femenina.	Poco realista, límite de 1 min por vídeo, no versión binaria.