PERCEPCIÓN PÚBLICA DE LOS SESGOS DE LA IA EN ESPAÑA: EL ROL DE LOS CIUDADANOS EN LA COMPRENSIÓN DE LOS SESGOS DE GÉNERO, RAZA, IDEOLOGÍA Y CULTURA

Germán Rodríguez-Wilches

Directores:
Carlos Arcila Calderón y
Patricia Sánchez Holgado

PLAN DE INVESTIGACIÓN

PROGRAMA DE DOCTORADO FORMACIÓN EN LA SOCIEDAD DEL CONOCIMIENTO

UNIVERSIDAD DE SALAMANCA

04 de junio de 2025

1. INTRODUCCIÓN Y JUSTIFICACIÓN DEL TEMA OBJETO DE ESTUDIO

La inteligencia artificial (IA) generativa se ha integrado rápidamente en diversos ámbitos de la vida cotidiana, tales como la comunicación, la educación y el entretenimiento; estas tecnologías, que incluyen asistentes de redacción como ChatGPT y generadores de imágenes como DALL·E, no solo automatizan tareas, sino que también influyen en la representación de identidades, corporalidades, culturas e ideologías. Este potencial transformador viene acompañado de riesgos éticos y sociales cada vez más documentados, como la amplificación de los sesgos presentes en los datos de entrenamiento (Kirk et al., 2021), lo que se manifiesta en resultados discriminatorios vinculados con la raza, el género y la perpetuación de estereotipos sociales. Como ha señalado la UNESCO (2022), es urgente que el desarrollo de la IA respete la diversidad cultural, los derechos humanos y los principios democráticos. En esta línea, la Unión Europea impulsa una IA confiable a través de marcos como la Assessment List for Trustworthy Artificial Intelligence (ALTAI) y la Ley de Inteligencia Artificial (EU AI Act), orientados a responder a las preocupaciones sociales que genera esta tecnología (Cerezo-Martínez et al., 2024); así como otras iniciativas, entre ellas el Manifiesto para una IA Segura en la Educación (Alier Forment et al., 2024; Garcia-Peñalvo et al., 2024), que establece principios éticos aplicables más allá del ámbito educativo, entre los que destacan la protección de la privacidad de los usuarios y la promoción de la transparencia y la minimización de sesgos.

En España, la IA se ha integrado de forma progresiva; los ciudadanos identifican con mayor claridad su presencia en asistentes virtuales como Siri o Alexa, así como en algoritmos de personalización de contenido en plataformas y tiendas online. No obstante, el uso específico de la IA generativa sigue siendo poco comprendido: la ciudadanía la percibe como algo técnico y distante, con dificultades para entender su funcionamiento o evaluar sus efectos (Arcila-Calderón et al., 2023a). Esta desconexión entre su uso creciente y la escasa comprensión ciudadana genera una percepción pública fragmentada, marcada por el entusiasmo ante su potencial innovador y, a la vez, por la preocupación ante sus implicaciones éticas y sociales (Sánchez-Holgado et al., 2022).

Esta brecha cognitiva es especialmente relevante cuando se trata de identificar sesgos discriminatorios; porque la mayoría de los usuarios pueden no estar formados para reconocer cómo una imagen generada por IA puede reforzar estereotipos de género o cómo un texto aparentemente neutro puede reproducir visiones ideológicas sesgadas. Esta dificultad se agrava por factores estructurales del contexto español, como la diversidad cultural y regional (INE, 2024), la polarización política y el acceso desigual a la tecnología, que condicionan la interpretación de los contenidos algorítmicos; lo que una persona percibe como sesgo puede depender de su marco ideológico, entorno sociocultural o experiencias previas de discriminación. Por esta razón, fomentar la participación ciudadana en la evaluación crítica de estos sistemas resulta clave para empoderar a las comunidades y descentralizar el poder en el diseño y la gobernanza algorítmica (Young et al., 2024); y aunque Europa ha avanzado en marcos normativos para una IA ética, en España aún faltan mecanismos que permitan a la ciudadanía intervenir activamente.

En los últimos años, el campo de estudio sobre IA ha avanzado de forma significativa, no solo desde la ingeniería o la informática, sino también desde las ciencias sociales y la comunicación. Las investigaciones recientes han comenzado a centrarse en aspectos como la percepción pública (Arcila-Calderón et al., 2023b; Brauner et al., 2023; Sánchez-Holgado & Arcila-Calderón, 2024; Sartori & Bocca, 2023), los sesgos algorítmicos (Casals Creus, 2024; Chauhan et al., 2024; Cheong et al., 2024; Perdomo Reyes, 2024; Zhou et al., 2024) y la gobernanza de la IA (Medina Plasencia, 2025), construyendo una visión más compleja sobre el papel de la ciudadanía en relación con la IA. Sin embargo, este corpus sigue siendo limitado, sobre todo en contextos específicos como el español, y carece de enfoques que integren la experiencia ciudadana en la comprensión de los sesgos con el análisis sociotécnico de la IA generativa.

Estas limitaciones se pueden abordar por medio de diferentes teorías metodológicas, como la Teoría de la Difusión de Innovaciones (TDI) (Rogers, 2003), que puede ser útil para entender cómo la IA generativa se propaga en la sociedad española, y cómo factores culturales, sociales y estructurales influyen en la adopción y aceptación de esta tecnología. Complementariamente, un estudio centrado en Brasil ofrece un marco valioso al encontrar que, aunque la percepción de discriminación algorítmica puede ser baja, aumenta entre mujeres y personas de grupos históricamente oprimidos, lo que subraya la importancia de considerar la dimensión subjetiva y social de los sesgos (Borba et al., 2024). En paralelo, la revisión sistemática de Emilio Ferrara (2023) ofrece un análisis crítico, donde destaca que los sesgos en la IA afectan tanto a los datos sintéticos como a los modelos de lenguaje e imagen, con impactos en áreas sensibles como el empleo, la representación de género o la diversidad cultural; este marco técnico y conceptual resulta clave para comprender el impacto que pueden tener los sesgos en los resultados generados por la IA.

En este panorama, la presente investigación se propone ocupar un espacio aún poco explorado: analizar la percepción pública de los sesgos de género, raza, ideología y cultura en la IA generativa en el contexto español, poniendo especial énfasis en formatos de texto e imagen. Para ello se adopta un enfoque metodológico mixto, escalonado y longitudinal, que permite examinar cómo distintos grupos sociales perciben, interpretan y responden a estos sesgos discriminatorios, integrando tanto análisis técnicos de contenido generado por IA como experiencias perceptivas ciudadanas.

Este proyecto aporta originalidad al articular campos diversos, tales como los estudios sobre percepción pública, comunicación digital, sesgos algorítmicos y ciencia ciudadana (Vohland et al., 2021). La metodología propuesta, que combina análisis de formatos, encuestas, grupos focales y experimentos sociales, permitirá examinar rigurosamente cómo se perciben los sesgos según el tipo de contenido y el perfil del receptor. A nivel institucional, esta investigación se alinea con las agendas europeas sobre ética y gobernanza de la IA, aportando evidencia empírica para orientar futuras estrategias de participación pública. Y los resultados no solo tendrán impacto académico, sino que podrán traducirse en materiales educativos, contenidos divulgativos y estrategias pedagógicas para públicos no especializados, facilitando procesos de sensibilización y alfabetización crítica sobre la IA y sus implicaciones sociales.

2. HIPÓTESIS DE TRABAJO Y PRINCIPALES OBJETIVOS A ALCANZAR

La delimitación del problema y el análisis teórico han evidenciado la necesidad de explorar la percepción pública de los sesgos de género, raza, ideología y cultura en la IA generativa en España. Ante la escasez de estudios sistemáticos en este contexto, se ha optado por un enfoque exploratorio y comparativo que integra las dimensiones perceptiva, sociocultural y participativa de la ciudadanía frente a los contenidos generados en formatos textuales y visuales. Para orientar este análisis, se formulan un conjunto de preguntas de investigación que buscan dar respuesta a los principales vacíos identificados en la literatura sobre percepción pública, sesgos algorítmicos y participación ciudadana en el ámbito de la IA generativa.

Diversos trabajos recientes muestran que, aunque la ciudadanía reconoce la creciente presencia de estas tecnologías en su vida diaria, persisten limitaciones significativas en cuanto a su comprensión, así como preocupaciones éticas relacionadas con sesgos y falta de neutralidad(Arcila-Calderón et al., 2023b; Brauner et al., 2023; Sánchez-Holgado & Arcila-Calderón, 2024). Además, hasta el momento no se han encontrado estudios en España que exploren de manera específica la percepción de sesgos discriminatorios en contenidos generados por IA en formatos de texto e imagen. Esto plantea la necesidad de formular la primera pregunta de investigación:

RQ1 - ¿Cómo percibe la ciudadanía española la presencia de sesgos discriminatorios (de género, raciales, ideológicos y culturales) en contenidos generados por IA generativa en formatos de texto e imagen?

Otro aspecto crucial es considerar cómo esta percepción varía entre distintos segmentos de la población. Investigaciones previas han evidenciado que los colectivos históricamente discriminados presentan una mayor sensibilidad para identificar sesgos en sistemas de IA (Borba et al., 2024), incluso cuando estos son sutiles o difíciles de identificar. Sin embargo, en el contexto español no existen estudios empíricos que analicen cómo estas variaciones afectan la percepción de sesgos en contenidos generados por IA. En respuesta a este vacío, se plantea la segunda pregunta:

RQ2 - ¿De qué manera varía la percepción de los sesgos de la IA generativa entre distintos grupos sociodemográficos de la ciudadanía española?

Asimismo, la alfabetización tecnológica y el conocimiento crítico sobre IA se presentan como factores que potencian la capacidad de los individuos para identificar riesgos y sesgos (Borba et al., 2024; Sánchez-Holgado et al., 2022; UNESCO, 2022). Sin embargo, se desconoce en qué medida esta relación se manifiesta en el caso de la ciudadanía española. Esto lleva a plantear la tercera pregunta:

RQ3 - ¿Cómo influye el nivel de familiaridad tecnológica o la alfabetización digital de los ciudadanos en su capacidad para detectar los sesgos discriminatorios presentes en contenidos de IA generativa?

Además de analizar la percepción de sesgos, resulta crucial explorar las reacciones actitudinales de la ciudadanía española al identificarlos en contenidos generados por IA, y en este marco, el formato del contenido adquiere relevancia. Si bien existen estudios

que analizan la presencia de sesgos en imágenes generadas por IA (AlDahoul et al., 2025; Casals Creus, 2024; Cheong et al., 2024), no se han encontrado investigaciones en España que comparen las reacciones ciudadanas frente a contenidos textuales e icónicos. La importancia de esta comparación se apoya en teorías de la comunicación que indican que los estímulos visuales provocan respuestas emocionales más inmediatas que los textuales (Clark & Paivio, 1991). Con base en estas consideraciones, se plantea la siguiente pregunta:

RQ4 - ¿Cómo reaccionan los ciudadanos españoles ante la identificación sesgos discriminatorios en contenidos generados por IA, y en qué medida dichas respuestas varían según el formato del contenido (texto versus imagen) y el perfil sociodemográfico?

Estas preguntas de investigación abordan de manera integrada la percepción y las reacciones de la ciudadanía española ante los sesgos discriminatorios presentes en los contenidos generados por IA generativa. Al captar no solo la conciencia crítica, sino también las implicaciones emocionales y sociales de dicha percepción, el proyecto contribuye a llenar un vacío importante en el campo de la comunicación y la tecnología. Además, se alinea con perspectivas de ciencia ciudadana y gobernanza participativa que promueven una implicación activa del público en la evaluación ética de los desarrollos tecnológicos (Vohland et al., 2021; Young et al., 2024).

Objetivos de la Investigación

En consonancia con las preguntas formuladas, la investigación se estructura en torno a un objetivo general que sintetiza el propósito global del estudio, y cuatro objetivos específicos que orientan sus distintas dimensiones:

OG1 - Analizar cómo la ciudadanía española percibe, interpreta y responde a los sesgos discriminatorios de género, raza, ideología y cultura presentes en contenidos generados por IA generativa en texto e imagen, considerando las diferencias entre formatos y las variaciones según el perfil sociodemográfico y el nivel de familiaridad tecnológica.

OE1 – Explorar cómo percibe la ciudadanía española los sesgos discriminatorios de género, raza, ideología y cultura en contenidos generados por IA generativa, tanto en formato textual como en formato visual.

OE2 – Comparar la percepción de los sesgos de la IA generativa entre distintos grupos sociodemográficos de la ciudadanía española, en función de variables como la edad, el género y la ideología política.

OE3 – Evaluar en qué medida el nivel de familiaridad tecnológica o alfabetización digital influye en la capacidad de los ciudadanos españoles para detectar y comprender los sesgos presentes en contenidos generados por IA.

OE4 – Analizar las reacciones de ciudadanos españoles ante la identificación de sesgos en contenidos generados por IA, evaluando las diferencias en función del formato del contenido (texto versus imagen) y del perfil sociodemográfico.

3. METODOLOGÍA A UTILIZAR

La investigación se desarrollará mediante un enfoque metodológico mixto (Creswell & Plano Clark, 2017), de carácter secuencial y longitudinal (Hernández Sampieri et al., 2014), con el propósito de analizar cómo los ciudadanos españoles perciben, interpretan y valoran los sesgos de género, raza, ideología y cultura presentes en los contenidos generados por IA generativa. La estrategia combina técnicas cualitativas (Krueger & Casey, 2015) y cuantitativas (Igartua Perosanz, 2006), distribuidas en tres fases progresivas que abordan tanto contenidos textuales como visuales, integran componentes experimentales (Shadish et al., 2002) de análisis perceptivo, y culminan en acciones de transferencia social orientadas a la alfabetización mediática y la gobernanza inclusiva de la IA.

Fase 1 – Exploración de percepciones ciudadanas ante sesgos en contenidos textuales generados por IA

La primera fase se centrará en el análisis de percepciones frente a contenidos textuales generados por IA. Esta etapa combinará técnicas cualitativas (grupos focales) y cuantitativas (encuesta nacional), estructurándose en dos estudios complementarios:

F1.1 – Estudio cualitativo: Grupos focales exploratorio

Se realizarán tres grupos focales segmentados por edad (18–30 años, 31–55 años y 56 años en adelante), con 10 participantes en cada grupo, asegurando el equilibrio de género. Durante las sesiones se explorará el conocimiento espontáneo sobre IA, la identificación de posibles sesgos, las emociones que estos provocan y las valoraciones éticas asociadas. Las sesiones serán grabadas, transcritas y analizadas mediante codificación temática cualitativa. Los resultados cualitativos servirán de base para el diseño del cuestionario estructurado de la siguiente subfase.

F1.2 – Estudio cuantitativo: Encuesta nacional

A partir de los hallazgos cualitativos, se diseñará una encuesta estructurada que incluirá fragmentos textuales con sesgos predefinidos. Se aplicará a una muestra de 400 ciudadanos españoles, seleccionados mediante cuotas representativas de edad, género y comunidad autónoma. La encuesta evaluará la capacidad de detección de sesgos, las emociones asociadas y las valoraciones éticas, así como el nivel de familiaridad tecnológica de los participantes. El análisis de los datos incluirá estadísticas descriptivas y pruebas inferenciales para identificar patrones y diferencias entre grupos.

Fase 2 – Exploración y experimentación sobre percepción de sesgos en imágenes generadas por IA

Esta segunda fase amplía el análisis a contenidos visuales generados por IA, con el objetivo de estudiar cómo se perciben los sesgos según su nivel de explicitud y el perfil del receptor. Consta de un estudio cualitativo y un experimento controlado:

F2.1 – Estudio cualitativo: Grupos focales sobre percepción de sesgos en imágenes

Se realizarán tres grupos focales con el mismo esquema de segmentación que en la Fase 1.1. Se presentarán imágenes generadas por IA con sesgos discriminatorios en dos niveles: explícitos y sutiles, mostrando un máximo de dos imágenes por tipo de sesgo. Cada sesión permitirá explorar la detección espontánea de los sesgos, la carga emocional provocada y las valoraciones éticas asociadas. Este estudio servirá también como pretest cualitativo para la selección y ajuste de las imágenes que serán utilizadas en el experimento posterior.

F2.2 – Estudio experimental: Experimento online basado en percepción de imágenes

El diseño experimental contempla un factor manipulado: el nivel de explicitud del sesgo (explícito vs implícito), y variables sociodemográficas observadas, como edad, género y familiaridad tecnológica. Cada participante visualizará de forma aleatoria un conjunto de imágenes (máximo dos imágenes por tipo de sesgo), respondiendo tras cada estímulo a preguntas breves sobre la detección de sesgos, la identificación del tipo de sesgo percibido y la valoración emocional. La muestra estará compuesta por 400 ciudadanos españoles, seleccionados mediante cuotas representativas por edad, género y comunidad autónoma. El análisis de resultados se realizará mediante técnicas estadísticas descriptivas e inferenciales, permitiendo comparar tasas de detección entre imágenes explícitas e implícitas, así como analizar diferencias significativas según grupos sociodemográficos. Este análisis posibilitará contrastar las hipótesis de investigación relacionadas con la sensibilidad hacia los sesgos visuales y la influencia de las características sociodemográficas.

Fase 3 – Transferencia social y construcción de recursos para la alfabetización mediática y la gobernanza inclusiva de la IA generativa

La tercera fase transformará los hallazgos empíricos en herramientas accesibles para la ciudadanía y los actores institucionales, materializando el compromiso de retribución social del conocimiento.

F3.1 – Producción de materiales para la alfabetización mediática y digital

Se diseñará una guía visual de identificación de sesgos en IA generativa, infografías didácticas sobre ejemplos de sesgos, un glosario ilustrado de conceptos clave y recomendaciones pedagógicas para contextos educativos. Estos materiales estarán basados en los resultados de la investigación y serán distribuidos en formato digital abierto, orientados a estudiantes, docentes, bibliotecas y colectivos ciudadanos. También, se redactará un documento de recomendaciones dirigido a instituciones públicas, diseñadores de tecnología, educadores y comunicadores, abordando estrategias de transparencia, control de sesgos y alfabetización mediática. Este documento se alineará con marcos normativos internacionales y con los principios de gobernanza participativa y ciencia ciudadana.

4. MEDIOS Y RECURSOS MATERIALES DISPONIBLES

Este proyecto doctoral se apoyará en una combinación de recursos humanos, institucionales y tecnológicos que permitirán desarrollar de forma estructurada y rigurosa las distintas fases de la investigación.

Recursos humanos

- La investigación está dirigida por un equipo académico vinculado al **Observatorio de los Contenidos Audiovisuales (OCA)**, Grupo de Investigación Reconocido de la Universidad de Salamanca y Grupo de Investigación de Excelencia de Castilla y León (GR319).

- Se prevé la colaboración con expertos en comunicación digital, análisis de datos y estudios culturales, dentro del entorno académico de la universidad.

- La participación ciudadana será clave en la fase empírica, a través de encuestas, grupos focales y experimentos sociales, apoyándose en plataformas especializadas para su implementación y análisis.

Recursos institucionales

- El proyecto se desarrollará en el marco del **Programa de Doctorado en Formación en la Sociedad del Conocimiento** de la Universidad de Salamanca (https://knowledgesociety.usal.es), que ofrece un enfoque interdisciplinar y acceso a espacios académicos y recursos compartidos.

- El **Grupo de Investigación Observatorio de los Contenidos Audiovisuales (OCA)** (https://www.ocausal.es/es/), con sede en la Facultad de Ciencias Sociales, proporcionará apoyo metodológico, acceso a infraestructura tecnológica, espacios de trabajo y acompañamiento especializado en áreas vinculadas a la comunicación y la cultura digital.

Recursos tecnológicos

- **Fuentes de información**: acceso a bases de datos bibliográficas especializadas mediante los servicios de la Universidad de Salamanca, así como recursos en acceso abierto.

- **Organización y tratamiento de bibliografía**: uso de gestores como Mendeley.

- **Herramientas para análisis de datos y procesamiento de lenguaje natural**: *SPSS* para el análisis estadístico, y *R Studio* y *Python* para análisis computacional y textual.

- **Plataformas para recolección de datos**: Qualtrics y servicios asociados a paneles online.

- **Generación de contenidos con IA**: uso controlado de herramientas como *ChatGPT, Gemini, Claude, DALL·E, MidJourney y Leonardo* para la creación de estímulos visuales y textuales.

- **Software de apoyo**: *Microsoft Office* y Aplicaciones de *Adobe* para diseño gráfico, edición y maquetación.

- **Entornos digitales y redes sociales**: se prevé el uso de redes sociales y herramientas asociadas para experimentos digitales, recolección de datos y validación participativa.

5. PLANIFICACIÓN TEMPORAL AJUSTADA A CUATRO AÑOS

Planificación Tesis Doctoral Germán Rodríguez-Wilches	2024	2025				2026				2027				2028		
	4T	1T	2T	3T	4T	1T	2T	3T	4T	1T	2T	3T	4T	1T	2T	3T
PRIMER AÑO (Marco teórico, Revisión de Literatura y Plan de Investigación)																
Planteamiento del tema de investigación																
Revisión exploratoria y bibliografía base																
Delimitación del tema y problema																
Justificación, objetivos e hipótesis																
Preguntas de investigación																
Marco teórico: selección y análisis crítico																
Construcción del plan de investigación																
Entrega del plan																
Estado del arte: sistematización y síntesis																
Diseño metodológico definitivo																
Diseño y estructuración FASE 1																
F1.1 – Grupos focales exploratorio																
Análisis F1.1 y ajustes diseño																
SEGUNDO AÑO (Estudio exploratorio y Estado del Arte)																
Capítulo Estado del Arte																
Envió de ponencia a Congreso																
F1.2 – Encuesta nacional																
Análisis F1.2																
Publicación artículo FASE 1																
Revisión Capítulo Estado del Arte																
Diseño y estructuración FASE 2																
TERCER AÑO (Estudio experimental, validación y publicaciones)																
Envió de ponencia a Congreso																
F2.1 – Grupos focales sesgos en imágenes																
Análisis F2.1 y ajustes diseño																
F2.2 – Estudio experimental																
Análisis F2.2																
Publicación artículo FASE 2																
Envió de ponencia a Congreso																
Diseño y estructuración FASE 3																
CUARTO AÑO (Redacción, cierre de tesis y defensa)																
Desarrollo F3.1 y F3.2																
Actualización del estado del arte																
Redacción final de tesis																
Revisión y presentación final																
Defensa de tesis doctoral																

6. PLAN DE FORMACIÓN PERSONAL

El plan de formación previsto en el marco de esta tesis doctoral se orienta al fortalecimiento de competencias metodológicas, teóricas y científicas específicas para el análisis de la percepción ciudadana de sesgos discriminatorios en inteligencia artificial generativa.

En el ámbito metodológico, se contempla la participación en cursos especializados de análisis cualitativo asistido por software, análisis estadístico aplicado en ciencias sociales y formación en diseño de experimentos sociológicos online. Este itinerario se complementará con la asistencia a las actividades del "Programa de formación transversal de la Escuela de Doctorado", que ofrecen actualización continua en metodologías de investigación, competencias digitales, comunicación científica y buenas prácticas académicas. Estas acciones permitirán profundizar en las herramientas requeridas para la implementación rigurosa de las fases cualitativas, cuantitativas y experimentales del estudio.

En el plano teórico, se prevé la actualización continua en temas vinculados a sesgos algorítmicos, percepción pública de tecnologías emergentes, alfabetización mediática y gobernanza ética de la inteligencia artificial. La formación se articulará mediante seminarios especializados, cursos de formación transversal ofertados por la Escuela de Doctorado, y actividades académicas relacionadas con los ejes centrales de la investigación.

La movilidad académica también constituye un componente esencial del plan de formación, mediante la colaboración con otros grupos de investigación nacionales o internacionales especializados en IA y sociedad, comunicación digital o ciencias sociales computacionales, contribuyendo así al enriquecimiento interdisciplinar del proyecto.

Finalmente, la participación en congresos nacionales e internacionales permitirá presentar avances de la investigación, intercambiar conocimientos y fortalecer las competencias de comunicación científica. Asimismo, se incentivará la publicación de resultados en revistas académicas de impacto, con el fin de contribuir a la difusión de los hallazgos y fortalecer el impacto científico y social de la tesis.

7. REFERENCIAS BIBLIOGRÁFICAS

AlDahoul, N., Rahwan, T., & Zaki, Y. (2025). AI-generated faces influence gender stereotypes and racial homogenization. *Scientific Reports*, *15*(1), 1-15. https://doi.org/https://doi.org/10.1038/s41598-025-99623-3

Alier Forment, M., Garcia Peñalvo, F., Casañ Guerrero, M. J., Pereira, J. A., & Llorens Largo, F. (2024, octubre 8). *Safe AI in Education Manifesto*. Version 0.4.0. https://manifesto.safeaieducation.org/

Arcila-Calderón, C., Igartua, J. J., Sánchez Holgado, P., Amores, J. J., Marcos Ramos, M., de Garay, B., Piñeiro Naval, V., Rodríguez Contreras, L., Blanco-Herrero, D., & Frías Vázquez, M. (2023a). *IA Spain 2023: Informa Público de «Percepción social de la Inteligencia Artificial en España»*. https://www.ocausal.es/investigacion/proyectos/percepcion-ia/percepcion-ia/

Arcila-Calderón, C., Igartua, J. J., Sánchez-Holgado, P., Amores, J. J., Blanco-Herrero, D., Gomes-Barbosa, M., Marcos Ramos, M., Piñeiro-Naval, V., & González de Garay, B. (2023b). *IA Ciudadana*. https://www.ocausal.es/investigacion/proyectos/ia/ia-ciudadana/

Borba, R. L., de Paula Ferreira, I. E., & Bertucci Ramos, P. H. (2024). Addressing discriminatory bias in artificial intelligence systems operated by companies: An analysis of end-user perspectives. *Technovation*, *138*, 103118. https://doi.org/10.1016/J.TECHNOVATION.2024.103118

Brauner, P., Hick, A., Philipsen, R., & Ziefle, M. (2023). What does the public think about artificial intelligence?—A criticality map to understand bias in the public perception of AI. *Frontiers in Computer Science*, *5*, 1113903. https://doi.org/10.3389/FCOMP.2023.1113903/BIBTEX

Casals Creus, L. (2024). Del dato al píxel: sobre los sesgos de género y raciales en la IA generativa Dall-E. *Premi Rosalind Franklin al millor Treball Final de Màster amb perspectiva de gènere*. https://diposit.ub.edu/dspace/handle/2445/219908

Cerezo-Martínez, P., Nicolás-Sánchez, A., & Castro-Toledo, F. J. (2024). Analyzing the European institutional response to ethical and regulatory challenges of artificial intelligence in addressing discriminatory bias. *Frontiers in Artificial Intelligence*, *7*, 1393259. https://doi.org/10.3389/FRAI.2024.1393259/BIBTEX

Chauhan, A., Anand, T., Jauhari, T., Shah, A., Singh, R., Rajaram, A., & Vanga, R. (2024, febrero 7). Identifying Race and Gender Bias in Stable Diffusion AI Image Generation. *2024 IEEE 3rd International Conference on AI in Cybersecurity (ICAIC)*. https://doi.org/10.1109/ICAIC60265.2024.10433840

Cheong, M., Robinson, P., Byrne, J., Ruppanner, L., Klein, C., Abedin, E., Ferreira, M., Reimann, R., Chalson, S., Alfano, M., Cheong, M., Byrne, J., Ruppanner, L., Ferreira, ; M, Reimann, R., Alfano, M., Chalson, S., Robinson, P., & Klein, C.

(2024). Investigating Gender and Racial Biases in DALL-E Mini Images. *ACM Journal on Responsible Computing*, *1*(2), 1-20. https://doi.org/10.1145/3649883

Clark, J. M., & Paivio, A. (1991). Dual coding theory and education. *Educational Psychology Review*, *3*(3), 149-210. https://doi.org/10.1007/BF01320076/METRICS

Creswell, J. W., & Plano Clark, V. L. (2017). *Designing and Conducting Mixed Methods Research* (3.ª ed.). SAGE Publications.

Shadish, W. R., Cook, T. D., & Campbell, D. T. (2002). *Experimental and quasi-experimental designs for generalized causal inference*. Houghton Mifflin Company.

Ferrara, E. (2024). Fairness and Bias in Artificial Intelligence: A Brief Survey of Sources, Impacts, and Mitigation Strategies. *Sci*, *6*(1), 3. https://doi.org/10.3390/SCI6010003

Garcia-Peñalvo, F. J., Alier, M., Pereira, J., & Casany, M. J. (2024). Safe, Transparent, and Ethical Artificial Intelligence: Keys to Quality Sustainable Education (SDG4). *IJERI: International Journal of Educational Research and Innovation*, *22*, 1-21. https://doi.org/10.46661/ijeri.11036

Hernández Sampieri, R., Feránadez Collado, C., & Baptista Lucio, M. D. P. (2014). *Metodología de la investigación* (6.ª ed.). McGraw Hill España.

Igartua Perosanz, J. J. (2006). *Métodos cuantitativos de investigación en comunicación* (1.ª ed.). Bosch.

INE. (2024). *España en cifras 2024*. http://publicacionesoficiales.boe.eshttp://www.ine.es/prodyser/espa_cifras

Kirk, H. R., Jun, Y., Iqbal, H., Benussi, E., Volpin, F., Dreyer, F. A., Shtedritski, A., & Asano, Y. M. (2021). Bias Out-of-the-Box: An Empirical Analysis of Intersectional Occupational Biases in Popular Generative Language Models. *Advances in Neural Information Processing Systems*, *34*, 2611-2624. https://doi.org/10.48550/arXiv.2102.04130

Krueger, R. A., & Casey, M. A. (2015). *Focus Groups: A Practical Guide For Applied Research* (5.ª ed.). SAGE Publications.

Medina Plasencia, F. de G. (2025). IA y privacidad: Armonización Normativa y Retos Regulatorios en la Unión Europea y España. *IDP. Revista de Internet, Derecho y Política*, *42*, 1-12. https://doi.org/10.7238/IDP.V0I42.432084

Perdomo Reyes, I. (2024). Injusticia epistémica y reproducción de sesgos de género en la inteligencia artificial. *Revista iberoamericana de ciencia tecnología y sociedad*, *19*(56), 89-100. https://doi.org/10.52712/ISSN.1850-0013-555

Rogers, E. M. (2003). *Diffusion of Innovations* (5.ª ed.). Free Press.

Sánchez-Holgado, P., & Arcila-Calderón, C. (2024). Adoption and use factors of artificial intelligence and big data by citizens. *Communication & Society*, *37*(2), 227-246. https://doi.org/10.15581/003.37.2.227-246

Sánchez-Holgado, P., Arcila-Calderón, C., & Blanco-Herrero, D. (2022). Conocimiento y actitudes de la ciudadanía española sobre el big data y la inteligencia artificial. *Revista ICONO 14. Revista científica de Comunicación y Tecnologías emergentes*, *20*(1), 2022. https://doi.org/10.7195/RI14.V21I1.1908

Sartori, L., & Bocca, G. (2023). Minding the gap(s): public perceptions of AI and socio-technical imaginaries. *AI and Society*, *38*(2), 443-458. https://doi.org/10.1007/S00146-022-01422-1/TABLES/9

UNESCO. (2022). *Recomendación sobre la ética de la inteligencia artificial*. https://unesdoc.unesco.org/ark:/48223/pf0000381137_spa

Vohland, K., Land-Zandstra, A., Ceccaroni, L., Lemmens, R., Perelló, J., Ponti, M., Samson, R., & Wagenknecht, K. (2021). *The Science of Citizen Science* (1.ª ed.). Springer Cham. https://doi.org/10.1007/978-3-030-58278-4

Young, M., Ehsan, U., Singh, R., Tafesse, E., Gilman, M., Harrington, C., & Metcalf, J. (2024). Participation versus scale: Tensions in the practical demands on participatory AI. *First Monday*, *29*(4). https://doi.org/10.5210/FM.V29I4.13642

Zhou, M., Abhishek, V., Derdenger, T., Kim, J., & Srinivasan, K. (2024). Bias in Generative AI. *ArXiv*, *abs/2403.02726*. https://arxiv.org/abs/2403.02726v1

VNiVERSiDAD D SALAMANCA