Gender Bias in Generative Artificial Intelligence Systems: A Systematic Literature Review

Rueda Ñopo, Melisa Isabel[1]; Ponte Córdova, Santos[2]; Rojas, Fanny Giovana[3]; Rufino Gabriel, Margarita Mercedes[4]; Lozano Mejía, Hilter Jamess[5];Chávez Ortiz, Patricia Gladys [6] ; Rosario Pacahuala, Emilio Augusto[7]

[1] Universidad Tecnológica del Perú, Lima, Perú, c18750@utp.edu.pe, [2 4 7] Universidad Privada del Norte, Lima, Perú, santos.ponte@upn.edu.pe, margarita.rufino@upn.edu.pe, emilio.rosario@upn.edu.pe [3] Universidad Científica del Sur, Lima, Perú, Fgabriel@cientifica.edu.pe,[5] Universidad Nacional Federico Villarreal, Lima, Perú, hlozanome@unfv.edu.pe, [6] Universidad Cesar Vallejo, Lima, Perú, pchavezort@ucvvirtual.edu.pe

Abstract– The objective of this systematic review was to analyze the available scientific evidence on gender bias in generative artificial intelligence systems, to understand how it manifests, its consequences, and strategies aimed at developing more inclusive systems. To this end, a systematic literature search was conducted in Scopus, applying the PICOC methodology for formulating the research question and the PRISMA guidelines for the study selection and analysis process. Initially, 106 articles were retrieved, which were subjected to inclusion and exclusion criteria, resulting in a final sample of 40 studies for analysis. The results show that generative artificial intelligence systems (both text-based and image-based) reproduce gender bias through various manifestations, such as stereotypes and traditional gender roles, underrepresentation and invisibility of women, and discrimination in evaluation and selection processes, among others. Furthermore, it was identified that these biases particularly affect women in STEM fields, thus reinforcing pre-existing educational and professional gaps. Mitigation strategies include diversifying training data, human oversight, and critical AI literacy. The study concludes that these biases continue to pose a challenge to developing more equitable and inclusive AI systems, particularly in a society where structural inequalities affecting women can amplify the impact of these technologies.

.

Keywords-- Artificial intelligence, gender bias, women in STEM, Chat GPT, DALL-E.

Sesgos de género en sistemas de inteligencia artificial generativa: una revisión sistemática de literatura

Resumen*– El objetivo de esta revisión sistemática fue analizar la evidencia científica disponible acerca de los sesgos de género en sistemas de inteligencia artificial generativa, conocer cómo se manifiesta, sus consecuencias y las estrategias orientadas al desarrollo de sistemas más inclusivos. Para ello, se realizó una búsqueda sistemática de literatura en Scopus, aplicando la metodología PICOC para la formulación de la pregunta de investigación y las directrices de PRISMA para el proceso de selección y análisis de estudio. Inicialmente eran 106 artículos, los cuales fueron sometidos a criterios de inclusión y exclusión, obteniéndose como muestra final 40 estudios para el análisis. Los resultados evidencian que los sistemas de inteligencia artificial generativa (escrita y de imagen) reproducen sesgos de géneros a través de diversas manifestaciones como estereotipos y roles tradicionales de género, subrepresentación e visibilización femenina, discriminación en procesos de evaluación y selección, entre otros. Asimismo, se identificó que estos sesgos afectan especialmente a las mujeres en ámbitos de STEM, reforzando así brechas educativas y profesionales preexistentes. Entre las estrategias de mitigación se destaca la diversificación de los datos de entrenamientos, la supervisión humana y la alfabetización crítica en IA. Se concluye que estos sesgos continúan representando un desafío para el desarrollo de sistemas de IA más equitativos e inclusivos sobre todo en una sociedad en donde las desigualdades estructurales que afectan a las mujeres pueden amplificar los efectos de estas tecnologías.

Palabras clave--Inteligencia artificial generativa, sesgos de género, mujeres en STEM, Chat GPT, DALL-E.

I. Introducción (Heading 1)

En el siglo XXI, el gran descubrimiento de la humanidad fue el uso de la inteligencia artificial (IA) en las diversas áreas de conocimiento, principalmente en la Ingeniería, que abrió la puerta a la Medicina, a la Educación, a la Administración, al Derecho y a otras disciplinas. Hoy se emplea la Inteligencia Artificial Generativa (IAG), para gestionar los procesos de manera más rápida y eficaz en la toma de decisiones, debido a que se ha ido mejorando de forma constante los modelos de lenguaje, dándoles no solo una función. El uso de herramientas de la IA como Chat GPT, Copilot, Gemini y otras, favorece el pensamiento crítico, la creatividad y la innovación en las distintas disciplinas de la Ingeniería, las cuales permiten la resolución de problemas complejos dentro del aprendizaje automático conocido como "machine learning" [1]. Actualmente, encontramos IAG como Chat GPT, Midjourney, DALL-E, Gemini, Claude, LLaMA, Deepseek, Wordtune, Adobe Firefly, Grok, Copilot, Bing Image Creator, Doubao, Kimi, Gans, etc. han generado diversos sesgos de género favoreciendo a representaciones masculinas invisibilizando a las mujeres en las áreas de Ciencia, Tecnología, Ingeniería y Matemática (STEM) debido a la falta a que la misma IAG genera sesgos algorítmicos, y sus patrones de automatización amenazan con profundizar las brechas laborales de género preexistentes. Por otro lado, también se observa sesgos étnicos que representan a mujeres de raza blanca [2].

El informe de la Unesco en el año 2024 reveló que los modelos de inteligencia artificial de código abierto más antiguos, como LLaMA 2 o GPT-2, exhiben sesgos de género sustancialmente más marcados y regresivos que los modelos comerciales cerrados [3]. Además, en el lenguaje masivo afirma que términos como "familia", "hogar", "niños" se asocian más a las mujeres, en cambio, vocablos como "carrera", "salario", "ejecutivo" se atribuye a los hombres, creando ciertos estereotipos dentro de la sociedad. Las investigaciones analizadas han provocado un impacto en las carreras de Ciencia, Tecnología, Ingeniería y Matemática (STEM) donde se visualiza que tienden a masculinizar las carreras según el perfil de género. Los organismos internacionales como la UNESCO y el Programa de las Naciones Unidas para el Desarrollo (PNUD) junto al Banco Mundial incentivan promover liderazgos femeninos. Asimismo, auditar a la IA en la emisión de sesgos externos en la educación superior buscando desarrollar la neutralización de los datos a través del entrenamiento sintácticamente del texto modificando sustantivos y pronombres de género al ejecutar el prompt [4].

Un estudio comparativo en Brasil y Estados Unidos evidencia la presencia de sesgos raciales y laborales asociados a la bibliotecología. Al solicitar imágenes sobre esta profesión a sistemas de inteligencia artificial, el 62,8 % de los resultados representó exclusivamente a personas blancas. Asimismo, la investigación destaca una marcada disparidad de género: aunque las mujeres constituyen el 82 % de la fuerza laboral en esta disciplina, perciben salarios inferiores a los de sus pares masculinos [5]

En experimentos realizados en Turquía, empleando las herramientas Chat GPT y DALL-E, cuando se les pidió dibujar a expertos en STEM, el 100 % de resultados fueron hombres. Sin embargo, cuando se pidió dibujar profesores de STEM, representó a mujeres dentro de un aula interactiva con estudiantes, lo cual muestra un sesgo de género. Podemos visualizar que estas herramientas representan al hombre con un conocimiento técnico y científico, mientras que a la mujer con las profesiones de servicio social [6].

En Reino Unido [7], se ha detectado que existe sesgo de género en el proceso de reclutamiento utilizando el chat GPT para evaluar los CV de los candidatos, el cual se asocia con nombres preferencialmente masculinos llegando a una calificación alta y sobresaliente reforzando estereotipos laborales existentes y como efecto trae contrataciones injustas.

En Estados Unidos (EEUU), se realizó un estudio que evalúa las herramientas Chat GPT, Midjourney, DALL-E y LLaMA, las cuales, al ser consultadas por imágenes predeterminadas, presentan un sesgo androcéntrico mostrando un 88% de hombres y solo 10% de mujeres. Lo mismo sucede con las

imágenes predeterminadas generadas por estas IA en el aspecto racial, pues favorecen a la raza blanca en un 86%, excluyendo casi por completo a las demás razas [8]. La falta de supervisión humana podría generar que se publique información errónea invisibilizando la diversidad de raza o género.

El sector salud no está ajeno al desarrollo de sesgos de género. Un estudio realizado por especialistas de obstetricia y ginecología de Brasil, Estados Unidos y Pakistán identifican un sesgo de género en los Modelos de Lenguaje Grande (LLM) al ser consultados por los especialistas de la salud. En la mayoría de casos representan al sector sanitario como hombres, mientras que a las mujeres se les asigna roles como obstetras o personal de primera línea [9]. Otra investigación, empleando GPT-4, señala que otro sesgo que puede desarrollarse es el estigma de la enfermedad mental, ya que los médicos son más proclives a diagnosticar a los hombres como enfermedades físicas y a las mujeres como enfermedades psicosociales [10].

Los modelos de lenguaje GPT-4 y DeepSeek-V3 exhiben sesgos de género en la asignación de responsabilidades parentales al replicar estereotipos tradicionales, en los que las madres asumen el rol principal de cuidadoras y los padres un rol secundario y de protectores. Esta discrepancia responde a las pautas de los datos de entrenamiento y no a una postura ideológica del propio modelo.[11]

Existe, también, una brecha de género en la industria tecnológica. Algunos datos obtenidos de una investigación realizada por la UNESCO, señalan que en las grandes empresas de machine learning, solo el 20% de empleos técnicos son ocupados por mujeres. Además, escasamente se permite un 12% de féminas en la investigación con IA y tan solo un 6% en desarrolladores de software [12].

El análisis de los sesgos de género en la inteligencia artificial generativa exige abordar dos dimensiones fundamentales: por un lado, su impacto en disciplinas como la Ingeniería, la Salud y la Educación; por otro lado, su delimitación geográfica, predominantemente enfocada en Estados Unidos, Europa y Asia. En contraste, esta problemática permanece escasamente explorada en América Latina y el Caribe, lo que invisibiliza las brechas de género locales y la perpetuación de arquetipos discriminatorios producidos por la IAG en la región. En este sentido, el objetivo de la investigación es determinar los sesgos de género en sistemas de inteligencia artificial generativa identificando sus manifestaciones, consecuencias y estrategias de solución en el campo STEM.

II. Metodología

Con respecto al diseño, la presente revisión sistemática de la literatura tuvo como finalidad identificar, evaluar y sintetizar la evidencia científica sobre los sesgos de género asociado con las herramientas de la inteligencia artificial generativa basadas en grandes modelos de lenguaje (LLM). Así también, la investigación se edificó en base a las directrices Preferred Reporting Items for Systematic Reviews and Meta Analyses (PRISMA), para con ello, asegurar la rigurosidad, transparencia y trazabilidad de la selección de artículos que inicia desde la identificación luego el cribado, continuando con la elegibilidad y finalmente la inclusión de los artículos para analizar. Con los artículos incluidos se realizó primero un análisis de datos según la distribución por años, por filiación del primer autor, artículos más citados, coocurrencia de palabras clave y distribución de estudios por cuartil. Luego, se continuó con el análisis cualitativo donde se identificaron y analizaron categorías emergentes de la información en base a las preguntas de investigación formuladas en el trabajo.

En la Tabla 1, se muestran las preguntas de investigación formuladas mediante el modelo PICOT. donde la (P) se refiere al problema, es decir a los tipos de sesgos, discriminación, estereotipos de género que se identificó en la información. La (I) corresponde a la intervención, el cual trata de las herramientas de IA generativa que han sido evaluadas en las investigaciones incluidas en el trabajo. También se trabajó el componente (C) que trata la comparación o diferencias que se observan en los datos provenientes de la inteligencia artificial generativa (Generative Artificial Intelligence, GenAI) sobre los hombres y mujeres. Y finalmente la (O) que se pregunta por los resultados como consecuencias y estrategias para mitigar los sesgos de género en la GenAI. El tiempo (T) se filtró mediante la base de datos Scopus delimitada en 5 años de antigüedad, y la (C) de contexto determinó en base al ámbito femenino.

Así también, las palabras claves se determinaron usando la lógica del PICOT. Para ampliar la extracción de artículos, se empleó los términos en inglés.

TABLA I
COMPOSICIÓN DE LA PREGUNTA PICOC

PICOTC	PREGUNTA	PALABRAS CLAVES
P	¿Qué tipos de sesgos, discriminación, estereotipos o desigualdades de género identifica la literatura científica?	"gender bias" OR "gender biases" OR "gender stereotype*" OR "gender discrimination" OR "sexism" OR "gender inequality" OR "algorithmic bias" OR "AI bias" OR "gender representation"
I	¿Qué herramientas de IA generativa han sido evaluadas respecto a sesgos de género?	"generative AI" OR "generative artificial intelligence" OR "GenAI" OR "generative models" OR "large language models" OR "LLM" OR "ChatGPT" OR "AI-generated content"
C	¿Qué diferencias se observan entre hombres y mujeres respecto a la presencia de sesgos en la IA generativa?	"gender differences" OR "gender comparison" OR "male-female differences" OR "gender equality" OR "gender equity" OR "gender fairness" OR "gender parity"
O	¿Qué consecuencias generan los sesgos de género y qué estrategias se proponen para mitigarlos?	"gender discrimination" OR "gender inequality" OR "gender stereotypes" OR "gender disparities" OR "gender gap"OR "bias mitigation" OR "bias reduction" OR "bias

		detection" OR "algorithmic fairness" OR "debiasing"
C	¿En qué contextos educativos, académicos y profesionales relacionados con mujeres, estudiantes mujeres y mujeres en STEM se han identificado sesgos y discriminación de género asociados con la IA?	"women" OR "woman" OR "girls" OR "female" OR "female students" OR "women in engineering" OR "women in STEM"

La búsqueda de los artículos se realizó en base a la siguiente interrogante general: ¿Qué evidencia científica existe sobre los sesgos, discriminación, estereotipos o desigualdades de género asociado con las herramientas de la inteligencia artificial generativa basadas en grandes modelos de lenguaje (LLM) considerando las diferencias, consecuencias y estrategias de mitigación entre los años 2023 al 2026?

Para realizar la búsqueda se utilizó las palabras claves donde los términos correspondientes a un mismo componente o grupo conceptual se relacionaron mediante el operador booleano OR, mientras que los principales bloques conceptuales se enlazaron con AND en base al siguiente modelo: ((P) AND (I) AND (C) AND(O)). La exploración bibliográfica se realizó en la base de datos Scopus, el factor que llevó a su elección fue por su nivel de impacto en la comunidad científica.

En la tabla 2, se encuentra la ecuación de búsqueda utilizada, donde se resalta las palabras clave, los operadores booleanos, además también se aplicaron los filtros: years(2023 a 2026), document type (article), source type(journal), publication stage (final) y open access(all open access). Los componentes utilizados en la búsqueda fueron (P): "gender bias" OR "gender biases" OR "gender stereotype*" OR "gender discrimination" OR "sexism" OR "gender inequality" OR "algorithmic bias" OR "AI bias" OR "gender representation", (I): "generative AI" OR "generative artificial intelligence" OR "GenAI" OR "generative models" OR "large language models" OR "LLM" OR "ChatGPT" OR "AI-generated content"(C)"women" OR "woman" OR "girls" OR "female" OR "female students" OR "women in engineering" OR "women in STEM". Tanto los filtros como los términos clave fueron aplicados de manera progresiva para obtener la mayor cantidad de evidencia relevante en la identificación de investigaciones para luego poder aplicar los criterios de selección.

TABLA II
ECUACIÓN DE BÚSQUEDA

BASE DE DATOS	ECUACIÓN
SCOPUS	(TITLE-ABS-KEY("WOMEN" OR "WOMAN" OR "GIRLS" OR "FEMALE" OR "FEMALE STUDENTS" OR "WOMEN IN ENGINEERING" OR "WOMEN IN STEM") AND TITLE-ABS-KEY("GENERATIVE AI" OR "GENERATIVE ARTIFICIAL INTELLIGENCE" OR "GENAI" OR "GENERATIVE MODELS" OR "LARGE LANGUAGE MODELS" OR "LLM" OR "CHATGPT" OR "AI-GENERATED CONTENT") AND TITLE-ABS-KEY("GENDER BIAS" OR "GENDER BIASES" OR "GENDER STEREOTYPE*" OR "GENDER DISCRIMINATION" OR "SEXISM" OR "GENDER INEQUALITY" OR "ALGORITHMIC BIAS" OR "AI BIAS" OR "GENDER REPRESENTATION")) AND (LIMIT-TO (DOCTYPE,"AR")) AND (LIMIT-TO (SRCTYPE,"J")) AND (LIMIT-TO (PUBSTAGE,"FINAL")) AND (LIMIT-TO (OA,"ALL")) AND PUBYEAR > 2022 AND PUBYEAR < 2027

Continuando la investigación, como se describe en la Tabla 3, se planteó los criterios de inclusión y exclusión que sirve para definir que investigaciones cumplen con el objetivo de la revisión sistemática de la literatura. Estos se aplicaron de manera secuencial durante la evaluación de elegibilidad en el PRIMA.

TABLA III
CRITERIOS DE INCLUSION Y EXCLUSION

COMPONENTE	CI	CRITERIOS DE INCLUSIÓN	CE	CRITERIOS DE EXCLUSIÓN
P (Población)	CI1	Estudios relacionados con los sesgos de género, discriminación, estereotipos, desigual en sistemas de IA generativa o herramientas basadas en LLM	CE1	Estudios que no tratan la dimensión de sesgos de género, discriminación, estereotipos, desigual en sistemas de IA generativa
I (Intervención)	CI2	Estudios que evalúen la IA generativa, LLM	CE2	Estudios centrados en machine learning o en IA no generativa
C (Comparación)	CI3	Estudios que comparen datos sobre género, sesgo, discriminación o estereotipos generados en IA generativa	CE3	Estudios que no realizan ningún análisis relacionado con las diferencias de género, sesgos, discriminación o estereotipos
O (Resultado)	CI4	Estudios que expliquen consecuencias, así también como estrategias de mitigación	CE4	Estudios que solo evalúan rendimiento técnico de la IA y no analiza efectos ni estrategias
C (Contexto)	CI5	Estudios educativos, académicos y profesionales relacionados con mujeres, estudiantes mujeres y mujeres en STEM identificando sesgos y discriminación de género asociados con la IA generativa	CE5	Analizan otros contextos asociados con o sin la IA generativa

Como se refleja en la Fig. 1, el proceso de selección se desarrolló mediante 4 fases del método PRISMA: identificación, cribado, elegibilidad e inclusión. En la fase identificación con la fórmula de búsqueda y filtros correspondientes se recuperó 106 investigaciones. Luego, en el

cribado donde se revisó los artículos recuperados considerando el abstract, título, palabras clave en correspondencia con los criterios de inclusión como resultado se obtuvieron 60 artículos pertinentes, excluyéndose 46. Posteriormente, las 60 investigaciones a texto completo fueron analizadas en la fase de elegibilidad aplicando meticulosamente los criterios de inclusión y exclusión. En esta etapa se excluyeron 20 artículos. De este total, 10 investigaciones fueron excluidas por el CE1, 1 por CE2, 6 por CE3, 2 por CE4 y 1 por CE5. Finalmente, se incluyeron 40 artículos en la RSL.

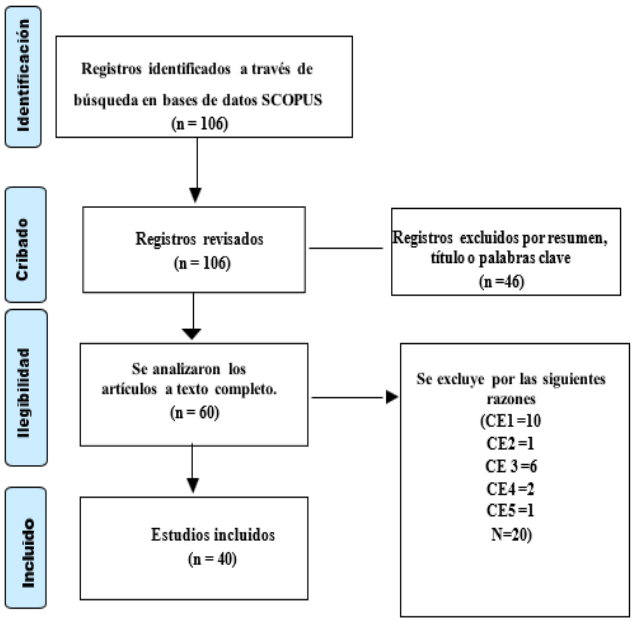


FIG. 1 MÉTODO PRISMA

III. RESULTADOS

A. *Organización de la información*

En base a la fig. 2. La distribución temporal demuestra una ascendencia reciente y acelerada de la producción científica sobre los sesgos de género en los sistemas de inteligencia artificial generativa. Del total de 40 investigaciones analizadas, 15 fueron publicadas en el 2025 (37,5%), siendo el año con mayor producción. Le sigue 2026 hasta el mes de julio, con 13 artículos (32,5%), mientras que 2024 concentra 9 publicaciones (22,5%). Por el contrario, la producción correspondiente al 2023 fue menor, con 3 artículos (7,5%).

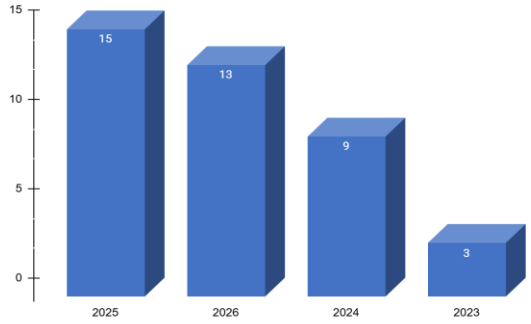


Fig 2. Distribución por años

B. *País de afiliación del primer autor*

Teniendo en cuenta la Tabla 4, de las 40 investigaciones analizadas, según la filiación del primer autor, se presenta una mayor concentración de investigaciones en EEUU en una cantidad de 12 artículos (30,0%). Le sigue España con 8 artículos (20,0%) y Australia con 5 (12,5%). Así también, Reino Unido con 4 artículos (10%), mientras que China tenía 3 (7,5%). Turquía e Italia ambos con 2 artículos cada uno (5,0%) y Austria, Suiza, Noruega y Corea del Sur representan un artículo cada uno (2,5%). Esto evidencia carencia de investigaciones sobre este tema en Latinoamérica.

TABLA IV
PAÍS DE AFILIACIÓN

País	N.º de artículos	%
Estados Unidos	12	30,0 %
España	8	20,0 %
Australia	5	12,5 %
Reino Unido	4	10,0 %
China	3	7,5 %
Turquía	2	5,0 %
Italia	2	5,0 %
Austria	1	2,5 %
Suiza	1	2,5 %
Noruega	1	2,5 %
Corea del Sur	1	2,5 %
Total	40	100 %

C. *Artículos más citados*

La Tabla 5, evidencia 5 artículos más citados y de mayor impacto de los 40 seleccionados en la revisión sistemática de la literatura. El artículo con más citas fue "Large language models show human-like content biases in transmission chain experiments" con 125 citas. Seguido por, "Smiling women pitching down: auditing representational and presentational gender biases in image-generative AI" con 108 citas. Luego está, "How AI can perpetuate – Or help mitigate – Gender bias in leadership" con 47 citas. También se presenta "Gender bias in generative artificial intelligence text-to-image depiction of medical students" con 29 citas. Finalmente, se destaca "Is generative AI reshaping academic practices worldwide? A survey of adoption, benefits, and concerns" con 26 citas.

TABLA V
ARTÍCULOS MÁS CITADOS

Título	N° citas
Is generative AI reshaping academic practices worldwide? A survey of adoption, benefits, and concerns	26
Gender bias in generative artificial intelligence text-to-image depiction of medical students	29
How AI can perpetuate – Or help mitigate – Gender bias in leadership	47
Smiling women pitching down: auditing representational and presentational gender biases in image-generative AI	108

Large language models show human-like content biases in transmission chain experiments	125

D. Mapa Bibliométrico de coocurrencia de palabras clave

La Fig.3 sobre las palabras clave fue generada en VOSviewer con 139 keywords diferentes de las 40 investigaciones incluidas. En ella se muestra una estructura temática considerablemente articulada alrededor del término inteligencia artificial (artificial intelligence), el cual representa el principal nodo de la red. Alrededor de esta definición se conectan generative AI (Nodo central secundario), large language models (Nodo importante), gender (Nodo central), bias (Nodo intermedio), fairness (Nodo emergente), race (Nodo conectado con género y sesgo), historical inequity (Nodo relativamente reciente), chat gpt (Nodo relevante), implicit bias (Nodo periférico), DALL-E3 (Nodo reciente), AI in higher education(Nodo reciente), Academic guidance (Nodo reciente), Claude 3.5 (Nodo reciente) y chatbots (Nodo superior).

Analizando las palabras clave se demuestra que las investigaciones iniciaron en base a la temática de la IA y chatbot pero luego se fue desplazando a temas de género y discriminación en la IA estableciéndose la siguiente cadena:

Inteligencia artificial-IA generativa-modelos de lenguaje-generación de contenidos-representación social-sesgos-género-equidad.

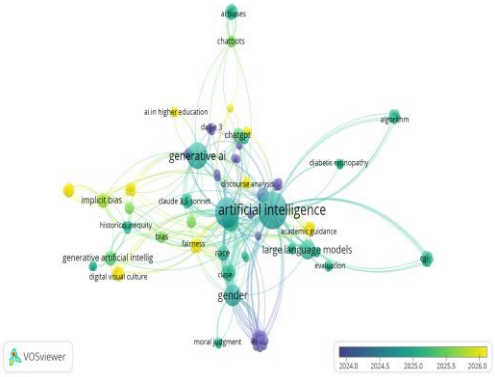


Fig 3. Nubes de palabras en VOSviewer

E. Distribución de estudios según cuartil

En la Fig. 4, se muestran los estudios según el cuartil de las revistas donde una mayor concentración de las investigaciones son fuentes científicas de alta visibilidad. De los 40 artículos incluidos en la RSL, 27 (67,5%) fueron publicadas en revistas clasificadas en Q1, así también, 9 (22, 5%) son de Q2 y 4 (10%) fueron publicados en Q3. Demostrando calidad y rigor en las investigaciones analizadas.

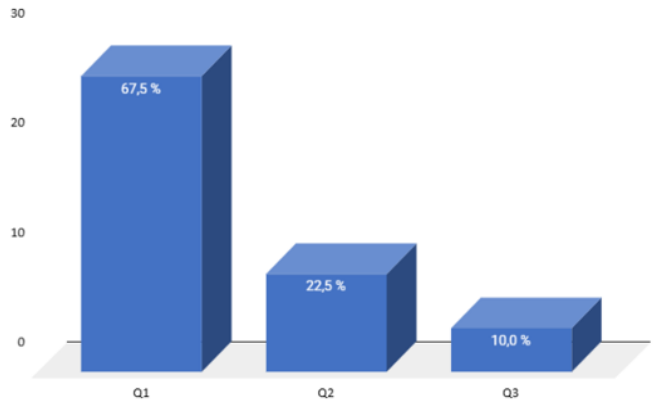


Fig 4. Porcentajes de distribución por cuartil

F. *Población: ¿Qué tipos de sesgos, discriminación, estereotipos o desigualdades de género identifica la literatura científica?*

Como se observa en la Fig. 5. La problemática o población se categoriza en 5 criterios que se extraen exclusivamente de las 40 investigaciones seleccionadas en el PRISMA. En una primera categoría (C1), 16 (40%) investigaciones resaltan la reproducción algorítmica de estereotipos de género, roles tradicionales y segregación ocupacional, donde se identificaron asociaciones continuas entre mujeres y roles de cuidado, familia, servicios y labores domésticas. En contraste con los hombres que lo relacionan con el liderazgo, tecnología, ingeniería, finanzas, profesiones técnicas, puestos ejecutivos. [13]; [6]; 14]; [11]; [15]; [16]; [17]; [18]; [19]; [20]; [21]; [22]; [4]; [3]; [23]. Así también, en una segunda categoría (C2) 7 (17,5%) investigaciones tratan sobre la subrepresentación, visibilización en determinadas profesiones y diversidad de género. Es decir, se reporta una baja participación femenina en las ingenierías y se resalta a la representación masculina por defecto. [12]. [24]; [9]; [2]; [25]; [26]; [27]. En una tercera categoría (C3) el problema que emerge en 9 (22,5%) investigaciones es la discriminación y sesgos en la evaluación, selección, contratación, liderazgo, resolución de problemas y toma de decisiones profesionales y académicas. [28]; [10]; [7]; [29]; [30]; [31]; [32]; [33]; [34]. También, una cuarta categoría (C4) correspondiente a 2 (5.0 %) investigaciones se observan sesgos de capacidades (agencia), personalidad, sexualidad y comportamiento, así también como valoración moral entre hombres y mujeres. [35]; [36]. Finalmente, en una quinta categoría (C5), 6 (15.0%) artículos muestran otros sesgos, brechas y manifestaciones relacionadas con el género como cuestiones raciales y étnicos, avatares o influencer virtuales. [37]; [8]; [5]; [38]; [39]; [40]; [41].

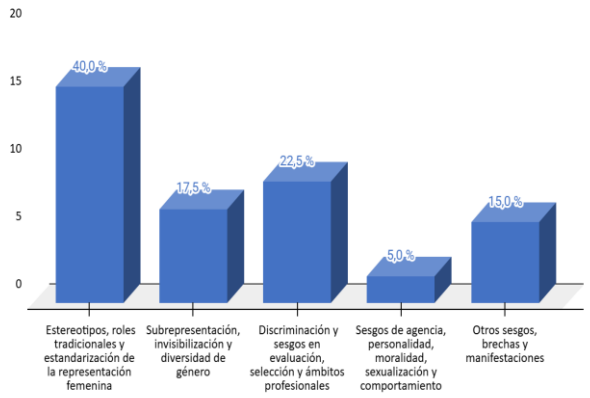


Fig 5. Categorización

G. Intervención: ¿Qué herramientas de IA generativa O LLM han sido evaluadas respecto a sesgos de género?

En la Tabla 5, se observa el análisis de las herramientas de la IA generativa y MML evaluadas en las 40 investigaciones seleccionadas teniendo en cuenta las categorías emergentes.

TABLA VI
IAGen/MML POR CATEGORÍA

C	IAgen/MML	Investigaciones
C1	GPT-4; DALL-E 3; Chat GPT; DALL-E; DALL-E 2; Bing Image Creator; Chat GPT-3; SiktChat (GPT-4o); GPT-4.1; DeepSeek-V3; BART-large-MNLI; DeBERTa-v3-base-mnli-fever-anli; scandi-nli-large; GPT-4o; Gemini 1.5 Pro; Claude 3.5 Sonnet; LLaMA 3.1:8b; Chat GPT o1; Claude 3.7 Sonnet; Gemini 2.0 Flash; Chat GPT-4; Gemini; Llama 2; Llama 3; Midjourney; Copilot; LLM	[13]; [6];[14]; [16]; [17]; [11]; [15]; [18]; [19]; [20]; [21]; [22]; [4]; [3]; [23]; [40]
C2	Chat GPT; Gemini; DeepSeek; o3-mini; DeepSeek-R1; GPT-3.5; GPT-4; Meditron-70B; DALL-E 3; DALL-E 2; Midjourney; DALL-E; Microsoft; Claude	[12]; [24]; [9]; [2]; [25]; [26]; [27]
C3	Chat GPT; GPT-4; Wordplay.ai; Chat GPT-4o; ChatGPT 3.5; Claude	[28]; [10]; [7]; [29];[30]; [31]; [32]; [33]; [34].
C4	Sistemas de IA y LLM	[35]; [36].
C5	ChatGPT; Google Gemini; Wordtune; ChatGPT-4.0; Midjourney V.6; DALL-E 3; GPT-2; Llama 2; Llama 3; Adobe Firefly; Claude; Grok; Qwen-Plus; DeepSeek; Doubao; Kimi 2; DALL-E	[37]; [8]; [5]; [38]; [39]; [41]

En la Fig. 6 se encontró que Chat Gpt es la herramienta más empleada, apareciendo en 21 (52,5%) de los 40 artículos, luego DALL-E se lo evaluó en 10 artículos (25%). Así también 9 investigaciones revisaron GPT (22,5%), 7 investigaciones mencionaron Gemini (17,5%), Llama, Deep Seek, Claude fueron analizadas en 5 artículos (12,5%), Midlourney con 4 artículos (10%) los demás fueron evaluados una sola vez. Se resalta que las mismas herramientas se analizan en varios artículos.

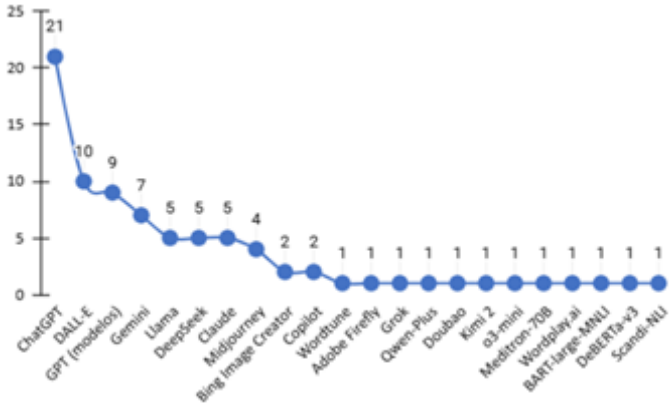
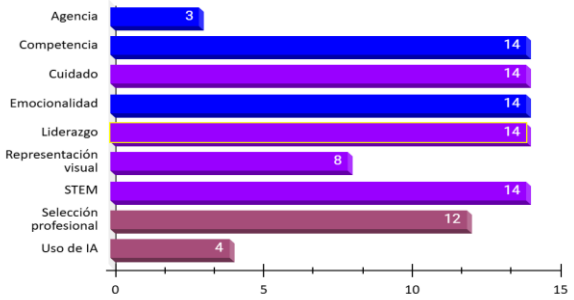


Fig 6. Herramientas evaluadas

H. Comparación ¿Qué diferencias reporta la literatura entre hombres y mujeres, o entre distintos modelos de GenAI, respecto a la presencia de sesgos?

Como se evidencia en la Fig. 7, las principales diferencias encontradas en las 40 investigaciones incluidas se sintetizan en tres bloques: a) Representación, es la que tiene mayor tendencia y significa "quien aparece y con qué frecuencia" esto se da en especial en profesiones que tradicionalmente han presentado desigualdad de género, generando mayor o menor asociación con la figura masculina y femenina, para el liderazgo, representación visual y la formación STEM hay mayor asociación con la figura masculina. Y el cuidado tiene mayor representación con la mujer. También se encontró la b) Atribución donde se establecen características que se atribuyen a cada género relacionando a los hombres la competencia, la agencia (capacidad) y a la mujer con la emocionalidad, la maternidad, la calidez. Por último, el desempeño (12), este afecta particularmente a las decisiones para la selección profesional y contratación hasta en el uso de la IA que se asocia más con la imagen masculina.



a

Fig. 7: Diferencias

I. Resultados ¿Qué consecuencias generan los sesgos de género en la educación superior y qué estrategias se proponen para mitigarlos?

Después de analizar en los apartados superiores los sesgos, brechas, discriminación, estereotipos que se asocian con la inteligencia artificial generativa es necesario establecer las posibles consecuencias que estos generarían según las 40 investigaciones analizadas en la RSL. Con respecto a C1, es posible que se intensifiquen los estereotipos, que se asocie las profesiones y roles con un género, que se refuercen los roles tradicionales sesgados, desincentivo de vocaciones STEM en mujeres.([37]; [8]; [12]; [13]; [6]; [14]; [17]; [11]; [15]; [20]; [21]; [22]; [4]; [3]; [23]). En C2, se tendría como efectos la invisibilidad de mujeres, exclusión simbólica de la mujer, la reproducción de desigualdades, mayor afectación a la inclusión, disminución de la identificación de la mujer con determinadas carreras. [8]; [24]; [9]; [2]; [25]; [26]; [19]; [27]; [18]; [34]. En C3, se observa la desigualdad en contratación y selección; decisiones injustas; afectación de oportunidades académicas y profesionales; barreras acumulativas para investigadoras; inequidad en revisión por pares; afectación del ascenso profesional; reproducción de desigualdades en carreras STEM. [28]; [38]; [10]; [7]; [29]; [30]; [32]; [33]; [36].
Para C4, se reproduce la injusticia relacionada con la atribución diferenciada de agencia; tipificación moral; sexualización y objetivación; violencia simbólica; asignación diferenciada de culpa, limitación de la expresión diversa de género. [28]; [13]; [18]; [4]; [23]; [39]; [35]. Finalmente, en C5 se profundiza la desventaja competitiva para investigadoras; discriminación interseccional; amplificación de desigualdades estructurales; problemas de acceso a información; brechas educativas y profesionales. [28]; [37]; [8]; [5]; [31]; [27]; [41].

De las 40 investigaciones 35 estudios contienen propuestas. En 23 (65,7%) artículos de los 35 resaltan la auditoría, evaluación y monitoreo, para detectar, medir y realizar seguimiento a los sesgos de género. Así también, 22 (62,9 %), estudios consideran que el uso de datos diversos y representativos, es decir equitativos y equilibrados en el entrenamiento de la IAgen va a reducir los sesgos. 14 (40%) estudios proponen la Mitigación técnica, aquí se aplica el proceso de fine-tuning o reentrenamiento, para realizar la corrección de asociaciones algorítmicas o utilizar mecanismos automáticos para detectar las brechas. 12 (34,3%) artículos, se refieren a las instrucciones proporcionadas mediante prompts (neutrales) para reducir respuestas estereotipadas.

No obstante, 11 (31,4%) estudios consideran que la solución está en la formación y alfabetización crítica de las personas que utilizan e interpretan la IAgen. De tal manera que tengan la competencia de reconocer, cuestionar y evaluar los sesgos. En la misma línea, 10 (28,6%) estudios consideran que debe mejorarse el diseño volviéndolo inclusivo y ético desde las etapas iniciales incorporando datos sobre mujeres científicas, mujeres ingenieras para tener respuestas más equilibradas. 9 artículos (25,7) proponen la Human- in-the- loop que se centra en la revisión crítica. Finalmente 7 (20%) estudios mencionan la transparencia y gobernanza para informar, reglamentar, supervisar y asignar responsabilidad y protocolos del manejo de los datos.

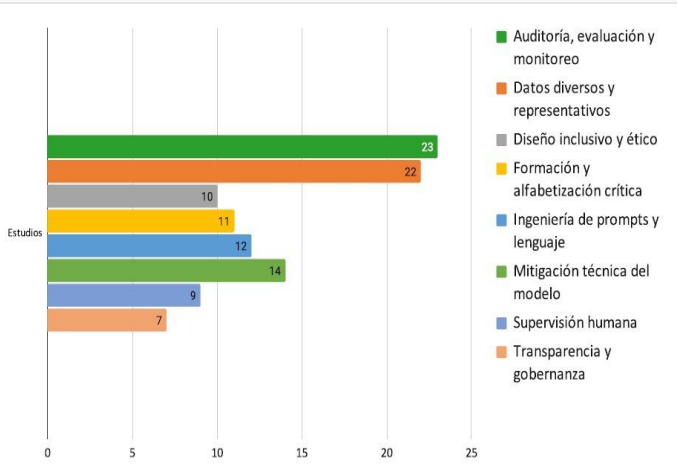


Fig. 8 Mitigación de sesgos de género en la GenAI

IV. Discusión

Por medio de este trabajo se ha analizado la evidencia científica que nos ha permitido entender que los sistemas de inteligencia artificial generativa continúan reproduciendo sesgos de géneros presentes en los datos históricos con los que fueron entrenados, a pesar de que la realidad actual los difiere. Esta tendencia es una problemática nacida de una limitación estructural de la IAG más que una característica aislada de una herramienta específica.
La predominancia de sesgos dirigidos hacia mujeres y niñas sugiere que los sistemas de IA no solo reproducen información histórica, sino que también amplifican patrones de desigualdad presentes en las estructuras sociales que alimentaron sus procesos de entrenamiento. Este hallazgo resulta principalmente relevante en las disciplinas STEM, donde la persistencia de representaciones masculinizadas puede reforzar barreras simbólicas ya existentes y limitar las expectativas de participación de futuras generaciones de mujeres. Autores [6], [13], [22], [8], [5] y [14] identificaron que las féminas son representadas de forma recurrente con atributos asociados al cuidado, la sensibilidad, la belleza o la dependencia emocional, mientras que los hombres aparecen vinculados al liderazgo, la razón, la competencia técnica y la autoridad. Esto podría generar una distorsión de la realidad que impacta de manera directa en una percepción negativa acerca de las capacidades de la mujer dentro de las disciplinas STEM. Mas esto no solo las afecta a ellas sino también a las poblaciones con identidades interseccionales como minorías étnicas, culturales o sexuales experimentan invisibilización y exclusión simbólica ([41], [27], [19] y [34])
Los modelos de inteligencia artificial generativa en donde se evidenciado esta problemática son dos los de lenguaje (Chat GPT, GPT-4, Gemini, Claude, LLaMA y DeepSeek) y los

generadores de imágenes (DALL-E 3, Midjourney y herramientas similares). Esto permite demostrar que los sesgos aparecen de manera consistente independiente de la arquitectura utilizada, aunque con diferencias entre la magnitud de los modelos [24]. Se encontraron variaciones de representación entre DeepSeek-R1y o3-mini, mientras que [19] reportaron que GPT-4 presentaba una distribución de género relativamente más equilibrada que Gemini, Claude, LLaMA, pero que ninguno logró eliminar los sesgos completamente. Aunque los sesgos se observan en todos los modelos analizados, su intensidad varía entre sistemas, lo que evidencia avances parciales en mitigación, pero no la eliminación de sus causas subyacentes. La mayor tendencia a reproducir representaciones masculinas en profesiones STEM se da en los sistemas de generación de imagen. [6] evidenció que Chat GPT y DALL-E representan en su totalidad de expertos STEM a varones, mientras que [2] reportaron una sobrerrepresentación masculina en estudiantes de medicina y personal paramédico. Sin contar que [19], [27] y [34] demostraron que en la traducción de texto a imagen se intensifica la hipersexualiación e infantilización de la mujer, al mismo tiempo que se hipermasculiniza las representaciones corporales de autoridad. Estos hallazgos indican que la IA puede reforzar patrones de representación desequilibrados a pesar que las instrucciones sean neutrales. Esto es confirmado por [40], [2] y [26]. Pues identifican que algunos modelos utilizan la masculinidad como una categoría predeterminada, cuando el prompt no indica el género.

Entonces se puede afirmar que el hallazgo más consistente de esta revisión corresponde a las diferencias observadas entre las representaciones masculinas y femeninas generadas por la IA. Siendo la categoría más relevante la de estereotipos y roles tradicionales de género. Estudios de [8], [12], [13], [11], [22] y [35] mostraron que la figura masculina es presentada predominantemente como líder, innovador, científico, ingeniero y expertos técnicos. Mientras que las mujeres aparecen asociadas a roles de cuidado, apoyo emocional, docencia, maternidad y actividades domésticas. Paralelamente los estudios de [17], [11], [18], [22] y [35] coinciden que la IAG incurre en una homogeneización estética e ideológica del género feminino restringiendo así la diversidad de experiencias y reduciendo a la mujer a cánones estereotipos tradicionales. Esto debido a que los generadores visuales objetivan y sexualizan los cuerpos femeninos en proporción muy superiores a los masculinos, despojándolas así de agencia profesional y moral. Otra categoría es la subrepresentación e invisibilización dirigida a mujeres ([24]; [9]) a las identidades no binarias y diversas ([2]; [26]; [19]) esto debido a que se las IAG promueven una visión monocular masculina ([27]; [34]). La discriminación y la desigualdad profesional es otra categoría en la que se evidencia sobre todo en el ámbito de evaluación y selección profesional. [30] encontraron que los hombres tienen una mayor probabilidad de ser recomendados en especialidades quirúrgicas de alta competitividad. Además, que según [29] señala que los hombres en su mayoría son presentados como buenos líderes mientras que una proporción importante de mujeres aparece asociada a un liderazgo deficiente. También [7] observaron que los atributos masculinos favorecen determinadas decisiones de selección. Además, que, ante perfiles idénticos, los sistemas evalúan de manera más favorable a los hombres para puestos directivos o técnicos pues penaliza a las mujeres por brechas temporales de maternidad ([10]; [7]; [29]; [17]; [15]; [30]; [32])

También se ha hallado una doble moral donde la asertividad masculina se valora como visión estratégica mientras que estos comportamientos en mujeres se tachan de agresividad o de falta de empatía ([29]; [4]; [23]) Estos resultados indican que la IAG no solo reproduce diferencias simbólicas sino también desigualdades potencialmente relevantes para la toma de decisiones.

Esta situación agranda la lucha por la igualdad de género pues está generando consecuencias educativas, profesionales, sociales y culturales reproducidas por la IAG. Ante la perpetuación de estereotipos de género, los estudios de [37], [6], [14], [17], [11], [15], y [3] sugieren que estas representaciones influyen en las aspiraciones profesionales y académicas de las mujeres y niñas en el área STEM. Además, ante la invisibilización femenina se limita la disponibilidad de referentes y modelos de identificación, generando una reducción de pertenencia a las mujeres en determinados espacios profesionales. ([2], [24], [26] y [19]) Esto también afecta a la igualdad de oportunidades en los procesos de contratación, promoción profesional, evaluación académica y selección de candidatos generando así la discriminación por género y perpetuando la brecha salarial. ([38], [30], [33], [32] y [36]). Asimismo, los hallazgos de [23], [2], [33], [36]., [18], y [41] documentan que la propagación de imágenes objetivizadas y de violencia simbólica exacerba la discriminación interseccional al cruzar el género con raza, etnia u orientación sexual. Estos hallazgos sugieren que los sesgos de la IA no solo afectan a las representaciones simbólicas, sino que también pueden influir en decisiones de selección, promoción y evaluación profesional.

Asimismo, los estudios analizados evidencian diferencias contextuales relevantes. Mientras algunas investigaciones son desarrolladas en ámbitos académicos enfatizan los efectos de los sesgos sobre las aspiraciones educativas y vocacionales. En aquellos que fueron centrados en escenarios laborales destacan sus consecuencias sobre los procesos de contratación, evaluación y promoción profesional. Esto sugiere que los efectos de la IAG pueden manifestarse de manera distinta según el contexto de aplicación, pero mantiene la reproducción de la desigualdad de género. Además, la literatura estudiada proviene del norte global (Europa, Asia y EE.UU) generando así un vacío acerca de la situación de América Latina y esto es alarmante pues los modelos entrenados bajo hegemonía anglosajona ignoran las realidades y lenguas locales e imponen estereotipos ajenos que exigen una evaluación empírica regional.

Entonces es necesario desarrollar sistemas más inclusivos y responsables para evitar estos sesgos. Entre las principales recomendaciones destacan la diversificación de prompts, curación de los datos de entrenamiento, auditorías periódicas, supervisión humana, mecanismo de transparencia,

alfabetización crítica en IA y formación de usuarios y desarrolladores. Asimismo, autores como [37], [31], [14] y [19] enfatizan la importancia de enfoques interseccionales y equipos de desarrollo diversos para reducir la reproducción sistemática de desigualdades.

Esta revisión presenta algunas limitaciones relacionadas con la selección de estudios en una determinada base de dato, el predominio de publicaciones en inglés y la exclusión de literatura gris. Además, la rápida evolución de la inteligencia artificial generativa exige una actualización constante de la evidencia. También hay una ausencia de estudios identificados en América Latina, evidenciando un vacío de investigación que limita la comprensión de los sesgos de género en la IAG en los contextos sociales, culturales y lingüísticos de la región.

V. Conclusión

Se ha evidenciado en este estudio que los sistemas de IAG continúan reproduciendo sesgos de género, mediante estereotipos, sub representaciones femeninas, desigualdades en la evaluación y asignación de roles tradicionales. Estos pueden influir negativamente en las oportunidades académicas y profesionales de las mujeres especialmente en los campos STEM, reforzando las brechas históricas de género.

Por lo que es necesario que se implementen técnicas educativas y regulatorias que promuevan sistemas de IAG más equitativos, transparentes e inclusivos.

Además, es urgente la necesidad de aumentar las investigaciones en contextos latinoamericanos que nos permitan comprender y abordar estas problemáticas desde las realidades regionales en donde las mujeres latinoamericanas enfrentan desafíos particulares asociados a las desigualdades estructurales de género, menor participación en el campo científico y tecnológico, en donde existen brechas de acceso a oportunidad educativas y digitales. Pues estos factores podrían amplificar los efectos de los sesgos presentes en IAG. Por lo que se necesita garantizar que la inteligencia artificial no represente un desafío tecnológico, sino una herramienta que permita construir un futuro en el que niñas y mujeres puedan reconocerse, participar y liderar en los espacios donde se genera el conocimiento y la innovación.

Agradecimiento/Reconocimiento

En reconocimiento al equipo de trabajo del presente trabajo.

References

[1] M. M. Rufino Gabriel, M. Rueda Ñopo, F. G. Gabriel Rojas, S. Ponte Cordova, E. Rosario Pacahuala, y P. G. Chavez Ortiz, "El Uso de la Inteligencia Artificial para la Formación Profesional de los Estudiantes de Ingeniería," en *Proc. 23rd LACCEI Int. Multi-Conf. Eng. Educ. Technol.*, Mexico City, Mexico, Jul. 16-18, 2025, doi: 10.18687/LACCEI2025.1.1.737.

[2] G. Currie, J. Currie, S. Anderson y J. Hewis, "Gender bias in generative artificial intelligence text-to-image depiction of medical students," Health Education Journal, vol. 83, no. 7, pp. 732–746, nov. 2024, doi: 10.1177/00178969241274621.

[3] L. Betti, P. Bajardi, y G. De Francisci Morales, «Moral judgments in online discourse are not biased by gender», Sci. Rep., vol. 15, art. 8749, 2025, doi: 10.1038/s41598-025-08749-x.

[4] L. Sun, M. Wei, Y. Sun, Y. J. Suh, L. Shen y S. Yang, "Smiling women pitching down: Auditing representational and presentational gender biases in image-generative AI," Journal of Computer-Mediated Communication, vol. 29, no. 1, art. no. zmad045, ene. 2024, doi: 10.1093/jcmc/zmad045.

[5] V. Ito y L. Grimes, «Occupational identities in the age of generative AI: the case of librarianship», Information Research, vol. 31, n.º iConf, art. 64153, 2026. doi: 10.47989/ir31iConf64153

[6] F. Gülhan, «STEM professionals and teacher perception in images created with ChatGPT and DALL-E», Discover Education, vol. 5, art. 447, 2026. doi: 10.1007/s44217-026-01439-4.

[7] S. S. Sivakaminathan y E. Musi, "Chat GPT is a gender bias echo-chamber in HR recruitment: an NLP analysis and framework to uncover the language roots of bias," AI & SOCIETY, vol. 41, no. 4, pp. 2841–2861, 2026, doi: 10.1007/s00146-025-02564-8.

[8] K. Heley, J. K. Hom, y L. Laestadius, "AI-Generated Images of Substance Use and Recovery: Mixed Methods Case Study," JMIR AI, vol. 5, p. e81977, Feb. 2026, doi: 10.2196/81977.

[9] H. A. Lima, P. H. F. S. Trocoli-Couto, Z. Moazzam, L. C. D. Rocha, A. Pagano, F. F. Martins, L. T. Brabo, Z. S. N. Reis, L. Keder, A. Begum, M. H. Mamede, T. M. Pawlik, y V. Resende, "Quality assessment of large language models' output in maternal health," Scientific Reports, vol. 15, no. 1, art. no. 22474, Jul. 2025, doi: 10.1038/s41598-025-03501-x

[10] M. Achtari, A. Salihu, O. Muller, E. Abbé, C. Clair, J. Schwarz, y S. Fournier, "Gender Bias in AI's Perception of Cardiovascular Risk," Journal of Medical Internet Research, vol. 26, p. e54242, Oct. 2024, doi: 10.2196/54242.

[11] J. Xiu y Y. Sun, "Reinforcing intensive motherhood: A study of gender bias in parental responsibilities allocation by large language models," PLoS ONE, vol. 20, no. 11, p. e0335706, Nov. 2025, doi: 10.1371/journal.pone.0335706.

[12] N. Erogul, "Gender perception in artificial intelligence: Deepseek, Gemini and ChatGPT comparison", The Journal of Education Culture and Society, no. 2, pp. 87–116, 2025.

[13] M. Wölfle y D. Schmuck, "Are bots more biased than humans? Assessing gender stereotypes of warmth and competence in AI-generated ratings, stories, and images," *Computers in Human Behavior Reports*, vol. 18, p. 101145, ago. 2026, doi: 10.1016/j.chbr.2026.101145

[14]. M. Castro Souto y G. Padilla Castillo, "Biases and barriers in generative artificial intelligence from a gender perspective: Analysis of responses in academic orientation," *Texto Digital*, vol. 21, e103518, 2025, doi: 10.5007/1518-2924.2025.e103518. .

[15]. K. Leong y A. Sung, «Gender stereotypes in artificial intelligence within the accounting profession using large language models», *Humanit. Soc. Sci. Commun.*, vol. 11, no. 1, art. no. 1141, sep. 2024, doi: 10.1057/s41599-024-03660-8.

[16] A. Acerbi y J. M. Stubbersfield, "Large language models show human-like content biases in transmission chain experiments," *Proceedings of the National Academy of Sciences*, vol. 120, no. 44, e2313790120, 2023, doi: 10.1073/pnas.2313790120.

[17] Ø. O. Skaar y C. M. Skaar, «The Gentler Researcher: Examining Research Bias and Gender Stereotypes in Sikt AI Chat», *Nordic Journal of Digital Literacy*, vol. 19, n.º 1, pp. 31–43, mar. 2024, doi: 10.18261/njdl.21.1.3.

[18] D. Sandoval-Martín y C. Martínez-Sanzo, "Gender and artificial intelligence: A systematic literature review," *Social Sciences*, vol. 13, no. 5, art. 250, 2024, doi: 10.3390/socsci13050250.

[19] I. Mirza, A. A. Jafari, C. Ozcinar, and G. Anbarjafari, "Quantifying gender bias in large language models using information-theoretic and statistical analysis," *Information*, vol. 16, no. 5, Art. no. 358, 2025, doi: 10.3390/info16050358.

[20] B. D. Menz *et al.*, "Evaluating the Potential of Reasoning Large Language Models to Perpetuate Racial and Gender Disease Stereotypes in Health Care," *JMIR Medical Informatics*, vol. 12, p. e82256, 2025, doi: 10.2196/82256.

[21] G. Wu, S. Tewari, A. Wong, E. Chung, I. Chim, B. Hoang, N. Mansubi, A. Shams, M. Del Buono, S. Srinivasan, H. Paliath-Pathiyal, and O. Khan, "Chatbots and Diabetes: Is There Gender Bias?," *Journal of Patient Experience*, vol. 12, pp. 1–7, 2025, doi: 10.1177/23743735251380954.

[22] Y. Wang «Limitations and potentials of ChatGPT-4 in gender expression: A future research outlook from a feminist perspective», *Edelweiss Applied*

Science and Technology, vol. 8, no. 6, pp. 5855–5868, 2024, doi: 10.55214/25768484.v8i6.3267.

[23] S. Dueñas Mohedas, F. Jiménez Alcarria, and A. Mejón, "Sesgos de género en vídeos con IA Gen: el caso de #WomenWarriors en YouTube Shorts," *Cuadernos.info*, no. 64, pp. 144–166, 2026, doi: 10.7764/cdi.64.98934.

[24] J. J. Docking, L. X. Li, B. D. Menz, S. Bacchi, A. M. Hopkins, and M. J. Sorich, "Evaluating the Potential of Reasoning Large Language Models to Perpetuate Racial and Gender Disease Stereotypes in Health Care," *Journal of Medical Internet Research*, vol. 28, Art. no. e82256, 2026, doi: 10.2196/82256

[25] G. Currie, J. Hewis, y P. Ebbs, «Gender bias in text-to-image generative artificial intelligence depiction of Australian paramedics and first responders», *Australas. Emerg. Care*, vol. 28, n.º 2, pp. 103–109, jun. 2025, doi: 10.1016/j.auec.2024.11.003.

[26] P. Sutera, R. Bhatia, T. Lin, L. Chang, A. Brown, and R. Jagsi, "Generative AI in Medicine: Pioneering Progress or Perpetuating Historical Inaccuracies? Cross-Sectional Study Evaluating Implicit Bias," *JMIR AI*, vol. 4, Art. no. e56891, 2025, doi: 10.2196/56891.

[27] A. Castro-Martínez, J. L. Torres-Martín, and C. Pérez-Ordóñez, "Class, gender and race stereotypes reproduced by generative AI: Recommendations for users," *Encontros Bibli: Electronic Journal of Library Science, Archival Science and Information Science*, vol. 30, pp. 1–26, 2025, doi: 10.5007/1518-2924.2025.e103518.

[28] B. Sanz Urquijo, M. López Belloso, y A. Izaguirre-Choperena, «Empathy, bias, and data responsibility: evaluating AI chatbots for gender-based violence support», *Front. Polit. Sci.*, vol. 7, art. no. 1631881, jul. 2025, doi: 10.3389/fpos.2025.1631881.

[29] T. Newstead, B. Eager, and S. Wilson, "How AI can perpetuate – Or help mitigate – Gender bias in leadership," *Organizational Dynamics*, vol. 52, no. 4, Art. no. 100998, 2023, doi: 10.1016/j.orgdyn.2023.100998.

[30] O. Allam *et al.*, «When Artificial Intelligence Fails the Future: How Unchecked Artificial Intelligence Could Amplify Inequality in Medical Education», *Plast. Reconstr. Surg. Glob. Open*, vol. 14, art. no. e7544, 2026, doi: 10.1097/GOX.0000000000007544.

[31] Z. W. Petzel y L. Sowerby, «Prejudiced interactions with large language models (LLMs) reduce trustworthiness and behavioral intentions among members of stigmatized groups», *Comput. Hum. Behav.*, vol. 165, art. no. 108563, abr. 2025, doi: 10.1016/j.chb.2025.108563.

[32] J. L. Farlow, M. Abouyared, E. M. Rettig, A. Kejner, R. Patel, and H. A. Edwards, "Gender bias in artificial intelligence-written letters of reference," *Otolaryngology–Head and Neck Surgery*, vol. 171, no. 4, pp. 1027–1032, 2024, doi: 10.1002/ohn.806.

[33] J. P. H. Verharen, «ChatGPT identifies gender disparities in scientific peer review», *eLife*, vol. 12, art. no. RP90230, nov. 2023, doi: 10.7554/eLife.90230.

[34] S. J. Villén-Higueras and A. Mata-Núñez, "Influencers virtuales femeninas y representación de género: análisis de su imagen en Instagram," *European Public & Social Innovation Review*, vol. 10, pp. 1–22, 2025, doi: 10.31637/epsir-2025-568.

[35] M. Villalba-Osés, J. P. Consuegra-Ayala, y M. Palomar, «Understanding Gender Bias in Text-to-Image Models Through Quantitative and Interpretable Analysis: A Fashion Case Study», *Expert Syst.*, vol. 43, n.º 4, 2026, doi: 10.1111/exsy.70232.

[36] D. Guilbeault, S. Delecourt, and B. S. Desikan, "Age and gender distortion in online media and large language models," *Nature*, vol. 646, no. 8087, pp. 1129–1137, 2025, doi: 10.1038/s41586-025-09581-z.

[37] E. Mohammadi, M. Thelwall, Y. Cai, T. Collier, I. Tahamtan, and A. Eftekhar, "Is generative AI reshaping academic practices worldwide? A survey of adoption, benefits, and concerns," *Information Processing & Management*, vol. 63, no. 1, Art. no. 104350, 2026, doi: 10.1016/j.ipm.2025.104350. .

[38] A. Xiang, «Diagnosing the bias iceberg in large language models: A three-level framework of explicit, evaluative, and implicit gender bias», *Inf. Process. Manage.*, vol. 63, n.º 3, art. no. 104554, may. 2026, doi: 10.1016/j.ipm.2025.104554.

[39] R. Torronteras Manzano, L. Andaluz Antón, and I. Sacaluga Rodríguez, "IA e Influencers: El uso de la inteligencia artificial para perpetuar los estereotipos de género y la hipersexualización de la mujer," *European Public & Social Innovation Review*, vol. 10, pp. 1–19, 2025, doi: 10.31637/epsir-2025-1017.

[40] R. Huang *et al.*, «Evaluation and Bias Analysis of Large Language Models in Generating Synthetic Electronic Health Records: Comparative Study», *J. Med. Internet Res.*, vol. 27, art. no. e65317, may. 2025, doi: 10.2196/65317.

[41] H.-W.-S. Bao and P. Gries, "Intersectional race–gender stereotypes in natural language," *British Journal of Social Psychology*, vol. 63, no. 4, pp. 1771–1786, 2024, doi: 10.1111/bjso.12748.