Epistemologías críticas y ciudadanía comunicacional: Pensamiento y praxis para la transformación

Colección Libertad y Conocimiento No. 13

Coordinadores
Edizon León Castro
Efendy Maldonado Gómez de la Torre

ISBN Digital: 978-9978-55-237-7

Edición General
Gissela Dávila Cobo
Gestión editorial
Diego S. Acevedo A.

Centro Internacional de Estudios Superiores de Comunicación para América Latina
Av. Diego de Almagro N32-133 y Andrade Marín • Quito, Ecuador
Teléfonos: (593 2) 254 8011
www.ciespal.org
https://ediciones.ciespal.org/

Av. Colón ES-56 y Juan León Mera
Edificio Ave María, Torre B
Telf.: (+593) 963918707
www.uaw.edu.ec
Quito-Ecuador

Este libro fue revisado por pares académicos en su totalidad

Ediciones Ciespal, 2025

La Representación de Colectivos Socioculturalmente Vulnerables en la Era de la Inteligencia Artificial: Desafíos y Oportunidades desde la Antropología Social y Cultural[89].

Jorge Grau Rebollo
Departamento de Antropología Social - GRAFO
Universitat Autònoma de Barcelona

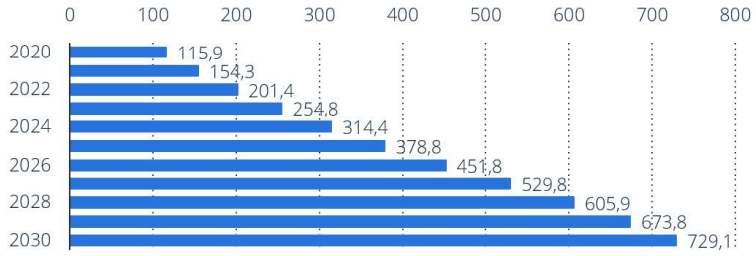
Resumen
En la actualidad, la inteligencia artificial (IA) desempeña un papel crucial en la conformación de los imaginarios sociales, incluyendo a grupos considerados socioculturalmente vulnerables. Desde una perspectiva antropológica, es crucial examinar cómo los sesgos presentes en los algoritmos que alimentan los sistemas de IA influyen en estas representaciones y afectan las percepciones sociales y culturales sobre estos colectivos. Si bien las narrativas audiovisuales generadas mediante IA presentan un gran potencial para desafiar estereotipos e imaginarios distorsionados, no puede obviarse el riesgo que comportan las diversas brechas y barreras que conllevan un acceso desigual a las tecnologías de IA, lo que exige intervenciones críticas que orienten una representación más equitativa y justa en el ámbito digital.

89 Esta investigación se ha realizado en el marco del proyecto *Infancias vulnerables en contextos actuales de multi-crisis: mecanismos institucionales, recursos familiares y estrategias comunitarias de afrontamiento* (AFRONTA, PID2022-139502OB-I00), financiado por la Agencia Estatal de Investigación española (Ministerio de Ciencia e Innovación), dentro del programa de Proyectos de generación de Conocimiento 2022, cofinanciado por la Unión Europea.

Introducción

La inteligencia artificial (IA) se ha consolidado en muy poco tiempo como una de las tecnologías más influyentes en el ámbito global, transformando desde las dinámicas cotidianas hasta las estructuras sociales y políticas fundamentales. Según datos del portal STATISTA[90], el número de usuarios de inteligencia artificial a nivel mundial ha aumentado desde los 115,9 millones de usuarios en el año 2020 hasta los 314,4 en 2024, estimándose un total de 729 millones para el año 2030 (Gráfico 1).

Gráfico 1: Número de usuarios de Inteligencia Artificial a nivel mundial de 2020 a 2030 (en millones)



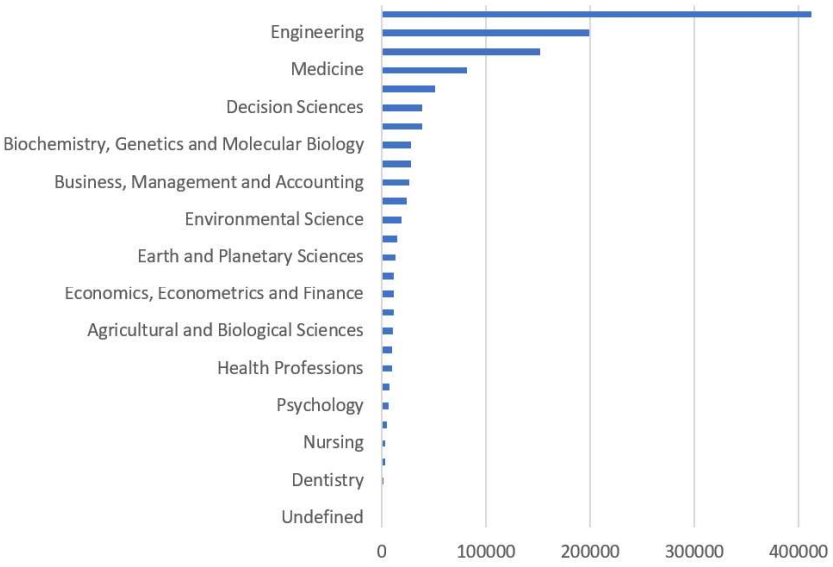
Fuente: STATISTA (https://es—statista—com.eu1.proxy.openathens.net/estadisticas/1535028/inteligencia—artificial—usuarios—mundiales).

En los últimos años, la IA ha permeado sectores clave de la sociedad, como la salud, la educación o la economía (algunas estimaciones calculan su contribución a la economía mundial en el año 2030 en un 3,5% del Producto Interior Bruto global - Fioretti et al. 2024) y ha despertado un notable interés académico: la base de datos SCOPUS cuenta con 185 documentos que abordaban esta cuestión en 1980 frente a los 97.267 en 2024, totalizando 655.704 documentos en el conjunto de sus registros[91]; aunque, como muestra el Gráfico 2, con una distribución disciplinar muy dispar:

90 https://www.statista.com/

91 https://www.scopus.com

Gráfico 2: Distribución de publicaciones sobre Inteligencia Artificial en función de la disciplina. Serie histórica de Scopus (1911—2025)



Fuente: Adaptado de los datos extraídos de SCOPUS (https://www.scopus.com).

Si bien buena parte de esta literatura constata que esta tecnología promete optimizar procesos y mejorar la eficiencia —además de redefinir la manera en que las personas interactúan con el mundo digital—, también ponen de relieve cómo, a medida que su presencia se expande, surgen nuevas preocupaciones, también en lo que respecta a su impacto en la representación de los colectivos socioculturales vulnerables (Wang et al., 2024). Estos grupos, que incluyen ente otros a las minorías raciales o racializadas (Mickel, 2024), las mujeres (Makhortykh et al., 2021), las personas con discapacidad (Mack et al., 2024) y aquellos en situaciones de pobreza (Curto et al., 2024), enfrentan riesgos significativos derivados de la manera en que la IA maneja sus datos y los utiliza posteriormente en la generación de información.

Uno de los temas clave para el análisis y la reflexión es precisamente cómo los algoritmos que sustentan los sistemas de IA, lejos de ser neutrales, están inevitablemente marcados por los sesgos presentes en los datos utilizados para su entrenamiento (Roselli et al., 2019). Estos sesgos, ya sean explícitos o implícitos, tienen el potencial de perpetuar estereotipos existentes y agravar las desigualdades sociales, generando resultados que acaban por reforzar las distorsiones en lugar de corregirlas. De este modo, los algoritmos de IA pueden terminar orientando decisiones que no solo excluyan a los grupos vulnerables, sino que los etiqueten o los presenten de manera distorsionada y deshumanizante (Schultz et al., 2024).

Estos sesgos, que a menudo se presentan de manera invisible y difícil de detectar, son amplificados por el alcance global y la automatización de los sistemas de IA. En la medida en que esta tecnología se inserta en múltiples capas de la vida social y económica, sus implicaciones para la equidad y la justicia social se tornan cada vez más evidentes. Algunos colectivos vulnerables, históricamente marginados en muchas sociedades, corren el riesgo de aumentar su distorsión representacional aún más drásticamente, pudiendo sentirse las consecuencias también en el acceso a recursos y oportunidades. En este sentido, la IA podría llegar a actuar como un resorte detonante o amplificador de las desigualdades existentes si no se desarrolla bajo una perspectiva inclusiva, crítica y reflexiva sobre las realidades socioculturalmente diversas.

En contraste, la visibilización de estos colectivos en los espacios digitales podría beneficiarse de la IA —e incluso en algunos casos servir como un medio de empoderamiento—, proporcionando plataformas para la expresión de voces tradicionalmente orilladas o silenciadas. Sin embargo, esta capacidad de amplificación depende en gran medida de cómo los sistemas sean diseñados y entrenados: si en los procesos de entrenamiento no se toman adecuadamente en cuenta las especificidades culturales, raciales y de género (o si, simplemente, se replican acríticamente los sesgos que puedan contener algunas fuentes

de datos), la IA no haría sino perpetuar las narrativas hegemónicas en lugar de fomentar la visibilidad social de la diversidad sociocultural presente en nuestro entorno.

Pero estos sesgos no se inducen desde el vacío. Por ejemplo, en lo que concierne al género según constatan West et al. (2019), solo el 18 % de los autores de las principales conferencias sobre IA son mujeres y más del 80 % de los profesores de IA son hombres (p.10), lo que, junto con la escasa diversidad interna en los equipos de trabajo, puede generar una desventaja significativa en el diseño de sistemas de IA inclusivos (Fosch-Villaronga y Poulsen, 2022). Esta situación incrementa la probabilidad de que los algoritmos desarrollados reproduzcan sesgos que afecten particularmente a los grupos que, debido a su condición socialmente vulnerable, ya enfrentan diversas brechas y barreras en el acceso a la tecnología y a sus potenciales beneficios. *Así, la estereotipación de los resultados obtenidos en las consultas en línea no sería más que una consecuencia de los sesgos inducidos por las desigualdades estructurales presentes en el contexto social* (Noble, 2018).

Es imprescindible, pues, examinar cómo la inteligencia artificial influye en la representación de los colectivos vulnerables, moldeando las percepciones públicas sobre estos grupos y afectando, en consecuencia, sus oportunidades de mejora social. En este sentido, la IA no es únicamente una herramienta tecnológica, sino un escenario sociotecnológico donde se dirimen los valores culturales que orientarán la inclusión o exclusión de estos colectivos. No es extraño que numerosos estudios hayan puesto sobre la mesa la cuestión de las desigualdades sociales reproducidas o inducidas por los algoritmos, pero también han subrayado la necesidad de usar esta tecnología para combatir y tratar de corregir los sesgos existentes (Shahvaroughi Farahani y Ghasemi, 2024; Zajko, 2021). Así, aun asumiendo que los algoritmos que alimentan la IA suelen ser creados a partir de datos que no reflejan la diversidad real de las sociedades:

> Rather than being concerned over how socio-technical systems reproduce pre-existing biases, we can actually name what we want to avoid reproducing: identifying processes, structures, hierarchies and concepts that have already been articulated by critical AI scholars and those in the social sciences and humanities. Being specific also helps us to name desirable alternatives to reproducing injustice, and orients us to where our actions can have meaningful impact. (Zajko, 2021, p. 1054)

2. Principales desafíos

2.1. Los sesgos algorítmicos en la Inteligencia Artificial: ¿una consecuencia de las desigualdades sociales?

Una de las principales preocupaciones sobre la inteligencia artificial en el ámbito de la justicia social gira en torno al fenómeno de los sesgos algorítmicos. Estos sesgos, producto de datos distorsionados, mal representados o insuficientemente contrastados, pueden dar lugar a decisiones que refuercen las desigualdades preexistentes. En su trabajo sobre los "algoritmos de la opresión", Noble (2018) destaca cómo se refuerzan los estereotipos raciales y de género mediante el uso de algoritmos diseñados sin tener en cuenta los impactos sociales y culturales de los supuestos ideológicos subyacentes. Los resultados de los motores búsqueda priorizan ciertos contenidos por encima de otros y pueden contribuir de este modo a invisibilizar a las comunidades marginadas y a perpetuar narrativas dominantes que las deshumanizan. Esto se refleja, por ejemplo, en el caso de la distorsión que se aprecia en herramientas como los sistemas de reconocimiento facial, donde se ha observado que los algoritmos presentan tasas de error más altas al identificar rostros de mujeres y personas de piel oscura (Buolamwini y Gebru, 2018), lo que parece apuntar a distorsiones inducidas por la interseccionalidad de variables socialmente significativas (género y fenotipo, por ejemplo).

Ahondando en este ámbito, el trabajo de O'Neil (2016) sobre los "armamentos de la destrucción matemática" (*weapons of math destruction*) subraya cómo los sistemas de IA, al estar construidos sobre grandes bases de datos, pueden replicar o incluso agravar significativamente las desigualdades sociales. Por ejemplo, los sistemas de calificación crediticia, basados en IA, pueden discriminar a las personas de bajos ingresos o a las minorías racializadas, lo que les impide acceder a oportunidades económicas (ídem). La falta de transparencia y la opacidad de los modelos hacen aún más difícil identificar y corregir estos sesgos, lo que da lugar a una perpetuación de las desigualdades estructurales. Por esta razón, Capraro et al (2024) advierten del potencial de la inteligencia artificial generativa (IAG) tanto para exacerbar como para mitigar las desigualdades socioeconómicas existentes, poniendo especial énfasis en la perpetuación de sesgos con un elevado potencial para reproducir y amplificar las desigualdades sociales existentes:

> The data used to train AI models could suffer from bias, if those data are based on past human decision making (a notoriously biased process). An example is the translation bias observed in tools like Google Translate, where gender stereotypes are inadvertently perpetuated in language translations. Translating the phrase "she/he is a nurse" from Turkish (which is "genderless") to English (which is "gendered") yielded the feminine form (i.e. "she is a nurse"), while the phrase "she/he is a doctor" yielded the masculine form (i.e. "he is a doctor" (91)). Failing to account for these biases could amplify inequalities and injustices, specifically towards historically marginalized groups. (Capraro et al., 2024, p. 7)

Diversos estudios que han abordado el contenido de conjuntos de datos utilizados para entrenar las IA confirman este riesgo (Cabrera, 2024; George, 2025), mostrando hasta qué punto se reflejan y, potencialmente, amplifican dichos sesgos a lo largo del proceso, además de infringir las leyes de propiedad intelectual.

2.2. La discriminación algorítmica y la brecha digital

La discriminación algorítmica es una extensión de los sesgos de IA, que se manifiesta en la manera en que los sistemas de inteligencia artificial afectan a los colectivos vulnerables de manera diferenciada. Eubanks (2018) advierte en este sentido que la IA no solo reproduce distorsiones en las decisiones automatizadas, sino que también puede contribuir a la exclusión social de ciertos colectivos al dejar atrás a quienes no tienen acceso a la tecnología o a los recursos necesarios para comprenderla y manejarla. La brecha digital se revela entonces como un factor crítico que ahonda las desigualdades: las personas en situación de vulnerabilidad no solo se ven afectadas por los sesgos algorítmicos, sino que también enfrentan dificultades en el acceso a los sistemas que podrían mejorar su calidad de vida.

Esta exclusión digital es particularmente visible en las comunidades rurales o en los grupos de bajos ingresos, donde la falta de acceso a la tecnología limita las oportunidades educativas, laborales y sociales. Como recuerdan Choudhary y Bansal (2022) o Chen et al. (o 2024) la brecha digital no solo implica la insuficiencia o falta de disponibilidad de dispositivos tecnológicos, sino también la carencia de una alfabetización digital adecuada, lo que yugula aún más, si cabe, las oportunidades para que los colectivos afectados puedan participar plenamente en la esfera social y económica digital.

Adicionalmente, esta brecha puede constituir un doble obstáculo. Por un lado, debido a que las políticas y soluciones basadas en IA a menudo ignoran las necesidades de estos grupos al no tener en cuenta las diferencias socioculturales que deben ser abordadas para comprender su realidad. Por otro lado, y a un nivel distinto, a causa de la no inclusión de estas voces y situaciones en el diseño de los algoritmos, lo que sería crucial para prevenir el agravamiento de la desigualdad (Sandvig et al., 2014).

Abundando en este extremo, estudios recientes (Wang et al., 2024) confirman que un factor determinante en la definición de la

vulnerabilidad de una persona o colectivo es su nivel de competencia digital, de modo que cuanto menor sea dicha solvencia el riesgo de desproteger su información personal en las interacciones con herramientas de IA aumenta exponencialmente:

> We argue that people's ability to protect their information privacy is related to their AI competencies. AI skills and privacy protection skills seem to be intertwined with each other closely [...] Users with a higher level of AI knowledge could be more capable of managing and protecting online information privacy when interacting with AI systems. Users with a high level of AI skills are also more likely to have a high level of privacy protection skills, such as adjusting privacy settings. (Wang et al., 2024)

En este ámbito, el escepticismo tecnológico y la escasa comprensión de la tecnología pueden aliarse con otras variables sociodemográficas estructurales, como la edad o el género, para incrementar el perfil vulnerable de estos segmentos de población:

> Specifically, compared with the average user, the most vulnerable groups with the lowest levels of AI knowledge and AI skills (i.e. unskilled skeptics and neutral unskilled) were mostly older, with lower levels of education and privacy protection skills. In addition, expert skeptics appear to be mainly males and had higher levels of privacy protection skills than the average user. As we expect that this explanatory model may change over time along with the rapid development of AI technologies, we recommend future studies to further examine whether the model in the context of the AI divide will differ from conventional digital divide studies in the future. (Ídem)
>
> The digital divide for the older female is not only a matter of technology but also a matter of society. It is widely accepted that women are the major caregivers, the central part of self-care, and the most essential care recipients, owing to their long-life expectancy and socio-historical background. Consequently, women need to take advantage of digital for Elder care. Nevertheless, if there is a disparity in digital accessibility between genders, it would be a major obstacle to further utilizing digital Elder care. Concerning sustainable aging, the issue of the digital divide by gender should be paid attention. (Chen et al., 2024, p. 15)

> The inscription of masculinity in technology affected women's self-confidence to the extent that they felt less able to learn more about technologies, but such "masculine" framing of ICTs also formed a barrier for men who are not yet digitally connected. For some, this was another reason to feel shame, which prevented them from asking for (community) support. (Goedhart et al., 2022, p. 839)

2.3. La representación de los colectivos vulnerables en la IA: más allá de la visibilidad

Otra de las dimensiones problemáticas de la IA en relación con los colectivos vulnerables refiere a la gestión de las representaciones digitales de estos grupos. Así, como hemos mencionado anteriormente, los sistemas de reconocimiento facial y otros algoritmos de clasificación tienden a "ver" a las personas vulnerables de una manera estereotipada o simplificada, lo que puede reducir su humanidad y aumentar su marginación (Raji & Buolamwini, 2023). Estos sistemas, además, pueden ser entrenados con datos poco representativos que omiten la diversidad real de las experiencias humanas, lo que resultaría en la toma de decisiones erróneas. Por consiguiente, la representación mediada por IA puede constituir tanto un problema de visibilidad como de invisibilidad o infrarrepresentación: si bien la IA permite visibilizar ciertas realidades que antes no se reflejaban en los medios, esta presencia a menudo puede estar filtrada por un prisma ideológico que distorsione o simplifique a algunos grupos étnicos y culturales.

Por otro lado, los algoritmos de IA tienden a infrarrepresentar o excluir a ciertos colectivos, como las personas con discapacidad. Así, Mack et al. (2024) muestran cómo las herramientas de generación de imagen a partir de instrucciones de texto (*text-to-image AI models*) muestran un imaginario reduccionista que no hace sino "[...] *propagate broader societal misconceptions and biases about disability*" (p. 16). La pobre representación digital de estos grupos puede reforzar su condición marginada al carecer de la misma visibilidad en condiciones similares a otros sectores de la población. En esta línea abundan Toorn

y Scully (2024), quienes sostienen que "*When it comes to disability, the injustices of algorithmic social sorting are representational, material and epistemic*" (p. 3023), apuntando singularmente a la dificultad de reconstruir o capturar la dimensión socio-relacional de la discapacidad.

2.4. Ética, responsabilidad y regulación en la IA

De este modo, el creciente poder de la IA plantea una serie de cuestiones fundamentales en torno a la ética y la responsabilidad en la investigación, la docencia y la divulgación a distintos niveles. ¿Cómo garantizar, por ejemplo, que los tan referidos algoritmos no perpetúen las desigualdades y sean utilizados de manera equitativa? La respuesta, según Buolamwini y Gebru (2018), pasaría por una mayor responsabilidad de empresas y organizaciones implicadas sobre los impactos sociales de sus productos y servicios, lo que, entre otras cosas, aumentaría la transparencia en los procesos de entrenamiento de los modelos empleados, así como en la posterior toma de decisiones a partir de los mismos.

En estrecha conexión con este asunto, se reclama con frecuencia que los marcos regulatorios deben garantizar que esta tecnología se utilice de manera inclusiva, protegiendo a los colectivos vulnerables de las posibles consecuencias adversas que pueden inducir estos sistemas y garantizando en la medida de lo posible la equidad, la transparencia y el control necesarios para un uso más responsable de los mismos (Olatunji Akinrinola et al., 2024). Por esta razón, la implementación de políticas que aseguren la diversidad en los datos, el diseño ético de los algoritmos y la responsabilidad en el uso de la IA son pasos cruciales para evitar la perpetuación de estereotipos y de desigualdades en una sociedad crecientemente digital. Es en esta línea que estudios como el de Stahl et al. (2022) vienen reclamando la necesidad de que las agencias reguladoras se integren en el "ecosistema de gobernanza" (p. 23) de la inteligencia artificial.

3. Principales oportunidades

Sin embargo, la adopción de tecnologías de Inteligencia Artificial ofrece también oportunidades para mejorar la presentación y la inclusión de colectivos en situación de vulnerabilidad en condiciones más acordes a la realidad y a los contextos de cada uno de ellos. Como señalan García Peñalvo et al. (2023):

> *Para poder utilizar con criterio y conocimiento de causa una tecnología en los procesos de enseñanza y aprendizaje, primero se deben conocer sus posibilidades y límites sin dejarse llevar por los extremismos, que suelen estar especialmente sesgados cuando una tendencia potencialmente disruptiva hace su aparición, como ha sucedido con la IA generativa, cuya penetración ha sido especialmente acelerada.* (p. 7)

Sin ánimo de ser exhaustivo —y con la cautela a que obliga una tecnología que, sobre todo en lo referente a la IA generativa, ha eclosionado socialmente a gran escala hace relativamente muy poco tiempo[92] y se encuentra en un proceso de expansión y asentamiento con muy escasos precedentes—, algunas de las líneas en las que puede preverse una oportunidad de avance social en el ámbito de la vulnerabilidad son las siguientes:

3.1. Incremento de la visibilidad de colectivos mediáticamente infrarrepresentados o escasamente visibles

Como es bien sabido, el procesamiento de información basado en sistemas de IA atesora un enorme potencial para analizar ingentes volúmenes de datos provenientes de diferentes fuentes y en distintos formatos. Esta capacidad sería idónea para identificar patrones y situaciones específicas de vulnerabilidad que pudieran estar pasando desapercibidas en contextos socioculturalmente específicos. Además, la implementación de algoritmos avanzados podría facilitar una

92 ChatGPT, uno de los modelos de lenguaje generativo más populares en el mundo, salió a la luz pública en noviembre de 2022.

segmentación más afinada de la información recabada, facilitando de este modo la identificación de correlaciones complejas y de asociaciones significativas entre variables que sin estas herramientas resultarían más difíciles de detectar. Este potencial analítico podría revertir positivamente en la orientación de políticas públicas y abrir –o impulsar– nuevas vías de investigación que promuevan enfoques holísticos.

Paralelamente, en el terreno académico, la IA es un soporte enormemente útil para trabajar con cantidades considerables de literatura científica, contribuyendo así a la identificación de áreas poco desarrolladas o incluso de lagunas teóricas, por ejemplo. Todo ello redundaría en el incremento del impacto social de los estudios académicos, ya fuesen de investigación aplicada o básica, sobre el bienestar público. En cualquier caso, una condición necesaria para consolidar este potencial sería la consideración de las implicaciones éticas del uso de IA en estos contextos, sobre todo en lo referente a garantizar la transparencia de los procesos y evitando la reproducción de sesgos estructurales.

3.2. Detección y corrección de distorsiones en los sistemas existentes

En esta línea, es crucial prevenir una de las consecuencias más nocivas que puede comportar sobre los colectivos vulnerables la deformación inducida por sesgos ideológicos: los procesos de alterización sobrevenida. De este modo, si tradicionalmente los grupos distintos a los considerados mayoritarios o dominantes suelen verse expuestos a procesos de extrañamiento, deformación e incluso exclusión social (Boivin et al., 2004), la adopción de tecnologías de IA permitiría cuestionar determinadas representaciones dominantes sobre la otredad y generar opciones más inclusivas por medio de narrativas que restituyan las voces de colectivos históricamente excluidos, o que ofrezcan alternativas que contrarresten ciertos estereotipos

y constructos ideológicos hegemónicos sobre la diferencia y la desigualdad social.

3.3. Facilitación de la accesibilidad y optimización del acceso a servicios y recursos culturalmente adaptados

Mediante la detección temprana de necesidades (a través del procesamiento de datos a gran escala para identificar patrones y perfiles susceptibles de intervención a que me refería anteriormente), podrían personalizarse servicios y atender a necesidades más adaptadas a cada persona o grupo social. Esto podría llevarse a cabo mediante la optimización logística de recursos, la configuración de asistentes virtuales y *chatbots* (que, complementando la acción humana, proporcionasen información y asistencia personalizada a los usuarios), o identificando condiciones potencialmente discriminatorias en el acceso a ciertos servicios sociales.

A un nivel más inmediato, podrían optimizarse los mecanismos de asistencia rápida a personas en situaciones de riesgo, facilitando la información y los recursos necesarios para atender y manejar la emergencia. En esta línea, estudios como los de Cedeno-Moreno & Millan (2023) han mostrado cómo la IA permite implementar procedimientos de detección precoz de problemas de salud mental mediante el procesamiento de lenguaje natural (PLN), con el propósito último de identificar riesgos en un estadio temprano, antes de que puedan agravarse.

Debe tenerse en cuenta, no obstante, que la eficiencia final de esta vía de apoyo social dependerá en buena medida de la gestión de las situaciones concretas de diversidad cultural que impliquen cada contexto cada y colectivo susceptible de ser atendido, puesto que la comunicación intercultural puede verse limitada por condicionantes diversos (como por ejemplo la disparidad de sistemas simbólicos, códigos comunicativos o las pautas de interacción social, entre otros).

3.4. Mejora de los sistemas de detección de alertas y atención temprana en situaciones de riesgo

Refinar los sistemas de detección de alertas y atención temprana en contextos de riesgo para las personas constituye una prioridad en el ámbito de la protección de derechos humanos y la seguridad pública; por ejemplo, aunque no únicamente en estos supuestos, en contextos de violencia, sea cual sea su naturaleza: de género, vicaria o incluso institucional. Se requiere para ello optimizar los procesos de identificación de patrones de riesgo mediante el empleo de tecnologías avanzadas, como los sistemas de análisis predictivo, que permitan anticipar situaciones de violencia y ofrecer intervenciones adecuadas. En este sentido, la implementación de algoritmos de detección basados en *big data* debe acompañarse de una robusta infraestructura de soporte que incluya la capacitación de profesionales y la integración de los sistemas de alerta con resortes de apoyo multidisciplinares y multisectoriales (a nivel judicial, policial, sanitario, educativo, etc.).

Pese al carácter de urgencia que suele conllevar una actuación inmediata, para asegurar que la respuesta sea eficiente y adecuada a las necesidades de las víctimas deben fortalecerse y refinarse los protocolos de intervención. Dicha afinación exige una colaboración interinstitucional entre organismos de seguridad, salud, servicios sociales y entidades jurídicas, con el fin de crear un entramado sistémico de atención que garantice una evaluación integral de las necesidades atendidas. La eventual aplicación de modelos de análisis del riesgo, basados en la evaluación contextual y en el historial de las víctimas, debe complementarse con enfoques de protección que prioricen el bienestar y la seguridad de las personas afectadas, especialmente en situaciones de violencia vicaria, donde se pone en riesgo la integridad física y emocional de niños/as y adolescentes menores de edad.

Por otra parte, no debemos perder de vista el papel potencialmente inductor de vulnerabilidad sobrevenida de la violencia institucional (Grau-Rebollo et al., 2024). Para paliarla —y en último término, evitarla—,

deben revisarse desde una perspectiva crítica las estructuras de poder y las prácticas normativas que perpetúan (o acentúan) la desprotección de las víctimas. Una vez más, la mejora de estos sistemas debe integrar enfoques interseccionales que tomen en consideración las múltiples facetas y dimensiones de la violencia, atendiendo a factores de género, etnia, clase social o discapacidad, entre otros.

3.5. Prevención de la discriminación social

La prevención de la discriminación sobre colectivos socioculturalmente vulnerables implica el reconocimiento de las diferentes formas de diversidad y complejidad social y cultural presentes en los contextos de interacción humana. Desde una perspectiva antropológica, es fundamental comprender cómo la reproducción o difusión de cualquier distorsión, sea explícita o implícita, sedimenta en las estructuras sociales, se transfiere al imaginario colectivo y se perpetúa en la práctica cotidiana. Los fenómenos de discriminación (y, subsiguientemente, de exclusión) no pueden entenderse únicamente como acontecimientos individuales, sino que son producto de dinámicas sociales concretas, históricamente construidas, que afectan de manera diversa y concurrente al acceso equitativo que a ciertos grupos se les concede sobre determinados bienes, recursos o derechos sociales.

Desde esta perspectiva, las estructuras y mecanismos de exclusión operan de manera sistémica y requieren un abordaje holístico (integral a las diversas dimensiones implicadas) a la vez que específico (adaptado a cada contexto en que se manifiestan). En base a esta identificación, se pueden diseñar posteriormente intervenciones más focalizadas que no solo busquen la corrección de prácticas discriminatorias puntuales (tarea imprescindible, por otro lado), sino que también promuevan una transformación cultural y estructural a mayor escala. Para lograrlo, es necesario que cualquier intervención involucre a los diferentes actores sociales e institucionales implicados y que en la medida de lo posible cuente con un enfoque participativo que confiera agencia a los propios

colectivos vulnerabilizados. Desde esta perspectiva, la erradicación de los sesgos y la discriminación no puede limitarse únicamente a la modificación de actitudes individuales, sino que debe buscar también una reestructuración de las dinámicas sociales que apuntalan la marginalización y la exclusión.

3.6. Diseño de plataformas de capacitación inclusivas

Usando herramientas de IA, se pueden crear, por ejemplo, plataformas educativas que faciliten materiales en múltiples idiomas, en formatos accesibles (como audio o video) y con contenido adaptado a diferentes contextos socioculturales, lo que es crucial para algunos colectivos (migrantes, comunidades indígenas, o personas con discapacidades, entre otros). En esta línea, en el ámbito formativo pueden diseñarse itinerarios personalizados atendiendo a necesidades individuales o características culturalmente específicas, con finalidades que pueden abarcar desde una mejor integración en el mercado laboral de población inmigrante (Sydoruk, 2024), hasta diseños de aprendizaje optimizados para alumnado con necesidades especiales (Ayobami O Ayeni et al., 2024).

3.7. Detección automatizada de desinformación y generación de información veraz

En la tarea de identificación y filtrado de información no veraz o distorsionada que afecte negativamente a colectivos vulnerables, el análisis de patrones de contenido en redes sociales y otros medios mediante herramientas basadas en IA pueden contribuir a detectar esta desinformación antes de que se propague. También puede asistir en la creación de contenido educativo que explique de forma clara y asequible a diferentes niveles formativos temas complejos o controvertidos, facilitando el acceso a información precisa y confiable.

Por descontado, la sofisticación de estos procesos de verificación y propuesta de información dependen, entre otras cosas, del

refinamiento de los sistemas de procesamiento del lenguaje natural y de la fiabilidad de los marcos de evaluación para la detección automatizada de desinformación, integrando diferentes fuentes de información y técnicas de verificación diversas (Bonet-Jover, 2023). En este proceso, pueden usarse herramientas de cribaje de información existente, como los instrumentos de prevención y verificación recopilados en el banco de recursos del proyecto AFRONTA[93].

4. Una propuesta de integración de la IA en la investigación audiovisual aplicada

Desde el proyecto AFRONTA[94] tratamos de avanzar hacia la búsqueda de soluciones a los desafíos de la sociedad presentados por las consecuencias que tienen sobre familias vulnerables (y en particular sobre la crianza de sus hijos e hijas) la concurrencia de diversas crisis (económica, energética, medioambiental, sanitaria y de fiabilidad de la información), en escenarios donde la posible cronificación de condiciones estructurales de desigualdad, así como el afloramiento de circunstancias sobrevenidas, pueden agravan su condición vulnerable. Ello exige atender a la existencia de brechas y barreras de diversa naturaleza (sociotecnológicas y de recursos, género o digital, además de las que imponen una inadecuada comprensión —y consiguiente gestión— de la diversidad cultural) y de los obstáculos que resultan de una insuficiente competencia tecnológica e informacional en las relaciones con entidades y administraciones potencialmente proveedoras de bienestar.

Para ello proponemos abordar esta concurrencia desde una perspectiva interdisciplinar (Antropología, Medicina, Psicología, Pedagogía, Trabajo Social, Educación Social y Comunicología) e intersectorial (academia, instituciones, entidades, agentes y sujetos

93 https://webs.uab.cat/afronta/recursos/prevencion-de-fakes

94 https://webs.uab.cat/afronta

implicados) que permita incidir en cuatro ámbitos de interpenetración disciplinar: (1) salud, (2) apoyo formal e informal, (3) protección institucional sobre mujeres madres víctimas de violencia de género y sobre sus hijos e hijas, y (4) comunicación, transmisión cultural y formación de profesionales. Dentro de este cuarto ámbito incidiré aquí en una propuesta de investigación audiovisual aplicada que entronca con la educomunicación mediante la creación de recursos audiovisuales. Concretamente, proponemos trabajar sobre contenidos audiovisuales específicos orientados a cuestionar y deconstruir ideas preconcebidas que se realimentan de (y proporcionan combustible para) ciertos estereotipos sobre grupos sociales vulnerables. En concreto, nos interesa fomentar el aprendizaje reflexivo mediante actividades interactivas orientadas a alumnado de educación secundaria (12-16 años) que ayude a poner en cuestión potenciales prejuicios y estereotipos que inducen una visión negativa y sesgada sobre la realidad social de colectivos socialmente desfavorecidos.

Así, en la línea apuntada por (Hermann et al., 2025) proponemos usar la IA para que el alumnado pueda tomar consciencia de las distorsiones que puedan albergar ciertos estereotipos culturales y adopte una actitud crítica frente a la presencia de prejuicios que socaven la empatía hacia algunos grupos o colectivos. El nivel al que cada docente decida adaptar y aplicar estas propuestas dependerá de los propósitos específicos de cada caso y de los requisitos y proyecto docente de cada centro. En el caso concreto que presento a continuación, el objetivo es proporcionar al profesorado de secundaria herramientas que permitan: (a) personalizar el aprendizaje, (b) identificar sesgos mediante el análisis de datos, (c) fomentar el pensamiento crítico y (d) generación de contenidos inclusivos para abordar estas cuestiones.

4.1. La generación de vídeo por IA como potencial recurso pedagógico

El primer ejemplo escogido en este caso es el cortometraje *A Secret in the Dark*[95], generado por Joan Riedweg (director de cine y realizador de televisión con una dilatada trayectoria profesional en el ámbito audiovisual[96]), que presenta la historia de un abuso infantil, guardado en secreto durante años, por parte de un familiar próximo. Pese a la muerte del agresor y de la imposibilidad de obtener justicia y reparación del daño causado, la joven narradora reafirma su necesidad de explicar la experiencia vivida para poder finalmente llegar a superarla. Esta historia, que alude a un contexto referencial de extrema dureza, puede abordarse por medio de un formato que permita trabajar, entre otras, cuestiones como el abuso infantil, el trauma, el miedo y el secretismo, la ausencia de reparación legal o los procesos de resiliencia y autoafirmación.

De hecho, Riedweg, colaborador en el proyecto AFRONTA, diseñó y ejecutó esta pieza en el marco de la tercera edición de *Gen:48*, un concurso patrocinado por Runway ML (una plataforma que permite trabajar con herramientas de Inteligencia Artificial para la generación audiovisual[97]) en la que cada participante debía crear una producción audiovisual de corta duración en un plazo máximo de 48 utilizando la tecnología Gen3[98]. Es, por lo tanto, el recurso a un material que no se genera directamente desde un proyecto de investigación, pero que se alinea con sus objetivos y que amplía el repertorio de herramientas didácticas para uso docente. Por otro lado, el idioma original de esta pieza es el inglés, lo que no constituye un obstáculo para su manejo en

95 https://youtu.be/a5WY_YlF1KI?feature=shared.

96 https://www.riedweg.cat/joan

97 https://runwayml.com/

98 "El Gen-3 Alfa de Runway, el primero de una serie de nuevos modelos, se centra en mejorar la fidelidad, la consistencia y el movimiento con respecto a su predecesor. Está entrenado en una nueva infraestructura para el aprendizaje multimodal a gran escala, que combina el entrenamiento de vídeo e imagen. Gen-3 Alpha dispone de varias herramientas, como texto a vídeo, imagen a vídeo y texto a imagen, así como modos de control como el pincel de movimiento y los controles de cámara avanzados". (https://www.datacamp.com/es/blog/what—is—runway—gen—3)

contextos lingüísticamente diversos (gracias, entre otras cosas, a los mecanismos y recursos de traducción automática disponibles en la actualidad).

Todo el contenido de *A Secret in the Dark* se ha concebido mediante diferentes herramientas de IA (chatGPT, Suno, Runway...), incluidas la música y la letra de la canción a través de la cual se articula la historia. El texto de la canción (tabla 1) se pensó, al igual que las imágenes que lo componen (tabla 2), de modo que sugiriese los acontecimientos que estructuran la trama, pero sin llegar explicitar o revelar abiertamente los detalles del abuso:

Tabla 1. Letra de la canción *A Secret in the Dark* (2024).

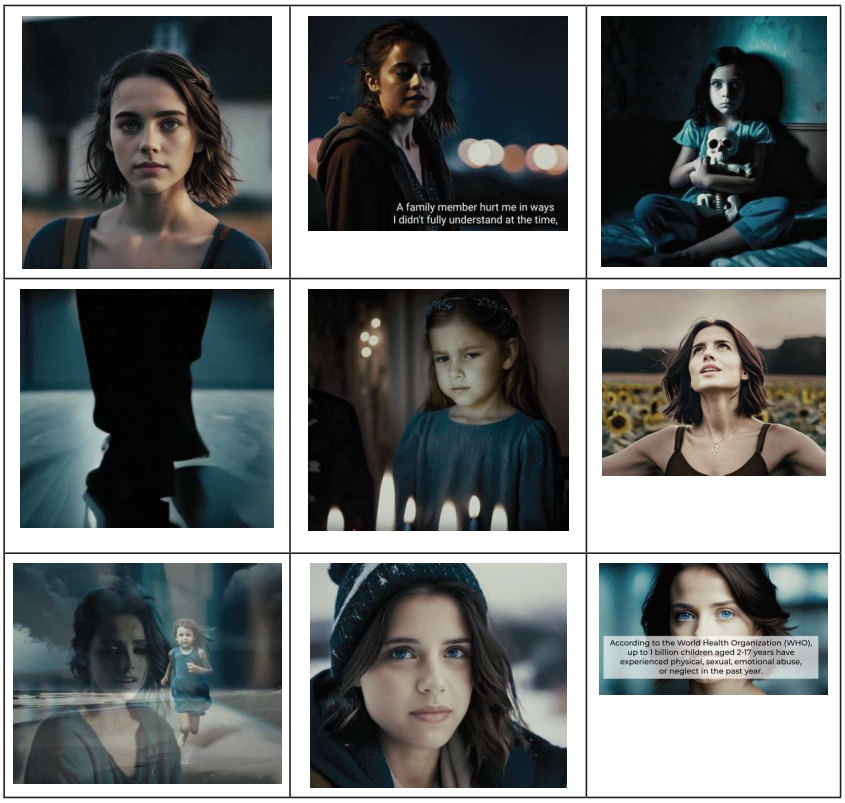
Versión original (inglés)	Versión traducida (español)
Okay, I'm going to tell you a secret.	Vale, te voy a contar un secreto.
When I was a child, I experienced things I never had the courage to talk about.	Cuando era niña, viví cosas de las que nunca tuve el valor de hablar.
A family member hurt me in ways I didn't fully understand at the time.	Un miembro de mi familia me hizo daño de formas que en ese momento no podía entender del todo.
But I knew it was unfair and painful.	Pero sabía que era injusto y doloroso.
For many years, I kept this secret inside.	Durante muchos años, guardé este secreto dentro de mí.
Thinking no one would believe me if I spoke up.	Pensando que nadie me creería si hablaba.
Also, just thinking about it made me feel	Además, solo pensar en ello me hacía sentir
an overwhelming sadness and a fear I couldn't put into words.	una tristeza abrumadora y un miedo que no podía expresar con palabras.
Now, 15 years later, all those memories have come flooding back, as if it happened yesterday.	Ahora, 15 años después, todos esos recuerdos han regresado de golpe, como si hubiera pasado ayer.
I felt the need to speak out.	Sentí la necesidad de hablar.
To give voice to what I went through.	De dar voz a lo que pasé.
I finally decided to report it, because I know I deserve to be heard.	Finalmente decidí denunciarlo, porque sé que merezco ser escuchada.
I thought the silence would protect me.	Pensaba que el silencio me protegería.

Versión original (inglés)	Versión traducida (español)
But it only held me down.	Pero solo me retuvo.
I've carried all this weight so long.	He llevado este peso tanto tiempo.
Now it's time to let it drown.	Ahora es el momento de dejarlo ahogarse.
The pain they caused, it won't define.	El dolor que me causaron no me definirá.
The woman I am, this heart of mine.	La mujer que soy, este corazón mío.
Even though they've left this place.	Aunque ellos ya no estén en este lugar.
I'll find my strength. I'll find my grace.	Encontraré mi fuerza. Encontraré mi gracia.
Cause the memories, they won't stay.	Porque los recuerdos no se quedarán.
But I'll be stronger every day.	Pero cada día seré más fuerte.
Now the truth, it's mine to tell.	Ahora la verdad es mía para contarla.
I'll break the silence, break the spell.	Romperé el silencio, romperé el hechizo.
Fifteen years have passed me by.	Quince años han pasado.
But I'm still standing, reaching high.	Pero sigo de pie, alcanzando más alto.
The one who hurt me's gone away.	La persona que me hizo daño se ha ido.
But I've got something left to say.	Pero tengo algo que decir.
So the justice may not come.	Así que tal vez la justicia no llegue.
I'll rise, I'll shine, I won't succumb.	Me levantaré, brillaré, no sucumbiré.
For the little girl who couldn't speak,	Por la niña pequeña que no podía hablar,
I'll be strong. I'll find what I seek.	seré fuerte. Encontraré lo que busco.
I'm not alone, I'm not afraid.	No estoy sola, no tengo miedo.
I'll carry on, I've got a way.	Seguiré adelante, tengo un camino.
In every tear, I see the light.	En cada lágrima veo la luz.
And through the dark, I'll win the fight.	Y a través de la oscuridad, ganaré la lucha.
But what happened is that the person who hurt me took their own life before I could even share my story.	Pero lo que pasó es que la persona que me hizo daño se quitó la vida antes de que pudiera contar mi historia.
Now, I find myself in a situation where I can't prove what I carry inside.	Ahora, me encuentro en una situación en la que no puedo probar lo que llevo dentro.
But that doesn't mean it's not real to me.	Pero eso no significa que no sea real para mí.
I know what I went through.	Sé lo que viví.
And even though I may never find justice the way I imagined.	Y aunque tal vez nunca encuentre justicia como la imaginaba.

Versión original (inglés)	Versión traducida (español)
I refuse to let my voice go unheard.	Me niego a dejar que mi voz quede en el silencio.
I'm doing this for myself.	Lo hago por mí misma.
For the little girl I was.	Por la niña que fui.
And for the woman I've become.	Y por la mujer en la que me he convertido.

Fuente: Transcripción obtenida de la versión de *A Secret in the Dark* en la plataforma YouTube (https://youtu.be/a5WY_YlF1KI?feature=shared). La traducción al español se ha realizado con revisión de idioma por medio de IA (ChatGPT)

Tabla 1. Fotogramas de *A Secret in the Dark* (2024)



Fuente: A Secret in the Dark en la plataforma YouTube (https://youtu.be/a5WY_YlF1KI?feature=shared)

Entre muchas posibilidades, la selección de esta pieza audiovisual permite trabajar sobre los fundamentos culturales de las distintas etapas de la vida, el género, las nociones de abuso y de reparación, el imaginario social, los estereotipos que puedan orientarlo y los lugares comunes —o los vacíos representacionales— al respecto. Tanto el conjunto de la producción como cada uno de los fotogramas seleccionados puede usarse como material auxiliar en el planteamiento de la vulnerabilidad a causa de las diferentes violencias ejercidas sobre la infancia y servir como instrumento de elicitación de información.

De este modo, la interacción visual puede facilitar al alumnado la comprensión y fomentar la empatía, promoviendo un entorno de trabajo inclusivo y participativo que ayude, por ejemplo, a valorar mejor las perspectivas y experiencias de grupos vulnerables (Pepe, 2023).

4.2. Propuesta didáctica basada en imágenes generadas mediante IA

En líneas generales, desde el proyecto AFRONTA tratamos de contribuir con estos objetivos mediante diversas líneas de actuación, una de las cuales se sustancia en una propuesta pedagógica[99] —en actualización permanente a lo largo del proyecto— que ofrece recursos didácticos para docentes (principalmente de Educación Secundaria Obligatoria, aunque puede escalarse a otros niveles educativos), enfocándose en metodologías activas y colaborativas.

En este momento, la propuesta cuenta con dos actividades en catalán (están orientadas a centros aunadas en la propuesta: *Menys vulnerables!* (¡Menos vulnerables!), cuyo objetivo principal es fortalecer el pensamiento crítico y prevenir la desinformación entre el alumnado (imagen 1). Este material ofrecerá una serie de actividades prácticas basadas en ejemplos visuales para identificar mecanismos que distorsionan la realidad. La estructura de estas actividades se ha diseñado pensando en la flexibilidad y la posibilidad de adaptación a diversos contextos educativos, abordando temas de vulnerabilidad sociocultural.

99 https://webs.uab.cat/afronta/recursos/proposta-didactica-catala

Imagen 1. Infografía informativa de la propuesta didáctica *Menys Vulnerables!* (¡Menos vulnerables!)

Fuente: elaboración propia en el marco del proyecto AFRONTA.

Los ejes fundamentales de la propuesta son los siguientes:

- *Fomentar el aprendizaje activo y colaborativo*: estimulando la participación activa del alumnado y su capacidad para trabajar en equipo, desarrollando habilidades sociales y comunicativas.
- *Desarrollar el pensamiento crítico*: valorando su capacidad para analizar, reflexionar y tomar decisiones informadas sobre los temas trabajados a partir de una serie de imágenes y pósteres que impulsen la reflexión.
- *Favorecer la creatividad*: sugiriendo actividades para que cada estudiante pueda expresarse de manera original, reforzando habilidades como la empatía y la comprensión.
- *Promover una actitud responsable y autónoma* en el proceso educativo mediante el diálogo y la argumentación.

Todo ello se articula a través de un conjunto de documentos interactivos en diversos formatos, con un objetivo central: aprender a interpretar imágenes desde una perspectiva social y contextual. De este modo, a partir del análisis de una imagen —en el caso de estas actividades, en formato póster— el alumnado debe reflexionar sobre la diversidad sociocultural, el papel de estereotipos e ideas preconcebidas en la representación de ciertas personas o colectivos y sobre el impacto de las imágenes generadas por IA.

Por ejemplo, a partir de un cartel con una imagen generada por IA y un texto que alude alegóricamente a un juego infantil (imagen 2), se plantean dos cuestiones centrales: ¿qué conexión pueden tener las imágenes con el juego? Y ¿qué podría hacerse para cambiar la situación a la que refiere? A partir de las respuestas que el alumnado, individualmente o en pequeños grupos, dé a estas respuestas, el o la docente puede abordar y orientar distintas cuestiones relativas a la vulnerabilidad (con el apoyo de materiales y recursos incluidos en otros apartados de la misma actividad).

Imagen 2. Uno de los pósteres incluidos en la propuesta didáctica: Menys vulnerables! (¡Menos vulnerables!)

Fuente: elaboración propia, utilizando herramientas de IA.

En definitiva, estas propuestas tratan de favorecer la formación de una ciudadanía crítica y consciente del alcance y de los límites de las representaciones que la envuelven, incidiendo en las etapas educativas regladas y proporcionando un entorno didáctico estructurado, pero flexible, que combina formatos y soportes diversos para avanzar en este objetivo. La celeridad que evidencia el desarrollo de tecnologías de IA (especialmente de IA generativa, aunque no sólo) inducirá con toda seguridad múltiples cambios en los procesos de enseñanza y aprendizaje a diversos niveles. Desde el entorno académico, es responsabilidad nuestra tratar de aprovecharlos para contribuir a la formación de una sociedad menos desigual y, por consiguiente, mejor cohesionada.

Bibliografía

Ayobami O Ayeni, Rodney E Ovbiye, Ayomide S Onayemi, & Kayode E Ojedele. (2024). AI-driven adaptive learning platforms: Enhancing educational outcomes for students with special needs through user-centric, tailored digital tools. *World Journal of Advanced Research and Reviews*, 22(3), 2253-2265. https://doi.org/10.30574/wjarr.2024.22.3.0843

Boivin, M., Rosato, A., & Arribas, V. (2004). *Constructores de Otredad. Una introducción a la Antropología Social* (3.ª ed.). Antropofagia.

Bonet-Jover, A. (2023). *Abordando el tratamiento automático de la desinformación: modelado de la confiabilidad en noticias mediante Procesamiento del Lenguaje Natural* [Tesis doctoral]. Universidad de Alicante.

Buolamwini, J., & Gebru, T. (2018). Gender Shades: Intersectional Accuracy Disparities in Commercial Gender Classification. *Proceedings of Machine Learning Research*, *81*, 1-15. https://proceedings.mlr.press/v81/buolamwini18a/buolamwini18a.pdf

Cabrera, C. (2024, diciembre 9). La IA genera audios plagados de machismo, racismo e infracciones de derechos de autor. *El País (Edición digital)*. https://elpais.com/tecnologia/2024-12-09/la-ia-genera-audios-plagados-de-machismo-racismo-e-infracciones-de-derechos-de-autor.html

Capraro, V., Lentsch, A., Acemoglu, D., Akgun, S., Akhmedova, A., Bilancini, E., Bonnefon, J.-F., Brañas-Garza, P., Butera, L., Douglas, K. M., Everett, J., Gigerenzer, G., Greenhow, C., Hashimoto, D., Holt-Lunstad, J., Jetten, J., Johnson, S., Longoni, C., Lunn, P., … Viale, R. (2024). The Impact of Generative Artificial Intelligence on Socioeconomic Inequalities and Policy Making. *SSRN Electronic Journal*. https://doi.org/10.2139/ssrn.4666103

Cedeno-Moreno, D. E., & Millan, A. (2023). Arquitectura de PLN aplicada al contexto de la salud mental. *I+D Tecnológico*, *19*(2), 24-29. https://doi.org/10.33412/idt.v19.2.3770

Chen, D., Han, J., & Song, Y. (2024). The efficacy and tendency of the gender digital divide and Smart Aging Policy in China. *Health Care for Women International*, 1-22. https://doi.org/10.1080/07399332.2024.2385326

Choudhary, H., & Bansal, N. (2022). Addressing Digital Divide through Digital Literacy Training Programs: A Systematic Literature Review. *Digital Education Review*, *41*, 224-248. https://doi.org/https://doi.org/10.1344/der.2022.41.224-248

Curto, G., Jojoa Acosta, M. F., Comim, F., & Garcia-Zapirain, B. (2024). Are AI systems biased against the poor? A machine learning analysis using Word2Vec and GloVe embeddings. *AI and Society*, *39*(2), 617-632. https://doi.org/10.1007/S00146-022-01494-Z/TABLES/3

Eubanks, V. (2018). *Automating Inequality: How High-Tech Tools Profile, Police, and Punish the Poor*. St Martin's Press.

Fosch-Villaronga, E., & Poulsen, A. (2022). *Diversity and Inclusion in Artificial Intelligence* (pp. 109-134). T.M.C. Asser Press. https://doi.org/10.1007/978-94-6265-523-2_6

García Peñalvo, F. J., Llorens-Largo, F., & Vidal, J. (2023). La nueva realidad de la educación ante los avances de la inteligencia artificial generativa. *RIED-Revista Iberoamericana de Educación a Distancia*, *27*(1), 9-39. https://doi.org/10.5944/ried.27.1.37716

George, A. (2025). *Thwarting bias in AI systems*. Carnegie Mellon University, College of Engineering. https://engineering.cmu.edu/news-events/news/2018/12/11-datta-proxies.html

Goedhart, N. S., Verdonk, P., & Dedding, C. (2022). "Never good enough." A situated understanding of the impact of digitalization on citizens living in a low socioeconomic position. *Policy & Internet*, *14*(4), 824-844. https://doi.org/10.1002/poi3.315

Grau-Rebollo, J., García-Tugas, L., & García-García, B. (2024). Induced vulnerability: the consequences of racialization for African women in an emergency shelter in Catalonia (Spain). *Ethnic and Racial Studies*, *47*(7), 1510-1527. https://doi.org/10.1080/01419870.2023.2289135

Hermann, E., De Freitas, J., & Puntoni, S. (2025). Reducing prejudice with counter stereotypical AI. *Consumer Psychology Review*, *8*(1), 75-86. https://doi.org/10.1002/arcp.1102

Mack, K. A., Qadri, R., Denton, R., Kane, S. K., & Bennett, C. L. (2024). "They only care to show us the wheelchair": disability representation in text-to-image AI models. *Proceedings of the CHI Conference on Human Factors in Computing Systems*, 1-23. https://doi.org/10.1145/3613904.3642166

Makhortykh, M., Urman, A., & Ulloa, R. (2021). Detecting Race and Gender Bias in Visual Representation of AI on Web Search Engines. *Communications in Computer and Information Science*, *1418*, 36-50. https://doi.org/10.1007/978-3-030-78818-6_5

Mickel, J. (2024). Racial/Ethnic Categories in AI and Algorithmic Fairness: Why They Matter and What They Represent. *2024 ACM Conference on Fairness, Accountability, and Transparency, FAccT 2024*, 2484-2494. https://doi.org/10.1145/3630106.3659050/SUPPL_FILE/RACIAL_CLASSIFICATION_IN_AI_AND_ALGORITHMIC_FAIRNESS_APPENDIX.PDF

Noble, S. U. (2018). *Algorithms of Oppression. How Search Engines Reinforce Racism*. New York University Press.

Olatunji Akinrinola, Chinwe Chinazo Okoye, Onyeka Chrisanctus Ofodile, & Chinonye Esther Ugochukwu. (2024). Navigating and reviewing ethical dilemmas in AI development: Strategies for transparency, fairness, and ac-

countability. *GSC Advanced Research and Reviews*, *18*(3), 050-058. https://doi.org/10.30574/gscarr.2024.18.3.0088

O'Neil, C. (2016). *Weapons of math destruction: how big data increases inequality and threatens democracy*. Crown Publishing Group.

Pepe, A. (2023). Using Image-Based Research Methods in Vulnerable Populations as a Culturally Sensitive Approach: Ethical and Methodological Aspects. En D. Villa & F. Zuccoli (Eds.), *Proceedings of the 3rd International and Interdisciplinary Conference on Image and Imagination* (pp. 11-17). Springer. https://doi.org/10.1007/978-3-031-25906-7_2

Raji, I. D., & Buolamwini, J. (2023). Actionable Auditing: Investigating the Impact of Publicly Naming Biased Performance Results of Commercial AI Products. *Communications of the ACM*, *66*(1), 101-108.

Roselli, D., Matthews, J., & Talagala, N. (2019). Managing Bias in AI. *Companion Proceedings of The 2019 World Wide Web Conference*, 539-544. https://doi.org/10.1145/3308560.3317590

Sandvig, C., Hamilton, K. , Karahalios, K., & Langbort, C. (2014). *Auditing Algorithms: Research Methods for Detecting Discrimination on Internet Platforms*.

Schultz, M. D., Clegg, M., Hofstetter, R., & Seele, P. (2024). Algorithms and dehumanization: a definition and avoidance model. *AI & SOCIETY*. https://doi.org/10.1007/s00146-024-02123-7

Shahvaroughi Farahani, M., & Ghasemi, G. (2024). Artificial Intelligence and Inequality: Challenges and Opportunities. *Qeios*. https://doi.org/10.32388/7HWUZ2

Stahl, B. C., Rodrigues, R., Santiago, N., & Macnish, K. (2022). A European Agency for Artificial Intelligence: Protecting fundamental rights and ethical values. *Computer Law & Security Review*, *45*, 1-25. https://doi.org/10.1016/j.clsr.2022.105661

Sydoruk, T. (2024). How AI can reduce unemployment rate among vulnerable population of new immigrants: Assimilation issues resolved with AI. *Economics & Education*, *9*(2), 20-25. https://doi.org/10.30525/2500-946X/2024-2-3

van Toorn, G., & Scully, J. L. (2024). Unveiling algorithmic power: exploring the impact of automated systems on disabled people's engagement with social services. *Disability & Society*, *39*(11), 3004-3029. https://doi.org/10.1080/09687599.2023.2233684

Wang, C., Boerman, S., de Vreese, K., Kroon, A., & Möller, J. (2024). The artificial intelligence divide: Who is the most vulnerable? *New Media & Society*. https://journals.sagepub.com/doi/full/10.1177/14614448241232345

West, S., Whittaker, M., & Crawford, K. (2019). Discriminating systems: Gender, Race, and Power in AI. En *AI Now Institute* (Número April). https://ainowinstitute.org/wp-content/uploads/2023/04/discriminatingsystems.pdf

Zajko, M. (2021). Conservative AI and social inequality: conceptualizing alternatives to bias through social theory. *AI & SOCIETY*, *36*(3), 1047-1056. https://doi.org/10.1007/s00146-021-01153-9