

# INTELIGENCIA ARTIFICIAL PARA LA INCLUSIÓN Y LA EQUIDAD DIGITAL

**Formación Emprendedora  
para La Generación de Ingresos  
Em Comunidades Vulnerables:**

**El Uso De La Inteligencia Artificial  
Como Recurso Educativo En El Tercer  
Sector**

PÀG. 7

**Alfabetización digital algorítmica  
y búsqueda de empleo:**

**una revisión de la literatura**

PÀG. 15

**Com podeu saber si feu  
un ús responsable de la vostra  
aplicació d'ia?**

PÀG. 23

**La interseccionalidad  
como herramienta analítica**

**para la auditoría de sesgos en  
inteligencia artificial y modelos  
de lenguaje a gran escala (LLM)**

PÀG. 31

**Justicia climática  
en tiempos de IA:**

**entre la rapidez, el saber científico  
y los desafíos de la alfabetización  
algorítmica**

PÀG. 43

**Entre la promesa y la amenaza:  
la inteligencia artificial generativa  
en la educación universitaria**

PÀG. 53

**Inteligencia Artificial  
en la universidad:**

**percepciones del profesorado  
sobre su impacto en las competencias  
investigadoras y docentes**

PÀG. 59

## Equipo editorial

### DIRECCIÓN

Núria Valls Carol

### EQUIPO EDITOR

Caterine Fagundes Vila

*EDUCADORA E INVESTIGADORA. DOCTORA EN MÉTODOS DE INVESTIGACIÓN Y DIAGNÓSTICO EN EDUCACIÓN POR LA UNIVERSIDAD DE BARCELONA.*

Francisco Javier Poleo Gutiérrez

*LICENCIADO EN DERECHO Y EN COMUNICACIÓN ORGANIZACIONAL POR LA UNIVERSIDAD DE MÁLAGA. PRESIDENTE DE INCIDE.*

María Teresa Castilla Mesa

*PROFESORA E INVESTIGADORA TITULAR DE LA UNIVERSIDAD DE MÁLAGA, DEL DEPARTAMENTO DE DIDÁCTICA Y ORGANIZACIÓN ESCOLAR.*

Cecilia Seabra Gomes da Silva

*MESTRE EN COMUNICACIÓN SOCIAL POR LA UNIVERSIDAD DEL ESTADO DE RIO DE JANEIRO/ BRASIL.*

### COMITÉ CIENTÍFICO

José M<sup>a</sup> Fernández Batanero

*UNIVERSIDAD DE SEVILLA (ESPAÑA)*

Julio Cabero Almenara

*UNIVERSIDAD DE SEVILLA (ESPAÑA)*

Miquel Essomba Gilabert

*UNIVERSIDAD AUTÓNOMA DE BARCELONA (ESPAÑA)*

Margarita Aravena Gaete

*UNIVERSIDAD ANDRÉS BELLO. (CHILE)*

Damarys Roy

*UNIVERSIDAD ANDRÉS BELLO. (CHILE)*

Tiberio Feliz Murias

*UNIVERSIDAD NACIONAL DE EDUCACIÓN A DISTANCIA (ESPAÑA)*

Giovanni Arduini

*UNIVERSITÀ DE CASSINO (ITALIA)*

Diletta Chiusaroli

*UNIVERSITÀ DE CASSINO (ITALIA)*

Fabrizio Pizzi

*UNIVERSITÀ DE CASSINO (ITALIA)*

Víctor Manuel Martín Solbes

*UNIVERSIDAD DE MÁLAGA (ESPAÑA)*

María Teresa Castilla Mesa

*UNIVERSIDAD DE MÁLAGA (ESPAÑA)*

Kilian Wirthwein Vega

*UNIVERSIDAD CARLOS III (ESPAÑA)*

Mateus Saraiva

*UNIVERSIDAD FEDERAL DO RIO GRANDE SUL (BRASIL)*

Katia Cristian Puente Muniz

*UNIVERSIDAD VEIGA DE ALMEIDA (BRASIL)*

Raquel Poch Pallarols

*UNIVERSIDAD DE BARCELONA (ESPAÑA)*

## Índice

1. Bienvenida	
2. Presentación	
3. Formación emprendedora para la generación de ingresos em comunidades vulnerables: el uso de la inteligencia artificial como recurso educativo en el tercer sector .....	7
<i>Tiago Cortinaz da Silva y Douglas Severo Ferreira</i>	
4. Alfabetización digital algorítmica y búsqueda de empleo: una revisión de la literatura.....	15
<i>Stefano De Marco, Guillaume Dumont, Ellen Helsper y Daniel Calderón-Gómez</i>	
5. Com podeu saber si feu un ús responsable de la vostra aplicació d'IA?.....	23
<i>Arlet Brufau i Centelles y Albert Sabater Coll</i>	
6. La interseccionalidad como herramienta analítica para la auditoría de sesgos en inteligencia artificial y modelos de lenguaje a gran escala (LLM) .....	31
<i>José María Regalado López</i>	
7. Justicia climática en tiempos de IA: entre la rapidez, el saber científico y los desafíos de la alfabetización algorítmica .....	43
<i>Anne Morais y Cláudia Giesel</i>	
8. Entre la promesa y la amenaza: la inteligencia artificial generativa en la educación universitaria.....	53
<i>Albert García-Arnau y Carla Barrio</i>	
9. Inteligencia Artificial en la universidad: percepciones del profesorado sobre su impacto en las competencias investigadoras y docentes.....	59
<i>Yaritza Garcés-Delgado, Arminda Suárez-Perdomo y Carmen Nuria Arvelo-Rosales</i>	

# Bienvenida

---

Queridas lectoras y lectores,

Es un placer daros la bienvenida al segundo número de la *Revista de Tecnología para la Inclusión Social (RETIS)*. Con esta nueva edición, reafirmamos el compromiso de **Fundación Esplai** y del **Observatorio de las Brechas Digitales** con la construcción de una sociedad digital más justa, inclusiva y equitativa. En un momento histórico en el que las tecnologías emergentes, y en especial la Inteligencia Artificial, están transformando profundamente nuestras formas de vivir, aprender, trabajar y relacionarnos, creemos más necesario que nunca impulsar un espacio de reflexión crítica, conocimiento riguroso y diálogo abierto desde el Tercer Sector y en colaboración con el ámbito académico, el sector empresarial, la administración pública y la ciudadanía.

Este número está dedicado a la “Inteligencia Artificial para la Inclusión y la Equidad Digital”. Con él, queremos alejarnos de los discursos extremos, ya sean los alarmistas o los excesivamente entusiastas, para centrarnos en un análisis sereno y comprometido con las personas y los territorios más vulnerables. Las contribuciones que aquí se recogen exploran el impacto de la IA en ámbitos tan diversos como la educación, el acceso al empleo, la alfabetización algorítmica o la participación democrática, con especial atención a las oportunidades que pueden surgir cuando estas tecnologías se diseñan e implementan con enfoque inclusivo, ético y socialmente responsable.

A través de esta publicación, seguimos avanzando en nuestros objetivos: dar visibilidad a experiencias innovadoras que emplean la tecnología como herramienta de transformación social, analizar políticas públicas que busquen reducir las desigualdades digitales, presentar casos de éxito que inspiren nuevas prácticas en el trabajo comunitario, reflexionar críticamente sobre los desafíos que plantea el uso de la tecnología en contextos de vulnerabilidad y contribuir con investigaciones empíricas al debate sobre el acceso, el uso y el impacto real de la tecnología en nuestras sociedades.

RETIS quiere ser, una vez más, una invitación a pensar colectivamente. A conectar saberes y prácticas. A generar pensamiento desde una mirada científica, pero profundamente comprometida con los derechos humanos y la justicia social. Os animamos a seguir acompañándonos en este viaje de aprendizaje compartido, construyendo, debatiendo y mejorando, para hacer de la inclusión digital no solo un objetivo, sino una realidad tangible.

Atentamente,

**Núria Valls Carol**

*Directora del Observatorio de las Brechas Digitales*

# Presentación

El segundo número de la Revista de Tecnología para la Inclusión Social, titulado “Inteligencia Artificial para la Inclusión y la Equidad Digital”, tiene como objetivo reflexionar sobre las transformaciones que las aplicaciones y herramientas de Inteligencia Artificial están produciendo en distintos ámbitos de la vida social, entre los que destacan el mundo del trabajo, el acceso y difusión del conocimiento y la información o el sistema educativo, así como los retos y dilemas éticos emergentes. Del mismo modo, más allá de los peligros y riesgos asociados a esta nueva revolución tecnológica, sobre lo que se ha escrito mucho en los últimos años, queremos poner el foco en las implicaciones y oportunidades de la aplicación de la Inteligencia Artificial desde la perspectiva de la equidad, la inclusión de los grupos vulnerables y la justicia social. Por ello, intentamos huir tanto de perspectivas catastrofistas y apocalípticas que predicen el fin de la sociedad debido al cambio tecnológico –algo que no es nuevo– como de las ensoñaciones ciberutópicas que ven en la Inteligencia Artificial la solución a todos nuestros problemas.

Así, como punto de partida, comenzamos el monográfico reflexionando sobre la irrupción de la Inteligencia Artificial en el ámbito profesional. En el primer artículo, Tiago Cortinaz da Silva y Douglas Severo Ferreira presentan una investigación sobre la utilización de la Inteligencia Artificial como recurso educativo en el tercer sector dentro de comunidades vulnerables. Particularmente, se enfatiza el potencial de las herramientas de IA en la personalización pedagógica orientada al emprendimiento, en la medición del impacto y en el fortalecimiento de la inteligencia colectiva de las comunidades, evitando las ópticas extractivistas de los saberes locales y comunitarios. Continuando con el énfasis en la influencia de los avances tecnológicos dentro del mercado de trabajo, en el segundo artículo Stefano De Marco, Guillaume Dumont, Ellen Helsper y Daniel Calderón-Gómez plantean una revisión de la bibliografía sobre la alfabetización digital relativa a la gestión de los algoritmos que rigen el funcionamiento de las plataformas y herramientas de búsqueda de empleo. Específicamente, se constata la emergencia una nueva dimensión de la brecha digital relacionada con el conocimiento y el manejo de los algoritmos presentes en estos servicios digitales, lo que puede suponer una barrera a la hora de gestionar la imagen digital y las oportunidades laborales de la ciudadanía.

Transitando del mercado de trabajo al ámbito educativo, en el tercer artículo Arlet Brufau i Centelles y Albert Sabater Coll desarrollan un análisis del PIO, una herramienta de autoevaluación sobre el uso de aplicaciones de Inteligencia Artificial en el contexto educativo desde una perspectiva ética, que tenga en cuenta la transparencia, equidad, seguridad, responsabilidad, privacidad, autonomía y sostenibilidad. Profundizando en los problemas éticos y sociales de la IA, en el cuarto artículo José María Regalado López presenta una revisión sistemática de la bibliografía sobre los sesgos presentes en el uso de Inteligencia Artificial y modelos de lenguaje a gran escala (LLM), proponiendo la necesidad de optar por un enfoque teórico interseccional para auditar de forma efectiva los sesgos presentes en este tipo de herramientas con el fin de garantizar la equidad y la justicia social. A continuación, en el quinto artículo, Anne Morais y Cláudia Giesel desarrollan un análisis crítico de los discursos producidos por aplicaciones conversacionales de Inteligencia Artificial (como ChatGPT y Gemini) sobre temáticas de justicia climática, destacando los problemas de rigor, profundidad y fundamentación teórica de estos textos, lo que incide en la necesidad de realizar un uso crítico de las aplicaciones de IA, basado en una correcta alfabetización digital y algorítmica de la población.

Finalmente, los últimos dos artículos se centran específicamente en el sistema educativo universitario, uno de los contextos en los que la utilización de la Inteligencia Artificial ha generado controversia en los últimos años. En el sexto artículo, Albert García-Arnau y Carla Barrio reflexionan sobre los cambios que la inteligencia artificial generativa (IAG) está produciendo en la educación universitaria, así como los retos, oportunidades e implicaciones éticas de la irrupción de este tipo de herramientas. A continuación, en el séptimo artículo, Yaritza Garcés-Delgado, Arminda Suárez-Perdomo y Carmen Nuria Arvelo-Rosales presentan los resultados de un estudio cuantitativo a profesorado universitario sobre el conocimiento y uso de herramientas de Inteligencia Artificial en el aula, así como los miedos y experiencias derivados de su uso, con el objetivo de promover una integración crítica y responsable de la IA en la educación superior.

En resumen, los siete artículos incluidos en el presente monográfico abordan diferentes dimensiones, contextos y perspectivas teóricas y empíricas de aplicación de la Inteligencia Artificial en la actualidad, pero tienen como punto en común su carácter práctico, centrado en el reconocimiento de los desafíos que plantea el cambio tecnológico, pero también de sus potencialidades y oportunidades para mejorar la inclusión y condiciones de vida de la población.

Daniel Calderón Gómez  
Universidad Complutense de Madrid



# Formación Emprendedora para la Generación de Ingresos em Comunidades Vulnerables: el Uso de la Inteligencia Artificial como Recurso Educativo en el Tercer Sector

Entrepreneurial Education for Income Generation in Vulnerable Communities: Using Artificial Intelligence as an Educational Tool in the Third Sector

Capacitação em Empreendedorismo para Geração de Renda em Comunidades Vulneráveis: o Uso da Inteligência Artificial como Recurso Educacional no Terceiro Setor

Tiago Cortinaz da Silva  
 Douglas Severo Ferreira

## PALABRAS CLAVE:

Inteligencia Artificial  
 Emprendimiento Social  
 Educación Popular  
 Inclusión Digital  
 TercerSector.

## RESUMEN:

Este artículo presenta una investigación teórico-empírica sobre la aplicación de la Inteligencia Artificial (IA) como recurso educativo en programas de capacitación en emprendimiento orientados a la generación de ingresos en comunidades vulnerables, con énfasis en la actuación del Tercer Sector. La investigación parte de la metodología *By Necessity - Plan de Negocios por Necesidad* e integra marcos de referencia de educación popular, capital cultural, desarrollo humano y sociedad en red. El análisis combina una revisión bibliográfica y el análisis documental de los programas del Instituto Besouro (2016-2025). El estudio propone un modelo híbrido - tecnología y mediación comunitaria - que posiciona la IA como herramienta de personalización pedagógica, medición de impacto e inteligencia colectiva, preservando el protagonismo del saber empírico local. Asimismo, se discuten los desafíos éticos, de infraestructura y de gobernanza de datos necesarios para una implementación inclusiva y transformadora.

## KEYWORDS

Artificial Intelligence  
 Social Entrepreneurship  
 Popular Education  
 Digital Inclusion  
 Third Sector

## ABSTRACT:

This article presents a theoretical-empirical investigation on the application of Artificial Intelligence (AI) as an educational resource in entrepreneurship training programs aimed at income generation in vulnerable communities, with an emphasis on the role of the Third Sector. The research is based on the *By Necessity - Business Plan by Necessity* methodology and integrates frameworks of popular education, cultural capital, human development, and network society. The analysis combines a bibliographic review and a documentary analysis of the Besouro Institute's programs (2016-2025). The study proposes a hybrid model - technology and community mediation - that positions AI as a tool for pedagogical personalization, impact measurement, and collective intelligence, while preserving the protagonism of local empirical knowledge. It also discusses the ethical, infrastructural, and data governance challenges required for an inclusive and transformative implementation.

## PALAVRAS-CHAVE:

Inteligência Artificial  
 Empreendedorismo Social  
 Educação Popular  
 Inclusão Digital  
 Terceiro Setor.

## RESUMO:

Este artigo apresenta uma investigação teórico-empírica sobre a aplicação da Inteligência Artificial (IA) como recurso educacional em programas de capacitação em empreendedorismo voltados à geração de renda em comunidades vulneráveis, com foco na atuação do Terceiro Setor. A pesquisa parte da metodologia *By Necessity - Plano de Negócios por Necessidade* e integra referenciais de educação popular, capital cultural, desenvolvimento humano e sociedade em rede. A análise combina revisão bibliográfica e análise documental dos programas do Instituto Besouro (2016-2025). O estudo propõe um modelo híbrido - tecnologia e mediação comunitária - que posiciona a IA como ferramenta de personalização pedagógica, mensuração de impacto e inteligência coletiva, preservando o protagonismo do saber empírico local. Discute-se, ainda, os desafios éticos, de infraestrutura e de governança de dados necessários para uma implementação inclusiva e transformadora.

**CÓMO CITAR:** Cortinaz da Silva, T., & Ferreira, D. S. (2025). Formación emprendedora para la generación de ingresos en comunidades vulnerables: El uso de la inteligencia artificial como recurso educativo en el tercer sector. *RETIS. Revista de Tecnología para la Inclusión Social*, 2(1), 7-15, DOI: 10.70664/retis.v2i1.001

## 1. Introdução

A emergência do empreendedorismo nas periferias e nas comunidades vulneráveis constitui um fenômeno multifacetado, atravessado por dimensões econômicas, sociais, de raça, de gênero e simbólicas. Trata-se tanto de uma resposta às ausências do mercado formal de trabalho e do Estado, quanto de uma expressão cultural e de resistência identitária (Lima, 2016; 2017; Bourdieu, 1979). Nessas territorialidades, o “fazer econômico” é indissociável da experiência cotidiana, da rede de sociabilidades e da necessidade de sobrevivência, que se convertem em potenciais de inovação e criatividade. O empreendedorismo periférico, portanto, não emerge apenas como alternativa econômica, mas como forma de agência social, um modo de reconfigurar a exclusão socioeconômica em potência produtiva e cultural.

No Brasil, o papel do Terceiro Setor tem sido central na promoção de práticas empreendedoras inclusivas, principalmente em regiões caracterizadas pela informalidade econômica e pela escassez de políticas públicas efetivas (Lima, 2017; 2018). Nesse contexto, metodologias como a *By Necessity* oferecem um arcabouço pedagógico e prático para a construção de negócios sustentáveis a partir dos saberes empíricos dos próprios participantes. Diferente das abordagens tradicionais de ensino de empreendedorismo, a *By Necessity* parte da valorização do “saber de cada um”, do conhecimento local e da experiência comunitária como capital cultural legítimo (Bourdieu, 1979; Lima, 2017). Assim, transforma habilidades locais - como culinária, costura, mecânica ou estética - em oportunidades concretas de geração de renda, conectando teoria administrativa à prática social e cultural dos territórios.

A efetividade dessa metodologia se comprova pela atuação de instituições como o Instituto Besouro de Fomento Social, que já impactou mais de 100 mil empreendedores em 29 países, alcançando periferias urbanas, comunidades indígenas, quilombolas e povos tradicionais da Amazônia. Os programas desenvolvidos pelo Instituto Besouro demonstram que a capacitação empreendedora pode ser um instrumento de transformação social quando articulada com princípios de inclusão, identidade e empoderamento comunitário.

No entanto, a difusão de tecnologias digitais e a expansão da inteligência artificial (IA) na educação abrem novas possibilidades e desafios para o campo do empreendedorismo social. Ferramentas de IA podem ampliar o alcance dos cursos de capacitação para geração de renda, personalizar processos de aprendizagem e fortalecer a autonomia dos participantes, desde que implementadas de forma ética e contextualizada. **A questão central que orienta este ensaio é, portanto, como integrar a IA à capacitação empreendedora do Terceiro Setor sem comprometer o caráter comunitário, empírico e identitário das práticas de ensino.**

A hipótese que se apresenta é que a integração da IA, quando guiada por princípios de inclusão digital, pedagogia comunitária e respeito ao capital cultural local, pode potencializar - e não substituir - a pedagogia empírica das periferias. A articulação entre o humano e o tecnológico pode ampliar o alcance das metodologias sociais, fortalecer o acompanhamento personalizado dos beneficiários e promover redes de aprendizagem colaborativas. Assim, este ensaio busca compreender os impactos da integração da IA na aprendizagem, na autonomia e na geração de renda em contextos vulneráveis, propondo um modelo de **Inteligência Artificial Educacional de Base Comunitária**, inspirado nas práticas da metodologia *By Necessity* e nas teorias de Bourdieu (1979) sobre *habitus* e capital cultural.

## 2. Revisão de Literatura: O uso da Inteligência Artificial na Educação Empreendedora

A incorporação da Inteligência Artificial (IA) na educação tem transformado profundamente os modelos de ensino-aprendizagem, inaugurando uma nova fase do que se convencionou chamar de Educação 4.0 - uma abordagem que integra tecnologias digitais, análise de dados e personalização pedagógica (Schwab, 2016; Luckin et al., 2016). Nesse cenário, a IA não apenas automatiza processos, mas redefine o papel do conhecimento e do educador, propondo um ecossistema de aprendizagem mais conectado, adaptativo e centrado no aluno (Holmes et al., 2019).

Na educação empreendedora, o uso da IA desponta como um campo emergente de grande potencial. Segundo Marques e Batista (2023), a IA tem contribuído para o desenvolvimento de competências empreendedoras ao oferecer simulações de negócios, diagnósticos preditivos e suporte à tomada de decisão com base em dados. Plataformas baseadas em *machine learning* são capazes de analisar perfis de estudantes e sugerir trilhas de aprendizado personalizadas, estimulando a autonomia e o protagonismo - valores centrais à formação empreendedora contemporânea.

A IA, quando integrada a metodologias ativas, possibilita a criação de ambientes de aprendizagem inteligentes, nos quais os alunos podem testar hipóteses, planejar negócios simulados e receber *feedbacks* automatizados e contextualizados. Luckin et al. (2016) destacam que o potencial da IA na educação está na sua capacidade de ampliar o alcance e a qualidade do ensino, reduzindo desigualdades e democratizando o acesso à informação. A personalização é um dos pilares do ensino mediado por IA. Segundo Anderson et al. (2020), algoritmos de aprendizagem permitem analisar o ritmo, o estilo e as preferências de aprendizagem dos alunos, oferecendo conteúdos e desafios ajustados às suas necessidades. Essa lógica se aproxima do ensino contextualizado proposto por Paulo Freire (1996), no qual a educação deve partir da realidade concreta do sujeito e promover a leitura crítica do mundo.

Para além da eficiência técnica, o uso da IA na educação empreendedora deve ser analisado como um instrumento de inclusão social e democratização do conhecimento. Boaventura de Sousa Santos (2018) e Pierre Bourdieu (1979) alertam que a tecnologia, quando dissociada da cultura e do capital social dos territórios, tende a reproduzir desigualdades. Por isso, a integração entre IA e pedagogia comunitária deve respeitar a diversidade cultural e os saberes locais. Nesse sentido, a IA aplicada à educação empreendedora torna-se um vetor de desenvolvimento local, pois contribui para mapear demandas, gerar dados sobre microempreendimentos e apoiar políticas públicas de inclusão produtiva (UNESCO, 2021).

Entre os principais desafios apontados pela literatura, estão: a necessidade de formação docente para lidar com ferramentas de IA (Holmes et al., 2019); o risco de vieses algorítmicos reproduzirem desigualdades sociais (O'Neil, 2016); e a urgência de políticas públicas que promovam acesso equitativo às tecnologias educacionais (UNESCO, 2021). Contudo, o horizonte é promissor: a IA tem o potencial de transformar a educação empreendedora em um processo de inovação social contínua, no qual dados, saberes e experiências se unem para criar soluções locais com impacto global.

### 3. Empreendedorismo por necessidade (By Necessity) e saber empírico

O empreendedorismo por necessidade constitui uma resposta direta às lacunas estruturais do mercado de trabalho formal e às ausências do Estado em prover oportunidades equitativas. Segundo o SEBRAE (2023), a maior parte dos empreendedores brasileiros inicia atividades autônomas por ausência de alternativas formais de emprego. Essa dinâmica, muitas vezes desconsiderada pelas análises econômicas tradicionais, revela um empreendedorismo de sobrevivência que, embora informal, é sustentado por racionalidades próprias, estratégias de rede e saberes locais (Lima, 2016; 2018).

Vinicius Mendes Lima, ao desenvolver a metodologia *By Necessity*, propõe que esse fazer econômico nas periferias seja compreendido como forma legítima de produção de conhecimento. Inspirado no conceito bourdieusiano de capital cultural (Bourdieu, 1979), Lima reinterpreta o saber empírico como capital empírico, ou seja, a acumulação de experiências, práticas e repertórios de ação construídos na vida cotidiana, que servem de base para a criação de negócios sustentáveis e socialmente enraizados. O capital empírico, nesse sentido, opera como uma forma de resistência simbólica: transforma aquilo que o sistema social considera “deficiência” - a falta de formação acadêmica formal - em potência produtiva e criativa. Tal conceito reaproxima a economia popular das teorias sociológicas da prática, situando o empreendedor periférico como agente de transformação e produtor de valor simbólico e econômico em territórios de exclusão (Lima, 2017; Bourdieu, 1979).

A pedagogia crítica de Paulo Freire (1987) oferece o alicerce teórico para compreender a formação empreendedora como um processo de libertação e conscientização. Freire defende que as pessoas se educam em comunhão, a partir da sua relação com o mundo. Essa perspectiva rompe com a lógica bancária da educação, substituindo a transmissão de conteúdos pela problematização da realidade vivida. A metodologia *By Necessity* incorpora esse princípio ao reconhecer o aluno como protagonista de sua trajetória. O processo formativo parte do que o indivíduo já sabe fazer, valorizando seu contexto, seus afetos e suas condições de vida como matéria-prima para o aprendizado (Lima, 2017; 2020). Trata-se de uma pedagogia dialógica e horizontal, na qual o educador é também mediador, e o conhecimento é construído coletivamente.

Essa abordagem aproxima-se da concepção freireana de educação popular, que busca a emancipação por meio da práxis - ação e reflexão transformadora - e resgata a dignidade do saber local. Nas periferias, tal perspectiva tem se mostrado especialmente eficaz, pois a aprendizagem ocorre a partir da prática empreendedora e do reconhecimento identitário, e não da imposição de modelos externos de gestão.

Na perspectiva de Amartya Sen (2000) e Martha Nussbaum (2011), o desenvolvimento humano é definido pela ampliação das *capabilities* - as liberdades reais das pessoas para escolher e agir. Mais do que crescimento econômico, o desenvolvimento implica aumentar as oportunidades de vida digna, de autonomia e de participação cidadã. No campo do empreendedorismo social, isso significa oferecer instrumentos para que os indivíduos ampliem suas capacidades de decisão; tais como: formalizar um negócio, gerir recursos, acessar crédito e inovar em seus territórios. A metodologia *By Necessity* dialoga diretamente com esse enfoque, pois sua meta não é apenas criar empreendimentos, mas expandir as possibilidades de existência de quem empreende - sobretudo entre jovens, mulheres e populações racializadas atendidas por programas e projetos do Instituto Besouro.

Nesse contexto, a **Inteligência Artificial Educacional de Base Comunitária** surge como uma ferramenta potencializadora das capacidades humanas, desde que usada de forma crítica e ética. Quando inserida em processos de formação empreendedora, a IA pode ampliar o acesso a informações, personalizar trajetórias de aprendizagem e fortalecer a autonomia de decisão - sem substituir o vínculo humano e comunitário que sustenta a aprendizagem social.

### 4. Sociedade em rede, inteligência coletiva e Educação 4.0

A partir das análises de Manuel Castells (1999), vivemos em uma sociedade em rede, na qual as relações de poder, produção e conhecimento são estruturadas em fluxos informacionais. Nessa configuração, o conhecimento torna-se o principal recurso estratégico e a inclusão digital passa a ser um elemento de cidadania e de emancipação social. Pierre Lévy (1998) amplia essa perspectiva ao introduzir o conceito de inteligência coletiva, emergente da cooperação entre pessoas interconectadas, potencializada pela partilha de experiências e saberes mediados pelas tecnologias digitais.

No campo educacional, essas ideias dialogam com o paradigma da Educação 4.0, que, segundo Schwab (2016) e Luckin et al. (2016), articula cultura *maker*, aprendizagem por projetos e personalização do ensino via tecnologias emergentes - incluindo a inteligência artificial (IA), a Internet das Coisas (*IoT - Internet of Things*) e a análise de dados educacionais. A educação, nessa perspectiva, não é um espaço de mera transmissão de conhecimento, mas sim um ecossistema de aprendizagem colaborativa, voltado à criatividade, à resolução de problemas e à inovação social.

Autores como Paulo Freire (1970, 1996) e Boaventura de Sousa Santos (2018) reforçam que a educação emancipatória precisa considerar o saber popular como fonte legítima de conhecimento. Essa concepção se conecta diretamente à pedagogia comunitária defendida por Lima (2017, 2018, 2021), especialmente em sua metodologia *By Necessity*, que reconhece o saber empírico e as vivências dos sujeitos periféricos como ponto de partida para a construção de saberes aplicáveis à vida real. A proposta metodológica *By Necessity* traduz essa pedagogia em prática social e econômica. Baseada em princípios da aprendizagem significativa, a metodologia parte das competências que cada participante já possui - culinária, costura, estética, marcenaria, entre outras - para estruturar um plano de negócios sustentável, combinando saberes empíricos e gestão estratégica. Essa abordagem evidencia

o papel da inteligência coletiva e da cognição distribuída (Engeström, 2001; Wenger, 1998) como pilares do aprendizado em rede e da criação de valor comunitário.

Ao integrar IA, conectividade e saber comunitário, vislumbra-se um ecossistema de aprendizagem descentralizado, em que o conhecimento é cocriado e disseminado entre pares, docentes e sistemas inteligentes. Nessa configuração, a IA deve atuar como mediadora da inteligência coletiva, e não como substituta do educador ou do vínculo humano que ancora o processo formativo. Essa perspectiva reflete a obra de Bourdieu (1979), ao considerar o capital cultural como elemento estruturante das desigualdades de acesso ao saber e, portanto, um desafio a ser superado pela educação tecnológica inclusiva. A união desses referenciais sustenta uma epistemologia da educação empreendedora que reconhece o território e a experiência como *locus* de produção de conhecimento. Trata-se de uma educação em rede, por necessidade e para a autonomia, capaz de gerar transformação social, inclusão produtiva e cidadania digital.

Apesar de seu potencial inclusivo, a IA carrega riscos de reprodução das desigualdades sociais e epistemológicas. Autores como Selwyn (2019) e Williamson (2021) alertam que algoritmos educacionais podem reforçar hierarquias existentes ao traduzirem vieses socioeconômicos em padrões de decisão automatizados. Para mitigar esses riscos, organismos internacionais como a UNESCO (2022) e a OCDE (2019) estabeleceram princípios éticos para o uso da IA; a saber: transparência, justiça, responsabilização, inclusão e respeito à privacidade. No contexto da educação social, tais princípios devem ser reinterpretados sob a ótica da soberania de dados comunitários, garantindo que as tecnologias utilizadas respeitem a autonomia dos territórios e fortaleçam à autodeterminação digital. Dessa forma, a integração entre a metodologia *By Necessity* e a IA educacional exige uma abordagem crítica e decolonial da tecnologia - que reconheça o valor dos saberes periféricos e evite transformar a inclusão digital em nova forma de dependência estrutural. O desafio é fazer da IA uma ferramenta de emancipação e não de controle, mantendo o protagonismo humano como eixo central do processo de desenvolvimento.

## 5. Ensaio e Considerações Iniciais

A coleta e análise de dados baseou-se em dois eixos principais: revisão de literatura e análise documental. A análise documental consistiu no estudo de materiais pedagógicos produzidos pelo Instituto Besouro entre 2016 e 2025 - incluindo os livros *A riqueza das favelas* (Lima, 2016); *By Necessity: Plano de negócios por necessidade* (Lima, 2017); e *Canvas das favelas* (Lima, 2018) - e relatórios de impacto social do Instituto Besouro. Esses documentos foram examinados para identificar fundamentos pedagógicos, indicadores de resultado e referências à tecnologia digital no processo formativo.

Os dados foram tratados por meio da análise de conteúdo (Bardin, 2011), em três etapas: 1) Pré-análise: organização e leitura flutuante do corpus empírico (documentos e observações); 2) Exploração do material: categorização temática de unidades de registro, com base em eixos teóricos definidos a priori - pedagogia crítica, empoderamento, uso de IA, capital empírico e inclusão digital; e 3) Tratamento e interpretação: triangulação entre dados empíricos e referenciais teóricos, buscando identificar padrões, convergências e contradições.

O processo analítico permitiu identificar quatro núcleos principais: (i) práticas pedagógicas eficientes na formação empreendedora; (ii) possibilidades concretas de integração da IA aos processos formativos; (iii) barreiras de infraestrutura e de letramento digital nas periferias; e (iv) impactos econômicos e socioculturais pós-capacitação, incluindo indicadores de geração de renda, formalização de negócios e fortalecimento de capital social local.

Os experimentos conceituais com plataformas de IA educacional - especialmente sistemas de *learning analytics* e tutores adaptativos - demonstraram alto potencial de personalização. Esses sistemas foram capazes de mapear perfis de aprendizagem, sugerir itinerários formativos específicos e antecipar dificuldades recorrentes, como precificação incorreta ou falhas no controle de caixa. A personalização aumentou a velocidade de aquisição de competências tanto técnicas (gestão financeira e marketing digital, por exemplo) quanto socioemocionais (liderança e comunicação, por exemplo). Esses resultados estão alinhados com a literatura de Educação 4.0 (Luckin et al., 2016; Schwab, 2016), que enfatiza o papel da tecnologia como mediadora da autonomia e não como substituta da interação humana. No contexto periférico, o diferencial está em adaptar os algoritmos para reconhecer linguagens populares, ritmos de aprendizagem variados e trajetórias de vida não lineares.

A aplicação da IA à mensuração de impacto social mostrou-se promissora para o Terceiro Setor. Sistemas baseados em *data analytics* permitiram acompanhar indicadores em tempo real, como evolução do faturamento, número de clientes, taxa de formalização, empregos gerados e cumprimento de metas individuais. Contudo, emergiu uma preocupação recorrente: a governança dos dados. A coleta de informações pessoais e empresariais exige protocolos claros de consentimento e segurança. Em consonância com a UNESCO (2022) e a OCDE (2019), propõe-se o uso de modelos de consentimento dinâmico e dados abertos comunitários, que garantam transparência e soberania local sobre as informações. Assim, a IA não se torna instrumento de controle, mas ferramenta de autogestão e *accountability* social.

A infraestrutura tecnológica continua sendo o principal gargalo para a implementação de soluções de IA nos territórios periféricos. A conectividade limitada (muitas vezes irregular, conhecida como “gato net”), o uso compartilhado de dispositivos móveis por diversas pessoas de uma mesma família e a falta de letramento digital são desafios a serem superados. Essas limitações indicam que políticas públicas de inclusão digital comunitária e parcerias com provedores locais de internet são pré-condições para qualquer avanço sustentável. Como destaca Castells (1999), a exclusão informacional se traduz em exclusão social. Portanto, a transformação digital do empreendedorismo periférico depende tanto de infraestrutura física quanto de letramento tecnológico e cultural, entendido como a capacidade crítica de uso e criação de tecnologias (Selwyn, 2019).

A principal lição deste ensaio é que a IA deve ser entendida como amplificadora das capacidades humanas e não como substituta das interações pedagógicas. A mediação dos educadores locais - pessoas que compartilham vivências territoriais com os alunos - é decisiva para contextualizar o uso da tecnologia e interpretar seus resultados. Essa perspectiva sustenta-se nos estudos

de Amartya Sen (2000), ao postular que o desenvolvimento humano ocorre quando tecnologias expandem liberdades e não quando as restringem. A IA deve servir à pedagogia comunitária e não o contrário.

A incorporação da IA nos programas sociais exige que ela se apoie em epistemologias periféricas - formas locais de produzir conhecimentos, linguagens e estéticas. **Modelos genéricos, treinados com dados de economias formais, não captam a complexidade dos mercados populares. Assim, é necessário treinar algoritmos com dados locais, representativos das realidades das favelas e das comunidades atendidas.** Essa abordagem decolonial da tecnologia (Mignolo, 2017) permite construir sistemas que respeitem a diversidade cultural e a linguagem dos usuários, validando seus saberes e modos de fazer. Em vez de impor padrões, a IA pode aprender com os territórios, tornando-se um espelho de suas inteligências coletivas (Lévy, 1998).

A consolidação de um ecossistema de IA socialmente justo requer modelos participativos de governança de dados. Recomenda-se a criação de comitês locais de ética e tecnologia, com representação comunitária, que deliberem sobre coleta, uso e compartilhamento de dados. Inspirado nos princípios da UNESCO (2022), propõe-se o desenvolvimento de protocolos de consentimento dinâmico - nos quais os empreendedores possam autorizar, revisar e revogar o uso de seus dados a qualquer momento. Tal modelo fortalece a soberania digital e reduz assimetrias de poder entre provedores tecnológicos e usuários vulneráveis.

A sustentabilidade técnica e financeira da **Inteligência Artificial Educacional de Base Comunitária** no Terceiro Setor depende de alianças multissetoriais. Parcerias entre ONGs, universidades, governos locais, provedores de conectividade e empresas de tecnologia podem viabilizar soluções escaláveis e de baixo custo. Modelos híbridos de financiamento - combinando financiamento público, patrocínios privados e assinaturas simbólicas dos próprios beneficiários - mostram-se promissores para garantir autonomia financeira sem comprometer o acesso gratuito. Além disso, iniciativas como o *BRDE Labs* e o *Programa VOA AMBEV* demonstram que a inovação social pode ser acelerada quando há colaboração entre setores público e privado em torno de objetivos de impacto.

Por fim, a IA se revela uma ferramenta poderosa para fomentar aprendizagem contínua e construção de ecossistemas locais de inovação. Plataformas inteligentes podem oferecer microcredenciais, mentorias automatizadas e itinerários formativos personalizados de evolução profissional. Entretanto, para que isso se consolide, é necessário investir em infraestruturas de aprendizagem comunitária - espaços *maker*, centros digitais e laboratórios de inovação social conectados às realidades territoriais. Esses espaços devem operar como nodos de uma rede de inteligência coletiva periférica, articulando tecnologia, cultura e empreendedorismo (Castells, 1999; Lévy, 1998). Em síntese, os resultados demonstram que a integração entre *By Necessity* e IA educacional é viável e desejável, desde que respeite o princípio central da metodologia: partir do que o aluno já sabe e da realidade que o cerca. Quando a tecnologia reconhece o valor do saber empírico e o amplia, ela deixa de ser instrumento de exclusão e passa a ser ferramenta de emancipação.

## 6. Considerações Finais e Recomendações

A Inteligência Artificial (IA) desponta como ferramenta estratégica para transformar os processos de capacitação empreendedora em contextos de vulnerabilidade. Quando bem implementada, ela potencializa três dimensões centrais do desenvolvimento humano: personalização da aprendizagem, mensuração precisa de impacto e fortalecimento da inteligência coletiva. No entanto, seu uso não pode ser reduzido à eficiência técnica - ele exige uma pauta política, ética e pedagógica comprometida com a inclusão digital, a justiça cognitiva e o protagonismo comunitário (UNESCO, 2022; Lima, 2018).

Este ensaio demonstra que a metodologia *By Necessity* oferece um modelo concreto de integração entre tecnologia, mediação humana e saber empírico. Em vez de substituir o educador, a IA pode ser configurada para ampliar a capacidade pedagógica das metodologias sociais, personalizando o acompanhamento e sistematizando aprendizados locais. O resultado é um ciclo virtuoso: o saber empírico alimenta a IA e a IA devolve à comunidade conhecimento estruturado e útil, ao ser aplicável.

Essa dinâmica confirma que o Terceiro Setor - pela sua proximidade territorial, capilaridade e legitimidade social - está em posição privilegiada para liderar a transição digital inclusiva. Organizações como o Instituto Besouro já demonstram que é possível articular inovação tecnológica com pedagogia popular, gerando impactos econômicos mensuráveis e transformações sociais duradouras. No entanto, para que essa integração se consolide em escala nacional e internacional, são necessárias ações coordenadas entre sociedade civil, Estado e iniciativa privada. As recomendações a seguir sintetizam diretrizes derivadas dos achados empíricos e teóricos deste estudo.

### 6.1. Recomendações Políticas

#### 6.1.1. Política Nacional de Inclusão Empreendedora Digital

Propõe-se a criação de uma política pública específica para integrar capacitação tecnológica, letramento digital e empreendedorismo de base comunitária. Essa política deve incluir recursos para infraestrutura (acesso à internet, dispositivos móveis etc.), formação de educadores locais e apoio a incubadoras sociais em territórios vulneráveis;

#### 6.1.2. Marco Ético-Comunitário para o Uso da IA Social

Inspirado nas diretrizes da UNESCO (2022) e da OCDE (2019), recomenda-se o desenvolvimento de um marco ético nacional que regulamente o uso da IA em programas sociais. O modelo deve priorizar soberania de dados comunitários, consentimento dinâmico, transparência algorítmica e representação popular nos comitês de governança tecnológica;

### 6.1.3. Fomento a Ecossistemas Locais de Inovação Popular

Incentivar a criação de centros de inovação comunitária, espaços *maker* e laboratórios de prototipagem social em favelas e periferias. Tais ecossistemas permitiriam o desenvolvimento de soluções tecnológicas locais, integrando empreendedores, educadores e pesquisadores em redes territoriais de aprendizagem e produção.

### 6.1.4. Integração com o Plano Nacional de Startups e o Sistema S

O empreendedorismo por necessidade deve ser reconhecido como parte da política de inovação brasileira. Recomenda-se integrar as metodologias educacionais - como a *By Necessity* - às ações do Plano Nacional de Startups e Empreendedorismo Jovem, criando linhas de crédito e aceleração específicas para negócios oriundos da base da pirâmide (Lima, 2016; 2018).

## 6.2. Recomendações Práticas e Institucionais

### 6.2.1. Formação de Facilitadores Digitais Comunitários

Capacitar educadores e agentes sociais comunitários para o uso ético e pedagógico da IA. Esses profissionais atuariam como mediadores culturais e tecnológicos, adaptando conteúdos e interpretando os dados gerados pelos sistemas inteligentes segundo a realidade local.

### 6.2.2. Governança Participativa de Dados

Implementar modelos de gestão compartilhada da informação, em que comunidades possam deliberar sobre o uso de dados gerados em suas formações. Essa governança deve envolver conselhos locais, ONGs, universidades e órgãos públicos, garantindo transparência e autonomia.

### 6.2.3. Monitoramento e Avaliação em Tempo Real

Desenvolver painéis de acompanhamento integrados, alimentados por IA, que permitam mensurar resultados socioeconômicos em tempo real - como variação de renda, formalização e geração de empregos. Tais sistemas podem servir de base para políticas públicas baseadas em evidências, fortalecendo a transparência do impacto social.

### 6.2.4. Intercâmbio Internacional de Metodologias

Estimular o compartilhamento de experiências do Terceiro Setor, como a implementação da metodologia *By Necessity* em territórios periféricos, com outros países como a Espanha, ampliando redes de cooperação entre o Terceiro Setor, universidades, iniciativa privada e governos. O compartilhamento de boas práticas e o desenvolvimento de metodologias internacionais reforçam o papel do Brasil como exportador de tecnologias sociais.

A integração entre Inteligência Artificial e empreendedorismo por necessidade não representa apenas uma inovação pedagógica, mas uma nova fronteira da inclusão social. A tecnologia, quando moldada por epistemologias periféricas e guiada por valores comunitários, pode ser mais uma ferramenta de emancipação e cidadania. O desafio central está em humanizar a IA - fazê-la aprender com as favelas, com os territórios, com o saber empírico. O futuro da capacitação empreendedora depende de um pacto entre técnica e cultura, algoritmo e afeto, eficiência e solidariedade.

A metodologia *By Necessity* demonstra que não há contradição entre o digital e o popular, entre inovação e periferia; há, sim, um campo fértil onde o conhecimento comunitário se torna o motor da transformação tecnológica e social. Como sintetiza Lima (2017), “o empreendedorismo é o nome moderno do instinto de sobrevivência”. Com a IA a serviço desse instinto, o potencial de geração de riqueza - simbólica, econômica e coletiva - torna-se ilimitado.

## 7. Referências bibliográficas

- Bardin, L. (2011). *Análise de conteúdo*. Lisboa: Edições 70.
- Bourdieu, P. (1979). *La distinction: Critique sociale du jugement*. Paris: Éditions de Minuit.
- Castells, M. (1999). *A sociedade em rede*. São Paulo: Paz e Terra.
- Engeström, Y. (2001). Expansive learning at work: Toward an activity theoretical reconceptualization. *Journal of Education and Work*, 14(1), 133–156. <https://doi.org/10.1080/13639080020028747>
- Freire, P. (1987). *Pedagogia do oprimido*. Rio de Janeiro: Paz e Terra.
- Freire, P. (1996). *Pedagogia da autonomia: Saberes necessários à prática educativa*. São Paulo: Paz e Terra.
- Glaser, B., & Strauss, A. (1967). *The discovery of grounded theory: Strategies for qualitative research*. Chicago: Aldine.
- Lévy, P. (1998). *A inteligência coletiva: Por uma antropologia do ciberespaço*. São Paulo: Loyola.
- Lima, V. M. (2016). *A riqueza das favelas*. Porto Alegre: Editora Besouro.
- Lima, V. M. (2017). *By Necessity: Plano de negócios por necessidade*. Porto Alegre: Editora Besouro.
- Lima, V. M. (2018). *Canvas das favelas*. Porto Alegre: Editora Besouro.
- Luckin, R., Holmes, W., Griffiths, M., & Forcier, L. B. (2016). *Intelligence unleashed: An argument for AI in education*. Pearson Education.
- Mignolo, W. (2017). *Coloniality: The darker side of modernity*. Durham, NC: Duke University Press.
- Nussbaum, M. (2011). *Creating capabilities: The human development approach*. Cambridge, MA: Harvard University Press.
- Organisation for Economic Co-operation and Development. (2019). *Artificial intelligence in society*. OECD Publishing. <https://doi.org/10.1787/eedfee77-en>
- Santos, B. de S. (2018). *O fim do império cognitivo: A afirmação das epistemologias do Sul*. Belo Horizonte: Autêntica.

- Schwab, K. (2016). *The fourth industrial revolution*. Geneva: World Economic Forum.
- SEBRAE. (2023). *Relatório anual de empreendedorismo no Brasil*. Brasília: SEBRAE.
- Selwyn, N. (2019). *Should robots replace teachers? AI and the future of education*. Cambridge: Polity Press.
- Sen, A. (2000). *Desenvolvimento como liberdade*. São Paulo: Companhia das Letras.
- UNESCO. (2022). *Ethics of artificial intelligence: Global framework for education*. Paris: UNESCO.
- Wenger, E. (1998). *Communities of practice: Learning, meaning, and identity*. Cambridge: Cambridge University Press.
- Williamson, B. (2021). *Algorithmic governance and education*. London: Routledge.
- Yin, R. K. (2015). *Case study research: Design and methods* (5th ed.). Thousand Oaks, CA: SAGE.



# Alfabetización digital algorítmica y búsqueda de empleo: una revisión de la literatura

## Algorithmic digital literacy and job search: a review of the literature

**Stefano De Marco**

*Universidad de Salamanca (s.demarco@usal.es)*

**Guillaume Dumont**

*EM Lyon Business School (Francia)*

**Ellen Helsper**

*London School of Economics (Reino Unido)*

**Daniel Calderón-Gómez**

*Universidad Complutense de Madrid*

### PALABRAS CLAVE:

Alfabetización  
Algorítmica  
Alfabetización Digital  
Alfabetización en IA  
Búsqueda de Empleo  
Desigualdad Digital

### RESUMEN:

En este artículo se presenta una revisión de la bibliografía sobre el concepto de alfabetización digital algorítmica y su aplicación al ámbito de la búsqueda de empleo. Para ello, se plantea una discusión del concepto de alfabetización algorítmica y su vinculación con la desigualdad digital, la alfabetización digital y los estudios sobre competencias digitales. A continuación, se presenta una revisión de los estudios recientes sobre gestión y manejo de algoritmos para la búsqueda de empleo, un campo de investigación emergente en los últimos años pero que aún requiere de mayor desarrollo empírico.

### KEY WORDS:

Algorithmic Literacy  
Digital Literacy  
AI Literacy  
Job Search, Digital  
Inequality

### ABSTRACT:

This article presents a review of the literature on the concept of algorithmic digital literacy and its application to the field of job seeking. To this end, it discusses the concept of algorithmic literacy and its connection with digital inequality, digital literacy, and studies on digital competencies. Therefore, we present a revision of recent research about the management and use of algorithms in job searching, an emerging field of study in recent years that still requires further empirical development.

## 1. INTRODUCCIÓN

En la actualidad, asistimos a una tendencia hacia la digitalización de la vida social, de forma que la mayoría de los servicios ofertados por empresas, administraciones e instituciones están progresivamente trasladándose al ámbito digital, al mismo tiempo que se restringen y limitan las opciones offline de prestación de servicios. Asimismo, esta transformación digital se fundamenta en la lógica y las restricciones de las plataformas online, cuyo funcionamiento se rige en el uso de algoritmos (Plantin et al., 2018). Un algoritmo se define como un conjunto de procedimientos codificados que, a partir de cálculos específicos, consiguen transformar grandes cantidades de *inputs* en resultados o *outputs* deseados (Zarouali, Helberger y de Vreese, 2021). El diseño de estos algoritmos y su aplicación al ámbito de las plataformas, herramientas y ecosistemas sociotécnicos resulta en la creación de las “plataformas algorítmicas [...] que transmiten información de acuerdo con las lógicas sociales, culturales, políticas y comerciales dominantes” (Cotter y Reisdorf, 2020, p. 746), reproduciendo las dinámicas y procesos de estratificación social que encontramos en el mundo offline.

**CÓMO CITAR:** De Marco, S., Dumont, G., Helsper, E., & Calderón-Gómez, D. (2025). Alfabetización digital algorítmica y búsqueda de empleo: Una revisión de la literatura. *RETIS. Revista de Tecnología para la Inclusión Social*, 2(1), 15-22.  
DOI: [10.70664/retis.v2i1.002](https://doi.org/10.70664/retis.v2i1.002)

\* C-e: [s.demarco@usal.es](mailto:s.demarco@usal.es) / [g dumont@em-lyon.com](mailto:g dumont@em-lyon.com) / [e.j.helsper@lse.ac.uk](mailto:e.j.helsper@lse.ac.uk) / [danielcalderon@ucm.es](mailto:danielcalderon@ucm.es)

Este es el caso, por ejemplo, de los motores de búsqueda y de las redes sociales, pues el funcionamiento de sus *feeds* (o contenido sindicado que aparece a los usuarios) depende de algoritmos de los cuales, a menudo, los usuarios no son conscientes o, de serlo, desconocen su funcionamiento. Al mismo tiempo, por razones comerciales, las compañías tienden a no difundir los códigos que rigen el funcionamiento de los algoritmos. Así, se ha generado un contexto opaco en el cual la voluntad de los usuarios no es suficiente, muchas veces, para condicionar los resultados obtenidos a partir del uso de determinadas plataformas, derivando en lo que Kordzadeh y Ghasemaghei (2022) definen como sesgo algorítmico (*algorithmic bias*), que ocurre cuando “los resultados de un algoritmo benefician o perjudican a determinados individuos o grupos más que a otros sin que exista una razón justificada para tal impacto desigual” (p. 388).

Así pues, se está observando como, por una parte, la aplicación de la Inteligencia Artificial (IA) a servicios públicos puede generar desigualdades y cometer injusticias a la hora de repartir recursos, por ejemplo, entre usuarios de servicios públicos (Eubanks, 2018). Por otra parte, hay estudios que evidencian los problemas de privacidad derivadas del uso de la información personal de los usuarios de las plataformas (Micheli, Lutz y Büchi, 2018) o por la generación de burbujas informativas que sesgan la percepción de los acontecimientos económicos y políticos por parte de los usuarios (Colleoni, Rozza y Arvidsson, 2014; Yarchi, Baden y Kligler-Vilenchik, 2021). Particularmente, en este artículo presentamos una revisión teórica sobre la importancia que tiene la alfabetización algorítmica de los usuarios en el uso que realizan de las plataformas digitales de búsqueda de empleo, pues se trata de uno de los sectores en el que las competencias y capacidad de agencia de los usuarios a la hora de manejar e influir en los algoritmos resulta crucial para mejorar sus oportunidades laborales y profesionales.

## 2. LA ALFABETIZACIÓN DIGITAL ALGORÍTMICA

En un contexto social y tecnológico como el actual, de omnipresencia de las tecnologías digitales en nuestra vida cotidiana, la capacidad para manejar y sacar provecho de las mismas constituye uno de los aspectos nodales de la desigualdad digital. Por lo tanto, más allá del acceso a las tecnologías, en la actualidad el principal foco de desigualdad sociodigital tiene que ver con las competencias digitales –segunda brecha digital (Castaño, 2008; Ruii y Ragnedda, 2020)– y con la capacidad de aprovechar las tecnologías para mejorar las propias oportunidades de vida –tercera brecha digital (Ragnedda, 2016; Scheerder, Van Deursen y Van Dijk, 2017)–.

En la Unión Europea, el enfoque más consensuado para dimensionar las competencias digitales es el *Marco europeo de competencias digitales DigComp 2.2* (Vuorikari, Kluzer, y Punie, 2022), que incluye una serie de tareas relacionadas con el ámbito de las tecnologías digitales que las personas saben realizar dentro de cinco áreas: información, comunicación, creación de contenido, seguridad y resolución de problemas. Sin embargo, la alfabetización digital (Erstad, 2011; Sefton-Green, Nixon y Erstad, 2009) no puede materializarse en un conjunto estable e inmutable de destrezas, sino que tiene que ver con procesos informales de aprendizaje a partir del uso de los dispositivos, con la adquisición de familiaridad y confianza en las tecnologías que forman parte de la experiencia cotidiana de los sujetos y, en definitiva, con la propia autopercepción sobre la capacidad para desenvolverse satisfactoriamente como usuario en el ecosistema digital (Cotter y Reisdorf, 2020; Jordan & Helsper, 2023; Ragnedda, Ruii y Calderón-Gómez, 2024).

Una de las dimensiones de la alfabetización digital que ha ganado protagonismo en los últimos años tiene que ver con las competencias necesarias para manejarse con los algoritmos que rigen el funcionamiento de las plataformas digitales, emergiendo el concepto de *alfabetización digital algorítmica* (Gruber y Hargittai, 2023; Oeldorf-Hirsch y Neubaum, 2025). Así, numerosos autores han planteado la existencia de diferencias entre internautas respecto al conocimiento y manejo de los algoritmos, generando así patrones diferentes de interacción y, por ende, de obtención de beneficios tangibles a partir de su uso (Cotter and Reisdorf, 2020; Lutz, 2019; Zarouali, Helberger y de Vreese, 2021). Es este un ámbito de investigación incipiente, caracterizado por la peculiaridad de que gran parte de los investigadores sociales desconocen el contenido y el funcionamiento de los algoritmos de las diferentes plataformas digitales (Hargittai et al., 2020). A pesar de esta limitación, se han realizado diferentes investigaciones, sobre todo en relación con plataformas de búsqueda y con las redes sociales online, que han permitido elaborar numerosas definiciones, tanto teóricas como empíricas, de las habilidades en el manejo de los algoritmos (Cotter y Reisdorf, 2020; DeVito, 2021; Dogruel, Masur y Joeckel, 2022; Jarrahi et al., 2021; Savolainen y Ruckenstein, 2022).

Así, la alfabetización algorítmica se plantea como un concepto paraguas que engloba una amplia diversidad de términos (Oeldorf-Hirsch y Neubaum, 2025), entre los que se incluyen alusiones al conocimiento –*algorithmic awareness, knowledge, understanding, imaginary*–, a la experiencia en el manejo de los algoritmos –*algorithmic experience*– o a las desigualdades vinculadas con su uso –*algorithmic divide*–. Un análisis atento de estas definiciones permite observar cómo se basan en dos constructos principales: la consciencia de la existencia de los algoritmos o *algorithmic awareness* (por ejemplo: saber que el orden de los resultados en un motor de búsqueda puede ser influenciado por factores de diferente índole, incluidos los de tipo comercial) y la capacidad de “manipular” el algoritmo para obtener resultados deseados por el usuario (por ejemplo: dar ‘me gusta’ a determinados contenidos de manera consciente y así influenciar el *feed*). Consecuentemente, cuando hablamos de alfabetización algorítmica (*algorithmic literacy*) nos referimos a un conjunto de disposiciones, conocimientos y habilidades prácticas que pueden ayudar a los usuarios a conocer los sesgos algorítmicos de las plataformas, interactuando con ellas de manera consciente y siendo capaces de manipular el tipo de resultados obtenidos a partir de dicha interacción. Esta definición abarca dimensiones cognitivas, afectivas y conductuales que afectan al desempeño de los usuarios con las plataformas (Oeldorf-Hirsch y Neubaum, 2025):

- *Dimensión cognitiva*: tiene que ver con el conocimiento de la existencia de ciertos algoritmos que rigen las plataformas (*awareness, knowledge*) y la comprensión de la capacidad para utilizarlos adecuadamente (*understanding*).
- *Dimensión afectiva*: tiene que ver con las actitudes, intereses y motivaciones de los usuarios hacia la manipulación activa de los algoritmos, que va del rechazo a la utilización deliberada y consciente (ver Logg, Minson y Moore, 2019).

- **Dimensión conductual:** tiene que ver con la capacidad de manipulación, gestión y utilización real de los algoritmos en beneficio propio.

Asimismo, el conocimiento de los usuarios sobre los algoritmos es de tipo “interactivo”, se adquiere con el uso y con la interacción cotidiana (Cotter y Reisdorf, 2020), pues los sesgos algorítmicos, en la mayor parte de plataformas, están deliberadamente ocultos por motivaciones económicas y comerciales, por lo que deben ser intuidos por parte del usuario mediante procesos de carácter informal y colectivo, basados más en el “aprendizaje durante el uso” y en la observación de la conducta de otros internautas que en cursos o procesos de capacitación formal (Green et al., 2024). A medida que los usuarios interactúan con las plataformas que se basan en algoritmos, también reflexionan sobre lo que observan, generando así creencias intuitivas sobre su funcionamiento, una imaginaria algorítmica (*algorithmic imaginary*) referida a “la manera en que las personas imaginan, perciben y experimentan los algoritmos y lo que esos imaginarios hacen posible” (Butcher, 2017, p. 31). Esta interacción con la plataforma (endógena) y con otros usuarios (exógena) lleva al desarrollo de teorías intuitivas sobre el funcionamiento de los algoritmos (*algorithmic folk theories*), las cuales son siempre imperfectas y parciales, pero ayudan a orientar las prácticas de los usuarios a través de dinámicas de ensayo y error, de “hacking” de las plataformas y herramientas digitales (DeVito, 2021; Siles et al, 2020). La inconsistencia de estas teorías intuitivas se pone de manifiesto cuando se observa una violación de las expectativas en cuanto a los resultados que deberían obtenerse a partir del uso de las plataformas (Swart, 2021).

Desde el punto de vista de la investigación, es importante señalar que la operacionalización y medición de la alfabetización algorítmica es complicada, pues las propuestas que se han desarrollado para su conceptualización todavía son muy recientes. Por una parte, existen propuestas cualitativas que se basan, principalmente, en el uso de entrevistas y grupos de discusión, que tienen como objetivo comprender el funcionamiento de la experiencia algorítmica entre los usuarios, aplicándose a contextos especialmente relevantes, principalmente al ámbito de las redes sociales y plataformas digitales (Alvarado y Waern, 2020; Butcher, 2017; Klug et al., 2021; Powers, 2017). Por otro lado, desde la perspectiva cuantitativa, se ha planteado la elaboración de índices o indicadores que midan las diferencias competencias algorítmicas de los usuarios y que puedan relacionarse con otras competencias digitales o beneficios tangibles del uso de las TIC. A modo de ejemplo, podemos citar la *Algorithmic Literacy Scale* (Dogruel, Masur y Joeckel, 2022) o la *Algorithmic Media Content Awareness (AMCA) scale* (Zarouali et al., 2021), si bien se trata de propuestas emergentes que deben ser validadas a través de mayor investigación aplicada.

Más recientemente, un área de investigación que ha comenzado a desarrollarse en los últimos años tiene que ver con la alfabetización digital vinculada con la Inteligencia Artificial (*AI Literacy*), que intenta identificar las desigualdades, riesgos y vinculadas con la enorme generalización de uso de este tipo de herramientas (Chiur, 2025; Lintner, 2024; Stoilova y Livingstone, 2025). Se trata de un campo de investigación emergente e incipiente que, sin embargo, ha crecido mucho en poco tiempo, especialmente en el sector de la educación (Tadimalla y Maher, 2025; Su et al., 2023), pero que todavía está pendiente de consolidarse e integrarse con otras perspectivas teóricas relativas a la alfabetización digital y la socialización tecnológica.

### 3. LA GESTIÓN DE LOS ALGORITMOS EN LA BÚSQUEDA DE EMPLEO

Durante las últimas tres décadas, el desarrollo de las TIC ha transformado drásticamente la manera de buscar trabajo (Backman y Hedenus, 2019; Bonet et al., 2013; Coverdill y Finlay, 2017; Garg y Telang, 2018; Janta y Ladkin, 2013). Mas concretamente, la digitalización de los canales de búsqueda de empleo ha otorgado a las plataformas digitales un rol central en el acceso al mercado laboral, reconfigurando la forma en que las personas gestionan y experimentan este proceso (De Marco, Dumont y Helsper, 2025; Gandini, 2016; Petriglieri et al., 2019; Stark & Pais, 2020; Vallas & Schor, 2020).

En este contexto, los intermediarios clásicos de acceso al trabajo han empezado a usar masivamente las plataformas digitales, facilitando así las transacciones entre las personas en búsqueda de empleo y los empleadores (Bonet et al., 2013; Coverdill & Finlay, 2017; Sharone, 2013). Plataformas como Infojobs, Jobtoday, LinkedIn, Indeed, Monster y Glassdoor, entre otras, facilitan la inclusión en el mercado laboral de un gran número de candidatas (Coverdill & Finlay, 2017; Gershon, 2017). A su vez, las empresas han abierto canales directos para la búsqueda y solicitud de empleo a través de sus páginas web, donde los candidatos pueden, después de haberse registrado, ver y solicitar vacantes. Grandes intermediarios, como Randstad y Adecco, han replicado este proceso, multiplicando así el número de plataformas disponibles.

Todo ello ha modificado radicalmente la experiencia de la búsqueda de trabajo. Los candidatos disponen de nuevas herramientas y oportunidades para encontrar vacantes y, eventualmente, acceder a empleos mejor remunerados (Kroft y Pope, 2014), especialmente en el caso de las personas jóvenes y aquellas con mayor confianza en el uso de las TIC (Piercy y Kyong Lee, 2019). A su vez, los reclutadores disponen de grandes bases de datos de candidatos y se están planteando el uso de Inteligencia Artificial para implementar procesos de selección automatizados, emergiendo nuevos dilemas éticos sobre la igualdad de oportunidades y los sesgos de estos procedimientos (Dubovitskaya, Mazzola y Denzler, 2020). Además, hay un desequilibrio entre la oferta y la demanda fomentando la competencia entre los candidatos con perfiles medios o bajos (McDonald et al., 2019), pues existe un nicho de trabajadores cuyos perfiles son muy demandados por las empresas y que no necesitan competir para asegurarse un empleo (Jarrahi y Sutherland, 2019).

En este contexto, las habilidades digitales son fundamentales para buscar empleo en las plataformas, ya que proporcionan mayores oportunidades de obtenerlo (Laukkarinen, 2023; Dumont et al., 2023; Van Dijk et al, 2017). Estas habilidades se concentran principalmente entre candidatos con mayor nivel de recursos económicos y educativos (De Marco et. al, 2023; Karaoglu, Hargittai, and Nguyen, 2021). Consecuentemente, asistimos a un fenómeno de brecha digital de tercer nivel: los beneficios tangibles de la búsqueda online de empleo se distribuirán entre las personas con mayor nivel de recursos materiales. Así, las posibilidades que ofrecen las tecnologías digitales para la búsqueda de empleo son mayores para personas en mejor situación socioeconómica.

Este nuevo fenómeno de exclusión digital tiene consecuencias considerables. Entre ellas, el hecho que personas con menos habilidades abandonan la búsqueda (Dumont et al., 2023) y que personas con menor nivel de destreza sufren mayor estrés psi-

cológico (De Marco et al., 2023). Estos resultados están en línea con los estudios realizados en el ámbito de la brecha de tercer nivel, en base a los cuales se considera que las habilidades digitales también son una forma para modular los efectos indeseados que desprenden del uso de internet (Helsper, 2021). En efecto, esta forma de buscar trabajo va asociada a un nuevo conjunto de presiones con las cuales los candidatos tienen que lidiar, tal y como se ha explicitado anteriormente, y puede comportar niveles elevados de desgaste psicológico (Pultz y Sharone, 2020; Vallas y Cummins, 2015; Van Oort, 2013).

Estas consideraciones, sin embargo, no han sido acompañadas por un estudio sistemático de cómo los candidatos interactúan con los algoritmos que rigen las plataformas para buscar empleo. Las únicas investigaciones en este ámbito se centran en la *gig economy* (o economía de plataformas) e investigan cómo los algoritmos ejercen como agentes de recursos humanos en encubierta, monitoreando el trabajo realizado por los usuarios (Duggan et al., 2023; Tassinari et al., 2020; Veen et al., 2019). De este modo, generan normas de conducta a las que los trabajadores deben atenerse para poder conseguir empleos, obligándoles a trabajar siguiendo pautas determinadas establecidas por el algoritmo y empujándoles a mejorar su reputación de “buenos profesionales” (Duffy, 2020; Dumont, 2017).

Si las habilidades digitales facilitan el buen desenlace de la búsqueda de empleo online, todavía desconocemos el rol de la *Algorithmic Literacy* en este proceso. Sin embargo, es fundamental conocer estrategias para manejar al algoritmo y hacer más visible una candidatura, por ejemplo, escogiendo bien las palabras claves a insertar en un CV (Dumont et al. 2023). Es por ello que, por ejemplo, los profesionistas que se anuncian en plataformas para el empleo intentan desarrollar un conocimiento intuitivo del algoritmo, para así dar visibilidad a su perfil, obtener más clientes y eludir el control del algoritmo sobre su actividad profesional (Jarrahi y Sutherland, 2019). De la misma forma, los trabajadores de las industrias creativas consideran las habilidades para lidiar con los algoritmos como parte de sus competencias laborales, así que invierten recursos económicos en intentar dar sentido a los sistemas algorítmicos de las plataformas en las que se implican (Duffy, 2020).

En este sentido, la necesidad de generar una buena “reputación digital”, mediante a un comportamiento activo en las redes sociales y la continua creación de contenidos digitales, choca con el nivel bajo de resultados tangibles que estos profesionales obtienen en el ámbito laboral (Duffy y Meisner, 2022). La presión por ganar en visibilidad publicando contenido, entrelazando información personal y profesional, genera dificultades y malestar entre los usuarios (Sharone, 2017). Todo ello, puede dar lugar a la autoculpabilización, la pérdida de confianza en uno mismo y la ansiedad (Pultz y Sharone, 2020; Vallas y Cummins, 2015; Van Oort, 2013). Además, la falta de contacto con personas físicas en el proceso de selección también es fuente de frustración para los candidatos, ya que no tienen ningún tipo de retroalimentación acerca del proceso en el que participan (Rieucan, 2014).

En su conjunto, estas investigaciones apuntan a que el uso de algoritmos en plataformas de búsqueda de empleo podría estar agudizando la estratificación social de dos maneras. Por una parte, las personas con menor nivel de comprensión del funcionamiento de dichos algoritmos se verían perjudicados en el proceso de búsqueda. En segundo lugar, los mecanismos en los que se basa el funcionamiento de las plataformas de empleo podrían generar mayor malestar y agotamiento psicológico en aquellas personas con menos recursos digitales y materiales. Profundizar en estas cuestiones requiere desarrollar conocimiento empírico y teórico en tres ámbitos. Primero, se requiere describir los componentes de la *Algorithmic Literacy* y analizar en profundidad cómo las personas en búsqueda de empleo adquieren este tipo de destrezas. En segundo lugar, es necesario explicar si (y cómo) este tipo de destrezas influyen en los resultados del proceso de búsqueda y si ayudan a prevenir el desgaste psicológico a ello asociado. Tercero, es imprescindible entender si los recursos materiales de los internautas influyen en la adquisición de la *Algorithmic Literacy*, generando así mecanismos de desigualdad digital.

#### 4. CONCLUSIONES E IMPLICACIONES

Este trabajo se ha centrado en la *Algorithmic Literacy*, es decir, las habilidades de los internautas para interactuar con los algoritmos que rigen las plataformas digitales. Más en concreto, se ha querido profundizar, mediante análisis bibliográfico, en la relación que este tipo de destreza mantiene con la búsqueda de empleo a través de plataformas. Las evidencias empíricas apuntan a que la capacidad para manipular los algoritmos puede otorgar alguna ventaja competitiva a los candidatos a la hora de optar para un puesto de trabajo. Esto se traduce, por una parte, en una mayor visibilidad para sus candidaturas (De Marco, Dumont y Helsper, 2025; Jarrahi y Sutherland, 2019; Duffy, 2020). Además, podría prevenir el malestar psicológico que las personas en búsqueda de empleo pueden sufrir como consecuencia del propio proceso de búsqueda (Duffy y Meisner, 2022; Pultz y Sharone, 2020). Por ejemplo, por las largas esperas, la falta de información por parte de las empresas o la falta de transparencia acerca del proceso de selección etc. Lo cual aumenta, además, la probabilidad de abandono de la búsqueda, alargando así las temporadas de desempleo (Dumont, De Marco y Helsper, 2024).

A lo largo del texto, este tipo de evidencias se ha enmarcado en el seno de la tercera brecha digital (van Deursen et al., 2017), es decir, en cómo las personas con mayor nivel de recursos offline obtienen un mayor nivel de beneficios tangibles a partir del mismo uso de internet (van Deursen y Helsper, 2015). Además, estas mismas personas tendrían una mayor probabilidad de evitar los efectos indeseados del uso de internet (Helsper, 2021). En este caso concreto, las personas con un mayor nivel de recursos materiales y educativos tendrían un mayor nivel de destreza para interactuar con los algoritmos. Como consecuencia, serían capaces de dar mayor visibilidad a sus candidaturas y de evitar el malestar psicológico asociado al proceso de búsqueda. Vista de esta manera, pues, la tercera brecha digital es un potenciador de la estratificación social, ya que permite a las personas con mayores niveles de recursos offline aumentar dichos recursos gracias a la búsqueda online de empleo (Ragnedda, 2016).

Esta perspectiva es muy relevante porque plantea la posibilidad de que el acceso al mercado laboral esté sesgado por la desigualdad digital. Es decir, las personas que ya gozan de una posición privilegiada tendrían un acceso más fácil al mercado laboral gracias a las plataformas digitales. Ello llevaría a la necesidad de comprender qué tipo de competencias digitales necesitan los ciudadanos para poder enfrentar los retos que los cambios sociales y tecnológicos están planteando en cuanto al acceso al mercado laboral. No se trata solo de darles conocimientos para el correcto desempeño de las tareas laborales, sino darles herramientas

para competir en equidad con otros candidatos que optan a los mismos puestos de trabajo. Permitiéndoles, además, sobrellevar el proceso evitando la componente de desgaste psicológico que lleva consigo la opacidad

Sin embargo, esta reseña bibliográfica ha puesto de manifiesto la falta de literatura especializada sobre el tema. Pese a la gran relevancia de este objeto de investigación, faltan todavía trabajos que, con un mayor nivel de profundidad, permitan comprender cómo la *Algorithmic Literacy* pueda o no generar una posición de ventaja a la hora de buscar empleo online. Es este un reto de gran relevancia que, sin embargo, se torna más difícil por las propias características del objeto de investigación. Es decir, los algoritmos son algo de por sí opaco, que las empresas, plataformas o administraciones tienden a no compartir con el público. No es posible, por tanto, conocer a priori y de manera "objetiva" cómo funcionan. Así pues, su estudio debe servirse de las percepciones de los usuarios, analizando también las formas más o menos exitosas de cómo éstos interactúan con los algoritmos. A partir de ahí se podrá tener alguna pista de cuáles son las maneras más beneficiosas de relacionarse con las plataformas para el empleo, pudiendo así generar conocimiento para la formación e inserción laboral de las personas.

## 5. REFERENCIAS BIBLIOGRÁFICAS

- Alvarado, O. y Waern, A. (2018) Towards algorithmic experience. Proceedings of the 2018 CHI conference on human factors in computing systems. Association for Computing Machinery, New York, paper 286, 1-12. <https://doi.org/10.1145/3173574.3173860>
- Backman, C. y Hedenus, A. (2019), Online privacy in job recruitment processes? Boundary work among cybervetting recruiters. *New Technology, Work and Employment*, 34, 157-173. <https://doi.org/10.1111/ntwe.12140>
- Bonet, R.; Cappelli, P. y Hamori, M. (2013). Labor Market Intermediaries and the New Paradigm for Human Resources, *ANNALS*, 7, 341-392, <https://doi.org/10.5465/19416520.2013.774213>
- Bucher, T. (2016). The algorithmic imaginary: exploring the ordinary affects of Facebook algorithms. *Information, Communication & Society*, 20(1), 30-44. <https://doi.org/10.1080/1369118X.2016.1154086>
- Calderón Gómez, Daniel. 2020. "The Third Digital Divide and Bourdieu: Bidirectional Conversion of Economic, Cultural, and Social Capital to (and from) Digital Capital among Young People in Madrid." *New Media & Society*, 23(9), 2534-2553. <https://doi.org/10.1177/1461444820933252>
- Castaño, C. (2008). *La segunda brecha digital*. Cátedra.
- Chiu, T. K. F. (2025). AI literacy and competency: definitions, frameworks, development and future research directions. *Interactive Learning Environments*, 33(5), 3225-3229. <https://doi.org/10.1080/10494820.2025.2514372>
- Colleoni, E., Rozza, A. y Arvidsson, A. (2014). Echo chamber or public sphere? Predicting political orientation and measuring political homophily in Twitter using big data. *Journal of Communication*, 64(2), 317-332. <https://doi.org/10.1111/jcom.12084>
- Cotter K. and Reisdorf B. C. (2020) Algorithmic knowledge gaps: a new dimension of (digital) inequality. *International Journal of Communication*, 14, 745-765.
- Coverdill, J. E. y Finlay, W. (2017). *High Tech and High Touch: Headhunting, Technology, and Economic Transformation*. Cornell University Press.
- De Marco, S., Dumont, G., & Helsper, E. J. (2025). The reproduction of structural inequalities in online job search strategies and outcomes. *New Media & Society*, 0(0). <https://doi.org/10.1177/14614448241310497>
- DeVito, M. A. (2021) Adaptive folk theorization as a path to algorithmic literacy on changing platforms. Proceedings of the ACM on Human-Computer Interaction 5(CSCW2), 339. <https://doi.org/10.1145/3476080>
- Dogruel, L., Masur, P. y Joeckel, S. (2021). Development and Validation of an Algorithm Literacy Scale for Internet Users. *Communication Methods and Measures*, 16(2), 115-133. <https://doi.org/10.1080/19312458.2021.1968361>
- Dubovitskaya A, Mazzola L, Denzler A (2020) Towards a trusted support platform for the job placement task. In Schwarzmann U, Boehme C, Heras DB, et al. (Eds.). *Euro-Par 2019: Parallel Processing Workshops*. Springer International Publishing (pp. 203-215).
- Duffy, B. E. (2020). Algorithmic precarity in cultural work. *Communication and the Public*, 5(3-4), 103-107. <https://doi.org/10.1177/2057047320959855>
- Duffy, B. E. y Meisner, C. (2022). Platform governance at the margins: Social media creators' experiences with algorithmic (in) visibility. *Media, Culture & Society*, 45(2), 285-304. <https://doi.org/10.1177/01634437221111923>
- Duggan, J., Sherman, U., Carbery, R. y McDonnell, A. (2021). Boundaryless careers and algorithmic constraints in the gig economy. *The International Journal of Human Resource Management*, 33(22), 4468-4498. <https://doi.org/10.1080/09585192.2021.1953565>
- Dumont, G. (2017). The labor of reputation building: Creating, developing and managing individual reputation. *Consumption Markets & Culture*, 21(6), 515-531. <https://doi.org/10.1080/10253866.2017.1373253>
- Dumont, G., De Marco, S. y Helsper, E.J. (2024) Online job search discouragement: how employment platforms and digital exclusion shape the experience of low-qualified job seekers? *New Technology, Work and Employment*, 39, 89-108. <https://doi.org/10.1111/ntwe.12279>
- Erstad, O. (2011). Citizens navigating in literate worlds: The case of digital literacy. En M. Thomas (Ed.). *Deconstructing digital natives: Young people, technology, and the new literacies*. Routledge (pp. 99-118) <https://doi.org/10.4324/9780203818848>
- Eubanks, B. (2018). *Artificial Intelligence for HR: Use AI to Support and Develop a Successful Workforce*. Kogan Page
- Gandini, A. (2015). Digital work: Self-branding and social capital in the freelance knowledge economy. *Marketing Theory*, 16(1), 123-141. <https://doi.org/10.1177/1470593115607942>
- Garg, R. y Telang, R. (2017) To Be or Not to Be Linked: Online Social Networks and Job Search by Unemployed Workforce. *Management Science*, 64(8), 3926-3941. <https://doi.org/10.1287/mnsc.2017.2784>

- Gershon, I. (2017). *Down and Out in the New Economy How People Find (or Don't Find) Work Today*. The University of Chicago Press.
- Green, L., Haddon, L., Livingstone, S., O'Neill, B., Stevenson, K.J. y Holloway, D. (2024). Parenting and Digital Media. En *Digital Media Use in Early Childhood: Birth to Six* Bloomsbury Academic (pp. 51–68). <http://doi.org/10.5040/9781350120303.ch-4>
- Gruber, J., y Hargittai, E. (2023). The importance of algorithm skills for informed Internet use. *Big Data & Society*, 10(1). <https://doi.org/10.1177/20539517231168100>
- Hargittai, E.; Gruber, J.; Djukaric, T.; Fuchs, J. y Brombach, L. (2020). Black box measures? How to study people's algorithm skills. *Information, Communication & Society*, 23(5), 764–775. <https://doi.org/10.1080/1369118X.2020.1713846>
- Helsper, E. J. (2021) *The Digital Disconnect: The Social Causes and Consequences of Digital Inequalities*. SAGE.
- Janta, H. y Ladkin, A. (2013), Online technologies and Polish migrants. *New Technology, Work and Employment*, 28: 241-253. <https://doi.org/10.1111/ntwe.12018>
- Jarrahi, M. H., Newlands, G., Lee, M. K., Wolf, C. T., Kinder, E. y Sutherland, W. (2021). Algorithmic management in a work context. *Big Data & Society*, 8(2). <https://doi.org/10.1177/20539517211020332>
- Jarrahi, M.H. y Sutherland, W. (2019). Algorithmic Management and Algorithmic Competencies: Understanding and Appropriating Algorithms in Gig Work. In Taylor, N., Christian-Lamb, C., Martin, M., Nardi, B. (Eds.). *Information in Contemporary Society*. iConference 2019. Lecture Notes in Computer Science, vol 11420. Springer. [https://doi.org/10.1007/978-3-030-15742-5\\_55](https://doi.org/10.1007/978-3-030-15742-5_55)
- Karaoglu, G., Hargittai, E. y Nguyen, M. H. (2021). Inequality in online job searching in the age of social media. *Information, Communication & Society*, 25(12), 1826–1844. <https://doi.org/10.1080/1369118X.2021.1897150>
- Klug, D., Qin, Y., Evans, M. y Kaufman, G. (2021). Trick and Please. A Mixed-Method Study On User Assumptions About the TikTok Algorithm. *Proceedings of the 13th ACM Web Science Conference 2021 (WebSci '21)*. Association for Computing Machinery, New York, 84–92. <https://doi.org/10.1145/3447535.3462512>
- Kordzadeh, N., y Ghasemaghahi, M. (2021). Algorithmic bias: review, synthesis, and future research directions. *European Journal of Information Systems*, 31(3), 388–409. <https://doi.org/10.1080/0960085X.2021.1927212>
- Kroft, K. y Pope, D. G. (2014) Does Online Search Crowd Out Traditional Search and Improve Matching Efficiency? Evidence from Craigslist. *Journal of Labor Economics*, 32(2), 259-303. <https://doi.org/10.1086/673374>
- Laukkarinen, M. (2023). Social media as a place to see and be seen: Exploring factors affecting job attainment via social media. *The Information Society*, 39(4), 199–212. <https://doi.org/10.1080/01972243.2023.2199418>
- Lintner, T. (2024). A systematic review of AI literacy scales. *NPJ science of learning*, 9, 50. <https://doi.org/10.1038/s41539-024-00264-4>
- Logg, J. M., Minson, J. A. y Moore, D. A. (2019) Algorithm appreciation: people prefer algorithmic to human judgment. *Organizational Behavior and Human Decision Processes*, 151, 90–103. <https://doi.org/10.1016/j.obhdp.2018.12.005>
- Lutz, C. (2019). Digital Inequalities in the Age of Artificial Intelligence and Big Data. *Human Behavior and Emerging Technologies*, 1, 141-148. <https://doi.org/10.1002/hbe2.140>
- McDonald S, Damarin AK, Lawhorne J, et al. (2019) Black holes and purple squirrels: a tale of two online labor markets. In Vallas SP, Kovalainen A (Eds.). *Work and Labor in the Digital Age*. Emerald Publishing Limited (pp. 93–120).
- Micheli, M.; Lutz, C. y Büchi, M (2018). Digital footprints: an emerging dimension of digital inequality. *Journal of Information, Communication and Ethics in Society*, 16(3), 242–251. <https://doi.org/10.1108/JICES-02-2018-0014>
- Oeldorf-Hirsch, A. y Neubaum, G. (2025). What do we know about algorithmic literacy? The status quo and a research agenda for a growing field. *New Media and Society*, 27(2), 681–701. <https://doi.org/10.1177/14614448231182662>
- Petriglieri, G., Ashford, S. J. y Wrzesniewski, A. (2018). Agony and Ecstasy in the Gig Economy: Cultivating Holding Environments for Precarious and Personalized Work Identities. *Administrative Science Quarterly*, 64(1), 124-170. <https://doi.org/10.1177/0001839218759646>
- Piercy, C. W. y Lee, S. K. (2018). A typology of job search sources: Exploring the changing nature of job search networks. *New Media & Society*, 21(6), 1173-1191. <https://doi.org/10.1177/1461444818808071>
- Plantin, J.C., Lagoze, C., Edwards, P. N. y Sandvig, C. (2018). Infrastructure studies meet platform studies in the age of Google and Facebook. *New Media & Society*, 20(1), 293-310. <https://doi.org/10.1177/1461444816661553>
- Powers, E. (2017). My news feed is filtered? Awareness of news personalization among college students. *Digital Journalism*, 5(10), 1315–1335. <https://doi.org/10.1080/21670811.2017.1286943>
- Pultz, S. y Sharone, O. (2020). The intimate dance of networking: A Comparative Study of the Emotional Labor of Young American and Danish Jobseekers. In E. H. Gorman, & S. P. Valas (Eds.), *Professional Work: Knowledge, Power and Social Inequalities* (Vol. 34). Emerald Group Publishing (pp. 33-58). <https://doi.org/10.1108/S0277-28332020000034006>
- Ragnedda, M. (2016). *The third digital divide: A Weberian approach to digital inequalities*. Routledge. <https://doi.org/10.4324/9781315606002>
- Ragnedda, M., Ruiiu, M. L. y Calderón-Gómez, D. (2024). Examining the Interplay of Sociodemographic and Sociotechnical Factors on Users' Perceived Digital Skills. *Media and Communication*, 12, article 8167. <https://doi.org/10.17645/mac.8167>
- Rieucan, G. (2015). Getting a low-paid job in French and UK supermarkets: from walk-in to online application?. *Employee Relations: The International Journal*, 37(1), 141–156. <https://doi.org/10.1108/ER-03-2014-0022>
- Ruiiu, M. L. y Ragnedda, M. (2020). Digital capital and online activities: An empirical analysis of the second level of digital divide. *First Monday*, 25(7). <https://doi.org/10.5210/fm.v25i7.10855>
- Savolainen, L. y Ruckenstein, M. (2022). Dimensions of autonomy in human–algorithm relations. *New Media & Society*, 26(6), 3472-3490. <https://doi.org/10.1177/14614448221100802>

- Scheerder, A., Van Deursen, A. y Van Dijk, J. (2017). Determinants of Internet skills, uses and outcomes: A systematic review of the second- and third-level digital divide. *Telematics and Informatics*, 34(8), 607-624. <https://doi.org/10.1016/j.tele.2017.07.007>
- Sefton-Green, J., Nixon H. y Erstad, O. (2009). Reviewing approaches and perspectives on 'digital literacy'. *Pedagogies*, 4(2), 107-125. <https://doi.org/10.1080/15544800902741556>
- Sharone, O. (2017). LinkedIn or LinkedOut? How Social Networking Sites are Reshaping the Labor Market. In S. Vallas (Ed.). *Emerging conceptions of work, management and the labor market* (Vol. 30). Emerald Group Publishing. <https://doi.org/10.1108/S0277-2833201730>
- Siles, I., Segura-Castillo, A., Solís, R. y Sancho, M. (2020). Folk theories of algorithmic recommendations on Spotify: Enacting data assemblages in the global South. *Big Data & Society*, 7(1). <https://doi.org/10.1177/2053951720923377>
- Stark, D. y Pais, I. (2020). Algorithmic Management in the Platform Economy. *Sociologica*, 14(3), 47-72. <https://doi.org/10.6092/issn.1971-8853/12221>
- Stoilova, M. y Livingstone, S. (2025). AI literacy in an unequal world: pitfalls and promises. London School of Economics. <https://blogs.lse.ac.uk/mediase/2025/10/27/ai-literacy-in-an-unequal-world-pitfalls-and-promises/> (acceso: 29-10-2025).
- Swart, J. (2021). Experiencing Algorithms: How Young People Understand, Feel About, and Engage With Algorithmic News Selection on Social Media. *Social Media + Society*, 7(2). <https://doi.org/10.1177/20563051211008828>
- Su, J., Ng, D. T. K. y Chu, S. K. W. (2023). Artificial Intelligence (AI) Literacy in Early Childhood Education: The Challenges and Opportunities. *Computers and Education: Artificial Intelligence*, 4, 100124. <https://doi.org/10.1016/j.caeai.2023.100124>
- Tadimalla, S. Y. y Maher, M. L. (2025). AI literacy as a core component of AI education. *AI Magazine*, 46, e70007. <https://doi.org/10.1002/aaai.70007>
- Tassinari, A. y Maccarrone, V. (2020). Riders on the Storm: Workplace Solidarity among Gig Economy Couriers in Italy and the UK. *Work, Employment and Society*, 34(1), 35-54. <https://doi.org/10.1177/0950017019862954>
- Vallas, S. P. y Cummins, E. R. (2015). Personal Branding and Identity Norms in the Popular Business Press: Enterprise Culture in an Age of Precarity. *Organization Studies*, 36(3), 293-319. <https://doi.org/10.1177/0170840614563741>
- Vallas, S. y Schor, J. B. (2020). What Do Platforms Do? Understanding the Gig Economy, *Annual Review of Sociology*, 46, 273-294. <https://doi.org/10.1146/annurev-soc-121919-054857>
- Van Deursen, A., Helsper, E. J., Eynon, R. y Van Dijk, J. (2017) The compoundness and sequentiality of digital inequality. *International Journal of Communication*, 11, 452-473.
- Van Deursen, A. J., & Helsper, E. J. (2015). The third-level digital divide: Who benefits most from being online?. In *Communication and information technologies annual*, 10, 29-52. <https://doi.org/10.1108/S2050-206020150000010002>
- Van Oort, M. (2013). Making the neoliberal precariat: Two faces of job searching in Minneapolis. *Ethnography*, 16(1), 74-94. <https://doi.org/10.1177/1466138113506636>
- Veen, A., Barratt, T. y Goods, C. (2019). Platform-Capital's 'App-etite' for Control: A Labour Process Analysis of Food-Delivery Work in Australia. *Work, Employment and Society*, 34(3), 388-406. <https://doi.org/10.1177/0950017019836911>
- Yarchi, M., Baden, C. y Kligler-Vilenchik, N. (2020). Political Polarization on the Digital Sphere: A Cross-platform, Over-time Analysis of Interactional, Positional, and Affective Polarization on Social Media. *Political Communication*, 38(1-2), 98-139. <https://doi.org/10.1080/10584609.2020.1785067>
- Vuorikari, R., Kluzer, S. y Punie, Y. (2022). DigComp 2.2: The Digital Competence Framework for Citizens - With new examples of knowledge, skills and attitudes. Publications Office of the European Union. <https://doi.org/10.2760/115376>
- Zarouali B., Boerman S. C. y De Vreese C. H. (2021) Is this recommended by an algorithm? The development and validation of the algorithmic media content awareness scale (AMCA-scale). *Telematics and Informatics*, 62, 101607. <https://doi.org/10.1016/j.tele.2021.101607>

## FINANCIACIÓN

Este trabajo se enmarca en el proyecto "Algorithmic literacy, búsqueda de empleo y exclusión digital" (PID2023-146586OB-100), financiado por el Ministerio de Ciencia, Innovación y Universidades en el Programa de Proyectos de Generación de Conocimiento 2023.



# Com podeu saber si feu un ús responsable de la vostra aplicació d'ia?

How can you know if you are making responsible us of your ai application?

**Arlet Brufau i Centelles**

*Universitat de Girona*

**Albert Sabater Coll**

*Universitat de Girona*

## PARAULES CLAU:

Intel·ligència artificial  
educació  
ètica  
legalitat  
avaluació.

## RESUM:

El Model PIO Educació té la voluntat d'esdevenir una eina d'autoavaluació referent en l'àmbit educatiu a través del qual la comunitat educativa pugui conèixer el grau d'adequació ètic i legal del seu sistema d'IA. L'avaluació consisteix en una suma de preguntes binàries fonamentals en la normativa i estàndards ètics i legals aplicables en matèria d'IA i educació que segueix l'estructura dels set principis ètics (transparència, justícia, seguretat, responsabilitat, privacitat, autonomia i sostenibilitat). Durant el seu ús permet a la persona usuària adquirir la capacitat i alfabetització en matèria d'IA i finalment, conèixer de manera fidedigna l'ur sistema. Per tant, es tracta d'un model de capacitació i autoconeixement del sistema d'IA per a la comunitat educativa fonamentat en la legalitat i l'ètica.

## PALAVRAS-CLAVE:

Inteligência artificial  
educação  
ética  
legalidade  
autoavaliação.

## RESUMO:

*Traduzido com DeepL.com*

O Modelo Educacional PIO pretende tornar-se uma ferramenta de autoavaliação de referência no campo da educação, através da qual a comunidade educativa pode verificar a adequação ética e jurídica do seu sistema de IA. A avaliação consiste em perguntas binárias baseadas nos regulamentos e normas éticas e jurídicas aplicáveis à IA e à educação, seguindo a estrutura dos sete princípios éticos (transparência, equidade, segurança, responsabilidade, privacidade, autonomia e sustentabilidade). Durante a sua utilização, permite ao utilizador adquirir formação e literacia em IA e obter uma compreensão fidedigna do seu sistema. É, portanto, um modelo de formação e autoconsciência do sistema de IA para a comunidade educativa, baseado na legalidade e na ética.

## KEYWORDS:

Inclusive education  
critical thinking  
higher education  
reflection  
educational policies

## ABSTRACT:

The PIO Education Model aims to become a benchmark self-assessment tool in the field of education, through which the educational community can ascertain the ethical and legal adequacy of its AI system. The assessment consists of binary questions based on the ethical and legal regulations and standards applicable to AI and education, following the structure of the seven ethical principles (transparency, fairness, security, accountability, privacy, autonomy and sustainability). During its use, it allows the user to acquire training and literacy in AI and to gain a reliable understanding of their system. It is therefore a model of training and self-awareness of the AI system for the educational community based on legality and ethics.

**CÓMO CITAR:** Brufau i Centelles, A., & Sabater Coll, A. (2025). Com podeu saber si feu un ús responsable de la vostra aplicació d'IA? *RETIS. Revista de Tecnología para la Inclusión Social*, 2(1), 23-30.  
DOI: [10.70664/retis.v2i1.003](https://doi.org/10.70664/retis.v2i1.003)

## 1. INTRODUCCIÓ

La creixent importància dels sistemes d'IA per a nombrosos aspectes de la nostra vida quotidiana (Floridi et al. 2018) i, especialment en l'àmbit educatiu (UNESCO, 2021), ha fet que es demani una major inclusió de consideracions ètiques per incrementar la confiança i avaluar els mateixos sistemes d'IA (Kastania, 2024) i aplicar-los de manera senzilla i pràctica (Mittelstad, 2019).

Si bé hi ha diversos sistemes d'IA d'automatització de criteris, de predicció o de percepció que introdueixen capacitats i potencialitats en diversos àmbits, tots ells requereixen una reflexió profunda sobre el seu ús ètic i el seu impacte social. En aquest sentit, des de l'Observatori d'Ètica en Intel·ligència Artificial de Catalunya (OEIAC), hem portat a terme una reflexió profunda en la que no tan sols considerem que és important establir uns principis ètics sòlids que guiïn la integració de la tecnologia, sinó també que la seva aplicació en àmbits crítics com l'educatiu sigui responsable i, per tant, beneficiï de manera justa i equitativa a tant professorat com alumnat.

El principal problema és que les directrius sobre les consideracions ètiques que estan disponibles són sovint massa complexes i difícils d'entendre, ja que estan escrites per a persones usuàries acadèmiques i tècnics (Memarian & Doleck, 2023). Això no només en dificulta la comprensió per a diferents parts interessades sinó que, a més a més, endarrerix l'adopció i l'aplicació com un element de valor d'innovació responsable (Morley et al., 2019). En conseqüència, s'ha convertit en una prioritat obrir el focus de les consideracions ètiques al voltant de la IA i traduir principis ètics en indicadors i observables de fàcil interpretació que proporcionin un marc d'acció, particularment a l'àmbit de l'educació (Chaudhry et al., 2022).

A més, en l'actualitat no podem fer referència als sistemes d'IA com a novetat perquè les aplicacions d'intel·ligència artificial generativa més mediàtiques ja fa tres anys que van irrompre mundialment i fins i tot, s'han presentat versions posteriors. És doncs, innegable, que la IA forma part de la nostra quotidianitat i és fonamental comprendre-la per poder-hi interactuar, sense tenir-li por, ni participar de l'exageració de llurs habilitats. Cal tenir en compte que ens trobem en una situació d'impàs perquè tots els elements que acompanyen la IA, necessaris per establir uns fonaments clars i sòlids per a integrar-la, s'estan desenvolupant o creant. Ens referim a elements com la manca de regulació integral, ja sigui internacional o normativa interna, de directrius provinents d'organismes internacionals, d'estudis rigorosos i fiables que també parlin de llurs efectes i conseqüències, així com l'existència d'organismes públics que acompanyin i aconsellin sobre la integració de la IA.

Per aquesta raó, és fonamental que totes les persones usuàries dels sistemes d'IA durant aquests anys d'adequació i integració, ens formem i autoavaluem els sistemes que volem utilitzar o ja utilitzem com a hàbit compromès que contribueix a un ús responsable de la IA. Actuant activament i assumint la responsabilitat, alhora que ens capacitem i desenvolupem l'habilitat d'utilitzar-la.

Conseqüentment, en aquest article, presentem el Model PIO Educació de l'OEIAC, el qual es tracta d'una eina d'autoavaluació que permet a les persones usuàries conèixer el grau d'adequació ètica i legalitat de llur sistema d'IA. El Model PIO plasma la governança en IA que està en consonància amb l'informe presentat pel mateix Jobin et al. (2019) sobre els valors globals de les estratègies d'ètica en l'IA, on la transparència i la responsabilitat, entre altres, són els principis més sol·licitats a la literatura actual (Floridi, 2019). Però més enllà d'una visió única de responsabilitat entorn de la transparència, aquesta es pot entendre també de dues maneres més. D'una banda, fa referència al fet que la pròpia tecnologia de la IA sigui translúcida i que hi hagi una obertura de la informació que s'utilitza. D'altra banda, hi ha la transparència de les organitzacions que desenvolupen i fan servir sistemes d'IA. Si bé la primera és discutida regularment i directament, o en relació amb els processos requerits per garantir-la en termes d'explicitabilitat, comprensió i comunicació; la segona està en un segon pla i ha d'avançar per garantir i protegir altres requisits com els drets humans fonamentals, la privadesa, la dignitat, l'autonomia i el benestar (UNI Global Union, 2017).

És innegable que l'ús i impacte de la IA introdueix capacitats i potencials tecnològics inèdits esdevenint una eina clau en l'evolució de la societat i per extensió, també, a l'àmbit educatiu. Algunes d'aquestes qualitats poden ser facilitar l'accés al coneixement i la diversificació de mètodes d'ensenyament, però al mateix temps hem de poder avaluar el paper crític d'aquestes tecnologies en l'entorn d'aprenentatge (Riley & Bruno, 2024). Ara bé, hem de matisar dues qüestions al voltant de la tecnologia que genera confusió i dificulta a la comunitat educativa comprendre i acompanyar aquest procés. En primer lloc, la tecnologia i la IA són diferents i no ens hi podem referir com a sinònims perquè no és el mateix una calculadora, on conservem tota la nostra agència, que els sistemes d'IA on la interlocució està limitada degut a la quantitat de dades d'entrenament que utilitza. En segon lloc, contribuïm ha abandonar el discurs intencionat que la tecnologia és neutre per naturalesa perquè incorpora i arrossega valors i creences des que es programa i es desenvolupa els quals es reflectiran al resultat. Així mateix, sabem que l'ús de la tecnologia i de la IA no sempre és positiva ni desitjable i per això, hem de poder establir un principi marc de prudència (Steels y López de Mántras, 2018).

Tot seguit, afegim la importància dels principis ètics els quals contribueixen a l'adopció de la IA com una base sòlida que orienten l'ús i garanteix que la tecnologia serveixi com a eina per a enriquir l'educació respectant els drets i la dignitat de la comunitat educativa. En conseqüència ha esdevingut una prioritat traduir els principis ètics en indicadors i observables de fàcil interpretació que proporcionin un marc d'acció (Chaudhry & al, 2022). A més a més, la suma de la regulació és fonamental i necessària per a l'ús responsable i en benefici de la societat dels sistemes d'IA jugant el doble paper de promoure la confiança de la societat i tutelar els drets de les persones usuàries (Croley, 2008).

Tanmateix, l'adequació amb l'ètica permet establir estàndards al voltant dels set principis dotant-los de reconeixement i esdevenint transcendents. Aquests són la transparència i explicabilitat, la justícia i equitat, la seguretat i no maleficència, la responsabilitat i retiment de comptes, la privacitat, l'autonomia i la sostenibilitat (AIEI Group, 2020) els quals estan reconeguts als estàndards internacionals per a les fases de disseny, implementació, comercialització i ús dels sistemes d'IA.

La composició d'ambdós elements donen lloc a accions i mesures les quals estableixen el camí cap a una IA responsable i sostenible. I per fer-ho possible, traduïm els principis ètics en indicadors clars i mesurables a nivell qualitatiu i quantitatiu que ens permet avaluar els sistemes d'IA en un marc regulador i ètic. D'una banda, això permet eliminar la dificultat de comprensió que acompanya l'àmbit jurídic i facilita que la comunitat educativa adopti l'ètica i la regulació; i d'altra banda, desenvolupa la tasca de

capacitació i alfabetització de tots els subjectes que interactuen amb el sistema d'IA complint amb l'obligació establerta a l'article 4 del Reglament (UE) 2024/1680 del Parlament Europeu i el Consell sobre Intel·ligència Artificial (RIA).

És per això que el Model PIO Educació dota a la comunitat educativa d'una eina gratuïta i d'accés públic que ajuda a conèixer el sistema d'IA que volen integrar al procés d'aprenentatge. A més a més, a través de l'avaluació es produeix un procés d'educació i capacitació de tots els subjectes que interactuen, sigui alumnat, professorat, famílies o altres persones de l'entorn educatiu.

La següent secció tractarà els antecedents on recollim l'anàlisi conjuntural i la diagnosi que vam realitzar. La tercera secció tractarà de les raons per les quals proposem el Model PIO Educació fruit de l'impacte dels sistemes d'IA a l'entorn educatiu. La quarta secció tractarà de l'estructura que segueix, el mètode que hem dut a terme per a elaborar-lo i els fonaments jurídics i ètics que fonamenten el model. La cinquena secció tractarà del model pròpiament explicant-lo i tot seguit, presentem altres recursos que hem elaborat. Per acabar, la secció de discussió i conclusions on finalitzaran amb les nostres contribucions i futures línies de recerca.

## 2. PER QUÈ EL MODEL PIO EDUCACIÓ?

El terme *Human Centred Artificial Intelligent* (HCAI) o intel·ligència artificial centrada en les persones garanteix una IA responsable on les persones proveïdores i les desenvolupadores han dut a terme llurs tasques d'acord amb els valors ètics i prioritant el benestar de les persones usuàries garantint el desenvolupament sostenible (Dignum, V, 2019). Aquest enfocament en l'aplicació de la IA permetrà minimitzar l'existència dels riscos dels sistemes els quals cal conèixer per a ponderar i prendre decisions. Alguns d'aquests riscos són els perills que la IA pot suposar pels drets més essencials, i conèixer les precaucions que s'haurien d'adoptar per tal de garantir que es converteix en una eina en favor de la humanitat, i no en contra d'ella (Comissari per als Drets Humans del Consell d'Europa, 2023). Així mateix, la intersecció entre la IA i l'àmbit educatiu posa de relleu la consideració de l'educació com un factor fonamental pel desenvolupament d'un país. Tal i com queda palès en el document de Nacions Unides número quatre sobre l'Objectiu de Desenvolupament Sostenible (ODS) el qual té la proposta de garantir una educació inclusiva, equitativa i de qualitat que promou l'oportunitat d'aprenentatge durant tota la vida.

L'impacte per a la comunitat educativa dels sistemes d'IA es produeix de manera multifactorial i diferent en funció del subjecte destinatari amb la característica d'evolució ràpidament.

D'una banda, ens trobem amb un context desigual respecte a l'ús i al coneixement sobre la IA de tota la comunitat educativa. Les eines o els mitjans per a poder-la utilitzar a l'aprenentatge està estretament vinculada a la motivació de cada docent sense unes directrius clares i uniformes per part d'organismes superiors. De l'altra, l'alumnat presenta casuístiques i una relació molt diversa amb els sistemes d'IA on també opera l'esclat digital la qual consisteix tant en l'accés a la connexió internet, com en la desinformació i desconeixement de l'ús responsable i ètic de la IA. Finalment, la funció dels protocols de centre on establir l'ús i els criteris de contractació dels sistemes d'IA esdevé fonamental per a evitar una tria arbitrària i subjectiva que fonamentarà aquest ús ètic i responsable.

A més a més, hem de sumar el factor de la desconfiança de les persones usuàries vers els sistemes d'IA degut a que els models actualment els hi manca transparència o presenten dubtes sobre el seu funcionament i la programació dels algorismes. Així mateix, el desconeixement de la normativa pertinent i la complexitat per a comprendre la legislació aplicable allunyen aquestes eines de la comunitat educativa.

En un tercer nivell, avaluem la societat la qual es mostra prudent en el fet que dita innovació es dugui a terme amb un grup vulnerable de la població com és l'alumnat i als centres educatius perquè ens manquen estudis rigorosos i informació sobre les conseqüències. Sobre tot, una identificació d'aquells efectes secundaris o subtils que no podem detectar fins a mitjà o llarg termini de l'ús de sistemes d'IA en l'aprenentatge.

El punt següent a tractar són els debats al voltant del funcionament i l'estructura del sistema educatiu que ens planteja l'adequació de la IA. D'entrada, plantejar-nos si la IA pot substituir al professorat, tot abordant les capacitats i els límits de la tecnologia en equilibri amb els avantatges de la innovació i els valor insubstituïble de les relacions humanes per a l'aprenentatge. En aquest punt, hi ha un seguit de raons que fan impossible que això passi com la capacitat d'emetre judicis des d'una perspectiva professional, el valor emocional i pedagògic i la retroalimentació constant de la tasca dels docents.

En primer lloc, el professorat té la capacitat de prendre decisions informades i sensibles a l'aprenentatge i el benestar de l'alumnat gràcies a les metodologies pedagògiques i didàctiques, el coneixement de la matèria, l'experiència professional i el domini de la dinàmica de la seva aula. En segon lloc, la relació de confiança i la comprensió que s'estableix entre el docent i l'alumne (i tota la comunitat educativa) crea un entorn segur i de suport on l'alumnat es sent valorat i motivat per desenvolupar el seu procés d'aprenentatge cometent errors i emetent dubtes. En últim terme, la tasca del professorat es fomenta una atmosfera que propici l'esperit crític, el creixement intel·lectual i l'estimulació en l'aprenentatge impulsant a l'alumnat a millorar i per tant, va molt més enllà de corregir errors.

Tot seguit, i molt enllaçat amb el dilema anterior, apuntem la capacitat de més o menys personalització de l'aprenentatge dels sistemes d'IA adaptat al ritme de cada alumne. Ara bé, presenta possibilitats com l'adaptació i acompanyament, i també riscos com la individualització, l'estandardització, l'homogeneïtzació i en definitiva, l'automatització de l'aprenentatge (Lara, T. & Magro C., 2025).

Es a dir, si la personalització de l'aprenentatge es redueix a oferir opcions i continguts que s'activen segons les respostes de l'alumnat en un sistema d'opcions múltiples, ens trobem davant d'un conjunt limitat d'alternatives preestablertes. Per tant, més que personalització esdevé estandardització que pretén simular un tracte individualitzat de l'aprenentatge allunyat de la diversificació de l'ensenyament. (Farelli, M, 2024, Intel·ligència artificial y educación: insumos para su abordaje desde Iberoamérica).

I encara més, si el professorat passa a ser facilitador de l'aprenentatge degut a l'automatització d'aquest, caracteritzat per la planificació i el control, es reduirà les incerteses i els dubtes de l'alumnat i provocarà una autolimitació d'allò que pensen i les opcions possibles que tenen. Aquest canvi de llenguatge no és baladí i apropa l'educació a una lògica instrumental i tecnocràtica allunyant-la del seu sentit.

En aquest mateix punt, procedim a analitzar l'impacte cognitiu de la IA, sobre tot dels sistemes d'intel·ligència artificial generativa (IAG). Referenciem diferents estudis sobre l'augment de la delegació del procés d'avaluació crítica als sistemes d'IAG (Hao-Ping, L & al., 2025) els quals impacten a les àrees del record i la comprensió, la manera de recopilar informació i verificar-la i finalment, la percepció d'un problema i la possible solució. Per tant, es produeix l'impacte a les capacitats de l'alumnat durant l'ús de les plataformes educatives perquè disposen una confiança absoluta i cega en els sistemes d'IAG com una guia veraç i infal·lible indispensable per a llur aprenentatge. Aquesta situació rep el nom de biaix d'automatització (Stadler, M. & al., 2024).

En últim terme i connectat amb aquest biaix, tot i que de manera més general, apuntem el fenomen "delegació de la màquina" (Köbis, N. & al., 2025). Aquest consisteix en la delegació que les persones fan de llurs tasques o automatització d'aquestes, de la presa de decisions i consultes o dubtes relatius a tots els àmbits de la vida als sistemes d'IA. La capacitat humana de prendre decisions en els debats ètics i morals que puguem trobar-nos, no es assumible pels sistemes d'IA els quals executen llur programació sense reflexió ni ponderació dels diferents dilemes. Per tant, a causa de l'augment d'utilització dels sistemes d'IA, es constata que les persones prenen decisions de manera automàtica sense realitzar els debats ètics i morals pertinents donant lloc a comportaments negatius.

Davant d'aquests dilemes, l'aplicabilitat dels principis ètics és la garantia que els sistemes d'IA s'adeqüen als estàndards d'una eina responsable i de confiança. Això voldrà dir: son utilitzats de manera transparent i explicable; en defensa de la justícia i conscients dels perjudicis que pot ocasionar; segurs, garantint el benestar i amb un retiment de comptes; respectuosos amb la privacitat de les dades; no influiran en l'autonomia ni l'agència de les persones usuàries; i duren a terme un ús sostenible dels recursos naturals i garanteixen els drets humans. Aquesta confiança fomenta un entorn d'aprenentatge on el sistema d'IA millora l'experiència educativa i acompanya a l'alumnat.

En conseqüència, el Model PIO Educació és una eina de capacitació i alfabetització per a tota la comunitat educativa en l'ús dels sistemes d'IA que li permet conèixer-lo i de manera general entendre'n el funcionament.

### 3. L'ESTRUCTURA

El Model PIO Educació consisteix en un sistema d'avaluació que té l'estructura d'una llista de verificació o *checklist* de preguntes binàries les respostes de les quals reflecteix el grau d'adequació del sistema d'IA al estàndards ètics i legals. L'objectiu és poder auditar el sistema d'IA i les dades que s'utilitzen per tal d'anticipar o avaluar els possibles impactes en l'àmbit educatiu.

Un altre punt és el fonament del Model PIO Educació, ja que es tracta d'un model d'autoavaluació crític i honest que permet a la persona usuària del sistema d'IA conèixer-lo i saber si s'adequa als principis ètics i la normativa legal existent. Es du a terme partint de l'auto-responsabilitat com a element fonamental per a la veracitat del resultat, ja que les persones usuàries responen al qüestionari sense supervisió. A més a més, les preguntes del qüestionari parteixen del principi de senzillesa sempre al voltant de la pregunta "ho hem fet?".

Pel que fa a la justificació del Model PIO Educació recopila la normativa i els textos i recomanacions que interaccionen ambdues matèries, sistemes d'IA i educació.

Per això, respecte a la normativa, a part del ja esmentat RIA, hem inclòs: el Conveni marc del Consell d'Europa sobre IA i drets humans, democràcia i estat de dret; Conveni 108 del Consell d'Europa per a la protecció de les persones físiques pel que fa al tractament de dades personals; la Directiva 2024/2853 sobre responsabilitat pels danys causats per productes defectuosos; Directiva 2019/882, de 17 d'abril del 2019 [Directiva sobre els requisits d'accessibilitat dels productes i serveis]; la Directiva 2019/790 del Parlament Europeu i del Consell, de 17 d'abril de 2019, sobre els drets d'autor i drets afins en el mercat únic digital; la Directiva 2016/2102 de 26 d'octubre de 2016 [Directiva sobre l'accessibilitat dels llocs webs i aplicacions per a dispositius mòbils del sector públic]; la Directiva 2001/29/CE, de 22 de maig del 2001 [Directiva dels drets d'autor i drets afins en la societat de la informació]; el Reglament 2016/679, del Parlament Europeu i del Consell, de 27 d'abril de 2016, relatiu a la protecció de les persones físiques pel que fa al tractament de dades personals i a la lliure circulació d'aquestes dades [Reglament general de protecció de dades]; i el Reglament 2022/868 del Parlament Europeu i del Consell, de 30 de maig del 2022, relatiu a la governança europea de dades [Llei Europea sobre la Governança de dades].

Respecte a l'àmbit de regulació de l'estat membre, l'educació té la naturalesa de dret fonamental i està garantida a l'article 27 de la Constitució espanyola on a l'apartat primer recull que es tracta d'un dret inherent a totes les persones. I per aquesta raó, es regula de manera integral a la Llei Orgànica 3/2020, del 29 de desembre, per la qual es modifica la Llei Orgànica 2/2006, de 3 de maig, d'Educació.

Tanmateix, hem referenciat la literatura acadèmica especialitzada i les recomanacions i llibres blancs en matèria d'IA i educació que contribueixen a enriquir l'avaluació i son respectuosos amb els acords i normes socials. Com l'eina creada per la Comissió amb el títol "Espai europeu de l'educació" (*European Education Area*); les Directrius ètiques per a una IA fiable (*Ethics guidelines for trustworthy AI*) de la Comissió Europea; o les Recomanacions sobre l'ètica de la IA el qual consisteix en una guia global adoptada pels cent noranta-tres estats membres de la UNESCO establint pautes per assegurar que el desenvolupament i ús de la IA segueixi els principis ètics.

En últim terme, el subjecte al qual va dirigit aquest model és tota la comunitat educativa entesa com les persones i les institucions que intervenen en el procés educatiu de manera directa o no, independentment del seu paper. Així doncs, en formen part: l'alumnat, tota persona que rebí una educació reglada i no reglada independentment de l'edat i l'etapa educativa en la qual es trobin; el professorat qui integra i utilitza els sistemes d'IA, ja sigui per a crear llurs seqüències didàctiques, el material que utilitzarà

per a les sessions o en el transcurs de les mateixes; dins dels centres educatius, els professionals d'atenció educativa i el personal d'administració i serveis, ja que es tracten de professionals que també intervenen en l'educació i cal garantir que també gaudeixen d'aquesta formació i coneixement dels sistemes d'IA; i també inclouem les famílies de l'alumnat, sobretot en els casos que sigui menor d'edat, per tal que puguin tutelar llurs drets, a més a més, d'adquirir la capacitat i formació adient en els sistemes d'IA.

## 4. EL MODEL PIO EDUCACIÓ

### 4.1. L'AVALUACIÓ

El Model PIO Educació és una eina d'avaluació del sistema d'IA utilitzat en l'àmbit educatiu gratuïta i consultable a la pàgina web de l'OEIAC ([oeiac.cat](http://oeiac.cat)) que permet a les persones usuàries conèixer quin és el grau d'adequació a l'ètica i la legalitat del seu sistema d'IA.

Aquest formulari d'avaluació consisteix en un llistat de verificació o *checklist* de preguntes binàries que es responen en funció de si el sistema compleix o no amb la informació que es pregunta. Sempre hi ha la tercera opció de "no aplica" per als casos que el sistema d'IA no ho prevegi. Així mateix, les cent onze preguntes s'elaboren i es fonamenten a partir de la recopilació de les obligacions extretes del RIA, dels altres textos normatius i específics per a l'àmbit educatiu, de les recomanacions ètiques i legals i també els textos acadèmics.

Pel que fa a l'estructura, s'articula seguint els set principis ètics a través dels quals s'agrupen les preguntes i el significat de cadascun és: El primer, la transparència i l'explicabilitat, engloba la comunicació clara i la comprensió de com funcionen les eines d'IA, com s'utilitzen en entorns educatius i les implicacions del seu ús per a totes les parts interessades, inclosos professorat, alumnat, famílies o tutors. A més a més, asseguren que el desplegament dels sistemes d'IA no és una caixa negra, sinó un llibre obert que convida a l'escrutini, la comprensió i el consentiment informat. Implica revelar les capacitats i limitacions en funció de cada procés educatiu, com també, qualsevol biaix potencial inherent als seus algorismes, que podria influir de manera negativa en el contingut educatiu i les interaccions de l'alumnat; El segon principi és la justícia i l'equitat el qual posa èmfasi en la inclusió i els resultats equitatius per a tots els membres de la comunitat educativa i requereix que les tecnologies d'IA es dissenyin, implementin, comercialitzin i s'utilitzin de manera que tinguin en compte i es respectin la diversitat d'origens, identitats i necessitats dins de l'entorn escolar; El tercer és la seguretat i la no maleficència el qual requereix la robustesa del funcionament i resistència en vers d'atacs maliciosos o intent de pirateig. A més a més, suposa el desplegament i l'ús dels sistemes d'IA no afectin negativament el benestar de l'alumnat, professorat i la comunitat educativa, en general. Principalment, implica un anàlisi de com els sistemes d'IA afecten a la salut mental i prendre les mesures proactives necessàries per mitigar els possibles efectes negatius garantint el pensament crític i la creativitat respectant l'alumnat i el professorat; El quart principi és la responsabilitat i el retiment de comptes el qual implica garantir el compliment dels principis ètics i coneixement de l'impacte dels sistemes d'IA en l'ensenyament, l'aprenentatge i el benestar de la comunitat educativa; El cinquè principi és la privacitat la qual garanteix que es respecti i protegeixi la informació personal i el dret a la confidencialitat i gestió de les dades de totes les persones de l'ecosistema educatiu que s'han introduït. Això suposa implementar mesures sòlides de protecció de dades, polítiques transparents d'ús de dades i donar als individus control sobre la seva pròpia informació; El sisè principi és l'autonomia el qual implica que els sistemes d'IA respectin i promoguin la independència en la presa de decisions de totes les persones de l'ecosistema educatiu. Això vol dir que la IA ha de donar suport a la capacitat de presa de decisions informades sobre els processos d'ensenyament i aprenentatge del professorat i l'alumnat i en cap cas, no pot ser per menystenir l'autoritat del professorat o el control de l'estudiant sobre el seu recorregut d'aprenentatge; Finalment, el principi de sostenibilitat el qual implica garantir que el desplegament es realitzi d'una manera ambientalment responsable, econòmicament viable i socialment equitativa respectant els drets i la regulació, tant en la defensa de les persones treballadores, en matèria de propietat intel·lectual com dels drets d'autoria.

Així mateix, cada principi es divideix en taules segons la temàtica que aborden, de manera que cada pregunta es troba inclosa en una casella. El requadre de la qüestió a respondre ve acompanyat de dues altres caselles. La primera conté les referències legals, ètiques i acadèmiques en què es basa la pregunta, i la segona exposa una recomanació sobre la pregunta que permeti a les usuàries comprendre millor el contingut que es demana sent una guia.

Pel que fa a la suma de les respostes a totes les preguntes donen lloc a uns percentatges que atribueixen el grau d'adequació a cada un dels set principis i en general, avalua el grau de proximitat del sistema d'IA de l'ètica i la legalitat. Aquests percentatges es tradueixen en un conjunt de mètriques les quals són el reflex fidedigne del sistema d'IA partint d'una confiança implícita de la declaració de responsabilitat sobre la informació donada al qüestionari.

En finalitzar l'avaluació s'atribueix una insígnia gràfica descarregable que reflecteix la suma de les mètriques obtingudes amb les respostes donades. Com també s'obté la informació detallada per a cadascun dels principis i el percentatge assolit d'aquest mitjançant un informe que es rep a l'adreça de correu de registre que permetrà ser coneixedors dels punts forts i els febles del sistema d'IA per a poder millorar-los i mitigar-ne els possibles resultats esbiaixats.

### 4.2. LES PREGUNTES INICIALS

Aquesta eina permet a les persones usuàries de la comunitat educativa aproximar-se al sistema d'IA que volen fer servir o que ja utilitzen mitjançant deu preguntes bàsiques. A través del contingut d'aquestes es dibuixa les característiques essencials que ha de tenir l'eina i permeten a la persona usuària aproximar-se al model d'autoavaluació i familiaritzar-se abans de dur-lo a terme.

Aquestes preguntes són les següents: Coneixeu les oportunitats i limitacions dels sistemes d'IA per entendre i respondre a la dinàmica que genera amb cada persona usuària des de l'òptica de les emocions?; Coneixeu les conseqüències subtils dels sistemes d'IA per a les persones usuàries?; Sabeu si l'ús de la IA a l'aula pot generar un impacte per al sentiment de pertinença i comunitat de l'alumnat?; Podeu identificar les oportunitats i limitacions de la IA per oferir suport específic, estímuls i intervencions edu-

catives específiques?; Coneixeu de manera òptima els sistemes d'IA per a poder seleccionar i aplicar la més adequada per a cada acció formativa considerant la perspectiva ètica?; Podeu guiar l'alumnat a través dels reptes d'aprenentatge i desenvolupament personal que presenta la IA a l'educació?; Podeu assegurar que amb l'ús de sistemes d'IA us centreu en els objectius d'aprenentatge específics i consideracions ètiques?; Coneixeu les amenaces i reptes dels sistemes d'IA per a l'alumnat?; Podeu assegurar que la divulgació i l'explicació dels sistemes d'IA es du a terme de manera responsable i eficaç en els entorns acadèmics?; i Compteu amb un mecanisme de responsabilitat en cas que l'ús del sistema d'IA causi algun mal a l'alumnat?.

Per tant, veiem com el contingut de les preguntes van des del coneixement del sistema d'IA, concretament les oportunitats i les limitacions que presenta i llurs conseqüències. Una segona esfera analitza la relació que desenvolupa amb l'alumnat, es a dir, l'impacte als seus sentiments, els beneficis i les limitacions que li suposarà per a l'aprenentatge. Des d'una perspectiva més genèrica s'analitza el domini que te el professorat del sistema d'IA que permet adaptar-lo a cada alumne i guiar-los. Així mateix, es pot dedicar a objectius d'aprenentatge específic i de l'altra, estableix una relació transparent entre l'alumnat i el sistema d'IA. I finalment, la importància d'una assegurança de responsabilitat civil per tal de donar resposta a les conseqüències negatives que puguin sorgir.

### 4.3. GUIES RÀPIDES DE PREGUNTES I RECOMENACIONS

Un dels neguits més freqüents de la comunitat educativa es la manca de directrius d'actuació i capacitació que els permeti establir un criteri de presa de decisions i un ús clar dels sistemes d'IA. Tot això amb un llenguatge comprensible i dinàmic que elimini les barreres i distàncies amb la tecnologia fent-la realment entenedora.

Per aquesta raó, vam elaborar les guies que són una eina pensada per a uns subjectes molt concrets, el professorat i l'alumnat. Alhora, es componen de dos apartats, d'una banda, les preguntes essencials de cadascun dels set principis que s'han de formular on les respostes afirmatives implicaran un sistema d'IA ètic i legal. I de l'altra banda, el llistat de les recomanacions, seguint la mateixa estructura, amb la voluntat d'aportar unes directrius d'actuació relacionades amb les preguntes que proposa millores al sistema d'IA per tal d'assolir els estàndards desitjats.

Una de les principals característiques d'aquestes guies és la síntesis d'aquells elements fonamentals de cada principi, en funció del subjecte, creant un recull de la informació essencial dels sistemes d'IA ètics i legals.

Encara més, l'ús d'un llenguatge formulat en primera persona i un ritme dinàmic del redactat contribueixen a trencar la barreira que el llenguatge tecnològic moltes vegades estableix fent particip i responsable a la persona usaria, ja que es sent interpel·lada. En aquest sentit, i amb el mateix objectiu, el vocabulari emprat és formal, tot i que, no tècnic perquè no sigui un impediment o fre per a la capacitació i alfabetització dels sistemes d'IA.

Fent referència a la guia dirigida al professorat destaquem les preguntes amb llurs recomanacions corresponents com: Tens garanties per a protegir-vos de la informació inexacta o interaccions que poden ser perjudicials? Prioritzar la tutela de la transparència del sistema i no acceptar la informació inexacte o les interaccions perjudicials per a l'alumnat; Coneixes la informació detallada de les dades i llur procedència que alimenten els sistemes d'IA? El sistema d'IA ha de proporcionar un mecanisme de seguretat clar respecte al risc que pot ocasionar als drets fonamentals; El sistema d'IA no afecta negativament a la salut mental de l'alumnat, el professorat ni la comunitat educativa? L'ús del sistema d'IA no pot afectar negativament en cap cas a la salut, ja sigui física o, especialment, mental, els drets de l'alumnat i la seguretat del sistema. Estàs exagerant el que pot fer el teu sistema d'IA? No exagereu la hipèrbola dels sistemes d'IA sumant-vos als discursos que els hi atribueixen capacitats màgiques amb funcions il·limitades, ja que en cap cas podran substituir les vostres funcions essencials ni aptituds. La persona responsable està capacitada i formada per a exercir aquestes funcions? Establiu qui serà la persona responsable de les dades introduïdes al sistema i llur responsabilitat legal. En aquest sentit la manera de designar-la, la durada del càrrec i les funcions. Has pres mesures que garanteixin que l'alumnat sigui conscient que interactua amb una IA? En tot moment ho ha de ser i se l'ha d'informar amb mecanismes que assegurin la seva consciència i comprensió; Has tingut en compte l'accés social, incloent-hi l'esclatxa digital, en l'ús i implementació del sistema d'IA? Ha de ser una prioritat abans d'introduir els sistemes d'IA a l'aprenentatge.

Pel que fa a la guia dirigida a l'alumnat destaquem les següents: Pots identificar si el contingut que obtens amb un sistema ha estat creat per una IA? El sistema d'IA sempre ha d'indicar que el resultat ha estat elaborat mitjançant intel·ligència artificial; El sistema d'IA s'adapta a les teves preferències i habilitats? Tan en l'ús com en el resultat el sistema d'IA s'ha de poder adaptar a cada alumne; Coneixes els possibles riscos o danys que et provocaria el sistema d'IA si el fas servir per a usos diferents als indicats pel professorat? L'ús del sistema d'IA més enllà dels previstos generarà riscos i pot causar danys; Saps com i de quina manera influeix el sistema d'IA en la presa de les teves decisions? El sistema d'IA ha d'estar sempre sota supervisió i mesures de correcció humanes per poder controlar llurs resultats; Coneixes el risc que comporta cedir les teves dades en sistemes d'IA? Has de conèixer el teu sistema d'IA, es a dir, quin risc tenen les dades una vegada introduïdes, el risc de la pèrdua de les dades, l'impacte a la privacitat i la traçabilitat; El sistema d'IA t'encoratja a desenvolupar una simpatia, empatia o inclinació que sobreestimi llurs capacitats? En cap cas el sistema d'IA et pot encoratjar ni despertar el sentiment de sobreestimar les capacitats humanes en favor d'aquest; Coneixes la mitjana de vida útil del sistema d'IA? Has de poder saber-la i el cost econòmic i l'impacte ambiental de llur manteniment.

En conclusió, les guies son el recull d'aquell contingut de l'avaluació considerat essencial i constitutiu del nucli essencials dels sistemes d'IA responsables recollits en forma de pregunta i recomanació perquè es pugui consultar de manera més dinàmica.

## 5. DISCUSIÓ I CONCLUSIONS

L'aplicació del Model PIO és important des d'un punt de vista de governança i d'aplicació d'una IA responsable, on l'aplicació de principis ètics ha de servir també per desmitificar el funcionament de la IA i, alhora, per avaluar críticament el paper d'aquestes tecnologies en els entorns d'aprenentatge (Riley & Bruno, 2024). Això fomenta un entorn on es poden prendre decisions informades sobre la integració de la tecnologia i, en última instància, permet el desplegament del consentiment informat, ja que garanteix que els membres de la comunitat escolar siguin conscients de com afecten les tecnologies de l'IA el procés educatiu.

Així doncs, recollint el més important, l'ús de les aplicacions d'IA ha de ser responsable i això vol dir, complir amb els estàndards ètics i la normativa legal aplicable en totes les seves etapes, és a dir, des del disseny, la programació, la comercialització i finalment, la utilització. A més a més, l'alfabetització i la capacitat de la comunitat educativa serà essencial per aprofitar els beneficis i detectar els riscos, i sobretot, poder-los mitigar.

Convé ressaltar l'impacte dels set principis ètics pel seu valor al procés de comprensió de la IA, com també, la funció que fan com a directrius d'actuació i integració en l'aprenentatge. Son una brúixola en un context d'impàs mentre la normativa es consolida i es publiquen regulacions extenses, els organismes públics ofereixen directrius concises per a tots els àmbits, l'acadèmia elabora investigacions profundes sobre l'evolució i conseqüències i la societat es capacita.

En particular, l'impacte dels sistemes d'IA a l'aprenentatge col·lisiona amb les capacitats i aptituds que l'alumnat adquireix i desenvolupa al llarg d'aquests anys, com és la capacitat cognitiva, l'autonomia i la creativitat. Alhora, posa en risc principis com l'equitat, la transparència i la confiança, els quals han esdevingut valors fonamentals de l'educació. A més a més, no podem oblidar factors com l'edat de l'alumnat que utilitzarà aquest sistema d'IA, ni les capacitats econòmiques que l'acompanyen i condicionarà l'ús donant lloc a l'esclatxa digital.

En conseqüència, el Model PIO Educació es tracta d'una eina d'autoavaluació que permet a les persones usuàries conèixer el propi sistema d'IA i capacitar-les en l'ús, beneficis i riscos per poder actuar conseqüentment. Globalment, permet a les usuàries detectar les limitacions o punts de millora i utilitzar-lo d'acord amb això extraient el màxim benefici.

Finalment, en aquest present curs escolar iniciarem l'estudi i avaluació pràctica del model amb professionals i centres educatius per avaluar-ne l'impacte per tal que el Model PIO Educació sigui adaptat a la realitat de la comunitat educativa i la puguin utilitzar com una eina essencial que acompanya al seu sistema d'IA.

## 6. REFERENCIES BIBLIOGRÀFIQUES

- Alon-Barkat, S. & Busuioc, M. (2022). *Human-AI interactions in public sector decision making: "Automation bias" and "selective adherence" to algorithmic advice*. Journal of Public Administration Research and Theory. [https://www.researchgate.net/publication/358515845\\_Human-AI\\_Interactions\\_in\\_Public\\_Sector\\_Decision-Making\\_%27Automation\\_Bias%27\\_and\\_%27Selective\\_Adherence%27\\_to\\_Algorithmic\\_Advice](https://www.researchgate.net/publication/358515845_Human-AI_Interactions_in_Public_Sector_Decision-Making_%27Automation_Bias%27_and_%27Selective_Adherence%27_to_Algorithmic_Advice).
- AI Ethics impact group. (2020). *From Principles to Practice – An interdisciplinary framework to operationalise AI ethics*. <https://www.ai-ethics-impact.org/en>.
- Aragüez Valenzuela, L. (coord); Sabater, A. & Brufau, A.; González Hernández, A.; Lozano Lares, F.; Salas Porras, M.; Rueda Monroy, J.A.; Giménez Murugarren, M.P.; Sánchez González, M.; Benítez Saña, R.M.; & Ruiz Mosquera, A.C. (2025). *La IA en la docència universitària. Una guia de buenas prácticas basada en la ética, responsabilidad e innovación*. Tirant humanides. <https://editorial.tirant.com/es/ebook/la-ia-en-la-docencia-universitaria-una-guia-de-buenas-practicas-basada-en-la-etica-responsabilidad-e-innovacion-lucia-aragüez-valenzuela-9788410813083>.
- Bastani, H.; Bastani, O.; Sungu, A.; Ge, H.; Kabakci, O.; Mariman, R. (2024). *Generative AI Can Harm Learning*. The Wharton School Research. <https://scale.stanford.edu/ai/repository/generative-ai-can-harm-learning>.
- Bauer, E.; Greiff, S.; Graesser, A. C.; Sailer, M.; & Scheiter, K. (2025). *Looking Beyond the Hype: Understanding the Effects of AI on Learning*. Springer Nature. <https://link.springer.com/article/10.1007/s10648-025-10020-8>.
- Brufau i Centelles, A.; Sabater Coll, A.; Gisbert Cervera, M.; Ametller, J.; Lillo Campoy, A. (2025). *El Model PIO (Principis, Indicadors i Observables) Educació: Un enfocament d'autoavaluació per al compliment dels estàndards ètics i normatius aplicables al sistema d'intel·ligència artificial en l'àmbit de l'educació*. Oficina Edicions UdG : Observatori d'Ètica en Intel·ligència Artificial de Catalunya. *El Model PIO (Principis, Indicadors i Observables) Educació: Un enfocament d'autoavaluació per al compliment dels estàndards ètics i normatius aplicables al sistema d'intel·ligència artificial en l'àmbit de l'educació*.
- Chaudhry, M.A.; Cukurova, M.; & Luckin, R. (2022). A Transparency Index Framework for AI in Education. *In International conference on artificial intelligence in education (pp. 195-198)*. <https://arxiv.org/abs/2206.03220>.
- Cortes Generales. (2024). *Ley Orgánica 2/2006, de 3 de mayo, de Educación*. <https://www.boe.es/buscar/pdf/2006/BOE-A-2006-7899-consolidado.pdf>.
- Cortes Generales. (1978). *Constitució Espanyola*. [Constitución Española](https://www.boe.es/buscar/pdf/1978/BOE-A-1978-2853-consolidado.pdf).
- Council of Europe. (2024). *Council of Europe Framework Convention on Artificial Intelligence and Human Rights, Democracy and the Rule of Law*. <https://rm.coe.int/1680afae3c>.
- Council of Europe. (2018). *Convention 108 + Convention for the protection of individuals with regard to the processing of personal data*. <https://rm.coe.int/convention-108-convention-for-the-protection-of-individuals-with-regard-to-the-processing-of-personal-data>.
- Croley, S.P. (2008). *Regulation and Public Interests: The Possibility of Good Regulatory Government*. Princeton University Press. <https://www.jstor.org/stable/j.ctt7rsdk>.
- Dignum, V. (2019). *Responsible artificial intelligence: How to develop and use AI in a responsible way* Cham: Springer. <https://link.springer.com/content/pdf/10.1007/978-3-030-30371-6.pdf>.
- European Commission. (2024). *European Education Area*. [Página inicial - European Education Area](https://ec.europa.eu/education/european-education-area).
- European Commission. (2019). *Ethics guidelines for trustworthy AI*. [Ethics guidelines for trustworthy AI | Shaping Europe's digital future](https://ec.europa.eu/artificial-intelligence/ethics-guidelines-for-trustworthy-ai).
- European Parliament and of the Council. (2022). *Regulation (EU) 2022/868 of the European Parliament and of the Council of 30 May 2022 on European data governance and amending Regulation (EU) 2018/1724 (Data Governance Act)*. <https://eur-lex.europa.eu/legal-content/EN/TXT/PDF/?uri=CELEX:32022R0868>.
- European Parliament and the Council of the European Union. (2024). *Directive (EU) 2024/2853 of the European Parliament and of the Council of 23 October 2024 on liability for defective products and repealing Council Directive 85/374/EEC*. [https://eur-lex.europa.eu/legal-content/EN/TXT/PDF/?uri=OJ:L\\_202402853](https://eur-lex.europa.eu/legal-content/EN/TXT/PDF/?uri=OJ:L_202402853).

- European Parliament and the Council of the European Union. (2019). *Directive (EU) 2019/882 of the European Parliament and of the Council of 17 April 2019 on the accessibility requirements for products and services*. <https://eur-lex.europa.eu/legal-content/EN/TXT/HTML/?uri=CELEX:32019L0882>.
- European Parliament and the Council of the European Union. (2019). *Directive (EU) 2019/790 of the European Parliament and of the Council of 17 April 2019 on copyright and related rights in the digital single market and amending directives 96/9/EC and 2001/29/EC*. <https://eur-lex.europa.eu/legal-content/EN/TXT/PDF/?uri=CELEX:32019L0790>.
- European Parliament and the Council of the European Union. (2016). *Directive (EU) 2016/2102 of the European Parliament and of the Council of 26 October 2016 on the accessibility of the websites and mobile applications of public sector bodies*. <https://eur-lex.europa.eu/legal-content/EN/TXT/PDF/?uri=CELEX:32016L2102>.
- European Parliament and the Council of the European Union. (2001). *Directive 2001/29/EC of the European Parliament and of the Council of 22 May 2001 on the harmonisation of certain aspects of copyright and related rights in the information society*. <https://eur-lex.europa.eu/legal-content/EN/TXT/PDF/?uri=CELEX:32001L0029>.
- European Parliament and the Council of the European Union. (2024). *Regulation (EU) 2024/1689 of the European Parliament and of the Council of 13 June 2024 laying down harmonised rules on artificial intelligence and amending regulations (EC) No 300/2008, (EU) No 167/2013, (EU) No 168/2013, (EU) 2018/858, (EU) 2018/1139 and (EU) 2019/2144 and directives 2014/90/EU, (EU) 2016/797 and (EU) 2020/1828 (Artificial Intelligence Act)*. [https://eur-lex.europa.eu/legal-content/EN/TXT/HTML/?uri=OJ:L\\_202401689](https://eur-lex.europa.eu/legal-content/EN/TXT/HTML/?uri=OJ:L_202401689).
- European Parliament and the Council of the European Union. (2016). *Regulation (EU) 2016/679 of the European Parliament and of the Council of 27 April 2016 on the protection of natural persons with regard to the processing of personal data and on the free movement of such data, and repealing Directive 95/46/EC (General Data Protection Regulation)*. <https://eur-lex.europa.eu/legal-content/EN/TXT/PDF/?uri=CELEX:32016R0679>.
- Farelli, M. (2024). *Inteligencia artificial y educación: insumos para su abordaje desde Iberoamérica*. <https://oei.int/oficinas/argentina/publicaciones/inteligencia-artificial-y-educacion-insumos-para-su-abordaje-desde-iberoamerica/>.
- Floridi, L.; Cowls, J.; Beltramini, M.; Chatila, R.; Chazerand, P.; Dignum, V.; Luetge, C.; Madelin, R.; Pagallo, U.; Rossi, F.; Schafer, B.; Valcke, P.; & Vayena, E. (2018). *AI4People-An Ethical Framework for a Good AI Society: Opportunities, Risks, Principles, and Recommendations*. PubMed. <https://pubmed.ncbi.nlm.nih.gov/30930541/>.
- Hao-Ping, L.; Drosos, I.; Sarkar, A.; Rintel, S.; Wilson, N.; Tankelevitch, L.; & Banks, R. (2025). *The Impact of Generative AI on Critical Thinking: Self-Reported Reductions in Cognitive Effort and Confidence Effects From a Survey of Knowledge Workers*. Microsoft Research. [https://www.microsoft.com/en-us/research/wp-content/uploads/2025/01/lee\\_2025\\_ai\\_critical\\_thinking\\_survey.pdf](https://www.microsoft.com/en-us/research/wp-content/uploads/2025/01/lee_2025_ai_critical_thinking_survey.pdf).
- Jobin, A.; Ienca, M.; & Vayena, E. (2019). *The global landscape of AI ethics guidelines*. <https://www.nature.com/articles/s42256-019-0088-2>.
- Kastania, N. P. P. (2024). Building Trust in AI Education: Addressing Transparency and Ensuring Trustworthiness. In *Trust and Inclusion in AI-Mediated Education: Where Human Learning Meets Learning Machines* (pp. 73-90). [https://link.springer.com/chapter/10.1007/978-3-031-64487-0\\_4](https://link.springer.com/chapter/10.1007/978-3-031-64487-0_4).
- Köbis, N.; Rahwan, Z.; Rilla, R.; Ibrahim Supriyatno, B.; Bersch, C.; Tamer, A.; Bonnefon, J.-F.; & Rahwan, I. (2025). *Delegation to artificial intelligence can increase dishonest behaviour*. Nature. <https://www.nature.com/articles/s41586-025-09505-x>.
- Lara, T. & Magro, C. (2025). *IA y Educación. Una relación con costuras*. Biblioteca digital journey. <https://digitaljourney.es/una-invita-cion-a-pensar-las-costuras-de-la-educacion/>.
- Memarian, B.; & Doleck, T. (2023). *Fairness, Accountability, Transparency and Ethics (FATE) in Artificial Intelligence (AI) and Higher Education: A systematic review*. <https://www.sciencedirect.com/science/article/pii/S2666920X23000310>.
- Mittelstand, B. (2019). *Principles alone cannot guarantee ethical AI*. <https://www.nature.com/articles/s42256-019-0114-4>.
- Mondal, H. (2025). *Ethical engagement with artificial intelligence in medical education*. PubMed. <https://pubmed.ncbi.nlm.nih.gov/39749996/>.
- Morley, J.; Floridi, L.; Kinsey, L.; & Elhalal, A. (2019). *From What to How: An Initial Review of Publicly Available AI Ethics Tools, Methods and Research to Translate Principles into Practices*. <https://pmc.ncbi.nlm.nih.gov/articles/PMC7417387/>.
- Naciones Unidas. (2023). *Educació: Objectiu de Desenvolupament Sostenible 4. Educació - Desenvolupament Sostenible*.
- Prather, J.; Reeves, B.; Leinonen, J.; Macneil, S.; Randrianasolo, A.S.; Becker, B.; Kimmel, B.; Wright, J.; Briggs, B. (2024). *The Widening Gap: The Benefits and Harms of Generative AI for Novice Programmers*. arXiv. <https://export.arxiv.org/pdf/2405.17739v1>.
- Riley, B.; & Bruno, P. (2024). *Education hazards of generative AI. Cognitive Resonances*. <https://cognitiveresonance.net/EducationHazardsofGenerativeAI.pdfv>.
- Sabour, S.; Liu, J.M.; Liu, S.; Yao, C. Z.; Cui, S.; Zhang, X.; Zhang, W.; Cao, Y.; Bhat, A.; Guan, J.; Wu, W.; Mihalcea, R.; Wang, H.; Althoff, T.; Lee, T. M. C.; & Huang, M. (2025). *Human Decision-making is Susceptible to AI-driven Manipulation*. arXiv. <https://arxiv.org/abs/2502.07663>.
- Stadler, M. & Bannert, M. & Sailer, M. (2024). *Cognitive ease at a cost: LLMs reduce mental effort but compromise Depth in student scientific inquiry*. Science Direct. <https://www.sciencedirect.com/science/article/pii/S0747563224002541>.
- Steels, L.; & López de Mantras, R. (2018). *The Barcelona Declarations for the Proper Development and Usage of Artificial Intelligence in Europe*, *AI Communications*, 31 (6), 485 – 494.
- UNESCO. (2021). *Reimagining our futures together: A new social contract for education*. <https://unesdoc.unesco.org/ark:/48223/pf0000381137>.
- UNI Global Union. (2017). *Top 10 Principales for Ethical Artificial Intelligence*. [https://uniglobalunion.org/wp-content/uploads/uni\\_ethical\\_ai.pdf](https://uniglobalunion.org/wp-content/uploads/uni_ethical_ai.pdf).

# La interseccionalidad como herramienta analítica para la auditoría de sesgos en inteligencia artificial y modelos de lenguaje a gran escala (LLM)

## Intersectionality as an analytical tool for bias auditing in artificial intelligence and large language models (LLM)

Jose María Regalado López  
Universidad Pontificia de Comillas

### PALABRAS CLAVE:

Interseccionalidad  
auditoría algorítmica  
sesgo algorítmico  
inteligencia artificial  
equidad

### RESUMEN:

El objetivo de este artículo es postular la necesidad urgente de adoptar un marco de análisis interseccional para identificar y comprender las formas complejas de sesgo que emergen en la inteligencia artificial (IA) y los modelos de lenguaje a gran escala (LLM). La metodología examina la arquitectura de la inequidad algorítmica a lo largo del ciclo de vida de la IA (datos, modelado e implementación), utilizando la teoría interseccional como herramienta analítica crítica y describiendo una metodología en tres fases para la auditoría: mapeo del contexto situado, desagregación de métricas de equidad y comunicación de la complejidad. Se discute que los análisis unidimensionales son insuficientes, pues la discriminación es compuesta y afecta de forma más grave a grupos en la intersección de identidades. Se concluye que la interseccionalidad es una herramienta esencial para una auditoría de sesgos efectiva que aspire a la justicia social, requiriendo un enfoque interdisciplinario para garantizar sistemas de IA equitativos y confiables.

### KEYWORDS:

Intersectionality  
algorithmic auditing  
algorithmic bias  
artificial intelligence  
equity

### ABSTRACT:

The objective of this article is to postulate the urgent need to adopt an intersectional analysis framework to identify and understand the complex forms of bias that emerge in Artificial Intelligence (AI) and Large Language Models (LLM). The methodology examines the architecture of algorithmic inequity throughout the AI lifecycle (data, modeling, and implementation), utilizing intersectional theory as a critical analytical tool and describing a three-phase methodology for the audit: mapping of the situated context, disaggregation of equity metrics, and communication of complexity. It is discussed that unidimensional analyses are insufficient, since discrimination is composite and affects groups at the intersection of identities more severely. It is concluded that intersectionality is an essential analytical tool for effective bias auditing that aspires to social justice, requiring an interdisciplinary approach to guarantee equitable and reliable AI systems.

## 1. INTRODUCCIÓN

El auge de los Modelos de Lenguaje a Gran Escala (LLM), como GPT-4 de OpenAI y LLaMA 2 de Meta, marca un punto de inflexión en la integración de la Inteligencia Artificial (IA) en la sociedad. Estos sistemas están siendo rápidamente incorporados en ámbitos sociales, económicos y culturales, desde la generación de contenido hasta la toma de decisiones críticas en áreas como la salud, el derecho y la economía. Sin embargo, a pesar de su impresionante potencial, estos sistemas no son neutrales.

Los LLM son entrenados con vastos *corpus* de texto extraídos de internet, lo que inevitablemente provoca que absorban, reproduzcan e incluso amplifiquen los sesgos y estereotipos presentes en la sociedad y en los datos históricos. Esta incorporación de prejuicios inherentes en los datos de entrenamiento constituye una de las principales fuentes de sesgo algorítmico (Navigli et al., 2023; Ovalle et al., 2023).

Las consecuencias negativas de esta realidad son palpables y tienen un impacto directo en la justicia y la equidad social. Los sesgos se manifiestan en forma de daños asignativos, donde se distribuyen recursos u oportunidades de manera injusta (Ovalle et

**CÓMO CITAR:** Regalado, J. M. (2025). La interseccionalidad como herramienta analítica para la auditoría de sesgos en inteligencia artificial y modelos de lenguaje a gran escala (LLM). *RETIS. Revista de Tecnología para la Inclusión Social*, 2(1), 31-41, DOI:10.70664/retis.v2i1.004

al., 2023). Un ejemplo notorio es el caso del software COMPAS, utilizado en el sistema judicial de EE.UU., donde se documentó que el algoritmo perpetuaba la discriminación racial al predecir incorrectamente una mayor tasa de reincidencia para acusados negros en comparación con los blancos. De igual manera, se ha evidenciado el sesgo de género en sistemas de selección de personal, como el caso de Amazon, que desechó un algoritmo de contratación al descubrir que éste favorecía consistentemente a los hombres basándose en patrones lingüísticos sesgados (Costanza-Chock et al., 2022). Estos modelos también generan daños representacionales, perpetuando estereotipos, por ejemplo, al asociar roles estereotípicos de género como "programador" con el hombre y "ama de casa" con la mujer (Kotek et al., 2023).

Ante esta complejidad, los análisis de sesgo unidimensionales, que se centran en un único eje de identidad (como el género o la raza), han demostrado ser insuficientes para capturar la naturaleza compuesta y entrelazada de la discriminación algorítmica (Ovalle et al., 2023; Simó, 2024). La desigualdad social está determinada por múltiples ejes (raza, género, clase) que actúan de manera conjunta y se influyen entre sí (Hill Collins y Bilge, 2019).

Por lo tanto, este artículo postula la necesidad urgente de adoptar un marco de análisis interseccional para identificar y comprender las formas complejas y compuestas de sesgo que emergen en la intersección de múltiples identidades. El enfoque interseccional, acuñado originalmente por Kimberlé Williams Crenshaw (1990), ofrece la herramienta analítica crítica para desvelar la Matriz de Dominación que subyace a la exclusión (Crenshaw, 1989; Hill Collins, 2017; Ovalle et al., 2023).

Describimos unas pautas metodológicas para integrar sistemáticamente la perspectiva interseccional en las prácticas de auditoría algorítmica, transformando el proceso de una simple verificación técnica a un análisis sociotécnico robusto (Raji et al., 2020). Al adoptar la auditoría de sesgo en conjunto con el pensamiento complejo, se garantiza que los sistemas de IA no solo sean técnicamente robustos, sino también deseables y éticamente alineados con los principios de justicia y equidad social (Bustelo, 2024).

Este artículo se estructura de la siguiente manera: primero, se examina cómo los sesgos sociales, como el sesgo de género y el sesgo racial, se introducen y se propagan a lo largo de todo el ciclo de vida de los LLM. A continuación, se profundizará en la teoría interseccional como una herramienta analítica crítica para superar las deficiencias de los enfoques de eje único. Posteriormente, se propondrá un modelo práctico para integrar el análisis interseccional en la metodología de las auditorías algorítmicas, un mecanismo de gobernanza clave para la rendición de cuentas (Éticas Research and Consulting, 2021). Finalmente, se discutirán los desafíos culturales, lingüísticos y metodológicos que esta integración plantea.

## 2. LA ARQUITECTURA DE LA INEQUIDAD ALGORÍTMICA. UNA TAXONOMÍA DE SESGOS A LO LARGO DEL CICLO DE VIDA DE LA INTELIGENCIA ARTIFICIAL

La creciente integración de la Inteligencia Artificial en sistemas sociales y económicos críticos ha puesto de manifiesto la inevitable transferencia y amplificación de los prejuicios humanos y las desigualdades sociales históricas dentro de los sistemas algorítmicos. Eliminar este fenómeno, conocido como sesgo de la IA o sesgo de *machine learning*, es un reto no solo técnico, sino profundamente social, ya que requiere un entendimiento profundo de las fuerzas sociales existentes y de las técnicas de la ciencia de datos. Los sesgos pueden encontrarse en cada etapa del ciclo de vida de la IA.

El análisis de estos sesgos en los LLM debe ser una tarea estratégica que abarque todas las fases de su ciclo de vida, desde la concepción y la recopilación de datos hasta su implementación y evaluación. Los sesgos no son un problema aislado que se corrige en una única etapa, sino un fenómeno sistémico que puede ser introducido, reforzado o amplificado en múltiples puntos del proceso de desarrollo.

A continuación, se presenta una taxonomía del sesgo algorítmico estructurada en tres fases principales a lo largo del desarrollo y la implementación del sistema.

### 2.1. FASE 1: SESGOS PRE-ALGORÍTMICOS EN LA PREPARACIÓN Y DATOS

El sesgo pre-algorítmico surge antes de la construcción del modelo, principalmente en la selección, recopilación y etiquetado de los datos utilizados para entrenar los modelos de *machine learning*<sup>1</sup>. Puesto que los sistemas de IA aprenden a tomar decisiones basándose en estos datos, es crucial que los conjuntos de datos sean evaluados para detectar la presencia de sesgos. Si la calidad del dato es deficiente, la calidad del resultado del algoritmo se verá comprometida. (Navigli et al., 2023).

#### 2.1.1. Fuentes fundamentales del sesgo de datos

El problema central radica en que los datos de entrenamiento a menudo reflejan y perpetúan los prejuicios sociales preexistentes.

##### **Representación desigual o sesgo de muestreo**

Este es un factor crítico. Se produce cuando ciertos grupos están sobrerrepresentados o infrarrepresentados en el conjunto de datos de entrenamiento con respecto a la población real a la que se aplicará el sistema. (Chu et al., 2024).

Un ejemplo notable es el de los algoritmos de reconocimiento facial, donde la sobrerrepresentación de personas de raza blanca en los datos de entrenamiento puede llevar a errores significativos en el reconocimiento de personas de color. De manera similar, se observaron disparidades en conjuntos de datos como IJB-A y Adience<sup>2</sup>, donde los individuos de piel clara constituían una vasta mayoría, sesgando el análisis contra grupos subrepresentados de piel oscura.

1 El Aprendizaje de Máquinas (Machine Learning) es un subcampo de la Inteligencia Artificial (IA) que utiliza algoritmos y técnicas de aprendizaje automático para que los sistemas puedan aprender a tomar decisiones o detectar patrones y reglas latentes directamente a partir de la ingestión de datos de entrenamiento.

2 Los conjuntos de datos IJB-A y Adience son bancos de pruebas (benchmarks) de imágenes faciales no restringidas utilizados en visión por computadora para tareas de clasificación y reconocimiento, y son notables por exhibir un sesgo demográfico significativo con una sobrerrepresentación de hombres de piel clara y una infrarrepresentación de mujeres de piel oscura

### Sesgo de etiquetado humano

Los prejuicios inherentes en las percepciones o juicios de los etiquetadores (o anotadores) pueden manifestarse en los datos de entrenamiento, perpetuando estereotipos y decisiones discriminatorias (Crawford, como se citó en Bustelo, 2024). El uso de etiquetado incoherente o la inclusión/exclusión desproporcionada de ciertas características en los datos pueden llevar a la exclusión de solicitantes de empleo cualificados, por ejemplo (IBM Data and AI Team, 2023).

### Sesgo histórico y social

La realidad capturada en los datos históricos a menudo está marcada por la injusticia y la desigualdad (por ejemplo, si los datos muestran que "todos los programadores son hombres y todas las enfermeras son mujeres," el modelo interiorizará y perpetuará estos sesgos ocupacionales y de género) (Chu et al., 2024). El vocabulario y los indicadores socioeconómicos (como los ingresos) utilizados por el algoritmo pueden discriminar involuntariamente por raza o sexo.

#### 2.1.2. Tipos de sesgos sociales incorporados

Los modelos de lenguaje grande (LLM) y otros sistemas de IA demuestran incorporar una amplia gama de sesgos sociales a partir de sus datos de entrenamiento, afectando especialmente a las minorías y los grupos marginados (Navigli et al., 2023), los más comunes son:

- **Sesgo de género:** Es uno de los más estudiados. Los sistemas asocian roles estereotípicos (como "programador" para hombre y "ama de casa" para mujer) (Kotek et al., 2023). Este sesgo se ha cuantificado en incrustaciones de palabras, revelando estereotipos étnicos y de género. Estudios sobre ChatGPT han encontrado sesgos de género en documentos generados, como cartas de referencia, donde los textos para sujetos masculinos muestran una formalidad, positividad y capacidad de acción significativamente mayores que los de sujetos femeninos (Wan et al., 2023).
- **Sesgo de raza y etnia:** Se evidencia en algoritmos que favorecen a ciertos grupos raciales. (Kotek et al., 2024) Los sistemas de diagnóstico asistido han mostrado menor precisión para pacientes negros que blancos. En modelos de lenguaje, la raza muestra el mayor sesgo de asociación, seguida por género, salud y religión (Bai et al. 2025).
- **Sesgo cultural y de valores:** Los LLM, a menudo exhiben un sesgo cultural con "acento americano". Los modelos reflejan valores de la sociedad donde se entrenaron, generando conflictos al implementarse en otros contextos (Jiao et al., 2023).
- **Sesgo de discapacidad y socioeconómico:** Los modelos pueden reflejar prejuicios hacia personas con discapacidad o estatus socioeconómico bajo, impactando procesos cruciales como decisiones judiciales (Navigli et al., 2023).
- **Sesgo interseccional:** Se refiere a las formas de discriminación que ocurren en la intersección de múltiples ejes de identidad y opresión, basados en el concepto de interseccionalidad Crenshaw (1990) y las propuestas de Hill Collins (2017), en las que profundizaremos más adelante. Como por ejemplo el hecho de ser mujer y negra. Este sesgo es crucial porque el rendimiento de los algoritmos empeora significativamente para grupos que están en los márgenes de varias categorías. Investigaciones demuestran que las incrustaciones de palabras contextualizadas contienen una distribución de sesgos humanos interseccionales (Simó, 2024). Es esencial que la auditoría de los sistemas de IA adopte esta perspectiva, ya que una evaluación no interseccional resulta incompleta.

## 2.2. FASE 2: SESGOS EN EL DESARROLLO Y MODELADO DEL ALGORITMO

Esta fase se centra en cómo el propio diseño y proceso de entrenamiento del algoritmo introduce o amplifica los sesgos, incluso cuando los datos se consideran aceptables. Este es un punto donde las decisiones de diseño pueden tener consecuencias éticas y sociales directas.

Los sesgos en esta fase resultan de las elecciones técnicas hechas por los desarrolladores.

### 2.2.1. Priorización de la Precisión sobre la Equidad

Los algoritmos suelen estar optimizados para maximizar la precisión general. Sin embargo, esta búsqueda de precisión puede llevar a los modelos a "capitalizar correlaciones casuales o anomalías estadísticas" presentes en el conjunto de datos. (Chu, et al., 2024). Un modelo puede producir resultados precisos basándose en justificaciones incorrectas (como usar el género como una característica discriminatoria si todos los ejemplos positivos de un *dataset*<sup>3</sup> provienen de hombres), lo que resulta en discriminación. (Chu, et al., 2024).

### 2.2.2. Errores de Programación y Ponderación

El sesgo puede surgir de errores técnicos o de que el desarrollador pondere injustamente ciertos factores en la toma de decisiones del algoritmo, reflejando sus propios sesgos (conscientes o inconscientes).

### 2.2.3. Amplificación del Sesgo

El proceso de modelado, especialmente en modelos grandes de lenguaje (LLM), puede amplificar los sesgos inherentes presentes en los datos de entrenamiento (IBM Data and AI Team, 2023). El concepto de recursividad, esencial en el pensamiento complejo, explica cómo los modelos entrenados con datos ya influenciados por el mismo algoritmo pueden generar un ciclo de retroalimentación que exacerba los sesgos. (Bustelo, 2024).

### 2.2.4. Sesgo de Selección o Asunciones

3 Un conjunto de datos (*dataset*) es una colección masiva de textos o información estructurada o no estructurada que se utiliza para entrenar y evaluar sistemas de Inteligencia Artificial y modelos de machine learning.

Este sesgo puede provenir de las asunciones hechas por el creador del modelo. En el procesamiento de lenguaje natural<sup>4</sup>, los modelos entrenados con grandes corpus de texto aprenden las semánticas y, con ellas, los prejuicios de la sociedad que generó ese lenguaje (Caliskan, como se citó en Kotek, 2023).

### 2.3. FASE 3: SESGOS EN EL USO, IMPLEMENTACIÓN Y RESULTADOS (POST-ALGORÍTMICOS)

En la fase de implementación, los sesgos se reflejan en los resultados y las consecuencias prácticas del uso del sistema en un entorno real. Estos sesgos se clasifican según el tipo de daño que causan. Los tipos de daños y sus manifestaciones son los siguientes:

#### 2.3.1. Daños Asignativos

Se presentan cuando un sistema automatizado distribuye recursos u oportunidades de manera injusta entre diferentes grupos sociales, como vemos en los siguientes ejemplos.

- **Justicia penal:** Un ejemplo es el algoritmo COMPAS, utilizado para predecir la reincidencia, que mostró un sesgo racial al discriminar a las personas afroamericanas (simó, 2024).
- **Contratación:** Amazon eliminó un algoritmo de selección de personal al descubrir que favorecía a los hombres basándose en el lenguaje de sus currículos (Costanza-Chock, et al., 2022).
- **Finanzas y servicios:** Los algoritmos utilizados para la aprobación de préstamos pueden introducir sesgos invisibles que afectan desproporcionadamente a ciertos grupos.
- **Salud:** Los LLM empleados para gestionar la salud de poblaciones han demostrado sesgo racial, afectando la asignación de recursos (Tao, et al., 2024).

#### 2.3.2. Daños Representacionales

Aparecen cuando un sistema representa a grupos sociales de manera desfavorable, los denigra, perpetúa estereotipos o directamente no reconoce su existencia (Blodgett et al., 2020).

- **Generación de Imágenes:** Se ha encontrado que aplicaciones de IA generativa (como Midjourney) perpetúan el sesgo de género. Al generar imágenes de profesionales especializados, los mayores de edad eran consistentemente representados como hombres, reforzando estereotipos sobre el rol laboral de la mujer (IBM Data and AI Team, 2023).
- **Estereotipos Lingüísticos:** Los modelos de lenguaje pueden crear narrativas que reflejan sesgos persistentes, como el sesgo anti-musulmán (Bai, et al., 2025).
- **Sesgo Estadístico:** Se produce cuando la discriminación grupal se basa en un hecho estadísticamente relevante en el mundo real, pero cuyo uso por el algoritmo resulta en un tratamiento desventajoso hacia un grupo vulnerable (Éticas Research and Consulting, 2021).

#### 2.3.3. Sesgo Cognitivo en la Implementación

El sesgo cognitivo de los humanos que diseñan y utilizan la IA también influye en la implementación. Los desarrolladores pueden preferir conjuntos de datos recopilados de estadounidenses en lugar de tomar muestras de poblaciones globales diversas (IBM Data and AI Team, 2023). Incluso en la interacción post-algorítmica, los equipos humanos pueden incorporar estos sesgos a través de la selección o ponderación de datos.

El análisis del contexto social, económico y cultural en el que opera el algoritmo es esencial para interpretar los resultados y asegurar la deseabilidad del sistema, garantizando que no incurra en formas de discriminación o impacte negativamente en grupos vulnerables (Éticas Research and Consulting, 2021).

### 2.4. LA NECESIDAD DE UNA AUDITORÍA INTEGRAL

Para enfrentar y mitigar los sesgos en la inteligencia artificial, es necesario adoptar un enfoque que vaya más allá de lo puramente técnico y se alinee con el pensamiento complejo y la metodología de auditoría de sesgos. La auditoría de sesgos se considera una práctica esencial para verificar sistemáticamente el cumplimiento de criterios objetivos y evaluar la confiabilidad de los algoritmos. Esto implica el uso de métricas de equidad grupal para determinar si el algoritmo opera de manera diferente entre grupos protegidos, así como el análisis de sensibilidad para observar cómo las variaciones en los datos afectan los resultados.

Es crucial que esta auditoría sea un proceso continuo y no un evento único, incorporando un monitoreo constante para detectar sesgos emergentes y asegurar que el modelo siga siendo justo y efectivo con el tiempo. Para lograr un desarrollo ético y equitativo de la inteligencia artificial, el enfoque debe ir más allá de los sesgos tradicionales (como raza y género) e incluir la interseccionalidad.

Es fundamental que los desarrolladores y los profesionales del ámbito social colaboren para que los sistemas de inteligencia artificial no solo sean técnicamente sólidos, sino también socialmente responsables y alineados con los principios de justicia y equidad. Aunque hemos identificado que el sesgo se origina y se amplifica en estas etapas, su detección y corrección requieren un marco analítico más sofisticado que el simple análisis técnico, como hemos mencionado anteriormente. A continuación, se presenta la perspectiva interseccional como la herramienta crítica necesaria para este propósito.

4 El Procesamiento del Lenguaje Natural (NLP) (Natural Language Processing) es un área de la Inteligencia Artificial que se enfoca en el desarrollo de modelos computacionales con la capacidad de comprender y generar lenguaje humano. El campo, que ha experimentado un cambio de paradigma con el advenimiento de los Modelos de Lenguaje a Gran Escala (LLM), aborda una amplia variedad de tareas, incluyendo la traducción, la generación de código y el análisis de sentimientos.

### 3. LA PERSPECTIVA INTERSECCIONAL Y LA MATRIZ DE DOMINACIÓN COMO HERRAMIENTAS DE ANÁLISIS CRÍTICO

Para superar las limitaciones de los análisis de sesgo unidimensionales, que examinan categorías como la raza o el género de manera aislada, es esencial adoptar la interseccionalidad como una herramienta tanto teórica como metodológica. Este enfoque es vital para hacer visibles las experiencias de opresión que son complejas y cualitativamente distintas, las cuales los métodos tradicionales suelen pasar por alto, permitiendo así un diagnóstico más preciso y profundo de los sesgos en la IA.

#### 3.1. FUNDAMENTOS DE LA INTERSECCIONALIDAD

El término interseccionalidad fue introducido por la académica y jurista Kimberlé Crenshaw en 1990 para explicar cómo diversos sistemas de opresión, como el racismo, el sexismo y el clasismo, se entrelazan y generan experiencias de discriminación únicas y complejas.

Crenshaw emplea la metáfora de una encrucijada para mostrar que la discriminación puede surgir desde múltiples direcciones. En este contexto, la interseccionalidad no es simplemente la suma de opresiones (por ejemplo, racismo + sexismo), sino una interacción sinérgica que origina una nueva forma de subordinación. Esto implica que el efecto combinado de varios sesgos no es meramente aditivo (por ejemplo, sesgo (raza) + sesgo (género)), sino multiplicativo o transformador, creando una forma de discriminación cualitativamente distinta y, a menudo, más grave.

Un ejemplo concreto que ilustra esta dinámica, tomado de los estudios de Crenshaw, es el caso de las mujeres inmigrantes que sufren violencia doméstica. Estas mujeres enfrentan una doble subordinación: por un lado, la violencia de género por parte de sus parejas, y por otro, el temor a la deportación debido a su estatus migratorio. Esta situación genera una vulnerabilidad única que no puede ser comprendida si se analiza únicamente desde la perspectiva del género o de la clase social por separado.

#### 3.2. LA INSUFICIENCIA DEL ANÁLISIS DE EJE ÚNICO EN LA INTELIGENCIA ARTIFICIAL

Los enfoques que auditan el sesgo de forma aislada son incapaces de detectar las formas más graves de discriminación algorítmica. El trabajo pionero de Buolamwini y Gebu (2018) es un ejemplo paradigmático. En su estudio sobre clasificadores faciales comerciales, demostraron que estos sistemas obtenían las tasas de error más altas con mujeres de piel oscura. Este fallo garrafal, que afectaba a un grupo en la intersección de raza y género, solo pudo ser detectado gracias a una auditoría interseccional que desagregó los resultados por ambas categorías simultáneamente (Buolamwini y Gebu, 2018).

Investigaciones técnicas posteriores han corroborado estos hallazgos. Por ejemplo, Tan y Celis (2019) demostraron que las identidades interseccionales (cómo ser mujer y afroamericana) sufren sesgos significativamente mayores en las representaciones de palabras contextualizadas generadas por LLM, en comparación con los sesgos asociados a sus identidades constituyentes por separado.

La creciente ubicuidad de la inteligencia artificial en diversos ámbitos de la sociedad, enfatiza la imperiosa necesidad de abordar su desarrollo desde una perspectiva interseccional. Este enfoque implica considerar la interrelación de las diversas identidades y experiencias sociales en la interacción con la IA.

Garantizar un desarrollo de la IA con mayor integridad requiere el reconocimiento y abordaje de las disparidades existentes en términos de género, raza, clase social, discapacidad y otras dimensiones de diversidad. Esto no solo contribuiría a mitigar los sesgos inherentes en los sistemas de IA, sino que también fomentaría soluciones más equitativas y representativas. La adopción de esta visión interseccional en el desarrollo de la IA es esencial para la creación de tecnologías verdaderamente inclusivas y beneficiosas para el conjunto de la sociedad, evitando así la perpetuación o exacerbación de las desigualdades existentes.

#### 3.3. APLICANDO LA INTERSECCIONALIDAD ESTRUCTURAL, POLÍTICA Y REPRESENTACIONAL A LA IA

La aplicación del enfoque interseccional en el desarrollo de la IA requiere analizar cómo las dimensiones de identidad y experiencia social se entrelazan y afectan la interacción con sistemas de IA. La adaptación del marco interseccional al contexto de la IA, implica examinar cómo las estructuras, políticas y representaciones pueden perpetuar o desafiar desigualdades existentes. Esto incluye analizar las prácticas de contratación en la industria, las políticas de desarrollo y despliegue de sistemas, y la representación de grupos en los conjuntos de datos para entrenar modelos de lenguaje.

Para aplicar este marco a los modelos de lenguaje de gran escala (LLM), adaptamos la triple clasificación de Crenshaw (1990) al contexto de la IA.

- **Interseccionalidad estructural:** Se refiere a cómo la estructura de los sistemas de IA, desde bases de datos hasta arquitectura, puede excluir a grupos en las intersecciones. Si los datos no reflejan la realidad de mujeres de color con bajos ingresos, el LLM será incapaz de generar respuestas pertinentes a sus necesidades, perpetuando su invisibilidad estructural.
- **Interseccionalidad política:** Expone cómo las agendas de mitigación de sesgos, al operar sobre un único eje, marginan a quienes están en las intersecciones. La política antirracista tiende a centrarse en hombres negros, mientras la feminista en mujeres blancas, volviendo invisibles a las mujeres negras.
- **Interseccionalidad representacional:** Analiza cómo los LLM generan estereotipos culturales dañinos que afectan a grupos interseccionales. Un modelo que asocia mujeres negras con agresividad o latinas con hipersexualización perpetúa una violencia representacional con consecuencias tangibles. Comprendido el poder analítico de la interseccionalidad, el siguiente paso es integrarla en un marco práctico de gobernanza como la auditoría algorítmica.

### 4. INTEGRACIÓN DE LA INTERSECCIONALIDAD EN LA AUDITORÍA ALGORÍTMICA

La auditoría algorítmica es un proceso sistemático de evaluación y análisis de los algoritmos utilizados en sistemas de inteligencia artificial y toma de decisiones automatizada. Su objetivo principal es examinar la transparencia, equidad y responsabilidad de

estos algoritmos, identificando posibles sesgos, errores o impactos negativos en diferentes grupos sociales. Este proceso implica la revisión del código fuente, los datos de entrenamiento y los resultados producidos por el algoritmo, así como la evaluación de su conformidad con estándares éticos y regulaciones legales. La auditoría algorítmica es fundamental para garantizar que los sistemas basados en algoritmos sean confiables, justos y no discriminatorios, especialmente en áreas sensibles como la justicia penal, la contratación laboral o la asignación de recursos públicos.

La auditoría algorítmica se posiciona como un mecanismo de gobernanza crucial para la rendición de cuentas de los sistemas de IA. Sin embargo, para que estas auditorías sean verdaderamente efectivas, deben trascender las métricas técnicas estándar y adoptar un análisis sociotécnico robusto. En este contexto, la interseccionalidad no es solo un concepto teórico deseable, sino una categoría de análisis indispensable para evaluar la equidad real de un sistema.

#### 4.1. PRINCIPIOS DE UNA AUDITORÍA ALGORÍTMICA

De acuerdo con la Guía de Auditoría Algorítmica de Ética (2021), una auditoría exhaustiva debe estar orientada por un conjunto de principios fundamentales que aseguren su integridad y su impacto social positivo. Estos principios incluyen:

- Cumplimiento legal y ético: Garantizar que el sistema algorítmico se adhiera a la normativa vigente y a los estándares éticos aplicables.
- Deseabilidad: Verificar que el algoritmo no incurra en formas de discriminación, especialmente contra individuos o grupos vulnerables.
- Aceptabilidad: Asegurar que el funcionamiento del sistema sea transparente y comprensible para la ciudadanía, permitiendo una supervisión y un control social efectivos.

#### 4.2. UN MARCO INTERSECCIONAL PARA LA AUDITORÍA. PUNTOS DE INTEGRACIÓN

El auge de los sistemas de Inteligencia Artificial para el tratamiento de información ha coincidido con la creciente relevancia empírica del enfoque interseccional en el ámbito social, incitando la búsqueda de un punto de confluencia entre ambos métodos de análisis. La interseccionalidad, ofrece una herramienta analítica esencial para comprender y examinar la complejidad del mundo y de las experiencias humanas, advirtiendo que la desigualdad social está determinada por múltiples ejes (como raza, género o clase) que actúan de manera conjunta y se influyen entre sí (Hill Collins y Bilge, 2019).

Dada la injerencia de los modelos de aprendizaje automático en esferas críticas de la vida social y profesional, resulta pertinente interrogarse acerca de la posible relación simbiótica entre interseccionalidad e IA (Simó, 2024). El enfoque interseccional es crucial para operar la justicia de manera efectiva en la IA (Ovalle et al., 2023), y su ausencia en el proceso auditor resulta en una evaluación incompleta (Costanza-Chock, 2022; Simó, 2024).

La literatura especializada, en consonancia con las directivas de buenas prácticas y los estudios de equidad interseccional, propone que esta perspectiva debe ser tejida en el ciclo completo de la IA (Wang et al., 2023; Ovalle et al., 2023). Específicamente, los estudios sugieren tres momentos clave en el ciclo de vida de un sistema de IA donde la interseccionalidad debe ser evaluada de forma explícita (Simó, 2024), que se describen a continuación:

##### 4.2.1. En la configuración de las bases de datos de entrenamiento

La auditoría debe verificar si los datos de entrenamiento representan de manera adecuada la diversidad de la población, prestando especial atención a los subgrupos interseccionales (Simó, 2024). Esto se alinea con la práctica cuantitativa esencial de los auditores, quienes informan que evalúan la representatividad de los datos y el sesgo en los datos de entrada (Costanza-Chock et al., 2022).

La meta es conseguir una representación fiel del entorno mediante la obtención de datos desagregados (Simó, 2024). Es crucial evitar que los datos de grupos marginalizados sean descartados o malinterpretados como "ruido" estadístico, ya que la heterogeneidad de los datos puede traducirse en una pérdida de precisión en el entrenamiento de los sistemas, disminuyendo el rendimiento para estas categorías (Buolamwini y Gebru, 2018). Esta etapa busca evitar la sobrerrepresentación de categorías hegemónicas y abordar el problema de desequilibrio de clases, frecuente al intentar configurar una base de datos interseccional (Simó, 2024).

##### 4.2.2. En el descubrimiento de correlaciones

Durante la fase de desarrollo y entrenamiento del modelo, se debe investigar activamente si el algoritmo está aprendiendo correlaciones espurias o patrones emergentes que perjudican a subgrupos interseccionales específicos (Simó, 2024).

El potencial de los modelos de *machine learning* para el tratamiento de datos multidimensionales y la estimación de correlaciones no aparentes puede ser aprovechado por la interseccionalidad. Sin embargo, esta fase es donde se manifiesta que los modelos de lenguaje contextualizados contienen una distribución de sesgos humanos interseccionales (Tan y Celis, 2019). Por ello, se han desarrollado pruebas de asociación de *embeddings*<sup>5</sup> dirigidas específicamente a identidades interseccionales (Tan y Celis, 2019), siendo necesario examinar si el algoritmo está "capitalizando correlaciones casuales o anomalías estadísticas" presentes en los datos (Bustelo, 2024).

5 El concepto de embedding (o incrustación de palabras) es una técnica de aprendizaje automático utilizada en el procesamiento de lenguaje natural (NLP) donde las palabras o frases se representan vectorialmente con números reales, lo que permite capturar los matices del lenguaje y sirve como un componente fundamental en los Modelos de Lenguaje a Gran Escala (LLM)

### 4.2.3. En la fase de auditoría como categoría de fiabilidad

La interseccionalidad debe ser un criterio fundamental para certificar la fiabilidad y equidad de un sistema (Simó, 2024). Las auditorías algorítmicas, que consisten en la verificación sistemática del cumplimiento de criterios objetivos, buscan certificar la confiabilidad de los algoritmos.

Un algoritmo no puede ser considerado fiable si demuestra un rendimiento deficiente o sesgado para cualquier subgrupo interseccional relevante (Simó, 2024). Esto se demuestra empíricamente: al examinar el rendimiento de los algoritmos asociados a diferentes atributos en la intersección de raza y género, se ha comprobado que los peores resultados se obtuvieron consistentemente con mujeres negras, remarcando la importancia de realizar auditorías interseccionales para abordar las disparidades en la precisión (Buolamwini y Gebru, 2018; Simó, 2024).

Por lo tanto, es crucial que la auditoría de sesgo se integre como una práctica continua (Raji et al., 2020), y no se limite a medir la equidad en función de subgrupos aislados, sino que incorpore la perspectiva interseccional como una sensibilidad analítica para garantizar que los sistemas de IA no refuercen las desigualdades existentes (Ovalle et al., 2023). Para conseguir un desarrollo ético y equitativo de la IA, el foco debe ir más allá de los ejes de sesgo tradicionales para incluir la interseccionalidad (Ovalle et al., 2023)

## 5. PROPUESTA METODOLÓGICA PARA UNA AUDITORÍA INTERSECCIONAL

Integrar la interseccionalidad como marco de análisis es crucial para la operatividad efectiva de la equidad en la Inteligencia Artificial (Ovalle et al., 2023; Simó, 2024). Este enfoque es vital para entender cómo la desigualdad se configura a través de múltiples ejes, como la raza, el género y la clase, que actúan conjuntamente e influyen mutuamente (Hill Collins y Bilge, 2019). La falta de esta perspectiva lleva a una evaluación incompleta de los sistemas algorítmicos (Costanza-Chock, 2022; Simó, 2024). Al aplicar una lente interseccional a las fases de la auditoría propuestas en la Guía de Auditoría Algorítmica (Éticas Research and Consulting, 2021), se obtiene una metodología más robusta y completa, diseñada para asegurar que la IA se desarrolle de manera ética y equitativa (Ovalle et al., 2023; Simó, 2024).

### 5.1. FASE DE ESTUDIO PRELIMINAR Y MAPEO. EL CONTEXTO SITUADO Y LA INTERSECCIONALIDAD ESTRUCTURAL

La etapa inicial de la auditoría, dedicada a la identificación de grupos vulnerables, debe trascender las categorías amplias y genéricas (como "inmigrantes" o "mujeres") (Éticas Research and Consulting, 2021). Es menester, en cambio, definir y analizar subgrupos interseccionales específicos y relevantes para el contexto de aplicación del algoritmo (Simó, 2024; Éticas Research and Consulting, 2021).

El concepto de interseccionalidad estructural (Crenshaw, 1990) es crucial en esta fase, pues subraya cómo la posición de las personas en la intersección de múltiples sistemas de opresión (como raza y género) hace que sus experiencias de subordinación sean cualitativamente distintas (Crenshaw, 1989; Hill Collins, 2017). Por ejemplo, en un sistema de asignación de ayudas sociales, un subgrupo relevante podría ser "mujeres inmigrantes de bajos ingresos que no hablan el idioma local". Esta definición situada reconoce que factores como el idioma, la clase y el género convergen para estructurar las vivencias de la pobreza de una manera concreta.

La auditoría debe adoptar un enfoque intra categórico, que se centra en los grupos sociales situados en la intersección de distintas categorías para comprender la complejidad con la que estas se interconectan en sus vidas (Martínez-Palacios, 2017). Para ello, es indispensable que los auditores se sitúen en el contexto sociohistórico para identificar los ejes de desigualdad que están actuando con mayor o menor preponderancia.

La literatura sobre auditoría reconoce esta necesidad, pues uno de los principales desafíos es la subrepresentación de grupos interseccionales o raros en los datos (Raji et al., 2020; Costanza-Chock et al., 2022). Los auditores, al definir los grupos a monitorear, deben reconocer que estos pueden ser definidos por la pertenencia a múltiples atributos protegidos (Éticas Research and Consulting, 2021) y deben ser identificados de manera dinámica durante el proceso, en correspondencia con la realidad en la que se inscribe el algoritmo (Éticas Research and Consulting, 2021).

### 5.2. FASE DE PLAN DE ANÁLISIS Y EJECUCIÓN. DESAGREGACIÓN DE MÉTRICAS DE EQUIDAD

Durante la fase de análisis cuantitativo, el objetivo es identificar el sesgo algorítmico (Éticas Research and Consulting, 2021). Es fundamental desagregar las métricas de rendimiento y equidad, tales como las tasas de falsos positivos (FPR) y falsos negativos (FNR), considerando sus intersecciones y no únicamente ejes individuales como género o raza (Éticas Research and Consulting, 2021).

Este enfoque de equidad interseccional permite identificar disparidades que podrían permanecer ocultas en un análisis unidimensional (Simó, 2024). El análisis debe comparar el desempeño del modelo entre subgrupos demográficos (Bustelo, 2024), revelando si el algoritmo se comporta de manera diferente entre grupos protegidos (Éticas Research and Consulting, 2021).

La investigación ha demostrado que los algoritmos presentan un rendimiento inferior para subgrupos en la intersección (Buolamwini y Gebru, 2018; Ovalle et al., 2023). Los peores resultados en clasificadores de género se observaron en mujeres negras, lo que evidencia la necesidad de auditorías interseccionales (Buolamwini y Gebru, 2018; Simó, 2024). El análisis debe examinar variables proxy<sup>6</sup> que, de manera aislada, carecen de interés, pero que en conjunto revelan información sensible (Éticas

6 En un sistema de inteligencia artificial, una variable proxy es una medición indirecta que se utiliza en lugar de la variable real que se desea optimizar o predecir, generalmente porque esta última es difícil de observar o cuantificar. Por ejemplo, un algoritmo puede usar el nivel de ingresos como variable proxy de la capacidad adquisitiva, o el tiempo de permanencia en una web como proxy del interés del usuario. El uso de variables proxy puede introducir sesgos o distorsiones, ya que no siempre representan fielmente la realidad que se intenta modelar.

Research and Consulting, 2021). El informe debe explicitar qué variables proxy captura el sistema y cómo las combina, evitando que se capitalicen correlaciones casuales (Éticas Research and Consulting, 2021; Bustelo, 2024).

Este enfoque se alinea con la flexibilización de la vigilancia intelectual (Ovalle et al., 2023) mediante el uso de metodologías que preserven las características sociales e históricas de los grupos interseccionales (Ovalle et al., 2023).

Las variables proxy son indicadores que, de manera aislada, pueden carecer de interés, pero que, al ser examinados junto a otras variables, revelan información sensible mediante inferencia (Éticas Research and Consulting, 2021).

### 5.3. FASE DE INFORME DE AUDITORÍA. COMUNICACIÓN DE LA COMPLEJIDAD

El Informe de auditoría (Éticas Research and Consulting, 2021) debe comunicar de manera explícita y clara los resultados del análisis interseccional. Esto es parte de las responsabilidades de transparencia y rendición de cuentas de los sistemas de IA (Éticas Research and Consulting, 2021).

El informe debe destacar qué variables proxy o qué intersecciones de identidades fueron más determinantes para los resultados del sistema, y señalar los riesgos de discriminación asociados a estos hallazgos (Éticas Research and Consulting, 2021). La publicación de la información detallada ayuda a combatir la opacidad algorítmica y permite a los equipos humanos comprender el impacto social de los modelos (Éticas Research and Consulting, 2021).

Además, la interseccionalidad debe ser un criterio fundamental para certificar la fiabilidad y equidad de un sistema (Simó, 2024). Un algoritmo no puede ser considerado fiable si demuestra un rendimiento deficiente o sesgado para cualquier subgrupo interseccional relevante (Simó, 2024).

## 6. DESAFÍOS PRÁCTICOS Y CONCEPTUALES DE LA AUDITORÍA INTERSECCIONAL

Si bien la integración de la interseccionalidad en la auditoría algorítmica es una estrategia poderosa para promover la equidad (Ovalle et al., 2023), su implementación no está exenta de importantes desafíos prácticos y conceptuales (Simó, 2024).

### 6.1. REDUCCIONISMO Y SIMPLIFICACIÓN

Existe una tendencia en la investigación de IA a reducir la interseccionalidad a una simple "equidad de subgrupos", lo que se limita a optimizar métricas sobre grupos demográficos sin abordar la realidad social subyacente que impulsa la opresión.

Esta simplificación puede llevar a la despolitización del enfoque, olvidando su raíz genealógica en el feminismo negro y su énfasis en la justicia social (Ovalle et al., 2023).

### 6.2. PROBLEMAS DE MUESTREO Y DATOS RAROS

Los auditores señalan que los problemas de tamaño de muestra son muy reales para las clases demográficas pequeñas, que a menudo son los grupos interseccionales más vulnerables (Costanza-Chock et al., 2022). Si los datos de estos grupos marginalizados son escasos, pueden ser malinterpretados como "ruido" estadístico (Simó, 2024), lo que dificulta la modelización robusta y la capacidad del algoritmo para generalizar a estas poblaciones (Buolamwini y Gebru, 2018).

### 6.3. DIFICULTADES LEGALES Y DE RECOLECCIÓN DATOS

Algunos auditores expresan preocupación sobre la posibilidad de infringir la legislación antidiscriminación al intentar recopilar los datos demográficos necesarios para realizar un análisis interseccional detallado, especialmente en categorías que no están legalmente protegidas (Costanza-Chock et al., 2022).

Para mitigar estos desafíos, se requiere que la auditoría no solo se concentre en el análisis técnico, sino que también incorpore el pensamiento complejo para comprender la interconexión y las múltiples dimensiones de los sesgos algorítmicos (Bustelo, 2024). La auditoría debe ser un proceso continuo, incluyendo la revisión constante del impacto del modelo en entornos dinámicos y fomentando la participación de las comunidades que podrían ser directamente perjudicadas por el sistema (Costanza-Chock et al., 2022).

## 7. DESAFÍOS METODOLÓGICOS EN LA INTEGRACIÓN

La operativización de la interseccionalidad en la auditoría algorítmica se confronta con varios retos metodológicos clave, muchos de los cuales reflejan tensiones epistemológicas históricas entre las ciencias sociales críticas y las ciencias computacionales.

### 7.1. LA TENSION ENTRE LA TEORÍA CRÍTICA Y LA CUANTIFICACIÓN TÉCNICA

El desafío más acuciante radica en cómo los investigadores de IA, que a menudo operan bajo epistemologías arraigadas en el positivismo y el empirismo (Ovalle et al., 2023), interpretan y operacionalizan un marco conceptual cuyo objetivo fundacional es la justicia social y la praxis (Crenshaw, 1990; Ovalle et al., 2023).

#### 7.1.1. Reducción a la equidad de subgrupos

Cómo incide Ovalle (2023), una revisión crítica de la literatura revela que la mayoría de los investigadores reducen abrumadoramente la interseccionalidad a la optimización de métricas de equidad sobre subgrupos demográficos. Esto conlleva a menudo a la confusión entre la interseccionalidad y la equidad de subgrupos.

### 7.1.2. Desconexión estructural

Esta reducción a la cuantificación simplifica la crítica, neutralizando el potencial del concepto que fue diseñado para examinar las relaciones de poder que crean y refuerzan las desigualdades. Al operar de esta manera, los investigadores fallan en conectar explícitamente los subgrupos a las estructuras sociales, el contexto o la historia que causan la desigualdad (Ovalle et al., 2023).

## 7.2. EL PROBLEMA DE LAS CATEGORIZACIONES RÍGIDAS

La interseccionalidad, en su origen, buscó romper la visión monolítica de la mujer, evitando la esencialización de las categorías y la cosificación de las identidades (Simó, 2024). Sin embargo, al aplicarse a la IA, surge el riesgo de la simplificación y neutralización del concepto (Ovalle et al., 2023; Parra y Busquier, 2022).

### 7.2.1. Neutralización y despolitización

El concepto se ha expandido y masificado de manera acrítica en el ámbito académico y en las políticas públicas, lo que conlleva su simplificación y neutralización (Parra y Busquier, 2022). Se busca una interpretación heurística y vaga, fácil de operacionalizar, lo que neutraliza la capacidad del concepto para confrontar las estructuras de poder.

### 7.2.2. La Necesidad de flexibilidad

La complejidad intrínseca de la interseccionalidad obliga a rechazar las categorías estáticas y las categorizaciones rígidas. La IA, al reducir las identidades a atributos fijos para su modelado, a menudo diluye la intencionalidad original del marco, que consiste en cuestionar cómo las estructuras sociales influyen en las diferencias grupales percibidas (Ovalle et al., 2023).

### 7.2.3. Retos de datos raros

Incluso cuando los auditores desean implementar el análisis interseccional (el 65% de los auditores expresan interés) (Costanza-Chock et al., 2022), se enfrentan a desafíos prácticos como la escasez de datos en los subgrupos interseccionales (Simó, 2024; Costanza-Chock et al., 2022). Los datos de grupos marginalizados con baja prevalencia pueden ser malinterpretados como "ruido" estadístico, haciendo que la modelización sea inservible para estas poblaciones (Buolamwini & Gebru, 2018; Simó, 2024; Eticas Research and Consulting, 2021).

## 8. CONCLUSIONES E IMPLICACIONES FUTURAS

Este análisis revela que los sesgos en los Modelos de Lenguaje a Gran Escala constituyen un problema complejo y multifacético, presente a lo largo de todo su ciclo de vida, desde la recopilación de datos hasta su implementación final. Se ha argumentado que los enfoques unidimensionales para la detección de sesgos son insuficientes y, en ocasiones, contraproducentes. La conclusión principal es que la interseccionalidad no debe ser considerada simplemente como una perspectiva teórica deseable, sino como una herramienta analítica esencial para una auditoría y mitigación de sesgos verdaderamente efectiva, que aspire a la justicia social.

Para lograr una inteligencia artificial más equitativa, es imprescindible adoptar un enfoque interdisciplinario que integre de manera genuina los conocimientos de la sociología, la ética y las ciencias de la computación. La solución a los sesgos algorítmicos no es puramente técnica; requiere un compromiso profundo con valores como la justicia, la transparencia, la explicabilidad y el diseño centrado en el ser humano. Solo a través de la colaboración entre disciplinas y un enfoque crítico y reflexivo podremos construir sistemas de inteligencia artificial que sirvan a toda la humanidad y no solo a sus segmentos más privilegiados.

Las implicaciones de este trabajo señalan varias líneas de investigación futuras que son cruciales para avanzar en este campo:

- **Desarrollo de métricas de equidad explícitamente interseccionales.** Es necesario ir más allá de las métricas existentes y crear nuevas formas de cuantificar el sesgo que capturen las disparidades en las intersecciones de identidades.
- **Creación de corpus de entrenamiento más diversos y representativos a nivel global.** Se deben realizar esfuerzos concertados para recopilar y curar datos que reflejen la diversidad lingüística, cultural y social del mundo, prestando especial atención a las comunidades históricamente marginadas.
- **Investigación sobre metodologías de auditoría cualitativas y mixtas.** Se necesitan nuevos métodos de auditoría que puedan manejar la fluidez, el contexto y la naturaleza relacional de las identidades sociales, complementando los análisis cuantitativos con enfoques cualitativos que capturen las experiencias vividas de las personas afectadas por estos sistemas.

## 9. BIBLIOGRAFÍA

- Adilazuarda, M. F., Mukherjee, S., Lavania, P., Singh, S., Aji, A. F., O'Neill, J., Choudhury, M. (2024). Towards measuring and modeling "culture" in LLMs: A survey. *arXiv preprint arXiv:2403.15412*. <https://doi.org/10.48550/arXiv.2403.15412>
- Bai, X., Wang, A., Song, I., & Griffiths, T. L. (2025). Explicitly unbiased large language models still form biased associations. *Proceedings of the National Academy of Sciences*. <https://doi.org/10.1073/pnas.2416228122>
- Blodgett, S. L., Barocas, S., Daumé III, H., & Wallach, H. (2020). Language (technology) is power: A critical survey of "bias" in NLP. In *Proceedings of the 58th Annual Meeting of the Association for Computational Linguistics*. <https://doi.org/10.48550/arXiv.2005.14050>
- Buolamwini, J., & Gebru, T. (2018). Gender shades: Intersectional accuracy disparities in commercial gender classification. In *Conference on fairness, accountability and transparency* (pp. 77-91). PMLR. <https://proceedings.mlr.press/v81/buolamwini18a.html>

- Bustelo, J. L. (2024). La intersección entre la inteligencia artificial (IA), el pensamiento complejo y la metodología de auditoría de sesgo. *Revista Iberoamericana De Complejidad Y Ciencias Económicas*, 2(4), 5-16. <https://doi.org/10.48168/ricce.v2n4p5>
- Chu, Z., Wang, Z., & Zhang, W. (2024). Fairness in large language models: A taxonomic survey. *ACM SIGKDD Explorations Newsletter*, 26(1), 34-48. <https://doi.org/10.1145/3682112.3682117>
- Costanza-Chock, S., Raji, I. D., & Buolamwini, J. (2022). Who Audits the Auditors? Recommendations from a field scan of the algorithmic auditing ecosystem. In *Proceedings of the 2022 ACM Conference on Fairness, Accountability, and Transparency* (pp. 1571-1583). ACM. <https://doi.org/10.1145/3531146.353321>
- Crenshaw, K. (1989). Demarginalizing the intersection of race and sex: A Black feminist critique of antidiscrimination doctrine, feminist theory and antiracist politics. *University of Chicago Legal Forum*, 1989(1), 139-167. Recuperado de <https://chicagounbound.uchicago.edu/uclf/vol1989/iss1/8>
- Crenshaw, K. (1990). Mapping the margins: Intersectional identity politics, and violence against women of color. *Stanford Law Review*, 43(6), 1241-1299. <https://doi.org/10.3917/cdge.039.0051>
- Devinney, H., Björklund, J., & Björklund, H. (2024). We don't talk about that: case studies on intersectional analysis of social bias in large language models. In *Workshop on Gender Bias in Natural Language Processing (GeBNLP)* (pp. 33-44). Association for Computational Linguistics. <https://aclanthology.org/2024.gebnlp-1.3/>
- Éticas Research and Consulting. (2021). *Guía de Auditoría Algorítmica*. Agencia Española de Protección de Datos. <https://eticas.ai/wp-content/uploads/2024/04/Guide-to-Algorithmic-Auditing-SP.pdf>
- Hill Collins, P., & Bilge, S. (2019). *Interseccionalidad* (ed. en español). Ediciones Morata.
- Hill Collins, P. (2017). La diferencia que crea el poder: interseccionalidad y profundización democrática. *Investigaciones Feministas*, 8(1), 19-39. <https://doi.org/10.5209/INFE.54888>
- Hu, T., Kyrzhenko, Y., Rathje, S., Collier, N., van der Linden, S., & Roozenbeek, J. (2025). Generative language models exhibit social identity biases. *Nature Computational Science*, 5(1), 65-75. <https://doi.org/10.48550/arXiv.2310.15819>
- IBM Data and AI Team. (2023, 16 de octubre). El sesgo de la IA, a la luz de ejemplos reales [Página Web]. <https://www.ibm.com/es-es/think/topics/shedding-light-on-ai-bias-with-real-world-examples>
- Jiao, W., Wang, W., Huang, J., Wang, X., & Tu, Z. (2023). Is ChatGPT a good translator? a preliminary study. *arXiv preprint arXiv:2301.08745*. <https://doi.org/10.48550/arXiv.2303.17466>
- Kotek, H., Dockum, R., & Sun, D. (2023). Gender bias and stereotypes in large language models. In *Proceedings of the ACM collective intelligence conference* (pp. 12-24). ACM. <https://doi.org/10.1145/3582269.3615599>
- Kotek, H., Sun, D. Q., Xiu, Z., Bowler, M., & Klein, C. (2024). Protected group bias and stereotypes in large language models. *ArXiv Preprint*. <https://doi.org/10.48550/arXiv.2403.14727>
- Martínez-Palacios, J. (2017). Exclusión, profundización democrática e interseccionalidad. *Revista de Investigaciones Feministas*, 8(1), 53-71. <https://doi.org/10.5209/INFE.54827>
- Naous, T., Ryan, M. J., Ritter, A., & Xu, W. (2023). Having Beer after Prayer? Measuring Cultural Bias in Large Language Models. *arXiv preprint arXiv:2310.13627*. <https://doi.org/10.48550/arXiv.2305.14456>
- Navigli, R., Conia, S., & Ross, B. (2023). Biases in large language models: Origins, inventory, and discussion. *ACM Journal of Data and Information Quality*, 15(2), 1-27. <https://doi.org/10.1145/35973>
- Ovalle, A., Subramonian, A., Gautam, V., Gee, G., & Chang, K.-W. (2023). Factoring the Matrix of Domination: A Critical Review and Reimagination of Intersectionality in AI Fairness. *AAAI/ACM Conference on AI, Ethics, and Society (AIES '23)*. <https://doi.org/10.1145/3600211.3604705>
- Parra, F., & Busquier, L. (2022). Retrospectivas de la interseccionalidad a partir de la resistencia desde los márgenes. *Las Torres de Lucca. Revista internacional de filosofía política*, 11(1), 23-35. <https://doi.org/10.5209/ltld.77044>
- Raji, I. D., Smart, A., White, R. N., Mitchell, M., Gebru, T., Hutchinson, B., Smith-Loud, J., Theron, D., & Barnes, P. (2020). *Closing the AI accountability gap: Defining an end-to-end framework for internal algorithmic auditing*. In *Proceedings of the Conference on Fairness, Accountability, and Transparency* (pp. 33-44). ACM. <https://doi.org/10.1145/3351095.3372873>
- Rebolledo, J., & Galaz, C. (2022). Interseccionalidad: Aspectos conceptuales y recomendaciones para las políticas públicas. Dirección de Estudios de PRODEMU. <https://www.prodemu.cl/wp-content/uploads/2023/05/Interseccionalidad12.pdf>
- Sakib, S. K., & Das, A. B. (2024). Challenging fairness: A comprehensive exploration of bias in LLM-based recommendations. In *2024 IEEE International Conference on Big Data (BigData)* (pp. 1585-1592). IEEE. <https://doi.org/10.1109/BgData62323.2024.10825082>
- Simó, E. (2024). Hacia una Inteligencia Artificial Interseccional para el tratamiento de información. *Investigaciones Feministas*, 15(1), 137-144. <https://doi.org/10.5209/infe.81954>
- Tan, Y. C., & Celis, L. E. (2019). Assessing social and intersectional biases in contextualized word representations. *Advances in Neural Information Processing Systems*, 32. <https://proceedings.neurips.cc/paper/2019/hash/201d546992726352471cfea6b0df0a48-Abstract.html>
- Tao, Y., Viberg, O., Baker, R. S., & Kizilcec, R. F. (2024). Cultural bias and cultural alignment of large language models. *PNAS nexus*, 3(9), pgae346. <https://doi.org/10.1093/pnasnexus/pgae346>
- Wan, Y., Pu, G., Sun, J., Garimella, A., Chang, K.-W., & Peng, N. (2023). "Kelly is a Warm Person, Joseph is a Role Model": Gender Biases in LLM-Generated Reference Letters. *Findings of the Association for Computational Linguistics: EMNLP 2023*, 3730-3748. <https://doi.org/10.48550/arXiv.2310.09219>
- Zhang, D., Zhang, Y., Bihani, G., & Rayz, J. (2024). Hire Me or Not? Examining Language Model's Behavior with Occupation Attributes. *arXiv preprint arXiv:2405.06687*. <https://doi.org/10.48550/arXiv.2405.06687>

## **NOTA BIOGRÁFICA**

Formador en competencias digitales e Inteligencia Artificial para la intervención social. Profesor del Grado de Trabajo Social en la Universidad Pontificia de Comillas. Graduado en Trabajo Social por la Universidad Complutense de Madrid, Máster en Redes Sociales y Aprendizaje Digital por la UNED, y Máster en Innovación e Inteligencia Artificial por Founderz-Microsoft. Desde 2017 ejerciendo como consultor para organizaciones del Tercer Sector y la administración pública, tras un bagaje de 10 años en el Tercer Sector como Trabajador Social.



# Justiça climática em tempos de IA: entre a rapidez, o saber científico e os desafios da alfabetização algorítmica

Justicia climática en tiempos de IA: entre la rapidez, el saber científico y los desafíos de la alfabetización algorítmica

Climate Justice in the Age of AI: Between the Speed, the Scientific Knowledge, and the Challenges of Algorithmic Literacy

Claudia Mendes Giesel  
Anne Caroline de Moraes Santos

## PALAVRAS-CHAVE:

Justiça Climática  
Inteligência Artificial  
Análise Crítica do Discurso  
Alfabetização Algorítmica.

## RESUMO:

A pesquisa tem como objetivo analisar criticamente discursos sobre justiça climática produzidos por chatbots de inteligência artificial (ChatGPT e Gemini), comparando-os a artigo científico publicado em periódico qualificado. Para isso, utilizou-se a Análise Crítica do Discurso (ACD) de Norman Fairclough como instrumento de análise desses textos, estudando-os a partir de três dimensões: textual, discursiva e prática-social. Além disso, deu-se ênfase aos níveis de profundidade, criticidade e fundamentação teórica dessas produções. Os resultados indicam que os textos automatizados apresentam linguagem coesa e acessível, mas carecem de densidade argumentativa e sustentação teórica, diferindo significativamente do rigor das produções científicas humanas. Conclui-se que o uso de inteligências artificiais deve ser acompanhado de práticas de alfabetização digital e algorítmica, que favoreçam a diversificação de fontes e o fortalecimento do pensamento crítico diante da crescente influência dos sistemas automatizados na mediação do conhecimento.

## PALABRAS CLAVE

Justicia Climática  
Inteligencia Artificial  
Análisis Crítico del Discurso  
Alfabetización Algorítmica.

## RESUMEN:

La investigación tiene como objetivo analizar críticamente los discursos sobre justicia climática producidos por chatbots de inteligencia artificial (ChatGPT y Gemini), comparándolos con un artículo científico publicado en una revista académica calificada. Para ello, se utilizó el enfoque de Análisis Crítico del Discurso (ACD) de Norman Fairclough como instrumento de análisis de los textos, examinándolos a partir de tres dimensiones: textual, discursiva y práctica social. Además, se enfatizaron los niveles de profundidad, criticidad y fundamentación teórica de dichas producciones. Los resultados indican que los textos automatizados presentan un lenguaje cohesionado y accesible, pero carecen de densidad argumentativa y sustento teórico, diferenciándose significativamente del rigor de las producciones científicas humanas. Se concluye que el uso de inteligencias artificiales debe ir acompañado de prácticas de alfabetización digital y algorítmica que fomenten la diversificación de fuentes y el fortalecimiento del pensamiento crítico ante la creciente influencia de los sistemas automatizados en la mediación del conocimiento.

## KEYWORDS:

Climate Justice  
Artificial Intelligence  
Critical Discourse Analysis  
Algorithmic Literacy.

## ABSTRACT:

The research aims to critically analyze discourses on climate justice produced by artificial intelligence chatbots (ChatGPT and Gemini), comparing them with a scientific article published in a qualified academic journal. To this end, Norman Fairclough's Critical Discourse Analysis (CDA) framework was used as the analytical instrument, examining the texts through three dimensions: textual, discursive, and social practice. In addition, emphasis was placed on the levels of depth, criticality, and theoretical grounding of these textual productions. The results indicate that automated texts present cohesive and accessible language but lack argumentative density and theoretical support, differing significantly from the rigor of human scientific productions. It is concluded that the use of artificial intelligences should be accompanied by digital and algorithmic literacy practices that promote the diversification of sources and the strengthening of critical thinking in the face of the growing influence of automated systems in the mediation of knowledge.

**CÓMO CITAR:** Moraes, A., & Giesel, C. (2025). Justiça climática em tempos de IA: Entre a rapidez, o saber científico e os desafios da alfabetização algorítmica. *RETIS. Revista de Tecnología para la Inclusión Social*, 2(1), 43-52.  
DOI: 10.70664/retis.v2i1.005

\* C-e: [anne.santos@uva.br](mailto:anne.santos@uva.br) / [claudia.giesel@uva.br](mailto:claudia.giesel@uva.br)

## 1. INTRODUÇÃO

Analisar discursos é de suma importância para que possamos dismantlar hegemonias e ideologias de poder existentes na sociedade. Discursos são construídos socialmente e historicamente. Na era digital e algorítmica, os discursos da Inteligência Artificial (IA) podem reforçar narrativas hegemônicas e fortalecer desigualdades. O avanço vertiginoso da IA Generativa (IAGen) – com ferramentas como o ChatGPT e o Gemini – tem redefinido os processos de produção e consumo de conhecimento, tornando as máquinas mediadoras ativas de narrativas sociais.

Neste cenário, a velocidade e a opacidade dos algoritmos encontram o debate complexo e urgente sobre a Justiça Climática. Este último exige a desnaturalização de estruturas de poder globais que perpetuam a vulnerabilidade das comunidades mais pobres. A emergência climática, portanto, não é apenas uma crise física, mas uma crise de justiça, onde a responsabilidade histórica dos países e corporações do Norte Global colide com a vulnerabilidade desproporcional do Sul Global.

Assim, o presente artigo explora como a Análise Crítica do Discurso (ACD), na perspectiva tridimensional de Norman Fairclough, pode contribuir para o avanço dos estudos sobre IA e Justiça Climática, destacando os impactos desses novos discursos automatizados na manutenção do *status quo*. A ACD, historicamente voltada para textos e interações humanas, revela-se um instrumento metodológico fundamental para analisar como os discursos da IA podem reproduzir e reforçar desigualdades existentes.

O objetivo é, portanto, analisar criticamente discursos sobre justiça climática produzidos por chatbots de inteligência artificial, comparando-os a um artigo científico de referência. Essa comparação visa aferir os níveis de profundidade, criticidade e fundamentação teórica, desvendando o viés ideológico da rapidez algorítmica.

Para isso, a ACD é utilizada a partir de suas três dimensões: textual (focada na nominalização e escolhas lexicais que apagam agentes), discursiva (analisando a intertextualidade e a superficialidade argumentativa) e prática-social (examinando como o discurso da IA contribui para o silenciamento de vozes e a manutenção de estruturas de poder). A hipótese central é a de que os textos automatizados, apesar de sua coesão linguística, carecem da densidade argumentativa e da sustentação teórica do saber científico, falhando em nomear e responsabilizar os agentes causadores da crise.

O artigo se desenvolve em três seções: a Seção 2 apresenta a ACD e as reflexões sobre Justiça Climática na era digital; a Seção 3 detalha a análise empírica, confrontando os recortes textuais da IA com o texto científico; e, por fim, a Seção 4 discute a urgência da Alfabetização Algorítmica e do Letramento de Prompt como caminhos possíveis para formar cidadãos críticos capazes de interagir com o conhecimento mediado pela IA de forma ética e reflexiva.

## 2. ANÁLISE CRÍTICA DO DISCURSO E REFLEXÕES SOBRE JUSTIÇA CLIMÁTICA NA ERA DIGITAL

Historicamente, teorias de análise de discurso sempre foram voltadas a um trabalho com textos e interações humanas. Essa realidade tem mudado com o avanço das tecnologias de inteligência artificial (IA). Com o surgimento da IA, muitos discursos, que eram produzidos e interpretados pelos seres humanos, atualmente são interpretados por algoritmos, que passam a construir formas de discursos próprias. Ao contrapor os discursos construídos por humanos com os processados pelas máquinas, entendemos que a IA tem uma capacidade de difusão em tempo real bem maior. Isso afeta a forma como discursos são disseminados e manipulados pelo *status quo*.

Entender que a IA não é neutra é de suma importância para que possamos avançar nos estudos com teorias de análises do discurso na sociedade contemporânea. Segundo Fernandes e Mozart (2025, p. 3),

o uso recente e constante da Inteligência Artificial (IA) em atividades e ferramentas cotidianas como redes sociais e trabalhos acadêmicos tem aberto a necessidade de diálogos éticos, críticos, sociais e econômicos levando-se em conta seu poderio semiótico e possibilidades da Análise do Discurso em busca de como esses discursos constroem ideologias, significados, dúvidas e assertivas de poder e manipulação.

A IA pode amplificar narrativas dominantes, que podem contribuir significativamente para percepções e decisões relacionadas a políticas públicas e conscientização ambiental em geral. Essas narrativas que silenciam sujeitos e agentes causadores dos graves problemas ambientais devem ser analisadas para que consigamos avançar em prol de uma justiça climática sustentável.

Integrar a ACD em pesquisas sobre discursos da IA pode contribuir para abordar questões relacionadas às dinâmicas de poder que são impregnadas de pressupostos ideológicos. Assim, o uso dessa abordagem crítica é um instrumento metodológico importante para analisar como a IAGen podem reproduzir e reforçar desigualdades existentes.

### 2.1. JUSTIÇA CLIMÁTICA: UMA REFLEXÃO CRÍTICO-REFLEXIVA

Não há como questionar que mudanças climáticas e o aquecimento global trazem importantes desafios para o mundo. Pensar nesses desafios engloba refletir criticamente sobre justiça climática. Segundo Sultana (2021, p.118),

a justiça climática trata fundamentalmente de prestar atenção em como a mudança climática afeta as pessoas de maneira diferente, desigual e desproporcional, bem como corrigir as injustiças resultantes de maneiras justas e equitativas. Os objetivos são reduzir a marginalização, a exploração e a opressão e aumentar a equidade e a justiça.

O conceito de justiça social está, portanto, inerentemente relacionado à práxis, que segundo Freire (1970, p. 51) nos remete a ideia de “reflexão e ação sobre o mundo a fim de transformá-lo”. Em termos amplos, uma reflexão crítica sobre justiça climática pode contribuir para que possamos dismantlar desigualdades estruturais que produzem injustiças (Foran, 2016; Mehta et al., 2021; Rice et al., 2015).

Nesta toada, é necessário tratar sobre formas de resistência de comunidades marginalizadas que sempre foram oprimidas por desigualdades históricas frutos do sistema capitalista. Diante disso, é indispensável abordar o conceito de justiça climática para

discutir as perspectivas Norte Global, focada no uso do desenvolvimento e sustentabilidade, e Sul Global, centrada na soberania dos povos, equidade social, história colonial etc.

O dilema entre Norte-Sul com relação à política climática está ligado a impasses significativos entre nações desenvolvidas e em desenvolvimento. Em geral, no Hemisfério Norte, a industrialização é mais acentuada e o efeito estufa é mais alto. Não à toa as nações do Norte possuem mais recursos financeiros e tecnológicos, que podem mitigar os efeitos das mudanças climáticas. Já as nações do Sul, como destaca Sultana (2022), são mais pobres financeiramente e tecnologicamente e acabam enfrentando os problemas climáticos de maneira desigual.

Jacobi, Filho e Pierro (2022, p. 36) alertam que “os padrões de consumo humano, associado ao meio de produção predatório vigente, acabaram por aumentar em 1,07°C a temperatura na Terra, entre os anos de 2011 e 2020. A previsão é a de que a média de temperatura ultrapasse 1,5°C até 2050 (IPCC 2022)”. Além disso, mencionam o levantamento da OCHA: “mais de 152 milhões de pessoas foram afetadas por 1.205 catástrofes ocorridas na América Latina e Caribe. [Essa] região é uma das mais sujeitas a eventos extremos, como ondas de calor, secas severas, chuvas volumosas e deslizamentos de terra”. Esses dados reforçam a necessidade de políticas ambientais capazes de enfrentar essa problemática de uma maneira mais justa e equitativa.

Jacobi et al. (2021, p.163) evidenciam a justiça climática no Brasil e argumentam sobre as fragilidades das populações mais necessitadas pela óptica dos direitos humanos. O debate precisa acontecer tanto por parte dos grandes empreendimentos econômicos poluidores quanto pelo poder público e movimentos sociais. Na prática, o que se vê são discursos demagogos, poucas ações concretas e efetivas, além do descumprimento de acordos.

No que tange às construções discursivas sobre esse tema, o discurso da academia, como qualquer outra forma de discurso, não é isento de ideologias. Segundo van Dijk (2016, p. 56), ideologias não são inatas,

elas são adquiridas gradualmente por pessoas como membros de grupos sociais, mediadas por experiências pessoais (modelos mentais subjetivos) exemplificadas por, ou generalizadas como atitudes socialmente compartilhadas em relação a assuntos políticos ou sociais relevantes. Para que tais atitudes sociais e as ideologias a elas subjacentes sejam adquiridas e compartilhadas em um grupo, elas geralmente precisam, antes de tudo, ser expressas, formuladas ou de alguma forma comunicadas entre os membros do grupo ou defendidas ou legitimadas fora do grupo.

A Inteligência Artificial também faz uso de discursos para legitimar hegemonias que perpetuam o *status quo*. A ACD se apresenta aqui como uma ferramenta que pode contribuir de maneira significativa para a desconstrução desses discursos.

## 2.2. A METODOLOGIA DA ANÁLISE CRÍTICA DO DISCURSO

Esta pesquisa pretende responder à seguinte questão: Como os discursos da IA são superficiais, sem fundamentação teórica e não responsabilizam os reais agentes causadores da crise climática?

Para isso, este estudo será centrado na ACD de Norman Fairclough e articulado em três dimensões do discurso presentes nos trabalhos do autor. Na dimensão textual, focaremos no processo de nominalização e escolhas lexicais. Para Fairclough (2003, p.13), “uma consequência comum da nominalização é que os agentes dos processos, ou seja, as pessoas que iniciam os processos ou agem sobre outras pessoas ou objetos, estão ausentes dos textos”. Para Fairclough (2003), as escolhas lexicais não são neutras e refletem relações de poder e ideologia.

Na dimensão discursiva, usaremos a intertextualidade como categoria de análise. De acordo com Fairclough (2004, p. 39), “a intertextualidade é a presença de elementos reais de outros textos dentro de um texto – citações. Mas existem várias formas menos óbvias de incorporar elementos de outros textos. Exemplos dessa forma mais velada são textos que trazem alusões implícitas de vozes que podem legitimar certas manutenções de *status quo* ou até mesmo pela omissão de representações sociais que lutam pela justiça climática e criticam abertamente os interesses dos grandes grupos. Dessa forma, vê-se que o uso de intertextualidade não é neutro e pode contribuir para que ideologias invisíveis sejam perpetuadas.

Na dimensão prática social, abordaremos como construções ideológicas e hegemônicas podem contribuir para silenciar vozes historicamente vulneráveis, fortalecendo, assim, um enfrentamento sustentável dos problemas climáticos. Segundo Fairclough (2003, p. 35),

A conexão entre o texto e a prática social é vista como mediada pela prática discursiva: de um lado, os processos de produção e interpretação são formados pela natureza da prática social, ajudando também a formá-la e, por outro lado, o processo de produção forma (e deixa vestígios) no texto, e o processo interpretativo opera sobre ‘pistas’ no texto.

Analisar discursos sob a ótica da ACD pode contribuir para que tenhamos um entendimento mais crítico e reflexivo sobre textos gerados pela IA e textos de referência no tocante às mudanças climáticas.

## 3. ANÁLISE EMPÍRICA: A REPRODUÇÃO DE VIESES NA ESCRITA DA INTELIGÊNCIA ARTIFICIAL

Nesta seção, realizaremos as análises de: 1. trechos de textos produzidos pela IA, mais especificamente de dois chatbots de IA generativa, ChatGPT e Gemini. Para dar mais amplitude à análise, foram escolhidas as versões gratuitas do ChatGPT e paga do Gemini; 2. partes de um texto científico publicado em 2023 na Revista Direito e Práxis, da Universidade do Estado do Rio de Janeiro (UERJ).

### 3.1. PARÂMETROS DE ANÁLISE

Os textos escolhidos possuem a mesma temática: as lutas por justiça socioambiental diante da emergência climática. Aqueles produzidos pela IA, tanto no ChatGPT quanto no Gemini, foram criados em 26 de setembro de 2025. Já o artigo científico da Re-

vista Direito e Práxis é de 2023. Fizemos busca pelos mesmos prompts no Gemini gratuito e não houve mudanças significativas para a versão paga, por isso a decisão foi por não analisar esses textos.

O objetivo da pesquisa é analisar como textos acadêmicos sobre justiça climática, criados por chatbots de IA generativa, estão sendo construídos em comparação com textos acadêmicos publicados em revistas científicas nacionais, como a Revista Direito e Práxis. Não usamos revista internacional, pois, como estamos trabalhando com ACD, entendemos ser melhor analisar textos de uma mesma língua, e escolhemos a língua portuguesa.

Para fazer comparação válida, partimos da temática presente no artigo publicado na Revista Direito e Práxis, em 2023, intitulado *As lutas por justiça socioambiental diante da emergência climática*. Essa temática foi usada para a produção dos textos pela IAGen.

A hipótese da pesquisa é a de que os textos criados por IA são mais superficiais, com pouco desenvolvimento e fundamentação teórica, sem citação de vozes que legitimam discursos de comunidades e movimentos sociais empenhados na justiça climática e com raras ou vagas menções às grandes lideranças e corporações mundiais e sua responsabilidade sobre a situação climática do planeta.

Sobre o texto acadêmico, o artigo foi escrito pela pós-doutora em Direito pela PUC Paraná e doutora em Meio Ambiente e Desenvolvimento pela Universidade Federal do Paraná (UFPR), Katya Regina Isaguirre-Torres, e pela doutora em Direitos Humanos, também pela UFPR, Tchenna Fernandes Maso. A Revista Direito e Práxis possui Qualis A1 e está indexada em importantes bases, portais e bibliotecas nacionais, regionais e internacionais.

Para os chatbots, foram criados dois prompts distintos dados de forma idêntica ao ChatGPT e ao Gemini: 1. prompt solicitando produção de artigo científico sobre a temática escolhida; 2. prompt solicitando a produção de artigo científico em doze páginas com fundamentação teórica atual. Destaca-se que o Gemini, quando solicitadas as doze páginas, afirmou que “é impossível para mim gerar um artigo científico de 12 páginas em um único turno de resposta, especialmente considerando a profundidade e o rigor acadêmico exigidos, além da formatação ABNT para referências, que exige detalhes específicos de cada fonte”. Já o ChatGPT criou o artigo. Por isso, mantivemos o prompt 2 e retiramos, para ambos os chatbots, a determinação das doze páginas, mas mantivemos a solicitação de artigo fundamentado.

O objetivo é, portanto, realizar análise de todos os textos mencionados anteriormente, utilizando os parâmetros da ACD para avaliar as três dimensões discursivas dos textos selecionados.

### 3.2. ANÁLISE DOS RECORTES TEXTUAIS A PARTIR DA ANÁLISE CRÍTICA DO DISCURSO (ACD)

Logo abaixo apresentamos os dois prompts utilizados para a construção dos artigos científicos por parte do ChatGPT e do Gemini e algumas informações importantes sobre os textos produzidos.

Tabela 1. Prompts utilizados

Prompt 1	Prompt 2
escreva um artigo científico sobre “As lutas por justiça socioambiental diante da emergência climática”.	escreva um artigo científico, com fundamentação teórica atual, sobre “As lutas por justiça socioambiental diante da emergência climática”

Fonte: autoral

Vale destacar que criamos os prompts com nível básico de complexidade, pois o objetivo dessa pesquisa era estudar o comportamento desses chatbots diante de possíveis comandos feitos por usuários médios, ou seja, ainda com pouco letramento algorítmico.

A diferença central entre os dois prompts está atrelada ao pedido de fundamentação teórica à IA com o intuito de percebermos se haveria diferenças significativas na produção com a solicitação direta do comando para a fundamentação atual. No Gemini, a fundamentação bibliográfica e a menção a relatórios atuais sobre a temática foram maiores do que o ChatGPT, que pouco fundamentou o texto mesmo com o prompt 2 solicitando essa fundamentação.

Para o prompt 1, o ChatGPT fez menção a duas fontes. Já o Gemini fez a seis. No caso do prompt 2, Gemini usou nove fontes bibliográficas, principalmente de teóricos da área ambiental, como o pesquisador sueco Andreas Malm, o americano Robert Bullard e as próprias autoras do artigo científico da Revista Direito e Práxis, que estamos analisando nessa pesquisa, Tchenna Maso e Katya Isaguirre-Torres.

No caso do ChatGPT, no prompt 2, amplia-se a fundamentação para seis, porém não são feitas ao longo do texto; links de sites e indexadores são inseridos após cada parágrafo. A menção, dessa forma, não é à pesquisa direta de teóricos especialistas no assunto, mas a sites em que os textos se encontram sem nenhuma menção autoral, como *Reserch gate*, *ScienceDirect* e *Climate in the Courts*. O próprio IPCC, Painel Intergovernamental sobre Mudanças Climáticas (*Intergovernmental Panel on Climate Change*) é mencionado dessa forma, sem menção direta aos dados disponibilizados, apenas com a sigla do IPCC na linha seguinte ao término do parágrafo com hiperlink para acessá-lo.

### 3.3. A LUTA PELA JUSTIÇA CLIMÁTICA SOB O OLHAR DAS DIMENSÕES DA ANÁLISE CRÍTICA DO DISCURSO: A DIMENSÃO TEXTUAL

Analisar discursos sob uma ótica crítica é desvendar relações de poder que contribuem para a manutenção de hegemonias capazes de silenciar vozes de grupos minoritários que historicamente foram e são silenciados. É importante ressaltar que todo texto seleciona categorias linguísticas aceitáveis em determinados contextos de produção de discurso. Na dimensão textual da ACD,

a nominalização, recurso linguístico em que uma ação ou processo é transformada em um substantivo, é usada para despersonalizar e naturalizar processos sociais, ou seja, esse processo cria um imaginário de neutralidade para ações que dependem de agentes sociais e decisões políticas. No prompt 1, podemos observar, com o uso do sujeito oracional “políticas públicas”, que há uma transformação de processos em substantivos, os quais apagam o agente que realiza a ação. Isso faz com que a ação discursiva seja impessoal, o que oculta responsabilidades específicas: “Políticas públicas devem reconhecer a centralidade das comunidades mais afetadas, promovendo reparação, redistribuição e reconhecimento” (Trecho criado a partir do Prompt 1 para ChatGPT).

Por outro lado, no texto científico, observamos que a ordem do discurso é organizada de maneira distinta daquela organizada pela IA, pois podemos observar a presença de agentes, o que, neste caso, claramente responsabiliza países desenvolvidos como aqueles que não cumprem suas obrigações históricas. Ao mencionar “os países desenvolvidos, apesar dos discursos de seus representantes, não assumiram o papel histórico que tiveram na emissão de poluentes”, o texto deixa explícito os agentes responsáveis. A escolha de como os discursos são organizados pela IA e pelo texto científico evidencia como a língua é utilizada como meio de poder para construir sentidos sobre obrigação e responsabilidade.

Na verdade, as propostas apresentadas na COP 26 assumiram compromissos muito mais singelos que a COP 21, que resultou no Acordo de Paris (2015). Os países desenvolvidos, apesar dos discursos de seus representantes, não assumiram o papel histórico que tiveram na emissão de poluentes. E se recusaram a respeitar as metas de redução, à medida que isso implicava qualquer ação que pudesse afetar as expectativas de lucros de suas corporações. Basta observar, que tais países não assumiram seu papel na dívida climática ao não transferirem recursos para o fundo climático, uma proposta de apoio aos países em desenvolvimento que sofrem com os impactos das mudanças climáticas. (Isaguirre-Torres; Maso, 2023, p. 460)

O uso do processo de nominalização contribui, portanto, para a construção de significados ideológicos. Ao transformarmos ações em substantivos, podemos ocultar quem está realizando a ação, o que acarreta implicações significativas na forma como atribuímos responsabilidades. (Fairclough, 2001).

Além do processo de nominalização, as escolhas lexicais feitas pela IA e pelo texto acadêmico divergem consideravelmente. Segundo a ACD, escolhas lexicais carregam significados ideológicos que afetam como os textos são recebidos e interpretados. Não existe neutralidade na escolha lexical já que essa escolha sempre irá privilegiar ou deslegitimar vozes. No prompt 2 do ChatGPT, as escolhas lexicais “políticas ambiciosas” atreladas aos jovens e com uso de aspas mostram uma clara tentativa de ironizar a luta tornando-a utópica e quicá inalcançável. O uso de aspas juntamente com a escolha das palavras pode contribuir para que o consumo do discurso sobre justiça climática e movimento juvenil seja enfraquecido.

Pesquisas atuais enfatizam que justiça climática não é apenas um princípio normativo, mas uma prática em evolução que envolve: estratégias legais (litígios estruturantes), resistência territorial (defesa de territórios indígenas e tradicionais), mobilização juvenil (pressão pública e demanda por “políticas ambiciosas”) e iniciativas de base para adaptação.

Para Fairclough (2001, p. 117) “[...] as ideologias são significações/construções da realidade que são construídas em várias dimensões das formas/sentidos das práticas discursivas e que contribuem para a produção, a reprodução ou a transformação das relações de dominação”. Assim, a escolha de palavras utilizadas em um texto pode perpetuar discursos hegemônicos e ideológicos que deslegitimam a luta em prol do avanço democrática de uma sociedade justa e sustentável.

No trecho abaixo, extraído do artigo científico, podemos observar que o uso das aspas em “Outros” tem um sentido ideológico diferente daquele representado no texto da IA. Esse uso sinaliza que esses grupos são representados como diferentes do padrão social hegemônico, centrado nas instituições, organizações e Estados. É importante ressaltar que essa escolha lexical enfatiza uma diferença que foi socialmente construída e legitimada pelo poder.

Esses elementos permitem definir o poder hegemônico, expresso nas negociações das Conferências das Partes, centrado nas negociações entre Estados, sendo que alguns deles possuem maior poder político do que outros, à medida que essas negociações reproduzem estruturas da dependência, assim como da colonialidade do poder. Assim, indígenas, quilombolas, camponeses, mulheres, esses povos “Outros” estão situados fora dos espaços de poder e decisórios, não à toa constituindo espaços alternativos paralelos. (Isaguirre-Torres; Maso, 2023, p. 460, grifo próprio)

De acordo com Fairclough (2003, p. 159) “[...] quem você é, é parcialmente uma questão de como você fala, como você escreve, bem como uma questão de corporificação - como você se parece, como você se posiciona, como você se movimenta, e assim por diante”. Assim, o uso de “Outros” mostra claramente a intenção das autoras do texto científico de fazer uso dessa escolha lexical para afirmar sua defesa das questões de justiça climática.

Na próxima seção, abordaremos como a dimensão discursiva se apresenta nos discursos analisados da IA e do texto científico.

### 3.4. A CONSTRUÇÃO DE SENTIDOS SOBRE A JUSTIÇA CLIMÁTICA: A DIMENSÃO DISCURSIVA

A estrutura de um texto científico e os estudos sobre gêneros acadêmicos mobilizam diferentes obras sobre o assunto todos os anos. No site da Estante Virtual, só em 2024, foram 73 obras produzidas envolvendo essas temáticas. Normas direcionadoras envolvendo número de páginas, regras de formatação a serem seguidas, apresentação clara da problemática, hipóteses, marco teórico, metodologia, resultados e impacto da pesquisa são exigências para que os textos sejam aprovados por avaliadores em revistas científicas de todo o Brasil e do mundo. Não à toa, periódicos sérios são cobrados de terem bem delineadas suas normas de submissão em seus sites. Tais revistas possuem alto rigor em seus critérios de submissão e aceitação do texto, e os autores que se submetem a esses processos de análise e publicação de seus textos entendem a necessidade desse rigor e atendem as solicitações de mudanças, que são recorrentes, para o melhor desenvolvimento do texto, seu grau de criticidade, análise e organização dos resultados.

Foi isso que vimos no artigo científico analisado nesta pesquisa. Primeiramente resumo em três línguas para ampliar o alcance linguístico do texto. Depois, estrutura dividida em Introdução, três seções de desenvolvimento e as Conclusões Finais. Todas as seções bem desenvolvidas e sem desequilíbrio de extensão. As referências bibliográficas estavam ao final do texto com 33 itens.

Já o que vemos na produção da IA é bem diferente: textos rasos, pouca fundamentação e desenvolvimento, seções pequenas e pouco desenvolvidas, apresentação de análises em tópicos. Para prompt 1 e 2, tanto ChatGPT como Gemini inseriram resumo ao texto apenas em português. Para prompt 1, Gemini criou Introdução com três parágrafos, dois capítulos de desenvolvimento, um com dois parágrafos e três tópicos numerados de 1 a 3 com três linhas cada para trazer formas de manifestação do capitalismo climático. O outro veio com a mesma estrutura, porém com três parágrafos. A Conclusão com dois parágrafos. Já as referências com seis itens. No prompt 2, o Gemini amplia a estrutura do texto com Introdução com três capítulos, dois capítulos de desenvolvimento com subitens (2.1, 2.2, 2.3, 3.1, 3.2 e 3.3). Os subcapítulos são curtos, dois parágrafos cada para seção 2 e um parágrafo para itens da seção 3.

No que toca às referências, como dito antes, o artigo científico apresenta 33 itens. Já o Gemini, para o prompt 1, usa seis referências e, para prompt 2, nove referências. O ChatGPT, para prompt 1, apresenta três referências, e, para prompt 2, sete.

Fairclough (2016, p. 96) salienta que a dimensão discursiva “é constitutiva tanto de maneira convencional como criativa: contribui para reproduzir a sociedade (identidades sociais, relações sociais, sistemas de conhecimento e crença) como é, mas também contribui para transformá-la”. Os discursos analisados da IA e o texto científico evidenciam essa reprodução social, através da intertextualidade, mencionada por Fairclough. O trecho abaixo, retirado do texto gerado pelo Gemini a partir do prompt 2, aponta para uma discussão que abordará questões econômicas e contestará o sistema capitalista como propulsor da crise climática vivenciada atualmente. Por outro lado, podemos observar um claro apagamento de nomeação de agentes.

**A emergência climática constitui o maior desafio civilizacional do século XXI.** Os dados científicos, consolidados nos relatórios do Painel Intergovernamental sobre Mudanças Climáticas (IPCC), apontam para um aumento irreversível na frequência e intensidade de eventos extremos (IPCC, 2022). No entanto, limitar a crise climática a uma questão meramente física ou tecnológica obscurece sua dimensão mais crítica: a crise de justiça. (grifo próprio)

Vemos, portanto, que, na passagem grifada, os agentes envolvidos nesse desafio não são mencionados. Além disso, fala-se em dados científicos e aumento irreversível na frequência de eventos extremos, mas não são feitas menções críticas aos países desenvolvidos responsáveis pelos graves problemas climáticos globais. Não trazer vozes que criticam realidades contribui para a manutenção do status quo.

Em contrapartida, o trecho do texto científico analisado nomeia e critica os países da ONU por apresentarem “propostas modestas” para a resolução do problema climático. Embora os dois trechos usem relatórios do IPCC, os dois fragmentos discursivos organizam a ordem do discurso de maneira distinta. No trecho científico, há uma intenção clara da autora de criticar às políticas governamentais existentes. Ao nomear os países da ONU, a autora destaca os agentes e os traz para o centro do debate sobre medidas ineficazes diante de um cenário de gravidade. O texto científico usa a intertextualidade para denunciar discursos hegemônicos.

As lideranças dos quase 200 países que participaram da 26ª Conferência das Nações Unidas sobre Mudanças Climáticas (COP26), em novembro de 2021, apresentaram propostas bastante modestas para responder à crise climática. Apenas alguns meses depois, em fevereiro de 2022, o Painel Intergovernamental da ONU sobre mudanças climáticas (IPCC) apresentou um relatório dramático, no qual aponta a ineficiência das medidas que vêm sendo implementadas pelos países, e a gravidade das consequências do aquecimento global para vários povos do mundo. (Isaguirre-Torres; Maso, 2023, p. 460)

Além disso, o trecho abaixo extraído do discurso da IA utiliza o trecho “países industrializados e corporações globais” sem mencionar os agentes dos problemas ambientais que causam injustiça climática. O trecho da IA até faz uma tentativa de salientar uma distribuição desigual, mas não aponta os responsáveis para compreender a crise climática.

O problema central reside na **distribuição desigual** das causas e dos impactos da crise. Países industrializados e corporações globais, historicamente responsáveis pela maior parte das emissões de gases de efeito estufa, impõem o ônus do ajuste e do sofrimento às comunidades do Sul Global e às populações de baixa renda, que menos contribuíram para o problema. Essa disparidade define o campo da **injustiça climática**. (grifo da IA)

Por outro lado, o artigo científico pormenoriza e nomeia os agentes ao mencionar que “nenhum dos maiores extrativistas aderiu à aliança”. Ao trazer vozes que colaboram para o fortalecimento de críticas sobre um trabalho sustentável com a justiça climática, as autoras deixam claro sua intenção de colocar os agentes no centro das discussões. Além disso, o trecho salienta responsabilidades históricas e estruturais de grandes emissores de poluentes.

Um outro documento assinado foi a “Aliança para além do petróleo e do gás”, que indica uma moratória na extração de hidrocarbonetos. A iniciativa foi promovida por Costa Rica e Dinamarca e conseguiu a adesão da França, Suécia, Irlanda, Groenlândia e País de Gales, e associados como o Estado da Califórnia nos EUA. Contudo, nenhum dos maiores extrativistas aderiu à aliança. Na conferência também foi proposto o Acordo de Glasgow sobre Emissão Zero de veículos, que pretende encerrar a venda de motores de combustão interna até 2040. O acordo contou com a participação de empresas que representam cerca de ¼ das vendas globais e foi assinado por cidades, como Buenos Aires, La Paz e São Paulo. Ficaram de fora países com os maiores mercados consumidores, como os EUA, China e Brasil e também as grandes montadoras (GUDYNAS, 2022). (Isaguirre-Torres; Maso, 2023, p. 466)

Na próxima seção, usaremos a dimensão prática para analisar os discursos objetos deste estudo.

### 3.5. DESCONSTRUINDO IDEOLOGIAS E HEGEMONIAS: A DIMENSÃO PRÁTICA SOCIAL

A mudança climática não está somente relacionada a problemas com o meio ambiente. Essas mudanças e estão também intrinsecamente enraizadas a processos históricos de relação de poder, que perpetuam e mantêm o *status quo*. É de suma importância entender como a prática social se relaciona com poder e ideologia e como esses discursos moldam os desafios estruturais de poder. Para Fairclough (2001, p. 121)

As ideologias surgem nas sociedades caracterizadas por relações de dominação com base na classe, no gênero social, no grupo cultural, e assim por diante, e à medida que os seres humanos são capazes de transcender tais sociedades, são capazes de transcender a ideologia.

O trecho extraído da IA menciona a luta dos povos indígenas, mas não nomeia os responsáveis pelo “desmatamento e infraestrutura que geram emissões e vulnerabilidades”. O apagamento desses agentes pode ser entendido como uma desigualdade de poder já que a não menção aos atores torna a luta dos povos indígenas vulnerável com um agente invisível. Como lutar com autores que são invisíveis? A quem isso fortalece e a quem legitima?

Em contextos como a Amazônia, povos indígenas e comunidades tradicionais combinam estratégias jurídicas, ocupações, comunicação pública e prática de manejo territorial para enfrentar projetos de desmatamento e infraestrutura que geram emissões e vulnerabilidades.

Além disso, outro trecho da IA evidencia uma lacuna de poder estrutural ao deixar de mencionar os atores da redistribuição de recursos: “estudos de caso mostram que sem redistribuição de recursos”. Não há aqui explícita a responsabilidade dessa redistribuição. A ausência de um agente específico reforça uma ideologia em prol de grandes grupos e corporações.

No entanto, a sustentabilidade e a justiça dessas iniciativas dependem de financiamento público estável e inclusão efetiva de comunidades na governança local. Estudos de caso mostram que sem redistribuição de recursos e participação real, adaptações locais podem reproduzir exclusões.

Em outro trecho, também podemos observar que essa ausência de nomeação dos agentes deixa claro o viés ideológico da IA. Quem seriam os atores hegemônicos? Qual a ideologia por trás desse apagamento? Qual a razão desse apagamento? Na prática social, essa relação do discurso com a sociedade deve ser compreendida ideologicamente para entendermos como discursos moldam pensamentos e contribuem para a manutenção de ideologias que comprometem o avanço de ações sustentáveis com relação à justiça climática.

Movimentos sociais e comunidades de base frequentemente enfrentam barreiras institucionais — burocracia, criminalização da ação coletiva, desinformação — e pressões de interesses econômicos organizados. A captura de espaços participativos por atores hegemônicos pode reduzir a eficácia de políticas alegadamente “justas”.

Por outro lado, ao analisarmos o trecho extraído do artigo científico, observamos uma crítica clara às disputas geopolíticas e a ordem neoliberal naturalizada na irresponsabilidade histórica dos países ricos com relação à crise climática. Fairclough (2021) argumenta que, na prática social, o discurso pode atuar como contra-hegemônico no sentido de dismantlar estruturas de poder que evidenciam um sujeito neutro ou apagado de suas responsabilidades históricas com relação à distribuição de responsabilidade global em relação a problemas de ordem climática.

Na verdade, as propostas apresentadas na COP 26 assumiram compromissos muito mais singelos que a COP 21, que resultou no Acordo de Paris (2015). Os países desenvolvidos, apesar dos discursos de seus representantes, não assumiram o papel histórico que tiveram na emissão de poluentes. E se recusaram a respeitar as metas de redução, à medida que isso implicava qualquer ação que pudesse afetar as expectativas de lucros de suas corporações. Basta observar, que tais países não assumiram seu papel na dívida climática ao não transferirem recursos para o fundo climático, uma proposta de apoio aos países em desenvolvimento que sofrem com os impactos das mudanças climáticas.

Em outro trecho do artigo científico, há uma clara articulação entre o debate sobre o meio ambiente com estruturas sociais e econômicas do modelo capitalista. Esse modelo, segundo o trecho, perpetua desigualdades entre os países desenvolvidos e em desenvolvimento e mantém o *status quo* de uma ordem mundial desigual, ou seja, o desenvolvimento de um é alicerçado na exploração do outro.

As nações ricas sustentam altos padrões de desenvolvimento social e preservação ambiental por dispõem de recursos extraídos de países do Sul global (ou para sermos mais precisas a teoria proposta, dependentes). Os países “subdesenvolvidos” sofrem de externalização dos problemas do desenvolvimento, sustentam os impactos ambientais: queimadas, barragens, agrotóxicos, transgênicos, contaminação das águas, trabalho escravo, retirada de direitos trabalhistas, superexploração do trabalho. Os países dependentes não produzem subdesenvolvimento por causalidade internas, mas por estarem submetidos a estrutura desigual de desenvolvimento pensada entre as nações, e consequentemente da distribuição dos ônus dos impactos.

A dimensão prática social critica um desenvolvimento que é ideológico e legitima desigualdades. Ao fazer isso, a ACD contribui para o avanço de pautas relacionadas a uma sustentabilidade global real e não somente aquelas imbuídas de narrativas dominantes. Com isso, faz-se necessário um debate sobre alfabetização algorítmica já que muitos discursos da IA são carregados de hegemonias e ideologias que enfraquecem as lutas em prol de uma mudança climática sustentável que realmente possa representar grupos historicamente silenciados.

#### 4. A ALFABETIZAÇÃO ALGORÍTMICA E O LETRAMENTO DE DADOS E DE PROMPT COMO CAMINHOS POSSÍVEIS

Após a apresentação da análise dos textos científicos criados pela IA em comparação ao texto produzido por pesquisadoras brasileiras, vemos a necessidade de tratar sobre alfabetização algorítmica e letramento digital para que as fontes de informação utilizadas dentro e fora de espaços educacionais sejam diversificadas, abrangendo não apenas pesquisa feitas em chatbots de IAGen, mas em textos acadêmicos publicados em revistas sérias com base comprovadamente científica e crítica.

De acordo com dados fornecidos pela Open AI e publicado pelo G1 (2025), o Brasil é um dos países que mais utilizam o ChatGPT, ficando atrás apenas de EUA e Índia. O estudo revelou também que os usuários brasileiros enviam cerca de 140 milhões de mensagens ao dia, em um total de 2 bilhões de comandos diários. Quando olhamos para o cenário nacional, os estados que lideram em uso são São Paulo, Distrito Federal e Santa Catarina, seguidos por Tocantins, Rio de Janeiro, Ceará, Paraná, Amapá, Mato Grosso e Pernambuco.

Sobre os principais usos, 20% são para redação e comunicação e 15% para programação, ciência de dados e matemática. Os principais usuários estão entre pessoas de 25 a 34 anos (33%) e de 18 a 24 anos (27%). Segundo a matéria do G1, isso pode indicar que estudantes e recém-formados são os que mais usam essas plataformas. Em setembro de 2025, a OpenAI também anunciou que o ChatGPT estava para alcançar 700 milhões de usuários ativos por semana, um aumento expressivo quando comparado aos 500 milhões registrados em março desse mesmo ano (FastCompany, 2025). Já o Gemini, soma mais de 450 milhões de usuários ativos mensais.

Esses dados evidenciam o crescimento exponencial da presença da IA no cotidiano nacional e internacional e reforçam a importância da alfabetização algorítmica. Compreender como essas ferramentas funcionam, suas possibilidades e limitações, torna-se essencial para formar usuários críticos, capazes de interagir com os sistemas de maneira ética, reflexiva e informada.

O termo letramento (em inglês *literacy*), segundo Magda Soares (2014), chega na área da Educação e das Ciências Linguísticas na metade dos anos 80. Para a estudiosa, letramento seria o estado ou condição que determinado grupo social desenvolve como consequência de ter se apropriado dos mecanismos de leitura e de escrita. Saber ler e escrever, portanto, não significa que o indivíduo está fazendo isso de forma crítica e contextualizada.

Nessa perspectiva, se estamos hoje tratando sobre diferentes conceitos de letramento, como o digital, o midiático e o algorítmico, é porque o Brasil e o mundo estão vivenciando há algumas décadas o envolvimento da sua produção linguística atrelada a plataformas e meios digitais, intensificados pelo uso da IA, e agora, mais precisamente, da IAGen. Raquel Timponi e Raquel Evangelista (2025, p. 115) apresentam o cenário em que esse crescimento se dá e sua relação com a produção científica mundial atual:

o panorama global da inteligência artificial adentra uma nova fase. Esse estágio é marcado pelo aprimoramento de sistemas e agentes inteligentes voltados para a pesquisa avançada, ampliando a capacidade de análise de grandes volumes de dados e promovendo a automatização de processos complexos de curadoria informacional. Mais do que um embate comercial, esse movimento reflete uma mudança paradigmática: a transição da IAG de um modelo meramente reativo para um papel ativo na produção do conhecimento científico, redefinindo os limites da investigação acadêmica e impulsionando a inovação tecnológica.

As pesquisadoras destacam também que os critérios rigorosos de revistas sérias, com revisão por pares e indexação em bases científicas, defrontam-se com buscas fáceis de textos criados pela IAGen, textos esses, em sua maioria, opacos, sem apresentação da origem dos dados usados. Isso se torna um problema ainda maior quando estamos falando, principalmente, de um país como o Brasil em que 53% dos brasileiros não leram sequer uma parte de uma obra nos três meses anteriores à pesquisa realizada (Retratos da Leitura, 2024). No mesmo ano, o Indicador de Analfabetismo funcional (Inaf) revelou que a maior parte da população brasileira (35%) está dentro do chamado analfabetismo consolidado, ou seja, possui capacidade intermediária de leitura e escrita.

Com base nesse cenário, é importante tratar sobre os conceitos de alfabetização e letramento digital, letramento de dados ou algorítmico (em inglês *data literacy*), letramento de prompts, e o que grandes órgãos governamentais e institucionais têm apresentado sobre esses conceitos e sua importância para o desenvolvimento humano e científico no século XXI.

A Unesco (2023), no *Guia para a IA generativa na educação e na pesquisa*, evidencia a importância de se pensar em alfabetização nas dimensões humanas tecnológicas da IA para os estudantes, contudo, dados de 2022 revelaram que apenas 15 países estavam, de fato, tornando curricular essa dimensão nas escolas. Em 2018, no *Global Framework of Reference on Digital Literacy Skills for Indicator 4.4.2*, também utilizou o termo alfabetização e, dessa vez, associando-o ao letramento digital:

o **letramento digital** pode ser definido como: [...] a capacidade de acessar, gerenciar, entender, integrar, comunicar, avaliar e criar informações de forma segura e apropriada por meio de tecnologias digitais para emprego, empregos decentes e empreendedorismo. Inclui competências que são diversas vezes referidas como **alfabetização em informática**, **alfabetização em TIC**, **alfabetização informacional** e **alfabetização midiática**. (Unesco, 2023, grifo próprio)

Esse uso ora de alfabetização ora de letramento sem especificar as diferenças entre os dois conceitos também aparece nesse documento quando o conceito de letramento (*literacy*) midiático e informacional é abordado:

O **Letramento Midiático e Informacional (MIL)** enfatiza uma abordagem crítica do letramento. O MIL reconhece que as pessoas aprendem tanto dentro quanto fora da sala de aula, por meio de plataformas informacionais, midiáticas e tecnológicas. Ele permite que as pessoas questionem de forma crítica aquilo que leem, ouvem e aprendem. (Unesco, 2023, grifo próprio)

Importante entendermos as diferenças e aproximações que esses dois termos possuem. Magda Soares (2004), em *Letramento e Alfabetização: as muitas facetas*, estabelece a diferença entre os dois conceitos: 1) alfabetização é a dimensão linguística e discursiva fundamental para o sujeito dominar o sistema de linguagem humana, processo que permite o indivíduo ser alfabetizado; 2) letramento ocupa-se da dimensão social e cultural, em que o indivíduo usa a leitura e a escrita para se inserir nas práticas sociais.

Mesmo fazendo a distinção, Magda Soares (2024, p. 14, grifo próprio), no entanto, aponta a indissociabilidade desses dois processos: “dissociar alfabetização e letramento é um equívoco porque [...] a alfabetização desenvolve-se no contexto de e por meio de práticas sociais de leitura e de escrita, isto é, através de atividades de **letramento**, e este, por sua vez, só se pode desenvolver se houver a aquisição e o domínio da **tecnologia da escrita – a alfabetização**”.

Assim como Magda, outras referências nos estudos de letramento, como Roxane Rojo, Eliana Albuquerque e Andrea Brito Ferreira, nos últimos cinco anos, estudam essa temática e dialogam com o pensamento de Soares. Para Rojo (2022), o processo de alfabetizar letrando é o ponto de partida para o multiletrar, ou seja, o letramento em suas mais diferentes formas, como o visual, digital e algorítmico, por exemplo. Albuquerque e Brito (2024) corroboram a ideia e defendem que as práticas em sala de aula devem refletir essa integração entre alfabetização e letramento a partir de um trabalho pedagógico que valorize a diversidade de saberes dos educandos e que se organize a partir do cotidiano. Dessa forma, usar letramento e alfabetização de forma integrada, com base no pensamento das autoras supracitadas, não seria um equívoco; no entanto, é importante não usar de forma indiscriminada sem entender o que as aproxima e também diferencia.

Dito isso, fica evidente que hoje as práticas cotidianas estão cada vez mais entrelaçadas ao uso das tecnologias digitais. Nessa toada, surgem então os termos mencionados anteriormente: letramento digital, informacional e midiático. Outro termo usado recentemente, a partir do uso da IAGen, é letramento de dados (data literacy) e letramento de prompt (prompt literacy). Yohan Hwang, Jang Ho Lee e Dongkwang Shin (2023) tratam sobre o conceito de letramento de prompt e afirmam que gerar prompt, interpretar *outputs* (saídas) e refinar *prompts* são habilidades linguísticas que precisam ser desenvolvidas nos usuários para que possam usar bem os dados e entender o modo de funcionamento da IAGen.

Diante desse cenário, é determinante que os governos insiram o investimento nesses letramentos em suas políticas públicas. Dados da Unesco de 2023 revelaram que apenas sete países (China, Finlândia, Geórgia, Catar, Espanha, Tailândia e Turquia) estavam desenvolvendo treinamento para professores sobre IA. No Brasil, temos documentos criados pelo Ministério da Educação, como o *Guia de Educação Digital e Midiática*, que orienta as escolas a incorporar as competências digitais e enfatiza a necessidade de o espaço escolar promover o entendimento sobre algoritmos, uso de dados e suas implicações éticas e sociais das diferentes formas de IA. Esse guia é parte da Política Nacional de Educação Digital (PNED), instituída pela Lei nº 14.533/2023, e da Estratégia Brasileira de Educação Midiática (EBEM), estabelecendo a base para as competências de Cidadania Digital.

Inserir como componente curricular transversal em todas as disciplinas escolares e também nas universidades a compreensão de algoritmos, uso de dados e implicações éticas, com respeito aos princípios de dignidade humana (como aponta a UNESCO) é determinante para que avancemos nessa questão. Professores e alunos devem ser letrados para o uso da IA Gen em seu dia a dia escolar, profissional e pessoal.

Ao problematizarmos nessa pesquisa a opacidade dos textos científicos criados pela IA, sua pouca fundamentação e crítica sobre os agentes causadores dos problemáticos climáticos mundiais, precisamos afirmar essa necessária ação governamental em formar sua população para as novas práticas sociais e culturais que se formam para que debates sobre justiça climática sejam produzidos de forma rica e com a complexidade que o tema exige. Para isso, alfabetização/letramento algorítmico é passo vital.

## 5. CONSIDERAÇÕES FINAIS

Este artigo utilizou a Análise Crítica do Discurso (ACD) de Norman Fairclough para confrontar discursos sobre Justiça Climática gerados por chatbots de Inteligência Artificial Generativa (ChatGPT e Gemini) com a produção de rigor científico, representada por um artigo publicado em periódico Qualis A1. A análise se concentrou em dismantlar a neutralidade aparente dos textos automatizados, revelando como a velocidade do algoritmo pode, paradoxalmente, travar o avanço do debate crítico.

Os resultados confirmaram a hipótese inicial da pesquisa. Na dimensão textual, as ferramentas de IA, embora eficientes na coesão e na acessibilidade da linguagem, mostraram uma tendência sistemática ao apagamento de agentes por meio da nominalização (“políticas públicas”, “interesses econômicos organizados”). Este recurso linguístico tem a função ideológica de despersonalizar a responsabilidade histórica e estrutural da crise climática, ocultando os grandes emissores de poluentes e os grupos de poder do Norte Global.

Na dimensão discursiva, a análise da intertextualidade evidenciou a superficialidade e a baixa fundamentação das produções da IA. Ao contrário do artigo científico, que mobiliza 33 fontes e articula teorias críticas sobre colonialidade do poder e dependência, os chatbots apresentaram referências genéricas ou incompletas, demonstrando que o discurso automatizado, embora acessível, carece de densidade argumentativa e rigor epistemológico, reproduzindo um saber que é mais informativo do que formativo.

Na dimensão da prática social, o contraste se tornou mais evidente: enquanto o discurso científico coloca no centro do debate as comunidades silenciadas (povos indígenas, quilombolas, camponeses) e denuncia a assimetria geopolítica da dívida climática, os textos da IA, ao apagarem os agentes hegemônicos, enfraquecem a possibilidade de uma ação política transformadora. A rapidez do algoritmo, neste caso, serve como um instrumento para a manutenção do status quo, naturalizando as relações de dominação que perpetuam a injustiça climática.

Conclui-se, portanto, que a Alfabetização Algorítmica e o Letramento de Prompt (a nova habilidade de interagir, interpretar e refinar a IAGen) não são mais competências opcionais, mas sim estratégias vitais e contra-hegemônicas para a Educação do século XXI.

O crescimento exponencial do uso da IA Generativa no Brasil (com milhões de comandos diários) exige uma ação governamental imediata para inserir o letramento algorítmico no cerne das políticas públicas educacionais, de forma transversal e crítica. Somente ao capacitar estudantes e professores para dismantlar os “discursos invisíveis” dos algoritmos – compreendendo seus vieses, sua opacidade e sua dimensão ideológica – será possível garantir a diversificação das fontes de conhecimento e fortalecer o pensamento crítico necessário para a construção de uma Justiça Climática verdadeiramente sustentável, equitativa e democrática.

## 6. REFERÊNCIAS BIBLIOGRÁFICAS

- Ação Educativa & Conhecimento Social. (2024). Indicador de Alfabetismo Funcional (Inaf). <https://alfabetismofuncional.org.br/>
- Albuquerque, E. B. C. de, & Ferreira, A. T. B. (2024). A organização do trabalho pedagógico no ciclo de alfabetização: Os conhecimentos dos residentes em ação. *Signum: Estudos da Linguagem*, 27(1), 158–172. <https://ojs.uel.br/revistas/uel/index.php/signum/article/view/49534>
- Fairclough, N. (2001). *Discurso e mudança social* (I. Magalhães, Trad.). Editora Universidade de Brasília. (Obra original publicada em 1992)
- Fairclough, N. (2003). *Analisando discursos: Análise textual para pesquisa social* (J. A. Vieira, R. L. Vieira, & F. C. O. da Silva, Coord.). Editora Universidade de Brasília.
- Fairclough, N. (2016). *Discurso e mudança social* (2. ed.; I. Magalhães, Trad.). Editora Universidade de Brasília.
- Fernandes, W. S., & Guimarães, M. E. L. (2025). Análise crítica do discurso em trabalhos sobre algoritmos digitais e inteligência artificial na Educação Matemática. *Revista NUPEM*, 17(41), 1–22. <https://doi.org/e-2025018>
- FastCompany. (2025). OpenAI projeta 700 milhões de usuários semanais no ChatGPT. *FastCompany Brasil*. <https://fastcompany-brasil.com/ia/openai-projeta-700-milhoes-de-usuarios-semanais-no-ChatGPT/>
- Foran, J. (2016). Reimagining radical climate justice. In P. Wapner & H. Elver (Eds.), *Reimagining climate change* (pp. 150–170). Routledge. <https://doi.org/10.4324/9781315671468-9>
- Freire, P. (1970). *Pedagogy of the oppressed*. Continuum.
- G1. (2025, 12 de agosto). Brasil está entre os 3 países que mais usam o ChatGPT, diz OpenAI. *G1 Tecnologia*. <https://g1.globo.com/tecnologia/noticia/2025/08/12/brasil-esta-entre-os-3-paises-que-mais-usam-o-ChatGPT-diz-openai.ghtml>
- Hwang, Y., Lee, J. H., & Shin, D. (2023). What is prompt literacy? An exploratory study of language learners' development of new literacy skill using generative AI (arXiv:2311.05373). Cornell University. <https://arxiv.org/pdf/2311.05373>
- Instituto Pró-Livro (IPL). (2024). *Pesquisa Retratos da Leitura no Brasil: 6ª Edição [Apresentação de resultados]*. [https://www.prolivro.org.br/wp-content/uploads/2024/11/Apresentac%CC%A7a%CC%83o\\_Retratos\\_da\\_Leitura\\_2024\\_13-11\\_SITE.pdf](https://www.prolivro.org.br/wp-content/uploads/2024/11/Apresentac%CC%A7a%CC%83o_Retratos_da_Leitura_2024_13-11_SITE.pdf)
- Isaguirre-Torres, K. R., & Maso, T. F. (2023). As lutas por justiça socioambiental diante da emergência climática. *Revista Direito e Práxis*, 14(1), 458–485. <https://doi.org/10.1590/2179-8966/2023/73122>
- Jacobi, P. R., Arruda Filho, M. T., & Pierro, B. de. (2022). Ambiente e sociedade em tempos de emergência climática: Do resgate histórico ao momento atual. *Revista Fronteiras*, 11(3), 35–46. <https://doi.org/10.21664/2238-8869.2022v11i3.p35-46>
- Mehta, L., Srivastava, S., Movik, S., Adam, H. N., D'Souza, R., Parthasarathy, D., ... et al. (2021). Transformation as praxis: Responding to climate change uncertainties in marginal environments in South Asia. *Current Opinion in Environmental Sustainability*, 49, 110–117. <https://doi.org/10.1016/j.cosust.2021.04.002>
- Rice, J. L., Burke, B. J., & Heynen, N. (2015). Knowing climate change, embodying climate praxis: Experiential knowledge in Southern Appalachia. *Annals of the Association of American Geographers*, 105(2), 253–262. <https://doi.org/10.1080/00045608.2014.985628>
- Rojo, R. (2022). Multiletramentos na escola: Uma entrevista com Roxane Rojo. *Educitec - Revista de Estudos e Pesquisas sobre Ensino Tecnológico*, 8(1). <https://sistemascmc.ifam.edu.br/educitec/index.php/educitec/article/view/1998>
- Rojo, R., & Moura, E. de. (2023). Pedagogia dos multiletramentos: Os “multi” que modificam (e atualizam) a prática docente – Resenha Crítica. *Revista Letrilhando*. <https://letrilhando.com/2023/12/01/pedagogia-dos-multiletramentos-os-multi-que-modificam-e-atualizam-a-pratica-docente-resenha-critica/>
- Soares, M. (2004). Letramento e alfabetização: As muitas facetas. *Revista Brasileira de Educação*, (25), 5–17. <https://www.scielo.br/j/rbedu/a/89tX3SGw5G4dNWdHRkRxrZk/?format=pdf&lang=pt>
- Sultana, F. (2021). Critical climate justice. *The Royal Geographical Society (with IBG)*, 188(1), 118–124. <https://doi.org/10.1111/geoj.12>
- Sultana, F. (2022). The unbearable heaviness of climate coloniality. *Political Geography*, 99, 102638. <https://doi.org/10.1016/j.polgeo.2022.102638>
- Timponi, R., & Evangelista, R. L. (2025). Letramento Midiático, Algorítmico e Inteligência Artificial: O papel dos agentes inteligentes na curadoria da pesquisa acadêmica. *Revista Eco-Pós*, 28(1), 110–132. [https://revistaecopos.eco.ufrj.br/eco\\_pos/article/view/28452](https://revistaecopos.eco.ufrj.br/eco_pos/article/view/28452)
- Torres, P. H. C., Urbinatti, A. M., Gomes, C., Schmidt, L., Leonel, A. L., Momm, S., & Jacobi, P. R. (2021). Justiça climática e as estratégias de adaptação às mudanças climáticas no Brasil e em Portugal. *Estudos Avançados*, 35(102), 159–176. <https://doi.org/10.1590/s0103-4014.2021.35102.010160>
- UNESCO. (n.d.). *Media and information literacy: A critical approach to literacy in the digital world*. Recuperado em 17 de outubro de 2025, de <https://www.unesco.org/en/articles/media-and-information-literacy-critical-approach-literacy-digital-world>
- Van Dijk, T. A. (2016). *Discurso-cognição-sociedade: Estado atual e perspectivas da abordagem sociocognitiva do discurso*. *Letrônica*, supl. (Diálogos), s8–s29. <https://doi.org/10.15448/1984-4301.2016.s.23189>

# Entre la promesa y la amenaza: la inteligencia artificial generativa en la educación universitaria

## Between Promise and Threat: Generative Artificial Intelligence in University Education

**Albert García-Arnau**

*Universidad Complutense de Madrid*

**Carla Barrio**

*Universidad Nacional de Educación a Distancia*

### PALABRAS CLAVE:

Inteligencia Artificial  
Generativa  
Educación Universitaria  
Docencia  
Evaluación  
Innovación Pedagógica.

### RESUMEN:

Este artículo reflexiona sobre los cambios que la inteligencia artificial generativa (IAG) está produciendo en la educación universitaria. A partir de varias experiencias de innovación docente en la Universidad Complutense de Madrid y la UNED, se analizan los retos que plantea su uso en la enseñanza y la evaluación, así como las oportunidades que ofrece para mejorar ciertos procesos educativos. El texto combina la revisión de estudios recientes con observaciones surgidas del trabajo en el aula. En el texto se discuten las implicaciones éticas, epistemológicas y estructurales de la irrupción de la IAG. A modo de conclusiones, se destaca que el principal desafío no reside tanto en el fraude académico, sino en la masificación de su potencial y la despersonalización del aprendizaje. Se concluye que la universidad debe promover una integración crítica y equilibrada de la IAG, fortaleciendo tanto las competencias digitales como las predigitales mediante una reforma estructural del modelo educativo que reduzca las ratios y aumente la inversión.

### KEYWORDS:

Generative Artificial  
Intelligence,  
Higher Education  
Teaching  
Evaluation  
Pedagogical Innovation

### ABSTRACT:

This article reflects on the changes that generative artificial intelligence (GAI) is bringing about in higher education. Drawing from several experiences of teaching innovation at the Complutense University of Madrid and UNED, it analyses the challenges posed by its use in teaching and assessment, as well as the opportunities it offers to enhance certain educational processes. The text combines a review of recent studies with observations arising from classroom practice. It discusses the ethical, epistemological, and structural implications of the advent of GAI. In conclusion, it highlights that the main challenge does not lie so much in academic fraud, but rather in the widespread use of its potential and the depersonalisation of learning. It concludes that universities should foster a critical and balanced integration of GAI, strengthening both digital and pre-digital competences through a structural reform of the educational model that reduces student-to-teacher ratios and increases investment.

## 1. INTRODUCCIÓN

El objetivo de este artículo es fomentar una reflexión conjunta en torno a los efectos positivos y adversos que la inteligencia artificial generativa (IAG) está produciendo en la educación universitaria de las ciencias sociales. Se trata de problemas emergentes y discusiones en construcción, por lo que lejos de cerrar estas cuestiones, es nuestra intención fomentar un debate necesario. Más allá de esbozar líneas que planteen soluciones taxativas a los problemas planteados, se trata de abrir interrogantes para problematizar cuestiones que la investigación aún no ha podido construir un corpus asentado.

Las presentes reflexiones se inscriben en una trayectoria académica que articula la investigación teórica con la experiencia pedagógica directa en el ámbito de la enseñanza universitaria. En particular este texto se nutre de los aprendizajes derivados del proyecto “Integrando las herramientas de Inteligencia Artificial Generativa (IAG) en la enseñanza universitaria de asignaturas de sociología” (Proyecto 333, convocatoria 2024-2025, Innova UCM), coordinado por Albert García Arnau y el proyecto dirigido

**CÓMO CITAR:** García-Arnau, A., & Barrio, C. (2025). *Entre la promesa y la amenaza: La inteligencia artificial generativa en la educación universitaria*. RETIS. *Revista de Tecnología para la Inclusión Social*, 2(1), 53-57.  
DOI: [10.70664/retis.v1i2.006](https://doi.org/10.70664/retis.v1i2.006)

por Carla Barrio “Pruebas de Evaluación Continua a Distancia con Inteligencia Artificial Generativa: Investigación Social desde un Aprendizaje Activo” (PEC-IAG, Innovación Docente, UNED). Finalmente, las consideraciones aquí desarrolladas también se nutren de fuentes más informales, como las experiencias docentes acumuladas en el aula a partir de actividades mediadas por herramientas de IAG, los debates mantenidos en seminarios, las controversias aparecidas en prensa y medios, los talleres donde se han compartido prácticas docentes, así como las conversaciones cotidianas con colegas que atraviesan procesos similares de adaptación pedagógica a las nuevas realidades tecnosociales.

Según expone uno de los más recientes informes publicados en España sobre educación e inteligencia artificial generativa (IAG) (Fundación Cotec, 2025), la mayor parte de los docentes (67%) percibe que la inteligencia artificial tiene un impacto negativo en los procesos cognitivos de los alumnos. También son mayoría los docentes encuestados (66%) que destacan que estos avances tecnológicos afectan negativamente a las habilidades de lectoescritura de los jóvenes. Más de la mitad (52%), también se encuentran preocupados por el posible impacto que la IAG pueda tener en la capacidad crítica del alumnado.

Más allá de la percepción del cuerpo docente, lo que sí parece claro a día de hoy es que la expansión del uso de la IAG en el ámbito universitario plantea transformaciones significativas en los modos de producción y circulación del conocimiento. Este fenómeno desafía, entre otras realidades, los marcos institucionales y éticos que sustentan la estructura de la educación superior. Lejos de constituir una herramienta más dentro del repertorio tecnológico disponible para la docencia y la investigación, la IAG se configura como un agente de transformación que amenaza con poner en jaque a nivel práctico gran parte de los procesos de enseñanza, aprendizaje y legitimación del saber. Su incorporación progresiva en los espacios de educación superior obliga a reconsiderar el papel y la responsabilidad de la universidad en un contexto social cada vez más mediado por tecnologías inteligentes y automatismos de creciente complejidad.

Autores como Selwyn (2022) tratando de situar la inteligencia artificial (IA) y las tecnologías basadas en datos como fenómenos que exceden el plano meramente instrumental dentro del ámbito educativo. Selwyn advierte que la expansión de la inteligencia artificial generativa y otros sistemas automatizados no puede entenderse al margen de sus implicaciones sociales, políticas y ecológicas, subrayando la necesidad de adoptar una mirada crítica hacia los discursos tecnoutópicos que acompañan su desarrollo. Desde esta perspectiva, Selwyn propone atender a los efectos estructurales de la IA en la educación —en la configuración del conocimiento, la vigilancia, la desigualdad y la deshumanización de los procesos pedagógicos— y recuerda que su implementación es, en última instancia, un acto político que refleja determinados intereses y valores institucionales.

Por su parte, Knox (2023) profundiza en esta dimensión política al cuestionar los enfoques éticos dominantes sobre las tecnologías educativas, centrados en principios abstractos de transparencia o equidad. Frente a estos marcos descontextualizados, propone una “repolitización” del debate sobre la IA educativa mediante formas de participación radical que involucren a docentes, estudiantes y comunidades en la definición de los criterios éticos y en la gobernanza tecnológica. De este modo, ambos autores coinciden en que el papel de la universidad ante la inteligencia artificial no debe limitarse a la adopción técnica o normativa de estas herramientas, sino orientarse hacia un proceso reflexivo y deliberativo que considere sus implicaciones sociales y su potencial para transformar, de manera desigual, las relaciones de poder y los modos de producción del conocimiento.

Desde perspectivas claramente críticas, ambos autores coinciden en la necesidad de implementar procesos reflexivos frente a la incorporación de la inteligencia artificial en la educación. Si bien no toda innovación tecnológica transforma de manera profunda los procesos formativos; sin embargo, algunas son capaces de redefinir las prácticas, los valores y las relaciones de poder que estructuran el propio sistema educativo. En este sentido, la inteligencia artificial generativa (IAG) —como anteriormente fue el caso de la invención del libro o de internet— no puede entenderse como una mera herramienta, sino como un agente con capacidad de alterar los equilibrios establecidos en la educación superior. Su potencial para modificar el statu quo invita a abordarla, siguiendo a Latour (2005), como un actante no humano: un ente —humano o no humano— capaz de producir una diferencia sustancial dentro de una red de relaciones. Desde esta perspectiva, la IAG se presenta como un artefacto híbrido, tecnológico pero configurado y constantemente entrenado mediante interacciones humanas, que posee la capacidad de transformar de manera significativa la red de relaciones que conforma el sistema universitario.

Desde el punto de vista del trabajo investigador, la IAG se presenta como una ayuda inestimable para las tareas más mecánicas. Quienes investigan reconocen su utilidad para la transcripción de entrevistas, la búsqueda bibliográfica, la revisión ortotipográfica o de estilos y formatos. Estas aplicaciones, lejos de suponer una amenaza a la calidad de la investigación, pueden entenderse como una acción que permite economizar el tiempo y concentrarse en las dimensiones analíticas del trabajo investigativo. Además, y aunque aún en fase de desarrollo y perfeccionamiento, cada vez es más habitual que la IAG se utilice para asistir el proceso de escritura académica (Khalifa y Albadawy, 2024), hacer las revisiones por pares en revistas de alto impacto (Kousha, 2024), o incluso para realizar análisis de datos —tanto cualitativos (Odden et al., 2024) como cuantitativos (Garg et al., 2018)— mediante *embeddings*. Los recientes resultados y aplicaciones resultan tan interesantes que parecen asegurar que el debate sobre el impacto de la IAG no puede centrarse únicamente en cuestiones como el “plagio académico”, sino que debemos empezar a pensar en sus posibilidades, límites e implicaciones desde una perspectiva más amplia.

Ahora bien, en el contexto de la educación universitaria stricto sensu, que es la cuestión sobre la que queremos que se centre este texto, la expansión de la IAG introduce una nueva línea de tensión que invita a repensar los modos de aprender y evaluar. Esta nueva presencia actúa como un espejo que revela las limitaciones de los modelos pedagógicos tradicionales y la necesidad de transformarlos. Incorporar la IAG de manera crítica implica reconocer su potencial para ampliar las formas de pensamiento, expresión y acompañamiento académico, pero también afrontar las cuestiones éticas y epistemológicas que suscita en torno a las formas de evaluación y las dinámicas de aprendizaje.

## 2. RETOS DE LA EDUCACIÓN UNIVERSITARIA EN LA ERA DE LA IAG

Sin embargo, más allá del pánico moral que parece cundir entre los educadores ante la irrupción de la inteligencia artificial generativa en los procesos educativos, resulta necesario analizar cuáles de los retos que plantea son verdaderamente novedosos y cuáles constituyen, en realidad, reformulaciones de problemas clásicos del sistema educativo. A bote pronto, desde el punto de vista del llamado plagio, nada parece indicar que nos encontremos ante fenómenos completamente nuevos. Históricamente, las universidades han convivido con la picaresca de una parte del alumnado que busca sortear las pruebas evaluativas vulnerando su propósito pedagógico. El plagio en trabajos textuales —especialmente en el momento de la incorporación masiva de los ordenadores con acceso a internet—, las diferentes estrategias desarrolladas para copiar en los exámenes o incluso la subcontratación de trabajos de fin de grado, máster o tesis doctorales ya generaban debates similares a los actuales. Con aquel fenómeno surgieron también los primeros sistemas antiplagio que fueron adoptados por las instituciones educativas. Entre estos sistemas destaca Turnitin, herramienta adoptada por numerosas universidades y contratada por la propia Universidad Complutense de Madrid, que ha destinado sustanciales recursos económicos a su implantación y mantenimiento en un intento de subcontratar la vigilancia y garantizar una supuesta integridad del sistema de evaluación.

Lo novedoso de la IAG no es la posibilidad del fraude, sino su potencial masificación debido a su exigua barrera de entrada. Hoy en día, gran parte de las IAG pueden usarse de forma gratuita si bien con limitaciones en sus funcionalidades, por lo que cualquier estudiante puede generar textos coherentes, verosímiles y adaptados al nivel esperado sobre las temáticas requeridas mediante una mínima interacción con un *chatbot*. Utilizando *prompts* no demasiado elaborados, es relativamente sencillo obtener resultados textuales aceptables en un tiempo muy reducido y con un esfuerzo que esté muy por debajo del requerido por el evaluador. Esta democratización del potencial “fraude” es la que obliga a la universidad a repensar su función: si la adquisición de conocimiento no puede ser realmente verificada mediante trabajos y entregas de ejercicios (el método mayoritario que ha venido paulatinamente sustituyendo el peso de los exámenes tradicionales), es posible que a medio-largo plazo, las credenciales mismas (los títulos) tiendan a perder valor simbólico y social.

Por añadidura, la irrupción de la inteligencia artificial generativa introduce un nuevo elemento de desigualdad dentro del aula universitaria. En la actualidad, algunos estudiantes utilizan estas herramientas y otros no, pero si se atiende únicamente al resultado final de los trabajos, los recursos disponibles no permiten distinguir con fiabilidad cuándo un texto ha sido generado o ampliamente asistido por una IAG. Esto genera una situación paradójica: es posible que las calificaciones más altas estén recayendo precisamente en quienes emplean la inteligencia artificial de forma más eficaz, o incluso en quienes pueden acceder a versiones de pago con mayores capacidades. De este modo, la IAG no solo desafía los criterios de evaluación tradicionales, sino que amenaza con erosionar el principio de supuesta justicia meritocrática sobre el que se asienta el sistema de calificación universitario. En última instancia, se abre una nueva brecha digital que puede traducirse en desigualdades educativas profundas entre el alumnado.

En busca de devolver integridad de los procesos de evaluación, el personal docente (o al menos aquel que ha tomado conciencia del potencial disruptivo de la IAG) se enfrenta a un dilema, pues debe decidir entre confiar a ciegas en la autoría del estudiante o asumir, con cierto cinismo, que la mayor parte del trabajo ha sido realizada por la IA. En ambos casos, su autoridad docente se ve erosionada y, lo que es más importante, se pone en cuestión si el aprendizaje se ha llevado a cabo de manera efectiva. Sin embargo, la respuesta institucional es tímida. En casos anteriores recientes de irrupción de tecnologías disruptivas (como la difusión de internet) las universidades han confiado en soluciones técnicas como los mencionados detectores de plagio. Sin embargo, en el presente caso, los programas de *software* antiplagio parecen incapaces de discernir con fiabilidad la autoría de los textos generados por IAG. Todo indica que, al no poder recurrir a las prácticas punitivistas, ya no se puede rehuir el hecho de que la cuestión debe ser abordada con profundidad estructural.

Este problema educativo no surge en el vacío. Desde la implantación del Espacio Europeo de Educación Superior (EEES), el modelo de Bolonia ha promovido una transición desde el modelo de clase magistral hacia la enseñanza basada en competencias, prácticas y evaluación continua (EC, 2024). Pero esta transformación no ha venido acompañada de los recursos necesarios para ponerla en marcha, sino más bien todo lo contrario. Los ratios de alumnado crecientes —que en los primeros cursos de algunos grados introducen a más de 80 estudiantes por aula— dificultan la personalización del aprendizaje y reducen la docencia a la gestión de tareas masivas. En este contexto, la IAG se convierte en una amenaza estructural porque revela el carácter insostenible del modelo. En otras palabras, las pruebas realizadas en diferido —como ensayos o ejercicios que el alumnado elabora fuera del aula y entrega en línea— se han vuelto fácilmente automatizables mediante el uso de inteligencia artificial generativa. Este hecho reduce de manera significativa su valor formativo y evaluador, al posibilitar la producción de respuestas sin que medie necesariamente un proceso genuino de reflexión o aprendizaje. Podría decirse que la universidad ha actuado como la “rana de Chomsky”, que se adapta sin resistencia a una temperatura creciente hasta que es demasiado tarde para reaccionar. La incorporación acrítica de nuevas metodologías careciendo de los recursos necesarios para llevarlas a cabo ha debilitado la capacidad de respuesta del sistema ante disrupciones como la que presenta hoy en día la IAG.

Llevando al límite la teoría del actor-red (Latour, 2005; Callon, 1986), la masificación del potencial “fraude” que comentábamos previamente, la IAG podría incluso concebirse como un “super-actante”, un agente no humano con la capacidad de reorganizar de forma profunda las lógicas de acción de los procesos educativos universitarios. Frente a este fenómeno, nos encontramos con respuestas diversas. Una parte del personal docente aboga por un retorno a formas de evaluación tradicionales compuestas por exámenes presenciales, defensas orales o ejercicios realizados en clase. Paradójicamente, la transformación digital de la educación —independientemente de su dotación e implicaciones— impulsada durante las últimas dos décadas, se ve ahora invertida por la necesidad de garantizar que las competencias son efectivamente adquiridas.

De nuevo, el problema de fondo no es tecnológico, sino estructural. Mientras la universidad pública mantenga ratios elevados de estudiantes por clase y continúe sin valorar debidamente la función docente (siempre a la sombra del reconocimiento de la

función investigadora), seguiremos sin disponer de las condiciones necesarias para acompañar realmente a cada estudiante en su proceso de aprendizaje. La docencia personalizada —la que parece la única estrategia docente eficaz frente al uso indebido de la IAG— exige tiempo, estabilidad laboral y valoración institucional de la función docente.

Por ello, la respuesta no puede centrarse en prohibir o detectar el uso de IA, sino que debe ser mucho más ambiciosa. Es necesario volver el rostro al pasado, pero quizás no para retomar métodos de evaluación más o menos arcaicos (exámenes en el aula o pruebas orales), sino para recuperar reivindicaciones educativas históricas que parecen haberse abandonado, pero que siguen siendo centrales. La incorporación de la IAG en las aulas requiere una revalorización de la labor docente y una actualización de los sistemas de evaluación hacia la personalización individualizada que atienda más al proceso de aprendizaje que al resultado empaquetado como producto (trabajo, ensayo, etc.) donde cada vez es más sencillo llegar a la meta sorteando el proceso.

Sin embargo, lejos de negar la inevitable realidad, es necesario asumir que la IAG esta va a ser utilizada de forma cotidiana por el alumnado. Nuestra propuesta pasaría por permitir que se integre como un asistente activo tanto de tareas mecánicas como en aquellas tareas más complejas y que requieran de una reflexión más profunda. Pero es importante recalcar que esa función asistencial no debe convertirse en sustituto del pensamiento propio, pues acabaría atrofiando la adquisición de competencias, algunas tan básicas como la propia lecto-escritura, o la capacidad crítica y, por tanto, pervirtiendo la propia lógica del sistema educativo. Para garantizar este proceso y velar por él, necesitamos poder realizar tutorías personalizadas y rutinarias, un seguimiento continuo de cada estudiante y disponer de herramientas pedagógicas dialógicas en las que el personal docente pueda mediar —ser guía, pero también centinela— del proceso de aprendizaje. El personal docente debe conocer las herramientas de IAG para enseñar a usarlas críticamente, guiando a sus estudiantes hacia un aprendizaje sustantivo y velando a la vez por que estos no caigan en la tentación del “camino fácil”.

Si la universidad pública continúa operando en un marco de precariedad y masificación, la automatización de procesos que introduce la IAG no será un apoyo, sino una sustitución. La historia reciente de profesiones afectadas por la IA —como la programación o el diseño gráfico— muestra que la sustitución comienza cuando las tareas humanas son rutinarias y despersonalizadas. La educación, sin embargo, no puede reducirse a una secuencia de tareas: es relación, acompañamiento y juicio crítico. En ese sentido, la universidad tiene aún la posibilidad de marcar la diferencia, pero solo si asume el desafío con visión política y apuesta por una históricamente necesaria reforma estructural.

### 3. CONCLUSIONES Y ALGUNAS PROPUESTAS

La inteligencia artificial generativa constituye un punto de inflexión para la educación universitaria. Más que como una simple amenaza, actúa como un espejo que refleja las contradicciones internas del sistema: la sobrecarga docente, la infravaloración de la enseñanza y la dependencia de modelos de evaluación obsoletos. Ante este escenario, la universidad debe recuperar su misión educativa y apostar por una pedagogía crítica de lo digital que no excluya a la tecnología, pero que tampoco se deje arrollar por ella.

Frente a las posturas que defienden un retorno a modelos pedagógicos anteriores —como volver a poner todo el peso en los exámenes escritos presenciales al estilo “oposición”— resulta necesario reconocer que cualquier intento de desterrar la inteligencia artificial generativa del proceso educativo está, de antemano, condenado al fracaso. La IAG no es una herramienta confinada al ámbito académico, sino un actante cada vez más presente en el entramado social cotidiano. Su influencia se extiende a las dinámicas laborales, comunicativas, artísticas y culturales, configurando un nuevo ecosistema de producción y circulación del conocimiento. En este contexto, la universidad no puede ni debe dar la espalda a esta irrupción disruptiva, pues hacerlo equivaldría a renunciar a su función social como espacio de mediación crítica entre la propia sociedad y el cambio tecnológico.

Lo que debe promoverse, como ya hemos esbozado, es una incorporación consciente, crítica y guiada de esta tecnología, esto es, una apropiación crítica de sus funciones. Esta apropiación necesita de una formación obligatoria tanto en la naturaleza de la IAG (entendimiento de los procesos) como en sus usos específicamente educativos. Esta formación debe dirigirse al personal docente que, independientemente de si decide incorporar o no la IAG en su docencia, debe conocer su funcionamiento y posibilidades. Pero, además, debe también destinarse al propio cuerpo estudiantil, quiénes deberían recibir capacitación en el uso avanzado y ético de la IAG como herramienta que le permita gestionar sus procesos de aprendizaje o delegar ciertas cuestiones mecánicas y tediosas.

Sin embargo, el sistema educativo universitario debería garantizar, a la par que una familiarización crítica con la IAG, competencias en escritura académica, lectura crítica y demás habilidades predigitales al margen de la IAG (talleres en aula, sin mediación digital). Adquirir ambas habilidades: trabajar con IAG y sin ella, resulta fundamental para garantizar el nivel de competencias que un/a graduado/a universitario/a debería manejar. Estos conocimientos como la escritura académica o la lectura crítica han sido los grandes olvidados de la estructura universitaria (Becker, 2007) tratadas frecuentemente como competencias aledañas al proceso de aprendizaje, pero nunca enseñadas y evaluadas específicamente. A su vez, según la sociología de la educación (Bourdieu, 1985) las competencias en expresión escrita han sido los grandes elementos de distinción y desigualdad en la evaluación del alumnado. Una apuesta fuerte por separar las competencias digitales y las predigitales e incluir a ambas con sus métodos específicos de evaluación pueden prevenir que el uso competente de la IAG fagocite procesos de aprendizaje también necesarios que podrían ser suplantados por lógicas mecánicas. El abordaje del problema de la IAG en el ámbito de la educación superior demanda, por lo tanto, una estrategia dual y equilibrada. Resulta imprescindible que el sistema universitario garantice, simultáneamente a la alfabetización crítica en IAG, así como la consolidación de competencias fundamentales de la era predigital, tales como la escritura académica, la lectura crítica y la comprensión analítica. Para asegurar que un graduado universitario maneje un nivel de competencias robusto y completo, es crucial que adquiera y demuestre destreza en ambos planos: operar eficientemente con herramientas de IAG y ejecutar tareas complejas sin su mediación. Esto implica la implementación de espa-

cios formativos específicos, como talleres presenciales y sin mediación digital, dedicados exclusivamente al desarrollo de estas habilidades, y de carácter formal y obligatorio.

Es imprescindible formar a profesores y estudiantes en el uso de la IAG con una visión equilibrada que fomente perspectivas que no renuncien a su potencial. Sin embargo, también hay que evitar que el uso de estas nuevas tecnologías digitales afecte negativamente a las capacidades cognitivas de las nuevas generaciones de estudiantes, en riesgo de atrofiarse ante la automatización de pensamiento y ejecución. Para ello, se requieren políticas públicas que reconozcan la importancia del tiempo docente, la personalización del aprendizaje y la centralidad del acompañamiento pedagógico, así como dotaciones presupuestarias que permitan a la educación de todos implementar de una vez por todas los cambios pedagógicos necesarios. También es imprescindible que, a la par que se integran competencias sobre el uso de la IAG, se recuperen espacios educativos libres de mediaciones digitales, que permitan trabajar aquello que no se trabaja en los espacios mediados digitalmente. Sólo atendiendo a estas necesarias reformas podrá preservarse el valor credencial de los títulos universitarios, así como el sentido humanista que debe seguir guiando la educación superior ante esta nueva disrupción tecnológica. No aprovechar la oportunidad de una incorporación consensuada, consciente y crítica de la IAG, puede producir redundar en notables desigualdades educativas, así como provocar que la propia institución se vea superada por su efecto disruptor.

#### 4. REFERENCIAS

- Becker, H. S. (2007). *Writing for social scientists: How to start and finish your thesis, book, or article* (2nd ed.). University of Chicago Press.
- Bourdieu, P. (1985). *¿Qué significa hablar? Economía de los intercambios lingüísticos*. Madrid: Akal.
- Callon, M. (1986). Some elements of a sociology of translation: Domestication of the scallops and the fishermen of St Brieuc Bay. In J. Law (Ed.), *Power, action and belief: A new sociology of knowledge?* (pp. 196–223). Routledge.
- European Commission / European Education and Culture Executive Agency. (2024). *The European Higher Education Area in 2024: Bologna Process Implementation Report*. Luxembourg: Publications Office of the European Union.
- Fundación Cotec para la Innovación. (2025). *IA y educación*. Cotec. <https://cotec.es/proyectos-cpt/ia-y-educacion/>
- Garg, N., Schiebinger, L., Jurafsky, D., & Zou, J. (2018). *Word embeddings quantify 100 years of gender and ethnic stereotypes*. *Proceedings of the National Academy of Sciences*, 115(16), E3635–E3644. <https://doi.org/10.1073/pnas.1720347115>
- Hernández-Ramírez, R., & Ferreira, J. B. (2024). The Future End of Design Work: A Critical Overview of Managerialism, Generative AI, and the Nature of Knowledge Work, and Why Craft Remains Relevant. *She Ji*, 10(4), 414–440. <https://doi.org/10.1016/j.sheji.2024.11.002>
- Khalifa, M., & Albadawy, M. (2024). *Using artificial intelligence in academic writing and research: An essential productivity tool*. *Computer Methods and Programs in Biomedicine Update*, 5, Article 100145. <https://doi.org/10.1016/j.cmpbup.2024.100145>
- Knox, J. (2023). *AI and Education: Critical Perspectives on Artificial Intelligence in Learning*. Routledge.
- Kousha, K., & Thelwall, M. (2024). *Artificial intelligence to support publishing and peer review: A summary and review*. *Learned Publishing*, 37(1), 4–12. <https://doi.org/10.1002/leap.1570>
- Latour, B. (2005). *Reassembling the social: An introduction to actor-network-theory*. Oxford University Press.
- Odden, T. O. B., Tyseng, H., Mjaaland, J. T., Kreutzer, M. F., & Malthe-Sørenssen, A. (2024). *Using text embeddings for deductive qualitative research at scale in physics education*. *Physical Review Physics Education Research*, 20, 020151. <https://doi.org/10.1103/PhysRevPhysEducRes.20.020151>
- Selwyn, N. (2022). *Should robots replace teachers? AI and the future of education*. Polity Press.
- Williamson, B. (2020). *Education governance and datafication*. Routledge.



# Inteligencia Artificial en la universidad: percepciones del profesorado sobre su impacto en las competencias investigadoras y docentes

## Artificial Intelligence in Higher Education: University Teachers' Perceptions of Its Impact on Research and Teaching Competences

Yaritza Garcés-Delgado  
Arminda Suárez-Perdomo  
Carmen Nuria Arvelo-Rosales  
*Universidad de La Laguna*

### PALABRAS CLAVE:

Inteligencia artificial  
Educación superior  
Competencias docentes  
Brecha digital  
Profesorado universitario

### RESUMEN:

El objetivo de este estudio fue analizar las competencias docentes e investigadoras en el uso de la Inteligencia Artificial (IA) en el profesorado universitario, considerando el nivel de conocimiento, la frecuencia de uso y factores sociodemográficos como género, área de conocimiento y tipo de universidad. Se llevó a cabo un estudio cuantitativo transversal con 564 docentes de 30 universidades españolas, utilizando la Escala de Competencias Investigadoras y Docentes en IA (CID-IA), compuesta por seis dimensiones. Los resultados evidencian asociaciones positivas entre competencias, nivel de conocimiento y frecuencia de uso, mientras que la ansiedad se relaciona de forma negativa. Además, se identificaron diferencias significativas según los factores sociodemográficos. Estos hallazgos confirman la utilidad de la CID-IA como instrumento diagnóstico y destacan la necesidad de programas formativos y políticas institucionales que promuevan una integración crítica, responsable y consecuente de la IA en la educación superior.

### KEYWORDS:

Inteligência artificial  
Educação superior  
Competências docentes  
Divisão digital  
Docentes universitários

### RESUMO:

O objetivo deste estudo foi analisar as competências docentes e investigadoras no uso da Inteligência Artificial (IA) entre o professorado universitário, considerando o nível de conhecimento, a frequência de uso e fatores sociodemográficos como gênero, área de conhecimento e tipo de universidade. Foi realizado um estudo quantitativo transversal com 564 docentes de 30 universidades espanholas, utilizando a Escala de Competências Docentes e Investigadoras em IA (CID-IA), composta por seis dimensões. Os resultados evidenciam associações positivas entre competências, nível de conhecimento e frequência de uso, enquanto a ansiedade se relaciona de forma negativa. Além disso, foram identificadas diferenças significativas de acordo com os fatores sociodemográficos. Esses achados confirmam a utilidade da CID-IA como instrumento diagnóstico e destacam a necessidade de programas formativos e políticas institucionais que promovam uma integração crítica, responsável e consequente da IA no ensino superior.

### PALAVRAS-CHAVE:

Artificial intelligence  
Higher education  
Teaching competencies  
Digital divide  
University faculty

### ABSTRACT:

The aim of this study was to analyse university teachers' teaching and research competences in the use of Artificial Intelligence (AI), considering their level of knowledge, frequency of use, and sociodemographic factors such as gender, field of knowledge, and type of university. A cross-sectional quantitative study was conducted with 564 faculty members from 30 Spanish universities, using the Teaching and Research Competence in AI Scale (CID-IA), composed of six dimensions. The results show positive associations between competences, knowledge level, and frequency of use, while anxiety is negatively related. Moreover, significant differences were identified according to sociodemographic factors. These findings confirm the usefulness of the CID-IA as a diagnostic instrument and highlight the need for training programmes and institutional policies that promote a critical, responsible, and coherent integration of AI in higher education.

**CÓMO CITAR:** Garcés-Delgado, Y., Suárez-Perdomo, A., & Arvelo-Rosales, C. N. (2025). Inteligencia artificial en la universidad: Percepciones del profesorado sobre su impacto en las competencias investigadoras y docentes. *RETIS. Revista de Tecnología para la Inclusión Social*, 2(1), 59-68, DOI: 10.70664/retis.v2i1.007

## 1. INTRODUCCIÓN

La educación superior se encuentra en un punto de inflexión por la convivencia de distintos procesos de cambio. Entre ellos, la Inteligencia Artificial (IA) se presenta como uno de los fenómenos más complejos de los últimos tiempos. Lo que hasta hace poco se entendía como una tecnología emergente, vinculada sobre todo a la informática, se ha convertido en el foco de atención de la transformación digital (Spivakovsky et al., 2023). Así, la llegada de la IA generativa y la consolidación de la IA analítica han revolucionado la forma de enseñar, aprender e investigar (Bearman et al., 2023; Cedeño et al., 2024; Crompton & Burke, 2025; González-Campos et al., 2024). De esta forma, ya no se habla de potenciales aplicaciones de la IA al contexto educativo, sino de una integración inminente en la vida académica universitaria (Brandão et al., 2024).

Las potencialidades de la IA son indudables y están siendo ampliamente documentadas (Kuleto et al., 2021; Mohamed et al., 2025). Entre los beneficios más señalados se incluyen, por ejemplo, la optimización de tareas administrativas y repetitivas, el aumento de la eficiencia en la gestión de contenidos y recursos, la personalización de los procesos de aprendizaje para adaptarlos a ritmos y estilos individuales y el apoyo a la innovación metodológica en el aula (Karimi & Khawaja, 2023). Asimismo, destaca su impacto en la producción científica al facilitar búsquedas bibliográficas más específicas, sugerir marcos analíticos o incluso colaborar en la redacción de borradores de diseños metodológicos (Luckin, 2023; Yau et al., 2023). En este sentido, la IA no solo incrementa la productividad, sino que abre la puerta a formas novedosas de colaboración interdisciplinar.

Sin embargo, a pesar de todas las bondades descritas también lleva consigo una serie de tensiones y dilemas éticos. La incorporación de estas herramientas plantea interrogantes sobre la autoría académica, la propiedad intelectual, la originalidad de las producciones científicas, la privacidad y la fiabilidad (Goodin & Kariotis, 2021; Holmes et al., 2022a). Además, hay una creciente preocupación por la transparencia y explicación de los algoritmos, especialmente aquellos relacionados con el refuerzo de sesgos de género, clase o cultura presentes en los datos de entrenamiento (Flores & García-Peñalvo, 2023).

En este escenario, el docente ocupa una posición central. No basta con disponer de la tecnología, sino, también, con la capacidad de integrar pedagógicamente, crítica y reflexivamente la IA en las prácticas docentes e investigadoras. Cabe destacar que la tecnología, por sí sola, no garantiza aprendizajes significativos ni mejora los procesos investigadores. Siendo tarea del profesorado definir si las herramientas se convierten en un recurso enriquecedor o en un dispositivo superficial y poco trascendente (Danault, 2025; Salas-Pilco et al., 2023). De esta forma, comprender cómo el profesorado universitario se aproxima a la IA, qué usos realiza, qué barreras experimenta y qué apoyos demanda se convierte en una cuestión estratégica para entender la universidad de hoy (Chan & Tsi, 2023; Hormazabal, 2025).

La literatura reciente señala que las percepciones y competencias del profesorado en torno a la IA son muy heterogéneas. Mientras algunos docentes se sienten cómodos explorando nuevas herramientas y las integran con naturalidad en sus prácticas, otros muestran ansiedad tecnológica, desconfianza o falta de formación (Ayuso & Gutiérrez, 2022; Galindo-Domínguez et al., 2024; Ghimire et al., 2024; Yau et al., 2023). Lo que pone de manifiesto la necesidad de la alfabetización digital de los docentes para incluir con éxito el uso de la IA en su labor educativa (Davies et al., 2020; Popenici & Kerr, 2017). Esta diversidad refleja la necesidad de políticas institucionales de acompañamiento que reconozcan distintos puntos de partida y ofrezcan itinerarios formativos diferenciados. En esta línea, Chan y Tsi (2023) plantean el debate sobre si la IA en la educación superior constituye un asistente que amplía las posibilidades docentes o un elemento que amenaza con reemplazar funciones académicas.

A nivel internacional, se han desarrollado diversos marcos de referencia para orientar la adquisición de competencias en IA y digitalización en educación. Entre ellos, el marco DigCompEdu (Redecker & Punie, 2017) ha sido particularmente influyente al proponer un modelo progresivo de desarrollo docente en seis áreas y veintidós competencias, desde niveles básicos (A1) hasta niveles especializados (C2). Por su parte, la UNESCO (2019; 2022) ha publicado documentos clave que subrayan la importancia de garantizar no solo la adquisición de competencias técnicas, sino también una alfabetización crítica y ética que permita al profesorado guiar al estudiantado en un uso responsable de estas tecnologías. De manera complementaria, la Comisión Europea (2019) y la Unión Europea (2022) han emitido normativas que vinculan la competencia digital y el uso de la IA con políticas de equidad y con la necesidad de cumplir estándares de transparencia, fiabilidad y no discriminación. Estos marcos coinciden en la idea de que la IA en la universidad no puede reducirse a una cuestión de “saber usar la herramienta”, sino que implica, más bien, la formación de un profesorado capaz de comprender los fundamentos, valorar los riesgos, contextualizar los usos y formar en ciudadanía digital crítica.

Pese a la relevancia de estos marcos, algunos estudios siguen mostrando importantes limitaciones. Por un lado, la investigación evidencia que la adopción de la IA en la universidad se caracteriza por un conocimiento parcial y desigual, con diferencias que atraviesan género, áreas de conocimiento y tipo de institución (Baque et al., 2024; Chambi & Choquetarqui, 2024; Ruiz-Lázaro et al., 2025). Estas diferencias no solo reflejan realidades individuales, sino también la existencia de entornos institucionales desiguales, donde no siempre se proporcionan recursos, formación o acompañamiento adecuados. Por otro lado, la escasez de herramientas de medición ha limitado la posibilidad de obtener resultados consistentes y significativos (Meyerhofer-Parra & González-Martínez, 2024). Hasta hace poco, los estudios se basaban en percepciones generales o en indicadores poco consistentes. Este vacío ha limitado la posibilidad de realizar comparaciones fiables entre contextos universitarios, de identificar patrones de competencias o de diseñar políticas de desarrollo profesional sustentadas en evidencia (Salas-Pilco et al., 2023).

Con el propósito de responder a estas carencias, investigaciones recientes desarrollaron la Escala de Competencias Investigadoras y Docentes en el Uso de la IA (CID-IA), que constituye un avance significativo en este campo (Garcés-Delgado et al., en prensa). Esta escala permite evaluar de manera integral seis dimensiones: (1) habilidades digitales para la búsqueda y análisis de la información, (2) uso ético de la IA en investigación, (3) aplicación de la IA en la docencia, (4) ansiedad o resistencia hacia la IA, (5) intención de adopción de herramientas de IA y (6) disponibilidad de recursos institucionales. La validación psicométrica de la CID-IA supuso un primer paso necesario, pues aportó un instrumento con garantías de fiabilidad y validez para evaluar el constructo. Sin embargo, persistía la necesidad de analizar con mayor profundidad cómo se relacionan estas competencias con el uso real de las herramientas de IA, tanto a nivel de conocimiento como de frecuencia de uso. Del mismo modo, era imprescindible

dible explorar cómo estas competencias varían según el género, el área de conocimiento y el tipo de universidad, factores que la literatura ha mostrado como relevantes para comprender las desigualdades en la adopción tecnológica (Oliinyk, 2024; Lee, 2024).

La relevancia de CID-IA radica en que permite detectar qué condiciones personales e institucionales favorecen o limitan la integración de la IA en la práctica docente e investigadora. Al analizar estas relaciones, es posible identificar perfiles competenciales diferenciados, rastrear las brechas o desigualdades y, lo más importante, generar recomendaciones basadas en evidencia para la formación del profesorado y el diseño de políticas universitarias (Alvarado-Peña et al., 2025; Hormazabal, 2025). El presente estudio plantea, por tanto, los siguientes objetivos:

1. Analizar la relación entre los factores de la Escala de Competencias Investigadoras y Docentes en el Uso de IA (Habilidades digitales para la búsqueda y análisis de la información, Uso ético de la IA en investigación Aplicación de la IA en la docencia, Ansiedad o resistencia al uso de la IA, Intención de adopción de herramientas de IA y Disponibilidad de recursos institucionales para IA) con respecto al nivel de conocimiento y frecuencia de uso de herramientas de IA.

2. Analizar las diferencias significativas entre los factores de la escala con respecto al género, el tipo de universidad, el área de conocimiento y el uso de herramientas de IA, controlando el efecto del nivel de conocimiento y frecuencia de uso de herramientas de IA.

De esta forma, la investigación pretende aportar una comprensión más profunda del perfil competencial del profesorado universitario español en relación con la IA. Este mapa no es un fin en sí mismo, sino un medio para orientar programas de formación docente ajustados, diseñar estrategias institucionales que reduzcan desigualdades y favorecer políticas que promuevan una integración de la IA crítica, responsable y consecuente en la educación superior.

## 2. MÉTODO

### 2.1. PARTICIPANTES

Participaron 564 docentes de 30 universidades españolas, de los cuales el 96.1% pertenecían a universidades públicas y el 3.9% restante de universidades privadas. Además, el 52.5% eran mujeres y el 47.5% hombres en edades comprendidas entre los 22 y 73 años (M 46.81, DT 10.9). El profesorado tenía una experiencia docente e investigadora que iba entre 1 año y los 48 años (M 16.41, DT 11.4). El 31.4% pertenecían al área de Ciencias Sociales y Jurídicas, el 21.7% Ciencias de la Salud, el 17.6% Artes y Humanidades, el 17% Ciencias y el 12,2% restante Ingeniería y Arquitectura. Por último, el 89.2% indicaron que utilizaban herramientas de IA y el 10.8% no las solían utilizar.

### 2.2. INSTRUMENTO

Para recabar información se utilizó la Escala de Competencias Investigadoras y Docentes en el Uso de IA (CID-IA) elaborado por Garcés-Delgado et al. (en prensa). A continuación, se describen las principales características del instrumento:

1. Datos sociodemográficos (7 ítems) en los que se recoge información sobre el género, la edad, la universidad, tipo de universidad, años de experiencia, rama de conocimiento y uso de herramientas de IA.
2. Nivel de conocimiento de herramientas de IA (1 ítem) Escala Likert 1-7 en el que 1 es Nivel básico y 7 Nivel avanzado.
3. Frecuencia de uso de herramientas de IA (1 ítem) Escala Likert 1-7 en el que 1 es Nunca y 7 es Siempre.
4. Escala CID-IA (Garcés-Delgado et al., en prensa) (44 ítems,  $\omega = .91$ , escala Likert 1-7 en el que 1 es Totalmente en desacuerdo y 7 Totalmente de acuerdo) dividida en 6 factores:
  - 4.a. Habilidades digitales para la búsqueda y análisis de la información (8 ítems,  $\omega = .89$ ) en el que se analizaba la percepción del profesorado sobre su nivel de habilidades en la búsqueda y análisis de información digital (Utilizo la IA para optimizar la recopilación y clasificación de información, Utilizo la IA para mejorar la precisión y fiabilidad de mis análisis de datos).
  - 4.b. Uso ético de la IA en investigación (7 ítems,  $\omega = .80$ ) en el que se analizaba la percepción docente sobre la importancia del uso ético de la IA en la investigación (Me preocupó por citar los contenidos generados con IA en mis trabajos académicos, Me preocupó por las implicaciones éticas del uso de IA en la producción de conocimiento).
  - 4.c. Aplicación de la IA en la docencia (8 ítems,  $\omega = .90$ ) en el que se analizaba las posibilidades que tiene la aplicación de herramientas IA en la docencia universitaria (Uso la IA porque mejora la calidad del proceso de enseñanza-aprendizaje, Uso la IA para automatizar tareas de evaluación o retroalimentación).
  - 4.d. Ansiedad o resistencia al uso de la IA (5 ítems,  $\omega = .77$ ) en el que se analizaba la posible ansiedad y resistencia que experimentan los docentes en el uso de la IA en su actividad académica (Me preocupa que la IA reduzca mi autonomía en la toma de decisiones académicas, Me genera ansiedad aprender a manejar herramientas de IA para docencia).
  - 4.e. Intención de adopción de herramientas de IA (8 ítems,  $\omega = .95$ ) en el que se analizaba la percepción del profesorado sobre la posibilidad de adoptar herramientas de IA en su actividad académica (Me siento motivada/o para aprender sobre IA por su potencial en la educación superior, Me gustaría usar la IA en futuros proyectos académicos-científicos).
  - 4.f. Disponibilidad de recursos institucionales para IA (8 ítems,  $\omega = .92$ ) en el que se analizaba el nivel de conocimiento del profesorado sobre los recursos de que disponen para el uso de la IA (Considero que en mi universidad apoyan el desarrollo de proyectos de innovación educativa con IA, Considero que en mi universidad proporcionan herramientas de IA).

### 2.3. PROCEDIMIENTO

Previo a la difusión del cuestionario se obtuvo el visto bueno del Comité de Ética de la Investigación y de Bienestar Animal de la Universidad de las autoras (CEIBA2025-3581). Dicha difusión se realizó en dos momentos: (1) se procedió a una recogida de datos en una muestra piloto, contando para ello con profesorado de la Facultad de Educación de la Universidad de las autoras y (2) se procedió a difundir de forma masiva el cuestionario a nivel nacional. Para ello se contactó con la dirección y administración

de departamentos de 30 universidades españolas, aunque solo fueron 25 las que participaron en su difusión y cumplimentación entre su profesorado. Con la finalidad de cumplir con lo establecido en la Ley Orgánica 3/2018, de 5 de diciembre, de Protección de Datos Personales y garantía de los derechos digitales, se aplicaron los procedimientos éticos de la investigación solicitando el consentimiento informado y garantizando el anonimato de las personas participantes.

## 2.4. PLAN DE ANÁLISIS

Los análisis estadísticos se realizaron utilizando el paquete estadístico *Statistical Package for the Social Sciences* (SPSS) versión 29. Se realizó un análisis de valores perdidos basado en el test MCAR (Little, 1988), en el que se indica que cuando falta por completar el 10% de los ítems se debe excluir el caso del análisis; así como, cuando los valores perdidos eran <10% se comprueba que dichos valores faltan por azar. Seguidamente, se realizó un análisis descriptivo preliminar, con el objetivo de hallar las medias y desviaciones típicas de los factores analizados. Para responder al primer objetivo, se realizó un análisis de coeficiente de correlación de Pearson entre los factores de la Escala CID-IA (Garcés-Delgado et al., en prensa) con respecto al nivel de conocimiento de las herramientas de IA y la frecuencia de uso de las herramientas de IA.

Por último, para responder al segundo objetivo, se realizó un análisis multivariante de covarianza (MANCOVA), con los factores de la Escala con respecto al género, el tipo de universidad, la rama de conocimiento y el uso de IA, considerando como covariable el nivel de conocimiento de las herramientas de IA y la frecuencia de uso de las herramientas de IA, tomando como indicador del tamaño del efecto (ES) el estadístico  $\eta^2$  (Cohen y Swerdlik, 2006). Para su interpretación se adoptaron los puntos de corte habitualmente empleados en la investigación educativa: .01 indica un efecto pequeño, .06 un efecto medio y .14 un efecto grande (Cohen, 1988). Asimismo, se tuvieron en cuenta las recomendaciones de Lakens (2013) de contextualizar la magnitud de los efectos en función del diseño y la relevancia teórica de los hallazgos.

## 3. RESULTADOS

A continuación, se presentan los resultados respondiendo a los objetivos planteados.

### 3.1. CORRELACIONES ENTRE LA PERCEPCIÓN DE LAS COMPETENCIAS INVESTIGADORAS Y DOCENTES EN EL USO DE LA IA Y EL NIVEL DE CONOCIMIENTO Y FRECUENCIA DE USO

En la Tabla 1 se muestran los principales resultados de los análisis descriptivos realizados (media, desviación típica y correlaciones). Las medias variaron entre 2.40 y 5.22 y las desviaciones típicas entre .92 a 1.55. Con respecto a las correlaciones variaron entre -.195 (frecuencia de uso de la IA con ansiedad o resistencia al uso de la IA) a .661 (aplicación de la IA en la docencia con habilidades para la búsqueda y análisis de la información).

**Tabla 1. Medias, Desviaciones típicas y Correlaciones de Pearson**

	M	DT	1	2	3	4	5	6	7	8
1 Habilidades digitales para la búsqueda y análisis de la información	2.83	1.55	-							
2 Uso ético de la IA en investigación	5.22	1.30	.387***	-						
3 Aplicación de la IA en la docencia	2.78	1.46	.661***	.356***	-					
4 Ansiedad o resistencia al uso de la IA	3.33	1.38	-.035	.134**	-.018	-				
5 Intención de adopción de herramientas de IA	4.95	1.53	.472***	.355**	.543**	-.023	-			
6 Disponibilidad de recursos institucionales para IA	3.23	1.47	.299***	.280**	.302**	.086*	.218**	-		
7 Nivel de conocimiento de la IA	2.40	.92	.274***	.221**	.229**	-.161*	.047	.100*	-	
8 Frecuencia de uso de la IA	3.04	1.07	.637***	.410**	.626**	-.195**	.487**	.231**	.431**	-

Nota. \* < .001

Se observaron correlaciones positivas entre los factores de la escala en cuanto al nivel de conocimiento de la IA y la frecuencia de uso. No obstante, se observó una correlación negativa entre la ansiedad o resistencia al uso de la IA con respecto al nivel de conocimiento de la IA y su frecuencia de uso. Es decir, a más ansiedad o resistencia al uso de las herramientas de IA, menos nivel de conocimiento y menos frecuencia de uso. Sin embargo, este factor también mostró una correlación positiva con respecto al uso ético de la IA y la disponibilidad de recursos institucionales para la IA.

### 3.2. EFECTOS EN LA PERCEPCIÓN DE LAS COMPETENCIAS INVESTIGADORAS Y DOCENTES EN EL USO DE LA IA

Para responder al primer objetivo, se realizaron análisis multivariantes de covarianza (MANCOVA) entre los factores de la Escala CID-IA. En el caso del uso del nivel de conocimiento con covarianza se observaron diferencias significativas con respecto al género ( $F(6, 545) = 2.94, p < .01, \eta^2 = .10$ ), al tipo de universidad ( $F(6, 545) = 7.50, p < .001, \eta^2 = .10$ ), al área de conocimiento ( $F(6, 545) = 6.29, p < .001, \eta^2 = .10$ ), y a si utiliza IA ( $F(6, 545) = 10.63, p < .001, \eta^2 = .11$ ).

Como se muestra en la Tabla 2, se observaron diferencias significativas en el uso ético de la IA en investigación, siendo las mujeres las que mayor importancia le daban a este aspecto. Con respecto a la disponibilidad de recursos institucionales de herramientas de IA, fueron los hombres los que mostraron un mayor conocimiento de los recursos disponibles en sus respectivas universidades.

**Tabla 2. Test MANCOVA con la comparación entre los factores de la Escala CID-IA con respecto al género**

	Género		F	$\eta^2$
	Mujer	Hombre		
Habilidades digitales para la búsqueda y análisis de la información	2.76(1.5)	2.96(1.5)	.358	.00
Uso ético de la IA en investigación	5.37(1.1)	5.15(1.4)	5.80**	.02
Aplicación de la IA en la docencia	2.74(1.4)	2.88(1.5)	.542	.00
Ansiedad o resistencia al uso de la IA	3.31(1.3)	3.29(1.3)	.449	.00
Intención de adopción de herramientas de IA	5.05(1.4)	4.85(1.5)	1.81	.01
Disponibilidad de recursos institucionales para IA	3.18(1.4)	3.35(1.5)	4.13*	.02

Con relación con los análisis según el tipo de universidad, como se muestra en la Tabla 3, se observaron diferencias significativas en diversos factores. El profesorado de las universidades privadas percibía mayor habilidad en el uso de la IA, dándole mayor importancia a su uso ético y aplicándola en su docencia, asimismo, se observó que la disponibilidades de recursos institucionales para el profesorado de las universidades privadas era superior al de las universidades públicas participantes.

**Tabla 3. Test MANCOVA con la comparación entre los factores de la Escala CID-IA con respecto al tipo de universidad.**

	Tipo de universidad		F	$\eta^2$
	Pública	Privada		
Habilidades digitales para la búsqueda y análisis de la información	2.38(1.5)	3.57(1.4)	5.27*	.01
Uso ético de la IA en investigación	5.23(1.2)	5.92(1.0)	6.61**	.012
Aplicación de la IA en la docencia	2.78(1.4)	3.43(1.4)	4.38*	.01
Ansiedad o resistencia al uso de la IA	3.29(1.3)	3.61(1.39)	1.18	.00
Intención de adopción de herramientas de IA	4.95(1.5)	5.17(1.36)	.433	.00
Disponibilidad de recursos institucionales para IA	3.17(1.4)	5.15(1.4)	41.39***	.07

En el caso del área de conocimiento (Tabla 4), se hallaron diferencias significativas en todos los factores de la escala. Destacando la percepción de las habilidades para la búsqueda y análisis de la información, la aplicación de la IA en la docencia universitaria y en la intención de adoptar las herramientas de IA en las actividades académicas e investigadoras.

**Tabla 4. Test MANCOVA con la comparación entre los factores de la Escala CID-IA con respecto al área de conocimiento.**

	Área de conocimiento					F	$\eta p^2$
	CC	CCS	CCSS	AyH	II		
Habilidades digitales para la búsqueda y análisis de la información	2.41(1.2)	3.17(1.6)	3.08(1.5)	2.59(1.4)	2.80(1.5)	7.08***	.05
Uso ético de la IA en investigación	4.94(1.2)	5.03(1.4)	5.52(1.1)	5.41(1.2)	5.24(1.3)	4.64**	.03
Aplicación de la IA en la docencia	2.31(1.1)	3.02(1.5)	3.03(1.4)	2.61(1.5)	2.83(1.5)	5.96***	.04
Ansiedad o resistencia al uso de la IA	4.92(1.5)	3.46(1.3)	3.36(1.3)	3.50(1.4)	2.84(1.3)	2.39*	.02
Intención de adopción de herramientas de IA	4.92(1.5)	5.09(1.5)	5.26(1.3)	4.42(1.68)	4.75(1.5)	5.59***	.04
Disponibilidad de recursos institucionales para IA	3.10(1.5)	3.06(1.3)	3.53(1.5)	3.15(1.4)	3.20(1.4)	2.39*	.02

Nota. CC- Ciencias; CCS- Ciencias de la Salud; CCSS- Ciencias Sociales; AyH- Artes y Humanidades; II- Ingenierías

Por último, en la Tabla 5 se presentan los análisis obtenidos del profesorado que ya utilizaba la IA. Se observaron diferencias significativas en casi todos los factores de la escala. En concreto, destacan las habilidades digitales con las que cuentan los docentes que hacen uso habitual de la IA, el uso ético, la aplicación en la docencia y, sobre todo, la intención de adoptar las herramientas de IA en las actividades académicas e investigadoras.

**Tabla 5. Test MANCOVA con la comparación entre los factores de la Escala CID-IA con respecto al uso de herramientas de IA.**

	Utiliza IA		F	$\eta p^2$
	No	Si		
Habilidades digitales para la búsqueda y análisis de la información	1.71(1.2)	3.01(1.52)	30.86***	.05
Uso ético de la IA en investigación	4.57(1.6)	5.35(1.2)	15.13***	.03
Aplicación de la IA en la docencia	1.75(1.2)	2.93(1.4)	28.94***	.05
Ansiedad o resistencia al uso de la IA	3.48(1.4)	3.28(1.3)	.231	.00
Intención de adopción de herramientas de IA	3.62(1.8)	5.12(1.4)	55.62***	.09
Disponibilidad de recursos institucionales para IA	2.66(1.2)	3.32(1.4)	8.91**	.02

Seguidamente, se realizó un MANCOVA con las mismas variables usadas previamente, pero utilizando como covariable la frecuencia de uso de la IA. En este caso, se observaron diferencias significativas en cuanto al género ( $F(6, 545) = 3.46, p < .01, \eta p^2 = .03$ ), al tipo de universidad ( $F(6, 545) = 6.95, p < .001, \eta p^2 = .07$ ), al área de conocimiento ( $F(6, 545) = 5.92, p < .001, \eta p^2 = .10$ ), y a si utiliza IA ( $F(6, 545) = 2.42, p < .05, \eta p^2 = .03$ ).

Así, se encontraron diferencias significativas en función del género en diversos factores (Tabla 6). Específicamente, se observaron diferencias con respecto al uso ético de la IA en la investigación. Además, se hallaron diferencias significativas en los factores de intención de adopción de herramientas de IA y en la disponibilidad de herramientas de IA en sus respectivas universidades.

**Tabla 6. Test MANCOVA con la comparación entre los factores de la Escala CID-IA con respecto al género.**

	Género		F	$\eta^2$
	Mujer	Hombre		
Habilidades digitales para la búsqueda y análisis de la información	2.73(1.5)	2.93(1.5)	.481	.00
Factor Uso ético de la IA en investigación	5.34(1.1)	5.11(1.4)	5.79**	.02
Aplicación de la IA en la docencia	2.71(1.4)	2.86(4.5)	1.13	.00
Ansiedad o resistencia al uso de la IA	3.35(1.3)	3.30(1.3)	.306	.00
Intención de adopción de herramientas de IA	5.03(1.4)	4.87(1.5)	3.44*	.01
Disponibilidad de recursos institucionales para IA	3.17(1.4)	3.33(1.5)	4.44*	.02

Con respecto a la variable tipo de universidad, en la Tabla 7 se muestran las diferencias significativas encontradas en cuanto al uso ético de la IA en la investigación y, sobre todo, con respecto a la disponibilidad de herramientas IA en la universidad.

**Tabla 7. Test MANCOVA con la comparación entre los factores de la Escala CID-IA con respecto al tipo de universidad**

	Tipo de universidad		F	$\eta^2$
	Pública	Privada		
Habilidades digitales para la búsqueda y análisis de la información	2.80(1.5)	3.57(1.45)	3.26	.00
Factor Uso ético de la IA en investigación	5.20(1.3)	5.92(1.0)	4.77*	.01
Aplicación de la IA en la docencia	2.75(1.46)	3.43(1.4)	2.46	.00
Ansiedad o resistencia al uso de la IA	3.31(1.3)	3.61(1.3)	1.69	.00
Intención de adopción de herramientas de IA	4.94(1.5)	5.17(1.3)	.983	.00
Disponibilidad de recursos institucionales para IA	3.15(1.4)	5.15(1.4)	39.32***	.10

En el caso del área de conocimiento (Tabla 8), se observaron diferencias significativas en todos los factores de la escala, destacando las habilidades digitales para la búsqueda y el análisis de la información y la intención de adoptar herramientas de IA en las actividades académicas y de investigación.

**Tabla 8. Test MANCOVA con la comparación entre los factores de la Escala CID-IA con respecto al área de conocimiento**

	Área de conocimiento					F	$\eta^2$
	CC	CCS	CCSS	AyH	II		
Habilidades digitales para la búsqueda y análisis de la información	2.41(1.1)	3.11(1.6)	3.02(1.5)	2.59(1.4)	2.75(1.5)	5.19***	.04
Factor Uso ético de la IA en investigación	4.94(1.1)	5.02(1.4)	5.43(1.2)	5.41(1.1)	5.18(1.3)	4.34**	.03
Aplicación de la IA en la docencia	2.31(1.1)	2.97(1.5)	2.99(1.4)	2.61(1.5)	2.79(1.5)	4.96**	.03
Ansiedad o resistencia al uso de la IA	3.11(1.2)	3.53(1.4)	3.37(1.3)	3.50(1.4)	2.89(1.3)	3.17*	.02
Intención de adopción de herramientas de IA	4.92(1.5)	5.07(1.5)	5.24(1.3)	4.42(1.6)	4.76(1.5)	5.23***	.04
Disponibilidad de recursos institucionales para IA	3.10(1.5)	3.04(1.3)	3.50(1.5)	3.15(1.4)	3.18 (1.5)	2.10*	.02

Nota. CC- Ciencias; CCS- Ciencias de la Salud; CCSS- Ciencias Sociales; AyH- Artes y Humanidades; II- Ingenierías

Por último, en la Tabla 9 se presentan las diferencias significativas encontradas entre los docentes que usan la IA y los que no, en concreto en la intención de adoptar herramientas de IA en las actividades académicas y de investigación.

**Tabla 9. Test MANCOVA con la comparación entre los factores de la Escala CID-IA con respecto al uso de herramientas de IA.**

	Utiliza IA		F	np2
	No	Si		
Habilidades digitales para la búsqueda y análisis de la información	1.71(1.2)	2.97(1.5)	.061	.00
Factor Uso ético de la IA en investigación	4.57(1.6)	5.30(1.2)	.258	.00
Aplicación de la IA en la docencia	1.75(1.2)	2.90(1.4)	.912	.00
Ansiedad o resistencia al uso de la IA	3.48(1.4)	3.31(1.3)	.854	.00
Intención de adopción de herramientas de IA	3.62(1.8)	5.11(1.4)	11.50**	.02
Disponibilidad de recursos institucionales para IA	2.66(1.2)	3.30(1.4)	1.43	.00

#### 4. DISCUSIÓN Y CONCLUSIONES

En el presente estudio se observa como la integración de la IA en la universidad se constituye como un proceso heterogéneo, ya que depende de las competencias, actitudes y condiciones institucionales que definen los distintos perfiles del profesorado universitario. Así, se han encontrado relaciones positivas entre las dimensiones de la Escala CID-IA -habilidades digitales, uso ético, aplicación docente e intención de adopción- y las variables nivel de conocimiento y frecuencia de uso, mientras que la ansiedad o resistencia emerge como una de las principales barreras que dificultan la adopción o el uso de la IA. La literatura actual explica este patrón a partir del papel facilitador de la autoeficacia y la experiencia práctica, así como del efecto restrictivo de las emociones negativas y la percepción de amenaza (Ayuso & Gutiérrez, 2022; Bearman et al., 2023; Galindo-Domínguez et al., 2024; Luckin, 2023; Yau et al., 2023).

Los resultados mostraron diferencias significativas en las variables sociodemográficas, resaltando el género como un elemento importante, ya que las profesoras otorgaban mayor importancia al uso ético de la IA, mientras que los profesores enfatizaban más los aspectos técnicos y de aplicación. Este resultado coincide con investigaciones recientes que muestran aproximaciones distintas a la tecnología y refuerza la necesidad de considerar la perspectiva de género en el diseño de la formación docente (Baque et al., 2024; Chambi & Choquetarqui, 2024). El tipo de universidad también condiciona la integración de la IA en la docencia e investigación, determinando que las instituciones privadas ofrecen más recursos y apoyos, dejando patente que la disponibilidad institucional es determinante en el grado de integración de la IA (Flores & García-Peñalvo, 2023; Ruiz-Lázaro et al., 2025). Asimismo, las diferencias por área de conocimiento ponen de relieve que las ciencias e ingenierías presentan un uso más intensivo, mientras que en humanidades y ciencias sociales prevalece una actitud más crítica y cautelosa (Holmes et al., 2022b; Lee, 2024; Salas-Pilco et al., 2023).

Este estudio aporta una contribución original al utilizar por primera vez un instrumento integral para medir competencias docentes e investigadoras en IA en el contexto universitario español. A diferencia de trabajos previos centrados en percepciones generales o en análisis fraccionados, la Escala CID-IA permite una evaluación estructurada en seis dimensiones y proporciona un mapa fiable de fortalezas y debilidades del profesorado en relación con la IA (Garcés-Delgado et al., en prensa; Oliinyk, 2024).

Los hallazgos se alinean con los marcos internacionales vigentes. La relevancia concedida al uso ético coincide con las directrices de la UNESCO (2019, 2022), incidiendo en la formación en IA crítica y responsable como un derecho del profesorado y del estudiantado. La asociación entre las habilidades digitales y la frecuencia de uso guarda relación con la evolución competencial propuesta en el DigCompEdu (Redecker & Punie, 2017), mientras que las diferencias encontradas por género y tipo de institución concuerdan con las políticas de equidad digital promovidas por la Comisión Europea (2019) y la Unión Europea (2022). En conjunto, estos resultados complementan las agendas internacionales, evidenciando cómo estas directrices globales se materializan en la práctica universitaria (Danault, 2025).

Atendiendo a los objetivos descritos: a) analizar la relación entre los factores de la Escala de Competencias Investigadoras y Docentes en el Uso de IA con respecto al nivel de conocimiento y frecuencia de uso de herramientas de IA; y, b) analizar las diferencias significativas entre los factores de la escala con respecto al género, el tipo de universidad, el área de conocimiento y el uso de herramientas de IA, controlando el efecto del nivel de conocimiento y frecuencia de uso de herramientas de IA. El primero, relativo a la relación entre las dimensiones de la CID-IA, el nivel de conocimiento y la frecuencia de uso, queda evidenciado, a través de las relaciones significativas, la coherencia entre los factores de la escala y el nivel de conocimiento y frecuencia de uso. El segundo, orientado al análisis de diferencias por género, tipo de universidad, área de conocimiento y uso de herramientas de IA, también se cumplió, ya que se encontraron diferencias relevantes que aportan nuevas evidencias al campo de estudio. Investigaciones realizadas en América Latina, Europa y Asia han señalado problemáticas similares: desigualdad en el acceso, ansiedad tecnológica y necesidad de marcos éticos consistentes (Alvarado-Peña et al., 2025; Bearman et al., 2023; Crompton & Burke, 2025; Ghimire et al., 2024). En este sentido, los resultados contribuyen al debate existente y ofrecen evidencias aplicables a otros sistemas universitarios que buscan una integración crítica de la IA.

Sin embargo, el presente estudio no está exento de limitaciones que deben ser reseñadas. En primer lugar, el diseño transversal utilizado impide establecer relaciones causales, limitándose a describir asociaciones entre competencias y uso de la IA. En segundo lugar, el uso de cuestionarios autoinforme puede introducir sesgos como la deseabilidad social en la manera en que el profesorado declara sus prácticas y competencias observadas en el uso de la IA. Por último, aunque participaron 564 docentes

de 30 universidades, los hallazgos se limitan al contexto español y necesitan contrastarse con investigaciones en otros países para validar su alcance (Chan & Tsi, 2023).

De estas limitaciones se desprenden varias perspectivas de futuro: 1) realizar estudios longitudinales para observar la evolución de las competencias, 2) aplicar la Escala CID-IA en contextos internacionales y, 3) utilizar metodologías mixtas que combinen cuestionarios, entrevistas y observación de prácticas, permitiendo la triangulación de los datos. También sería útil incorporar indicadores objetivos de uso, como registros en plataformas o analíticas institucionales y explorar variables mediadoras como la autoeficacia o la confianza en la tecnología (Hormazabal, 2025).

En cuanto a las implicaciones prácticas, los resultados apuntan a la necesidad de programas de formación híbridos que combinen alfabetización básica en IA con talleres adaptados a cada disciplina. En humanidades y ciencias sociales se deberían priorizar aspectos como la autoría y la ética académica, mientras que en ingenierías convendría centrarse en aplicaciones técnicas avanzadas. Asimismo, se recomienda promover comunidades de prácticas interuniversitarias que permitan compartir experiencias, reducir la ansiedad y favorecer el aprendizaje conjunto (González-Campos et al., 2024). A nivel institucional, los hallazgos señalan la importancia de garantizar recursos tecnológicos básicos, acompañamiento pedagógico y protocolos explícitos sobre el uso de la IA en docencia e investigación.

En definitiva, en este estudio se observa un claro potencial, así como el riesgo de la popularización de la IA en la universidad. El profesorado es un actor clave, cuyas competencias y condiciones de trabajo determinarán si la IA se convierte en una oportunidad para la enseñanza y la investigación o, por el contrario, en un nuevo generador de desigualdades. Así, se evidencia que la adopción tecnológica depende tanto de factores personales como institucionales y culturales. Por lo tanto, para asegurar la integración crítica de la IA en educación superior se requieren programas formativos específicos, políticas institucionales de acompañamiento y estrategias de gestión alineadas con los marcos internacionales.

## 5. REFERENCIAS BIBLIOGRÁFICAS

- Alvarado-Peña, A., Rodríguez, M., & Sánchez, L. (2025). Una mirada desde la inteligencia artificial al desarrollo de competencias investigadoras en docentes universitarios latinoamericanos. *Revista Latinoamericana de Educación Superior*, 12(1), 45-63. <https://dialnet.unirioja.es/descarga/articulo/10278461.pdf>
- Ayuso, D., & Gutiérrez, P. (2022). La Inteligencia Artificial como recurso educativo durante la formación inicial del profesorado. *RIED-Revista Iberoamericana de Educación a Distancia*, 25(2), 347-362. <https://doi.org/10.5944/ried.25.2.32332>
- Baque, V. E., Zavala, M. I., Mendoza, V. P., Recalde, E. B., Nevares, M. de J., Castillo, N. E., & Barreto, W. W. (2024). Percepciones y experiencias de docentes universitarios sobre la inteligencia artificial: transformación, ética y desafíos en el uso académico por estudiantes. *LATAM Revista Latinoamericana de Ciencias Sociales y Humanidades*, 5(6), 2763-2773. <https://doi.org/10.56712/latam.v5i6.3204>
- Bearman, M., Ryan, J., & Ajjawi, R. (2023). Discourses of artificial intelligence in higher education: A critical literature review. *Higher Education*, 86, 369-385. <https://doi.org/10.1007/s10734-022-00937-2>
- Brandão, A., Pedro, L., & Zagalo, N. (2024). Teacher professional development for a future with generative artificial intelligence—an integrative literature review. *Digital Education Review*, (45), 151-157. <https://doi.org/10.1344/der.2024.45.151-157>
- Cedeño, J. G., Maitta, I. S., Vélez, M. L., & Palomeque, J. Y. (2024). Investigación universitaria con inteligencia artificial. *Revista Venezolana De Gerencia*, 29(106), 817-830. <https://doi.org/10.52080/rvgluz.29.106.23>
- Chambi, M. C., & Choquetarqui, C. M. (2024). Implementación de herramientas basadas en inteligencia artificial en el ámbito de la educación superior. *Educación Superior*, 11(1), 81-92. <https://doi.org/10.53287/ueay5969vp97x>
- Chan, R., & Tsi, W. (2023). The AI revolution in education: Will AI replace or assist teachers in higher education? *arXiv preprint*. <https://arxiv.org/abs/2305.01185>
- Cohen, J. (1988). *Statistical Power Analysis/or the Behavioral Sciences (Rev.Ed.)*. Academic Press.
- Cohen, R., & Swerdlik, M. (2006). *Pruebas y evaluación psicológicas: introducción a las pruebas y a la medición* (6a ed.). McGraw.
- Comisión Europea. (2019). *Directrices Éticas para una IA fiable*. Grupo de Expertos de Alto Nivel sobre IA. <https://goo.su/L37YU>
- Crompton, H., & Burke, D. (Eds.). (2025). *Artificial Intelligence Applications in Higher Education: Theories, Ethics, and Case Studies for Universities*. Routledge.
- Danault, R. R. (2025). Adapting to AI: The evolving role of faculty in higher education. *University of New Hampshire Scholars' Repository*. [https://scholars.unh.edu/cgi/viewcontent.cgi?article=3166&context=faculty\\_pubs](https://scholars.unh.edu/cgi/viewcontent.cgi?article=3166&context=faculty_pubs)
- Davies, H. C., Eynon, R., & Salveson, C. (2020). The Mobilisation of AI in Education: A Bourdieusean Field Analysis. *Sociology*, 55, 539-560. <https://doi.org/10.1177/0038038520967888>
- Flores, J. M., & García-Peñalvo, F. J. (2023). Reflexiones sobre la ética, potencialidades y retos de la Inteligencia Artificial en el marco de la Educación de calidad (ODS4). *Comunicar*, 31(74), 37-47. <https://doi.org/10.3916/C74-2023-03>
- Galindo-Domínguez, H., Delgado, N., Campo, L., & Losada, D. (2024). Relationship between teachers' digital competence and attitudes towards artificial intelligence in education. *International Journal of Educational Research*, 126(September 2023). <https://doi.org/10.1016/j.ijer.2024.102381>
- Garcés-Delgado, Y., Suárez-Perdomo, A., & Arvelo-Rosales, C. N. (in press). Validation of a scale on artificial intelligence in university faculty competences: Assessment of teaching and research competences. In J. Sánchez-Santamaría (Ed.), *Research skills in lifelong education*. IGI Global.
- Ghimire, B., Prather, J., & Edwards, S. H. (2024). Generative AI in education: A study of educators' awareness, sentiments, and influencing factors. *arXiv preprint*. <https://arxiv.org/abs/2403.15586>

- González -Campos, J., López - Núñez, J., & Araya - Pérez, C. (2024). Educación superior e inteligencia artificial: desafíos para la universidad del siglo XXI. *Aloma: Revista De Psicología, Ciències De l'Educació I De l'Esport*, 42(1), 79–90. <https://doi.org/10.51698/aloma.2024.42.1.79-90>
- Gooding, P., & Kariotis, T. (2021). Ethics and law in research on algorithmic and data-driven technology in mental health care: scoping review. *JMIR Mental Health*, 8(6), e24668. <https://doi.org/10.2196/24668>
- Holmes, W., Bialik, M., & Fadel, C. (2022a). *Artificial Intelligence in Education: Promises and Implications for Teaching and Learning*. Center for Curriculum Redesign.
- Holmes, W., Porayska-Pomsta, K., Holstein, K., Sutherland, E., & Rienties, B. (2022b). Ethics of AI in Education: Towards a Community-Wide Framework. *International Journal of Artificial Intelligence in Education*, 32, 504–526. <https://doi.org/10.1007/s40593-021-00239-1>
- Hormazabal, R. G. (2025). Sesgos en la IA y educación superior: tipologías, impactos y retos. *Revista de Estudios en Educación (REXE)*, 20(2), 155-172. <https://www.rexe.cl/index.php/rexe/article/view/3062>
- Karimi, H., & Khawaja, S. (2023). The Impact of Artificial Intelligence on Higher Education in England. *Creative Education*, 14, 2405-2415. <https://doi.org/10.4236/ce.2023.1412154>
- Kuleto, V., Ilić, M., Dumangiu, M., Ranković, M., Martins, O. M., Păun, D., & Mihoreanu, L. (2021). Exploring opportunities and challenges of artificial intelligence and machine learning in higher education institutions. *Sustainability*, 13(18), 10424. <https://doi.org/10.3390/su131810424>
- Lakens, D. (2013). Calculating and reporting effect sizes to facilitate cumulative science: a practical primer for *t*-tests and ANOVAs. *Frontiers in Psychology*, 4 (863). <https://doi.org/10.3389/fpsyg.2013.00863>
- Lee, D. (2024). The impact of generative AI on higher education learning and teaching: Perspectives from Australian universities. *Education and Information Technologies*, 29(3), 2225-2244. <https://www.sciencedirect.com/science/article/pii/S2666920X24000225>
- Ley Orgánica 3/2018, de 5 de diciembre, de Protección de Datos Personales y garantía de los derechos digitales.
- Little, R. J. A. (1988). A Test of Missing Completely at Random for Multivariate Data with Missing Values. *Journal of the American Statistical Association*, 83(404), 1198–1202. <https://doi.org/10.1080/01621459.1988.10478722>
- Luckin, R. (2023). *Handbook of Artificial Intelligence in Education*. Edward Elgar Publishing.
- Meyerhofer-Parra, R., & González-Martínez, J. (2024). Teachers' perceptions of digital competences and their use for digital well-being: A mixed analysis on the extension of the DigCompEdu framework. *EduTec*, 87, 115–133. <http://doi.org/10.21556/edutec.2024.87.2967>
- Mohamed, A. M., Shaaban, T. S., Bakry, S. H., Guillén-Gámez, F. D., & Strzelecki, A. (2025). Empowering the faculty of education students: Applying AI's potential for motivating and enhancing learning. *Innovative Higher Education*, 50(2), 587-609. <https://doi.org/10.1007/s10755-024-09747-z>
- Oliinyk, I. (2024). Inteligencia artificial en el desarrollo de competencias investigadoras de los futuros doctores. *Revista EduWeb*, 18(1), 88-102. <https://revistaeduweb.org/index.php/eduweb/article/view/654>
- Popenici, S. A. D., & Kerr, S. (2017). Exploring the impact of artificial intelligence on teaching and learning in higher education. *Research and Practice in Technology Enhanced Learning*, 12, 22. <https://doi.org/10.1186/s41039-017-0062-8>
- Redecker, C., & Punie, Y. (2017). *European framework for the digital competence of educators: DigCompEdu*. Luxembourg: Publications Office of the European Union. <https://doi.org/10.2760/178382>
- Ruiz-Lázaro, P., Fernández, J., & Martínez, C. (2025). Análisis de las guías de uso de inteligencia artificial en educación superior: comparación entre las universidades españolas. *Bordón. Revista de Pedagogía*, 77(1), 55-78. <https://recyt.fecyt.es/index.php/BORDON/article/view/110638>
- Salas-Pilco, S. Z., Yang, Y., & Zhang, Z. (2023). Teachers' perceptions of artificial intelligence in education: A systematic review of empirical studies. *Computers and Education: Artificial Intelligence*, 5, 100147.
- Spivakovsky, O. V., Omelchuk, S. A., Kobets, V. V., Valko, N. V., & Malchukova, D. S. (2023). Institutional policies on artificial intelligence in university learning, teaching and research. *Information Technologies and Learning Tools*, 97(5), 181. <https://doi.org/10.33407/itlt.v97i5.5395>
- UNESCO. (2019). *Beijing consensus on artificial intelligence and education*. <https://unesdoc.unesco.org/ark:/48223/pf0000368303>
- UNESCO. (2022). *Recomendación sobre la ética de la inteligencia artificial*. [https://unesdoc.unesco.org/ark:/48223/pf0000379920\\_spa.page=15](https://unesdoc.unesco.org/ark:/48223/pf0000379920_spa.page=15)
- Unión Europea. (2022). *Directrices éticas sobre el uso de la inteligencia artificial (IA) y los datos en la educación y formación para los educadores*. <https://data.europa.eu/doi/10.2766/898>
- Yau, K. W., Chai, C. S., Chiu, T. K. F., Meng, H., King, I., & Yam, Y. (2023). A phenomenographic approach on teacher conceptions of teaching artificial intelligence (AI) in K-12 schools. *Education and Information Technologies*, 28(1), 1041–1064. <https://doi.org/10.1007/s10639-022-11161-x>



Observatorio  
Brechas  
Digitales

[observatoriobrechasdigitales.org](http://observatoriobrechasdigitales.org)



Financiado por:



✓ POR SOLIDARIDAD  
OTROS FINES DE INTERÉS SOCIAL