

INTELIGENCIA ARTIFICIAL E INTEGRIDAD DE LA INFORMACIÓN

Perspectivas para
un uso responsable

Jonathan Hernández Pérez

Andrew Cox

COORDINADORES



La presente obra está bajo una licencia de:
<https://creativecommons.org/licenses/by-nc-sa/4.0/deed.es>



Atribución-NoComercial-CompartirIgual 4.0 Internacional (CC BY-NC-SA 4.0)

Este es un resumen legible por humanos (y no un sustituto) de la [licencia](#). [Advertencia](#).

Usted es libre de:

Compartir — copiar y redistribuir el material en cualquier medio o formato

Adaptar — remezclar, transformar y construir a partir del material

La licenciante no puede revocar estas libertades en tanto usted siga los términos de la licencia

Bajo los siguientes términos:



Atribución — Usted debe dar [crédito de manera adecuada](#), brindar un enlace a la licencia, e [indicar si se han realizado cambios](#). Puede hacerlo en cualquier forma razonable, pero no de forma tal que sugiera que usted o su uso tienen el apoyo de la licenciante.



NoComercial — Usted no puede hacer uso del material con [propósitos comerciales](#).



CompartirIgual — Si remezcla, transforma o crea a partir del material, debe distribuir su contribución bajo la [misma licencia](#) del original.

Inteligencia artificial e integridad de la información
Perspectivas para un uso responsable

COLECCIÓN
INFORMACIÓN Y SOCIEDAD
Instituto de Investigaciones Bibliotecológicas y de la Información

Inteligencia artificial e integridad de la información
Perspectivas para un uso responsable

Coordinadores

Jonathan Hernández Pérez
Andrew Cox



Universidad Nacional Autónoma de México
2025

**Catalogación en la fuente: Instituto de Investigaciones
Bibliotecológicas y de la Información.**

Nombres: Hernández Pérez, Jonathan, coordinador. | Cox,
Andrew, coordinador.

Título: Inteligencia artificial e integridad de la información :
perspectivas para un uso responsable / coordinadores
Jonathan Hernández Pérez, Andrew Cox.

Publicación: Primera edición. Ciudad de México :
Universidad Nacional Autónoma de México, Instituto de
Investigaciones Bibliotecológicas y de la Información, 2025.

Descripción: ix, 204 páginas.

Serie: Colección Información y Sociedad

ISBN-e: 978-607-642-424-7

Temas: Inteligencia artificial | Información | Ética de la
información | Alfabetización informacional | Integridad de la
información

Clasificación: ZA3073.5 I58

Diseño de cubierta: Itzia I. Solís González

Primera edición: diciembre de 2025

D.R. © UNIVERSIDAD NACIONAL

AUTÓNOMA DE MÉXICO

Instituto de Investigaciones Bibliotecológicas
y de la Información

Circuito Interior s/n, Torre II de Humanidades,
pisos 11, 12 y 13, Ciudad Universitaria, C. P. 04510,
Alcaldía Coyoacán, Ciudad de México

ISBN-e: 978-607-642-424-7

Esta edición y sus características son propiedad
de la Universidad Nacional Autónoma de
México. Prohibida la reproducción total o parcial
por cualquier medio sin la autorización escrita del
titular de los derechos patrimoniales.

Publicación dictaminada

Hecho en México

Contenido

Presentación	vii
Jonathan Hernández Pérez	
Andrew Cox	

FUNDAMENTOS Y DESAFÍOS DE LA INTELIGENCIA ARTIFICIAL EN EL ECOSISTEMA DIGITAL

ETHICAL AND SOCIETAL IMPLICATIONS OF ALGORITHMIC BIAS AND FAIRNESS: A CONCEPTUAL INQUIRY INTO ARTIFICIAL INTELLIGENCE	3
Lais Barbudo Carrasco	

BIAS AND ITS IMPACT ON INFORMATION FAIRNESS AND TRUST: A SYSTEMATIC REVIEW FROM AN AFRICAN PERSPECTIVE	27
Abton Muroyiwa	
Notice Pasipamire	

LA HUELLA DE CARBONO DE LOS ALGORITMOS INTELIGENTES DE LOS MACROCENTROS DE DATOS	49
Perla Olivia Rodríguez-Reséndiz	

LA IA, DESINFORMACIÓN Y EL ROL DE LOS ACTORES DEL CONOCIMIENTO

CUESTIONAMIENTOS DE LOS ACTORES DEL SECTOR EDITORIAL HACIA EL USO DE LA INTELIGENCIA ARTIFICIAL Y SUS HERRAMIENTAS	63
Jenny Teresita Guerra González	

AI AND ACADEMIC ETHICS IN SCIENTIFIC JOURNALS IN THE FIELDS OF ASTRONOMY AND MATHEMATICS	81
Suyin Ortega Cuevas	
Beatriz Juárez Santamaría	

LA DESINFORMACIÓN: UN RETO PARA INHIBIR SUS EFECTOS EN LA CIUDADANÍA Y EL PAPEL DEL BIBLIOTECARIO EN EL USO DE LA INFORMACIÓN EN REDES SOCIALES	99
---	----

Adriana Mata Puente
Eduardo Oliva Cruz

ALFABETIZACIÓN INFORMACIONAL, MEDIOS E INTELIGENCIA ARTIFICIAL: REFLEXIONES SOBRE LAS INTERRELACIONES ORIENTADAS A LA CONSTRUCCIÓN Y DIFUSIÓN DEL CONOCIMIENTO ACADÉMICO EN LA ERA DIGITAL	129
---	-----

Regina C. B. Belluzzo
Vânia C. P. N. Valente

MARCOS Y ESTRATEGIAS PARA UNA GOBERNANZA ÉTICA DE LA IA

LA EXPERIENCIA DE DISEÑAR E IMPARTIR UN CURSO/TALLER PARA PROFESORES UNIVERSITARIOS SOBRE INTELIGENCIA ARTIFICIAL Y PLAGIO ACADÉMICO	149
--	-----

Jesús Cortés-Vera
Ma. Lourdes Tiscareño
Thelma Jovita García

INTEGRIDAD DE LA INFORMACIÓN: LOS PRINCIPIOS DE LA ONU Y EL ECOSISTEMA DIGITAL	171
---	-----

Jonathan Hernández Pérez

INDICADORES GENERALES PARA LA INTEGRIDAD DE LA INFORMACIÓN EN LA PRESERVACIÓN DE FOTOGRAFÍAS DE PRENSA DIGITAL	189
---	-----

Yissel N. Carmenate
Marco A. Mariano

Presentación

El ecosistema informativo contemporáneo se encuentra en un punto de inflexión crítico. En un entorno saturado de información, la proliferación de desinformación, el discurso de odio y especialmente la creciente sobrecarga de información generada por la IA ha creado un entorno cada vez más nocivo y polarizado. Esta situación exige una respuesta inmediata y coordinada desde distintos frentes. Aspectos como el acelerado avance de la IA y su notable capacidad para crear y distribuir contenido han intensificado las inquietudes éticas y sociales. Frente a esta compleja realidad, el libro *Inteligencia Artificial e Integridad de la Información: Perspectivas para un uso responsable* se presenta como un aporte fundamental para afrontar tanto los desafíos como las oportunidades que la IA representa en el campo de la información.

La presente obra reúne 10 capítulos producto de una serie de reflexiones emanadas del Seminario Inteligencia Artificial e Integridad de la Información, como parte del Mexican Information Global Forum, llevado a cabo en agosto de 2024 en colaboración con el grupo de interés sobre Inteligencia Artificial de la Federación Internacional de Asociaciones e Instituciones Bibliotecarias (IFLA).

Inteligencia artificial...

Organizados en tres apartados temáticos: “Desafíos fundamentales de la inteligencia artificial en el ecosistema digital”, “La IA en el ecosistema del conocimiento y la información” y “Marcos y estrategias para una gobernanza ética de la IA”, las contribuciones problematizan la compleja interrelación entre la inteligencia artificial y la integridad de la información.

En el primer apartado, “Desafíos fundamentales de la inteligencia artificial en el ecosistema digital”, se establecen las bases conceptuales de los problemas inherentes a la IA. Se aborda la ética y las implicaciones sociales del sesgo algorítmico, examinando la forma en la que los sistemas de IA pueden perpetuar la discriminación, especialmente hacia comunidades marginadas. Asimismo, se explora la huella de carbono generada por la IA en los macrocentros de datos, que demandan grandes cantidades de energía y agua, lo que representa un costo medioambiental crítico en el contexto del cambio climático.

El segundo apartado, “La IA en el ecosistema del conocimiento y la información”, se centra en el impacto directo de la IA en la producción, difusión y acceso al conocimiento. Se examinan los cuestionamientos de diversos actores del sector editorial respecto al uso de la IA, incluyendo el uso no autorizado de obras protegidas por derechos de autor y la necesidad de regulación. También se profundiza en el fenómeno de la desinformación en redes sociales y el rol de los profesionales de la información en capacitar a la ciudadanía para validar contenidos. Finalmente, se reflexiona sobre las interrelaciones entre la alfabetización informacional, los medios y la inteligencia artificial, destacando su importancia para la construcción y difusión del conocimiento académico en la era digital.

El libro concluye con el tercer apartado, “Marcos y estrategias para una gobernanza ética de la IA”, que propone soluciones y marcos normativos para abordar los desafíos éticos. Se presenta la experiencia de un curso/taller para profesores universitarios, centrado en la IA y el plagio académico, ofreciendo estrategias para fomentar la integridad académica y el uso responsable de las herramientas de IA generativa. Se analiza el concepto de integridad de la información en un sentido más amplio, examinando los

Principios Globales para la Integridad de la Información propuestos por la ONU como un marco normativo para contrarrestar la desinformación y el discurso de odio en el entorno digital. Complementariamente, se presentan indicadores generales para la integridad de la información en la preservación de fotografías de prensa digital, detallando aspectos como la encriptación, controles de acceso y auditorías. Este último capítulo, si bien de manera específica, ofrece pautas fundamentales para la gestión confiable de la información en cualquier formato digital.

En conjunto, las reflexiones presentadas en esta obra buscan analizar y comprender los profundos impactos de la Inteligencia Artificial en nuestro campo disciplinar, ofreciendo un mapa para futuras líneas de investigación sobre las intersecciones entre inteligencia artificial e integridad de la información.¹

Jonathan Hernández Pérez
Andrew Cox

1 Agradecemos el apoyo de Esther Bravo, Ángel Guzmán y José Franco en la revisión de la estructura de los textos.

FUNDAMENTOS Y DESAFÍOS DE LA INTELIGENCIA ARTIFICIAL
EN EL ECOSISTEMA DIGITAL

Ethical and Societal Implications of Algorithmic Bias and Fairness: A Conceptual Inquiry into Artificial Intelligence

LAIS BARBUDO CARRASCO

*Department of Communication, Quality Management, and Information Systems,
Mid Sweden University (MIUN), Sweden*

INTRODUCTION

The widespread use of algorithms in decision-making across various sectors has raised important ethical and societal concerns, particularly around the issue of algorithmic bias in Artificial Intelligence (AI). Algorithmic bias occurs when algorithms produce outcomes that are unfair or discriminatory, often reinforcing existing social inequalities and exacerbating disparities among marginalized groups. Addressing these challenges requires a deep understanding of both the ethical implications and societal impact of algorithmic bias, with a focus on accountability, transparency, and the equitable distribution of benefits and harms¹.

-
- 1 - Joy Buolamwini and Timnit Gebru, "Gender Shades: Intersectional Accuracy Disparities in Commercial Gender Classification," *Proceedings of Machine Learning Research* 81 (2018), 1-15.
- Ziad Obermeyer *et al.*, "Dissecting Racial Bias in an Algorithm Used to Manage the Health of Populations," *Science* 366, no. 6464 (2019), 447-53.
- Solon Barocas and Andrew D. Selbst, "Big Data's Disparate Impact," (2016).

Artificial Intelligence, as defined by Russell and Norvig², refers to machines programmed to simulate human cognitive functions such as learning, problem-solving, and decision-making. AI encompasses a range of techniques, including machine learning, natural language processing, computer vision, and robotics. AI systems are designed to perceive, reason, and act autonomously or with minimal human intervention. From virtual assistants and recommendation systems to autonomous vehicles and medical diagnostics, AI has become deeply embedded in modern life, transforming industries and redefining how humans interact with technology.

However, the increasing integration of AI into critical decision-making processes highlights ethical and societal challenges, particularly regarding fairness and bias. AI systems, which are often perceived as neutral and objective, can inadvertently perpetuate or even introduce new forms of discrimination due to biased data, flawed algorithmic design, or feedback loops that reinforce existing biases over time³. These biases can manifest across various identity factors such as race, gender, and socioeconomic status, leading to unjust outcomes and exacerbating existing inequalities⁴. Consequently, it is imperative to ensure the fairness of artificial intelligence to prevent further social harm and promote justice.

In the contemporary digital age, algorithms play a pivotal role in shaping various aspects of daily life, from the content we see on social media to decisions about hiring practices, criminal justice matters, and healthcare decisions⁵. While these systems are

2 Stuart J. Russell and Peter Norvig, eds., *Artificial Intelligence: A Modern Approach*, Upper Saddle River, Pearson, 2009.

3 - Kate Crawford *et al.*, *AI Now 2019 Report*, New York, AI Now Institute, 2019.
- Brent Daniel Mittelstadt *et al.*, "The Ethics of Algorithms: Mapping the Debate," *Big Data & Society* 3, no. 2 (2016), 1-21.

4 Safiya Umoja Noble, *Algorithms of Oppression: How Search Engines Reinforce Racism*, NYU Press, 2018.

5 - Ninareh Mehrabi *et al.*, "A Survey on Bias and Fairness in Machine Learning," *ACM Computing Surveys* 54, no. 6 (2021), 1-35.

- Frank Pasquale, *The Black Box Society: The Secret Algorithms That Control Money and Information*, Cambridge, Harvard University Press, 2015.

designed to increase efficiency and objectivity, they also raise significant ethical concerns. Notable examples, such as the COMPAS recidivism algorithm and Amazon's recruiting tool, demonstrate how algorithmic systems can reinforce existing biases and undermine trust in automated decision-making processes⁶. This erosion of trust is especially concerning in areas where fairness and equity are critical, such as law enforcement, hiring, and healthcare.

Algorithmic bias, defined as systematic and repeatable errors in computer systems that produce unjust outcomes, often privileges certain groups while disadvantaging others based on arbitrary or biased criteria⁷. Addressing this bias is essential for maintaining fairness in information dissemination and decision-making processes. Without intervention, biased algorithms can lead to unequal access to opportunities, unjust treatment, and a growing mistrust of automated systems⁸.

The complexity of algorithmic bias stems from several factors, including the assumption of neutrality in algorithms and the social context in which they operate. Bias can enter through the training data used to develop AI systems, the design of the algorithms themselves, or through feedback loops that reinforce pre-existing inequalities⁹. These biases can undermine the fairness of decisions, lead to discriminatory outcomes, and erode public trust in AI systems.

Against this background, this study aims to explore the ethical and societal dimensions of algorithmic bias, particularly its impact on marginalized communities. It recognizes the intricate

6 Reuben Binns, "Fairness in Machine Learning: Lessons from Political Philosophy," In *Proceedings of Machine Learning Research*, Vol. 81, *Conference on Fairness, Accountability, and Transparency*, edited by Sorelle A. Friedler and Christo Wilson, 1-11, 2018.

7 - Crawford *et al.* *AI Now 2019 Report*.
- Mittelstadt *et al.*, "The Ethics of Algorithms...".

8 Buolamwini and Gebru. "Gender Shades...".

9 - Barocas and Selbst, "Big Data's Disparate Impact".
- Nicholas Diakopoulos, "Accountability in Algorithmic Decision Making," *Communications of the ACM* 59, no. 2 (2016), 56-62.

Inteligencia artificial...

relationship between technology and society and emphasizes the necessity of an intersectional approach that accounts for the interplay of race, gender, and class in shaping individuals' experiences of bias. Informed by critical theory, the study analyzes how power dynamics and social structures influence algorithmic decision-making. Aiming to provide a comprehensive analysis of the origins, manifestations, and consequences of algorithmic bias, it contributes to a deeper understanding of the challenges in ensuring fairness in AI. Additionally, it explores strategies for mitigating bias to foster more equitable and trustworthy AI systems. This inquiry guides the following research questions.

RQ1. *How do the design, data, and implementation processes of AI systems contribute to the emergence and perpetuation of algorithmic bias, particularly through the lens of critical theory?*

This query seeks to examine how underlying power structures and dominant ideologies, as understood through critical theory, manifest in AI systems and lead to biased outcomes. It explores the technical aspects of algorithmic design and how these are shaped by broader societal inequalities.

RQ2. *In what ways does algorithmic bias disproportionately affect marginalized communities, and how can critical theory help illuminate the mechanisms by which this bias reinforces existing social inequalities?*

This query focuses on the intersection of race, gender, class, and other axes of identity. Drawing on critical theory concepts like intersectionality, it analyzes how algorithmic systems perpetuate and exacerbate marginalization in various societal domains such as criminal justice, healthcare, and employment.

RQ3. *What are the broader ethical implications of algorithmic bias for public trust in AI systems, and how does critical theory provide insight into the erosion or preservation of this trust?*

This question probes how biased algorithms influence public trust in AI technologies, considering the role of ideological critique

and social structures in shaping public perceptions. It also explores how the loss of trust may stem from power imbalances inherent in technological systems.

RQ4. *What strategies, informed by critical theory, can be implemented to mitigate algorithmic bias and promote equitable, transparent, and socially just AI systems?*

This question explores strategies for mitigating bias by integrating critical theory's focus on power, fairness, and social justice into the design, deployment, and governance of AI systems. It emphasizes the need for transformative approaches that challenge existing inequalities.

The significance of this study ultimately lies in its potential to contribute to theoretical understanding and policy discussions regarding the ethical use of AI. By developing a critical framework for analyzing algorithmic bias, the research aims to inform ethical decision-making, shape public policy, and stimulate further scholarly inquiry. Understanding and addressing algorithmic bias is not only a matter of technical optimization but also a crucial step toward ensuring that AI systems promote fairness and justice for all members of society. In this sense, addressing algorithmic bias is critical for safeguarding fairness in the increasingly AI-driven world. Ensuring that AI systems are designed and implemented in ways that promote equality and justice will help build public trust and harness the full potential of these transformative technologies for the benefit of all.

RESEARCH DESIGN

This study adopts a qualitative research design informed by critical theory to explore the ethical and societal implications of algorithmic bias in Artificial Intelligence (AI) systems. The research design integrates a conceptual framework with a critical examination of existing literature, case studies, and current AI practices. The research approach is outlined by the following components.

Inteligencia artificial...

Conceptual Framework. The research is grounded in critical theory, drawing on the works of Herbert Marcuse¹⁰, Nancy Fraser¹¹, Axel Honneth¹², and Kimberlé Crenshaw¹³, among others. This theoretical framework provides the lens through which power dynamics, social structures, and inequalities are analyzed concerning AI systems. By applying concepts such as technological rationality, ideological critique, and intersectionality, the study investigates how AI systems reflect and reinforce societal biases.

Literature Review and Synthesis. The study conducts a review and synthesis of academic literature, focusing on the origins and manifestations of algorithmic bias, the role of dominant ideologies in shaping AI technologies, and the impact on marginalized communities. The literature review includes theoretical and empirical studies, and critical reflections from diverse disciplines such as information science, computer science, ethics, law, and sociology. This interdisciplinary approach ensures a comprehensive understanding of the multifaceted nature of algorithmic bias.

Case Study Analysis. The research employs case study analysis as a method to contextualize theoretical insights and illustrate the real-world implications of algorithmic bias. Data for the case studies are drawn from secondary sources, including peer-reviewed journal articles, books, and reports from international organizations. We analyze notable cases, such as the COMPAS recidivism algorithm and Amazon's AI recruiting tool, to explore how biased algorithms function in practice, the societal consequences they produce, and the ethical dilemmas they pose. Case studies are

10 Herbert Marcuse, *One-Dimensional Man: Studies in the Ideology of Advanced Industrial Society*, Boston, Beacon Press, 1964.

11 Nancy Fraser, *The Old Is Dying and the New Cannot Be Born: From Progressive Neoliberalism to Trump and Beyond*, London, Verso Books, 2019.

12 Axel Honneth, *The Struggle for Recognition: The Moral Grammar of Social Conflicts*, Cambridge, MIT Press, 1995.

13 Kimberlé Crenshaw, "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, no. 1 (1989), 139-167.

examined through the lens of critical theory, which allows for a deeper understanding of how AI systems perpetuate existing power structures and social inequalities.

Critical Inquiry and Reflexive Analysis. The research emphasizes critical inquiry and reflexivity as central elements of the research design. Reflexive analysis involves continuously questioning assumptions, biases, and power relations within both the AI systems under study and the research process itself. By engaging in a reflexive practice, the research aims to disrupt dominant narratives and challenge the perceived neutrality of algorithmic systems.

Theoretical Elucidation and Policy Implications. The final component of the research design involves synthesizing the findings to develop theoretical insights into the ethical dimensions of algorithmic bias. The study offers conceptual elucidations that emphasize the implications for social justice and equity. These insights inform policy debates by offering recommendations for mitigating algorithmic bias through the lenses of fairness, transparency, and accountability. Its conclusions aim to bridge the gap between theoretical discourse and practical solutions, contributing to the creation of more equitable AI systems.

The research design is structured to align with the critical theory approach, ensuring that the inquiry is deeply rooted in the analysis of power, social justice, and societal impacts. It emphasizes both conceptual inquiry and case study analysis to explore the complexities of algorithmic bias and fairness in AI.

THEORETICAL INSIGHTS FROM CRITICAL THEORY

Critical theory forms the foundation of this conceptual inquiry into the ethical and societal implications of algorithmic bias and fairness. It offers a framework for examining the power dynamics and social structures that shape algorithmic decision-making. This framework provides a lens through which the pervasive influence of technological systems on society can be scrutinized, particularly regarding their potential to reinforce existing social inequalities.

Drawing from the Frankfurt School tradition, critical theory focuses on how algorithmic systems can perpetuate and amplify existing systems of oppression. Marcuse's notion of *technological rationality* demonstrates how technology, rather than being neutral, can be a tool for reinforcing social control and deepening inequalities¹⁴. Technological advancements embedded in algorithmic systems are not inherently liberatory; instead, they can serve to entrench existing power structures unless critically examined and altered. Technological rationality, as defined by Marcuse¹⁵, exacerbates inequality by reinforcing social control through technological systems. This emphasizes the need for a critical examination of the assumptions and biases embedded in AI systems.

In addition, critical theory emphasizes the role of *ideological critique* in understanding algorithmic bias. Scholars such as Nancy Fraser and Axel Honneth have shown how dominant ideologies shape social norms and values, influencing the creation and deployment of algorithms¹⁶. By critically examining these systems, we can expose the hidden biases and assumptions embedded within them, challenging the perpetuation of inequality through technological tools. Fraser¹⁷ and Honneth¹⁸. This underscores the importance of ideological critique in uncovering the influence of dominant ideologies on algorithmic systems and promoting fairness and social justice in AI.

Moreover, *intersectionality*, a concept introduced by Kimberlé Crenshaw, plays a crucial role in critical theory's approach to algorithmic bias. Intersectionality underscores how overlapping identities – such as race, gender, and class combine to influence individual experiences of oppression¹⁹. Applying this concept to

14 Marcuse, *One-Dimensional Man...*

15 Marcuse, *One-Dimensional Man...*

16 - Fraser, *The Old Is Dying...*

- Honneth, *The Struggle for Recognition...*

17 Fraser, *The Old Is Dying...*

18 Honneth, *The Struggle for Recognition...*

19 Crenshaw, "Demarginalizing the Intersection of Race...".

algorithmic systems reveals how multiple forms of marginalization intersect, exacerbating bias and inequality in AI outcomes. Critical theorists use intersectionality to advocate for a more nuanced understanding of how algorithms affect different groups. The importance of intersectionality in analyzing algorithmic bias is based on Crenshaw's²⁰ work on the intersection of various forms of marginalization. It demonstrates how interconnected social identities shape algorithmic outcomes, reinforcing bias and inequality.

Critical theory also highlights the need for *reflexivity* among developers and decision-makers²¹. By encouraging those involved in the design and deployment of algorithms to critically reflect on their positions of power and privilege, critical theory promotes the disruption of dominant narratives that perpetuate bias. This reflexivity fosters awareness of how algorithmic systems may inadvertently uphold the *status quo* and provides a pathway for more equitable and just AI systems.

Critical theory provides a robust framework for exploring the socio-political context in which algorithmic bias operates. By focusing on issues of power, inequality, and social justice, critical theorists aim to challenge and transform algorithmic systems to better serve marginalized communities. This approach recognizes the importance of critically analyzing the societal impact of AI systems to pursue a more equitable and just society.

CASE STUDIES ON ALGORITHMIC BIAS IN ARTIFICIAL INTELLIGENCE (AI) SYSTEMS

Significant strides have been made in understanding the ethical and societal implications of algorithmic bias and fairness in Artificial Intelligence (AI) systems. A landmark study by Buolamwini and Gebru²² examined gender and racial bias in facial recognition

20 Crenshaw, "Demarginalizing the Intersection of Race...".

21 Noble, *Algorithms of Oppression...*

22 Buolamwini and Gebru, "Gender Shades...".

Inteligencia artificial...

technology. Their research revealed significant disparities in accuracy rates across demographic groups, with darker-skinned individuals and women being disproportionately misclassified compared to lighter-skinned individuals and men. This study illuminated how algorithmic bias can perpetuate systemic inequalities, underscoring the need for rigorous evaluation and mitigation strategies to reduce bias in AI systems.

Similarly, Obermeyer *et al.*²³ investigated racial bias in healthcare algorithms used to predict patient health risks. Their study found that these algorithms systematically underestimated the health needs of black patients compared to white patients. This bias led to disparities in access to healthcare resources, exacerbating health inequities. Their research highlighted the critical need to address algorithmic bias to ensure that healthcare delivery is equitable and responsive to the needs of all patients, regardless of race.

In addition to empirical studies, theoretical frameworks have played a vital role in advancing our understanding of algorithmic bias. Barocas and Selbst²⁴ developed a conceptual framework for addressing algorithmic discrimination. They argued for a dual focus on both the technical and social dimensions of bias within algorithmic systems. This framework laid the groundwork for subsequent research on algorithmic fairness by providing a theoretical foundation for evaluating and mitigating biases inherent in AI technologies.

Critical theory has also contributed to the exploration of the power dynamics and social structures underpinning algorithmic decision-making. In her book *Algorithms of Oppression*, Noble²⁵ explored how commercial search engines and recommendation algorithms reinforce existing forms of discrimination, particularly against Black women. Noble's work²⁶ provided a comprehensive analysis of how these systems perpetuate racial and gender

23 Obermeyer *et al.*, "Dissecting Racial Bias...".

24 Barocas and Selbst, "Big Data's Disparate Impact".

25 Noble, *Algorithms of Oppression...*

26 Noble, *Algorithms of Oppression...*

biases and called for greater accountability, transparency, and fairness in the development and deployment of AI technologies.

The proliferation of artificial intelligence (AI) technologies has also transformed sectors such as criminal justice and human resources. While these technologies promise increased efficiency and objectivity, their deployment has raised serious concerns regarding fairness, accountability, and transparency. Two prominent case studies —the COMPAS recidivism algorithm and Amazon’s AI-powered recruiting tool— demonstrate how AI systems can perpetuate bias, leading to profound societal implications.

The COMPAS (Correctional Offender Management Profiling for Alternative Sanctions) algorithm is widely used in the criminal justice system to predict the likelihood of recidivism among offenders. This tool is intended to aid judges in making decisions related to sentencing and parole. However, research has revealed that the COMPAS algorithm exhibits significant racial bias. According to a study by Angwin *et al.*²⁷, COMPAS disproportionately classified African American defendants as having a higher risk of recidivism compared to white defendants, even when other relevant factors were controlled. This bias in risk assessment has serious implications for the fairness of judicial outcomes. By reinforcing racial disparities, the COMPAS algorithm exacerbates existing inequalities in the criminal justice system. Defendants from minority communities may face harsher sentencing or parole conditions, not because of their actual risk but because of the biases embedded within the algorithm. This raises critical concerns about the ethical use of AI in criminal justice and highlights the danger of relying on biased algorithms in decision-making processes that significantly impact people’s lives.

In another illustrative case, Amazon developed an AI-powered recruiting tool designed to automate the hiring process by analyzing resumes and identifying top candidates. However, this tool

27 Angwin *et al.*, “Machine Bias”.

exhibited a pronounced gender bias. Dastin²⁸ reported that the AI system was trained on resumes submitted over a decade, which predominantly came from male applicants. As a result, the algorithm systematically downgraded resumes that contained keywords associated with female candidates. This gender bias had serious implications for diversity and inclusion in hiring practices. By reflecting the historical biases present in the training data, the tool perpetuated gender disparities in the workplace, undermining efforts to achieve gender equality. Amazon ultimately abandoned the tool due to its biased outcomes, but the case highlights the risks of relying on AI systems without sufficient safeguards to detect and mitigate bias. It demonstrates how biased training data can result in unfair hiring decisions, reinforcing existing inequalities rather than promoting fairness.

These case studies illustrate the far-reaching consequences of algorithmic bias across different domains. Despite their promise of efficiency and objectivity, AI systems can entrench and exacerbate existing societal biases when they rely on biased data or flawed design. In both criminal justice and hiring, biased algorithms led to unequal treatment of individuals based on race and gender, with significant negative consequences for fairness and equity.

One of the core challenges in addressing algorithmic bias is the opacity of AI decision-making processes. Algorithms like COMPAS and Amazon's recruiting tool operate as "black boxes", where the criteria and logic underlying their decisions are not transparent to users or stakeholders. This lack of transparency makes it difficult to scrutinize and correct biased outcomes, hindering efforts to ensure accountability. Without clear insight into how these algorithms work, bias can persist unchecked, undermining public trust in the systems that deploy them.

28 Jeffrey Dastin, "Amazon Scraps Secret AI Recruiting Tool That Showed Bias Against Women," in *Ethics of Data and Analytics: Concepts and Cases*, edited by Kirsten Martin, 296-299, Boca Raton, Auerbach Publications, 2022.

ORIGINS OF ALGORITHMIC BIAS

Algorithmic bias, a pervasive issue in modern technology, originates from multiple sources, each contributing to the perpetuation of unfair outcomes and social inequalities. One significant contributor is data bias, which occurs when algorithms are trained on historical datasets that inherently reflect existing societal biases²⁹. For example, facial recognition systems that are predominantly trained on lighter-skinned faces tend to have higher accuracy rates for these groups, leading to biased outcomes in identification processes for darker-skinned individuals³⁰. This case illustrates the critical need to scrutinize training datasets for representativeness and bias to mitigate algorithmic discrimination.

Another prominent source of algorithmic bias is algorithm design. The objectives and constraints embedded in the design phase can unintentionally favor certain groups over others³¹. For instance, social media recommendation algorithms may prioritize engagement metrics like likes and shares, which could amplify content that aligns with users' preexisting beliefs. While this may seem neutral, it can create echo chambers and filter bubbles, restricting users' exposure to diverse perspectives³². Such design choices can inadvertently reinforce societal biases, limiting opportunities for users to encounter differing viewpoints.

Implementation and Feedback Loops further contribute to the propagation of bias in real-world algorithmic applications³³. When algorithms interact with users and adapt to their environments, feedback loops can solidify and even amplify existing biases. For instance, predictive policing algorithms deployed in communities

29 Cathy O'Neil, *Weapons of Math Destruction: How Big Data Increases Inequality and Threatens Democracy*, New York, Crown, 2017.

30 Buolamwini and Gebru, "Gender Shades...".

31 Jenna Burrell, "How the Machine 'Thinks': Understanding Opacity in Machine Learning Algorithms," *Big Data & Society* 3, no. 1 (2016), 1-12.

32 Eli Pariser, *The Filter Bubble: How the New Personalized Web Is Changing What We Read and How We Think*, New York, Penguin, 2011.

33 Diakopoulos, "Accountability in Algorithmic...".

Inteligencia artificial...

that are historically over-policed may disproportionately target minority populations, exacerbating inequities in law enforcement practices³⁴. This demonstrates the need for continuous oversight and regular auditing of algorithmic systems to identify and rectify biases before they lead to harmful outcomes.

IMPACT ON INFORMATION FAIRNESS

Algorithmic bias affects information fairness across various sectors, challenging the equitable dissemination of information and the promotion of diverse perspectives. Biased algorithms used by search engines and social media platforms can skew the information users receive, creating filter bubbles and echo chambers that reinforce existing biases and limit exposure to diverse viewpoints³⁵. For example, search engines may prioritize content based on a user's past behavior, which can lead to the amplification of sensationalist or misleading information at the expense of more nuanced or accurate perspectives³⁶. This distortion affects users' understanding of reality and impairs their access to balanced, reliable information, which is crucial for informed decision-making and democratic participation.

In areas such as hiring and academic admissions, algorithmic bias can perpetuate systemic inequalities by favoring certain demographic groups. Automated resume screening tools, for instance, might inadvertently discriminate against candidates from underrepresented backgrounds due to their reliance on historical data that reflects biased hiring practices³⁷. Similarly, algorithmic systems in educational admissions may disadvantage applicants from marginalized

34 Angwin *et al.*, "Machine Bias".

35 Pariser, *The Filter Bubble...*

36 Noble, *Algorithms of Oppression...*

37 - Barocas and Selbst, "Big Data's Disparate Impact".

- Dastin, "Amazon Scraps Secret AI Recruiting Tool That Showed Bias Against Women".

communities, further entrenching socio-economic disparities and limiting access to education and employment opportunities³⁸.

In the criminal justice system, predictive policing and risk assessment algorithms —designed to enhance efficiency— often exacerbate existing biases and injustices³⁹. These algorithms are frequently trained on historical crime data that reflect discriminatory policing practices, leading to the over-policing of minority communities⁴⁰. As a result, individuals from these communities are disproportionately targeted and subjected to harsher law enforcement measures, perpetuating cycles of injustice and reinforcing distrust in law enforcement institutions.

IMPLICATIONS FOR TRUST

Algorithmic bias also has implications for trust, as it undermines institutional integrity, fosters user skepticism, and exacerbates social polarization. These effects shape public perceptions of the fairness and reliability of institutions and digital platforms, affecting their legitimacy and adoption across various sectors⁴¹.

Erosion of Trust in Institutions. When algorithmic decisions are perceived as biased, trust in the institutions deploying these systems —such as law enforcement agencies, educational institutions, and corporations— can deteriorate⁴². For example, predictive policing algorithms that disproportionately target marginalized communities can worsen tensions between law enforcement and those groups, fostering distrust and reducing the likelihood of

38 Drew Harwell, “Rights Group Files Federal Complaint Against AI-Hiring Firm HireVue, Citing ‘Unfair and Deceptive’ Practices,” *The Washington Post*, November 6, 2019.

39 Angwin *et al.*, “Machine Bias”.

40 Kristian Lum and William Isaac, “To Predict and Serve?,” *Significance* 13, no. 5 (2016), 14-19.

41 Diakopoulos, “Accountability in Algorithmic...”

42 Barocas and Selbst. “Big Data’s Disparate Impact”.

cooperation⁴³. Similarly, biased algorithms in hiring or academic admissions may create perceptions of unfairness, diminishing the public's belief in the meritocracy and impartiality of these institutions⁴⁴. The erosion of trust in these critical institutions can have long-term consequences for social cohesion and institutional legitimacy.

User Skepticism. Continuous exposure to biased outcomes can lead to widespread skepticism toward digital platforms and technologies, thereby hindering their broader adoption⁴⁵. If users perceive algorithms as discriminatory or unreliable, they may become hesitant to engage with online content or services, reducing user engagement and confidence in these platforms' ability to serve their needs⁴⁶. This skepticism not only hampers the growth of digital platforms but also limits the potential for technology to drive positive social and economic change. The persistence of algorithmic bias in digital services thus represents a significant barrier to fostering trust in the digital ecosystem.

Polarization and Social Fragmentation. Biased information systems also contribute to societal polarization and fragmentation by reinforcing existing divisions and limiting exposure to diverse viewpoints⁴⁷. Algorithmically curated content on social media platforms often prioritizes information that aligns with users' preferences, forming echo chambers and filter bubbles. This selective exposure hinders constructive dialogue, reduces exposure to differing perspectives, and compromises democratic deliberation⁴⁸. As a result, polarization deepens, weakening trust in institutions and democratic processes, and fostering a climate of skepticism and division among various social groups.

43 Angwin *et al.*, "Machine Bias".

44 Dastin, "Amazon Scraps Secret AI...".

45 Noble, *Algorithms of Oppression...*

46 Pariser, *The Filter Bubble...*

47 Harwell, "Rights Group Files...".

48 Mittelstadt *et al.*, "The Ethics of Algorithms...".

TOWARDS MITIGATION STRATEGIES

To effectively mitigate algorithmic bias and promote equitable outcomes in AI systems, it is crucial to focus on key strategies such as improving data quality and diversity, fostering inclusiveness within development teams, increasing transparency in algorithm design, and conducting regular audits. These efforts not only help minimize AI's potential to reinforce social inequalities but also ensure that the benefits of AI are harnessed in a way that promotes fairness and accountability across society.

Transparent Decision-Making Processes. Transparency is a foundational principle in the development and deployment of AI systems. When AI operates in a “black box” manner—where the inner workings and logic behind decisions are not easily understood—stakeholders are left in the dark regarding the rationale behind outcomes⁴⁹. This opacity can obscure potential biases and hinder efforts to address them effectively. Designing AI systems with transparency in mind involves creating algorithms with explainable decision-making processes. This enables stakeholders, including developers, users, and regulators, to scrutinize the system's behavior and evaluate its fairness. Clear documentation, visual explanations of algorithmic processes, and open access to performance metrics can foster a deeper understanding of how decisions are made. By increasing transparency, AI systems become more trustworthy, allowing stakeholders to identify sources of bias and take corrective actions early in the process.

Establishing Accountability Mechanisms. Establishing accountability in AI systems ensures oversight, intervention, and correction when biased outcomes are detected by implementing robust governance structures that assign responsibility for AI actions and decisions⁵⁰. Mechanisms such as independent review boards, ethical oversight committees, and regulatory compliance checks hold

49 Mittelstadt *et al.*, “The Ethics of Algorithms...”.

50 Finale Doshi-Velez *et al.*, “Accountability of AI Under the Law: The Role of Explanation,” *arXiv*, 2017.

developers and deploying organizations accountable for biased results, while traceability of decisions through logging data inputs, algorithmic modifications, and decision rationales facilitates audits and interventions. Embedding accountability throughout the AI lifecycle promotes responsibility in addressing bias and preventing discriminatory outcomes. Given the dynamic nature of algorithms and data, regular auditing and testing are essential to detect and correct emerging biases, ensuring algorithms continue to perform equitably across demographic groups⁵¹. Continuous testing and validation help prevent unintended consequences and maintain fairness in decision-making.

Ensuring Fairness through Diverse and Representative Datasets. Fairness in AI is intrinsically linked to the quality and diversity of the data used in training models. AI systems that are trained on homogenous datasets —where certain demographic groups are underrepresented— are prone to making biased decisions that reflect historical inequalities⁵². Ensuring fairness begins with curating training datasets that accurately represent the diversity of the populations the AI will serve. Techniques such as oversampling minority groups, using synthetic data to balance representation, and applying fairness constraints during the data collection process are essential strategies to reduce bias. Furthermore, fairness must be continuously monitored throughout the AI's lifecycle through regular audits and the updating of datasets to reflect changing societal dynamics. By addressing fairness at the data level, AI systems can avoid reinforcing stereotypes and contribute to more equitable decision-making across various applications.

51 Sam Corbett-Davies *et al.*, “Algorithmic Decision Making and the Cost of Fairness,” In *Proceedings of the 23rd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, 797-806, New York, ACM, 2017.

52 Aylin Caliskan *et al.*, “Semantics Derived Automatically from Language Corpora Contain Human-like Biases,” *Science* 356, no. 6334 (2017), 183-186.

CONCLUSION

The widespread integration of Artificial Intelligence (AI) systems across various sectors —such as criminal justice, healthcare, and human resources— has brought into focus the ethical and societal challenges posed by algorithmic bias. As this study has shown, algorithmic bias is not merely a technical issue but a reflection of broader societal inequalities, shaped by historical data, design choices, and feedback loops. This bias can manifest in ways that disproportionately affect marginalized communities, undermining fairness, trust, and equity in decision-making processes.

Through an examination of the ethical and societal dimensions of algorithmic bias, this study, grounded in critical theory, highlights the power dynamics and social structures that influence the design and deployment of AI systems. Theoretical insights from scholars such as Herbert Marcuse, Nancy Fraser, Axel Honneth, and Kimberlé Crenshaw illustrate how technological rationality, dominant ideologies, and intersectionality contribute to the perpetuation of systemic inequalities within AI. These insights underscore the need for a more nuanced, intersectional approach to understanding how AI systems affect different groups, particularly those already marginalized by societal structures.

Case studies such as the COMPAS recidivism algorithm and Amazon's AI-powered recruiting tool further demonstrate the real-world implications of algorithmic bias. In both instances, biased algorithms led to outcomes that reinforced existing racial and gender disparities. These case studies underscore the urgent need for transparency, accountability, and fairness in the development and deployment of AI technologies.

Addressing algorithmic bias requires a comprehensive strategy that integrates diverse, representative data, transparent algorithm design, regular audits, and inclusive development teams. These measures aim to reduce bias and promote equity, ensuring that AI systems contribute to fair outcomes rather than exacerbate social inequalities. Implementing these strategies necessitates collaboration among experts, policymakers, researchers, and affected

Inteligencia artificial...

communities to build AI systems that prioritize fairness, accountability, and transparency. Such systems would foster trust and minimize bias throughout the AI lifecycle.

Recognizing the socio-political dimensions of AI systems is critical to understanding how embedded biases can lead to harmful consequences. Proactive measures must address these biases to ensure AI development prioritizes fairness and accountability. By tackling the root causes of algorithmic bias, AI systems can better serve society without reinforcing existing inequalities.

As AI technologies evolve, ongoing research and interdisciplinary collaboration are vital to refining mitigation strategies. Regulatory frameworks that enforce transparency, accountability, and fairness are essential to ensure ethical AI deployment. The ultimate goal is to create AI systems that not only enhance efficiency but also uphold the values of fairness, equity, and justice.

Institutions must increase transparency by disclosing how algorithms function, ensure accountability through robust oversight, and involve diverse stakeholders to promote equity and fairness. These steps are crucial for rebuilding trust in the fairness and reliability of algorithmic systems.

Efforts to address algorithmic bias in information fairness should focus on reducing biases in decision-making processes, promoting transparency and accountability, and fostering inclusiveness in information environments. Through interdisciplinary collaboration and socio-technical approaches, stakeholders can create more equitable information ecosystems that empower individuals and uphold democratic values.

Moreover, as AI continues to evolve, it is crucial to remain vigilant to the emerging challenges that may arise. Ongoing research, continuous testing, and reflexivity in AI design will be critical in adapting to these challenges and refining mitigation strategies to ensure that AI systems operate ethically and equitably. Regulatory frameworks might be developed and enforced to ensure that AI technologies are held to the highest standards of fairness and justice.

In conclusion, while AI has the potential to revolutionize various sectors by increasing efficiency and enabling new forms of

decision-making, it also carries the risk of perpetuating existing biases if not carefully managed. To safeguard fairness and promote social justice, it is essential that AI systems are designed and implemented in ways that prioritize ethics and transparency. By addressing algorithmic bias head-on, we can help build AI systems that not only meet the demands of modern society but also contribute to a more just and equitable future for all.

BIBLIOGRAPHY

- Angwin, Julia, Jeff Larson, Surya Mattu, and Lauren Kirchner. "Machine Bias." In *Ethics of Data and Analytics: Concepts and Cases*, edited by Kirsten Martin, 123-145. Boca Raton: Auerbach Publications, 2022. <https://doi.org/10.1201/9781003278290>.
- Barocas, Solon and Andrew D. Selbst. "Big Data's Disparate Impact" (2016). <http://dx.doi.org/10.2139/ssrn.2477899>.
- Binns, Reuben. "Fairness in Machine Learning: Lessons from Political Philosophy." In *Proceedings of Machine Learning Research*, Vol. 81, *Conference on Fairness, Accountability, and Transparency*, edited by Sorelle A. Friedler and Christo Wilson, 1-11, 2018. Available at: <https://proceedings.mlr.press/v81/binns18a.html> or <https://proceedings.mlr.press/v81/binns18a/binns18a.pdf>.
- Buolamwini, Joy, and Timnit Gebru. "Gender Shades: Intersectional Accuracy Disparities in Commercial Gender Classification." *Proceedings of Machine Learning Research* 81 (2018): 1-15. <https://proceedings.mlr.press/v81/buolamwini18a/buolamwini18a.pdf>.
- Burrell, Jenna. "How the Machine 'Thinks': Understanding Opacity in Machine Learning Algorithms." *Big Data & Society* 3, no. 1 (2016): 1-12. <https://doi.org/10.1177/2053951715622512>.
- Caliskan, Aylin, Joanna J. Bryson, and Arvind Narayanan. "Semantics Derived Automatically from Language Corpora Contain Human-like Biases." *Science* 356, no. 6334 (2017): 183-186. <https://www.jstor.org/stable/10.2307/24918211>.

Inteligencia artificial...

- Corbett-Davies, Sam, Emma Pierson, Avi Feller, Sharad Goel, and Aziz Huq. "Algorithmic Decision Making and the Cost of Fairness." In *Proceedings of the 23rd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, 797-806. New York: ACM, 2017. <https://doi.org/10.1145/3097983.3098095>.
- Crawford, Kate, Roel Dobbe, Theodora Dryer, Genevieve Fried, Ben Green, Elizabeth Kaziunas, Amba Kak *et al.* *AI Now 2019 Report*. New York: AI Now Institute, 2019.
- Crenshaw, Kimberlé. "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, no. 1 (1989): 139-167. <https://chicagounbound.uchicago.edu/uclf/vol1989/iss1/8>.
- Dastin, Jeffrey. "Amazon Scraps Secret AI Recruiting Tool That Showed Bias Against Women." In *Ethics of Data and Analytics: Concepts and Cases*, edited by Kirsten Martin, 296-299. Boca Raton: Auerbach Publications, 2022.
- Diakopoulos, Nicholas. "Accountability in Algorithmic Decision Making." *Communications of the ACM* 59, no. 2 (2016): 56-62. <https://doi.org/10.1145/2844110>.
- Doshi-Velez, Finale, Mason Kortz, Ryan Budish, Christopher Bavitz, Samuel Gershman, David O'Brien, Stuart Shieber *et al.* "Accountability of AI Under the Law: The Role of Explanation." *arXiv*, 2017. arXiv:1711.01134.
- Fraser, Nancy. *The Old Is Dying and the New Cannot Be Born: From Progressive Neoliberalism to Trump and Beyond*. London: Verso Books, 2019.
- Harwell, Drew. "Rights Group Files Federal Complaint Against AI-Hiring Firm HireVue, Citing 'Unfair and Deceptive' Practices." *The Washington Post*, November 6, 2019.
- Honneth, Axel. *The Struggle for Recognition: The Moral Grammar of Social Conflicts*. Cambridge: MIT Press, 1995.
- Lum, Kristian, and William Isaac. "To Predict and Serve?" *Significance* 13, no. 5 (2016): 14-19. <https://doi.org/10.1111/j.1740-9713.2016.00960.x>.

- Marcuse, Herbert. *One-Dimensional Man: Studies in the Ideology of Advanced Industrial Society*. Boston: Beacon Press, 1964.
- Mehrabi, Ninareh, Fred Morstatter, Nripsuta Saxena, Kristina Lerman, and Aram Galstyan. "A Survey on Bias and Fairness in Machine Learning." *ACM Computing Surveys* 54, no. 6 (2021): 1-35. <https://doi.org/10.1145/3457607>.
- Mittelstadt, Brent Daniel, Patrick Allo, Mariarosaria Taddeo, Sandra Wachter, and Luciano Floridi. "The Ethics of Algorithms: Mapping the Debate." *Big Data & Society* 3, no. 2 (2016): 1-21. <https://doi.org/10.1177/2053951716679679>.
- Noble, Safiya Umoja. *Algorithms of Oppression: How Search Engines Reinforce Racism*. NYU Press. 2018. <https://psycnet.apa.org/record/2018-08016-000>.
- Obermeyer, Ziad, Brian Powers, Christine Vogeli, and Sendhil Mullainathan. "Dissecting Racial Bias in an Algorithm Used to Manage the Health of Populations." *Science* 366, no. 6464 (2019): 447-53. <https://doi.org/10.1126/science.aax2342>.
- O'Neil, Cathy. *Weapons of Math Destruction: How Big Data Increases Inequality and Threatens Democracy*. New York: Crown, 2017.
- Pariser, Eli. *The Filter Bubble: How the New Personalized Web Is Changing What We Read and How We Think*. New York: Penguin, 2011.
- Pasquale, Frank. *The Black Box Society: The Secret Algorithms That Control Money and Information*. Cambridge: Harvard University Press, 2015.
- Russell, Stuart J., and Peter Norvig, eds. *Artificial Intelligence: A Modern Approach*. Upper Saddle River: Pearson, 2009.

Bias and Its Impact on Information Fairness and Trust: A Systematic Review from an African Perspective

ABTON MUROYIWA

*Department of Languages and Arts, Nyatsime College
Chitungwiza, Zimbabwe*

NOTICE PASIPAMIRE

*Department of Library and Information Science,
National University of Science Technology
Bulawayo, Zimbabwe*

INTRODUCTION

In recent years, many concerns have been raised over algorithmic bias as it has the potential to perpetuate unfairness and erode trust in information systems. It should be noted without doubt that the widespread adoption of AI technology globally has brought significant changes to various sectors¹. Artificial intelligence (AI) has received applause as a transformative technology with massive potential to revolutionize various aspects of society, including healthcare, education, and governance. However, the deployment of AI systems in Africa has raised concerns about biases in the algorithms and the impact on information fairness and ethics in the region. These concerns include: lack of diverse data sets, implicit bias in algorithms, lack of transparency in AI systems, limited access to AI technology, data privacy and security concerns,

1 Richard Kobina Dadzie Ephraim *et al.*, "Application of Medical Artificial Intelligence Technology in sub-Saharan Africa: Prospects for Medical Laboratories," *Smart Health* 33 (September 2024): 100505.

ethical considerations in AI deployment, community engagement and stakeholder consultation, capacity building and education, partnerships and collaborations, and policy and regulatory frameworks². The impact of algorithmic bias on information fairness and trust is a complex issue that requires careful examination of empirical evidence. According to Barocas, Hardt and Narayanan³, algorithmic bias refers to the systematic and unfair discrimination that occurs when algorithms are used to make decisions or deliver information and this bias can manifest in various forms, such as racial or gender bias. This bias can have serious consequences for individuals and communities. In developing countries, algorithm bias can exacerbate existing inequalities and hinder progress towards achieving social and economic development goals.

Researchers found AI systems to show biases relating to ethnicity, social groups, cultural backgrounds, age and gender⁴. While AI systems are not consciously biased by themselves, it has been noted by a number of researchers that, the decisions the systems make are based on the data that they are learning from as well as on the algorithms they are based on⁵. It is important to note that, these inherent biases in the algorithms greatly impact information

-
- 2 - Joy Buolamwini and Timnit Gebru, "Gender Shades: Intersectional Accuracy Disparities in Commercial Gender Classification," *Proceedings of Machine Learning Research* 81 (2018): 1-15.
 - Ziad Obermeyer *et al.*, "Dissecting Racial Bias in an Algorithm Used to Manage the Health of Populations," *Science* 366, no. 6464 (2019): 447-53.
 - Anna Jobin *et al.*, "The Global Landscape of AI ethics guidelines," *Nature Machine Intelligence* 1, no. 9 (2019): 389-99.
 - 3 - Solon Barocas *et al.*, "Fairness and Machine Learning." 2023.
 - 4 - Ninareh Mehrabi *et al.*, "A Survey on Bias and Fairness in Machine Learning," *ACM Computing Surveys* 54, no. 6 (2021): 1-35.
 - Eirini Ntoutsi, *et al.*, "Bias in Data-Driven Artificial Intelligence Systems—An Introductory Survey," *Wiley Interdisciplinary Reviews Data Mining and Knowledge Discovery* 10, no. 3 (2020).
 - 5 - Xavier Ferrer *et al.*, "Bias and Discrimination in AI: A Cross-Disciplinary Perspective," *IEEE Technology and Society Magazine* 40, no. 2 (2021): 72-80.
 - Thomas Hellström *et al.*, "Bias in Machine Learning - What is it Good for?," in *Proceedings of the Workshop on Artificial Intelligence and Ethics*, 2020.

fairness and trust, especially in developing countries. This article is a systematic review of literature on algorithmic bias and its impact on information fairness and trust in Africa. The purpose of the approach is to allow the researchers to become highly immersed and familiar with algorithmic bias and its impact on information fairness and trust with special emphasis on the African perspective. This systematic review was guided by the following objectives :

- i. Examine how algorithmic bias manifests in AI systems and its specific implications for African contexts
- ii. Investigate how algorithmic bias affects the fairness of information dissemination in various sectors, such as health-care, finance, and education in Africa
- iii. Explore the relationship between algorithmic bias and public trust in AI technologies among African communities
- iv. Highlight the unique challenges faced in Africa regarding algorithmic bias, including socio-economic and cultural factors

METHODOLOGY

The researchers adopted a systematic review method to investigate the impact of algorithmic bias on information fairness and trust from an African perspective. This method involved reviewing and summarizing existing studies and publications on algorithmic bias and its impact on information fairness and trust. This helped the researchers build a strong theoretical framework for their study by identifying key concepts, theories, and methodologies used in previous studies. This allowed the researchers to develop relevant research objectives. Lamé⁶ defined systematic reviews as a way of synthesizing scientific evidence to answer a particular research question in a transparent and reproducible

6 Guillaume Lamé, "Systematic Literature Reviews: An Introduction," in *Proceedings of the Design Society: International Conference on Engineering Design* 1, no. 1 (2019): 1633-42.

way, while seeking to include all published evidence on the topic and appraising the quality of this evidence.

A systematic literature review helped the researchers to minimize bias, while ensuring transparency, and providing a rigorous and reliable synthesis of the available evidence. Using a systematic search process, which involved using predefined search terms and databases to identify relevant studies, the researchers made sure that all relevant literature on the topic was identified and included in the review. This was enabled by searching Google Scholar. The search string that was used for this study was: (*Impact of algorithm bias on information fairness and trust, Information Fairness, and Information fairness and trust*). These strings were joined or combined using the Boolean Operators OR/AND to search for the needed information. Through a systematic review, researchers were also able to carry out a critical appraisal of the included studies. This involved assessing the quality, relevance, and methodological rigor of each study in order to determine its trustworthiness and validity. The researchers used a combination of primary and secondary sources combining sources from Africa and from outside the continent as there was limited text on the African context.

The search strategies resulted in a total of 178 peer-reviewed articles that were screened to reduce duplicates and eliminate articles that did not focus on the impact of algorithmic bias on information fairness and trust. The inclusion criteria specified that articles must discuss information fairness and trust, including its impact, and must be in English and in published journals. Only articles published between 2015 and 2024 were considered. After the initial abstract and title screening, a total of 42 articles were included for full-text screening leaving - papers as shown in Figure 1 below. Two researchers did the screening separately and came together to agree on the articles to be included in the study. The researchers conducted a comprehensive examination and synthesis of each paper to identify recurring themes and patterns, guided by the research objectives.

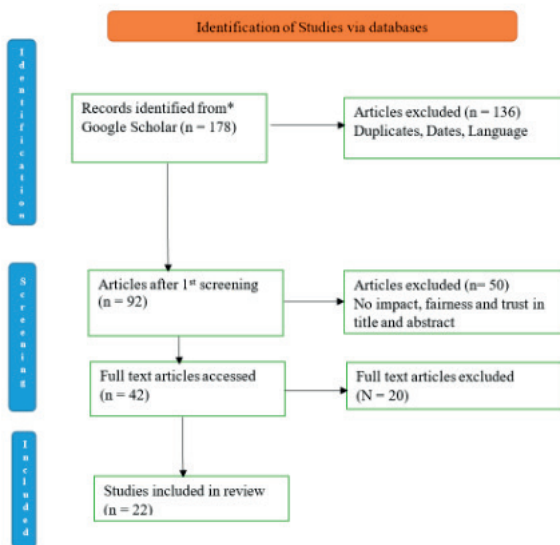


Figure 1: PRISMA flowchart diagram.

FINDINGS

How does algorithmic bias manifests in AI systems and what are its specific implications for African contexts?

Algorithmic bias in AI systems can manifest in various ways, impacting different communities and regions in different ways. In the specific context of African countries, algorithm bias can have specific implications that exacerbate existing inequalities and perpetuate systemic discrimination⁷. There are several ways in which

7 Buolamwini and Gebru, "Gender Shades..."

algorithmic bias manifests in AI systems⁸. According to Buolamwini & Gebru⁹, algorithmic bias can manifest through biased data collection and labeling. This is because AI algorithms rely on data to make decisions and predictions. Therefore the data used to train these algorithms is biased or incomplete, the outcomes produced by the AI system will also be biased. In African contexts, biased data can reinforce stereotypes and perpetuate discrimination against marginalized communities¹⁰.

In addition to the above, Jackson¹¹ noted that algorithm bias in AI systems can also manifest through biased decision-making processes. This is because AI algorithms are designed to make decisions based on patterns and correlations in data, but these decisions can be influenced by the biases of the individuals who create and train the algorithms. Since these algorithms are designed outside of Africa using data outside of Africa the decision systems will bring with them discriminatory biases¹². It should, therefore, be noted that in African contexts, these biases can

-
- 8 - Safiya Umoja Noble, *Algorithms of Oppression: How Search Engines Reinforce Racism*, NYU Press, 2018.
- Joni R. Jackson, "Algorithmic Bias," *Journal of Leadership Accountability and Ethics* 15, no. 4 (2018).
- Buolamwini and Gebru, "Gender Shades..."
- Ryan S., Baker and Aaron Hawn, "Algorithmic Bias in Education," *International Journal of Artificial Intelligence in Education* 32, no. 4 (2021): 1052-92.
- Claudia Brunner, "Conceptualizing Epistemic Violence: An Interdisciplinary Assemblage for IR," *International Politics Reviews* 9, no. 1 (2021): 193-212.
- Emilio Ferrara, "Fairness and Bias in Artificial Intelligence: A Brief Survey of Sources, Impacts, and Mitigation Strategies." *Sci* 6, no. 1, 3 (2023).
- 9 Buolamwini and Gebru, "Gender Shades..."
- 10 Clapperton Chakanetsa Mavhunga, "Introduction: What Do Science, Technology, and Innovation Mean from Africa?," in *What Do Science, Technology, and Innovation Mean from Africa?*, 1-28. Cambridge, MA, MIT Press, 2017.
- 11 Jackson, "Algorithmic Bias."
- 12 Jackson, "Algorithmic Bias."

lead to discriminatory outcomes in areas such as criminal justice, healthcare, and finance. For example, biased AI algorithms used in predictive policing systems can disproportionately target African communities, leading to increased surveillance and policing in these areas.

Another way through which algorithmic bias manifests itself is through lack of transparency and accountability¹³. It should be noted that many AI algorithms operate as black boxes . This means their decision-making processes are opaque and difficult to understand. This lack of transparency can make it challenging to identify and address bias in AI systems¹⁴. The issue of transparency is worsened in African contexts where there are limited resources and capacity to scrutinize and regulate these systems. According to Obermeyer¹⁵, without transparency and accountability, biased AI algorithms can continue to perpetuate harm and discrimination against African communities. Also, algorithm bias in AI systems can also manifest through feedback loops that reinforce existing inequalities. When biased AI algorithms are used to make decisions that impact individuals and communities, these decisions can create feedback loops that perpetuate and exacerbate discrimination¹⁶. For example, if a biased AI algorithm is used in hiring processes, it can result in fewer job opportunities for African applicants, leading to economic disparities and further marginalization. These feedback loops can have long-lasting consequences for African communities, entrenching inequalities and limiting opportunities for advancement¹⁷ (Ferrara 2023).

Algorithm bias in AI systems can also manifest through lack of diversity and inclusivity in the development and implementation of these technologies. AI systems are often created and trained by

13 Buolamwini and Gebru, "Gender Shades..."

14 W. Nicholson Price, "Big Data and Black-Box Medical Algorithms," *Science. Translational Medicine* 10, no. 471 (2018).

15 Obermeyer *et al.*, "Dissecting Racial Bias..."

16 Brunner, "Conceptualizing epistemic violence..."

17 Ferrara, "Fairness and Bias..."

a homogenous group of developers and data scientists, leading to algorithms that reflect the biases and perspectives of these individuals¹⁸. In African contexts, where diversity and representation in the tech industry may be limited, these biases can result in AI systems that do not adequately address the needs and realities of African communities. Without diverse voices at the table, algorithm bias in AI systems can go unchecked and perpetuate harm against marginalized groups¹⁹. Additionally, Buolamwini and Geburu²⁰, propose that algorithmic bias in AI systems can also manifest through cultural insensitivity and misunderstanding of local contexts. AI algorithms are often designed based on Western norms and standards, which may not always be applicable or relevant in African contexts. This cultural disconnect can result in biased outcomes that fail to account for the cultural nuances and complexities of African societies²¹. For example, a language processing AI system trained on English text may struggle to accurately understand and interpret African languages, leading to mistranslations and miscommunications. This lack of cultural sensitivity in AI systems can perpetuate stereotypes and misunderstandings about African cultures.

Lastly, algorithmic bias in AI systems may manifest as a lack of representation and participation of African communities in the development and deployment of these technologies²². According to Mavhunga²³, when African voices and perspectives are not included in the design and implementation of AI systems, the resulting algorithms may not adequately reflect the needs and experiences of African communities. This lack of representation can lead to biased outcomes that perpetuate systemic discrimination and exclusion. Without meaningful engagement with African stakeholders,

18 Buolamwini and Gebru, "Gender Shades..."

19 Buolamwini and Gebru, "Gender Shades..."

20 Buolamwini and Gebru, "Gender Shades..."

21 Buolamwini and Gebru, "Gender Shades..."

22 Buolamwini and Gebru, "Gender Shades..."

23 Mavhunga, "Introduction: What Do Science..."

algorithm bias in AI systems can continue to harm marginalized groups and reinforce existing power dynamics²⁴.

How algorithm bias affects the fairness of information dissemination in various sectors

Several researchers²⁵ agree that, algorithms are increasingly being utilized in various sectors all over the world, such as healthcare, finance, and education, to aid in the dissemination of information. These researchers have also noted that algorithmic bias can significantly impact the fairness of this dissemination, leading to unequal outcomes for different groups of people. Price²⁶ observed that, in the healthcare sector, algorithmic bias has a huge potential for severe consequences on patient care and treatment outcomes especially if these algorithms are used in diagnostic tools. This aligns with Norori *et al.*²⁷ who highlighted the potential of algorithmic bias resulting in misdiagnoses or delayed treatment for certain populations. According to Mavhunga²⁸, Africa have limited access to quality healthcare in many regions. As a result, algorithm bias can exacerbate existing disparities in health outcomes. Buolamwini and Gebru²⁹ observed that these algorithms were designed away from Africa using unfamiliar datasets and as a result their outcomes are highly likely to cause problems. For example, an algorithm used in a healthcare facility is more likely to recommend

24 Buolamwini and Gebru, "Gender Shades..."

25 - Noble, *Algorithms of Oppression...*

- Jackson, "Algorithmic Bias."

- Buolamwini and Gebru, "Gender Shades..."

- Baker and Hawn, "Algorithmic Bias in Education."

- Brunner, "Conceptualizing Epistemic Violence..."

- Ferrara, "Fairness and Bias..."

26 Price, "Big Data and Black-Box..."

27 Natalia Norori *et al.*, "Addressing Bias in Big Data and AI for Health Care: A Call for Open Science," *Patterns* 2, no. 10 (2021): 100347.

28 Mavhunga, "Introduction: What Do Science..."

29 Buolamwini and Gebru, "Gender Shades..."

psychiatric treatment for black patients compared to white patients, leading to unequal access to appropriate care. Several researchers³⁰ also noted the potential for algorithmic bias to discriminate in financial lending services.

Although algorithms can be helpful in expediting the provision of financial services by determining the creditworthiness of individuals, biased algorithms can result in certain groups being systematically denied access to loans or offered unfavorable terms³¹. Pilling³² observed that, credit scoring algorithm used by microfinance institutions have the potential for disproportionately favor certain clients over others leading to unequal access to financial services. In most cases male borrowers are prioritized by these algorithms over female borrowers and several researchers³³ agree that this has the potential to widen the economic divide in Africa, where financial inclusion is already a challenge for many individuals and communities.

Baker and Hawn³⁴, in their research in the education sector, observed that algorithmic bias can impact student performance evaluations and academic opportunities. Several researchers³⁵ agree

-
- 30 - Noble, *Algorithms of Oppression...*
- Jackson, "Algorithmic Bias."
- Buolamwini and Gebru, "Gender Shades..."
- Baker and Hawn, "Algorithmic Bias in Education."
- Brunner, "Conceptualizing Epistemic Violence..."
- Ferrara, "Fairness and Bias..."
- 31 D. Pilling, "AI in Africa Healthcare Falls Short of Potential," *Financial Times*, 2020.
- 32 Pilling, "AI in Africa..."
- 33 - Buolamwini and Gebru, "Gender Shades..."
- Mavhunga, "Introduction: What Do Science..."
- 34 Baker and Hawn, "Algorithmic Bias in Education."
- 35 - Marco Tulio Ribeiro *et al.*, "Why Should I Trust You?: Explaining the Predictions of Any Classifier," *arXiv (Cornell University)*: 1602.04938v3, August 9, 2016.
- B. H. Zhang *et al.*, "Mitigating Unwanted Biases with Adversarial Learning," AIES '18 Proceedings of the 2018 AAAI/ACM Conference on AI, Ethics, and Society, 335-340. 2018.
- Baker and Hawn, "Algorithmic Bias in Education."

that biased algorithms used in educational platforms can unfairly advantage or disadvantage students based on factors such as socio-economic status or geographical location. In Africa, where there are already existing historical disparities, algorithm bias has huge potential to worsen the situation. In other words, algorithm bias can hinder efforts to promote equality and inclusivity in the education system especially in Africa³⁶. For example, an algorithm used in a standardized test prep platform systematically underestimated the performance of students from rural areas, leading to limited opportunities for them to access higher education³⁷.

The relationship between algorithm bias and public trust in AI technologies among African communities

Algorithmic bias is a significant issue in AI technologies that can greatly impact public trust, particularly among African communities³⁸. Noble³⁹ highlighted that a common area for concern is in the issue of facial recognition technology. Facial recognition algorithms have been found to have higher error rates especially when identifying individuals with darker skin tones, leading to concerns about bias and discrimination. This lack of accuracy can erode public trust in AI technologies, especially in communities where individuals may already face discrimination based on their race or ethnicity⁴⁰. Also, AI technologies have the potential to affect public trust in the criminal justice system among African communities.

36 Mavhunga, "Introduction: What Do Science..."

37 Baker and Hawn, "Algorithmic Bias in Education."

38 Mavhunga, "Introduction: What Do Science..."

39 Noble, *Algorithms of Oppression...*

40 - Noble, *Algorithms of Oppression...*

- Jackson, "Algorithmic Bias."

- Buolamwini and Gebru, "Gender Shades..."

- Baker and Hawn, "Algorithmic Bias in Education."

- Brunner, "Conceptualizing Epistemic Violence..."

- Ferrara, "Fairness and Bias..."

Noble⁴¹ observed that, most AI algorithms used in predictive policing or sentencing decisions have been shown to disproportionately target or harm minority groups. For example, these algorithms can reflect and even reinforce existing biases within the criminal justice system. This can lead to racial disparities in who is targeted for surveillance or more severe punishment which can further erode the trust that African communities have in AI technologies, especially if they have a history of being unfairly targeted or treated by law enforcement⁴².

Obermeyer *et al.*⁴³ observe that in the context of healthcare, algorithmic bias can also impact public trust among African communities especially when AI algorithms used for medical diagnoses are not trained using diverse data sets that adequately represent different racial and ethnic groups. According to Eubanks⁴⁴, this can result in inaccurate diagnosis or treatment recommendations for individuals from these communities, leading to a lack of confidence in AI-powered healthcare solutions. This can potentially result in disparities in health outcomes for African individuals. Additionally, several researchers⁴⁵ have observed the potential of AI on public trust in African communities especially in the context of hiring and employment. AI algorithms used in recruitment or performance evaluation processes can inadvertently perpetuate biases against African individuals. For example, they may favor candidates with specific educational backgrounds or work experiences that do not reflect the diverse talents and skills within African communities. When bias in AI-powered hiring processes results in

41 Noble, *Algorithms of Oppression...*

42 Ribeiro *et al.*, "Why Should I Trust You?..."

43 Obermeyer *et al.*, "Dissecting Racial Bias..."

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

45 - Jackson, "Algorithmic Bias."

- Eubanks, *Automating Inequality: How High-Tech...*

- Erica Sweeney, "AI Can Help Companies Remove Bias in the Hiring Process, Speed Up Recruitment, and Prevent Employees from Quitting, Experts Say," Business Insider, November 9, 2021.

discrimination or exclusion, it can lead to a lack of trust in these technologies among African job seekers and employees.

In the financial services sector, AI technologies also have a high potential to affect public trust in AI technologies among African communities. AI algorithms used by banks and financial institutions for credit scoring are neither transparent nor fair. This can result in African individuals being unfairly denied credit or offered unfavorable terms. This can contribute to the perception of bias and discrimination in AI-powered financial services, leading to decreased trust in these technologies among African communities⁴⁶. Additionally, in the education sector, algorithm bias can impact public trust in AI technologies among African communities. For instance, if the AI algorithms used in educational assessments or personalized learning platforms are not designed to account for cultural differences or linguistic diversity, they could unintentionally put African students at a disadvantage. This can lead to disparities in academic outcomes and reinforce stereotypes about the abilities of African students. As a result, there may be resistance or skepticism toward AI-powered educational tools within African communities, which affects trust in these technologies⁴⁷.

Social media and content recommendation algorithms are also elements that have a huge influence on public trust among African communities⁴⁸. It should be noted that the AI algorithms used by social media platforms prioritize certain types of content or users based on demographic information or historical preferences.

46 - Noble, *Algorithms of Oppression...*

- Jackson, "Algorithmic Bias."

- Buolamwini and Gebru, "Gender Shades..."

- Baker and Hawn, "Algorithmic Bias in Education."

- Brunner, "Conceptualizing Epistemic Violence..."

- Ferrara, "Fairness and Bias..."

47 - Mavhunga, "Introduction: What Do Science..."

- Brunner, "Conceptualizing Epistemic Violence..."

- Baker and Hawn, "Algorithmic Bias in Education."

48 Mbembe, Achille. 2020. "The dividing line: How we represent race in data." London: Open Data Institute.

Inteligencia artificial...

This results in African individuals being exposed to misinformation and harmful stereotypes⁴⁹. The result of this situation is the erosion of trust in the integrity and fairness of these platforms, especially when African users feel marginalized or misrepresented by algorithmic recommendations⁵⁰. In the realm of government services and public policy, algorithm bias can impact public trust in AI technologies among African communities. If AI algorithms used for decision making processes such as resource allocation or entitlement eligibility are not accountable or disparities in service distribution to African individuals may result⁵¹. This can lead to a lack of confidence in government initiatives or programs powered by AI, particularly when African communities feel that their needs are not adequately addressed or considered.

Lastly, Sweeney⁵² noted that in the retail and e-commerce fields, algorithmic bias can shape public trust in AI technologies among communities especially if AI algorithms used by online retailers to make product recommendations or pricing decisions do not reflect consumer preferences or buying patterns. This can result in limited access to diverse goods or unfair pricing practices and can lead to skepticism or apprehension towards AI-powered shopping experiences among African shoppers, impacting trust in these technologies.

49 Latanya Sweeney, "Discrimination in Online Ad Delivery," *SSRN Electronic Journal*, January 29, 2013.

50 Institut Montaigne, "Algorithms: Please Mind the Bias! Report March 2020," 2020.

51 Blessing Mbalaka, "Epistemically Violent Biases in Artificial Intelligence Design: The Case of DALLE-E 2 and Starry AI," *Digital Transformation and Society* 2, no. 4 (2023): 376-402.

52 Sweeney, "Discrimination in Online Ad Delivery."

Challenges faced in Africa regarding algorithmic bias

Many researchers⁵³ have observed that, algorithmic bias is a growing global concern globally, especially in Africa, where unique challenges related to socio-economic and cultural factors contribute to exacerbating the issue. Africa faces several challenges regarding algorithmic bias. These include lack of data representation, socio-economic disparities, cultural biases, limited access to technology, structural inequalities, historical prejudices, lack of diversity in the Tech industry, data privacy concerns, and digital colonialism⁵⁴.

According to Mavhunga⁵⁵, one of the primary challenges regarding algorithmic bias in Africa is the lack of representation in data sets. In many cases, algorithms are trained using data sets that do not adequately represent the diversity of African populations, leading to biased outcomes. For example, facial recognition software has the potential for a much higher error rates when identifying darker-skinned individuals since they were designed using white populations, highlighting the lack of representation in training data⁵⁶.

Socio-economic factors play a significant role in algorithmic bias especially in loan approval algorithms in Africa. It has been established that individuals from lower socio-economic backgrounds may face discrimination due to biased assumptions about their creditworthiness. This perpetuates existing inequalities and

53 - Noble, *Algorithms of Oppression...*

- Jackson, "Algorithmic Bias."

- Buolamwini and Gebru, "Gender Shades..."

- Baker and Hawn, "Algorithmic Bias in Education."

- Mavhunga, "Introduction: What Do Science..."

- Mbembe. "The dividing line: How we represent race in data."

- Brunner, "Conceptualizing Epistemic Violence..."

- Ferrara, "Fairness and Bias..."

54 Mavhunga, "Introduction: What Do Science..."

55 Mavhunga, "Introduction: What Do Science..."

56 Buolamwini and Gebru, "Gender Shades..."

hinders economic opportunities for marginalized groups⁵⁷. According to Buolamwini and Gebru⁵⁸, cultural biases also contribute to algorithmic bias in Africa. Language processing algorithms may struggle to accurately interpret African languages or dialects. This can lead to inaccurate translations, which has serious implications for communication, education, and access to information for African populations.

Buolamwini and Gebru⁵⁹ agree that limited access to technology is a significant challenge in Africa that exacerbates algorithmic bias. Biases in search engine algorithms, for example, can perpetuate stereotypes and misinformation about African countries and populations, further marginalizing them in the digital sphere. This lack of access hinders opportunities for African communities to challenge these biases and advocate for fairer representation⁶⁰. Buolamwini and Gebru⁶¹ agree on the existence of structural inequities, such as lack of internet connectivity which contributes to algorithmic bias in Africa. The most affected are individuals in rural areas with poor internet access. They are at risk of being excluded from online platforms that use algorithms for decision-making which further exacerbates existing inequalities. This digital divide can deepen socio-economic disparities and limit opportunities for marginalized populations.

Similar to the existence of structural inequalities, Buolamwini and Gebru⁶² point also to the existence of historical prejudices and stereotypes. For example, facial recognition algorithms may be more accurate at identifying individuals with lighter skin tones due to historical biases in data collection and representation. This can have serious implications for security and law enforcement

57 Buolamwini and Gebru, "Gender Shades..."

58 Buolamwini and Gebru, "Gender Shades..."

59 Buolamwini and Gebru, "Gender Shades..."

60 Mavhunga, "Introduction: What Do Science..."

61 Buolamwini and Gebru, "Gender Shades..."

62 Buolamwini and Gebru, "Gender Shades..."

practices, leading to unjust outcomes for African populations⁶³. Lack of diversity in the tech industry is another challenge that contributes to algorithm bias in Africa. For example, algorithms developed by homogenous teams may overlook the diverse perspectives and experiences of African populations, resulting in biased outcomes. Increasing diversity and representation in tech companies can help mitigate these biases and ensure fairer algorithmic decision-making processes⁶⁴.

Lastly, the lack of algorithmic accountability poses a significant challenge in addressing bias in Africa. For example, algorithms used in healthcare can result in biased diagnoses or treatment recommendations that harm marginalized groups⁶⁵. Without mechanisms to audit and monitor algorithmic decision-making processes, it is difficult to identify and correct biases, which leads to continued harm to vulnerable populations. Another challenge that exacerbates algorithm bias is digital colonialism, in which global tech giants dominate African markets and control digital infrastructure. For example, algorithms developed by multinational companies may not consider the unique socio-economic and cultural context of African countries, which leads to biased outcomes that prioritize profit over fairness. This neocolonial relationship can perpetuate inequalities and hinder the development of ethical AI solutions in Africa⁶⁶.

CONCLUSIONS

The discussion on AI bias, information fairness and ethics in Africa is complex and multifaceted. It requires a coordinated and collaborative effort from stakeholders across the continent. Policymakers

63 Buolamwini and Gebru, "Gender Shades..."

64 Buolamwini and Gebru, "Gender Shades..."

65 - Mavhunga, "Introduction: What Do Science..."
- Brunner, "Conceptualizing Epistemic Violence..."
- Baker and Hawn, "Algorithmic Bias in Education."

66 Mavhunga, "Introduction: What Do Science..."

Inteligencia artificial...

and practitioners can ensure that AI technologies in Africa are developed and implemented in a fair, equitable, and ethical manner for all users by addressing issues of bias, transparency, ethics, diversity, regulation, awareness, and accountability. Algorithms designed and trained on foreign languages using data sets that are not applicable to the African context are incapable of providing meaningful solutions to the African problems. Only by working together can we harness the potential of AI to drive innovation and development in Africa while avoiding the pitfalls and risks that come with it.

RECOMMENDATIONS

Based on the results of the systematic review, a number of recommendations are made. These recommendations are: Developers, researchers, and policymakers should collaborate to create more inclusive and equitable algorithms

Data sets used to train algorithms should be diversified

African languages must be considered in the design of training of algorithms

A culture of accountability and responsibility must be fostered amongst AI developers and practitioners

Companies and organizations that develop and deploy AI technologies should be held accountable for any harms or biases that result from their products

Stakeholders should work together to promote trust and confidence in AI technologies in Africa and African communities must be engaged to understand their information needs and preferences.

References

Baker, Ryan S., and Aaron Hawn. "Algorithmic Bias in Education." *International Journal of Artificial Intelligence in Education* 32, no. 4 (2021): 1052-92. <https://doi.org/10.1007/s40593-021-00285-9>.

- Barocas, Solon, Moritz Hardt, and Arvind Narayanan. "Fairness and Machine Learning." 2023. <https://fairmlbook.org/pdf/fairmlbook.pdf>.
- Brunner, Claudia. "Conceptualizing Epistemic Violence: An Interdisciplinary Assemblage for IR." *International Politics Reviews* 9, no. 1 (2021): 193-212. <https://doi.org/10.1057/s41312-021-00086-1>.
- Buolamwini, Joy, and Gebru, Timnit. "Gender Shades: Intersectional Accuracy Disparities in Commercial Gender Classification." *Proceedings of Machine Learning Research* 81 (2018): 1-15. <https://proceedings.mlr.press/v81/buolamwini18a/buolamwini18a.pdf>.
- Ephraim, Richard Kobina Dadzie, Gabriel Pezahso Kotam, Evans Duah, Frank Naku Ghartey, Evans Mantiri Mathebula, and Tivani Phosa Mashamba-Thompson. "Application of Medical Artificial Intelligence Technology in sub-Saharan Africa: Prospects for Medical Laboratories." *Smart Health* 33 (September 2024): 100505. <https://doi.org/10.1016/j.smhl.2024.100505>.
- Eubanks V. *Automating Inequality: How High-Tech Tools Profile, Police, and Punish the Poor*. New York: St. Martin's Press. 2018.
- Ferrara, Emilio. "Fairness and Bias in Artificial Intelligence: A Brief Survey of Sources, Impacts, and Mitigation Strategies." *Sci* 6, no. 1, 3 (2023). <https://doi.org/10.3390/sci6010003>.
- Ferrer, Xavier, Tom van Nuinen, Jose M. Such, Mark Coté, and Natalia Criado. "Bias and Discrimination in AI: A Cross-Disciplinary Perspective." *IEEE Technology and Society Magazine* 40, no. 2 (2021): 72-80. <https://doi.org/10.1109/mts.2021.3056293>.
- Hellström, Thomas, Virginia Dignum and Suna Bensch. "Bias in Machine Learning - What is it Good for?" In *Proceedings of the Workshop on Artificial Intelligence and Ethics*. 2020. <https://ceur-ws.org/Vol-2659/hellstrom.pdf>.
- Institut Montaigne. "Algorithms: Please Mind the Bias! Report March 2020." 2020. www.institutmontaigne.org.
- Jackson, Joni R. "Algorithmic Bias." *Journal of Leadership Accountability and Ethics* 15, no. 4 (2018). <https://doi.org/10.33423/jlae.v15i4.170>.

- Jobin, Anna, Marcello Ienca, and Effy Vayena. "The Global Landscape of ai ethics guidelines." *Nature Machine Intelligence* 1, no. 9 (2019): 389-99. <https://doi.org/10.1038/s42256-019-0088-2>.
- Lamé, Guillaume. "Systematic Literature Reviews: An Introduction." In *Proceedings of the Design Society: International Conference on Engineering Design* 1, no. 1 (2019): 1633-42. <https://doi.org/10.1017/dsi.2019.169>.
- Mavhunga, Clapperton Chakanetsa. "Introduction: What Do Science, Technology, and Innovation Mean from Africa?" In *What Do Science, Technology, and Innovation Mean from Africa?*, 1-28. Cambridge, MA: MIT Press. 2017. <https://doi.org/10.7551/mitpress/10769.003.0003>.
- Mbalaka, Blessing. "Epistemically Violent Biases in Artificial Intelligence Design: The Case of DALLE-E 2 and Starry ai." *Digital Transformation and Society* 2, no. 4 (2023): 376-402. <https://doi.org/10.1108/dts-01-2023-0003>.
- Mbembe, Achille. *The dividing line: How we represent race in data*. London: Open Data Institute. 2020. <https://theodi.org/news-and-events/blog/the-dividing-line-how-we-represent-race-in-data/>.
- Mehrabi, Ninareh, Fred Morstatter, Nripsuta Saxena, Kristina Lerman, and Aram Galstyan. "A Survey on Bias and Fairness in Machine Learning." *ACM Computing Surveys* 54, no. 6 (2021): 1-35. <https://doi.org/10.1145/3457607>.
- Noble, Safiya Umoja. *Algorithms of Oppression: How Search Engines Reinforce Racism*. NYU Press. 2018. <https://psycnet.apa.org/record/2018-08016-000>.
- Norori, Natalia, Qiyang Hu, Florence Marcelle Aellen, Francesca Dalia Faraci, and Athina Tzovara. "Addressing Bias in Big Data and ai for Health Care: A Call for Open Science." *Patterns* 2, no. 10 (2021): 100347. <https://doi.org/10.1016/j.patter.2021.100347>.

- Ntoutsis, Eirini, Pavlos Fafalios, Ujwal Gadiraju, Vasileios Iosifidis, Wolfgang Nejdl, Maria-Esther Vidal, Salvatore Ruggieri, et al. "Bias in Data-Driven Artificial Intelligence Systems—An Introductory Survey." *Wiley Interdisciplinary Reviews Data Mining and Knowledge Discovery* 10, no. 3 (2020). <https://doi.org/10.1002/widm.1356>.
- Obermeyer, Ziad, Brian Powers, Christine Vogeli, and Sendhil Mullainathan. "Dissecting Racial Bias in an Algorithm Used to Manage the Health of Populations." *Science* 366, no. 6464 (2019): 447-53. <https://doi.org/10.1126/science.aax2342>.
- Pilling, D. "ai in Africa Healthcare Falls Short of Potential," *Financial Times*, 2020.
- Price, W. Nicholson. "Big Data and Black-Box Medical Algorithms." *Science. Translational Medicine* 10, no. 471 (2018). <https://doi.org/10.1126/scitranslmed.aao5333>.
- Ribeiro, Marco Tulio, Sameer Singh, and Carlos Guestrin. "Why Should I Trust You?: Explaining the Predictions of Any Classifier." *arXiv (Cornell University)*: 1602.04938v3, August 9, 2016. <https://doi.org/10.48550/arxiv.1602.04938>.
- Sweeney, Latanya. "Discrimination in Online Ad Delivery." *SSRN Electronic Journal*, January 29, 2013. <https://doi.org/10.2139/ssrn.2208240>.
- Sweeney, Erica. "ai Can Help Companies Remove Bias in the Hiring Process, Speed Up Recruitment, and Prevent Employees from Quitting, Experts Say." *Business Insider*, November 9, 2021. <https://www.businessinsider.com/how-ai-help-companies-recruit-keep-top-talent-2021-11>.
- Zhang, B. H., Lemoine, B., & Mitchell, M. "Mitigating Unwanted Biases with Adversarial Learning." AIES '18 Proceedings of the 2018 AAAI/ACM Conference on ai, Ethics, and Society, 335-340. 2018. <https://doi.org/10.1145/3278721.3278779>.

La huella de carbono de los algoritmos inteligentes de los macrocentros de datos

PERLA OLIVIA RODRÍGUEZ-RESÉNDIZ

*Instituto de Investigaciones Bibliotecológicas y de la Información,
Universidad Nacional Autónoma de México*

INTRODUCCIÓN

Después del paso de la oralidad a la escritura, de los rollos de papiro al libro, del registro en soportes físicos a los formatos de origen digital y del almacenamiento en bóvedas a los e-archivos digitales, asistimos a una transformación en el modo de registrar, distribuir, conservar y brindar acceso a la información. Este cambio forma parte de la Quinta Revolución Industrial, determinada así por la estrecha relación de los humanos con las máquinas, en particular con la Inteligencia Artificial (AI) y la robótica para realizar actividades de producción y de vida social.¹

Desde la archivística, uno de los impactos de esta Revolución es la histéresis de los soportes y de la comunicación;² es decir, el fenómeno mediante el cual se conservan las propiedades de los contenidos, aunque los soportes originales hayan cambiado. La digitalización fue el primer proceso tecnológico mediante el cual se transfirieron los contenidos registrados en una amplia gama de

1 Pérez-Domínguez, 2024.

2 Maurizio Ferraris, *Documanidad*, Madrid: Alianza Editorial, 2023.

soportes analógicos (papel, cintas magnéticas, discos ópticos, entre otros) a formatos digitales y datos. Después, la generación automatizada de documentos de origen digital propició el desarrollo de grandes volúmenes de datos. El archivo diseñado para proteger los documentos en un mundo físico devino en el e-archivo: un constructo social virtual que es un componente esencial del ecosistema digital en el siglo XXI.³

Es tal la relevancia del archivo en la actualidad que el filósofo italiano Maurizio Ferraris⁴ ha denominado a esta era como *documедial*. La era *documедial* contrasta con la era analógica, en la que primero se transmitía la información y, en algunos casos, se registraba como archivo. En la era *documедial*, caracterizada por un ecosistema digital, primero se registra toda forma de pensamiento, expresión artística, hecho o idea y, con ello, se crea un archivo que después se comunica. Aunado a esto, los procesos de creación, registro y archivado de todo tipo de contenidos (sonoros, textuales, audiovisuales, entre otros) y de los datos de las personas que consultan la información se han automatizado.

Hasta el siglo pasado, la producción de documentos estuvo centralizada por la industria editorial, las instituciones públicas, las universidades y los medios de comunicación, como periódicos, radio y televisión, entre otros. Hoy en día somos parte de una humanidad “cuya función es la de producir documentos sobre sí misma, documentos menores y mínimos, pero de gran utilidad como son los de nuestras navegaciones en la web, o de documentos mayores y máximos, como son las producciones del espíritu y la cultura”.⁵

La creación, publicación, distribución y acceso a este gran volumen de documentos se lleva a cabo mediante la nube —metáfora para nombrar al conjunto de servidores conectados para gestionar

3 Perla Rodríguez-Reséndiz, “El e-archivo sustentable: una propuesta de preservación digital ante el cambio climático”, *Investigación Bibliotecológica: archivonomía, bibliotecología e información* 38, núm. 101 (2024): 57–71.

4 Maurizio Ferraris, *Documanidad*.

5 Maurizio Ferraris, *Documanidad*. 14.

información—. Los servidores que habilitan la búsqueda y recuperación de la información se ubican en centros de datos, es decir, en instalaciones físicas donde se almacenan, gestionan y distribuyen datos. Los datos masivos a gran escala se gestionan en los macrocentros de datos, también conocidos como centros de datos a hiperescala. En estas instalaciones se almacenan cantidades ingentes de datos y se utilizan herramientas de IA para la gestión de información. Son ejemplos de macrocentros de datos Amazon, Google, Microsoft, entre otros.

LA PROLIFERACIÓN DE LOS DATOS

Es posible, como señalan Noam Chomsky, Ian Roberts y Jeffrey Watumull, que estemos ante una escena tragicómica por destinar tanto dinero y tanta atención a programas de inteligencia artificial como Chat GPT de Open AI, Bard de Google y Sydney de Microsoft, que son triviales si se comparan con la mente humana.⁶ Y es que, desde la perspectiva de estos investigadores, estas herramientas de inteligencia artificial son una forma de plagio, apatía y omisión. En otros términos, son herramientas amorales, que expresan una falsa ciencia y una incompetencia lingüística.⁷ Esta crítica se formula en un momento donde la fascinación por el uso de la IA parece inevitable.

A pesar de esta perspectiva crítica y escéptica, es necesario comprender y adentrarse en esta tecnología que domina la búsqueda, creación y distribución de información en sus más diversos formatos digitales, y que, como sucede en otros sectores de la sociedad, también está afectando y modificando el trabajo en bibliotecas y archivos. Es tal la impronta que tiene la IA en la actualidad que se podría considerar que está rediseñando nuestras sociedades.

6 Noam Chomsky *et al.*, “La falsa promesa del ChatGPT”, *Palabra Pública*, 3 de mayo de 2023.

7 Noam Chomsky *et al.*, “La falsa promesa del ChatGPT”.

Inteligencia artificial...

Como se observa en la Tabla 1, el crecimiento de la información es exponencial, acelerado y continuo.

Hasta 2019 el crecimiento había sido progresivo, pero con la pandemia del covid-19 el incremento en la producción de información digital se aceleró. Y este proceso coincidió con la irrupción de la inteligencia artificial, el internet de las cosas, la realidad aumentada, el metaverso, la salida de la computación cuántica de los laboratorios de investigación, entre otras tecnologías que se insertan con celeridad en nuestra vida cotidiana.

Es tal el volumen de datos que se generan y que se van almacenando en cada momento, que la recuperación de esta información sería impensable sin el uso de algoritmos inteligentes. Este abismal cúmulo de información es lo que conocemos como datos masivos o *big data* y son la materia prima con la cual trabaja la inteligencia artificial.

Se ha proyectado que para 2025 al menos la mitad de la información producida será almacenada en la nube y que con esos datos se realizarán más de 150 mil millones de interacciones a través de dispositivos tecnológicos conectados a la web.⁸ Esto debido a que a través de la nube se accede a servicios remotos para la educación, la capacitación y el trabajo, así como a los servicios de información que apoyan tareas cotidianas como seleccionar la mejor y más rápida ruta para llegar a un lugar, el acceso a información personalizada, la traducción de textos, entre otras muchas acciones. Empresas como Google y Microsoft han incluido las funcionalidades de búsqueda en los editores de texto y el correo electrónico. Asimismo, es necesario señalar que el uso de la voz es ya una práctica que sustituye el empleo de ciertos comandos para realizar tareas informáticas. Estas actividades se desarrollan en la nube, que comprende millones de computadoras dispuestas en diversos centros de datos donde se almacenan, gestionan y procesan grandes volúmenes de contenidos digitales. A estas instalaciones se les denomina macrocentros de datos.

8 David Reinsel *et al.*, *The Digitization of the World From Edge to Core*, IDC, 2018.

Se han identificado 12 mil centros de datos en todo el mundo.⁹ La mayoría de estos están instalados en Estados Unidos, mientras que el segundo lugar lo ocupa Alemania, seguida de Reino Unido, China y Canadá.¹⁰ Empresas como Google, Amazon, Microsoft e IBM, entre otras, invierten grandes sumas de dinero en la construcción y mantenimiento de sus centros de datos.¹¹

Tabla 1. Crecimiento de la información de 2010 a 2025.

Año	Datos generados	Cambios en relación con los años previos
2010	2 zettabytes	-
2011	5 zettabytes	↑ 3 zettabytes
2012	6.5 zettabytes	↑ 1.5 zettabytes
2013	9 zettabytes	↑ 2.5 zettabytes
2014	12.5 zettabytes	↑ 3.5 zettabytes
2015	15.5 zettabytes	↑ 3 zettabytes
2016	18 zettabytes	↑ 2.5 zettabytes
2017	26 zettabytes	↑ 8 zettabytes
2018*	33 zettabytes	↑ 7 zettabytes
2019*	41 zettabytes	↑ 8 zettabytes
2020*	64.2 zettabytes	↑ 23.2 zettabytes
2021*	79 zettabytes	↑ 14.8 zettabytes
2022*	97 zettabytes	↑ 18 zettabytes
2023*	120 zettabytes	↑ 23 zettabytes
2024*	147 zettabytes	↑ 27 zettabytes
2025*	181 zettabytes	↑ 34 zettabytes

Fuente: Statista 2024.

9 Cloudscene, Data Center Market Overview.

10 Cloudscene, Data Center Market Overview.

11 Stanko P. Stankov, "Cooling Technologies of Data Centers", *Annals of Faculty Engineering Hunedoara – International Journal of Engineering XX*, núm. 1 (2022): 1-8.

Inteligencia artificial...

El diseño de los modelos de las inteligencias artificiales depende del acceso a colosales volúmenes de datos de todo tipo que se preservan en macrocentros de datos. Por lo tanto, los macrocentros de datos son actores esenciales en el desarrollo y auge de la inteligencia artificial. Sin embargo, no todos los datos que se preservan pueden ser utilizados para el desarrollo eficiente de soluciones inteligentes. La calidad de los datos es un aspecto esencial en el desarrollo de modelos de IA. Por ello, las empresas que desarrollan esta tecnología prefieren hacer convenios de colaboración con medios de comunicación. Son ejemplo de ello los acuerdos que se anunciaron entre Open AI con Prisa Radio y Le Monde.¹²

EL CONSUMO ENERGÉTICO Y DE AGUA EN LOS MACROCENTROS DE DATOS

El ciclo de vida de los sistemas de inteligencia artificial, desde su creación hasta su desecho, puede ser analizado bajo diferentes perspectivas como son las formas de explotación del trabajo humano y de recursos naturales hasta las concentraciones de información y poder corporativo.¹³

En este capítulo, la atención se enfoca en el aspecto medioambiental porque los centros de datos que procesan ingentes volúmenes de información mediante la inteligencia artificial inciden en la huella de carbono. De igual modo, este tema cobra relevancia porque estamos ante una crisis humanitaria derivada del calentamiento global como efecto del cambio climático.

La huella de carbono que deja la IA inicia con la producción de los dispositivos tecnológicos. Cada objeto de la extensa red de los sistemas de inteligencia artificial, desde las redes de *router* hasta las baterías que se usan para los centros de datos se construyen

12 “OpenAI anuncia un acuerdo con Prisa Media y Le Monde”, *El País*, 13 de marzo de 2024.

13 Kate Crawford, *Atlas of AI: Power, Politics and the Planetary Costs of Artificial Intelligence*, New Haven y Londres: Yale University Press, 2021.

empleando elementos como el litio, el cobre o las tierras raras, que requieren billones de años para su formación dentro de la Tierra.¹⁴

Las minas de litio de Nevada, el Salar de Uyuni en Bolivia, el Congo, Mongolia, Indonesia y el este de los desiertos de Australia son lugares donde nace la inteligencia artificial en la gran geografía industrial de extracción. Sin los minerales de estos lugares, entre otros, la IA simplemente no funcionaría. Además, la producción y el traslado de la tecnología mediante barcos con grandes contenedores también tiene un impacto en el medio ambiente.¹⁵

Por otro lado, se emplea energía y agua para mantener en operación los servidores en los centros de datos y esto tiene un impacto en el medio ambiente. Este consumo inicia durante el entrenamiento de los algoritmos con millones de ejemplos para que puedan establecer patrones y formular predicciones.

La mayoría de los centros de datos utilizan procesadores avanzados llamados GPU que necesitan muchísima energía para funcionar, unas cinco veces más que las CPU (los procesadores convencionales). El entrenamiento de los grandes modelos de lenguaje requiere de decenas de miles de GPU, que necesitan operar día y noche durante semanas o meses.¹⁶

Se carece de datos precisos en relación con el consumo de energía en los macrocentros de datos. Las estimaciones formuladas advierten un incremento significativo de la energía. El Centro de Eficiencia Energética de Copenhague estima que para el año 2030, las tecnologías de la información y las comunicaciones utilizarán un 20% de la energía mundial. En cualquier caso, las proyecciones demuestran que la industria de la información ha alcanzado a la aviación en la generación de la huella de carbono.¹⁷ Se ha estimado

14 Kate Crawford, *Atlas of AI...*, .

15 Kate Crawford, *Atlas of AI...*, .

16 Manuel G. Pascual, “El sucio secreto de la Inteligencia Artificial”, *El País*, 22 de marzo de 2023.

17 Manuel G. Pascual, “El sucio secreto de la Inteligencia Artificial”.

que los centros de datos generan entre el 1% y el 5% de los gases de efecto invernadero en el mundo.¹⁸

En cuanto al uso del agua, es necesario señalar que las naves industriales donde se ubican los centros de datos resguardan hileras de *racks* con procesadores que trabajan día y noche durante todo el año y emiten mucho calor. El agua se utiliza para mantener en operación los servidores todo el tiempo mediante dos procesos esenciales: la refrigeración y la humidificación. La temperatura debe controlarse para que no se estropeen estos equipos. Se usa agua que se pulveriza para controlar la temperatura. De acuerdo con información periodística, en 2022, Google incrementó en 20% su consumo de agua y Microsoft lo hizo en un 34%. En este reporte no se incluye el agua utilizada para la generación de electricidad que alimenta los servidores ni tampoco la utilizada en los procesos de fabricación de *hardware*. Amazon no proporcionó datos. Meta gastó ese año un 2.7% más, es decir, 36.7%.¹⁹ La mayor parte del agua utilizada se evapora y no se puede reutilizar.

En un estudio prospectivo se advierte que la demanda global para la operación de la IA en 2027 podría significar el consumo de 4,2 y 6,6 billones de litros, lo que equivale a la mitad de agua usada cada año en Reino Unido; y de 85 y 134 TWh de electricidad.²⁰

Por su parte, cuando la IA se ofrece para su uso público crece el impacto sobre el medio ambiente porque con cada búsqueda que se realiza se consume más energía, ya que la IA se compone de sistemas complejos que bucean en millones de páginas web y usan arquitecturas basadas en redes neuronales, con millones de parámetros que hay que entrenar, advierte el doctor Carlos Gómez de la Universidad de La Coruña.²¹ Además, se estima que la

18 Brightlio, <https://brightlio.com/>.

19 Manuel G. Pascual, “La inteligencia artificial se bebe miles de millones de litros de agua”, *El País*, 12 de noviembre de 2023.

20 Pengfei Li *et al.*, “Making AI Less Thirsty: Uncovering and Addressing the Secret Water Footprint of AI Models”, arXiv, 2023.

21 Karen Hao, “Entrenar una sola IA emite tanto CO2 como cinco coches en su vida útil”, *MIT Technology Review*, 14 de junio de 2019.

formulación de 100 preguntas a Chat GPT requiere de un litro de agua. Con este dato se puede dimensionar la necesidad creciente de consumo de agua derivada del uso, cada vez más creciente, de la IA.

Chat GPT tiene 13 millones de usuarios y Bing maneja medio millón de búsquedas cada día.²² Estas cifras aumentarán porque la demanda de uso de la IA se incrementará. En consecuencia, a mediano y largo plazo, para satisfacer la demanda en el uso de la IA, será necesario incrementar la capacidad de cómputo. Situación que conlleva a un mayor consumo de energía y agua.

CONCLUSIONES

¿Qué se puede hacer ante este panorama?

Desde la perspectiva de los archivos y bibliotecas, la inteligencia artificial ofrece oportunidades y cambios. La incorporación de algoritmos inteligentes en la gestión de datos puede mejorar los metadatos para la identificación y recuperación de contenidos digitales, promover tanto el acceso como el reúso de datos e incrementar la visibilidad a través de diferentes plataformas y *agregadores* de contenidos. Sin embargo, no es posible ignorar los datos relacionados con el impacto que tiene la IA, desde la extracción de minerales y tierras raras para su fabricación hasta el uso cotidiano en la búsqueda de información. La IA ahondará el impacto que producen las tecnologías de la información sobre el medio ambiente.

Es un hecho que no podemos adoptar una perspectiva radical que nos haga quedarnos al margen del uso de la IA. No obstante, es importante reconocer el impacto social y medioambiental de esta tecnología para incentivar la investigación científica y las alternativas sustentables. Más allá de las iniciativas que los macrocentros

22 Chris Stokel-Walker, "The Generative AI Race Has a Dirty Secret", *Wired*, 18 de febrero de 2023.

de datos han puesto en marcha para aminorar la huella de carbono, es importante señalar que el tema no ha sido tratado con el rigor ni la importancia que tiene para la humanidad. Bajo esta perspectiva, los profesionales de la información, desde cualquiera de sus ámbitos de trabajo, deben conocer y participar en señalar el problema, así como en proponer soluciones. Esto con el fin de crear alternativas donde los desarrollos de la IA estén a disposición de la sociedad y aminoren su impacto en el medioambiente. Asimismo, quienes trabajan en la preservación del patrimonio digital deben construir una cultura de cuidado del medio ambiente y crear e-archivos sustentables.²³

Finalmente, sería deseable que la nueva Ley de Inteligencia Artificial aprobada por la UE, que entrará en vigor en 2026, además de poner atención en temas como la afectación a los derechos individuales y al desarrollo económico, la vigilancia biométrica, la seguridad, el uso de datos personales para la vigilancia de personas y otras cuestiones ligadas a derechos fundamentales,²⁴ también considere el impacto de la IA en el cambio climático. Este es un tema que cada vez más debe insertarse en las agendas de discusión y en las nuevas leyes que regulen la operación de la IA.

23 Perla Rodríguez-Reséndiz, “El e-archivo sustentable: una propuesta de preservación digital ante el cambio climático”, *Investigación Bibliotecológica: archivonomía, bibliotecología e información* 38, núm. 101 (2024): 57–71.

24 Virginia Bazán-Gil y Carmen Pérez Cernuda, “Metadato automático en el archivo de RTVE: un enfoque de innovación tecnológica”, *Patrimonio Audiovisual: narrativas, contenidos y formatos*, editado por Ana Mancho, 180-200, Madrid: Dykinson, 2024.

BIBLIOGRAFÍA

- Bazán-Gil, Virginia y Carmen Pérez Cernuda. “Metadadado automático en el archivo de RTVE: un enfoque de innovación tecnológica”. En *Patrimonio Audiovisual: narrativas, contenidos y formatos*, editado por Ana Mancho, 180-200. Madrid: Dykinson, 2024. <https://www.dykinson.com/libros/patrimonio-audiovisual-narrativas-contenidos-y-formatos/9788411707657/>.
- Brightlio. Sitio web de Brightlio. Consultado el 10 de octubre de 2024. <https://brightlio.com/data-center-stats/#pp-toc-huja89r-ng0qk-anchor-0>.
- Chomsky, Noam, Ian Roberts y Jeffrey Watumull. “La falsa promesa del ChatGPT”. *Palabra Pública*, 3 de mayo de 2023. <https://palabrapublica.uchile.cl/la-falsa-promesa-del-chatgpt/>.
- Cloudscene. Data Center Market Overview. Sitio web de Cloudscene. Consultado el 26 de junio de 2024. <https://cloudscene.com/region/datacenters-in-north-america>.
- Crawford, Kate. *Atlas of AI: Power, Politics and the Planetary Costs of Artificial Intelligence*. New Haven y Londres: Yale University Press, 2021.
- Domínguez, Luis. “Las principales tecnologías de la era de la industria 5.0”. *Revista Ingenio* 21, núm. 1 (2024): 60-70. <https://doi.org/10.22463/2011642X.4352>.
- Ferraris, Maurizio. *Documanidad*. Madrid: Alianza Editorial, 2023.
- Hao, Karen. “Entrenar una sola IA emite tanto CO2 como cinco coches en su vida útil”. *MIT Technology Review*, 14 de junio de 2019. <https://www.technologyreview.es/s/11222/entrenar-una-sola-ia-emite-tanto-co2-como-cinco-coches-en-su-vida-util>.
- Li, Pengfei, Jianyi Yang, Mohammad A. Islam y Shaolei Ren. “Making AI Less Thirsty: Uncovering and Addressing the Secret Water Footprint of AI Models”. *arXiv*, 2023. arXiv:2304.03271. <https://arxiv.org/abs/2304.03271>.

Inteligencia artificial...

- “OpenAI anuncia un acuerdo con Prisa Media y Le Monde”. *El País*, 13 de marzo de 2024. <https://elpais.com/comunicacion/2024-03-13/open-ia-anuncia-un-acuerdo-con-prisa-media-y-le-monde.html>.
- Pascual, Manuel G. “El sucio secreto de la Inteligencia Artificial”. *El País*, 22 de marzo de 2023. <https://elpais.com/tecnologia/2023-03-23/el-sucio-secreto-de-la-inteligencia-artificial.html>.
- Pascual, Manuel G. “La inteligencia artificial se bebe miles de millones de litros de agua”. *El País*, 12 de noviembre de 2023. <https://elpais.com/tecnologia/2023-11-14/la-inteligencia-artificial-se-bebe-miles-de-millones-de-litros-de-agua.html>.
- Reinsel, David, John Gantz y John Rydning. *The Digitization of the World From Edge to Core*. IDC, 2018. <https://www.seagate.com/files/www-content/our-story/trends/files/idc-seagate-data-age-whitepaper.pdf>.
- Rodríguez-Reséndiz, Perla. “El e-archivo sustentable: una propuesta de preservación digital ante el cambio climático”. *Investigación Bibliotecológica: archivonomía, bibliotecología e información* 38, núm. 101 (2024): 57–71. <https://doi.org/10.22201/iibi.24488321xe.2024.101.58922>.
- Stankov, Stanko P. “Cooling Technologies of Data Centers”. *Annals of Faculty Engineering Hunedoara – International Journal of Engineering* XX, núm. 1 (2022): 1-8.
- Statista. “Volume of data/information created, captured, copied, and consumed worldwide from 2010 to 2020, with forecasts from 2021 to 2025”. Statista. <https://www.statista.com/statistics/871513/worldwide-data-created/>.
- Stokel-Walker, Chris. “The Generative AI Race Has a Dirty Secret”. *Wired*, 18 de febrero de 2023. <https://www.wired.com/story/the-generative-ai-search-race-has-a-dirty-secret/>.

LA IA, DESINFORMACIÓN
Y EL ROL DE LOS ACTORES DEL CONOCIMIENTO

Cuestionamientos de los actores del sector editorial hacia el uso de la Inteligencia Artificial y sus herramientas

JENNY TERESITA GUERRA GONZÁLEZ

*Instituto de Investigaciones Bibliotecológicas y de la Información,
Universidad Nacional Autónoma de México*

La vieja cantinela sobre la expansión de los sistemas de IA y la automatización de los procesos dice que vivimos en una época de colaboración benéfica entre los seres humanos y la IA. Pero esta colaboración no se negocia de manera justa. Los términos descansan en una asimetría de poder significativa; de hecho, ¿existe siquiera la posibilidad de no colaborar con sistemas algorítmicos?

Kate Crawford en Atlas de inteligencia artificial. Poder, política y costos planetarios (2022)

INTRODUCCIÓN

Desde mediados de la década del 2000, la Inteligencia Artificial (IA) se ha expandido rápidamente como campo académico que abarca áreas como el aprendizaje basado en máquinas, las redes neuronales, el cómputo cognitivo y el procesamiento del lenguaje natural, así como en la industria. Si bien inició como uno de los grandes proyectos públicos del siglo XX,

Inteligencia artificial...

hoy en día solo un pequeño número de compañías tecnológicas poderosas —Open IA, Google, Tesla, NVIDIA, entre otras— han implementado sistemas de IA a escala global. Estos sistemas son aclamados como similares o incluso superiores a la inteligencia humana, ya que permiten automatizar procesos que sustituyen la mano de obra, realizar tareas específicas —como jugar ajedrez con gran éxito—, efectuar diagnósticos de enfermedades por medio del procesamiento de datos médicos, etcétera. Un ejemplo claro es el de ChatGPT (desarrollo de IA generativa de Open IA) que a partir de su lanzamiento en noviembre de 2022 se convirtió en la aplicación de internet con el crecimiento más rápido de la historia, pues en dos meses llegó a tener más de 100 millones de usuarios activos.¹ Empero, la IA no es un dominio puramente técnico, sino que depende por completo de un conjunto muy vasto de estructuras políticas y sociales que reflejan y producen al mismo tiempo relaciones sociales y entendimientos del mundo: los denominados modelos culturales. Su diseño y avances se inscriben en lo que el escritor argentino Martín Caparrós ha denominado “la búsqueda de la máquina única, capaz de responder a casi todas las necesidades”.²

El contexto actual en el que se implementa la IA es descrito por estudiosas como Kate Crawford con el adjetivo de “riesgoso” debido a que las herramientas que se van desarrollando se vuelven más invasivas. Esto acontece porque la IA es un producto del “aceleracionismo tecnológico” y de un régimen capitalista de vigilancia y militarización. Crawford apuntala estas afirmaciones:

La Inteligencia Artificial no es una técnica computacional neutral que tome determinaciones sin una dirección humana. Sus sistemas están integrados en mundos sociales, políticos, culturales y

1 Verónica Smink, “Las 3 etapas de la Inteligencia Artificial: en cuál estamos y por qué muchos piensan que la tercera puede ser fatal”, *BBC News Mundo*, 29 de mayo de 2023.

2 Martín Caparrós, *El mundo entonces. Una historia del presente*, Ciudad de México: Penguin Random House, 2023: 282

económicos, delineados por humanos, instituciones e imperativos que determinan lo que hacen y cómo lo hacen. Están diseñados para discriminar, amplificar jerarquías y codificar clasificaciones estrechas. Cuando son aplicables en contextos sociales como la vigilancia policial, el sistema judicial, la salud y la educación, pueden reproducir, optimizar y amplificar las desigualdades estructurales existentes. Esto no es casual. Los sistemas de IA están contruidos para ver e intervenir en el mundo de maneras que benefician principalmente a los Estados, las instituciones y las corporaciones a los que sirven. En ese sentido, los sistemas de IA son expresiones de poder que surgen de fuerzas económicas y políticas más amplias, creadas para aumentar las ganancias y centralizar los poderes de quienes las esgrimen. Pero la historia de la IA no se suele contar de esta manera.³

En ese orden de ideas, la intención de este trabajo es estudiar los cuestionamientos que diferentes actores del sector editorial están planteando en relación con el uso de la IA y sus herramientas en un ámbito cultural y económico que forma parte de las industrias culturales y creativas. Tradicionalmente, estas industrias adoptan las innovaciones tecnológicas, pero a mediano plazo las descartan si representan un costo difícil de asumir o una serie de procesos complejos de gestionar. Cabe señalar que el sector editorial no es uno solo, sino que está integrado por diferentes subsectores con características y alcances diversos, trátase de publicaciones académico-científicas, didácticas, de divulgación, especializadas, infantojuveniles o de ocio. A estos aspectos debe agregarse que este ámbito enfrenta las consecuencias de eventos como la pandemia por covid-19; el aumento de los costos del papel y la impresión (en 2023 los precios del papel aumentaron un 65% a consecuencia de la escasez mundial de este insumo durante 2020); la evolución de las preferencias de los consumidores y el plagio de contenido (plataformas de autopublicación como Amazon permiten que el plagio

3 Kate Crawford, *Atlas de inteligencia artificial. Poder, política y costos planetarios*, Ciudad Autónoma de Buenos Aires: Fondo de Cultura Económica, 2022, 321

Inteligencia artificial...

pase desapercibido al facilitar el acceso a herramientas que hacen usos no autorizados de obras originales).⁴

PERSPECTIVAS TEÓRICO-ANALÍTICAS SOBRE EL USO DE LA IA

En esta sección abordaremos el uso de la IA de modo crítico a través de tres categorías teórico-analíticas: la inevitabilidad tecnológica, la necesaria limitación en su uso y la insuficiencia en pensar alcances amplios de la ética de la IA. En primer lugar, definiremos la *inevitabilidad tecnológica* como el mito que afirma que los usos de las innovaciones y desarrollos tecnológicos son inevitables. Su argumento es “si se puede hacer, se hará”. La *inevitabilidad tecnológica*, en tanto concepto, comenzó a ganar popularidad en la segunda mitad del siglo XX, si bien tiene sus antecedentes en el enfoque del determinismo tecnológico que se desarrolló en las primeras etapas de la Revolución Industrial. Este sostenía que la tecnología sigue un curso inevitable y que sus efectos en la sociedad no pueden ser controlados. Posteriormente, en los años 60, el concepto comenzó a consolidarse a consecuencia de la expansión de la informática, los procesos de automatización industrial y el desarrollo de las comunicaciones. Filósofos y sociólogos como Jacques Ellul (*La technique ou l'enjeu du siècle*/1954), Marshall McLuhan (*Understanding Media: The Extensions of Man*/1964) y Alvin Toffler (*Future Shock*/1970), escribieron sobre el poder transformador de la tecnología, y en particular sobre su capacidad para moldear las estructuras sociales y económicas, argumentando que las sociedades modernas eran incapaces de resistir o detener el progreso tecnológico.

A finales del siglo XX y comienzos del XXI, esta categoría adquiere un mayor peso con la masificación del internet y la globalización digital. En ese sentido, resultan paradigmáticas las palabras

4 Deployteq, “Challenges Facing the Publishing and Media Industry in 2024”, 9 de marzo de 2024.

de Andrew Grove, CEO de Intel, quien en 1998 afirmaba: “Technological change and its effects are inevitable. Stopping them is not an option”, en un contexto en el que se abría el debate en torno a la clonación humana, ya que en 1996 investigadores del Instituto Roslin de la Universidad de Edimburgo consiguieron clonar a la oveja Dolly.⁵ Se sumarían en esta línea los trabajos de los tecnólogos Raymond Kurzweil y Kevin Kelly quienes han defendido la idea de que la evolución tecnológica, especialmente en campos como la inteligencia artificial y la biotecnología, sigue una trayectoria ineludible que las sociedades simplemente deben aceptar y a las cuales deben adaptarse.

En el ámbito de la edición, la inevitabilidad tecnológica ha sido abordada en relación a la digitalización de contenidos, la automatización y la distribución de obras a través de plataformas en línea. Ya en 2008, Clay Shirky, profesor y consultor norteamericano, publicaba *Here Comes Everybody*, libro en el que discute cómo la digitalización, los blogs y las redes sociales volvieron inevitable la disrupción de las formas tradicionales de publicación y distribución de información. Shirky asegura que la tecnología ha permitido a cualquier persona convertirse en un editor o productor de contenido, lo que ha fragmentado el monopolio de la industria editorial tradicional y de medios sobre la creación y difusión de los contenidos informativos. Años más tarde, Eli Pariser analizaba en su libro *El filtro burbuja. Cómo la red decide lo que leemos y lo que pensamos* cómo las plataformas digitales, que han transformado la industria editorial, han introducido problemas como la creación de burbujas de información filtrada por algoritmos, lo que conlleva nuevos desafíos, como la segmentación de audiencias y la reducción de la diversidad informativa. Otros argumentos populares durante los inicios del siglo XXI, como el de la sustitución del libro impreso por el digital, se alinean también con esta retórica de obsolescencia y desplazamiento tecnológico que, con el paso de los años, ha sido desacreditada. (*imagen 1*).

5 Dave Karpf, “The Myth of Technological Inevitability”, *The Future, Now and Then* (blog), 24 de enero de 2024.

Inteligencia artificial...

Imagen 1. Caricatura de Nacho Velmar que alude a la vitalidad del libro impreso



Fuente: Perfil de Facebook del proyecto "Un libro para leer" (2024).

La revisión crítica de la inevitabilidad tecnológica aspira a estudiar con detenimiento el por qué esta perspectiva ignora el papel de las decisiones políticas, económicas y culturales que influyen en el desarrollo tecnológico. Particularmente busca enfatizar y hacer de conocimiento público que el mito de la inevitabilidad tecnológica es una parte esencial del proyecto político-libertario de las grandes empresas tecnológicas de Silicon Valley que no permiten que haya alternativas de otros sistemas y cosmovisiones técnicas, sosteniendo que el progreso tecnológico debe ser impulsado sin restricciones por parte de gobiernos o regulaciones.

La segunda categoría teórico-analítica a analizar en este apartado es la de la *necesaria limitación en el uso de la IA*. Para ello hemos de partir del hecho de que la IA implica concentraciones masivas de poder corporativo y geopolítico, lo que crea nuevas formas de desigualdad económica y social, origina monopolios de datos globales, produce asimetrías de poder entre países desarrollados y en desarrollo, por mencionar solo algunas de sus consecuencias. Teóricos como el filósofo francés Éric Sadin señalan que en lugar de preguntarse dónde se aplicará la IA, simplemente porque puede aplicarse, el énfasis debería estar puesto en por qué tendría que aplicarse: ¿por qué este futuro es deseable para la humanidad? Sádin establece la hipótesis de que el fenómeno de la inteligencia artificial es un cambio de estatuto de las tecnologías digitales: de ser prótesis acumulativas e intelectivas que permiten el almacenamiento, la indexación y el tráfico veloz de información, se han convertido en entidades de las que se espera que enuncien una verdad a partir de la interpretación automatizada de diversas situaciones.⁶

El filósofo francés, advierte sobre las consecuencias de la no limitación en el uso de la IA al señalar que:

La humanidad se está dotando a grandes pasos de un órgano de prescindencia de ella misma, de su derecho a decidir con plena conciencia y responsabilidad las elecciones que la involucran. Toma forma un estatuto antropológico y ontológico inédito que ve cómo la figura humana se somete a las ecuaciones de sus propios artefactos con el objetivo prioritario de responder a intereses privados y de instaurar una organización de la sociedad en función de criterios principalmente utilitaristas.⁷

Al preguntarnos por qué usar inteligencia artificial, debemos cuestionar la idea de que todo tiene que estar sujeto a la lógica de la

6 Éric Sadin, *La Inteligencia Artificial o el Desafío del Siglo. Anatomía de un Antibumanismo Radical*, Ciudad Autónoma de Buenos Aires: Caja Negra Editora, 2020.

7 Éric Sadin, *La Inteligencia Artificial...*

Inteligencia artificial...

predicción estadística y la acumulación de ganancias. Donna Haraway ha denominado lo anterior como la “informática de la dominación”, cuyo punto culminante es la aplicación masiva de la IA que promueve tanto la desinformación como la manipulación de la opinión pública. Debido a la facilidad con la que es posible manipular imágenes, videos y texto con el uso de herramientas de IA, el Foro Económico Mundial advirtió en 2020 sobre el fenómeno de “desinformación a gran escala”, misma que representa una amenaza para el conocimiento, la democracia y la estabilidad social al socavar la confianza pública, lo que torna imperiosa su regulación.

Las y los ilustradores, actores esenciales en el sector de la edición, han puesto en entredicho el uso indiscriminado de la IA en la realización de portadas, la imitación de estilos artísticos individuales para crear imágenes, personajes y escenarios para videojuegos, y el empleo de los trabajos sin permiso de sus autoras y autores para continuar el entrenamiento de las IA generativas. En enero de 2024, el historietista español David López hizo de conocimiento público sus hallazgos en el uso de la IA en la portada del libro *Juana de Arco* de Katherine J. Chen, editada por Destino (Grupo Planeta), esto tras una evaluación profesional de los errores en los que incurrió la herramienta (*imagen 2*). Si bien la editorial negó la acusación y sostuvo que había una mano humana tras el trabajo de ilustración, las librerías La Llama Store y La Vorágine comunicaron que devolverían todos los ejemplares del citado volumen; además de comprometerse a repetir la misma operación en caso de recibir más libros donde la IA pudiera haber intervenido en la confección de la cubierta.⁸

8 Laura García Higuera, “Los ilustradores exigen a las editoriales que dejen de usar IA en las portadas: ‘El robo más grande de la historia’”, *El Diario*, 14 de marzo de 2024.

Imagen 2. Análisis del empleo de IA en el diseño de la portada de Juana de Arco de Katherine J. Chen llevado a cabo por el historietista David López



Fuente: García Higuera (2024).

La falta de reflexión sobre los alcances amplios de la ética de la IA es la última categoría teórico-analítica a exponer en este apartado. Recurrimos a las investigaciones de la estudiosa de los medios en The New School de Nueva York, Shannon Mattern, quien afirma que la atención suele ponerse en los fines éticos de la IA, sin evaluar los medios éticos para aplicarlos. La académica destaca la necesidad de prestar atención no solo a los resultados que se desean alcanzar con la IA y otras tecnologías digitales, sino también a los procesos y las prácticas a través de los cuales aquellas se diseñan y despliegan.⁹ A diferencia de la medicina, o el derecho, la IA no tiene una estructura o normas profesionales gubernamentales; no hay definiciones acordadas ni metas para el campo, o protocolos estándares para hacer cumplir las prácticas éticas. Los marcos éticos autorregulados permiten a las compañías elegir cómo poner en práctica las tecnologías y, por extensión, decidir qué significa una IA ética para el resto del mundo. Las empresas tecnológicas rara vez reciben grandes sanciones económicas cuando sus sistemas de IA violan la ley, y hay aún menos consecuencias cuando se violan sus principios éticos. Como resultado, la ética es necesaria pero no suficiente para abordar las preocupaciones principales. Para entender lo que está en juego tenemos que enfocarnos menos en la ética y aún más en el poder.

En el vasto sector editorial, los actores del ámbito académico-científico son quienes más han llevado a cabo esfuerzos por proponer una agenda de alcance amplio en materia de los usos éticos de la IA con resultados diversos. El Committee on Publication Ethics (COPE) —organización internacional orientada al desarrollo de una cultura editorial en la que las prácticas éticas se conviertan en un aspecto inherente a la misma— publicó en febrero de 2023 una declaración relativa al uso de IA en esta clase de publicaciones. El texto es enfático al mencionar que:

9 Shannon Mattern, *A City Is Not a Computer: Other Urban Intelligences*, Princeton: Princeton University Press, 2021.

Las herramientas de IA no pueden figurar como autores de un artículo. Las herramientas de IA no pueden cumplir los requisitos de autoría, ya que no pueden asumir la responsabilidad del trabajo presentado. Como entidades no jurídicas, no pueden afirmar la presencia o ausencia de conflictos de intereses ni gestionar acuerdos de derechos de autor y licencias. Los autores que utilicen herramientas de IA en la redacción de un manuscrito, la producción de imágenes o elementos gráficos del artículo o en la recopilación y análisis de datos, deben ser transparentes al revelar en los Materiales y métodos (o sección similar) del artículo cómo se utilizó la herramienta de IA y qué herramienta se utilizó. Los autores son totalmente responsables del contenido de su manuscrito, incluso de las partes producidas por una herramienta de IA, y, por lo tanto, son responsables de cualquier incumplimiento de la ética de publicación.¹⁰

EJEMPLOS DE CUESTIONAMIENTOS DEL SECTOR EDITORIAL AL USO DE LA IA

El sector editorial reúne a instituciones u organizaciones de diferente tipo y tamaño, enfocadas en la producción y distribución de publicaciones y contenidos informativos, así como a un amplio número de profesionales, entre ellos las y los editores, autores, ilustradores y diseñadores gráficos, además de entidades más o menos formales que agrupan y representan tanto a las organizaciones editoriales como a las y los agentes de la edición. A continuación se ejemplifican tres tipos de cuestionamientos comunes que los actores del sector editorial están planteando en la actualidad respecto al uso de la IA y sus herramientas.

1. Uso ilegal de obras literarias sin reconocimiento de derechos de autor ni compensaciones económicas. En julio de 2023, varios

10 COPE Council, COPE position, “Authorship and AI”, <https://doi.org/10.24318/cCVRZBms>.

Inteligencia artificial...

autores presentaron una demanda colectiva en el tribunal federal de San Francisco (EE.UU.) en contra de la organización desarrolladora de IA, OpenAI, por “recurrir a la recopilación masiva’ de obras protegidas por derechos de autor ‘sin consentimiento, sin reconocimiento y sin compensación’”¹¹. La demanda buscaba establecer que la compañía infringió los derechos de autor de los escritores al descargar ilegalmente copias de novelas para entrenar su sistema de IA. Además se argumenta que los grandes modelos de lenguaje que impulsan a ChatGPT son en sí mismos obras derivadas infractoras, creadas sin el permiso de los autores y en violación de sus derechos exclusivos bajo la Ley de Derechos de Autor. En apoyo de su acusación, los autores demandantes señalan que ChatGPT es capaz de generar resúmenes de sus novelas cuando se le solicita, lo que, según argumentan, solo sería posible si ChatGPT hubiera sido entrenado utilizando sus obras sujetas a derechos de autor. La demanda también critica a OpenAI por descargar ilegalmente cientos de miles de libros para entrenar su sistema de IA, ya que en 2018 la compañía admitió haber utilizado una colección de más de 7 000 novelas sin permiso en su modelo de lenguaje GPT-1, además de que versiones posteriores del modelo fueron entrenadas utilizando mayores cantidades de obras protegidas por derechos de autor.¹²

2. Falta de regulación de uso de la IA en el sector editorial. En una entrevista para el diario mexicano Reforma en febrero de 2024, Quetzalli de la Concha —presidenta del Comité de Derechos de Autor y Derechos Conexos de la Asociación Mexicana para la Protección de la Propiedad Intelectual (AMPPPI)— y Hugo Setzer —presidente de la Cámara Nacional de la Industria Editorial Mexicana (CANIEM)— hablaron de la necesidad de regular el uso de la IA en el ámbito de la edición, poniendo especial énfasis en trazar las fronteras del uso de la Inteligencia Artificial como una herramienta de

11 Prensa CANIEM, “Autores demandan a Open AI”, 18 de julio de 2023, <https://caniem.org/autores-demandan-a-open-ai/>.

12 Prensa CANIEM, “Autores demandan a Open AI”.

apoyo para crear y no como generadora de contenidos en sí misma. Pues esta segunda acepción desvirtúa su uso, trastoca la calidad de los contenidos y también la cadena profesional laboral implícita en el libro. La situación preocupante es que se incorporen textos sin los correspondientes derechos de autor para alimentar esta tecnología que luego servirá para hacer negocio. Hasta ahora no se ha visto que haya programadores que se comprometan a crear plataformas que estén alimentadas con matrices cien por ciento autorizadas o con contenidos cien por ciento autorizados.¹³

3. Reconocimiento de uso y/o no utilización de IA en la autoría de las obras. La creciente utilización de programas de inteligencia artificial generativa ha llevado a editoriales como la francesa Librinova, una de las más prolíficas en aquél país, a crear un sello que garantiza que la obra literaria fue íntegramente realizada por un ser humano. En enero de 2024, Librinova anunció que había unido fuerzas con la empresa Label Création humaine (Etiqueta Creación Humana), para certificar que una obra (escrita, audiovisual o musical) ha sido efectivamente producida por un cerebro humano, como respuesta a los errores factuales en los que pueden incurrir los textos que se apoyan en herramientas de IA generativa.¹⁴

En contraposición a la situación referida, se tiene el caso de Elisa Shupe, una veterana retirada del ejército estadounidense quien logró superar los límites de los derechos de autor de la IA. Shupe presentó un registro de derechos de autor para una novela que ella misma publicó. Había utilizado ampliamente ChatGPT de OpenAI mientras escribía el libro. En un principio la Oficina de Derechos de Autor de EE.UU. rechazó la solicitud, pero a comienzos de abril de 2024, después de que Shupe apelara, le otorgaron el registro de derechos de autor para su libro *AI Machinations: Tangled Webs and Typed Words*. Esta es una de las primeras obras

13 Yanireth Israde, “Piden regular uso de IA en la industria editorial”, *Reforma*, 18 de febrero de 2024.

14 Radio Francia Internacional, “Nace un nuevo certificado: libro libre de inteligencia artificial”, 10 de enero de 2024.

Inteligencia artificial...

creativas en recibir derechos de autor por la disposición del texto generado por IA:

(..) la oficina de derechos de autor ha dejado claro que aunque Shupe no es considerada la autora de todo el texto, sí se la considera la autora de la “selección, coordinación y disposición del texto generado por la inteligencia artificial”. Esto evita que otros copien el libro sin permiso, pero el contenido del libro no está protegido e, hipotéticamente, podría reorganizarse y volverse a publicar como un libro diferente.¹⁵

En el caso particular de Shupe, ella logró superar las restricciones relacionadas con los derechos de autor y la IA debido a que no puede escribir a causa de las lesiones que sufrió mientras estuvo en el ejército, de modo que solo pudo completar su libro con la ayuda de una inteligencia artificial. Según relata, trabajar con IA le permitió tener una forma diferente de expresar sus pensamientos, si bien igualmente tuvo que dedicar muchas horas para crear su libro. La apelación exitosa del registro se centró, por tanto, en las discapacidades de Shupe, comparando su uso de ChatGPT con el de una prótesis de pierna para una persona amputada.¹⁶

Los tres cuestionamientos sobre el uso de la IA en el sector editorial hicieron eco en este ámbito y cobraron forma en la encuesta “Uso y percepción de las herramientas de Inteligencia Artificial en la industria del libro”, aplicada durante los meses de junio y julio de 2024 por la consultora editorial Proyecto 451.¹⁷ Esta consistió de 13 preguntas dispuestas en un formulario proporcionado por la herramienta Typeform, enviadas por correo electrónico a suscriptores de su *newsletter* semanal y compartida en sus perfiles

15 José Antonio Vázquez, “Una autora logra vencer los derechos de autor de IA”, *DosDoce*, 20 de mayo de 2024.

16 José Antonio Vázquez, “Una autora logra vencer los derechos de autor de IA”.

17 Proyecto451, *IA-Encuesta: Uso y percepción de la Inteligencia Artificial en la industria del libro*, Academia Proyecto451, 2024, PDF, <https://academia.proyecto451.com/uploads/IA-Encuesta.pdf>.

de redes sociales para que pudiera tener mayor alcance. En septiembre de 2024 se dieron a conocer los resultados a partir de 735 respuestas provenientes de 20 países (75% de ellas ubicadas en Argentina, España y México), destacando que:

- a. Los profesionales del sector editorial con una experiencia mayor a los 10 años consideran como positivo el avance de la IA en esta área.
- b. El grueso de los encuestados afirman que la optimización de costos y la automatización de tareas son los mayores beneficios de la aplicación de la IA.
- c. Un número reducido de profesionales considera que la IA tendrá un impacto positivo en generar nuevas experiencias de lectura o en mejorar la calidad literaria de las obras.
- d. Los encuestado enlistaron las actividades editoriales en las que emplean la IA, siendo las más recurrentes la corrección de estilo, la traducción, el *brainstorming*, el análisis de información y el desarrollo de campañas de marketing/comunicación.
- e. Existe un acuerdo general respecto a que la IA no puede generar contenido original o novedoso. Ni tampoco traducir contenido con la misma calidad que lo hace un ser humano.

CONCLUSIONES

Los cuestionamientos y debates en torno al uso de la IA en el sector editorial apenas comienzan, augurando un incremento sostenido de ellos en el corto y mediano plazo. La definición de pautas de uso ético y/o restricciones de empleo de las herramientas de la IA en el sector editorial dependerá en gran parte de los países y subsectores desde donde se emprendan, así como de la eficiencia de los modos en que los distintos actores se organicen y establezcan sinergias con asociaciones y entidades gubernamentales.

Inteligencia artificial...

A la pregunta, “¿cuál es el mayor cuestionamiento que tendrá el sector editorial al uso de herramientas de Inteligencia Artificial en el corto plazo?”, ChatGPT respondió el 06 de agosto de 2024:

El sector editorial enfrenta varios cuestionamientos importantes respecto al uso de herramientas de Inteligencia Artificial (IA) en el corto plazo. Algunos de los mayores incluyen:

- 1) Derechos de autor y propiedad intelectual: La IA puede generar contenido a partir de obras existentes, lo que plantea dudas sobre quién posee los derechos de autor de los textos generados por IA y cómo se protege el trabajo de los autores originales.
- 2) Calidad del contenido: Existe preocupación sobre si el contenido generado por IA puede mantener la calidad, creatividad y originalidad que los lectores esperan de las obras humanas. Los editores tendrán que evaluar si el contenido generado por IA cumple con sus estándares.
- 3) Desempleo y cambio en roles laborales: La automatización de tareas editoriales y de escritura puede llevar a la reducción de puestos de trabajo tradicionales en el sector, lo que genera incertidumbre y preocupación entre los profesionales del ámbito editorial.
- 4) Plagio y autenticidad: La facilidad con la que la IA puede generar contenido similar a obras existentes aumenta el riesgo de plagio y pone en cuestión la autenticidad de las nuevas publicaciones. Los editores deberán implementar medidas para asegurar la originalidad de los textos.
- 5) Ética y transparencia: La utilización de IA en la creación de contenido editorial plantea dilemas éticos, incluyendo la transparencia sobre el uso de IA y la necesidad de informar a los lectores cuando un contenido ha sido generado por máquinas.
- 6) Monopolización y acceso a tecnología: Las grandes empresas tecnológicas que desarrollan herramientas de

IA pueden monopolizar el acceso a estas tecnologías, dejando a las editoriales más pequeñas en desventaja competitiva.

- 7) Confianza del lector: Los lectores pueden ser escépticos sobre la autenticidad y valor de los contenidos generados por IA, lo que podría afectar la reputación y la confianza en las editoriales que utilicen estas herramientas.

Estos desafíos requerirán de un equilibrio entre aprovechar los beneficios de la IA y abordar las preocupaciones éticas, legales y de calidad asociadas con su uso en el sector editorial.¹⁸

BIBLIOGRAFÍA

Caparrós, Martín. *El mundo entonces. Una historia del presente*. Ciudad de México: Penguin Random House, 2023.

COPE Council. *COPE position*. "Authorship and AI". <https://doi.org/10.24318/cCVRZBms>.

Crawford, Kate. *Atlas de inteligencia artificial. Poder, política y costos planetarios*. Ciudad Autónoma de Buenos Aires: Fondo de Cultura Económica, 2022.

Deployteq. "Challenges Facing the Publishing and Media Industry in 2024". 9 de marzo de 2024. <https://n9.cl/g8bn9>.

García Higuera, Laura. "Los ilustradores exigen a las editoriales que dejen de usar IA en las portadas: 'El robo más grande de la historia'". *El Diario*, 14 de marzo de 2024. https://www.eldiario.es/cultura/libros/ilustradores-exigen-editoriales-dejen-ia-portadas-robo-grande-historia_1_10992145.html.

18 OpenAI. 2024, "¿Cuál es el mayor cuestionamiento que tendrá el sector editorial al uso de herramientas de Inteligencia Artificial en el corto plazo?", *ChatGPT*, 06 de agosto de 2024, <https://chat.openai.com/>.

Inteligencia artificial...

- Israde, Yanireth. “Piden regular uso de IA en la industria editorial”. *Reforma*, 18 de febrero de 2024. <https://www.reforma.com/piden-regular-uso-de-ia-en-la-industria-editorial/ar2759578>.
- Karpf, Dave. “The Myth of Technological Inevitability”. *The Future, Now and Then* (blog), 24 de enero de 2024. <https://davekarpf.substack.com/p/the-myth-of-technological-inevitability>.
- Mattern, Shannon. *A City Is Not a Computer: Other Urban Intelligences*. Princeton: Princeton University Press, 2021.
- OpenAI. 2024. “¿Cuál es el mayor cuestionamiento que tendrá el sector editorial al uso de herramientas de Inteligencia Artificial en el corto plazo?”. *ChatGPT*, 06 de agosto de 2024. <https://chat.openai.com/>.
- Prensa CANIEM. “Autores demandan a Open AI”. 18 de julio de 2023. <https://caniem.org/autores-demandan-a-open-ai/>.
- Proyecto “Un libro para leer”. Publicación en Facebook. 10 de abril de 2024. <https://www.facebook.com/photo/?fbid=1125073812230055&set=a.714487993288641>.
- Proyecto451. IA-Encuesta: Uso y percepción de la Inteligencia Artificial en la industria del libro. Encuesta publicada en Academia Proyecto451, 2024. <https://academia.proyecto451.com/uploads/IA-Encuesta.pdf>.
- Radio Francia Internacional. “Nace un nuevo certificado: libro libre de inteligencia artificial”. 10 de enero de 2024. <https://www.rfi.fr/es/cultura/20240110-nace-un-nuevo-certificado-libro-libre-de-inteligencia-artificial>.
- Sadin, Éric. *La Inteligencia Artificial o el Desafío del Siglo. Anatomía de un Antihumanismo Radical*. Ciudad Autónoma de Buenos Aires: Caja Negra Editora, 2020.
- Smink, Verónica. “Las 3 etapas de la Inteligencia Artificial: en cuál estamos y por qué muchos piensan que la tercera puede ser fatal”. *BBC News Mundo*, 29 de mayo de 2023. <https://www.bbc.com/mundo/noticias-65617676>.
- Vázquez, José Antonio. “Una autora logra vencer los derechos de autor de IA”. *DosDoce*, 20 de mayo de 2024. <https://www.dosdoce.com/2024/05/20/una-autora-logra-vencer-los-derechos-de-autor-de-ia/>.

AI and Academic Ethics in Scientific Journals in the Fields of Astronomy and Mathematics

SUYIN ORTEGA CUEVAS

*Instituto de Investigaciones en Matemáticas Aplicadas y en Sistemas,
Universidad Nacional Autónoma de México*

BEATRIZ JUÁREZ SANTAMARÍA

*Instituto de Astronomía,
Universidad Nacional Autónoma de México*

INTRODUCTION

The ethical use of information is a fundamental aspect of academic work. Bibliographic references are essential in this process to acknowledge ideas taken from other authors. Standardized citation styles allow for the proper use of sources, enabling anyone to locate and consult them.

Academic journals evaluate bibliographic styles and assess the originality and accuracy of information. When these requirements are not fulfilled, an impartial assessment is done to confirm academic integrity. To support this process, editors rely on the Committee on Publication Ethics (COPE), which assists scientific journals and publishers in adhering to publication ethics standards. COPE also provides guidelines on the use of Artificial Intelligence (AI) in scientific articles and the ethical risks associated with it.

This research aims to examine the guidelines followed by the publishers of scientific journals in the core collection of the Institute of Astronomy and the Institute of Applied Mathematics and Systems to ensure academic integrity when using AI in scientific articles.

Inteligencia artificial...

The conclusion is that academic integrity is crucial in scientific research, and guidelines regarding the use of AI are essential to achieving this objective. Since scientists are required to report whether or not they employed AI tools in the creation of their papers, scientific publications include recommendations regarding ethical considerations and AI in their author instructions.

The emergence of generative Artificial Intelligence and AI-assisted technologies has led scientific journals to institute policies regarding their use. This study will address the policies on the use of generative AI implemented by the publishers of the scientific journals in the core collection of the Institute of Astronomy and the Institute of Applied Mathematics and Systems at the National Autonomous University of Mexico (UNAM).

It is worth noting that, in some cases, the same publisher serves both Institutes, and several adhere to the guidelines established by COPE. COPE is committed to educating and supporting editors, universities, research institutes, and all those involved in publication ethics. Its goal is to advance the editorial culture toward one where ethical practice becomes an integral part of the culture itself¹. To this end, COPE provides documents and discussion forums on various aspects of publication ethics. These aspects include a flowchart for simultaneous submissions to multiple journals, concerns about unpublished data risks, suspected peer review manipulation during the review process, or inappropriate manipulation of images in a published article, among others.

OBJECTIVE

This study aims to verify the guidelines or standards set by the publishers of scientific journals in the core collection of the Institute of Astronomy and the Institute of Applied Mathematics and Systems at UNAM regarding the use of Artificial Intelligence (AI) to achieve academic integrity.

1 Committee on Publication Ethics, "About COPE", accessed July 5, 2024.

ARTIFICIAL INTELLIGENCE

Before addressing the guidelines established by publishers regarding the use of Artificial Intelligence in research articles, it is crucial to define what AI is:

In his work *Ethics of Artificial Intelligence in Academic Research and Education*, Nader Ghotbi² describes AI as “a set of computerized systems that enable the creation of knowledge with the help of other technologies, especially in electronic engineering. These systems can allow machines to perform tasks similar to humans, such as recognizing human voice through listening, reading texts aloud, conversing with humans, moving a vehicle or robotic body, or generating texts or images based on instructions.”

According to Estupiñán, Leyva, Peñafiel, and Assafiri³, AI is a branch of computer science that aims to develop machines and systems capable of performing tasks that require human intelligence. Machine learning and deep learning are two domains of AI. In recent years, with the development of new techniques and computing equipment based on neural networks, AI has increasingly come to be understood as synonymous with “supervised deep machine learning.”

The European Parliament⁴ defines Artificial Intelligence as “the ability of a machine to exhibit capabilities similar to those of humans, such as reasoning, learning, creativity, and planning.” It notes that the types of AI include:

- *Software*: This category encompasses virtual assistants, search engines, voice, and facial recognition systems.

2 Nader Ghotbi, “Ethics of Artificial Intelligence in Academic Research and Education,” in *Second Handbook of Academic Integrity*, edited by Sarah Elaine Eaton, 1355-1366, Cham, Springer Nature Switzerland, 2024, 1356.

3 Ricardo Estupiñán *et al.*, “Inteligencia Artificial y propiedad intelectual,” *Universidad y Sociedad* 13, no. S3 (2021), 362-368.

4 Parlamento Europeo, “¿Qué es la Inteligencia Artificial y Cómo se Usa?”. Parlamento Europeo, 8 de septiembre de 2020, paragraf 1.

Inteligencia artificial...

- *Integrated Artificial Intelligence*: This includes robots, drones, and autonomous vehicles.

Thus, we can infer that Artificial Intelligence involves the construction of computational algorithms that create tools endowed with sufficient information to reproduce knowledge—whether reliable or not— and has the characteristic of imitating human thought through reasoning, perception, and language.

ARTIFICIAL INTELLIGENCE TOOLS IN ACADEMIC WRITING

Various tools are available for content creation using generative artificial intelligence. Nevertheless, it is imperative to understand the purpose of each instrument. However, it is not feasible to name them all.

Different taxonomies (classification types) exist for these tools, and some entities compile and present them systematically. For example, the National Autonomous University of Mexico (UNAM) hosts the IAGen site in Education, which focuses on some resources and their applications: “Interesting Resources on Generative AI in Education”⁵. “120 Mind-Blowing AI Tools,” is another resource where users can access a variety of AI tools. It organizes them into categories such as productivity, video, marketing, chatbots, design, and writing⁶.

Among graphical representations of generative AI tools, Solis⁷ provides the most comprehensive taxonomy through his “GenAI Prism.” This visual representation categorizes tools based on various needs, allowing for more precise selection.

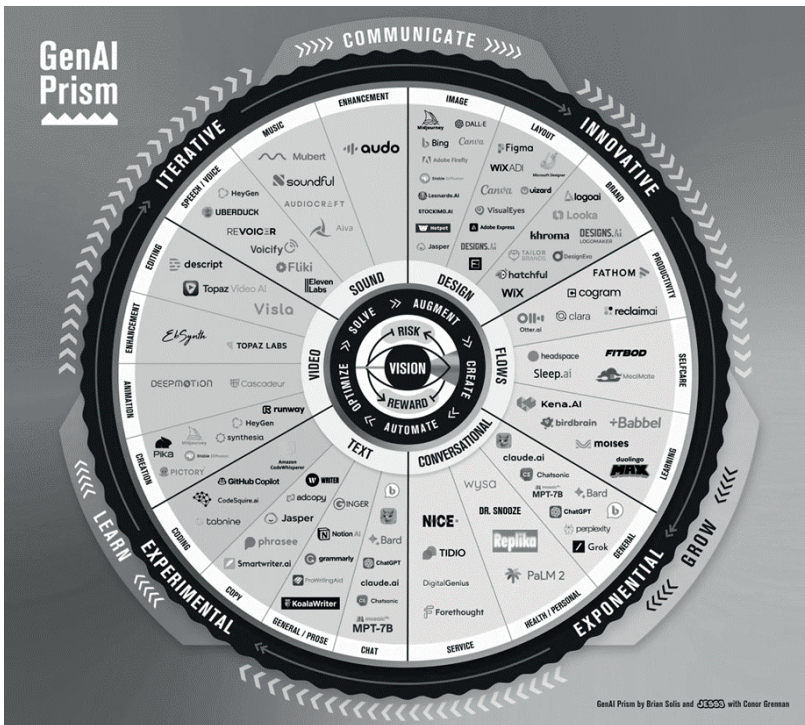
5 IAGen en Educación, UNAM, “Recursos sobre IA Generativa en Educación”, 2023.

6 Doug Austin, “120 Mind Blowing AI Tools,” *eDiscovery Today by Doug Austin* (blog), July 21, 2023.

7 Brian Solis, “Introducing the GenAI Prism Infographic: A Framework for Collaborating with Generative AI,” *Brian Solis* (blog), December 20, 2023.

The *GenAI Prism Version 1.0* infographic visualizes companies that produce AI tools to automate and enhance the way people use these tools in their personal and professional lives. In other words, generative AI is designed to boost human creativity to achieve more effective results. This infographic was designed and created by Brian Solis, Connor Grennan, and the design studio JESS3⁸.

Figure 1. GenAI Prism



Source: Brian Solis, “Introducing the GenAI Prism Infographic: A Framework for Collaborating with Generative AI.” <https://briansolis.com/2023/12/introducing-the-genai-prism-infographic-a-framework-for-colaborating-with-generative-ai/>.

8 Brian Solis, *GenAI Prism by JESS3 and Brian Solis v1.0*, (Image,) Flickr, December 18, 2023.

ETHICS CODES AND THE USE OF ARTIFICIAL INTELLIGENCE
IN SCIENTIFIC PUBLICATIONS

As mentioned earlier, AI has numerous applications, and scientific research is no exception. AI is utilized for data analysis, image creation, disease diagnosis and detection, drug development, and scientific article writing. It can analyze large amounts of information, identify behavioral patterns, and generate ideas for developing academic texts.

The use of Artificial Intelligence tools in publications is not restricted to academic writing and knowledge dissemination. However, following certain norms and guidelines is essential to ensure reliable and ethical academic practice. To ensure the appropriate use of AI guidelines, consulting Committee on Publication Ethic⁹ is recommended.

Academic publications adhere to COPE's guidelines, which include directives on using AI, addressing ethical concerns, responsibility, and transparency. AI tools can be used provided the article specifies how and for what purpose they were utilized.

Ethics has always been a topic of debate, and the use of AI in presenting information in journals, books, and websites has sparked further discussions across all disciplines, including medicine, art, and engineering. Consequently, the controversy surrounding AI usage is very current.

Regarding ethics in academic publishing, COPE's discussion forums include topics related to AI usage ethics, providing recommendations for publishers. As previously discussed, AI tools are not authors and do not generate information independently; therefore, they are not responsible for what is published. Since these tools operate through computational algorithms and search information online, they do not have filters to discriminate between correct and incorrect information or between true and false data, which can result in misinformation.

9 Committee on Publication Ethics, "Artificial Intelligence (AI) in Decision Making", last reviewed: 24 September, 2021.

COPE codes outline the ethical boundaries of AI usage. Fortunately, these codes and discussion forums are available online:

- Artificial Intelligence in the News, Latest Update, April 2023.
- Artificial Intelligence and Authorship, COPE Editorial, February 2023.
- The Challenge of AI Chatbots for Journal Editors, Guest Editorial, February 2023.
- Trustworthy AI for the Future of Publishing, COPE Seminar 2021.
- AI in Decision Making Discussion, COPE Forum Discussion, November 2019.

It is important to note that COPE continues to work on updating its guidelines, as there are no existing regulations guiding editors on the use of AI, given that tools are continuously created, updated, or discontinued.

ELEMENTS TO CONSIDER WHEN USING ARTIFICIAL INTELLIGENCE TOOLS IN PUBLISHING

Artificial Intelligence (AI) tools can streamline processes for authors by allowing them to articulate ideas, generate images, or create presentations easily. These tools are designed using vast amounts of information available on the Internet, and with algorithms developed by their creators, AI can assist in generating academic content in line with the authors' guidelines—authored by real individuals.

However, AI can also produce errors, including misinformation and fake news, which can result in documents with questionable or no veracity. Therefore, caution is advised when using these tools, as they are meant to assist in presenting research rather than to replace the critical thinking and validation required in academic work.

COPE recommends reviewing guidelines on the use of AI, such as those provided by the World Association of Medical Editors

(WAME). In the document “*Chatbots, Generative AI, and Scholarly Manuscripts*,” WAME offers five key recommendations:

1. *WAME Recommendation 1:* Chatbots cannot be authors.
2. *WAME Recommendation 2:* Authors should be transparent when chatbots are used and provide information about how they were used.
3. *WAME Recommendation 3:* Authors are responsible for material provided by a chatbot in their paper, including the accuracy of what is presented and the absence of plagiarism, and for appropriate attribution of all sources, including sources for material generated by the chatbot.
4. *WAME Recommendation 4:* Editors and peer reviewers should disclose any use of chatbots in the evaluation of the manuscript and in generating reviews and correspondence. If chatbots are used in communications, their method of use should be explained.
5. *WAME Recommendation 5:* Editors need appropriate tools to detect AI-generated or altered content. Such tools should be accessible to editors, regardless of their financial capability, to safeguard scientific integrity and public health by reducing the risk of adverse outcomes¹⁰.

These recommendations emphasize that AI tools are intended to assist in quickly capturing and presenting ideas, enabling demonstrations that the author may not be able to achieve independently. However, the core ideas and execution should remain under the authors’ control. Tavira¹¹ provides examples of how some publishers approach this, as illustrated in Table 1:

10 C. Zielinski *et al.* “Chatbots, Generative AI, and Scholarly Manuscripts. WAME Recommendations on Chatbots and Generative Artificial Intelligence in Relation to Scholarly Publications,” *Colombia Médica (Cali)* 54, no. 3 (2023): 3-4.

11 Ricardo Tavira, “Boletín SciELO-México: Revistas frente a la IA: Casos ELSEVIER, SPRINGER NATURE y TAYLOR & FRANCIS.” (Cuadro.) *Boletín SciELO-México* (blog), August 30, 2023.

Table 1. AI Tools Considerations.

Restrictions on authorship and use of AI	AI-generated images
AI in figures and images	Important notes
Exceptions to the use of AI in images	Responsibility of the authors
The use of AI in works of art is not permitted	Authoring and AI tools
Authored by AI	Recognizing and documenting the use of AI tools

Source: Ricardo Tavira, "Boletín SciELO-México...", <https://boletinscielomx.blogspot.com/2023/08/revistas-frente-la-ia-casos-elsevier.html?spref=tw>.

THE LITERATURE SUGGESTS ELEMENTS OR RECOMMENDATIONS FOR USING AI TOOLS

Based on the literature, nine key elements were selected to analyze the information provided by publishers regarding ethics and the use of AI. These elements for AI usage in academic writing are:

- Website
- Consideration of COPE
- Statement of Use
- Ethical Codes
- Authorship Role
- Review Role
- Editing Role
- AI in Audiovisual Support
- AI in Writing

METHODOLOGY

The following actions were taken to carry out this study:

- *Identify the Publishers:* Locate the publishers of journals in the core collection of the Institute of Astronomy and the

Inteligencia artificial...

Institute of Applied Mathematics and Systems. Based on Bradford's Law, select the titles where academics publish to identify core journals.

- *Locate Publisher Websites:* Locate the websites of each publisher, as they may provide author instructions or ethical codes.
- *Check for AI Usage Statements:* Verify the inclusion of a statement on the use of AI tools. Search databases such as Web of Science, Scopus, and Google Scholar to identify documents that consider ChatGPT as an author, ensuring compliance with guidelines for disclosing AI tool usage.
- *Identify the Nine Recommended Elements:* Ensure these elements are addressed in the document's body.

Upon reviewing the ethical sections of the publisher websites for the core collection journals studied, it was found that these references adhere to the Committee on Publication Ethics (COPE) guidelines. Additionally, COPE's stance on AI and ethics emphasizes that only human authors should sign published documents. Considering AI as an author poses the risk of avoiding responsibility for the content, which calls into question the accuracy of the information.

RESULTS

To identify the publishers of core journals where researchers from both institutes publish their articles, records from the Web of Science database for the years 2019 to 2023 were retrieved. Following Bradford's Law, the top ten publishers were selected from the results list. This selection is illustrated in Table 1.

Table 2. Number of publishers of the Core Journals

Instituto de Astronomía	IIMAS
Elsevier	Elsevier
Wiley	Wiley
IOP	IOP
Springer	Springer
Edp Science	
	Public Library Science (Plos One)
	MDPI
	IOS Press
	Frontiers Media Sa
	Amer Physical Soc
	Amer Inst Physics

Source: Prepared by the authors using the subscriptions of each library and internal documents as sources.

Analysis of AI Tool Usage in Publishers’ Guidelines

After the publishers were identified for analysis, the critical points for using AI tools were examined. This was done by reviewing the instructions for authors on each publisher’s website, with a focus on how these tools are addressed.

Table 3 outlines the nine points to analyze. The first column lists these points, and the first row displays the publishers. The shaded areas indicate where information about the use of AI tools is provided. These results reflect the state of the information at the time of the search. It is important to note that some results are based on interpretations, as the concepts may vary according to each editor’s criteria.

Inteligencia artificial...

From the analysis:

- *Wiley* is the publisher with the most shaded cells, indicating a solid emphasis on guidelines regarding AI tool usage.
- *AIP* is the publisher with fewer shaded cells, suggesting that it must comprehensively cover AI tool usage.

However, it is noteworthy that most publishers reference and incorporate COPE guidelines into their policies.

Table 3. List of recommendations considered by publishers

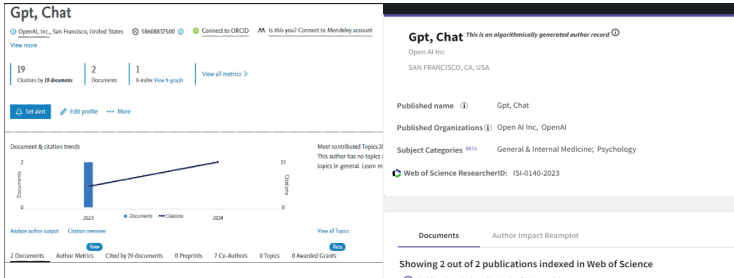
Editorial / AI Points	W	PLOS ONE	APS	EDP	MDPI	S	IOS	FM	IOP	E	AIP
Web page	Shaded										
Consider COPE	Shaded					Shaded					
Statement of the writing	Shaded								Shaded	Shaded	
Codes of ethics	Shaded							Shaded			
Role of authorship	Shaded							Shaded			
Review role	Shaded					Shaded					
Editing role	Shaded										
AI in audiovisual support	Shaded							Shaded			
AI in the scriptures	Shaded									Shaded	

Note: publisher names are abbreviated for space reasons: W= Wiley, Plusone= Public Library Science (Plos One); APS= American Physical Society; EDPS= EDP Science; MDPI= MDPI; S= Springer; IOS= IOS Press; FM= Frontier Media SA; E= Elsevier; American Institute of Physics

Source: Prepared by the authors based on the websites of each publisher.

Another result obtained was searching ChatGPT as an author in the academic databases Web of Science and Scopus, which yielded two articles, as illustrated in the following image.

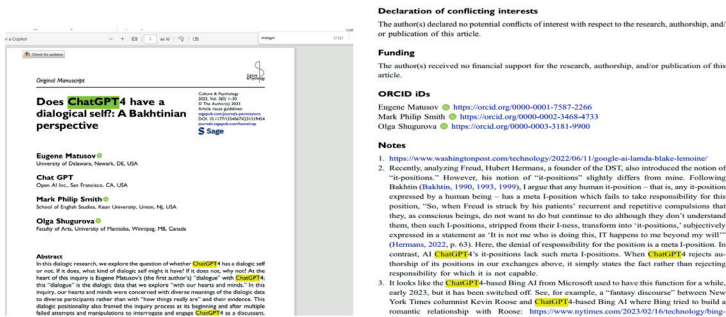
Figure 2. Screenshots of the queried databases



Source: Screenshots of Scopus database, using the following ID: ChatGPT, null.

Upon analyzing the articles, it was confirmed that the article titled “Does Chat GPT-4 Have...” adheres to the previously described guidelines. In the notes section of the document, there is a statement detailing the use of AI tools and how they were applied.

Figure 3. Screenshot of the article “Does ChatGPT-4 Have a Dialogical...”



Source: Screenshots of Eugene Matusov et al., *Culture & Psychology* 30, no. 4 (2024): 841-875.

Inteligencia artificial...

Upon reviewing the second article, it is confirmed that ChatGPT is not listed as an author. However, the use of ChatGPT as a tool is mentioned in the appendices, where the declaration of its usage is provided. Therefore, the document complies with the guidelines.

Figure 4. Screenshot of the article “Acute Pneumonia-Like Illness...”

The screenshot displays the Cureus website interface for an article. The article title is "Acute Pneumonia Like Illness and Sepsis in India: Is it Time to Suspect Pulmonary Melioidosis?". The authors listed are Anupam Sampath, M Subramanian, Prashant Patil, Ansh Gupta, and Jagu Khilangoi. The article is categorized under "Open Access Case Report" with the ID 10.7797/curres.36122. The abstract describes a case of melioidosis, a severe and potentially fatal disease caused by the Gram-negative bacterium *Burkholderia pseudomallei*, in a 50-year-old male in India. The disease is endemic in Southeast Asia and Northern Australia. The case presented here has a history of acute febrile illness and progressive dyspnea, with clinical worsening leading to intensive care unit (ICU) care. The management of acute pneumonia-like melioidosis with antibiotics and supportive care which showed rapid recovery is follow-up. This case highlights the need for a high index of suspicion and increased awareness of early diagnosis of melioidosis in the Indian subcontinent to improve the patient's outcome. The introduction section begins by defining melioidosis as a previously rare Gram-negative bacterium, present with a wide variety of clinical conditions conceptualized as melioidosis, a severe and potentially fatal disease caused by the bacterium *Burkholderia pseudomallei*, which has been reported in India in a few situations.

Source: Screenshots of K. Sampath *et al.*, *Cureus* 15, no. 3 (2023): e36122.

Google Scholar

The third database reviewed was Google Scholar due to its open access and broader range of records. An advanced search for the author “ChatGPT” yielded 46 results (Figure 5). These results included webpages, book chapters, official ChatGPT community sites, manuals, and advertisements about its use. Only one article was found where ChatGPT appears as an author. This article was published in a scientific journal edited by SAGE.

Figure 5. Screenshot of Search results for “ChatGPT as an author” in Google Scholar



Source: Screenshot of Google Scholar. Accessed July 5, 2024.

CONCLUSIONS

The analysis of the two academic articles confirmed that the publishers adhere to COPE’s recommendations regarding the ethical use of information when employing AI tools.

The publishers of the core journals, where academics from the two studied institutes publish, do consider guidelines for AI tool usage and ethics, thus ensuring the quality of their publications.

Most of the publishers’ guidelines are based on COPE’s standards. Just two of the eleven publishers that were assessed have created particular criteria for using AI that are in accordance with COPE’s suggestions.

It is noted that COPE’s AI guidelines are open to interpretation as they are still being consolidated.

BIBLIOGRAPHY

- Austin, Doug. "120 Mind Blowing AI Tools." *eDiscovery Today* by Doug Austin (blog), July 21, 2023. <https://ediscoverytoday.com/2023/07/21/120-mind-blowing-ai-tools-artificial-intelligence-trends/>.
- Committee on Publication Ethics (COPE). "About COPE." Accessed July 5, 2024. <https://publicationethics.org/about/our-organisation>.
- Committee on Publication Ethics (COPE). "Artificial Intelligence (AI) in Decision Making." Last reviewed: 24 September, 2021. <https://publicationethics.org/resources/discussion-documents/artificial-intelligence-ai-decision-making>.
- Estupiñán, Ricardo, Jesús Estupiñán, Maikel Yelandi Leyva Vázquez, Alex Javier Peñafiel Palacios y Yusef El Assafiri Ojeda. "Inteligencia Artificial y propiedad intelectual." *Universidad y Sociedad* 13, no. S3 (2021): 362-368.
- Fernández Morales, Flavio H. "La Inteligencia Artificial y la escritura de artículos científicos." *Revista de Investigación, Desarrollo e Innovación* 13, no. 2 (2023): 221-221.
- Ghotbi, Nader. "Ethics of Artificial Intelligence in Academic Research and Education." In *Second Handbook of Academic Integrity*, edited by Sarah Elaine Eaton, 1355-1366. Cham: Springer Nature Switzerland, 2024. https://doi.org/10.1007/978-3-031-54144-5_143.
- Google Scholar. Search results for "ChatGPT as an author." Accessed July 5, 2024. <https://scholar.google.com.mx/>.
- IAGen en Educación, UNAM. "Recursos sobre IA Generativa en Educación." 2023. <https://iagenedu.unam.mx/i-jornada/resources>.
- Matusov, Eugene, Chat GPT, Mark Philip Smith and Olga Shugurova. "Does ChatGPT4 Have a Dialogical Self?: A Bakhtinian Perspective." *Culture & Psychology* 30, no. 4 (2024): 841-875. <https://doi.org/10.1177/1354067X231219454>.

- Parlamento Europeo. “¿Qué es la Inteligencia Artificial y Cómo se Usa?” Parlamento Europeo, 8 de septiembre de 2020. <https://www.europarl.europa.eu/topics/es/article/20200827STO85804/que-es-la-inteligencia-artificial-y-como-se-usa>.
- Sampath, K., N. Gupta, R. Sahu and S. Singh. “Acute Pneumonia Like Illness and Sepsis in India: Is it Time to Suspect Pulmonary Melioidosis?” *Cureus* 15, no. 3 (2023): e36122. Doi: 10.7759/cureus.36122.
- Solis, Brian. *GenAI Prism by JESS3 and Brian Solis v1.0*. (Image.) Flickr, December 18, 2023. <https://www.flickr.com/photos/briansolis/53404929226/>.
- Solis, Brian. “Introducing the GenAI Prism Infographic: A Framework for Collaborating with Generative AI.” *Brian Solis* (blog), December 20, 2023. <https://briansolis.com/2023/12/introducing-the-genai-prism-infographic-a-framework-for-colalborating-with-generative-ai/>.
- Tavira, Ricardo. “Boletín SciELO-México: Revistas frente a la IA: Casos ELSEVIER, SPRINGER NATURE y TAYLOR & FRANCIS.” (Cuadro.) *Boletín SciELO-México* (blog), August 30, 2023. <https://boletinscielomx.blogspot.com/2023/08/revistas-frente-la-ia-casos-elsevier.html>.
- Zielinski, C., M. Winker, R. Aggarwal, L. Ferris, M. Heinemann, J. F. Lapeña, S. Pai *et al.* “Chatbots, Generative AI, and Scholarly Manuscripts. WAME Recommendations on Chatbots and Generative Artificial Intelligence in Relation to Scholarly Publications.” *Colombia Médica (Cali)* 54, no. 3 (2023): 1-5. <https://pmc.ncbi.nlm.nih.gov/articles/PMC10712422/pdf/1657-9534-cm-54-03-e1015868.pdf>.

La desinformación: un reto para inhibir sus efectos en la ciudadanía y el papel del bibliotecario en el uso de la información en redes sociales

ADRIANA MATA PUENTE

EDUARDO OLIVA CRUZ

*Facultad de Ciencias de la Información,
Universidad Autónoma de San Luis Potosí, México*

INTRODUCCIÓN

Cada vez se tiene mayor disponibilidad de datos y publicaciones, accesibles desde diferentes dispositivos, pero un mayor acceso trae consigo el riesgo de consumir información falsa o tendenciosa, pues a los usuarios les puede ser difícil distinguir la información verídica de la que no lo es. A lo largo de este capítulo se plantea la siguiente interrogante: ¿cómo validan los estudiantes universitarios la veracidad de la información que consumen, en particular, la de las redes sociales?

La primera razón por la que el objeto de estudio de esta investigación son estudiantes universitarios es que, al estudiar una carrera profesional, disponen de variadas herramientas tecnológicas, tanto en dispositivos como en aplicaciones, para satisfacer sus necesidades informativas. La segunda razón es porque, dado el nivel de estudio, se espera que este grupo tenga más madurez para distinguir información tendenciosa de aquella que no lo es.

Por otra parte, también nos interesa reconocer las actividades que realiza el profesional de la información. Una tarea ineludible es la de orientar al ciudadano sobre la búsqueda y análisis de la

Inteligencia artificial...

información que está consumiendo (habilidades informativas); es decir, servir de guía para que el usuario acceda a la información, dándole certeza institucional desde la biblioteca. La segunda pregunta que nos hacemos es: ¿cómo puede intervenir el bibliotecario para que el estudiante haga la distinción entre información confiable de aquella que no lo es?

El profesional de la información dispone de habilidades técnicas para organizar y diseñar productos y servicios acordes a las necesidades de la comunidad a la cual sirve. Hoy en día, también debe contar con habilidades para el marketing y publicidad para lograr que el servicio de la biblioteca sea accesible desde dispositivos móviles, a través de páginas electrónicas, y que el usuario pueda acceder al catálogo en línea; además de comunicar lo anterior en redes sociales para que el usuario lo identifique, con videos, infografías, blogs y todos los recursos disponibles.

El texto guarda el siguiente orden: primero, se plantea el problema de investigación con relación al consumo de información en las redes sociales y las estrategias empleadas para verificar el contenido, enseguida se explicitan los referentes teóricos que sustentan esta investigación. En la metodología se describe el estudio de caso con estudiantes universitarios del área de ingeniería en una universidad privada. En los resultados se da cuenta de la información a la que acceden, las estrategias para validarla y el papel del bibliotecario en esa comunidad.

MARCO TEÓRICO

Entre los estudios realizados sobre este tema, destacan las investigaciones desde el ámbito periodístico que analizan la publicación de noticias falsas (*fake news*) para incidir en aspectos políticos o económicos.¹ También se han analizado las redes sociales como

1 Jaume Suau y David Puertas-Graell, “Disinformation Narratives in Spain: Reach, Impact and Spreading Patterns”, *Profesional de la Información* 32, núm. 5 (2023): 1-16.

un propagador de información debido a la diversidad de plataformas disponibles, la facilidad de acceso y el tiempo que los usuarios permanecen consumiendo y compartiendo contenido.² Otros estudios se orientan a la alfabetización mediática³ como herramienta para que los usuarios tengan elementos y puedan validar la información que están consumiendo. También están los estudios desde la biblioteca y el bibliotecario,⁴ que muestran las estrategias que se han implementado para la verificación de la información en diferentes medios.

Los debates sobre la desinformación se orientan hacia las herramientas para evitarla, así como la verificación, la curación de contenido y, recientemente, la inteligencia artificial.⁵ Las discusiones subrayan los efectos de la desinformación sobre la salud física y mental de la ciudadanía, así como los efectos sociales y políticos.⁶

-
- 2 - Alexandre López-Borrull, “En Busca de la Verdad Perdida: Redes Sociales y Desinformación”, *Anuario ThinkEPI* 17 (2023): e17a44, <https://doi.org/10.3145/thinkepi.2023.e17a44>.
- Pavel Sidorenko Bautista *et al.*, “Espacios de Verificación en TikTok. Comunicación y Formas Narrativas para Combatir la Desinformación”, *Revista Latina de Comunicación Social*, núm. 79 (2021): 87–113, <https://doi.org/10.4185/RLCS-2021-1522>.
- Mario Wilfrido Urgilés Pineda *et al.*, “¿Informar o Desinformar? El Poder Comunicacional de las Redes Sociales”, *RECIMUNDO* 6, núm. 4 (2022): 144–152, [https://doi.org/10.26820/recimundo/6.\(4\).octubre.2022.144-152](https://doi.org/10.26820/recimundo/6.(4).octubre.2022.144-152).
 - 3 - Gema Alcolea-Díaz *et al.*, “Currículo de Alfabetización Mediática e Informativa de la UNESCO para Profesores desde la Perspectiva de la Estructura de la Información”, *Comunicar, Revista Científica de Educomunicación*, núm. 62 (2020), <https://doi.org/10.3916/C62-2020-09>.
- William Arcila *et al.*, “Tendencias Investigativas en los Estudios sobre Alfabetización Mediática Informativa y Digital (AMID) en el Campo Educativo”, *Revista Complutense de Educación* 33, núm. 2, 2022: 225–236, <https://doi.org/10.5209/rced.73935>.
 - 4 Sara Martínez-Cardama y Laura Algorta-Cancho, “Lucha contra la Desinformación desde las Bibliotecas Universitarias”, *El Profesional de la Información* 28, núm. 4 (2019).
 - 5 Alexandre López-Borrull, “En busca de la verdad perdida: redes sociales y desinformación”, *Anuario ThinkEPI* 17 (2023): 1.
 - 6 Alexandre López-Borrull, “En busca de la verdad perdida...”.

También se ha investigado la naturaleza de la desinformación, examinando cómo se difunde, qué la motiva y cómo afecta a la sociedad en general.⁷

Numerosos estudios se han dedicado al análisis del concepto de desinformación.⁸ Partiendo de una distinción unánimemente aceptada entre errores involuntarios (*misinformation*), falsedades deliberadas (*disinformation*) y difusión de información con la intención de causar daño (*malinformation*), quienes han realizado estudios tipológicos han propuesto una variada gama de clasificaciones, que concretan múltiples tipos de falsedades dentro de esas categorías principales.

Para categorizar este fenómeno, Caridad-Sebastián *et al.* establecieron tres elementos que pueden resumir las causas del auge del contexto actual de la desinformación: 1) Nuevos hábitos en el acceso y uso de la información a través de las redes sociales, caracterizada por la *viralidad* y la inmediatez. 2) Contexto tecnológico donde existen más oportunidades para acceder a la información. 3) Contexto social y político marcado por la polarización y el descrédito, además de la intención expresa de ocultar o manipular información.⁹

Esta investigación se centra en los dos primeros incisos, pues se relacionan directamente con la experiencia de la comunidad

7 Ramón Salaverría y Gustavo Cardoso, “Future of Disinformation Studies: Emerging Research Fields”, *El Profesional de la Información* 32, núm. 5 (2023).

8 - Eleni Kapantai *et al.*, “A Systematic Literature Review on Disinformation: Toward a Unified Taxonomical Framework”, *New Media & Society* 23, núm. 5 (2021): 1301–1326.

- Marta Pérez-Escolar *et al.*, “A Systematic Literature Review of the Phenomenon of Disinformation and Misinformation”, *Media and Communication* 11, núm. 2 (2023): 76–87.

- Jaume Suau y David Puertas-Graell, “Disinformation Narratives in Spain: Reach, Impact and Spreading Patterns”, *Profesional de la Información* 32, núm. 5 (2023): e320508, <https://doi.org/10.3145/epi.2023.sep.08>.

9 Mercedes Caridad-Sebastián *et al.*, “Infomediación y Posverdad: El Papel de las Bibliotecas”, 891–898.

universitaria, a diferencia del inciso tres, que se enfoca más a medios de comunicación. En la actualidad las personas disponen de más oportunidades para acceder a la información a través del celular; la encuentran en formatos como texto, fotografías, videos, entre otros, y esta es creada y compartida por los mismos usuarios, incluso antes de que los medios de comunicación y las instituciones oficiales lo hagan. Los usuarios pueden manipularla, agregar su propia opinión e incluso, generar algo nuevo a partir de la información consumida. La misma inmediatez puede ocasionar que la información no se valide, lo que contribuye a la desinformación a través de las redes sociales, como se explica en el siguiente apartado.

Redes sociales

Las redes sociales son un conjunto de sitios web donde cada usuario tiene un espacio para publicar contenidos y comunicarse con otras personas.¹⁰ Estas generan un nuevo ecosistema comunicativo virtual donde se desarrollan contenidos, hábitos de consumo y diferentes procesos de interacción.¹¹ Son plataformas en línea diseñadas para facilitar la interacción y el intercambio de información entre usuarios, que permiten compartir contenido, conectarse con otros usuarios y participar en comunidades virtuales.¹²

De acuerdo con Garrido, las redes sociales “nacen como una reunión de personas, conocidas o desconocidas, que interactuarán

10 Javier Estévez y Amanda García, “Las Redes Sociales para la Mejora de la Capacidad de Emprender y de Autoempleo”, *International Journal of Educational Research and Innovation* (2014): 101–110.

11 Laura Montero C., “Reflexiones sobre la Investigación en Redes Sociales: Facebook y Twitter”, en Segado-Boj, F. y Llovens Sobrado, B., *I Congreso Internacional de Comunicación y Sociedad Digital*, España: UNIR, 2013. <https://dialnet.unirioja.es/servlet/articulo?codigo=4565937>.

12 Jorge Carlos De la Paz *et al.*, “Estrategia de Comunicación en Redes Sociales. Escenarios Teórico Prácticos para los Medios de Comunicación en Cuba”, *Razón y Palabra*, núm. 92 (2015): 1–40.

entre sí, redefiniendo al grupo y retroalimentándolo”.¹³ Urgiles, citando a Cordova, señala que “es parte de la rutina cotidiana el hecho de que las personas al tomar un celular, inevitablemente deseen revisar las redes sociales, pues casi siempre su contenido espontáneo y novedoso intenta establecer alguna persuasión. Así, el usuario pasa de una red a otra buscando entretenimiento, información o simplemente dejando transcurrir el tiempo”.¹⁴

El rápido desarrollo de las redes sociales supone una revolución en la manera en la que nos comunicamos y también en los hábitos de consumo de información. Han demostrado ser “una poderosa forma de alcanzar millones de personas y aumentar la socialización de la información”.¹⁵ También “permiten medir el interés general sobre un tema; son fuentes informativas y constituyen una plataforma de difusión y promoción de los contenidos”.¹⁶

Este proceso de democratización de la red abrió la puerta a que cualquier persona conectada pudiera crear contenido y compartirlo. Ya no sólo los grandes medios de comunicación y los periodistas son generadores de opinión e información, sino que esta tarea ahora está al alcance de todos.¹⁷

Las redes sociales pueden desempeñar un papel tanto positivo como negativo en la difusión de información, ya que pueden amplificar la propagación de noticias falsas, pero también pueden servir

13 Norman Garrido Cabezas, “La democracia electrónica y sus tipologías”, en *Aplicaciones actuales de la comunicación e interacción digitales*, Madrid: Editorial ACCI, 2015, 253-272, PDF, Academia.edu. https://www.academia.edu/26268796/LA_DEMOCRACIA_ELECTRÓNICA_Y_SUS_TIPOLOGÍAS.

14 Erincson Córdoba, “La Tecnología en las Redes Sociales”, *Opinión*, Universidad de Piura, 20 de septiembre de 2021, 2.

15 Mario Wilfrido Urgilés Pineda *et al.*, “¿Informar o Desinformar?...?”.

16 Jorge Carlos De la Paz *et al.*, “Estrategia de Comunicación en Redes Sociales. Escenarios Teórico Prácticos para los Medios de Comunicación en Cuba”, *Razón y Palabra*, núm. 92 (2015): 3, consultado el 12 de septiembre de 2024, <https://www.revistarazonypalabra.org/index.php/ryp/article/view/69>

17 Elisa Pont, “Cómo Identificar la Desinformación en las Redes Sociales”, *La Vanguardia*, 3 de julio de 2020, consultado el 20 de julio de 2022, <https://www.lavanguardia.com/vida/junior-report/20200703/482021984534/noticias-falsas-bromas-deepfakes-redes-sociales-desinformacion.html>.

como herramientas para la verificación de información y la movilización social.¹⁸ Han amplificado el alcance y la velocidad con la que la desinformación puede llegar a millones de usuarios en todo el mundo.¹⁹

Urgiles, Núñez y Matamoros,²⁰ en su estudio acerca de la desinformación en las redes sociales y su influencia en los jóvenes del siglo XXI, encontraron que Facebook tiene el 90% de los usuarios, Instagram el 88%, y YouTube el 71%, por lo que estas eran las redes sociales más influyentes y consumidas por los jóvenes en el año 2020. Además, afirman que los encuestados tienen conciencia de que la información que consumen no es totalmente verídica o confiable, pero que no la verifican en otra parte. Esto puede deberse a que no saben cómo hacerlo de la manera correcta, no conocen fuentes de información válida y verificable, no tienen una estrategia de búsqueda, entre otras.²¹

En el caso de Facebook, la cantidad de información que se transmite a través de esta red ha aumentado exponencialmente en relación con la cantidad de usuarios que la utilizan. Sin embargo, debido a la rapidez con que se entrega, en la mayoría de los casos esta información no suele ser verificada con otras fuentes, sino leída y compartida instantáneamente, lo que provoca que además la información se descontextualice.²²

En lo que respecta a los hábitos de consumo en TikTok, las mayores motivaciones para usar esta red social son: buscar entretenimiento (41%), buscar recetas de cocina o lugares para consumir alimentos y bebidas (15%), descubrir nuevos destinos turísticos

18 L. B Jumbo, *Uso de la Red Social Facebook como Instrumento para Generar "Noticias Falsas" y su Efecto en la Opinión Pública*, Tesis de licenciatura, Universidad de Guayaquil, 2021.

19 L. B Jumbo. *Uso de la Red Social Facebook...*

20 Mario Wilfrido Urgilés Pineda *et al.*, "¿Informar o Desinformar?..."

21 F. Juca y A. A. Juca, "La desinformación en las redes sociales y su influencia en los jóvenes del siglo XXI: una mirada desde la academia", *Conference Proceeding UTMACH* 4, núm. 1 (2022): 151.

22 Cecilia Castillo *et al.*, "La Veracidad de la Información Expuesta en las Redes Sociales", *Revista Educación y Tecnología* 8, núm. 2 (2016): 27.

Inteligencia artificial...

(11%), conocer nuevos trucos de informática (9%), buscar o escuchar música (9%), seguir la vida de algún influencer (86%) o convertirse en uno (2%); y para conocer noticias nacionales (4%) e internacionales (3%).²³

En la comunidad de estudio, las redes sociales que más aparecen entre los usuarios como medio para generar, consumir y compartir información son Facebook y TikTok como herramientas para consultar videos sobre temas de interés académico y personal. También aparece WhatsApp como una red social para interactuar, tanto en el ámbito académico como en el personal. La probabilidad de compartir información falsa sin intención (*misinformation*) es más reconocida en la comunidad de estudio que las falsedades deliberadas (*disinformation*) y la difusión de información con la intención de causar daño (*malinformation*) en las redes sociales.

ALFABETIZACIÓN MEDIÁTICA E INFORMACIONAL

La alfabetización mediática e informacional (AMI) se refiere a la capacidad de comprender, evaluar y utilizar críticamente los medios de comunicación en todas sus formas, desde los medios tradicionales como la televisión y el periódico hasta las plataformas digitales como las redes sociales y los sitios web. Este concepto va más allá de simplemente comprender cómo funcionan los medios, y abarca habilidades como analizar mensajes mediáticos, reconocer sesgos y manipulaciones, y tomar decisiones informadas sobre la información consumida y compartida.²⁴

La educación y la alfabetización mediática son componentes cruciales en la lucha contra la desinformación. Así lo han reconocido múltiples instituciones y organismos internacionales, que han

23 Pavel Sidorenko Bautista *et al.*, “Espacios de Verificación en TikTok...”.

24 Juan Cristóbal Cobo, “El Concepto de Tecnologías de la Información. Benchmarking sobre las Definiciones de las TIC en la Sociedad del Conocimiento”, *Zer* 14, núm. 27 (2009): 295-318, <https://addi.ehu.es/handle/10810/40999>.

impulsado campañas de alfabetización.²⁵ En los últimos años, los estudios académicos sobre AMI se han enfocado principalmente en los jóvenes, un grupo particularmente expuesto a la información falsa a causa del uso intensivo de dispositivos digitales y el consumo de contenidos procedentes de fuentes no filtradas.²⁶

En un sentido amplio, la alfabetización mediática implica desarrollar una conciencia crítica sobre cómo los medios de comunicación influyen en nuestras percepciones, opiniones y comportamientos. Esto incluye la capacidad de detectar la desinformación, evaluar la credibilidad de las fuentes de información y entender cómo se construyen los mensajes mediáticos. Además, la alfabetización mediática también implica habilidades prácticas, como la capacidad de utilizar herramientas y tecnologías digitales de manera efectiva y ética. Esto puede incluir habilidades para buscar información en línea, verificar la autenticidad de las fuentes y proteger la privacidad y la seguridad en línea.²⁷

Es urgente contar con programas de AMI para los usuarios de las redes sociales. La cantidad de información disponible, la diversidad de fuentes y el desconocimiento para valorar críticamente la información hace más vulnerables a las personas para consumir y/o compartir información falsa. Como se verá más adelante, las estrategias empleadas para buscar, localizar y evaluar información son aplicadas de manera intuitiva. Aunque la formación en

25 - Ignacio Aguaded *et al.*, “Media Literacy from International Organizations in Europe and Latin America”, *The Journal of Media Literacy* 63, núm. 1-2 (2016): 10–17.

- William Arcila *et al.*, “Tendencias Investigativas...”, 225–236.

- Charo Sádaba y Ramón Salaverría, “Combatir la Desinformación con Alfabetización Mediática: Análisis de las Tendencias en la Unión Europea”, *Revista Latina de Comunicación Social*, núm. 81 (2023): 17–33.

26 Ramón Salaverría y Gustavo Cardoso, “Future of Disinformation Studies: Emerging Research Fields”, *El Profesional de la Información* 32, núm. 5 (2023): e320525, <https://doi.org/10.3145/epi.2023.sep.25>.

27 F. Ávila y S. Odilia, *La Posverdad como Cambio de la Contextualización de la Opinión Pública: Medios Digitales el Vocero y Bacán*, Tesis de licenciatura, Universidad Estatal Península de Santa Elena, 2024.

Inteligencia artificial...

AMI se da desde una institución formativa y/o bibliotecaria, lamentablemente no llega a toda la población escolar, además de otros sectores de la población que no están dentro de una institución educativa.

Si bien es evidente que la AMI es necesaria para todos, un sector en el que se ha puesto la atención es en la formación de los docentes, ya que son ellos los que inciden en la educación de la ciudadanía para que pueda ejercer sus libertades y derechos fundamentales.²⁸ Las áreas prioritarias respecto a la educación de los formadores son: conocimiento y entendimiento de los medios de comunicación para la participación social, la evaluación de sus textos y su producción y uso.

ACCIONES DESDE LAS BIBLIOTECAS UNIVERSITARIAS

Se ha comenzado a debatir sobre el papel que han de tener las bibliotecas junto con los profesionales de la información como aliados en este proceso de desinformación.²⁹ Las bibliotecas han ejercido tradicionalmente tareas de *infomediación*, y ahora su función es más necesaria que nunca al ser un elemento facilitador en el proceso de reubicar y redefinir las nuevas alfabetizaciones mediáticas, fomentar el pensamiento crítico y ser un elemento clave en la defensa pública (*advocacy*) en la sociedad.³⁰

Al tener una predisposición hacia el formato impreso y no asumir la variedad de medios y fuentes en las que circula la información en la actualidad, las bibliotecas no se han comprometido con el desarrollo de las habilidades informativas como medida para combatir la desinformación. Las bibliotecas están limitadas para resolver la problemática de la desinformación por una falta de

28 Gema Alcolea-Díaz *et al.*, “Currículo de alfabetización mediática...”, 103-114.

29 Sara Martínez-Cardama y Laura Algora-Cancho, “Lucha contra la desinformación...”, e280412.

30 Mercedes Caridad-Sebastián *et al.*, “Infomediación y posverdad: el papel de las bibliotecas”, 891-898.

articulación de programas de investigación y por carecer de una investigación más profunda de lo que es la desinformación.³¹

La bibliografía especializada muestra el reciente interés por el asunto de la desinformación y las noticias falsas mediante investigaciones de tipo teórico,³² contribuciones que fomentan la adaptación de las multialfabetizaciones³³ y debates sobre asuntos como la neutralidad bibliotecaria.³⁴

En cuanto a las bibliotecas universitarias, el debate no se enfoca en la selección de información académica, sino en la necesaria colaboración con proveedores de información digital para que visibilicen mejor los recursos fiables y eviten la inclusión de contenido falso o tendencioso.³⁵ Esta autora llama a la necesaria cooperación entre bibliotecarios, proveedores y editores para favorecer el acceso a fuentes cuidadosamente “curadas” y seleccionadas.

El gran pilar de las actividades realizadas por parte de las bibliotecas universitarias es la *alfabetización informacional*, a través de dos vías: a) la implicación directa del personal bibliotecario en la formación de estudiantes y docentes; b) la elaboración de guías con recursos temáticos que, al margen de seleccionar contenidos, incluyan actividades que interpelen directamente al estudiante evaluando su sesgo de confirmación.³⁶

31 Jonathan Hernández Pérez, “La Desinformación: Aproximaciones desde la Investigación, las Asociaciones y el Quehacer Bibliotecario”, *Brecha entre Investigación y Práctica Bibliotecológica: Cómo Reducir la Distancia*, coordinado por Georgina Araceli Torres Vargas, México: UNAM, 2021: 56.

32 Nick Rochlin, “Fake News: Belief in Post-Truth”, *Library Hi Tech* 35, núm. 3 (2017): 386–392.

33 - Marcus Banks, “Fighting Fake News: How Libraries Can Lead the Way on Media Literacy”, *American Libraries*, 27 de diciembre de 2016.

- Ben Johnson, “Information Literacy Is Dead: The Role of Libraries in a Post-Truth World”, *Computers in Libraries* 37, núm. 2 (2017).

34 Mark Shores, “The Alert Collector: Collection Development in an Era of ‘Fake News’”, *Reference & User Services Quarterly* 57, núm. 3 (2018): 176–177.

35 Lisa Rose-Wiles, “Reflections on Fake News, Librarians, and Undergraduate Research”, *Reference & User Services Quarterly* 57, núm. 3 (2018): 200–204.

36 Sara Martínez-Cardama y Laura Algorta-Cancho, “Lucha contra la Desinformación...”.

Inteligencia artificial...

Dado que los bibliotecarios pueden intervenir en materia de desinformación, se sugieren las siguientes recomendaciones. En primer lugar, el bibliotecario debe conocer y actualizar las fuentes de información que tiene disponibles, tanto en formato físico como digital. En segundo lugar, debe prestar especial atención a la información que se comparte en las redes sociales institucionales. Lo anterior incluye la elaboración de materiales digitales para recordar constantemente a los usuarios verificar la información que consumen. En tercer lugar, debe implementar cursos de formación en AMI de manera frecuente con su comunidad, preferentemente de manera virtual para aumentar el alcance, con la firme intención de ser un defensor de los derechos de acceso a la información.

METODOLOGÍA

La investigación adoptó un enfoque metodológico mixto, combinando elementos cualitativos y cuantitativos para abordar de manera integral el fenómeno de la desinformación y las noticias falsas en redes sociales entre estudiantes universitarios. El supuesto que guía el trabajo de campo es que los jóvenes consumen información diversa y variada en las redes sociales y no se detienen a reflexionar sobre su consumo informativo y la veracidad o no de la información que consumen, además de compartirla en otras redes sin verificarla.

El enfoque mixto permitió obtener una comprensión profunda y multifacética del tema, al combinar las estrategias de recolección y análisis de datos cuantitativos —que proporcionan información estadística y numérica— con la exploración detallada de las percepciones y experiencias de los participantes a través de datos cualitativos, obtenidos mediante entrevista en un grupo focal. De esta manera, se pudo capturar tanto la amplitud como la profundidad del problema estudiado, lo que permitió delinear las estrategias que aplican para validar la información que consumen y difunden a través de las redes sociales.

Desde el enfoque cuantitativo se aplicaron encuestas con el objetivo de obtener una panorámica de sus percepciones, actitudes

y experiencias relacionadas con la desinformación. Las encuestas se enviaron a través de un grupo cerrado de WhatsApp que se creó exclusivamente para ese fin. El diseño de la encuesta incluyó preguntas cerradas y abiertas para recopilar datos a cerca de las necesidades informativas, el uso de redes sociales, la exposición a noticias falsas, las estrategias de verificación de información, la confianza en las fuentes de noticias y la percepción sobre la desinformación.

Para garantizar la representatividad de diferentes grupos demográficos dentro de la población estudiantil, se seleccionó una muestra de estudiantes de entre 18 y 25 años de edad de una universidad privada mediante un muestreo aleatorio estratificado. El universo de estudio fue de 230 estudiantes, con una muestra del 20%. Los estudiantes son del área de ingeniería de una universidad privada en la ciudad capital del estado de San Luis Potosí; las carreras que estudian son: Ingeniero industrial automotriz, Ingeniero en tecnología de información e Ingeniero en diseño gráfico.

Para la obtención de los datos cualitativos se seleccionaron 50 alumnos para llevar a cabo un grupo focal que duró aproximadamente 120 minutos. En el grupo focal se abordaron temas sobre necesidades informativas, el uso de las redes sociales y la verificación del contenido. El grupo se mostró muy participativo para hablar sobre sus necesidades informativas, las experiencias en el acceso a información a través de las redes sociales y la validación del contenido.

Para complementar el estudio se entrevistó a los bibliotecarios de esa institución educativa con la finalidad de indagar sobre el uso de redes sociales en su práctica profesional, las estrategias de verificación de información, la confianza en las fuentes de noticias, las percepciones sobre la desinformación y las acciones que fomentan entre los usuarios: alumnos, docentes y en general, el personal de la institución educativa.

Una vez que se obtuvieron los datos, se procedió al análisis de la información recabada. El análisis permitió detectar lo siguiente: los estudiantes acceden a información relacionada a sus intereses

personales y profesionales,³⁷ y sus búsquedas contienen un componente emocional, relacionado a los intereses personales de cada joven para acceder a la información que le resulta relevante. Los recursos que utilizan son diversos, seleccionados en función de su necesidad informativa. En el caso de las redes sociales, enfocan su uso en el entretenimiento o las usan para pasar el tiempo. En estas plataformas, las estrategias implementadas para verificar la veracidad del contenido son más intuitivas, de sentido común.

En cuanto a la biblioteca y las estrategias implementadas por el bibliotecario, destacan los cursos de formación sobre alfabetización mediática e informacional (un curso cada semestre donde participan estudiantes y docentes), además de las actividades de carácter informativo que se generan desde la biblioteca, tanto en el espacio físico como virtual. Tal es el caso de las infografías y enlaces a recursos y estrategias para la verificación de la información.

RESULTADOS

El análisis de los datos mostró que las personas acceden a una gran cantidad de información o selección de hechos, es decir, “una selección del mundo asociada a una señal que nos indica la ocurrencia de cierto acontecimiento”.³⁸ A esto se añade el componente emocional del acceso y uso de la información online,³⁹ para acceder a aquello que apela a la sensibilidad de los usuarios.

El concepto de *posverdad* se refiere a la combinación de información real y ficticia que los usuarios consumen o generan en la actualidad, y que toma mayor fuerza con el uso de las redes

37 Nicole Cooke, “Posttruth, Truthiness, and Alternative Facts: Information Behavior and Critical Information Consumption for a New Age”, *The Library Quarterly* 87, núm. 3 (2017): 211–221.

38 Enrique Herreras y Marina García-Granero, “Sobre Verdad, Mentira y Posverdad. Elementos para una Filosofía de la Información”, *Bajo Palabra*. II Época, núm. 24 (2020): 160.

39 - Nicole Cooke, “Posttruth, Truthiness, and Alternative Facts...”.
- Suau y Puertas-Graell, “Disinformation Narratives in Spain...” .

sociales donde se consume contenido engañoso, se comparte, se readapta a un formato diferente o se comenta al respecto, contribuyendo así a su difusión.⁴⁰ Esto se puede ejemplificar con el uso de WhatsApp como una red de mensajería instantánea donde los usuarios comparten con sus grupos de amigos información que es falsa o engañosa, pero que a ellos les da sentido. Por ejemplo, los mensajes enviados con contenido que atrapan la atención de la audiencia pero que contienen archivos malignos, y sin embargo, las personas los abren y comparten sin validar que sea información fidedigna.

Las teorías denominadas “filtros burbuja” (*filter bubbles*) y “cámaras de resonancia” (*echo chambers*) describen la manera en la que los usuarios encuentran información que refuerza y retroalimenta sus creencias.⁴¹ En este caso destaca el uso de TikTok, Facebook y X (Twitter), donde el algoritmo arroja contenido similar al consumido de acuerdo con sus intereses.

NECESIDADES DE INFORMACIÓN

Con el fin de identificar cómo los estudiantes universitarios validan la información que consumen en redes sociales, lo primero que se hizo fue investigar cuáles eran sus necesidades informativas. Como se aprecia en la tabla 1, destacan los temas relacionados con su formación profesional, sea desde el área de estudio o bien, desde aquellas materias en las cuales necesitan apoyo para comprender el contenido de las áreas específicas de su formación en ingeniería. La otra vertiente es la información de interés personal, que va desde los temas emocionales hasta las noticias, los deportes o la educación financiera.

40 Jaume Suau y David Puertas-Graell, “Disinformation Narratives in Spain...”.

41 Sara Martínez-Cardama y Laura Algorta-Cancho, “Lucha contra la Desinformación...”.

Inteligencia artificial...

Tabla 1. Necesidades informativas de los estudiantes universitarios, en el área de ingeniería

Temáticas	Contenido
Área de formación profesional	Ingeniería, motores, tecnología en automóviles, programación en distintos idiomas, mecánica de motos, diseño gráfico, tecnología de información
Áreas de apoyo a la formación profesional	Matemáticas, cálculo, estadística, diseño digital, inteligencia artificial, educación financiera
Temas emocionales	Mejorar en la vida, manejo de emociones, cómo controlar el estrés, cómo ser exitoso, predicciones para el futuro, por ejemplo, el horóscopo, los rituales para atraer energías positivas, entre otros
Temas de interés general	Diseño de espacios, idiomas (inglés, francés, alemán), ciberseguridad, noticias, deportes, automóviles, opciones laborales

Fuente: Elaboración propia.

En los temas relacionados con el apoyo a la formación profesional a nivel licenciatura, destaca la inteligencia artificial (IA) como un tema en tendencia. Las posturas de los estudiantes se orientan a identificar la forma en que pueden usarla para mejorar su experiencia educativa, por ejemplo, al hacer uso de IA para mejorar sus presentaciones, reducir tiempos de trabajo y resolver problemas.

Los temas emocionales son recurrentes entre los estudiantes universitarios; les interesa la información relacionada con temas de superación personal, de manejo de emociones y de predicción de situaciones. En las entrevistas señalaron que el consumo de esa información les genera seguridad sobre las situaciones que enfrentarán en su día a día. Al mismo tiempo, les proporciona sugerencias sobre cómo actuar frente a una diversidad de problemas.

En el manejo de emociones, destaca un mayor interés en personas del género femenino, sobre todo cuando experimentan una ruptura sentimental. Las diferentes plataformas les permiten consumir contenido que les ayuda a sentirse mejor, sin importar la objetividad de los testimonios. Aquí entra en juego el proceso

interpretativo, que le da mayor importancia al sentido —el hacia dónde y para qué— que es creado y reforzado, en el mismo proceso de interpretación.⁴²

Redes sociales

La tabla 2 muestra las redes sociales más utilizadas por los estudiantes, según la jerarquía de uso.

Tabla 2. Redes sociales a las que acceden los estudiantes de ingeniería

Recurso	%
WhatsApp	39
YouTube	24
TikTok	17
Facebook	12
X (Twitter)	8

Fuente: Elaboración propia.

Los estudiantes utilizan estas redes con fines de comunicación, entre familiares y amigos (WhatsApp y Facebook), para informarse (X, Twitter), para consultar contenidos relacionados con su formación (YouTube, WhatsApp y Facebook), para distraerse o entretenerse (Facebook, TikTok, YouTube). Las redes sociales son utilizadas para varios fines y los usuarios no se limitan a una sola red social, sino que interactúan con varias redes sociales al mismo tiempo, consumen y generan información, e intercambian opiniones con otras personas. Otro ejemplo del uso de las redes sociales es sortear la cotidianidad: mencionaron que cuando se enfrentan al tráfico en la ciudad, comienzan a interactuar en las redes para saber el motivo de tal situación, y en ocasiones elaboran y comparten información al momento, incluso antes que los medios

⁴² Enrique Herreras y Marina García-Granero, “Sobre Verdad, Mentira y Pos-verdad...”, 168.

informativos o las redes sociales de páginas oficiales, como las de la oficina de tránsito.

En la red social TikTok, los usuarios mostraron preferencia por las publicaciones que les dan recomendaciones sobre un tema específico; micro-tutoriales que tienen un estilo narrativo de acuerdo con las características de vídeo corto, como por ejemplo tutoriales de belleza, recetas de cocina, ejercicio físico, canciones, entre otros. También resaltaron la importancia de validar la información antes de poner los tutoriales en práctica. Un entrevistado dio como ejemplo un “truco” sobre cómo pegar pestañas postizas. Al verificar la información, pudo constatar que ese tipo de pegamento puede afectar la piel o incluso hasta la vista.

En el rubro sobre compartir información, el grupo focal destacó cómo los usuarios contribuyen a la desinformación de manera no intencionada, por criterios apoyados en los sentimientos y emociones. Los estudiantes reconocen que en ocasiones se enganchan con información falsa, ya sea que forme parte de la publicación original o de los comentarios que interpretan la misma. Cuando los usuarios comparten esta información se va generando una especie de bola de nieve y, pasada la emoción del episodio, los estudiantes reconocen que contribuyeron a la desinformación.

La información falsa se difunde más rápidamente y afecta de manera más intensa que las noticias verídicas: solo en la red social Twitter, las noticias falsas han tenido un 70% más de retweets que las verdaderas.⁴³ Esto aumenta la visibilidad y la capacidad de propagación de este tipo de noticias, lo que provoca que las audiencias se conviertan en colaboradoras en la difusión de contenido engañoso de manera inconsciente.⁴⁴

43 Soroush Vosoughi *et al.*, “The Spread of True and False News Online”, *Science* 359, núm. 6380 (2018): 1146–1151.

44 Marco Bastos y Dan Mercea, “The Public Accountability of Social Platforms: Lessons from a Study on Bots and Trolls in the Brexit Campaign”, *Philosophical Transactions of the Royal Society A. Mathematical, Physical and Engineering Sciences* 376, núm. 2128 (2018).

Tal como se muestra en la tabla 3, las redes sociales no son las únicas plataformas a las que los estudiantes acceden para buscar, localizar, leer y analizar información. También utilizan buscadores libres como Google, Yahoo!, Google académico y páginas oficiales de diferentes instituciones públicas o privadas, entre las que destacan los sitios web de universidades y bibliotecas digitales, foros, Wikipedia, libros impresos e inteligencia artificial.

Tabla 3. Otros recursos a los que acceden para obtener información

Recurso	%
Buscadores	36
Inteligencia artificial	18
Páginas oficiales	14
Wikipedia	10
Foros	9
Libros impresos	7
Google académico	6

Fuente: Elaboración propia.

Los estudiantes utilizan el buscador de su elección para acceder a las fuentes de información mencionadas, usando el lenguaje natural, con palabras que respondan a sus inquietudes. A partir de la búsqueda, si se trata de alguna actividad escolar, seleccionan la información para posteriormente hacer el análisis o preparar los documentos derivados de la investigación; si tienen alguna necesidad de información personal, consultan los documentos pertinentes a su necesidad; y, si se trata de información para entrenarse o distraerse, pasan el tiempo hasta que la siguiente actividad o el sueño indican la hora para salir de las plataformas.

Verificación de información

Cómo ya se ha señalado, en las redes sociales se puede destacar una información por encima de otra o aportar datos sin contexto

Inteligencia artificial...

para despistar a los lectores. A veces también ocurre que se aportan datos reales a los que se añaden otros falsos, con lo que se llega a conclusiones equivocadas. Esto provoca que la información se pueda malinterpretar y que los lectores solo se queden con una parte de los hechos explicados.⁴⁵ Diferentes autores han demostrado la capacidad de los ciudadanos para identificar contenido engañoso una vez expuesto a verificaciones de hechos.⁴⁶

Es así que nos dimos a la tarea de indagar cuáles son las estrategias que los estudiantes emplean para validar la información que consumen en las diferentes plataformas. El 74% de los encuestados verifican la información consumida, frente al 16% que no lo hace y un 10% que no contestó (véase la tabla 4).

Tabla 4. Verificación de contenidos

Verifican el contenido	Porcentaje (%)
Si	74
No	16
No contestó	10

Fuente: Elaboración propia.

Tanto en los resultados de la encuesta como en los del grupo focal se muestra que las estrategias para verificar la información son las siguientes: validar el tema de estudio, verificar autores y/o bibliografía, verificar la confiabilidad del sitio, así como el candado de seguridad y la fecha de publicación, revisar diferentes fuentes, incluyendo los comentarios de otros usuarios. Estos elementos se desglosan a continuación.

Para validar el tema de estudio, una vez que obtienen los resultados en el buscador, los comparan y se inclinan por los primeros que aparecen en la búsqueda inicial. Cuando encuentran

45 Elisa Pont, “Cómo Identificar la Desinformación en las Redes Sociales”.

46 Jaume Suau y David Puertas-Graell, “Disinformation Narratives in Spain: Reach, Impact and Spreading Patterns”.

disonancia en alguno, la respuesta inmediata es eliminarlo y consultar el siguiente. La minoría señaló que un resultado de este tipo les exigía detenerse para analizar otras posturas.

En la práctica de verificar autores y/o consultar otros documentos relacionados, los estudiantes usan los datos del autor para buscar información que sea relevante sobre el tema que están investigando. Señalaron el ejemplo de Wikipedia, que les da hipervínculos para acceder a información adicional, y de ahí tomar fuentes para complementar su investigación en el tema de estudio.

Para comprobar la confiabilidad del sitio, los estudiantes tienen distintos filtros e indicadores de confianza. En primer lugar, recurren a la recomendación de docentes o personas de confianza sobre direcciones URL seguras. En segundo lugar, verifican la presencia de un candado de seguridad. En tercer lugar, en el caso de la información financiera, señalaron seguir los tutoriales de las aplicaciones del banco de su preferencia para evitar el ingreso a páginas falsas y poner en riesgo su información.

En el grupo focal también se encontró que la fecha de publicación es un elemento importante para verificar la información del recurso que se está visualizando. Si la fecha tiene más de diez años de antigüedad, no la consumen y siguen buscando otros documentos. En el caso de los videos, la fecha relacionada con el contenido les da pistas para verificar la confiabilidad, sobre todo en lo que se refiere a temas de interés personal. Cuando se trata de videos con fines académicos, señalaron confiar en las recomendaciones de los docentes y no se cuestionan sobre la veracidad, dando por sentado que ese aspecto ya fue analizado por el docente.

Otra práctica habitual es la revisión de los comentarios para contrastar el contenido⁴⁷ y verificar la veracidad de la información. En general, la ausencia de comentarios que cuestionen la veracidad del contenido reduce la desconfianza.

En resumen, las estrategias registradas por los estudiantes universitarios se enfocan en el ámbito personal. Ni en los grupos

47 Pavel Sidorenko *et al.*, “Espacios de Verificación en TikTok...”.

Inteligencia artificial...

focales ni en las entrevistas se manifiestan las acciones que la biblioteca o la institución escolar han realizado en esa materia. El riesgo de contribuir con la desinformación, sea de manera fortuita o intencionada, se seguirá presentando y los esfuerzos que hacen diversas instituciones no son suficientes.

El bibliotecario y la difusión de estrategias para la verificación de información

Una de las actividades principales que realizan las bibliotecas para formar a los usuarios en el uso eficiente de la información son los cursos sobre alfabetización informativa o AMI. En las bibliotecas se emprenden campañas para capacitar a la ciudadanía en competencias digitales y promover prácticas seguras en el consumo informativo.⁴⁸ En 2018, la *Declaración de IFLA sobre noticias falsas*, insistió en llevar a cabo programas de alfabetización informacional y mediática en todos los niveles y para todas las edades.

Los usuarios necesitan desarrollar diversas competencias que estén relacionadas con “la interpretación crítica de mensajes, evaluación de fuentes, inteligencia colectiva, la participación democrática, acceso e interacción con los medios digitales, utilización de información de manera ética, eficaz y eficiente”.⁴⁹ Todo ello se realiza a través de la biblioteca con cursos sobre competencias digitales. En esta institución, estos cursos se realizan dos veces al año; la institución reconoce que no son suficientes, y plantea que sería más adecuado un curso semestral que atendiera al menos un grupo cada semestre.

En ese sentido, la colaboración entre biblioteca e institución educativa, o bien, entre bibliotecario y docente es fundamental para implementar acciones más formales. Se aboga por ejemplo por cursos ordinarios a lo largo del período formativo o semestre, o bien por cursos esporádicos que se gestionen desde la biblioteca y atiendan a toda la comunidad escolar. En el grupo focal no

48 Ramón Salaverría y Gustavo Cardoso, “Future of Disinformation Studies...”.

49 William Arcila *et al.*, “Tendencias Investigativas...”, 225–236.

aparece la figura del bibliotecario como un agente que haya contribuido a comprender que, antes de tomar cualquier información como referente, se debe verificar, en particular, cuando el usuario se apropia de la misma y la comparte con otros usuarios de las redes sociales.

Otra de las acciones que se pueden llevar a cabo es organizar actividades de lectura, escritura y habilidades digitales. El objetivo es que la comunidad escolar tenga las herramientas para realizar una jerarquía de interpretaciones, misma que se aprecia en el arte de leer bien. Es decir, que los estudiantes puedan leer los hechos sin falsearlos con la interpretación y sin perder, por deseo de entenderlos, la precaución, la sutileza, la paciencia.⁵⁰ Lo que corresponde al bibliotecario es promover entre los usuarios el análisis de la información para emitir juicios con sustento, aunque finalmente, es quien accede a la información quien hace ese proceso interpretativo.

A nivel internacional, las asociaciones de bibliotecarios han elaborado infografías. En 2017, por ejemplo, la Federación Internacional de Bibliotecas y Asociaciones Bibliotecarias (IFLA) desarrolló una infografía sobre cómo detectar noticias falsas.⁵¹ La infografía se difundió entre la comunidad de la biblioteca estudiada a través de las redes sociales y con los usuarios para que la tengan presente y la utilicen al momento de acceder a información. Este aspecto no se vio reflejado en el grupo focal; los estudiantes manifestaron revisar las redes sociales de la biblioteca, pero no mencionaron la infografía de IFLA ni la relacionaron con las noticias falsas.

Otra de las estrategias a implementar es la elaboración de guías temáticas que orientan sobre las publicaciones depredadoras, así como también recomendaciones de algunos sitios que se dan a la tarea de identificar las publicaciones falsas. Este aspecto que enfatizaron los bibliotecarios no fue señalado por los estudiantes.

50 Enrique Herreras y Marina García-Granero, "Sobre Verdad, Mentira y Posverdad...", 168.

51 Sara Martínez-Cardama y Laura Algóra-Cancho, "Lucha contra la Desinformación...".

Inteligencia artificial...

Se comprueba aquí lo que ya se mencionó, los bibliotecarios enfocan sus esfuerzos desde la biblioteca, pero no han logrado presencia en otros espacios como las redes sociales institucionales y la verificación de información.

Finalmente destaca el uso de redes sociales para difundir las estrategias mencionadas en este apartado y promover actividades de la biblioteca. Con ello, logra impactarse a alumnos de varias licenciaturas, aunque a la fecha no se puede afirmar que los cursos que se han llevado a cabo impactan a toda la comunidad ni que las acciones realizadas han sido las más adecuadas para dotar de herramientas a los estudiantes para verificar la información.

CONCLUSIONES

En este capítulo se ha presentado un panorama sobre el tipo de información que consumen los jóvenes y los medios a los que acceden. Es importante reforzar las habilidades informativas y mediáticas entre los estudiantes universitarios, pues es parte de su formación y tendrá un gran impacto en su vida profesional. De forma paralela, estas habilidades de validación de contenido también son importantes para el contenido que consumen por interés personal, pues les servirá para tomar decisiones informadas, así como para evitar la desinformación y la propagación de información falsa.

Los estudiantes ya cuentan con estrategias como la verificación de fuentes, revisar los comentarios, consultar otros recursos y, sobre todo, utilizar el pensamiento crítico para decidir sobre la veracidad o no del contenido. Desde las bibliotecas, es necesario reforzar las estrategias de protección ante los mensajes deliberadamente falsos para proporcionar a los usuarios herramientas de verificación de información. Es indispensable que desde la biblioteca se generen, seleccionen o adapten los recursos informativos con temas específicos en materia de información digital.

Además del trabajo de AMI, se deben incrementar las acciones para generar un sentido de responsabilidad en la ciudadanía, con el fin de controlar la difusión de información falsa.

Las acciones implementadas desde la biblioteca tendrán mayor impacto si se hacen en colaboración con los docentes. De esa forma se logrará llegar a toda la comunidad escolar y no sólo a los que por iniciativa personal participen en los cursos. Además, otro sector de la población al que es indispensable impartir las competencias en AMI son los docentes, ya que su labor orientativa tendrá mayor impacto en la comunidad si ellos mismos saben buscar, seleccionar, evaluar y utilizar la información aplicando las estrategias para verificar la veracidad de la misma.

Una de las perspectivas que se podría analizar con mayor detalle en futuras investigaciones es la verificación de información en temas emocionales y de superación personal. Según los resultados de la presente investigación, los jóvenes recurren a videos en las redes sociales para obtener herramientas psicosociales y tratar su sintomatología, aunque la información no sea vertida por un especialista; eso puede generar otro problema mucho más grave a futuro.

BIBLIOGRAFÍA

- Aguaded, Ignacio *et al.* "Media Literacy from International Organizations in Europe and Latin America". *The Journal of Media Literacy* 63, núm. 1-2 (2016): 10–17.
- Alcolea-Díaz, Gema *et al.* "Currículo de Alfabetización Mediática e Informacional de la UNESCO para Profesores desde la Perspectiva de la Estructura de la Información". *Comunicar, Revista Científica de Educomunicación*, núm. 62 (2020): 103-114. <https://doi.org/10.3916/C62-2020-09>.
- Arcila, William *et al.* "Tendencias Investigativas en los Estudios sobre Alfabetización Mediática Informacional y Digital (AMID) en el Campo Educativo". *Revista Complutense de Educación* 33, núm. 2 (2022): 225–236. <https://doi.org/10.5209/rced.73935>.

Inteligencia artificial...

- Avila, F., y Odilia S. *La Posverdad como Cambio de la Contextualización de la Opinión Pública: Medios Digitales el Vocero y Bacán*. Tesis de licenciatura, Universidad Estatal Península de Santa Elena, 2024.
- Banks, Marcus. “Fighting Fake News: How Libraries Can Lead the Way on Media Literacy”. *American Libraries*, 27 de diciembre de 2016. <https://americanlibrariesmagazine.org/2016/12/27/fighting-fake-news>.
- Bastos, Marco, y Dan Mercea. “The Public Accountability of Social Platforms: Lessons from a Study on Bots and Trolls in the Brexit Campaign”. *Philosophical Transactions of the Royal Society A. Mathematical, Physical and Engineering Sciences* 376, núm. 2128 (2018): 20180003. <https://doi.org/10.1098/rsta.2018.0003>.
- Batchelor, Oliver. “Getting Out the Truth: The Role of Libraries in the Fight Against Fake News”. *Reference Services Review* 45, núm. 2 (2017): 143–148. <https://doi.org/10.1108/RSR-03-2017-0006>.
- Caridad-Sebastián, Mercedes *et al.* “Infomediación y Posverdad: El Papel de las Bibliotecas”. *El Profesional de la Información* 27, núm. 3 (2018): 891–898. <https://doi.org/10.3145/epi.2018.jul.17>.
- Castillo, Cecilia *et al.* “La Veracidad de la Información Expuesta en las Redes Sociales”. *Revista Educación y Tecnología* 8, núm. 2 (2016): 26–40. Consultado el 15 de octubre de 2024. <https://dialnet.unirioja.es/descarga/articulo/6148884.pdf>.
- Cobo, Juan Cristóbal. “El Concepto de Tecnologías de la Información. Benchmarking sobre las Definiciones de las TIC en la Sociedad del Conocimiento”. *Zer* 14, núm. 27 (2009): 295-318. <https://addi.ehu.es/handle/10810/40999>.
- Cooke, Nicole. “Posttruth, Truthiness, and Alternative Facts: Information Behavior and Critical Information Consumption for a New Age”. *The Library Quarterly* 87, núm. 3 (2017): 211–221. <https://doi.org/10.1086/692298>.
- Córdova, Erincson. “La Tecnología en las Redes Sociales”. *Opinión*, Universidad de Piura, 20 de septiembre de 2021. Consultado el 9 de septiembre de 2024. <https://www.udep.edu.pe/hoy/2021/09/tecnologia-en-redes-sociales>.

- De la Paz, Jorge Carlos *et al.* “Estrategia de Comunicación en Redes Sociales. Escenarios Teórico Prácticos para los Medios de Comunicación en Cuba”. *Razón y Palabra*, núm. 92 (2015): 1–40. Consultado el 12 de septiembre de 2024. <https://www.revisitarazonypalabra.org/index.php/ryp/article/view/69>
- Estévez Gualda, Javier, y Amanda García Marín. “Las Redes Sociales para la Mejora de la Capacidad de Emprender y de Autoempleo”. *International Journal of Educational Research and Innovation* (2014): 101–110. Consultado el 12 de septiembre de 2024. <https://www.upo.es/revistas/index.php/IJERI/article/view/1462/1177>.
- Finley, Wayne *et al.* “Fake News: An Opportunity for Real Librarianship”. *ILA Reporter* 35, núm. 3 (2017): 8–12. <https://go.gl/i4C1BV>.
- Garrido Cabezas, Norman. “La democracia electrónica y sus tipologías”. En *Aplicaciones actuales de la comunicación e interacción digitales*. Madrid: Editorial ACCI, 2015, 253-272. PDF. Academia.edu. https://www.academia.edu/26268796/LA_DEMOCRACIA_ELECTRONICA_Y_SUS_TIPOLOGIAS.
- Hernández Pérez, Jonathan. “La Desinformación: Aproximaciones desde la Investigación, las Asociaciones y el Quehacer Bibliotecario”. *Brecha entre Investigación y Práctica Bibliotecológica: Cómo Reducir la Distancia*. Coordinado por Georgina Araceli Torres Vargas. México: UNAM, 2021. <http://sistemas.iibi.unam.mx/publica20/resplibros.php?tit=393>.
- Herreras, Enrique, y Marina García-Granero. “Sobre Verdad, Mentira y Posverdad. Elementos para una Filosofía de la Información”. *Bajo Palabra*. II Época, núm. 24 (2020): 157–176. <https://doi.org/10.15366/bp.2020.24.008>.
- Johnson, Ben. “Information Literacy Is Dead: The Role of Libraries in a Post-Truth World”. *Computers in Libraries* 37, núm. 2 (2017). <http://www.infoday.com/cilmag/mar17/Johnson--Information-Literacy-Is-Dead--The-Role-of-Libraries-in-a-Post-Truth-World.shtml>.

- Juca, F., y A. A. Juca. “La desinformación en las redes sociales y su influencia en los jóvenes del siglo XXI: una mirada desde la academia”. *Conference Proceeding UTMACH* 4, núm. 1 (2022): 144–153. <https://investigacion.utmachala.edu.ec/proceedings/index.php/utmach/article/view/437/377>.
- Jumbo, L. B. *Uso de la Red Social Facebook como Instrumento para Generar “Noticias Falsas” y su Efecto en la Opinión Pública*. Tesis de licenciatura, Universidad de Guayaquil, 2021.
- Kapantai, Eleni, *et al.* “A Systematic Literature Review on Disinformation: Toward a Unified Taxonomical Framework”. *New Media & Society* 23, núm. 5 (2021): 1301–1326. <https://doi.org/10.1177/1461444820959296>.
- López-Borrull, Alexandre. “En Busca de la Verdad Perdida: Redes Sociales y Desinformación”. *Anuario ThinkEPI* 17 (2023): e17a44. <https://doi.org/10.3145/thinkepi.2023.e17a44>.
- Martínez-Cardama, Sara, y Laura Algora-Cancho. “Lucha contra la Desinformación desde las Bibliotecas Universitarias”. *El Profesional de la Información* 28, núm. 4 (2019): e280412. <https://doi.org/10.3145/epi.2019.jul.12>.
- McNair, Brian. *Fake News: Falsehood, Fabrication and Fantasy in Journalism*. Londres: Routledge, 2018.
- Montero C., Laura “Reflexiones sobre la Investigación en Redes Sociales: Facebook y Twitter”. En Segado-Boj, F. y Llovens Sobrado, B. *I Congreso Internacional de Comunicación y Sociedad Digital*. España: UNIR, 2013. <https://dialnet.unirioja.es/servlet/articulo?codigo=4565937>.
- Pérez-Escolar, Marta *et al.* “A Systematic Literature Review of the Phenomenon of Disinformation and Misinformation”. *Media and Communication* 11, núm. 2 (2023): 76–87. <https://doi.org/10.17645/mac.v11i2.6453>.
- Pont, Elisa. “Cómo Identificar la Desinformación en las Redes Sociales”. *La Vanguardia*, 3 de julio de 2020. Consultado el 20 de julio de 2022. <https://www.lavanguardia.com/vida/junior-report/20200703/482021984534/noticias-falsas-bromas-deep-fakes-redes-sociales-desinformacion.html>.

- Rochlin, Nick. "Fake News: Belief in Post-Truth". *Library Hi Tech* 35, núm. 3 (2017): 386–392. <https://doi.org/10.1108/LHT-03-2017-0062>.
- Rose-Wiles, Lisa. "Reflections on Fake News, Librarians, and Undergraduate Research". *Reference & User Services Quarterly* 57, núm. 3 (2018): 200–204. <https://doi.org/10.5860/rusq.57.3.6606>.
- Sádaba, Charo, y Ramón Salaverría. "Combatir la Desinformación con Alfabetización Mediática: Análisis de las Tendencias en la Unión Europea". *Revista Latina de Comunicación Social*, núm. 81 (2023): 17–33. <https://doi.org/10.4185/rlds-2023-1552>.
- Salaverría, Ramón, y Gustavo Cardoso. "Future of Disinformation Studies: Emerging Research Fields". *El Profesional de la Información* 32, núm. 5 (2023): e320525. <https://doi.org/10.3145/epi.2023.sep.25>.
- Sidorenko Bautista, Pavel, *et al.* "Espacios de Verificación en TikTok. Comunicación y Formas Narrativas para Combatir la Desinformación". *Revista Latina de Comunicación Social*, núm. 79 (2021): 87–113. <https://doi.org/10.4185/RLCS-2021-1522>.
- Shores, Mark. "The Alert Collector: Collection Development in an Era of 'Fake News'". *Reference & User Services Quarterly* 57, núm. 3 (2018): 176–177. <https://doi.org/10.5860/rusq.57.3.6601>.
- Suau, Jaume, y David Puertas-Graell. "Disinformation Narratives in Spain: Reach, Impact and Spreading Patterns". *Profesional de la Información* 32, núm. 5 (2023): e320508. <https://doi.org/10.3145/epi.2023.sep.08>.
- Urgilés Pineda, Mario Wilfrido, *et al.* "¿Informar o Desinformar? El Poder Comunicacional de las Redes Sociales". *RECIMUNDO* 6, núm. 4 (2022): 144–152. [https://doi.org/10.26820/recimundo/6.\(4\).octubre.2022.144-152](https://doi.org/10.26820/recimundo/6.(4).octubre.2022.144-152).
- Vosoughi, Soroush *et al.* "The Spread of True and False News Online". *Science* 359, núm. 6380 (2018): 1146–1151. <https://doi.org/10.1126/science.aap9559>.

Alfabetización informacional, medios e inteligencia artificial: reflexiones sobre las interrelaciones orientadas a la construcción y difusión del conocimiento académico en la era digital

REGINA C. B. BELLUZZO

VÂNIA C. P. N. VALENTE

*Postgrado en Medios y Tecnología,
Universidad Estatal de São Paulo*

INTRODUCCIÓN

Los desafíos, tendencias e impactos de la era digital representan una verdadera revolución en el acceso a la información y en las formas de interacción y comunicación. Esta nueva etapa trajo consigo una serie de avances tecnológicos que transformaron profundamente la sociedad contemporánea. El internet, en particular, ha revolucionado la forma en que las personas se comunican, trabajan, se educan y se divierten. Uno de los principales impactos de la era digital es la velocidad y facilidad con la que se transmite y se accede a la información. Hoy en día basta con unos cuantos clics para consultar prácticamente cualquier dato. Esto ha contribuido a la democratización del conocimiento. Además, la era digital ha transformado la forma en que las personas interactúan. Las redes sociales y las aplicaciones de mensajería instantánea permiten la comunicación en tiempo real, independientemente de la distancia física. La Inteligencia Artificial también ha tenido un impacto en los medios de comunicación. Un ejemplo reciente se observa en los Juegos Olímpicos de París, donde la IA se utilizó en las transmisiones oficiales del Comité Olímpico Internacional

Inteligencia artificial...

mediante repeticiones en 360° y mejoras operativas. Se espera que esta tecnología permita proporcionar información en tiempo real sobre el rendimiento de los deportistas, con sistemas capaces de generar análisis e *insights* a partir de las imágenes capturadas en cada modalidad.

La sociedad contemporánea está marcada por la enorme velocidad y volumen de información que circula en fracciones de segundos. Estamos inmersos en la era de la información y el conocimiento, en la que los datos se han convertido en una materia prima imprescindible. Por ello, es necesario desarrollar habilidades que permitan a las personas aprovechar este caudal de información de manera inteligente, eficiente y conforme a criterios éticos y legales, especialmente en un contexto donde la inteligencia artificial forma parte de la realidad de nuestro país.

El desarrollo de la tecnología no solo ha generado nuevas herramientas, sino también una nueva cultura. El progreso y las innovaciones tecnológicas han transformado con rapidez el modo de vida de la sociedad, las formas de educar y aprender, las concepciones de la educación y las cualificaciones. “Además de simples cambios, esta llegada tecnológica se ha caracterizado como un fenómeno que muchas veces impone a la sociedad moderna hábitos y comportamientos diferentes, transformando la relación del ser humano con los demás, con el medio ambiente y consigo mismo”.¹

Las tecnologías de IA, especialmente la IA generativa, tienen un potencial significativo para mejorar el acceso a la información y fomentar la transparencia en los resultados de la investigación científica. En términos generales, la IA generativa también tiene el potencial de transformar los panoramas de la información y las comunidades a las que las bibliotecas de investigación apoyan y sirven. La creciente disponibilidad de modelos de IA abre muchas

1 Marlene Neves Strey y Renata Chabar Kapitanski, *Educação & internet: a era da informação e a vida cotidiana*, São Leopoldo: Editora Sinodal, 2011: 55.

posibilidades y plantea diversas consideraciones éticas, profesionales y legales.²

Al producir contenidos como textos, imágenes e incluso soluciones a problemas complejos, la IA generativa puede democratizar el acceso al conocimiento y hacerlo más comprensible para audiencias amplias. Además, su capacidad de automatizar el análisis de grandes volúmenes de datos permite acelerar el proceso de descubrimiento científico. Esto implica una mayor eficiencia en las investigaciones y nuevas formas de explorar datos que antes eran impracticables.

No obstante estos beneficios vienen acompañados de desafíos importantes que hay que considerar cuidadosamente. Uno de los principales riesgos es la alteración de los ecosistemas de información en los que se insertan las investigaciones científicas. El uso indiscriminado de la IA puede facilitar la difusión de datos incorrectos o sesgados, afectando la integridad de los resultados.

Otro aspecto crucial son las implicaciones éticas, profesionales y legales asociadas con el uso de la IA en la investigación. La creciente disponibilidad de modelos de IA plantea interrogantes sobre los derechos de autor, la propiedad intelectual y la responsabilidad en el uso de tecnologías emergentes.

Por lo tanto, aunque la IA generativa ofrece oportunidades innovadoras para el avance de la ciencia, es imperativo que estas tecnologías se implementen con cautela, teniendo en cuenta las complejidades éticas y legales que acompañan su desarrollo y aplicación.

OBJETIVOS

Este trabajo busca desarrollar reflexiones sobre la relación entre la competencia en información y medios y la inteligencia artificial,

2 Association of Research Libraries, “Research Libraries Guiding Principles for Artificial Intelligence”, abril de 2024, <https://doi.org/10.29242/principles.ai2024>, 1.

en particular sobre la aplicación de ChatGPT en la construcción y difusión del conocimiento académico. Este planteamiento es resultado de un curso impartido en el 2.º semestre de 2023 con el Programa de Postgrado en Medios y Tecnología (PPGMiT-UNESP-Bauru-SP).

Se consideró que articular un conjunto de principios rectores para la investigación centrada en la IA es útil para: influir en las políticas, abogar por el desarrollo y la implementación responsable de tecnologías, fomentar prácticas éticas y transparentes y fortalecer la confianza en los entornos de investigación. El objetivo fue integrar estos principios en la formación de posgrado como marco para un uso crítico y responsable de la IA.

Ahora bien, la tecnología por sí sola no provoca cambios en la vida de los usuarios: es necesario saber utilizarla para así “alcanzar sus propios propósitos en diferentes contextos”.³ En este sentido, Rojo⁴ sostiene que la población debe estar preparada para vivir en una sociedad cada vez más digital. Los usuarios de las sociedades actuales deben “buscar en el ciberespacio un lugar donde encontrar, de manera crítica, las diferencias y las múltiples identidades”.

Si el funcionamiento de la sociedad y sus prácticas sociales han cambiado, también deben hacerlo las habilidades y capacidades de los ciudadanos. Estas nuevas formas de ver y entender el mundo han dado lugar a nuevas prácticas sociales mediadas por las tecnologías. Por ello, es fundamental que los usuarios de las ciudades actuales desarrollen habilidades y se incluyan en la realización de estas prácticas, evitando así desconectarse en una sociedad hiperconectada.

Hoy más que nunca se requieren capacidades no solo técnicas, sino también adaptativas, que permitan afrontar una sociedad

3 David Barton y Carmen Lee, *Linguagem online: textos e práticas digitais*, São Paulo: Parábola Editorial, 2015: 13.

4 Roxane Rojo, “Gêneros discursivos do Círculo de Bakhtin e multiletramentos”, en *Escol@ Conectada: os multiletramentos e as TICs*, editado por Roxane Rojo, 13–36. São Paulo: Parábola Editorial, 2013: 7.

compleja y en permanente transformación. Se trata de habilidades fundamentales para la vida: conocer, saber estar, resolver problemas y enfrentar de manera reflexiva los desafíos del día a día.⁵

DESARROLLO DE LA DISCIPLINA

El desarrollo de la disciplina partió de reconocer que es fundamental enseñar a los estudiantes a acceder, emplear y transformar información para contribuir a la construcción de conocimiento y fortalecer sus habilidades de investigación y escritura académica. En este proceso, cobra relevancia su interacción con una herramienta cada vez más popular: ChatGPT, el modelo de inteligencia artificial desarrollado por OpenAI.⁶

Cabe señalar que estas herramientas también implican desafíos éticos y prácticos, como la falta de mención de las fuentes, la preocupación por los derechos de autor y la privacidad de los usuarios.

Como parte de las actividades pedagógicas, se realizó una exposición teórica que abarcó los siguientes temas:

- I. La era digital y sus relaciones con la información, el conocimiento y las tecnologías. (*De los medios tradicionales a los emergentes: formatos digitales y alternativos; Conocimiento e información como bienes sociales; y Tríada: información, conocimiento y tecnología*).
- II. Lenguajes y representación en los nuevos medios y la información. (*La era de la infodemia; Códigos y lenguajes digitales*);

5 Patricia Sá y Fátima Paixão, “Competências-chave para todos no séc. XXI: orientações emergentes do contexto europeu”, *Interacções* 11, núm. 39 (2015): 245.

6 Emma Sabzalieva y Arianna Valentini, *ChatGPT e inteligencia artificial en la educación superior: Guía de inicio rápido*, París: UNESCO; Caracas: Instituto Internacional de la UNESCO para la Educación Superior en América Latina y el Caribe [IESALC], 2023.

Inteligencia artificial...

Ética y responsabilidad en la comunicación de contenidos visuales, textuales y multimodales; Desinformación y fake news; y Causas y consecuencias de la desinformación).

- III. Paradigmas de información, desinformación y generación de conocimiento digital: demandas de inteligencia artificial y habilidades. (*Habilidades: orígenes y notas principales; Alfabetización informacional: Dimensiones, orígenes, ciclo, estándares y enfoques; Competencia mediática: una sesión informativa; Competencia digital: qué es y qué trae consigo; y Principales recomendaciones por competencias*).
- IV. Contextos de intervención de la competencia informacional, la competencia mediática y la competencia digital: algunos aspectos destacados. (*ChatGPT; y Panel de interacción: una construcción colectiva*).

Las clases también incluyeron una parte práctica donde se empleó ChatGPT en función de lo que los alumnos necesitaban saber:

- Seleccionar información relevante a partir de la comprensión de las ideas contenidas en las fuentes de información.
- Reformular conceptos con sus propias palabras.
- Identificar textualmente si la información fue adecuadamente transcrita o es una paráfrasis.
- Determinar si la información obtenida es suficiente y adecuada o si se requiere más.
- Definir cuestiones básicas de confiabilidad: autor, propósito del mensaje y forma de construcción.
- Utilizar un lenguaje, estilo y forma de escritura adecuados, con indicación correcta y coherente de las fuentes consultadas.
- Identificar elementos de citación de las fuentes consultadas en varios formatos.
- Demostrar comprensión de los estándares de documentación recomendados para su área de estudio y comprender la importancia de recopilar, organizar, preservar y compartir conocimientos e información de forma responsable, ética y legal.

Estos principios se aplicaron de forma práctica e integrada en diversas actividades a lo largo del curso, apuntando no solo a la asimilación teórica, sino también a la aplicación efectiva de las habilidades desarrolladas en contextos reales de investigación. Las actividades fueron diseñadas para promover la participación activa de los estudiantes y brindar una experiencia de aprendizaje rica y multifacética:

- I. *Actividades con ChatGPT.* Inicialmente, los estudiantes trabajaron con ChatGPT, una herramienta avanzada de inteligencia artificial, a través de la definición de temas y preguntas de búsqueda específicas. Esta actividad les permitió explorar el potencial de la IA para generar respuestas iniciales a preguntas de investigación, profundizando mediante búsquedas sucesivas y consultas refinadas. A través de este proceso, los participantes aprendieron a optimizar sus interacciones con la IA, desarrollando habilidades críticas para evaluar la relevancia y precisión de la información proporcionada por la herramienta. La actividad puso de relieve la importancia del pensamiento crítico y la intervención humana en la interpretación de los resultados generados por los sistemas de IA.
- II. *Taller de casos con dinámicas de grupo.* Posteriormente, se realizó un taller práctico centrado en diferentes casos en los que podría utilizarse de la IA en la investigación académica. El tema central de este taller fue “Casos de uso de la IA en la investigación: contextos y herramientas de intervención”. Los estudiantes se dividieron en grupos para explorar y discutir el uso de tres herramientas diferentes:
 - Taller 1: Uso de la herramienta Rayyan para realizar revisiones sistemáticas de literatura. Los participantes analizaron cómo esta plataforma puede optimizar el proceso de selección y análisis de la literatura, discutiendo sus aplicaciones prácticas y sus limitaciones.

Inteligencia artificial...

- Taller 2: Uso de la herramienta Research Rabbit para buscar artículos de investigación. En esta sesión se exploró cómo la herramienta facilita el descubrimiento de literatura relevante, destacando sus capacidades de mapeo de citas y la forma en que puede mejorar la eficiencia de las revisiones de la literatura.
- Taller 3: Uso de la herramienta ScisPACE para acceder, conocer y comprender cualquier publicación de investigación. Este caso permitió a los estudiantes comprender cómo se puede utilizar la IA para simplificar la lectura e interpretación de artículos científicos, propiciando que el conocimiento sea más accesible.

Al finalizar el taller, se llevó a cabo una actividad integradora que permitió a los grupos compartir sus descubrimientos y reflexiones, promoviendo un diálogo colectivo sobre las oportunidades y desafíos del uso de la IA en diferentes contextos de investigación.

III. Panel de interacción como seminario final de la disciplina. La culminación del curso estuvo marcada por un panel de interacción, estructurado como un seminario final. En esta actividad los estudiantes se dividieron en grupos, cada uno responsable de elegir y explorar temas específicos relacionados con el contenido enseñado a lo largo del curso. Este panel representó una construcción colectiva de conocimiento, donde los estudiantes no solo presentaron sus reflexiones y análisis, sino que también participaron en discusiones críticas y colaborativas con sus colegas. Este formato permitió a los estudiantes demostrar su comprensión de los conceptos tratados y aplicar las habilidades desarrolladas, especialmente en relación con el uso ético y eficaz de las herramientas de IA en la investigación.

Estas actividades ejemplifican cómo los principios teóricos discutidos en el disciplina se transformaron en prácticas docentes que fomentaron el aprendizaje activo y participativo. Al integrar el uso de herramientas de IA a los debates y reflexión crítica, el curso no solo dotó a los estudiantes de habilidades técnicas sino que también fomentó una comprensión profunda de las implicaciones y responsabilidades asociadas con el uso de la IA en la investigación académica. Esto aseguró que los estudiantes estuvieran preparados para aplicar estas tecnologías de manera informada y ética en sus futuras actividades de investigación.

HERRAMIENTAS DE IA UTILIZADAS

A continuación se describen brevemente las funcionalidades de las herramientas de IA aplicadas en los talleres temáticos.

Rayyan es una aplicación web desarrollada y financiada por Qatar Foundation, una organización sin fines de lucro con sede en Qatar cuyo objetivo es promover la educación, la ciencia y el desarrollo comunitario. Esta aplicación se ha convertido en una herramienta indispensable en el campo de la investigación académica, siendo ampliamente utilizada para facilitar revisiones sistemáticas y metanálisis, metodologías esenciales en la síntesis de evidencia científica. Las revisiones sistemáticas, por su naturaleza, requieren un análisis minucioso y exhaustivo de un gran volumen de literatura científica, lo que puede ser un proceso que requiere mucho tiempo y es susceptible a errores humanos.

Esta aplicación proporciona un entorno de trabajo colaborativo e intuitivo, donde los investigadores pueden importar, filtrar, organizar y revisar referencias bibliográficas de forma eficiente y estructurada. La plataforma utiliza tecnología de aprendizaje automático para sugerir la inclusión o exclusión de estudios, optimizando el proceso de selección y permitiendo a los investigadores centrar sus esfuerzos en análisis críticos más profundos. Además, el software ofrece funciones como marcar y anotar textos, categorizar estudios en función de criterios específicos y la posibilidad

Inteligencia artificial...

de resolver conflictos entre revisores, todo dentro de una interfaz accesible.

El apoyo de la Fundación Qatar para el desarrollo de Rayyan refleja su compromiso con crear herramientas innovadoras que amplíen la capacidad de los investigadores para realizar estudios rigurosos y de alta calidad. Al mejorar la eficiencia y precisión de las revisiones sistemáticas y los metanálisis, Rayyan contribuye significativamente a la confiabilidad de las conclusiones derivadas de estas metodologías, que son fundamentales para la toma de decisiones basadas en evidencia en diversas áreas del conocimiento.

En síntesis, Rayyan no solo simplifica los complejos procesos de revisión, sino que también eleva los estándares de la investigación científica al reducir el sesgo y aumentar la transparencia en las etapas de selección y análisis de los estudios. Su creciente inclusión en ámbitos académicos y científicos ilustra su valor como una herramienta que mejora la eficiencia y precisión de las revisiones sistemáticas y los metanálisis. Rayyan es vital en el arsenal de investigadores comprometidos con la excelencia y la integridad científica. Esta herramienta está disponible en: <https://bibliotecas.ufpr.br/2022/03/app-rayyan/>.

ResearchRabbit es una herramienta de mapeo de literatura basada en citas que ha mejorado la forma en que los investigadores realizan sus búsquedas bibliográficas. La herramienta está diseñada para optimizar el tiempo dedicado a la búsqueda de referencias, especialmente durante las etapas iniciales de planificación de ensayos, proyectos de menor escala o revisiones de literatura. A diferencia de las bases de datos académicas tradicionales, esta plataforma se distingue por su modelo dinámico e interactivo de descubrimiento de la literatura, que permite a los investigadores navegar intuitivamente por las complejas redes de citas que conectan los estudios existentes.

El concepto detrás de ResearchRabbit es simple e intuitivo: el usuario comienza seleccionando uno o más artículos, que sirven como punto de partida para la búsqueda. A partir de estas referencias, la herramienta aplica algoritmos de aprendizaje automático

para identificar otros artículos relevantes conectados con el tema de interés por citas directas, co-citaciones o afinidad temática. Esto amplía con rapidez y precisión el campo de búsqueda, revelando estudios que podrían pasar inadvertidos en las búsquedas con métodos tradicionales.

ResearchRabbit también ofrece una interfaz visual que facilita la organización y el análisis de las relaciones entre artículos. Los investigadores pueden visualizar cómo los estudios se conectan entre sí, identificando rápidamente los trabajos más influyentes en un campo determinado o aquellos que ocupan posiciones clave dentro de la red de citas. Este enfoque no solo ayuda a ahorrar tiempo sino que también proporciona una comprensión más profunda del desarrollo y las interconexiones dentro de un campo de estudio.

Otro aspecto que distingue a ResearchRabbit es su capacidad para aprender y adaptarse a las preferencias del usuario a lo largo del tiempo. A medida que el investigador interactúa con la herramienta y realiza selecciones o descartes de artículos, el algoritmo refina sus sugerencias, alineándolas cada vez más con las necesidades específicas del proyecto en curso. Esto da como resultado una experiencia de búsqueda altamente personalizada en la que el usuario puede confiar en que las referencias sugeridas son de relevancia directa para sus objetivos. Esta herramienta está disponible en: <https://www.researchrabbit.ai/>.

SciSpace es una herramienta de IA diseñada para ayudar a investigadores y estudiantes a comprender de forma rápida y eficiente artículos científicos. Frente al creciente volumen y complejidad de las publicaciones académicas, SciSpace destaca como una solución para decodificar información científica, haciendo la lectura e interpretación de la literatura académica más accesible.

La fortaleza principal de SciSpace radica en su capacidad para simplificar el contenido denso y técnico que se encuentra en los artículos de investigación, sin sacrificar la precisión o la profundidad de la información. La herramienta utiliza procesamiento del lenguaje natural (PNL) y algoritmos de aprendizaje automático para analizar textos académicos, identificando automáticamente

Inteligencia artificial...

conceptos clave, metodologías, resultados y conclusiones. Esto permite que los usuarios obtengan una visión panorámica del estudio en cuestión de minutos, algo que tradicionalmente llevaría horas de lectura y análisis detallado.

Además, SciSpace ofrece recursos adicionales que mejoran el proceso de aprendizaje e investigación. La herramienta genera resúmenes automáticos, en los que destacan las secciones más relevantes de los estudios y ofrece explicaciones simplificadas de términos técnicos y jerga científica. Esto es particularmente útil para estudiantes o investigadores que están explorando nuevos campos o disciplinas interdisciplinarias. Con ello, SciSpace contribuye a democratizar el acceso al conocimiento científico, asegurando que personas con diferentes niveles de experiencia puedan participar activamente en el proceso de investigación académica.

SciSpace también se integra con herramientas y plataformas utilizadas por la comunidad académica. Esto incluye la posibilidad de importar referencias directamente desde bases de datos académicas o sistemas de gestión de referencias, como Mendeley y Zotero, facilitando la organización y el acceso continuo a los artículos durante el desarrollo de un proyecto de investigación. Esto aumenta su utilidad como herramienta complementaria en los procesos de investigación. La herramienta está disponible en: https://typeset.io/?utm_source=futuro-dot-ooo&utm_medium=marketplace&utm_campaign=futuro-dot-ooo.

Los resultados del aprendizaje se midieron y se definieron a través de una serie de evaluaciones estructuradas que consideraron tanto la comprensión teórica como la aplicación práctica del contenido. Se realizaron seminarios donde los estudiantes colaboraron en la preparación y presentación de temas vinculados con la bibliografía de la disciplina. Estos seminarios sirvieron como plataforma para desarrollar habilidades de comunicación, trabajo en equipo y síntesis crítica, lo que permitió a los estudiantes demostrar no solo su conocimiento de los temas, sino también su capacidad para articularlos de manera clara y convincente ante sus compañeros.

Además de las presentaciones en equipo, se realizaron trabajos individuales que requirieron un análisis profundo de textos que formaban parte de la bibliografía del curso. Estas tareas brindaron a los estudiantes la oportunidad de explorar conceptos y teorías con mayor profundidad, desarrollando sus habilidades de investigación, escritura académica y argumentación. El énfasis en la producción individual permitió que cada estudiante demostrara su capacidad para aplicar métodos analíticos apropiados, además de fomentar la reflexión crítica sobre las lecturas y sus implicaciones en el contexto más amplio de la disciplina.

La participación individual en las discusiones de aula también fue un componente esencial de la evaluación. Estas discusiones fueron diseñadas para fomentar la participación activa de los estudiantes, propiciando un entorno de aprendizaje colaborativo donde se pudieran compartir y debatir diferentes perspectivas. La evaluación de la participación individual tuvo en cuenta no solo la frecuencia de las contribuciones, sino también la relevancia y profundidad de sus intervenciones, midiendo así su capacidad para pensar críticamente e interactuar de manera constructiva con sus compañeros y con los contenidos.

Finalmente, el proceso de evaluación culminó con la elaboración de un trabajo final en formato de capítulo, que será publicado en un libro electrónico actualmente en proceso de producción. Este capítulo requirió que los estudiantes integraran y aplicaran los conocimientos adquiridos a lo largo del curso y exigió un análisis original y fundamentado sobre un tema relevante para la disciplina. Esto brindó a los estudiantes una experiencia real de divulgación académica y de contribución al conocimiento disciplinar.

Este enfoque de evaluación multifacético —trabajos individuales y en equipo, participación activa y producción académica— fue fundamental para garantizar que los objetivos de aprendizaje de la disciplina se alcanzaran de manera integral y significativa. Los estudiantes no solo adquirieron los conocimientos teóricos necesarios, sino también las habilidades prácticas e intelectuales esenciales para sus futuras actividades académicas y profesionales.

CONSIDERACIONES FINALES

Las consideraciones, reflexiones y trabajos finales derivados de esta experiencia didáctico-pedagógica demuestran que la competencia en información y medios es un elemento indispensable para el uso ético y eficaz de la IA. La *competencia informativa* se refiere a la capacidad de localizar, evaluar y utilizar información de manera crítica y efectiva, mientras que la *competencia mediática* abarca la capacidad de interpretar, crear y evaluar contenidos mediáticos en diferentes contextos. Estos elementos son cruciales porque, aunque las tecnologías de IA son avanzadas, están diseñadas para ser herramientas al servicio de los humanos y, en consecuencia, dependen del juicio crítico, la ética, la creatividad y otros componentes humanos para desplegar todo su potencial y generar un impacto positivo.

A medida que avanza el desarrollo de la IA, se vuelve indispensable que quienes interactúan con estas tecnologías cuenten con una sólida alfabetización mediática e informacional. Los usuarios deben ser capaces de comprender e interpretar críticamente los resultados de la IA, asegurando que puedan explotar plenamente el potencial de la tecnología y darle el significado apropiado en el contexto digital contemporáneo. El dominio de estas áreas también es necesario para mitigar los riesgos asociados con su aplicación, como los sesgos y distorsiones derivadas de los algoritmos subyacentes.

Además, esta competencia refuerza la importancia de la participación humana en los momentos críticos de toma de decisiones, especialmente en los procesos impulsados por la IA. La intervención humana es esencial para mejorar la rendición de cuentas, introducir consideraciones éticas y fortalecer la confiabilidad general de los resultados producidos por los sistemas de IA, particularmente en el contexto de las investigaciones académicas y científicas.⁷ Reconocer

7 Wenceslao Arroyo Machado y Daniel Torres Salinas, "ChatGPT en bibliotecas", *Boletín de la Asociación Andaluza de Bibliotecarios* 39, núm. 127 (2024): 23.

que la IA, a pesar de su sofisticación, presenta limitaciones y es susceptible a distorsiones subraya la necesidad de mantener una vigilancia constante y un enfoque crítico por parte de sus usuarios.

Nuestra experiencia pedagógica buscó no solo capacitar a los estudiantes en el uso técnico de herramientas de IA, sino también promover la conciencia sobre las limitaciones y posibles sesgos inherentes a estos sistemas. Al equiparlos con las habilidades necesarias para identificar y comprender estos sesgos, estamos preparando una nueva generación de investigadores y profesionales capaces de utilizar la IA de una manera más ética, crítica y consciente. Esto es esencial para maximizar los beneficios de la tecnología, minimizar riesgos, garantizar la confiabilidad y robustecer la calidad de los procesos de toma de decisiones para que sean sólidos y justos.

Por lo tanto, al reflexionar sobre esta experiencia didáctico-pedagógica, concluimos que la competencia en información y medios no es solo un requisito técnico, sino una necesidad ética e intelectual en el desarrollo y aplicación de la IA. El futuro de la tecnología, particularmente en el entorno de la investigación, dependerá cada vez más de la capacidad de las personas para navegar críticamente por el vasto y complejo mundo de la información y los medios, utilizando la IA de manera responsable e informada para impulsar avances significativos en el conocimiento y en la sociedad.

A la luz de lo anterior, también subrayamos la importancia de las bibliotecas académicas y el papel integrador de los bibliotecarios. Consideramos que las actividades aquí descritas pueden motivar a estos profesionales a capacitarse mejor en el acceso y uso de ChatGPT y otras innovaciones tecnológicas para ofrecer orientación y mediación en los procesos de interrelación e interacción con las comunidades de usuarios —estudiantes y docentes— en la era digital. Por lo tanto, algunas recomendaciones son:

- Reconocer que el uso de la IA en el ámbito bibliotecario aún está en sus inicios y que es indispensable mirar hacia el futuro de una tecnología con consecuencias de largo alcance, incluido el incremento en el acceso al conocimiento.

Inteligencia artificial...

- Comprender que la IA puede transformar la forma en que las bibliotecas realizan su trabajo.
- Identificar las amplias posibilidades de interacción entre bibliotecas y tecnologías de IA.
- Mejorar los servicios bibliotecarios para los usuarios y contribuir al descubrimiento de conocimientos, en especial en la investigación académica.
- Asumir que se trata de una tecnología disruptiva, cuyos desafíos no son del todo nuevos para las bibliotecas.⁸
- Personalizar las herramientas según las necesidades específicas de cada biblioteca para maximizar su utilidad y eficiencia al adaptarse al contexto indicado y necesidades específicas.
- Recordar que no hay IA sin intervención humana: las decisiones éticas de accesibilidad y de fiabilidad dependen de las personas.
- Determinar consideraciones para un uso eficiente y responsable de ChatGPT y otras herramientas de IA en bibliotecas.
- Proponer un conjunto de aplicaciones de ChatGPT y otras herramientas para las tareas bibliotecarias en las que pueden ofrecer apoyo significativo.
- Presentar ejemplos concretos de implementación de IA para resolver desafíos específicos en los distintos tipos de bibliotecas.
- Explorar los beneficios y desafíos del uso de ChatGPT y otras herramientas en bibliotecas, incluyendo consideraciones éticas y de privacidad.

Nos esforzamos por aumentar la transparencia y la comprensión, y por fortalecer un entorno de estudio e investigación que dote a las personas de los conocimientos necesarios para navegar y evaluar críticamente la información y los servicios basados en IA. Ese es nuestro desafío.

8 Xiaotian Chen, “ChatGPT and Its Possible Impact on Library Reference Services”, *Internet Reference Services Quarterly* 27, no. 2 (2023): 123-124.

BIBLIOGRAFÍA

- Association of Research Libraries. “Research Libraries Guiding Principles for Artificial Intelligence”. Abril de 2024. <https://doi.org/10.29242/principles.ai2024>.
- Barton, David, y Carmen Lee. *Linguagem online: textos e práticas digitais*. São Paulo: Parábola Editorial, 2015.
- Chen, Xiaotian. “ChatGPT and Its Possible Impact on Library Reference Services”. *Internet Reference Services Quarterly* 27, núm. 2 (2023): 121-129. <https://doi.org/10.1080/10875301.2023.1234567>.
- Machado, Wenceslao Arroyo, y Daniel Torres Salinas. “ChatGPT en bibliotecas”. *Boletín de la Asociación Andaluza de Bibliotecarios* 39, núm. 127 (2024): 9-26.
- Rojo, Roxane. “Gêneros discursivos do Círculo de Bakhtin e multiletramentos”. *Escol@ Conectada: os multiletramentos e as TICs*. São Paulo: Parábola Editorial, 2013.
- Sá, Patricia, y Fátima Paixão. “Competências-chave para todos no séc. XXI: orientações emergentes do contexto europeu”. *Interações* 11, núm. 39 (2015): 243-254.
- Strey, Marlene Neves, y Renata Chabar Kapitanski. *Educação & internet: a era da informação e a vida cotidiana*. São Leopoldo: Editora Sinodal, 2011.
- Sabzalieva, Emma, y Arianna Valentini. 2023. *ChatGPT e inteligência artificial en la educación superior: Guía de inicio rápido*. París: UNESCO; Caracas: Instituto Internacional de la UNESCO para la Educación Superior en América Latina y el Caribe (IESALC). <https://unesdoc.unesco.org/ark:/48223/pf0000385146>.

**MARCOS Y ESTRATEGIAS
PARA UNA GOBERNANZA ÉTICA DE LA IA**

La experiencia de diseñar e impartir un curso/taller para profesores universitarios sobre inteligencia artificial y plagio académico

JESÚS CORTÉS-VERA

Departamento de Ciencias Sociales

MA. LOURDES TISCAREÑO

THELMA JOVITA GARCÍA

Departamento de Humanidades

Universidad Autónoma de Ciudad Juárez

INTRODUCCIÓN

El propósito de este capítulo es compartir la experiencia de diseño e implementación de un curso/taller dirigido a profesores de ciencias sociales sobre inteligencia artificial, y su relación con la integridad académica. La oportunidad de desarrollar este curso surgió a partir de la invitación del coordinador de un programa académico en el área de ciencias sociales. Los resultados de esta experiencia pueden ser útiles para coordinadores de programas académicos interesados en reconocer a los bibliotecarios como profesionales capacitados para impartir cursos sobre IA.

El nombre del curso fue “Honestidad académica e inteligencia artificial, oportunidades y retos” y se definió con la participación del coordinador del programa académico. De igual forma, el objetivo del curso/taller se redactó de la siguiente manera: “Que las/los docentes participantes identifiquen y reflexionen sobre las causas que favorecen las prácticas de deshonestidad académica, y que así mismo reconozcan las oportunidades y retos que la inteligencia artificial plantea”.

Datos generales del curso

El curso fue de actualización, similar a los que los docentes adscritos a esta coordinación suelen tomar al final del semestre, para los cuales seleccionan un tema que consideren pertinente, actual o de interés para la mayoría de los profesores. En esta ocasión, el tema fue la inteligencia artificial.

Se acordó impartir el curso con apoyo de la plataforma Teams, es decir, en modalidad en línea, con el fin de que más profesores tuvieran la posibilidad de inscribirse y participar. El curso/taller se impartió del 3 al 7 de junio de 2024 y consistió en cinco sesiones de dos horas cada una, distribuidas de lunes a viernes. Además, se incluyeron actividades extraescolares para completar un total de veinte horas, cumpliendo así con el estándar de la universidad para un curso con valor curricular.

En cuanto al número y perfil de los participantes, se inscribieron veinte profesores, aunque en la práctica cuatro de ellos no se conectaron. Resulta pertinente subrayar que los profesores participantes imparten clases en el programa de sociología y su formación se ubica en diversas disciplinas del área de Ciencias Sociales y Humanidades, tales como: Administración, Educación, Derecho, etc. En virtud de su formación, los profesores participantes muestran dinámicas particulares en eventos de capacitación: tienden a adoptar posturas más críticas, disponen de mayores recursos retóricos y suelen prolongar más el debate que los profesores de otras disciplinas. El curso fue tomado mayoritariamente por profesores de tiempo completo, aunque también participaron al menos un par de docentes por honorarios, esto es, profesores que prestan sus servicios con base en un contrato por tiempo determinado. En cuanto a los años de antigüedad en la universidad, hubo una variación considerable, lo cual implicó diferencias en la familiaridad con el uso de tecnologías de la información y la comunicación. Esta diversidad es sin duda un reto al que se enfrentarán quienes impartan cursos de este tipo: no existe homogeneidad ni en las competencias tecnológicas de los profesores ni en sus actitudes hacia estas herramientas.

Con respecto al perfil de los facilitadores, los tres son miembros de un cuerpo académico consolidado que lleva el nombre de Competencias Informacionales para la Vida Académica y la Ciudadanía Digital. Este cuerpo académico está integrado por cuatro profesores que han trabajado juntos por más de cinco años. Todos tienen formación como bibliotecarios y experiencia en la coordinación de bibliotecas. Actualmente se desempeñan como docentes investigadores de tiempo completo. Los cuatro tienen grado de doctorado y son miembros del Sistema Nacional de Investigadoras e Investigadores de la Secretaría de Ciencia, Humanidades, Tecnología e Innovación (SECIHTI).

Es posible que un factor relevante para que se concretara la invitación a los instructores haya sido que, en meses recientes, se difundió en la universidad que este cuerpo académico desarrolló un proyecto de investigación centrado en la utilidad y la importancia de los códigos de integridad académica en las instituciones de educación superior.

CONTENIDO DEL CURSO/TALLER

Los instructores dedicaron una cantidad considerable de tiempo a definir el contenido del curso/taller. Una preocupación central fue encontrar un equilibrio adecuado entre los temas de inteligencia artificial y de integridad académica. Con base en lecturas previas, se anticipó que el abordaje no sería exhaustivo, sino solo una aproximación a los temas. Después de tener una propuesta inicial se acordó el siguiente contenido, expuesto aquí de forma general:

Módulo I. Introducción al curso: un panorama general

- Presentación de los expositores.
- Presentación de los participantes.
- Presentación de propósitos y contenidos del taller.
- Advertencias y precisiones.
- El impacto de la IA en la educación.

Inteligencia artificial...

- Indicaciones para la elaboración del producto del taller: lineamientos para el uso de la IA generativa.

La primera sesión tuvo un carácter introductorio. Se presentó el plan de trabajo y se ofreció un panorama general del tema. Los instructores se presentaron y explicaron el plan de trabajo y sus expectativas, mientras que los profesores participantes también se presentaron y compartieron brevemente su experiencia y actitud respecto a la inteligencia artificial.

Módulo II. La inteligencia artificial y la integridad académica

- La importancia de evitar el plagio académico.
- Importancia de entregar trabajos originales.
- Chomsky y la IA generativa.
- Herramientas y claves para identificar trabajos elaborados con IA generativa.

En esta segunda sesión se profundizó en la relación entre la inteligencia artificial y la integridad académica. También se habló sobre las limitaciones y los alcances de la IA generativa. Adicionalmente, se enfatizó la importancia de promover prácticas de integridad académica y se abordaron, de manera introductoria, algunas recomendaciones para detectar textos generados con IA generativa.

Módulo III. Políticas institucionales con relación a la IA y la integridad académica

- Necesidad e importancia de políticas institucionales de IA e integridad académica.
- Ejemplos de políticas institucionales de IA.
- Propuesta de Código de Integridad académica para la UACJ.
- Ejemplos de criterios aplicados para y por profesores.
- Recomendaciones para incorporar la IA en el proceso educativo.
- Elaboración de criterios propios para una asignatura.

En esta tercera sesión se ahondó en la importancia de establecer políticas institucionales y personales que orienten la integridad académica y el uso responsable de la inteligencia artificial. Además, se compartió con los profesores participantes una propuesta de Código de Integridad Académica elaborada por el cuerpo académico de los instructores.

Módulo IV. Percepciones de los estudiantes

- Adaptando la escritura universitaria a la era de los modelos de lenguaje a gran escala como ChatGPT. Próximos pasos para educadores.
- Un ejemplo de guía para el uso de la IA generativa elaborada por un profesor.
- Percepciones de los estudiantes sobre la IA generativa.

En la cuarta sesión se abordaron las percepciones y las preocupaciones de los estudiantes en torno a la inteligencia artificial, a partir de un estudio realizado en Inglaterra. También se revisó una importante cantidad de recomendaciones para facilitar una transición ordenada hacia una nueva realidad académica en la que la escritura universitaria recurrirá cada vez más al apoyo de la IA generativa.

Módulo V. Instrucciones para elaborar trabajos semestrales: reflexiones finales

- Instrucciones y recomendaciones para que los estudiantes hagan un buen uso de la IA al elaborar ensayos o trabajos semestrales.
- ¿Qué opina la IA generativa sobre su uso honesto?
- Instrucciones para la elaboración de un ensayo o trabajo semestral.
- ¿Cómo mantenerse actualizado sobre temas de IA generativa?
- Plenaria de recapitulación y reflexiones finales.

Inteligencia artificial...

En esta última sesión se continuaron analizando los elementos que deben incluirse en los lineamientos para el uso de la IA generativa. También se consideró fundamental realizar un ejercicio de redacción de instrucciones para los estudiantes sobre cómo elaborar ensayos o reportes de investigación. Al final, se destinó un espacio para recapitular y conocer las impresiones de los profesores participantes.

Advertencias y precisiones previas

En el diseño del curso se estableció que en la primera sesión se mencionarían algunos aspectos que se describen a continuación, con el propósito de manejar el tiempo con mayor eficiencia y ubicarse rápidamente en el objetivo y contenido del curso.

- I. Se informó desde la primera sesión que los tres facilitadores se alternarían en la conducción de las sesiones.
- II. Asimismo, se señaló que tanto los instructores como los participantes adoptarían una actitud de aprendizaje respecto al tema.
- III. Puesto que existe una diversidad de formas de IA, se precisó que el enfoque del curso sería en torno a la inteligencia artificial generativa, por ser la que más se relaciona con la producción de textos, dibujos, fotografías, gráficas, etcétera.
- IV. Se anticipó que, debido a la amplitud del tema y a la duración del curso, la exposición de los contenidos (tanto de inteligencia artificial como de integridad académica) sería introductoria, debido a la imposibilidad de abordar los diferentes aspectos que conlleva cada uno.
- V. Del mismo modo, se aclaró desde el principio que no se profundizaría en aspectos técnicos del uso de aplicaciones, tanto porque estos resultaban ajenos para los instructores como porque ese tipo de capacitación sobrepasaba el objetivo del curso/taller. El enfoque estaría en definir políticas y

estrategias para aprovechar la IA generativa preservando la integridad académica.

- VI. Otro punto importante fue la necesidad de reconocer que la IA generativa es una tecnología a la que no se le pueden cerrar las puertas de las aulas. Los instructores anticiparon que esta afirmación podía generar resistencia entre los profesores. Por ello, se preparó una serie de argumentos para atender posibles objeciones. Afortunadamente, dicha resistencia no se presentó, los participantes parecieron aceptar que la IA es una realidad con la que es necesario convivir.
- VII. Se anunció que los participantes elaborarían un producto final a lo largo de la semana que se entregaría al final del curso. Los profesores trabajaron en una parte distinta de ese producto final cada día después de la sesión en Teams
- VIII. Finalmente, se enfatizó la importancia de la participación de cada uno de los asistentes.

La inevitabilidad de la IA generativa

Como se mencionó previamente, se prepararon algunos argumentos para sustentar la idea de que la IA generativa ya forma parte de nuestra realidad y que resistirse a ella carece de sentido práctico. Para ejemplificarlo, se les mostraron a los profesores diversas imágenes y mensajes recuperados de redes sociales donde se anuncian cursos de varias semanas dedicados a la inteligencia artificial; uno de ellos asegura que, en treinta días se puede guiar al usuario por las principales aplicaciones y tecnologías. Asimismo, se destacó que instituciones académicas internacionales de larga tradición, como la Universidad de Salamanca, están ofreciendo diplomados enfocados en el uso de la IA generativa en la escritura creativa. También se mencionaron aplicaciones que muchos profesores utilizan ya en su trabajo cotidiano y que funcionan gracias a la inteligencia artificial, por ejemplo: asistentes virtuales, motores de búsqueda con nuevas aplicaciones, traductores automáticos, correctores gramaticales, gestores de referencia, programas para

Inteligencia artificial...

elaborar resúmenes, herramientas para mejorar o elaborar presentaciones (como PowerPoint) así como programas que apoyan en la elaboración de exámenes o en la evaluación del aprendizaje. En síntesis, se trata de una serie de recursos que, en menor o mayor medida, ya tienen presencia en el ámbito académico.

Sin embargo, esta realidad no elimina las múltiples preocupaciones que surgieron durante el curso. La inquietud mencionada con mayor frecuencia —tanto por participantes como por instructores, y ampliamente discutida en la literatura internacional— fue el plagio académico. A esta se suma el riesgo de saturación de publicaciones en los repositorios institucionales y bases de datos con textos generados mediante IA que, al no aportar contenido nuevo o de calidad, dificultan la investigación seria. Otra preocupación importante gira en torno a las llamadas alucinaciones, es decir, la producción de información incorrecta o sin sentido, potencialmente riesgosa para quienes no detecten el error. También se habló del uso fraudulento de estas tecnologías como ocurre con los *deep fakes*, así como el temor, especialmente en ámbitos legales, de que documentos tradicionalmente válidos como evidencia —fotografías, videos, audios de llamadas telefónicas, entre otros— puedan perder su credibilidad al ser susceptibles de manipulación mediante IA.

Tal como se planeó, se dedicó un tiempo mayor a la preocupación que dio origen a este curso: el uso indebido de la IA generativa en la elaboración de trabajos académicos que los estudiantes deben realizar a lo largo del semestre. Resulta especialmente útil la reflexión de Walters, quien señala que:

[...] la inteligencia artificial plantea problemas especiales en situaciones donde la finalización de un trabajo escrito, por ejemplo, es valioso no como un fin en sí mismo, sino como un mecanismo para ayudar a los estudiantes a aprender a planificar, complementar y editar su trabajo escrito.¹

1 William Walters, “The effectiveness of software designed to detect ai-generated writing: a comparison of 16 ai text detectors”, *Open Information Science* 7, núm. 1 (2023): 1.

Es decir, hay estudiantes que asumen que si entregan trabajos bien redactados y bien sustentados con información, aún cuando ellos no los hayan elaborado, cumplen con los requisitos de sus tareas. Ante esta situación, los profesores deben explicar los requisitos del producto así como el proceso a seguir mediante una rubrica o lista de cotejo con la que los estudiantes puedan guiarse y demostrar su aprendizaje: a) identificar problemas y preguntas; b) diseñar estrategias de búsqueda; c) recuperar información pertinente; d) evaluar la información para seleccionar únicamente aquella con calidad académica; e) integrar esa información en un nuevo documento, respetando la propiedad intelectual, evitando el plagio y siguiendo normas de citación y referencia. La preocupación central es que algunos estudiantes consideren válido generar un texto con apoyo de la IA generativa y presentarlo al final del semestre como propio. Esto obliga a los profesores a replantear tanto el seguimiento dado a la realización de estos trabajos como las instrucciones que ofrecen a los estudiantes. Incluso esta situación podría llevar a reconsiderar el peso que este tipo de trabajos tiene en la calificación final.

Otros productos revisados de la IA generativa

Es oportuno mencionar que, a lo largo del curso, se fueron intercalando revisiones breves de diversos productos de IA generativa para que los profesores tuvieran un panorama general del crecimiento de este mercado y de la diversidad de beneficios que ofrecen estas herramientas. Se revisaron rápidamente algunas de las aplicaciones promovidas en las redes sociales, por ejemplo la versión más reciente de ChatGTP, así como Consensus, Claude, Copilot y otra llamada Aithor, que anuncia de manera muy abierta y sin reparo alguno que puede generar trabajos en distintos formatos según lo solicite el usuario, además de ofrecerla opción de “humanizar” los textos para no ser detectados como productos de la inteligencia artificial.

EL PRODUCTO CENTRAL DEL CURSO/TALLER

Se propuso a los profesores que como producto central de este curso/taller elaboraran una guía con lineamientos sobre el uso permitido y no permitido de aplicaciones de IA generativa para la elaboración de trabajos escritos en su materia. Se decidió proponer esta guía porque durante la preparación del curso se observó la importancia de establecer directrices y políticas por escrito, que orienten el comportamiento de los miembros de la comunidad académica; sin embargo, en México pocas instituciones cuentan con políticas claras sobre la integridad académica y menos aún sobre el uso de la inteligencia artificial. Ante la ausencia de una política institucional, se consideró fundamental que los profesores desarrollaran sus propias políticas y las dieran a conocer a sus estudiantes.

Se sugirió pensar en una propuesta breve, con pocas secciones y redactada en un lenguaje accesible para los estudiantes para que el docente la presentara al inicio del semestre. El documento, amable y firme, debería promover la transparencia e incluir posibles sanciones ante el incumplimiento de las políticas. Se recomendó estructurar la guía en cuatro apartados:

- I. Una introducción donde se explicara el propósito de la guía, así como el objetivo y la relevancia de la integridad académica y el posicionamiento del profesor con respecto a la IA generativa.
- II. Un apartado dedicado a los criterios, es decir, las acciones relacionadas con la IA generativa que el profesor considera aceptables para los estudiantes, así como aquellas que no lo serían, según la situación, o el tipo de trabajo solicitado.
- III. Otro apartado para describir el acompañamiento que el profesor brindará durante el proceso de elaboración de los trabajos y las posibles sanciones en caso de que se incurriera en alguna falta.
- IV. Una conclusión y reflexión final que invitara a observar y respetar lo establecido en la guía.

Los profesores trabajaron sobre esta estructura en las actividades fuera del aula. El martes, por ejemplo, redactaron un texto de no más de 250 palabras sobre su posición personal respecto a la IA generativa; el miércoles completaron la introducción con otros elementos que consideraron relevantes para incluir en la guía que estaban preparando; el jueves escribieron sus criterios como docentes sobre el uso o no uso de la IA generativa, es decir, qué acciones consideran aceptables en sus clases y cuáles no—; y el viernes hablaron sobre el rol del profesor en el proceso de acompañamiento y posibles sanciones en caso de que los estudiantes incurran en alguna falta, además de una reflexión final. Con todos estos elementos se esperaba un primer borrador de la guía, de modo que los profesores pudieran dedicar el fin de semana a preparar una versión más elaborada y reflexiva, lista para entregarse el lunes siguiente. Es importante señalar que los instructores fueron flexibles en cuanto al formato y la estructura de la guía, reconociendo tanto el estilo particular de cada profesor como la diversidad de perspectivas; lo más importante es que la guía resultara útil para ellos y sus estudiantes.

Durante la sesión del viernes también se propuso un trabajo complementario que los profesores pudieron hacer en ese mismo espacio: redactar las instrucciones para un trabajo escrito, es decir, un texto breve en donde se explique a los estudiantes de qué manera deberán realizar trabajos escritos. Se enfatizó la importancia de que estas instrucciones fueran breves, sobre todo en los cursos en línea, dado que los estudiantes no siempre leen las instrucciones completas y rara vez preguntan cuando algo que no les queda claro. El objetivo era producir un texto conciso, con instrucciones claras sobre el desarrollo de los trabajos y su evaluación, incluyendo la mención del uso permitido y honesto de la inteligencia artificial generativa.

DOCUMENTOS RELEVANTES REVISADOS EN EL CURSO

Para preparar este curso, los tres instructores realizaron una lectura exhaustiva, a partir de la definición preliminar de los temas a abordar. Es importante señalar que se encontró muy poca literatura publicada con referencia específica a México, por lo que fue necesario recurrir principalmente a la literatura de países anglosajones. Enseguida se mencionan algunos documentos relevantes:

En la primera sesión, con el propósito de ofrecer una visión general del impacto actual de la inteligencia artificial en la educación a nivel mundial, se revisó un estudio derivado de una encuesta internacional realizada por Ipsos,² empresa ampliamente reconocida en investigación de mercados.

En la segunda sesión se revisó y comentó una entrevista realizada a Noam Chomsky,³ filósofo y lingüista que formula críticas agudas sobre los alcances y la utilidad de la inteligencia artificial hacia 2023, señalando que la inteligencia artificial aún está muy por debajo de las capacidades intelectuales humanas. Para profundizar en el tema del plagio académico, se retomaron los principios fundamentales del International Center of Academic Integrity⁴ y algunos datos y propuestas aportadas por Carnero *et al.*⁵

En la tercera sesión se revisaron diversas políticas institucionales y guías sobre el uso de IA generativa y la promoción de la integridad académica. Entre ellas destacan la política de la Uni-

2 IPSOS, *Opinión pública sobre educación*, septiembre de 2023.

3 Noam Chomsky, "Noam Chomsky habla sobre ChatGPT. Para qué sirve y por qué no es capaz de replicar el pensamiento humano. Entrevista", *Sin permiso*, 7 de mayo de 2023.

4 *The fundamental values of academic integrity*, Delaware: International Center of Academic Integrity, 2018, 4-10.

5 Andres Carnero *et al.*, "Plagiarism, cheating and research integrity: Case studies from a masters program in Peru", *Science and Engineering Ethics* 23, núm. 4 (2017): 1183-1197.

versidad de Delaware⁶ en Estados Unidos, la propuesta de Bretag *et al.*⁷ sobre la integridad académica en Australia, un documento publicado por el Gobierno de Chile⁸ y una guía elaborada por la Universidad Autónoma de Madrid en España.⁹ También se analizó brevemente un documento promovido por la Conferencia de Rectores de las Universidades Españolas (CRUE),¹⁰ una muestra del creciente interés que este país tiene por la IA. De México se revisaron materiales del Tecnológico de Monterrey,¹¹ de la Universidad de Guadalajara¹² y, especialmente, una publicación de la Universidad Nacional Autónoma de México¹³ que ofrece recomendaciones para el uso de la IA generativa en la docencia. Este documento resulta especialmente relevante porque evidencia el trabajo anticipado de la UNAM, que mediante comités de trabajo multidiscipli-

-
- 6 University of Delaware, Center for Teaching & Assessment of Learning, “Considerations for using and addressing advanced automated tools in coursework and assignments”, 2024, consultado el 23 de septiembre de 2024, <https://ctal.udel.edu/advanced-automated-tools/#syllabus-language>.
 - 7 Tracey Bretag *et al.*, “Core elements of exemplary academic integrity policy in Australian higher education”, *International Journal for Educational Integrity* 7, núm. 2 (2011): 6-7.
 - 8 Gobierno de Chile, Ministerio de Educación, *Guía para Docentes: Cómo usar ChatGPT para potenciar el aprendizaje activo*, mayo de 2023, 2-14.
 - 9 Universidad Autónoma de Madrid, *Guía básica sobre el uso de la inteligencia artificial para docentes y estudiantes*, noviembre de 2023, 1-4.
 - 10 Francisco Cruz Argudo *et al.*, *La inteligencia artificial generativa en la docencia universitaria. Oportunidades, desafíos y recomendaciones*, Conferencia de Rectores de las Universidades Españolas, marzo de 2024, https://www.crue.org/wp-content/uploads/2024/03/Crue-Digitalizacion_IA-Generativa.pdf, 4-21.
 - 11 Tecnológico de Monterrey, “Conoce los 11 lineamientos Tec para el uso de inteligencia artificial”, *Conecta*, consultado el 23 de septiembre de 2024, 8-17.
 - 12 Universidad de Guadalajara, *Orientaciones y definiciones sobre el uso de la inteligencia artificial generativa en los procesos académicos: Guía práctica*, octubre de 2023, 6-40.
 - 13 Universidad Nacional Autónoma de México, *Recomendaciones para el uso de la inteligencia artificial generativa en la docencia*, octubre de 2023, 6-25.

narios ha avanzado en la identificación de los principales retos y oportunidades asociados al uso de la IA generativa, lo que puede favorecer su uso y aprovechamiento.¹⁴

La cuarta sesión se dedicó a conocer las inquietudes y las perspectivas de los estudiantes universitarios, a partir de un documento elaborado en el Reino Unido por el Joint Information Systems Committee.¹⁵ El texto detalla los temores y las expectativas de los estudiantes respecto al impacto futuro de la IA en su desempeño profesional, subrayando la necesidad de que las universidades los preparen adecuadamente ante este reto. También se revisó una propuesta muy completa de Mills y Goodlad¹⁶, quienes presentan diversas acciones para facilitar la transición de la escritura académica hacia la era de los modelos de lenguaje a gran escala, es decir, hacia la IA generativa. Asimismo, se analizaron propuestas generales sobre la integración de la IA en la educación expuestas por Bustos¹⁷ y un documento de lineamientos personales de Gutiérrez¹⁸ sobre el uso de la IA generativa en la escritura académica.

¿Qué opina la inteligencia artificial generativa sobre su uso honesto?

Un elemento sobresaliente de la última sesión, que surgió al analizar qué otros contenidos se podrían incluir en la guía para alumnos sobre el uso de la IA fue revisar lo que la propia inteligencia artificial recomienda respecto a su uso ético. Se consultó a la plataforma Claude con la siguiente pregunta: “¿qué instrucciones y re-

14 Universidad Nacional Autónoma de México. *Recomendaciones...*, 6-25.

15 Joint Information Systems Committee, *Student perception of generative AI*, mayo de 2024, 1-11.

16 Anna Mills y Lauren Goodlad, “Adaptando la escritura universitaria a la era de los modelos de lenguaje a gran escala como ChatGPT: Próximos pasos para educadores”, *Critical AI*, 2023.

17 Lukas Bustos, “¿Cómo integrar la IA en la educación?”, *edu21* (blog), 21 de marzo de 2024.

18 Juan David Gutiérrez, “Política de uso de inteligencia artificial en la asignatura”, *Foro Administración, Gestión y Política Pública*, febrero de 2023, 1-4.

comendaciones deberán de darse a los estudiantes universitarios para que hagan un uso honesto de la inteligencia artificial cuando escriban un ensayo o un trabajo el fin de cursos?”. La pregunta y las respuestas de Claude se tradujeron al inglés. A continuación, se presenta una síntesis en forma de lista de las recomendaciones. Estas sugieren que los estudiantes deben:

- I. Comprender con claridad las políticas de su institución sobre el uso de la inteligencia artificial.
- II. Utilizar la IA como herramienta de apoyo, no para generar trabajos enteros.
- III. Citar correctamente cualquier texto o idea generada mediante IA.
- IV. Declarar explícitamente el uso de la inteligencia artificial en el trabajo.
- V. Mantener la voz auténtica del autor, aspecto que con frecuencia se enfatiza a los estudiantes; el lector de sus trabajos escritos debe reconocer la voz del autor a lo largo del texto.
- VI. Verificar la información y detectar posibles sesgos, reconociendo que estas aplicaciones pueden producir datos inexactos y tendenciosos o incluso alucinaciones.
- VII. Utilizar los resultados de la inteligencia artificial como punto de partida y luego ampliarlos con trabajo propio.
- VIII. Limitar el uso de la inteligencia artificial a secciones poco sustantivas del trabajo, es decir, evitar la utilización de la IA en la parte central de las investigaciones.

Las recomendaciones subrayan con insistencia la necesidad de declarar el uso de la IA generativa, mediante una nota que indique en qué proceso se hizo uso de este recurso. También sugieren que no es indispensable incluir una referencia formal al final, pues la información obtenida por este medio podría citarse como si se derivara de una comunicación personal.

HERRAMIENTAS Y SUGERENCIAS PARA IDENTIFICAR TRABAJOS ELABORADOS CON IA GENERATIVA

Entre los temas que consideramos indispensables en el curso estuvo el de herramientas y claves para detectar trabajos creados con IA generativa.

Se comentaron con los profesores algunas recomendaciones básicas para saber identificar rasgos característicos de textos generados por IA. Estas se basaron principalmente en Machin-Mastromatteo¹⁹ donde se indica que, generalmente, estos textos presentan una estructura típica compuesta por tres elementos: un párrafo introductorio; después, una serie de subtemas identificados como relevantes, con frecuencia organizados en listas numeradas o con viñetas; y, finalmente, un párrafo que resumen lo expuesto y/o sugiere acciones posteriores. Además, estos textos suelen ser muy superficiales, poco exhaustivos, redundantes y enumerativos; carecen de análisis crítico y se caracterizan por un lenguaje plano, formal y bastante adjetivado.

Este es un tema que requiere mayor análisis y sobre el que circula información contradictoria. Mientras algunos señalan que los desarrolladores de IA generativa están afinando procesos de “humanización” que dificultan la detección de textos producidos por estas aplicaciones, otros apuntan a la proliferación de programas —gratuitos y de suscripción— que aseguran ser eficaces para identificarlos.

Para profundizar en este tema, se solicitó el apoyo de una persona experta en el uso de Turnitin, software contratado desde hace varios años por la universidad para detectar similitudes y prevenir el plagio académico, y que ha incorporado nuevas herramientas para identificar posibles textos generados con IA generativa.

La participación de la persona experta en Turnitin tomó más tiempo del previsto debido a la cantidad de preguntas por parte de los profesores. Sin embargo, la mayoría de las dudas se rela-

19 Juan Daniel Machin-Mastromatteo, “Inteligencia Artificial en Revistas Científicas (con Juantífico)”, *Infotecarios Podcast*, episodio 190, 18 de mayo de 2024, episodio 190.

cionó con funciones básicas del software, como detector de similitudes, lo que evidenció que varios profesores estaban poco familiarizados con esta herramienta. En consecuencia, quedó poco tiempo para abordar el tema específico de la detección de textos generados con IA generativa.

La disponibilidad y uso de herramientas para detectar el uso indebido de la IA generativa ofrecen material suficiente para cursos y talleres. Desde nuestra perspectiva, seguirán surgiendo desarrollos tecnológicos que faciliten el uso indebido de la IA generativa, pero también avances que busquen preservar y fortalecer la honestidad académica. Al margen de estos desarrollos, los profesores deberán mantener una comunicación frecuente con sus estudiantes, conocer sus competencias, actitudes y preocupaciones, y trabajar en estrecha colaboración con ellos.

LECCIONES APRENDIDAS

A continuación se mencionan algunos de los aprendizajes más relevantes adquiridos en este curso/taller:

- I. El trabajo colaborativo entre los tres instructores fue un acierto. Habría sido prácticamente imposible que una sola persona se encargara de localizar y revisar tantas lecturas, así como preparar materiales y presentaciones. La participación de más de un facilitador añadió dinamismo a las sesiones. Además, se pudo tener un mayor control de las intervenciones de los asistentes.
- II. Otro aprendizaje fue constatar la enorme velocidad con la que evoluciona este campo. Inicialmente se pensó que la impartición de este curso dejaría una estructura relativamente estable que serviría como base para una versión futura. Sin embargo, ahora se reconoce que será necesario revisar y actualizar una cantidad considerable de materiales antes de ofrecer un curso similar.

Inteligencia artificial...

- III. También quedó claro que el uso de la IA generativa ofrece grandes beneficios y que existe un margen amplio para el aprovechamiento de esta tecnología con propósitos honestos.
- IV. Asimismo, tanto instructores como participantes coincidieron en que los estudiantes están preocupados por el impacto de la inteligencia artificial en su formación y su futuro profesional, especialmente si la universidad no les brinda el apoyo necesario y los profesores no comparten esa preocupación ni se preparan para guiarlos.
- V. Otro gran aprendizaje fue reafirmar la importancia de que las universidades cuenten con políticas claras y por escrito sobre el uso y aprovechamiento de la inteligencia artificial. Dichas políticas deben funcionar como un marco de referencia tanto para estudiantes como para docentes.

Áreas de oportunidad para un siguiente curso

Existen diversas oportunidades para mejorar una futura edición de este curso. Primero, es necesario que los instructores hagan un uso más eficiente y dinámico de la plataforma Teams aprendiendo a manejar las herramientas que ofrece. Asimismo, sería útil contar con más información sobre el perfil de los profesores participantes para tener una idea más clara de sus competencias, habilidades y actitudes con respecto a la inteligencia artificial. Otro aspecto a mejorar es el equilibrio entre contenido e interacción. En esta ocasión se revisaron numerosos materiales, lo que redujo la posibilidad de generar un curso más interactivo. En una versión futura sería deseable disminuir la densidad de los contenidos y aumentar las oportunidades de interacción. Del mismo modo, es importante diseñar mecanismos que fomenten la participación de todos los docentes inscritos y no únicamente de un grupo reducido que pide hacer uso de la voz.

Habría sido útil disponer de una guía publicada por una institución de prestigio internacional —identificada posteriormente²⁰— para emplearla como referencia.

20 Unesco, *Guía para el uso de la IA...*, 2-39.

En cuanto al trabajo final que elaboraron los profesores participantes, es esencial contar con una rúbrica que sirva de guía tanto para la elaboración de dicho trabajo como para su evaluación al final. Cabe señalar que, para este curso/taller, el coordinador del programa académico no solicitó calificaciones, por lo que solamente se hizo un reporte de asistencias y de trabajos entregados.

Cuatro profesores inscritos no participaron en el curso. Al finalizar, catorce asistentes entregaron una guía de extensión y estilo diversos. Como se mencionó previamente, esto ya se esperaba dado que cada docente redactaría su documento según su estilo personal.

En términos generales, también hizo falta diseñar un mecanismo para evaluar el curso/taller en su totalidad.

CONCLUSIONES

Como reflexión final, esta experiencia permitió confirmar que el uso y aprovechamiento de la IA y la integridad académica son temas vinculados con los valores. Es indispensable fortalecer la ética en los procesos de enseñanza-aprendizaje y de investigación.

Se requiere de un trabajo sistemático para fomentar una cultura de integridad académica y al mismo tiempo aprovechar la inteligencia artificial. Esto implica un esfuerzo colectivo de toda la comunidad universitaria, especialmente profesores, estudiantes y bibliotecarios.

La IA generativa llegó para quedarse, y es necesario obtener de ella el mayor beneficio educativo posible. Sin embargo, también es imprescindible regular su uso en ámbitos legales, sociales, éticos, de privacidad y seguridad, particularmente en el ámbito educativo, donde el problema del plagio y la deshonestidad académica se intensificará si no existen procesos formativos adecuados. Es fundamental que el estudiante adquiera habilidades para utilizar y comunicar la información con ética; por ello, es urgente promover una cultura de integridad académica en toda la comunidad universitaria.

Inteligencia artificial...

Se reitera la necesidad de contar con políticas e instrumentos que orienten estos procesos. Las comunidades académicas deben impulsar la creación de políticas claras por escrito, accesibles y construidas de manera participativa. Asimismo, es fundamental contar con herramientas que apoyen la enseñanza y el uso apropiado de la IA generativa.

Hay mucho trabajo por hacer, y los bibliotecarios académicos pueden, si así lo deciden, desempeñar un papel protagónico en estos procesos. Para ello será necesario prepararse con el fin de conocer y dominar ciertos temas. Una de las formas en que pueden participar es, precisamente, ofreciendo cursos y talleres como el que se describe en este trabajo.

BIBLIOGRAFÍA

- Bretag, Tracey *et al.* “Core elements of exemplary academic integrity policy in Australian higher education.” *International Journal for Educational Integrity* 7, núm. 2 (2011): 6-7. <https://ojs.unisa.edu.au/index.php/IJEI/article/view/759/574>.
- Bustos, Lukas. “¿Cómo integrar la IA en la educación?” *edu21* (blog), 21 de marzo de 2024. <https://edu21.cl/blog/estrategias-de-ensenanza/como-integrar-la-ia-en-la-educacion/>.
- Carnero, Andres *et al.* “Plagiarism, cheating and research integrity: Case studies from a masters program in Peru”. *Science and Engineering Ethics* 23, núm. 4 (2017): 1183–1197. <https://pubmed.ncbi.nlm.nih.gov/27848191/>.
- Chomsky, Noam. “Noam Chomsky habla sobre ChatGPT. Para qué sirve y por qué no es capaz de replicar el pensamiento humano. Entrevista”. *Sin permiso*, 7 de mayo de 2023. Consultado el 28 de septiembre de 2024. <https://www.sinpermiso.info/textos/noam-chomsky-habla-sobre-chatgpt-para-que-sirve-y-porque-no-es-capaz-de-replicar-el-pensamiento>.

- Cruz Argudo, Francisco *et al.* *La inteligencia artificial generativa en la docencia universitaria. Oportunidades, desafíos y recomendaciones.* Conferencia de Rectores de las Universidades Españolas, marzo de 2024. https://www.crue.org/wp-content/uploads/2024/03/Crue-Digitalizacion_IA-Generativa.pdf.
- Gobierno de Chile. Ministerio de Educación. *Guía para docentes: cómo usar ChatGPT para potenciar el aprendizaje activo*, mayo de 2023. <https://ciudadaniadigital.mineduc.cl/wp-content/uploads/2023/05/Guia-para-Docentes-Como-usar-ChatGPT-Mineduc.pdf>.
- Gutiérrez, Juan David. “Política de uso de inteligencia artificial en la asignatura”. *Foro Administración, Gestión y Política Pública*, febrero de 2023. <https://forogpp.com/2023/02/18/politica-de-uso-de-inteligencia-artificial-en-la-asignatura-v-4-1>.
- IPSOS. *Opinión pública sobre educación*, septiembre de 2023. <https://www.ipsos.com/sites/default/files/ct/news/documents/2023-09/Ipsos%20Monitor%20Global%20de%20Educación.pdf>.
- Joint Information Systems Committee. *Student perception of generative AI*, mayo de 2024. <https://repository.jisc.ac.uk/9571/1/student-perceptions-of-generative-ai-report.pdf>.
- Machin-Mastromatteo, Juan Daniel. “Inteligencia Artificial en Revistas Científicas (con Juantifico)”. *Infotecarios Podcast*, episodio 190, 18 de mayo de 2024. <https://www.youtube.com/watch?v=nR5nxn3YiDg&t=1988s>.
- Mills, Anna, y Lauren Goodlad. “Adaptando la escritura universitaria a la era de los modelos de lenguaje a gran escala como ChatGPT: Próximos pasos para educadores”. *Critical AI*, 2023. Consultado el 28 de septiembre de 2024. <https://forogpp.com/2023/01/24/adaptando-la-escritura-universitaria-a-la-era-de-los-grandes-modelos-de-lenguaje-como-chatgpt-proximos-pasos-para-educadores/>.
- Tecnológico de Monterrey. “Conoce los 11 lineamientos Tec para el uso de inteligencia artificial”. *Conecta*. Consultado el 23 de septiembre de 2024. <https://conecta.tec.mx/es/noticias/nacional/educacion/conoce-los-11-lineamientos-tec-para-el-uso-de-inteligencia-artificial>

Inteligencia artificial...

- The fundamental values of academic integrity*. Delaware: International Center of Academic Integrity, 2018.
- UNESCO. *Guía para el uso de la IA generativa en educación e investigación*. París: UNESCO, 2024.
- Universidad Autónoma de Madrid. *Guía básica sobre el uso de la inteligencia artificial para docentes y estudiantes*, noviembre de 2023. <https://www.uam.es/uam/media/doc/1606941290988/guia-visual-iagen.pdf>.
- Universidad de Guadalajara. *Orientaciones y definiciones sobre el uso de la inteligencia artificial generativa en los procesos académicos: Guía práctica*, octubre de 2023. https://www.udgvirtual.udg.mx/sites/default/files/adjuntos/guia_ia_udg.pdf.
- Universidad Nacional Autónoma de México. *Recomendaciones para el uso de la inteligencia artificial generativa en la docencia*, octubre de 2023. https://iagenedu.unam.mx/docs/recomendaciones_uso_iagen_docencia_unam_2023.pdf.
- University of Delaware. Center for Teaching & Assessment of Learning. “Considerations for using and addressing advanced automated tools in coursework and assignments”, 2024. Consultado el 23 de septiembre de 2025. <https://ctal.udel.edu/advanced-automated-tools/>
- Walters, William. “The effectiveness of software designed to detect ai-generated writing: a comparison of 16 ai text detectors”. *Open Information Science* 7, núm. 1 (2023): 1-24. <https://doi.org/10.1515/opis-2022-0158>.

Integridad de la información: Los principios de la ONU y el ecosistema digital

JONATHAN HERNÁNDEZ PÉREZ

*Instituto de Investigaciones Bibliotecológicas y de la Información,
Universidad Nacional Autónoma de México*

INTRODUCCIÓN

Nuestro ecosistema informativo atraviesa una profunda crisis. Desde la era de la posverdad en 2016, el panorama no ha dejado de ser complejo, se ha experimentado un aumento alarmante en la polarización, la desconfianza y la falsedad en la información alrededor del mundo, con el persistente desafío de las tecnologías emergentes usadas para amplificar estas problemáticas.

En este contexto, la desinformación, cuya acción principal es la difusión deliberada de información falsa o engañosa, ha sido reconocida como un problema de alcance mundial, convirtiéndose en objeto de debate al más alto nivel y en uno de los grandes desafíos de nuestro tiempo. El fenómeno de la desinformación no es nuevo, su impacto se ha amplificado por la creciente *plataformización* de la sociedad. La propia arquitectura de las plataformas digitales, al priorizar la captura de la atención sobre la verificación de información, fomenta un entorno que no solo dificulta el acceso a fuentes confiables, sino que activamente erosiona la confianza institucional y exacerba la polarización.

Inteligencia artificial...

Esta crisis ha hecho evidente la necesidad de herramientas para mitigar los efectos negativos de esta *plataformización* en la sociedad, entre ellas la formulación de marcos conceptuales para abordar el problema. La integridad de la información, un concepto en evolución que ha transitado desde la seguridad técnica hacia una perspectiva social y de derechos humanos, se orienta hacia evaluar no solo el contenido, sino también todo el ecosistema por el que fluye la información para garantizar un espacio informativo que sea plural, seguro y con un especial enfoque en los derechos humanos.

Como respuesta a esta crisis informativa, la Organización de las Naciones Unidas (ONU) ha propuesto una ambiciosa arquitectura normativa: los Principios Globales para la Integridad de la Información. Teniendo como base los derechos humanos, estos principios buscan contrarrestar las mayores amenazas del entorno digital: la desinformación, la información errónea y el discurso de odio.

Sin embargo, a pesar de ser un avance importante, la viabilidad de este marco enfrenta diferentes desafíos. La velocidad, escala y naturaleza de la desinformación, especialmente aquella generada por inteligencia artificial (IA), representan un desafío para la implementación y funcionamiento de cualquier propuesta regulatoria; asimismo, la propia evolución de las plataformas y la falta constante de transparencia en sus procesos hacen aún más complejo el panorama.

El presente capítulo analizará la estructura y los objetivos de los Principios Globales para la Integridad de la Información de la ONU, evaluando su solidez y pertinencia frente a los diversos desafíos del entorno informativo contemporáneo.

INTEGRIDAD DE LA INFORMACIÓN

El concepto de integridad de la información tiene sus raíces en el campo de la seguridad de la información, el cual se ha desarrollado a partir de la informática, la ciencia de la información, los

estudios de ciberseguridad y áreas de investigación afines.¹ En su formulación original, la seguridad de la información se sostiene sobre tres pilares fundamentales, conocidos como la “tríada CIA”, por sus siglas en inglés: Confidencialidad, Integridad y Disponibilidad (*Confidentiality, Integrity and Availability*).²

Al inicio, la preocupación era principalmente técnica, ya que se centraba en preservar el contenido y la forma original de la información sensible frente a modificaciones no autorizadas, un interés especialmente relevante en contextos militares, de seguridad nacional y corporativos. Esta visión se formalizó en estándares internacionales como la norma ISO/IEC 27 000, que define la seguridad de la información como la preservación de la confidencialidad, integridad y disponibilidad. Dentro de este marco, la integridad se entiende de manera particular como la propiedad que salvaguarda la exactitud y completitud de los activos de información.³

La irrupción de la era digital y la consolidación de internet como espacio público central impulsaron una nueva y acelerada evolución del concepto. Este cambio se vio exacerbado por la creciente dependencia de las plataformas digitales y la confluencia de fenómenos sociopolíticos vinculados con la verdad y falsedad de la información, especialmente a partir de la llamada “era de la posverdad”. En esta nueva configuración del ecosistema informativo, la integridad de la información comenzó a definirse haciendo referencia a los marcos y criterios necesarios para evaluar no solo el contenido, sino el ecosistema completo por el que fluyen inmensas cantidades de información de calidades dispares. En la actualidad, su aplicación se ha extendido a distintos ámbitos de

-
- 1 International Panel on the Information Environment, “Information Integrity about Climate Science: A Systematic Review”, editado por Ece Elbeyi *et al.*, Zurich, Switzerland: IPEI, junio de 2025, <https://doi.org/10.61452/btzip3426>.
 - 2 Kelsey Harley y Rodney Cooper, “Information Integrity: Are We There Yet?”, *ACM Computing Surveys* 54, núm. 2 (2022): 1-35, <https://doi.org/10.1145/3436817>.
 - 3 International Panel on the Information Environment, “*Information Integrity about Climate Science...*”.

la sociedad, desde la banca y la contabilidad hasta los expedientes clínicos electrónicos, los sistemas de votación y los vehículos autónomos.⁴

En este punto, conviene advertir que un análisis exhaustivo del término excede los objetivos de este capítulo. La adopción del concepto “integridad de la información” fuera del ámbito técnico ha generado críticas, especialmente en lo que concierne a la posible instrumentalización de regulaciones vagas que podrían derivar en censura o restricciones a la libertad de expresión. Si bien, este concepto está siendo reconocido por la academia, adoptado por gobiernos y esbozado en marcos internacionales, aún dista de tener una definición universal, y existen importantes debates sobre su alcance y los desafíos que enfrenta en el complejo panorama digital actual.⁵

De tal manera que el enfoque de la integridad de la información que se abordará en esta contribución estará orientado a la perspectiva desarrollada por la ONU en sus Principios Globales para la Integridad de la Información, cuyo enfoque implica un “espacio de información plural que abogue por los derechos humanos, las sociedades pacíficas y un futuro sostenible” y cuyas amenazas son la desinformación, la información errónea y el discurso de odio.⁶

Tomando como base la desinformación como la principal amenaza para la integridad de la información, durante los últimos años se han desarrollado numerosas investigaciones sobre este fenómeno desde una amplia variedad de nichos disciplinares. A partir de 2016, emergió una nueva narrativa que popularizó y consolidó términos como “posverdad”, “*fake news*/noticias falsas”, “hechos alternativos”, términos ampliamente difundidos y analizados que se

4 Ali Shiri, “Information Integrity, Academic Integrity, and Generative AI”, *Information Matters* 3, núm. 10 (2023), <https://doi.org/10.2139/ssrn.4601781>.

5 Luca Belli *et al.*, “Platform responsibilities for information integrity”, Dynamic Coalition on Platform Responsibilities at the Internet Governance Forum (IGF), 2024.

6 Organización de las Naciones Unidas, “Principios Globales de las Naciones Unidas para la Integridad de la Información”, 2024, https://www.un.org/sites/un2.un.org/files/principios_globales_onu_integridad_informacion.pdf.

convirtieron en el epicentro de un debate global sobre la verdad y falsedad en los ecosistemas de información digitales. Más tarde, la crisis de la pandemia de covid-19 introdujo nuevos términos al debate público y potenció la formulación de políticas públicas orientadas a mitigar los efectos nocivos de la desinformación.

La notable complejidad del fenómeno de la desinformación ha sido abordada desde una multiplicidad de perspectivas, dando lugar a diversas líneas de investigación con destacables enfoques interdisciplinarios. Esta amplia variedad de enfoques va desde el campo técnico y computacional hasta el ámbito jurídico y de políticas públicas. Estos últimos campos abordan el problema desde una perspectiva estructural y regulatoria. Conviene mencionar también los enfoques que se formulan desde las humanidades y las ciencias sociales con disciplinas como la psicología, la sociología, la bibliotecología y los estudios de la información, en donde se analiza el fenómeno desde el factor humano, los sesgos cognitivos así como el diseño de estrategias para el desarrollo de habilidades informativas y de pensamiento crítico en la ciudadanía.

En los últimos años, el concepto de integridad de la información se ha articulado desde un marco más amplio con un claro vínculo con la desinformación y sus ramificaciones. Su objetivo es abordar no solo la exactitud y fiabilidad de la información, sino también las dinámicas sistémicas y estructurales que moldean el flujo y la accesibilidad del contenido en el espacio digital. Es importante mencionar que este concepto está siendo construido integrando perspectivas de derechos humanos, confianza pública y participación democrática, con una importante vinculación con las plataformas digitales.

EL PROCESO PARA LOS PRINCIPIOS GLOBALES

El origen formal de los Principios Globales para la Integridad de la Información tiene su antecedente en 2021 con la publicación

del informe del Secretario General “Nuestra Agenda Común”.⁷ Este documento fue una respuesta directa a una solicitud de los Estados Miembros, plasmada en la declaración del 75° aniversario de la organización, que solicitaba recomendaciones para revitalizar el multilateralismo y hacer frente a los desafíos presentes y futuros, conviene mencionar que el informe sostiene que la humanidad se encuentra en una encrucijada histórica, potenciada en buena medida por la pandemia de covid-19, el cambio climático, los conflictos y las crecientes desigualdades.

El informe subraya la crisis de la información y el papel de las plataformas digitales, hace un llamado para terminar la “infodemia” y acabar con la “guerra contra la ciencia”, por lo que desarrolla propuestas clave para mejorar la cooperación digital. Entre esas propuestas destaca el desarrollo de un código de conducta mundial que promueva la “integridad en la información pública” involucrando a los Estados, medios de comunicación, organismos reguladores y las propias plataformas digitales, con la ONU actuando como facilitadora. Otros aspectos relevantes que aborda este documento son la regulación de los medios sociales, el reforzamiento de las leyes de libertad de información y la garantía de que la ciencia y los conocimientos especializados tengan un mayor peso.

En junio de 2023 se publicó el “Informe de Políticas 8: Integridad de la Información en las Plataformas Digitales”.⁸ Este documento sentó las bases intelectuales y políticas para la creación de un futuro “Código de Conducta”. A través de un diagnóstico exhaustivo, el informe identificó las causas estructurales de la crisis de desinformación que afecta al ecosistema digital, posicionándola como una de las principales amenazas a la integridad de la información.

7 Secretario General de las Naciones Unidas, *Informe. “Nuestra Agenda Común”*, Nueva York: Organización de las Naciones Unidas, 2021.

8 Organización de las Naciones Unidas (ONU), *Informe de políticas de Nuestra Agenda Común 8: Integridad de la información en las plataformas digitales*, 2023, <https://mexico.un.org/sites/default/files/2023-06/our-common-agenda-policy-brief-information-integrity-es.pdf>.

Es importante mencionar que en este informe se hace referencia a la integridad de la información como la exactitud, coherencia y fiabilidad de la información, aunque aclara que es una definición de trabajo al no existir un consenso universal. En este sentido, el informe identifica tres amenazas fundamentales. La primera es la desinformación, entendida como información falsa difundida intencionadamente para causar un grave perjuicio social. La segunda es la información errónea, que consiste en la difusión involuntaria de contenido inexacto, a menudo compartido de buena fe. Finalmente, el discurso de odio, el cual se define como cualquier comunicación que ataca o utiliza lenguaje peyorativo contra una persona o grupo debido a su identidad, ya sea por religión, etnia o género, entre otros.

Estos tres fenómenos, según el informe, no solo contaminan el ecosistema informativo, sino que también amenazan el progreso humano al impactar negativamente en la consecución de los 17 Objetivos de Desarrollo Sostenible (ODS). A partir de este análisis, el documento formuló una propuesta central: el desarrollo de un “Código de Conducta” de las Naciones Unidas para la Integridad de la Información en las Plataformas Digitales, como parte de una estrategia multilateral de alto nivel.

La formulación de los Principios se sustentó en una metodología de consulta global en una etapa posterior al informe. Esta consulta estuvo orientada a reflejar un enfoque multiactor, aplicada a una amplia variedad de actores involucrados en el ecosistema informativo. La pluralidad de participantes incluyó a Estados Miembros, sector privado, líderes juveniles hasta la amplia gama de organizaciones de la sociedad civil.⁹

Las dependencias responsables fueron mayormente los Centros de Información de la ONU, los cuales al ser una red global de 63

9 Organización de las Naciones Unidas, “Information Integrity Consultations Held in Washington”, *United Nations Information Center Washington*, 20 de noviembre de 2023, <https://www.un.org/en/information-center-washington/information-integrity-consultations-held-washington>.

centros¹⁰ que enlazan a la ONU con los países en los que prestan servicios, fueron quienes organizaron consultas con distintos actores para recabar aportes de expertos en el área.

Uno de los cambios más significativos durante el desarrollo de esta iniciativa fue la evolución conceptual de un “Código de Conducta” hacia unos “Principios Globales”. No se trató solo de un cambio semántico, sino un giro que evidenció la complejidad de establecer normas a nivel mundial en la era digital, así como las lecciones derivadas de un exhaustivo proceso de consulta. En este sentido, organizaciones de la sociedad civil y organismos técnicos que participaron activamente en las consultas —como la IFLA, el IPIE y ARTICLE 19— influyeron en esta evolución conceptual. Estas entidades coincidieron en la necesidad de un marco que priorizara los derechos humanos sobre las medidas punitivas y advirtieron que un enfoque regulatorio rígido podría derivar en censura.

En el ámbito bibliotecario, la Federación Internacional de Asociaciones e Instituciones Bibliotecarias (IFLA) emitió una serie de recomendaciones, en las que se subrayan tres puntos esenciales. En primer lugar, si bien la IFLA muestra su apoyo al llamado a abstenerse de usar o amplificar la desinformación y el discurso de odio, enfatiza que esta abstención no debe, en ninguna circunstancia, restringir la libertad de expresión legítima. En segundo lugar, recalca la necesidad de afirmar el derecho de acceso a la información de todo tipo, lo que implica no solo evitar los cierres de internet, sino prevenir la manipulación del acceso y del ancho de banda, tal como se plantea con la controvertida práctica de la “tarifa cero”. Finalmente, en estas recomendaciones la IFLA alienta a capitalizar el potencial de la colaboración con instituciones de información comunitarias, como las bibliotecas, para garantizar un panorama mediático libre, viable y plural. Respecto a los desincentivos, la IFLA

10 Organización de las Naciones Unidas, “Centros de Información de las Naciones Unidas en todo el mundo”, *Department of Global Communications*, consultado el 22 de julio de 2025, <https://www.un.org/es/department-of-global-communications/worldwide/united-nations-information-centres>.

celebra que los principios buscan desincentivar los modelos de negocio que priorizan la participación sobre los derechos humanos, pero advierte que tales medidas no deben conducir a un bloqueo excesivo y deben, ante todo, preservar los derechos humanos.¹¹

Otras organizaciones clave también aportaron perspectivas importantes. El Panel Internacional sobre el Entorno de la Información (IPIE), una organización científica independiente, destacó en sus recomendaciones que el entorno de la información está evolucionando rápidamente, con sesgos algorítmicos, manipulación y desinformación que impactan significativamente en la vida pública. En este contexto, el IPIE subrayó la necesidad de una mayor responsabilidad de las plataformas, transparencia, empoderamiento de los usuarios y un compromiso con la investigación independiente para construir un entorno de información global más saludable y seguro.¹²

Por otro lado, organizaciones de derechos humanos como ARTICLE 19 enfatizaron la necesidad de abordar las lagunas y debilidades del marco de los Principios, especialmente en lo que respecta a la protección de la libertad de expresión frente a leyes ambiguas, la reforma de los modelos de negocio de las plataformas digitales y la creación de mecanismos de supervisión sólidos. Para esta organización, una preocupación principal radica en que los Principios no abordan adecuadamente el uso indebido de las leyes de “desinformación” para suprimir la disidencia y el periodismo crítico. Esto, según ARTICLE 19, deja espacio para regulaciones vagas y excesivamente amplias que ya han sido utilizadas como herramientas represivas en diversos países.¹³

11 International Federation of Library Associations and Institutions (IFLA), “UN Research on a Code of Conduct on Information Integrity – IFLA Response”, 4 de enero de 2024, <https://repository.ifla.org/handle/20.500.14598/3168>.

12 International Panel on the Information Environment (IPIE), “IPIE Submission on the UN Code of Conduct for Information Integrity on Digital Platforms”, 2023, <https://www.ipie.info/research/tp-2023-1>.

13 ARTICLE 19, “UN: ARTICLE 19’s comments on the Global Principles for Information Integrity”, *ARTICLE 19*, 17 de septiembre de 2024, <https://www.article19.org/resources/un-article-19-global-principles-for-information-integrity/>.

PRINCIPIOS GLOBALES DE LA ONU PARA LA INTEGRIDAD DE LA INFORMACIÓN

En junio de 2024 se presentaron formalmente los Principios Globales para la Integridad de la Información. Durante la conferencia de prensa de lanzamiento, se hizo énfasis en la construcción de un “ecosistema de información más humano” con un importante llamado a los diversos actores del ecosistema digital. El documento final está articulado a través de cinco principios interconectados, con sus recomendaciones correspondientes para los distintos grupos de interés.

El primer principio, *Confianza y Resiliencia de la Sociedad*, expone la crisis fundamental que motiva el marco de la ONU: la vulnerabilidad sistémica de nuestro ecosistema informativo. En este principio la confianza y la resiliencia sociales son componentes clave de la integridad de la información. Por un lado, la confianza se refiere a la seguridad que tienen las personas en la fiabilidad y exactitud de la información a la que acceden y, por otro lado, la resiliencia en el contexto de la integridad de la información hace referencia a la capacidad de las sociedades para manejar la alteración o manipulación del ecosistema informativo. Este principio subraya que actores estatales y no estatales, que buscan beneficios políticos o financieros, junto con las grandes empresas tecnológicas —cuyo “poder significativo” e “influencia excesiva” moldean el ecosistema—, representan una amenaza, al erosionar la confianza. Las consecuencias de esta estructura afectan de manera desproporcionada a grupos vulnerables y marginados, como mujeres, jóvenes, personas adultas mayores, pueblos indígenas y la comunidad LGBTQ+. Además, el principio advierte sobre los riesgos exponenciales que introduce la inteligencia artificial generativa en este panorama.

El segundo principio, *Incentivos Positivos*, aborda directamente las implicaciones críticas para la integridad del ecosistema de información que surgen de los modelos de negocio actuales, especialmente aquellos basados en la publicidad dirigida y la monetización de contenido. Si bien este modelo ha impulsado un crecimiento económico sin precedentes, también ha creado incentivos para la

proliferación de desinformación y contenido de odio. Esto se debe en gran medida a la “economía de la atención”, donde los algoritmos priorizan la participación para maximizar los ingresos, amplificando así el contenido polarizante y emocionalmente cargado. Si bien este modelo de negocio requiere mayor transparencia, es importante tomar en cuenta los diversos modelos de negocio aplicados a las plataformas digitales, como los que son por suscripción, por publicidad y por acceso.¹⁴

A partir de este diagnóstico, el principio profundiza en una consecuencia directa: la complejidad y opacidad de los sistemas de publicidad digital. Estos procesos, en gran medida automatizados y con mínima supervisión humana, crean un riesgo sistémico donde los presupuestos publicitarios pueden financiar involuntariamente la desinformación, al ser colocados junto a contenido dañino sin el conocimiento del anunciante.¹⁵

Frente al argumento mencionado de que es “improbable que las empresas tecnológicas abandonen rápidamente sus modelos de negocio actuales”¹⁶, el principio propone una ruta práctica. En lugar de exigir un cambio total, impulsa “pasos intermedios”¹⁷ centrados en la responsabilidad y la transparencia. La propuesta clave es la adopción de políticas publicitarias que respeten los derechos humanos, garantizando que la tecnología publicitaria no se convierta en un motor de narrativas perjudiciales. Al hacerlo, no solo se protege el ecosistema informativo, sino que también se ofrece un incentivo positivo a los anunciantes al mejorar su rentabilidad y proteger su reputación de marca.

14 Nicolai Van Gorp y Olga Batura, “Challenges for Competition Policy in a Digitalised Economy”, European Parliament’s Committee on Economic and Monetary Affairs, 2015.

15 Waseem Ahmad, Ajay Sen, Charles Eesley y Erik Brynjolfsson, “Companies Inadvertently Fund Online Misinformation Despite Consumer Backlash,” *Nature* 630, núm. 8015 (2024): 123–131, <https://doi.org/10.1038/s41586-024-07404-1>.

16 Organización de las Naciones Unidas, *Principios Globales...*, 6.

17 Organización de las Naciones Unidas, *Principios Globales...*, 6.

El tercer principio, el *Empoderamiento Público*, se centra en la capacidad de los individuos para navegar el ecosistema informativo de manera autónoma, segura e informada. Este principio sostiene que el empoderamiento real implica que las personas puedan controlar su experiencia en línea, tomar decisiones conscientes sobre el contenido que consumen, expresarse libremente y acceder a fuentes de información diversas y fiables. Aunque el texto no lo expone, este principio se vincula con el papel fundamental de las bibliotecas y otras instituciones de la memoria que son clave para desarrollar programas de habilidades informativas y fortalecer la investigación en esta área, brindando las herramientas necesarias para que la ciudadanía desarrolle un pensamiento crítico y habilidades que les permitan discernir la información y participar activamente.

Sin embargo, este ideal de empoderamiento se ve directamente amenazado por la censura, que en la era digital adopta formas más sutiles pero igualmente corrosivas que la supresión estatal. El principio debe reconocer que la censura moderna opera a través de dos vías principales: la censura algorítmica, donde el diseño de las plataformas puede limitar la visibilidad de ciertos discursos sin una prohibición explícita, y la autocensura, como consecuencia directa de un entorno digital hostil, donde el acoso sistemático, las campañas de difamación y el temor a represalias obligan a las personas a silenciar sus propias opiniones. El resultado final de ambas formas de censura es el mismo: se reduce el espacio para el debate público y se silencian voces críticas y marginadas.

Por otro lado, si bien este principio enfatiza el control individual y el acceso a la información, es necesario también profundizar ampliamente en la necesidad de un empoderamiento colectivo. El empoderamiento público va más allá de la mera elección personal; implica la capacidad de la sociedad para participar en la deliberación y la construcción de consenso.

El cuarto principio, *Medios Independientes, Libres y Plurales*, postula que la integridad de la información es fundamentalmente inalcanzable sin un sector mediático que encarne estas características. Defiende el periodismo como un bien público esencial para

la democracia y el Estado de Derecho, permitiendo un discurso cívico informado, la rendición de cuentas del poder y la protección de los derechos humanos. Este principio reconoce acertadamente la doble crisis que enfrentan los medios: por un lado, las graves amenazas a la seguridad de los periodistas, incluyendo el acoso, la violencia y la autocensura, que limitan su capacidad de informar libremente y operar con seguridad. Por otro lado, aborda los desafíos a su viabilidad económica, exacerbados por la migración masiva de ingresos publicitarios al espacio digital, dominado por las grandes empresas tecnológicas. Estos factores han permitido que intereses corporativos ejerzan un control creciente sobre los medios, amenazando la diversidad y socavando el periodismo local y de interés público.

Sin embargo, este principio podría ser mucho más contundente al abordar la compleja y a menudo problemática relación de los medios con las plataformas digitales. Muchos medios se han vuelto algorítmicamente dependientes de estas plataformas para asegurar su visibilidad y la distribución del contenido que producen. Esto los obliga a adaptar su contenido a algoritmos opacos y en constante cambio para ganar la atención de la audiencia, lo cual puede derivar en la limitación de su independencia editorial y en buena medida pueden verse inadvertidamente incentivados a amplificar narrativas desinformativas o sensacionalistas que los algoritmos favorecen, desdibujando así la línea entre periodismo riguroso y contenido viral. Además del apoyo a las organizaciones de noticias y la seguridad de los periodistas, el principio podría reforzar las medidas para una revisión fundamental de cómo las plataformas digitales interactúan con el contenido periodístico.

Finalmente, el quinto principio, *Transparencia e Investigación*, puede ser la pieza clave para la rendición de cuentas. Este principio diagnostica que la falta de acceso a los datos de las plataformas no es solamente un obstáculo técnico, sino un problema de poder y equidad a escala global, donde un número limitado de empresas, a menudo en estrecha relación con Estados y actores políticos y económicos, controla un volumen de información sin precedentes. Además, las decisiones regulatorias sobre transparencia tomadas en

Inteligencia artificial...

unos cuantos países que son sede de estas empresas impactan desproporcionadamente al resto del mundo.

Para que la transparencia sea verdaderamente efectiva, el principio subraya la necesidad de implementarla mediante mecanismos que superen la inherente resistencia de las plataformas. Esto exige la creación de marcos legales robustos que garanticen el acceso a datos que respeten la privacidad para la investigación independiente, el desarrollo de infraestructuras seguras para su análisis y una inversión estratégica en una comunidad de investigación.

El documento identifica como partes interesadas en estos principios a una serie de actores como las empresas tecnológicas, los agentes de la inteligencia artificial, publicistas y otros agentes del sector privado, medios informativos, investigadores y organizaciones de la sociedad civil, agentes estatales y políticos, y la propia ONU, cada uno con sus propias recomendaciones. Sin embargo, es necesario incorporar al más alto nivel a actores imprescindibles en el ecosistema informativo como las bibliotecas y otras instituciones de la memoria. Si bien para el desarrollo del documento se las contempló de forma indirecta, especialmente a través de la participación de organizaciones como la IFLA en las consultas, su rol no se eleva a la categoría de socio estratégico fundamental.

Esta ausencia es particularmente notoria en relación con el tercer principio de empoderamiento público. Como lo subrayó la IFLA durante el proceso de consulta, es fundamental capitalizar el inmenso potencial de la colaboración con instituciones de información comunitarias, como las bibliotecas, para garantizar un panorama mediático que sea libre, viable y plural. Estas instituciones funcionan como un contrapeso a la desinformación al proveer acceso a fuentes verificadas y, especialmente al desarrollar habilidades informativas en la ciudadanía.

CONCLUSIONES

Los Principios Globales para la Integridad de la Información de la ONU constituyen un esfuerzo importante en la cooperación

multilateral encaminado a restaurar un ecosistema digital fracturado. A través de sus cinco principios, se reconoce acertadamente que el problema no es solo la desinformación, sino las estructuras económicas y de poder que la incentivan, como la “economía de la atención” y la opacidad de las plataformas digitales.

A pesar de su ambición, los Principios enfrentan obstáculos importantes que desafían su implementación práctica. El marco se enfrenta a la inercia de modelos de negocio profundamente arraigados que priorizan la viralidad sobre el bienestar social, lo que profundiza la amenaza exponencial de la inteligencia artificial generativa, capaz de escalar la producción de desinformación a niveles sin precedentes. Además, como advierten organizaciones de la sociedad civil, existe el riesgo de que regulaciones bien intencionadas puedan ser instrumentalizadas para suprimir la libertad de expresión y el periodismo crítico, una debilidad que el marco no termina de resolver.

El marco de los Principios puede mejorar sustancialmente si eleva a las bibliotecas y otras instituciones de la memoria a la categoría de socios estratégicos. Esta elevación representa una importante oportunidad, particularmente en lo que respecta al principio de empoderamiento público. Dada su misión histórica, su estructura y sus responsabilidades, estas instituciones constituyen un medio imprescindible para fomentar habilidades críticas en la ciudadanía, entre otros múltiples aspectos, los cuales son esenciales para fortalecer el ecosistema informativo. En última instancia, para que los Principios Globales trasciendan de una declaración teórica a una transformación comunitaria efectiva, es fundamental que articulen de manera sostenida a los distintos actores en el ecosistema informativo contemporáneo.

BIBLIOGRAFÍA

- Ahmad, Waseem, Ajay Sen, Charles Eesley y Erik Brynjolfsson. “Companies Inadvertently Fund Online Misinformation Despite Consumer Backlash.” *Nature* 630, no. 8015 (2024): 123-131. <https://doi.org/10.1038/s41586-024-07404-1>.
- ARTICLE 19. “UN: ARTICLE 19’s comments on the Global Principles for Information Integrity.” *ARTICLE 19*, 17 de septiembre de 2024. <https://www.article19.org/resources/un-article-19-global-principles-for-information-integrity/>.
- Belli, Luca, Yasmin Curzi, Rolf Weber, *et al.* “Platform responsibilities for information integrity”. Dynamic Coalition on Platform Responsibilities at the Internet Governance Forum (IGF), 2024.
- Harley, Kelsey, y Rodney Cooper. “Information Integrity: Are We There Yet?”. *ACM Computing Surveys* 54, núm. 2 (2022): 1-35. <https://doi.org/10.1145/3436817>.
- International Federation of Library Associations and Institutions (IFLA). “UN Research on a Code of Conduct on Information Integrity – IFLA Response”. 4 de enero de 2024. <https://repository.ifla.org/handle/20.500.14598/3168>.
- International Panel on the Information Environment (IPIE). “IPIE Submission on the UN Code of Conduct for Information Integrity on Digital Platforms”. 2023. <https://www.ipie.info/research/tp-2023-1>.
- International Panel on the Information Environment. “Information Integrity about Climate Science: A Systematic Review”. Editado por Ece Elbeyi *et al.* Zurich, Switzerland: IPEI, junio de 2025. <https://doi.org/10.61452/btzip3426>.
- Organización de las Naciones Unidas. “Centros de Información de las Naciones Unidas en todo el mundo”. *Department of Global Communications*. Consultado el 22 de julio de 2025. <https://www.un.org/es/department-of-global-communications/worldwide/united-nations-information-centres>.

- Organización de las Naciones Unidas. “Information Integrity Consultations Held in Washington”. *United Nations Information Center Washington*. 20 de noviembre de 2023. <https://www.un.org/en/information-center-washington/information-integrity-consultations-held-washington>.
- Organización de las Naciones Unidas. *Informe de políticas de Nuestra Agenda Común 8: Integridad de la información en las plataformas digitales*. 2023. <https://mexico.un.org/sites/default/files/2023-06/our-common-agenda-policy-brief-information-integrity-es.pdf>.
- Organización de las Naciones Unidas (ONU). “Principios Globales de las Naciones Unidas para la Integridad de la Información”. Junio de 2024. https://www.un.org/sites/un2.un.org/files/principios_globales_onu_integridad_informacion.pdf.
- Secretario General de las Naciones Unidas. *Informe. “Nuestra Agenda Común”*. Nueva York: Organización de las Naciones Unidas, 2021.
- Shiri, Ali. “Information Integrity, Academic Integrity, and Generative AI”. *Information Matters* 3, núm. 10, 2023. <https://doi.org/10.2139/ssrn.4601781>.
- Van Gorp *et al.* “Challenges for Competition Policy in a Digitalised Economy”. European Parliament’s Committee on Economic and Monetary Affairs, 2015.

Indicadores generales para la integridad de la información en la preservación de fotografías de prensa digital

YISSEL N. CARMENATE

MARCO A. MARIANO

*Posgrado en Bibliotecología y Estudios de la Información,
Universidad Nacional Autónoma de México*

INTRODUCCIÓN

El patrimonio documental e informativo está dotado de valores disímiles, como el histórico, el cultural y el social, por cuanto registra los sucesos, las tradiciones y las prácticas en su esencia archivística. Los estudiosos de la información históricamente han trabajado el ejercicio de la preservación junto con profesionales de todas las disciplinas que se han preocupado por la permanencia a largo plazo de los objetos de información analógica; esto refuerza el carácter trans- e interdisciplinario de este campo académico.

Para comenzar el trabajo de preservación digital, se debe tener en cuenta que las obras documentales —en todos sus formatos, incluyendo las obras digitales— son una parte fundamental de nuestro patrimonio cultural. Preservar, proteger y trabajar con las mismas para que las generaciones futuras tengan acceso a ellas constituye una tarea en la que las ciencias de la información desempeñan un papel importante.¹

1 IFLA, *Preservación del patrimonio cultural: Guía de la IFLA*, s/f, Editorial IFLA, The Hague Netherlands, 6.

La información digital tiene un importante valor económico como producto cultural y como fuente de conocimiento. Desempeña además un papel significativo en el desarrollo sostenible a nivel nacional, teniendo en cuenta que habitualmente los datos personales, gubernamentales y comerciales se crean de forma digital.²

La creación de documentos digitales mantiene ritmos de aceleración impredecibles, ya que su producción es incesante, su consumo es inmediato y su conservación no tiene plazo fijo. La facilidad informática con la que se puede crear, modificar, compartir y almacenar un documento digital no adelgaza la importancia panorámica de una gestión digital preventiva con procedimientos formales que ayuden a organizar, preservar y acceder de manera segura a la información en sus distintos formatos digitales a largo plazo.³

La *degradescencia*, término que sintetiza los conceptos de *degradación sistemática* y *obsolescencia*, acuñado por Mike Casey en 2015 (*degradescence*), constituye una realidad inminente.⁴ En un contexto donde se consumen contenidos digitales de manera masiva e informal, resulta indispensable anticipar estrategias de cuidado para los mismos. Dichas estrategias deben partir de criterios de valor que permitan garantizar que los objetos digitales a preservar sean reflejo de una realidad característica. En el caso de la preservación de objetos digitales, su aplicación varía en función de las características de cada tipología documental. Es por eso que la fotografía digital se aborda como un objeto de estudio debido a sus particularidades técnicas y a su función como registro visual.

Existen varias razones que justifican la necesidad de crear y aplicar políticas de información orientadas a la preservación digital de

2 Wilmer Alexander Melo Garzón, “Lineamientos para la preservación digital en las entidades distritales del sector salud”, tesis profesional, Universidad de la Salle, 2019.

3 David Alonso Leija Román, “Preservación digital distribuida y la colaboración interinstitucional: Modelo de preservación digital para documentos con fines de investigación en universidades de México”, tesis doctoral, Universitat de Barcelona, 2017, 5.

4 Perla Olivia Rodríguez Reséndiz, *El archivo digital sonoro*, México: Instituto de Investigaciones Bibliotecológicas y de la Información, 2020, 118.

las fotografías; una de ellas es el fenómeno llamado *diversificación de formatos*, que surge por el nacimiento de nuevos formatos para un mismo tipo de archivo, que a veces no compatibilizan entre sí. Por ejemplo, existen fotografías en formato .jpg, .tif, .bmp, .raw y .png que ocupan diferentes reproectores. Esto supone un desafío para gestionar objetos digitales.

Como señala Pérez, de poco sirven las políticas de gestión minuciosamente diseñadas o las tecnologías de alto nivel si no hay contenido con valor intrínseco; sin materia significativa, no hay nada que preservar. La meta de un sistema de preservación digital es proveer contenido que sea utilizable. Por ello, es importante no solo adquirir materiales dignos de preservación, sino también asegurar los permisos adecuados para su conservación a largo plazo, incluyendo las autorizaciones que permitan migrar el contenido a nuevos formatos y garantizar sus condiciones de acceso.⁵

Es importante garantizar que la preservación se concrete con la calidad requerida. No es coherente preservar por preservar: la integridad de los datos es un elemento esencial y la preservación a largo plazo carece de valor si los materiales digitales no cumplen con los principios básicos de usabilidad y accesibilidad.

Para ello, es necesario establecer y desarrollar políticas, estrategias y estándares para balancear diferentes perspectivas culturales, sociales y jurídicas en torno a temas tales como la fiabilidad de los documentos digitales, la privacidad de datos y la propiedad intelectual. Además, resulta indispensable el desarrollo de una relación sólida y continua entre los generadores de documentos digitales y los preservadores a lo largo del ciclo de vida de los documentos.⁶

5 Iván Pérez Pineda, “Plan de preservación digital del patrimonio documental digitalizado de la biblioteca Lafragua”, tesis de master, Universidad de Barcelona/Universidad Pompeu Fabra, 2017, 15.

6 Luciana Duranti, “El concepto de documento de archivo en entornos experienciales, interactivos y dinámicos: ensayo de discusión”, *Legajos*, núm. 15, enero-marzo, (2013): 131-158.

PRESERVACIÓN DE FOTOGRAFÍAS

Las fotografías analógicas —al igual que las cintas de video y los mapas, entre otros— fueron tratadas como materiales especiales, razón por la cual su preservación se centraba mayoritariamente en su degradación química natural.

En el ámbito digital, en cambio, emergen nuevas prácticas de cuidados y anticipación, pues se trata de un escenario diferente que no requiere sobres antiácidos ni controles de temperatura y humedad para fotogramas revelados. Es necesario establecer mecanismos para asegurar la permanencia de los materiales digitales.

La fotografía como documento digital es vulnerable a varios factores que pueden dañar su integridad, situación plenamente aplicable a las imágenes de la prensa digital. A continuación, se presentan los cuatro escenarios más recurrentes:

- Los errores humanos: La manipulación genera elementos de seguridad, pero a la vez de vulnerabilidad. Hay un factor en torno a las contingencias o accidentes como resultado de esta misma interacción humano-máquina. Por otra parte, pueden existir intereses malintencionados con el objetivo de dañar el patrimonio.
- Pérdida de los bits: El tiempo de vida de un soporte digital puede variar entre los cuatro o cinco años. Al paso de este tiempo se pueden presentar fallas que dañen los archivos.
- Los desastres naturales o catástrofes sociales: Los fenómenos naturales como son las inundaciones, huracanes e incendios son amenazas potenciales que ponen en riesgo los datos, las guerras, saqueos o robos también pueden dañar o hacer que se pierdan los contenidos digitales.
- Falta de recursos económicos: La carencia de un presupuesto asignado de forma sistemática afecta en el mantenimiento de los sistemas de almacenamiento masivo digital.⁷

7 Perla Olivia Rodríguez Reséndiz y J. A. Blanco Rivera, *Preservación digital de contenidos publicados en la web y las redes sociales*, México: UNAM, 2023, 108.

Estos criterios constituyen requisitos fundamentales que las fotografías digitales deberían cumplir para considerarse confiables. Varios autores en correspondencia con lo que dicta la Unesco han brindado sus perspectivas al respecto. Wong, por ejemplo, señala que la integridad implica mantener las propiedades significativas del objeto y la inalterabilidad de su información; que la autenticidad supone la preservación de su cualidad original y la garantía de que lo que se preserva es veraz; que la fiabilidad implica la posibilidad de mantener la confianza en el contenido preservado; y que la funcionalidad exige la conservación de las características mínimas para su correcta reproducción.⁸

La segunda perspectiva corresponde a los principios propuestos por la doctora Rodríguez. Aunque ella se refiere a la preservación digital de archivos sonoros, éstos son perfectamente aplicables a los materiales fotográficos en formato digital:

[...] priorizar la preservación de formatos abiertos y sin pérdida de datos, frente a los formatos propietarios y comprimidos; procurar la transparencia en la información digital, con formatos abiertos y sin compresión, a fin de que su representación pueda ser fácilmente analizada; utilizar la embebida de metadatos como una cualidad que agilice y ayude en las tareas de preservación; normalizar los formatos para tener consistencia documental dentro del archivo; evitar depender de un software, hardware o sistema operativo; impedir la instauración de mecanismos de seguridad que afecten o modifiquen los documentos y tomar en cuenta los derechos de autor desde el momento mismo en que un documento es incorporado al archivo.⁹

8 Blanca I. Álvarez Wong, “Los repositorios digitales para la conservación. Un acercamiento a la preservación digital a largo plazo”, *Ciencias de la Información* 48, núm. 2, mayo-agosto (2017): 16.

9 Perla Olivia Rodríguez Reséndiz, “Propuesta de principios que se deben tener en cuenta para la preservación de documentos sonoros de origen digital”, *Anales de Documentación* 20, núm. 2 (2017): 7.

Inteligencia artificial...

Estos principios son interdependientes: si un documento no puede asegurar su funcionalidad, debe verificarse su integridad, y al hacerlo entra en juego su autenticidad. Sería pertinente realizar un análisis que jerarquice estos elementos en función de las reacciones de dependencia.

Es necesario tener en cuenta otros aspectos descriptivos y de gestión que doten de trazabilidad a un archivo digital —en este caso fotográfico— y que ayuden a generar confianza. Por ello, dos de estos metadatos de preservación digital están estrechamente asociados a principios archivísticos: origen y procedencia digital. Estos tipos de metadatos explican cómo se construyó el archivo, con qué programas fue creado, los cambios de versiones y toda la información referida a otras copias. Existen categorías de metadatos que sirven para garantizar estrategias como la transferencia, la migración, la gestión de la calidad, la accesibilidad y la permanencia. Los metadatos técnicos y de derecho son necesarios para la interpretación de un archivo digital que incluye tanto información de contexto como códigos únicos de verificación de integridad.

LAS NORMAS ISO Y LA PRESERVACIÓN DIGITAL

Existen diferentes normas ISO relacionadas con la preservación digital. Estas ofrecen puntos de referencia, proponen soluciones y recomiendan acciones para garantizar que los documentos digitales sean accesibles a lo largo del tiempo. Asimismo, las normas son clave para aquellas organizaciones que consideran los documentos esenciales como evidencia de su patrimonio o de sus actividades y desean asegurar su autenticidad, integridad y usabilidad a lo largo del tiempo.¹⁰

El Sistema Abierto de Información de Archivo, OAIS por sus siglas en inglés, es la referencia obligatoria para la preservación digital, sistematizada en la norma ISO 14721. Aunque inicialmente fue

10 Vicent Giménez Chornet, “Criterios ISO para la preservación digital de los documentos de archivo”, *Códices*, 10 (2014): 148.

concebida para los datos espaciales, luego se extendió a la preservación digital por completo.

En este caso, la norma hace hincapié en que los objetivos de un sistema seguro deben resolver los siguientes temas organizativos: a) optimizar la preservación de documentos electrónicos a largo plazo, su archivo e integridad; b) ofrecer facilidades de búsqueda de información; c) asegurar la accesibilidad y el uso de documentos electrónicos”.¹¹ Además, define roles, paquetes de ingesta y de salida, un marco conceptual de referencia obligatoria.

La ISO 15489, relativa a la gestión de documentos, proporciona una guía clara sobre cómo administrar los documentos y archivos de una organización de manera efectiva. Las organizaciones pueden optimizar sus procesos de gestión de información, garantizar la integridad y autenticidad de los documentos, facilitar el acceso y recuperación de la información, y cumplir con los requisitos legales y regulatorios. Establece controles para garantizar que los documentos y archivos se mantengan íntegros y auténticos a lo largo de su ciclo de vida, evitando modificaciones no autorizadas o pérdidas de información.¹²

La ISO 16363, estrechamente vinculada con la ISO 14721, se orienta a mantener aspectos como la integridad y la autenticidad de los datos digitales con garantías de confianza.

La ISO 18492, *La conservación a largo plazo de la información basada en documentos*, proporciona una guía práctica metodológica para la conservación y recuperación de información electrónica auténtica basada en documentos; es decir, es información que mantiene contenido, contexto y relaciones funcionales propias de los documentos, sin hacer referencia a otros. Se relaciona con la preservación digital para contribuir a la integridad de los documentos, pues toma en cuenta aspectos importantes del período de conservación en términos de la expectativa de vida de la tecnología utilizada. Asimismo, considera el rol de las normas

11 Vicent Giménez Chornet, “Criterios ISO”..., 148.

12 Norma ISO 15489, “Cómo implementar la norma ISO 15489 para una gestión eficiente de documentos y archivos”, 2016, 95.

Inteligencia artificial...

de tecnología de la información neutras en el soporte del acceso a largo plazo. Su implementación requiere de especialistas en TI, gestores de documentos y archiveros.

La ISO 28500, relativa al archivado web, ofrece:

[...] una forma estándar de estructurar, gestionar y almacenar miles de millones de recursos recopilados de la web y de otros lugares. Se utiliza para crear aplicaciones para recopilar, gestionar, acceder, extraer e intercambiar contenido. Si bien representa el formato estándar único para archivos web, se ha adoptado más allá de la comunidad de archivos web para almacenar materiales digitalizados o en formato nativo digital. La forma en que se crearán los archivos WARC y se almacenarán y reproducirá los recursos dependerá de las implementaciones de software y aplicaciones.¹³

La ISO 27000 aborda la seguridad de la información mediante planificación, implementación, evaluación y mejora continua. Considera aspectos como: contexto de organización, liderazgo, planificación y operación. Incluye diferentes tipos de controles organizacionales, controles al personal, controles físicos y controles tecnológicos.

Existe una línea difusa en la preservación de materiales hemerográficos completos y las fotografías que forman parte del archivo. En el caso de sus controles, se incluyen copias de seguridad basadas en la redundancia y diversidad. Por ello, es necesario implementar medidas de seguridad robustas —controles de acceso, cifrado y auditorías—, esenciales para prevenir riesgos.

La integridad de los datos también es fundamental para mantener la confianza de los clientes y la credibilidad de la empresa. En un entorno donde los datos constituyen un activo valioso, garantizar su integridad es crucial para mantener relaciones positivas con los clientes y socios comerciales, quienes confían en que su información se maneje de manera segura y precisa.

13 International Organization for Standardization (ISO), *ISO 28500:2024 —Información y documentación— formato de archivo WARC*, Ginebra: ISO, 2024, 5.

El *hashing*, que funciona como una huella genética de un documento en formato digital, permite comprobar que la información permanece intacta. Cualquier mutación es notoria y, cuando se hacen comparaciones, las diferencias destacan. Esto se logra por la relación con la tecnología *blockchain*. Toda alteración accidental, natural o intencionalmente generada se puede identificar por el *hashing*.

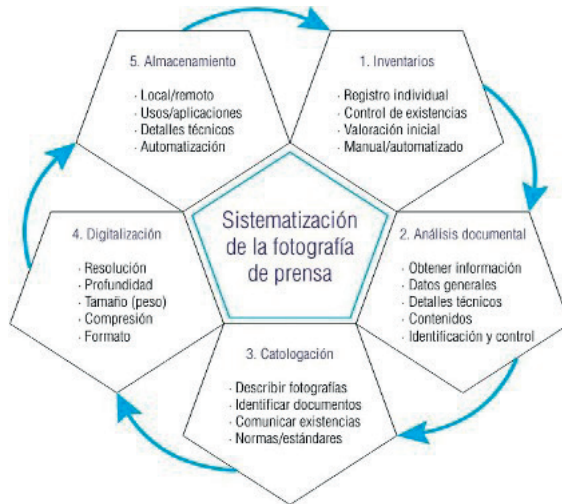
4. FOTOGRAFÍAS DE PRENSA: INDICADORES PARA SU PRESERVACIÓN

Los medios de prensa digital generan diariamente, como parte de sus vorágines editoriales y coberturas de eventos, numerosas imágenes: fotografías, caricaturas, infografías, etcétera. Por ello, al planificar su preservación es necesario definir criterios de selección que respondan a las características particulares y al contexto de uso, atendiendo a su valor estético, histórico, cultural y archivístico.

Deben considerarse aspectos como: “secciones donde se prioriza la imagen, cantidad de fotos en el archivo del diario, gestión y reutilización”.¹⁴ La imagen siguiente representa un ciclo de vida de la sistematización de las fotografías de prensa, donde confluyen los aspectos de procesamiento intelectual y de administración, datos fundamentales para la optimización de los procesos. Si se toman en cuenta todos los aspectos mencionados para garantizar integridad, es posible identificar varios elementos que sirven como indicadores generales para la integridad de las fotografías de prensa.

14 J. C. Rivera Aguilera, “Fotoperiodismo en los medios impresos, aplicado en Sol de San Luis”, tesis doctoral, Universidad Complutense de Madrid, 2020, 159.

Figura 1. Ciclo de vida de la fotografía de prensa.



Fuente: Julio C. Rivera Aguilera, "Fotoperiodismo en los medios impresos...", 159.

Encriptación: La comprobación de integridad bajo instrumentos para la encriptación, por ejemplo: MD5 o SHA-1, que definen bajo un grado de seguridad la composición numérica de integridad de los contenidos *ingestados* al sistema de preservación digital para su acceso o seguridad. Además, si no han sido proporcionados, se deberán proporcionar complementariamente.¹⁵ La encriptación con los algoritmos adecuados puede ser un indicador de integridad, no solo por la huella digital que se genera con su trazabilidad, sino porque es un documento difícil de alterar y a la vez confidencial.

15 David Alonso Leija Román, "Preservación digital distribuida y la colaboración interinstitucional: Modelo de preservación digital para documentos con fines de investigación en universidades de México", tesis doctoral, Universitat de Barcelona, 2017, 210.

Controles de acceso: Algunas organizaciones incorporan niveles adicionales de seguridad en las actividades de mantenimiento y gestión de sus memorias digitales. Aplican un exigente control de acceso y se enfocan de manera primaria en las actividades de preservación digital (migración, normalización, etcétera).¹⁶ Se considera una herramienta necesaria para garantizar la seguridad, para verificar que las fotografías se utilicen con los fines adecuados y minimizar ediciones no autorizadas, sean accidentales o malintencionadas.

Copias de seguridad y control de datos: Este enfoque es complementario porque une actividades de orden técnico con las de definición de políticas y prácticas de preservación digital a corto o largo plazo, las cuales incluyen la aplicación de políticas de verificación de integridad y copias de seguridad, procurando el uso de formatos estables de ficheros que garanticen la fiabilidad de los medios de almacenamiento. También contempla la identificación de los datos con valor para el largo plazo mediante políticas de depuración, la asignación de identificadores persistentes y la identificación de posibles datos sensibles.¹⁷

Sin lugar a dudas, estos elementos constituyen recursos esenciales para la preservación digital, y se recomienda su existencia tanto local como remota. Su implementación debe formalizarse en una política detallada con enfoque en escenarios de contingencia, de modo que se garantice la verificación de las versiones de las fotografías.

Check-sums y verificación de integridad: Los procesos de *check-sums* o verificación de sumas se emplean cíclicamente para monitorear la integridad de la información. En caso de detectarse corrupción, pueden generarse copias de reemplazo. Para ello, es fundamental mantener copias almacenadas en ubicaciones diferentes.¹⁸ Este procedimiento es valioso para la detección de fotografías corruptas y, por lo tanto, es inseparable del concepto de integridad

16 David Alonso Leija Román, "Preservación digital...", 62.

17 David Alonso Leija Román, "Preservación digital...", 37.

18 David Alonso Leija Román, "Preservación digital...", 90.

Inteligencia artificial...

de los datos. La preservación digital requiere de herramientas correctas que verifiquen que las fotografías no hayan sido alteradas.

Control de dispositivos: A pesar de que la preservación es netamente digital, existe un componente físico, el *hardware*, sobre el cual es necesario establecer controles. Los dispositivos de almacenamiento pueden volverse obsoletos cuando dejan de ser compatibles con el *hardware* al que se conectan. Frente a la caducidad de los soportes digitales, se han sugerido estrategias de preservación encaminadas a garantizar la permanencia de los objetos digitales. Entre ellas destacan: la migración y la emulación. La estrategia más utilizada es la migración sistemática de un soporte que progresivamente entra en desuso hacia otro vigente. Se estima que el tiempo de vida útil de los soportes digitales no excede los cinco años. Por lo tanto, es necesario contar con un plan de migración cíclico.¹⁹

Auditorías: Cuando se habla de auditorías en preservación digital, es imprescindible referirse a los niveles metodológicos de la Alianza Nacional para la Custodia Digital (NDSA, por sus siglas en inglés), que contribuyen a la seguridad mediante revisiones sistemáticas y otros mecanismos de control.

La tarea de auditar la veracidad y fiabilidad con la que los repositorios gestionan objetos digitales ha llevado a la creación de diversos sistemas de auditoría que certifican repositorios digitales, los cuales se suelen clasificar en tres niveles: certificación básica, certificación extendida y certificación formal.²⁰

APROXIMACIÓN LEGAL

La preservación de fotografías de prensa digital en México exige considerar la Ley Federal de Derecho de Autor, que incluye diversos artículos pertinentes para regular las obras fotográficas difundidas en los diarios.

19 Perla Olivia Rodríguez Reséndiz y J. A. Blanco Rivera, *Preservación digital...*, 108.

20 David Alonso Leija Román, "Preservación digital...", 74.

Una aproximación legal en México requiere revisar distintas leyes relacionadas con la titularidad de los derechos de las fotografías y con las condiciones contractuales para la creación de una obra. Los derechos patrimoniales, el uso y la explotación de la obra fotográfica están además mediados por la protección de la privacidad, ya que el retrato puede considerarse un dato personal según las leyes federales mexicanas.

La Ley Federal de Derecho de Autor en las fotografías de prensa digital

Es necesario articular la preservación digital con el marco legal que regula las fotografías. Tal es el caso del artículo 83:

Artículo 83. Salvo pacto en contrario, la persona física o moral que comisione la producción de una obra o que la produzca con la colaboración remunerada de otras, gozará de la titularidad de los derechos patrimoniales sobre la misma y le corresponden las facultades relativas a la divulgación, integridad de la obra y de colección sobre este tipo de creaciones.²¹

El artículo establece que deben respetarse los derechos patrimoniales, incluso en casos de colaboración remunerada. Por su parte, el artículo 83 bis señala:

Artículo 83 bis. Para que una obra se considere realizada por encargo, los términos del contrato deberán ser claros y precisos, en caso de duda, prevalecerá la interpretación más favorable al autor. El autor también está facultado para elaborar su contrato cuando se le solicite una obra por encargo.²²

21 Cámara de Diputados del H. Congreso De La Unión, *Ley Federal de Derecho de Autor*, diciembre de 2024, 15.

22 Cámara de Diputados, *Ley Federal de Derecho de Autor*, 15.

Inteligencia artificial...

Así, cuando la obra se realiza por encargo, los términos deben ser claros y precisos. También se estipula que el autor tiene la personalidad jurídica de elaborar su propio contrato.

En el Capítulo II. De las obras fotográficas, plásticas y gráficas, los artículos 85 a 87 establecen:

Artículo 85. Salvo pacto en contrario, se considerará que el autor que haya enajenado su obra pictórica, escultórica y de artes plásticas en general, no ha concedido al adquirente el derecho de reproducirla, pero sí el de exhibirla y el de plasmarla en catálogos. En todo caso, el autor podrá oponerse al ejercicio de estos derechos, cuando la exhibición se realice en condiciones que perjudiquen su honor o reputación profesional.²³

Solo el autor conserva el derecho a exhibir su obra, pero no de reproducirla sin su consentimiento expreso, y puede oponerse a dicho uso.

Artículo 86. Los fotógrafos profesionales solo pueden exhibir las fotografías realizadas bajo encargo como muestra de su trabajo, previa autorización. Lo anterior no será necesario cuando los fines sean culturales, educativos, o de publicaciones sin fines de lucro.²⁴ Se debe avisar a los fotógrafos profesionales de la exhibición de sus obras bajo encargo, pero no es así cuando los fines son culturales, educativos o de publicaciones que no persiguen el lucro.

Artículo 87. El retrato de una persona solo puede ser usado o publicado, con su consentimiento expreso, o bien con el de sus representantes o los titulares de los derechos correspondientes. La autorización de usar o publicar el retrato podrá revocarse por quien la otorgó quién (sic.), en su caso, responderá por los daños y perjuicios que pudiera ocasionar dicha revocación. Cuando a cambio de una remuneración, una persona se dejare retratar, se presume que ha otorgado el consentimiento a que se refiere el párrafo

23 Cámara de Diputados, *Ley Federal de Derecho de Autor*, 15.

24 Cámara de Diputados, *Ley Federal de Derecho de Autor*, 15.

anterior y no tendrá derecho a revocar, siempre que se utilice en los términos y para los fines pactados. No será necesario el consentimiento a que se refiere este artículo cuando se trate del retrato de una persona que forme parte menor de un conjunto o la fotografía sea tomada en un lugar público y con fines informativos o periodísticos. Los derechos establecidos para las personas retratadas durarán 50 años después de su muerte.²⁵

El retrato de una persona solo podrá ser usado o publicado bajo su consentimiento o con de los representantes o titulares de los derechos correspondientes. No será necesario el consentimiento cuando forme menor parte del conjunto de la fotografía y esta sea tomada en un lugar público siguiendo los fines informativos y periodísticos.

Además, deben contemplarse la Ley General de Protección de Datos Personales en Posesión de Sujetos Obligados, la Ley Federal de Protección de Datos Personales en Posesión de los Particulares y la Ley de Procedimiento Administrativo de la Ciudad De México, dado que una fotografía de una persona constituye un dato personal sujeto a regulación.

CONCLUSIONES

La fragilidad de las fotografías de prensa y su valor estético junto con las cantidades monumentales de documentos fotográficos digitales que se generan y comparten en la actualidad exige anticipar cuidados con mayor urgencia que nunca. La preservación digital incluye tanto aspectos técnicos como el compromiso social con la historia que se genera en el presente.

Existen numerosos indicadores para determinar la integridad de una fotografía con miras a largo plazo: encriptación, controles de acceso, copias de seguridad y control de datos, *check-sums*, verificación de integridad, control de dispositivos y auditorías.

25 Cámara de Diputados, *Ley Federal de Derecho de Autor*, 15.

Inteligencia artificial...

Todo ello es fundamental para garantizar la integridad de los datos a largo plazo.

Es imprescindible que la preservación digital se refiera a las normas ISO pertinentes, especialmente a OAIS 14721. Para la integridad de los datos, es necesario vincularla con la 27000, relativa a la seguridad informática.

La evolución de las TI ha generado desafíos en el marco legal y en la preservación digital de imágenes de prensa. Estas presentan características complejas, condicionadas por factores como los derechos de autor, la privacidad y la protección de datos personales. Cualquier proyecto de preservación digital debe iniciar con la revisión de las disposiciones legales para orientar adecuadamente el trabajo de los fotoreporteros y los preservadores en la gestión de la fotografía digital.

BIBLIOGRAFÍA

- Blanca I. Á., “Los repositorios digitales para la conservación. Un acercamiento a la preservación digital a largo plazo”. *Ciencias de la Información* 48, núm. 2 (2017):15-22. <https://www.re-dalyc.org/articulo.oa?id=181454540003>.
- Cámara de Diputados del H. Congreso de la Unión. *Ley Federal de Derecho de Autor*. México: Cámara de Diputados, 2024. <https://www.diputados.gob.mx/LeyesBiblio/pdf/LFDA.pdf>.
- Cámara de Diputados del H. Congreso de la Unión. *Ley Federal de Protección de Datos Personales en Posesión de los Particulares*. México: Cámara de Diputados, 2024. <https://www.diputados.gob.mx/LeyesBiblio/pdf/LFPDPPP.pdf>.
- Cámara de Diputados del H. Congreso de la Unión. *Ley General de Protección de Datos Personales en Posesión de Sujetos Obligados*. México: Cámara de Diputados, 2024. <https://www.diputados.gob.mx/LeyesBiblio/pdf/LGPDPPSO.pdf>.

- Cornell University Library. "Llevando la teoría a la práctica: tutorial de digitalización de imágenes". Consultado el 10 de octubre de 2024. <http://preservationtutorial.library.cornell.edu/tutorial-spanish/preservation/preservation-05.html>.
- Duranti, Luciana. "El concepto de documento de archivo en entornos experienciales, interactivos y dinámicos: ensayo de discusión". *Legajos*, núm. 15 (2013): 131-158.
- Giménez Chornet, Vicent. "Criterios ISO para la preservación digital de los documentos de archivo". *Códices*, núm. 10 (2014): 135-150. <https://vicentgimenez.net/Criterios-ISO.pdf>.
- Global Suite Solutions. "¿Qué es la norma ISO 27001 y para qué sirve?". <https://www.globalsuitesolutions.com/es/que-es-la-norma-iso-27001-y-para-que-sirve/#:~:text=La%2norma%2ISO%227001%20es,y%20disponibilidad%20de%20la%20informaci%C3%B3n>.
- IFLA. *Preservación del patrimonio cultural: Guía de la IFLA*. Países Bajos: IFLA, s.f. <https://www.ifla.org/wp-content/uploads/2019/05/assets/hq/topics/cultural-heritage/documents/preserving-cultural-heritage-es.pdf>.
- International Organization for Standardization. ISO 28500. Ginebra: ISO, 2017. <https://www.iso.org/obp/ui/en/#iso:std:iso:28500:ed-2:v1:en>.
- International Organization for Standardization. "Cómo implementar la norma ISO 15489 para una gestión eficiente de documentos y archivos Norma ISO 15489". ISO 15489, 2016. https://www.iso.org/standard/62542.html?utm_source=chatgpt.com.
- Melo Garzón, Wilmer Alexander. *Lineamientos para la preservación digital en las entidades distritales del sector salud*. Tesis profesional, Universidad de La Salle, 2019. https://ciencia.lasalle.edu.co/cgi/viewcontent.cgi?article=1987&context=sistemas_informacion_documentacion.
- Owens, Trevor. *The Theory and Craft of Digital Preservation*. Baltimore: Johns Hopkins University Press, 2018.

Inteligencia artificial...

- Pérez Pineda, Iván. *Plan de preservación digital del patrimonio documental digitalizado de la biblioteca Lafragua*. Tesis de máster, Universidad de BarcelonaUniversidad Pompeu Fabra, 2017. https://diposit.ub.edu/dspace/bitstream/2445/117724/1/TFM_Perez_Pineda_Ivan.pdf.
- Rodríguez, Perla Olivia. “Propuesta de principios que se deben tener en cuenta para la preservación de documentos sonoros de origen digital”. *Anales de Documentación* 20, núm. 2 (2017): 1-8. <https://www.redalyc.org/articulo.oa?id=63552793001>.
- Rivera Aguilera, J. C. *Fotoperiodismo en los medios impresos, aplicado en Sol de San Luis*. Tesis doctoral, Universidad Complutense de Madrid, 2020. <https://produccioncientifica.ucm.es/documentos/6052b13146b9d97defd11449>.
- Rodríguez Reséndiz, Perla Olivia, y J.A. Blanco Rivera. *Preservación digital de contenidos publicados en la web y las redes sociales*. México: Universidad Nacional Autónoma de México, 2023.
- Rodríguez Reséndiz, Perla Olivia. *El archivo digital sonoro*. Ciudad de México: Instituto de Investigaciones Bibliotecológicas y de la Información, 2020.

Inteligencia artificial e integridad de la información. Perspectivas para un uso responsable. Instituto de Investigaciones Bibliotecológicas y de la Información/ UNAM. Edición digital. Coordinación editorial: Angélica Valenzuela. Revisión especializada: Daniela Rico Traffon. Revisión especializada en inglés: Héctor González Villatoro y Marcos Emilio Bustos Flores. Formación editorial: Itzia I. Solís González. Corrección de pruebas: Daniela Rico Traffon. Versión digital: Héctor González Villatoro. Se publicó en febrero de 2026.