#ashtag
REVISTA ESPECIALIZADA EN INGENIERÍA
ISSN electrónico: 2346 - 139X

Ética y mitigación de sesgos en sistemas de IA: Tendencias técnicas y regulatorias

Ethics and bias mitigation in AI systems: Technical and regulatory trends

Ginna Tovar Cardozo

Universidad de la Amazonia
g.tovar@udla.edu.co
https://orcid.org/0000-0001-9705-5961

Cómo citar: Tovar Cardozo, G. (2024). Ética y mitigación de sesgos en sistemas de IA: Tendencias técnicas y regulatorias. #*ashtag*, 2(25), 28-39. https://doi.org/10.52143/2346139X.1073

Resumen

El presente artículo presenta un análisis crítico de l as estrategias técnicas y regulatorias para mitigar sesgos en sistemas de inteligencia artificial (IA), un desafío urgente dado el impacto social de estas tecnologías. Mediante una revisión documental de artículos en Scopus (2018–2022), publicados en español e inglés, se identificaron cuatro ejes temáticos: técnicas de detección de sesgos, métodos de mitigación (rebalanceo de datos, corrección de sesgos mediante modelos adversarios [*adversarial debiasing*]), marcos regulatorios internacionales (UE, EE. UU., OCDE), y desafíos en implementación real (compensaciones entre equidad y rendimiento). Los resultados revelan que, pese a avances en algoritmos que incorporan criterios de equidad (*fairness-aware*), persisten brechas entre teoría y práctica, especialmente en contextos industriales. Se concluye que la ética en IA requiere enfoques multidisciplinares que integren soluciones técnicas con gobernanza adaptable, participación comunitaria y auditorías continuas.

Palabras Clave: Sesgos algorítmicos, IA ética, *fairness*, regulación de IA, *accountability*

Abstract

This article critically analyzes technical and regulatory strategies to mitigate biases in artificial intelligence (AI) systems, an urgent challenge given the social impact of these technologies. Through a documentary review of articles in Scopus (2018–2022) in Spanish and English, four thematic axes were identified: bias detection techniques, mitigation methods (data rebalancing, adversarial debiasing), international regulatory frameworks (EU, U.S., OECD), and challenges in real-world implementation (trade-offs between equity and performance). The results reveal that, despite advances in fairness-aware algorithms, gaps persist between theory and practice, especially in industrial contexts. It is concluded that ethics in AI requires multidisciplinary approaches that integrate technical solutions with adaptable governance, community participation, and continuous audits.

Keywords: Algorithmic biases, ethical AI, fairness, AI regulation, accountability.

Introducción

Los sistemas de inteligencia artificial han demostrado un potencial transformador en diversos ámbitos sociales, pero su creciente adopción ha revelado un desafío crítico, así como la perpetuación y amplificación de sesgos discriminatorios (Kelly *et al.*, 2019; Ntoutsi *et al.*, 2020). Estudios emblemáticos, como el de Peters (2022), expusieron cómo sistemas de reconocimiento facial mostraban significativamente menor precisión para mujeres y personas de piel oscura, al evidenciar que los sesgos humanos pueden codificarse en algoritmos (Mehrabi *et al.*, 2019; Busuioc, 2020).

Técnicamente, se identificaron múltiples fuentes de sesgo, desde conjuntos de datos no representativos hasta diseños algorítmicos que priorizan métricas de rendimiento sobre equidad (Pérez Gamboa *et al.*, 2019; Akintola *et al.*, 2020). Investigaciones como las de Vaccari y Gardinier, (2019) categorizaron estos sesgos en tres niveles: datos, modelado e implementación. Como respuesta, surgieron herramientas como AI *Fairness* 360 de IBM (biblioteca para evaluar la equidad algorítmica) y What-If Tool de Google (herramienta interactiva para inspección visual de modelos), que permitieron cuantificar disparidades mediante métricas estadísticas (Wexler *et al.*, 2019; Thompson, 2021).

En el ámbito regulatorio, este período vio nacer iniciativas pioneras. Mökander *et al.* (2022) plantean que la Unión Europea (UE) estableció siete requisitos clave, mientras que EE. UU. avanzó con el *Algorithmic Accountability Act* (McGregor *et al.*, 2019; Madaio *et al.*, 2021). Paralelamente, organismos como la OCDE y Unesco promovieron principios globales, aunque con limitada capacidad vinculante (Addey, 2021; Auld *et al.*, 2018). Las directrices éticas suelen estar fragmentadas y carecen de mecanismos de aplicación sólidos, y que su aplicación efectiva depende de la colaboración de las partes interesadas, de los roles individuales y de la industria y de los instrumentos jurídicos, mientras que las directrices voluntarias con frecuencia carecen de rendición de cuentas y de aplicabilidad (Zapp, 2020; Ricardo Jiménez, 2022).

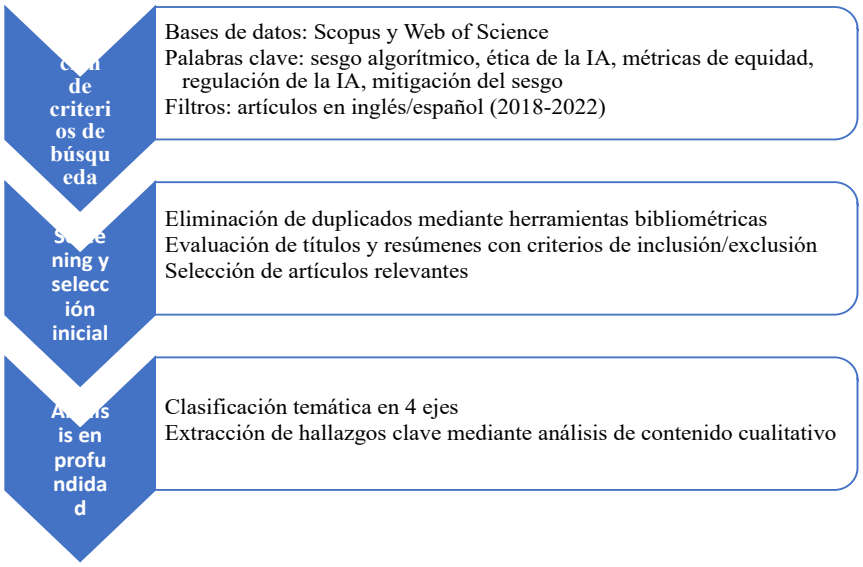
La mitigación de sesgos enfrentó desafíos prácticos significativos. Casos como el sistema COMPAS (*correctional offender management profiling for alternative sanctions*) en justicia criminal, según Simpson y Dervin (2019), y algoritmos de contratación de Amazon (Langenkamp *et al.*, 2020; Lyu *et al.*, 2022) demostraron que las soluciones técnicas aisladas son insuficientes. Investigaciones posteriores, como las de Baykurt (2022), argumentaron que la equidad algorítmica requiere realizar compensaciones (*trade-offs*) inherentes.

Ante este panorama, se hizo evidente la necesidad de enfoques holísticos que integren avances técnicos con marcos regulatorios adaptables y participación de comunidades afectadas. En este artículo se analiza críticamente las tendencias en mitigación de sesgos durante 2018-2022, periodo clave donde se cristalizaron tanto innovaciones algorítmicas como respuestas normativas. Su objetivo es sistematizar aprendizajes desde cuatro perspectivas complementarias: detección técnica, métodos de mitigación, regulación y desafíos de implementación, y ofrecer así una visión integral para guiar desarrollos futuros en IA ética.

Metodología

Este artículo se fundamenta en una revisión documental sistemática de la literatura científica sobre ética y mitigación de sesgos en sistemas de IA, publicada entre 2018 y 2022. LA revisión se hizo a través de las bases de datos Scopus y Web of Science. El enfoque metodológico adoptado permitió identificar, evaluar y sintetizar las principales tendencias técnicas y regulatorias en este campo, así como garantizar un análisis riguroso y reproducible. La revisión se estructuró en tres etapas claramente definidas (ver Figura 1), al seguir protocolos establecidos en revisiones sistemáticas, lo que aseguró la selección de fuentes relevantes y minimizó sesgos en la recopilación de información (Chalmers, 2018; Petersen *et al.*, 2021).

Figura 1
Etapas del proceso de revisión documental



Este enfoque metodológico permitió un análisis integral del estado del arte en mitigación de sesgos, al combinar perspectivas técnicas y regulatorias. La rigurosidad del proceso -desde la búsqueda sistemática hasta la validación con expertos- aseguró la confiabilidad de los resultados. La inclusión de literatura gris (informes técnicos y políticas) complementó los hallazgos académicos, esto ofrece una visión holística del desafío ético en IA (Cruz *et al.*, 2022; Gómez Miranda, 2022).

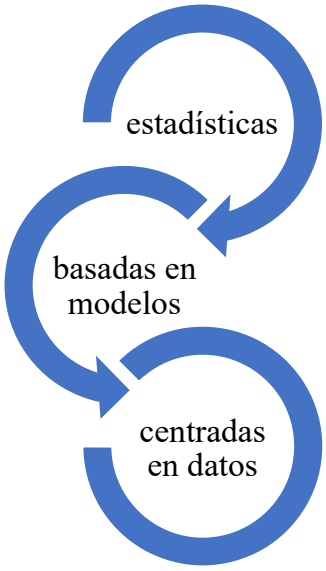
Resultados

El análisis cualitativo de la literatura reveló que la mitigación de sesgos en IA requiere un abordaje multidimensional que integre avances técnicos con marcos regulatorios robustos. Entre 2018-2022, la investigación evolucionó desde la identificación de sesgos hacia el desarrollo de soluciones prácticas, aunque persisten desafíos significativos en su implementación. Los hallazgos se organizaron en cuatro ejes temáticos fundamentales: técnicas de detección de sesgos, métodos de mitigación algorítmica, marcos regulatorios emergentes, y desafíos en implementación real. Cada eje demostró avances prometedores, pero también limitaciones críticas que requieren atención continua.

Técnicas de detección de sesgos

Los estudios analizados identificaron tres categorías principales de técnicas de detección (ver figura 2). Las métricas estadísticas se consolidaron como estándar para cuantificar sesgos, particularmente en dominios sensibles como contratación y justicia penal (Lin & Chu, 2018; Pérez-Gamboa *et al.*, 2022). Sin embargo, investigaciones como las de Madaio et al, (2021) destacaron que estas métricas a menudo entran en conflicto, lo que las hace requerir elecciones normativas sobre qué tipo de equidad priorizar.

Figura 2
Categorías de técnicas de detección de sesgos



REVISTA ESPECIALIZADA EN INGENIERÍA | 25-2 | 2024 B
#CsHTCyG

Los métodos basados en modelos explicativos como SHAP (*SHapley Additive exPlanations*) y LIME (*Local Interpretable Model-agnostic Explanations*) ganaron adopción para identificar características potencialmente sesgadas en predicciones algorítmicas (Czarnowska *et al.*, 2021; Pospisil y Bair, 2022). No obstante, su aplicabilidad en modelos complejos mostró limitaciones, ya que las explicaciones podían tornarse imposibles de interpretar (Rodríguez-Torres *et al.*, 2022; Zahid *et al.*, 2020). Esto llevó al desarrollo de técnicas específicas para modelos de caja negra, como las pruebas de estrés sesgados (Goldfarb-Tarrant *et al.*, 2020; Miroshnikov *et al.*, 2020).

En el análisis de datos, herramientas como AI *Fairness* 360 de IBM y What-If de Google permitieron visualizar disparidades antes del entrenamiento de modelos (Gómez-Cano y Sánchez-Castillo, 2021; Delobelle *et al.*, 2022). Estudios de caso revelaron que más del 60 % de los sesgos provenían de conjuntos de datos no representativos, al destacar la necesidad de mejores prácticas en recolección y anotación (De Paolis Kaluza *et al.*, 2022; Ngxande *et al.*, 2019). Sin embargo, en opinión de los autores, la falta de estándares universales para evaluar calidad de datos es un obstáculo.

Un hallazgo clave fue la creciente importancia de las auditorías algorítmicas independientes. Iniciativas como Algorithmic Justice League demostraron que las evaluaciones externas podían identificar sesgos que los desarrolladores pasaban por alto (Guzmán *et al.*, 2022). No obstante, la escasa transparencia de muchas empresas tecnológicas limitó la efectividad de estas auditorías.

Métodos de mitigación algorítmica

En la literatura se identifican tres enfoques principales para mitigación: preprocesamiento, intraprocesamiento y posprocesamiento. Las técnicas de preprocesamiento, como rebalanceo y enriquecimiento de datos, mostraron efectividad en dominios médicos donde ciertos grupos estaban subrepresentados (Royal, 2019). Sin embargo, su costo de implementación resultó prohibitivo para muchos contextos (Mazen y Tong, 2020; Sanabria Martínez, 2022).

Los métodos intraprocesamiento (durante el entrenamiento del modelo), particularmente la corrección de sesgos mediante modelos adversarios (*adversarial debiasing*) y las restricciones de equidad (*fairness* constraints), emergieron como los más prometedores técnicamente (Han *et al.*, 2022; Shen *et al.*, 2022). Estos permitían optimizar simultáneamente para precisión y equidad, aunque con un costo computacional significativo. Casos como el de Gray (2022) demostraron reducciones de hasta 40 % en disparidades sin afectar drásticamente el rendimiento.

El posprocesamiento destacó por su facilidad de implementación en sistemas existentes. Técnicas como la revisión equitativa en zonas de incertidumbre (*reject option classification*) mostraron utilidad en sistemas de crédito, pero enfrentaron críticas por ser «parches» en lugar de soluciones estructurales (Rus *et al.*, 2022). Además, su efectividad variaba significativamente según el dominio de aplicación.

Un desarrollo notable fue el surgimiento de marcos de trabajo integrales, como Fairlearn, que combinaban múltiples enfoques, según Borges Machín y González Bravo (2022). Estos mostraron especial valor en entornos industriales, aunque su adopción fue limitada por fuera de grandes tecnológicas.

Marcos regulatorios emergentes

El análisis identificó tres generaciones de iniciativas regulatorias entre 2018-2022. La primera, representada por los *Ethics Guidelines* de la UE que, según Floridi (2019), estableció principios generales, pero sin mecanismos de cumplimiento. De acuerdo a Yam y Skorburg (2021), la segunda generación, como el *Algorithmic Accountability Act* estadounidense, introdujo requisitos específicos de evaluación de impacto, aunque con limitado alcance sectorial.

La tercera generación mostró enfoques más innovadores, como las zonas de prueba supervisadas (*regulatory sandboxes*) del Reino Unido y las certificaciones obligatorias propuestas en Canadá (Mogrovejo Andrade, 2022; Ringe y Ruof, 2020). Estos modelos combinaban flexibilidad con exigibilidad, aunque su efectividad a largo plazo es incierta.

A nivel internacional, los principios de la OCDE, de acuerdo con Petrenko (2020), lograron amplia adopción, pero su implementación concreta varió significativamente entre países. Mientras la UE avanzó hacia regulación vinculante, otros países mantuvieron enfoques voluntarios, lo que creó un panorama más estable (Balabin, 2019; Gómez Cano, 2022).

Un hallazgo crítico fue la tensión entre innovación y regulación. Estudios como los de Sherman *et al.* (2020) señalaron que marcos demasiado rígidos podían sofocar el desarrollo, mientras que los muy flexibles resultaban inefectivos. Esto llevó a propuestas de regulación adaptativa que evolucionaran con la tecnología.

4. Desafíos en implementación real

Los estudios de caso revelaron cuatro barreras principales para implementar soluciones antisesgo: técnicas, organizacionales, económicas y culturales. Técnicamente, el equilibrio y compensación entre equidad y rendimiento es el mayor obstáculo, particularmente en aplicaciones de alto riesgo como diagnóstico médico (Carter *et al.*, 2020; Higuera Carrillo, 2022).

Las barreras organizacionales incluyeron falta de experiencia en equidad algorítmica y resistencia al cambio. Investigaciones como las de Morgan *et al.* (2018) mostraron que menos del 30 % de las empresas tenían equipos dedicados a ética de IA, al limitar su capacidad de implementación.

Económicamente, los costos de mitigación resultaron prohibitivos para pymes y países en desarrollo, lo que exacerbó desigualdades tecnológicas (Yarborough, 2021; Hoyos Chavarro *et al.*, 2022;). Esto planteó cuestiones éticas sobre quién debía asumir estos costos.

Culturalmente, persistió una brecha entre percepciones técnicas y sociales de equidad. Mientras ingenieros priorizaban métricas cuantitativas, comunidades afectadas enfatizaban necesidades contextuales (Dobler *et al.*, 2018; Kimura *et al.*, 2021). Este desajuste subrayó la necesidad de mayor participación comunitaria en diseño algorítmico (Ledesma y Malave-González, 2022).

Discusión

Los hallazgos revelan una paradoja fundamental en el campo. Mientras las técnicas de detección de sesgos han alcanzado un notable nivel de sofisticación teórica, su aplicación práctica es fragmentada y con frecuencia inefectiva (Fernández-Castilla *et al.*, 2019; Langenkamp *et al.*, 2020). Esta brecha se explica en parte por la complejidad técnica de los métodos más avanzados, que requieren conocimientos especializados no siempre disponibles en contextos industriales (Heiden *et al.*, 2019; Lin y Chu, 2018). Además, la falta de consenso sobre qué métricas de equidad priorizar en diferentes dominios ha generado confusión entre desarrolladores y reguladores (Baros *et al.*, 2022).

La evolución de los métodos de mitigación muestra un patrón similar de promesas teóricas frente a limitaciones prácticas. Si bien técnicas como la corrección de sesgos por oposición entre modelos (*adversarial debiasing*) demuestran capacidad para reducir disparidades en entornos controlados, su transferencia a sistemas reales a menudo enfrenta resistencia por las compensaciones en rendimiento y costos computacionales (De Paolis Kaluza *et al.*, 2022). Este desafío se agrava en aplicaciones donde la precisión es crítica, como diagnóstico médico o evaluación crediticia, donde incluso pequeñas reducciones en exactitud pueden tener consecuencias significativas (Ringe y Ruof, 2020; Rus *et al.*, 2022). Curiosamente, las soluciones más adoptadas no son necesariamente las más avanzadas técnicamente, sino aquellas con mejor equilibrio entre efectividad y facilidad de implementación, como ciertas formas de posprocesamiento.

En el ámbito regulatorio, el análisis muestra una tensión irresuelta entre la necesidad de estándares globales y la importancia de adaptaciones locales. Los principios generales propuestos por organismos internacionales han sido cruciales para establecer un lenguaje común, pero su implementación concreta varía dramáticamente según contextos jurídicos y culturales (Kinavey y Cool, 2019; Mogrovejo Andrade, 2022). Los enfoques más prometedores parecen ser aquellos que combinan requisitos básicos universales con mecanismos flexibles de adaptación sectorial, como las zonas de prueba supervisadas (*regulatory sandboxes*) (Ringe y Ruof, 2020). Sin embargo, persiste el riesgo de que la regulación beneficie a grandes empresas con recursos para cumplir requisitos complejos, mientras margina a actores más pequeños, pero potencialmente innovadores (Balabin, 2019; Hoyos Chavarro *et al.*, 2022).

Finalmente, los resultados destacan que los desafíos más persistentes no son técnicos sino sociotécnicos. La efectividad de cualquier solución depende de factores que trascienden lo algorítmico: estructuras organizacionales, incentivos económicos, dinámicas de poder y comprensión cultural del concepto de equidad (Kimura *et al.*, 2021; Orozco Castillo, 2022). Esto sugiere en opinión de los autores que el futuro del campo requiere no solo mejores herramientas técnicas, sino marcos interdisciplinares que integren perspectivas de ciencias sociales, derecho y ética aplicada desde las primeras etapas de diseño.

Conclusiones

El análisis evidencia que la mitigación de sesgos en IA requiere superar el enfoque puramente técnico para adoptar una perspectiva integral que combine innovación algorítmica, gobernanza adaptativa y transformación organizacional. Si bien se han logrado avances significativos en detección y mitigación de sesgos, su efectividad aun es limitada por barreras institucionales, económicas y culturales que requieren soluciones sistémicas. El camino hacia sistemas de IA más éticos y equitativos demandará mayor colaboración interdisciplinar, estándares regulatorios balanceados y mecanismos concretos de auditoría y rendición de cuentas (*accountability*) que trasciendan los principios declarativos.

La experiencia de este período sugiere que ninguna solución aislada (técnica, regulatoria u organizacional) puede resolver por sí sola el complejo desafío de los sesgos algorítmicos. Futuras investigaciones deberán profundizar en enfoques holísticos que consideren desde el diseño mismo de los sistemas hasta su impacto social real, al reconocer que la equidad algorítmica es tanto un desafío tecnológico como un imperativo ético y social.

Referencias

Addey, C. (2021). Passports to the Global South, UN flags, favourite experts: understanding the interplay between UNESCO and the OECD within the SDG4 context. *Globalisation, Societies and Education, 19,* 593 - 604. https://doi.org/10.1080/14767724.2020.1862643

Akintola, B., Jagboro, G., Ojo, G., & Odediran, S. (2020). Effectiveness of Mechanisms for Enforcement of Ethical Standards in the Construction Industry. *Journal of Construction Business and Management, 4*(1), 1–12. https://doi.org/10.15641/JCBM.4.1.530

Auld, E., Rappleye, J., & Morris, P. (2018). PISA for Development: how the OECD and World Bank shaped education governance post-2015. *Comparative Education, 55*, 197 - 219. https://doi.org/10.1080/03050068.2018.1538635

Balabin, A. (2019). The Implementation of Corporate Governance Standards in Large Russian Companies. *Proceedings of the International Scientific Conference "Far East Con" (ISCFEC 2018)*. https://doi.org/10.2991/iscfec-18.2019.24

Baros, J., Sotola, V., Bilik, P., Martínek, R., Jaros, R., Danys, L., & Simoník, P. (2022). Review of Fundamental Active Current Extraction Techniques for SAPF. *Sensors (Basel, Switzerland), 22*. https://doi.org/10.3390/s22207985

Baykurt, B. (2022). Algorithmic *accountability* in U. S. cities: Transparency, impact, and political economy. *Big Data & Society, 9*. https://doi.org/10.1177/20539517221115426

Borges Machín, A. Y. y González Bravo, Y. L. (2022). Educación comunitaria para un envejecimiento activo: experiencia en construcción desde el autodesarrollo. *Región Científica, 1*(1), 202212. https://doi.org/10.58763/rc202213

Busuioc, M. (2020). Accountable Artificial Intelligence: Holding Algorithms to Account. *Public Administration Review, 81*, 825 - 836. https://doi.org/10.1111/puar.13293

Carter, E., Onyeador, I., & Lewis, N. (2020). Developing & delivering effective anti-bias training: Challenges & recommendations. *Behavioral Science & Policy, 6*, 57 - 70. https://doi.org/10.1177/237946152000600106

Chalmers, P. (2018). Model-Based Measures for Detecting and Quantifying Response Bias. *Psychometrika, 83*, 696 - 732. https://doi.org/10.1007/s11336-018-9626-9

Cruz, I., Troffaes, M., Lindström, J., & Sahlin, U. (2022). A robust Bayesian bias-adjusted random effects model for consideration of uncertainty about bias terms in evidence synthesis. *Statistics in Medicine, 41*, 3365 - 3379. https://doi.org/10.1002/sim.9422

Czarnowska, P., Vyas, Y., & Shah, K. (2021). Quantifying Social Biases in NLP: A Generalization and Empirical Comparison of Extrinsic *Fairness* Metrics. *Transactions of the Association for Computational Linguistics, 9,* 1249-1267. https://doi.org/10.1162/tacl_a_00425

De Paolis Kaluza, M., Jain, S., & Radivojac, P. (2022). An Approach to Identifying and Quantifying Bias in Biomedical Data. Pacific Symposium on Biocomputing. *Pacific Symposium on Biocomputing, 28*, 311 - 322. https://doi.org/10.1142/9789811270611_0029

Delobelle, P., Tokpo, E., Calders, T., & Berendt, B. (2022). Measuring *Fairness* with Biased Rulers: A Comparative Study on Bias Metrics for Pre-trained Language Models. , 1693-1706. https://doi.org/10.18653/v1/2022.naacl-main.122

Dobler, C. C., Morrow, A. S., & Kamath, C. C. (2019). Clinicians' cognitive biases: a potential barrier to implementation of evidence-based clinical practice. *BMJ evidence-based medicine, 24*(4), 137–140. https://doi.org/10.1136/bmjebm-2018-111074

Fernández-Castilla, B., Declercq, L., Jamshidi, L., Beretvas, S., Onghena, P., & Van Den Noortgate, W. (2019). Detecting Selection Bias in Meta-Analyses with Multiple Outcomes: A Simulation Study. *The Journal of Experimental Education, 89*, 125 - 144. https://doi.org/10.1080/00220973.2019.1582470

Floridi, L. (2019). Establishing the rules for building trustworthy AI. *Nature Machine Intelligence, 1*, 261-262. https://doi.org/10.1038/S42256-019-0055-Y

Goldfarb-Tarrant, S., Marchant, R., Sánchez, R., Pandya, M., & Lopez, A. (2020). Intrinsic Bias Metrics Do Not Correlate with Application Bias. *ArXiv*, abs/2012.15859. https://doi.org/10.18653/v1/2021.acl-long.150

Gómez Cano, C. A. (2022). Ingreso, permanencia y estrategias para el fomento de los Semilleros de Investigación en una IES de Colombia. *Región Científica, 1*(1), 20226. https://doi.org/10.58763/rc20226

Gómez Miranda, O. M. (2022). La franquicia: de la inversión al emprendimiento. *Región Científica, 1*(1), 20229. https://doi.org/10.58763/rc20229

Gómez-Cano, C. y Sánchez-Castillo, V. (2021). Evaluación del nivel de madurez en la gestión de proyectos de una empresa prestadora de servicios públicos. *Económicas CUC, 42*(2), 133-144. https://doi.org/10.17981/econcuc.42.2.2021.Org.7

Gray, C. (2022). Overcoming Political Fragmentation: The Potential of Meso-Level Mechanisms. *International Journal of Health Policy and Management, 12*. https://doi.org/10.34172/ijhpm.2022.7075

Guzmán, D. L., Gómez-Cano, C. y Sánchez-Castillo, V. (2022). Construcción del Estado a partir de la participación ciudadana. *Revista Academia & Derecho, 14*(25). https://doi.org/10.18041/2215-8944/academia.25.10601

Han, X., Baldwin, T., & Cohn, T. (2022). Towards Equal Opportunity *Fairness* through Adversarial Learning. *ArXiv,* abs/2203.06317. https://doi.org/10.48550/arXiv.2203.06317

Heiden, B., Tonino-Heiden, B., Obermüller, T., Loipold, C., & Wissounig, W. (2020). Rising from systemic to industrial artificial intelligence applications (AIA) for predictive decision making (PDM): Four examples. En Y. Bi, R. Bhatia & S. Kapoor (Eds.), *Intelligent systems and applications. IntelliSys 2019 (Advances in Intelligent Systems and Computing, 1038*, pp. 1222–1233). Springer. https://doi.org/10.1007/978-3-030-29513-4_94

Higuera Carrillo, E. L. (2022). Aspectos clave en agroproyectos con enfoque comercial: Una aproximación desde las concepciones epistemológicas sobre el problema rural agrario en Colombia. *Región Científica, 1*(1), 20224. https://doi.org/10.58763/rc20224

Hoyos Chavarro, Y. A., Melo Zamudio, J. C., & Sánchez Castillo, V. (2022). Sistematización de la experiencia de circuito corto de comercialización estudio de caso Tibasosa, Boyacá. *Región Científica, 1*(1), 20228. https://doi.org/10.58763/rc20228

Kelly, C., Karthikesalingam, A., Suleyman, M., Corrado, G., & King, D. (2019). Key challenges for delivering clinical impact with artificial intelligence. *BMC Medicine, 17*. https://doi.org/10.1186/s12916-019-1426-2

Kimura, A., Antón-Oldenburg, M., & Pinderhughes, E. (2021). Developing and Teaching an Anti-Bias Curriculum in a Public Elementary School: Leadership, K-1 Teachers', and Young Children's Experiences. *Journal of Research in Childhood Education, 36*, 183 - 202. https://doi.org/10.1080/02568543.2021.1912222

Kinavey, H., & Cool, C. (2019). The Broken Lens: How Anti-Fat Bias in Psychotherapy is Harming Our Clients and What To Do About It. *Women & Therapy, 42*, 116 - 130. https://doi.org/10.1080/02703149.2018.1524070

Langenkamp, M., Costa, A., & Cheung, C. (2020). Hiring Fairly in the Age of Algorithms. *ArXiv*, abs/2004.07132. https://doi.org/10.2139/ssrn.3723046

Ledesma, F. y Malave-González, B. E. (2022). Patrones de comunicación científica sobre E-commerce: un estudio bibliométrico en la base de datos Scopus. *Región Científica, 1*(1), 202214. https://doi.org/10.58763/rc202214

Lin, L., & Chu, H. (2018). Quantifying publication bias in meta-analysis. *Biometrics, 74*. https://doi.org/10.1111/biom.12817

Lyu, Y., Lu, H., Lee, M., Schmitt, G., & Lim, B. (2022). IF-City: Intelligible Fair City Planning to Measure, Explain and Mitigate Inequality. *IEEE Transactions on Visualization and Computer Graphics, 30*, 3749-3766. https://doi.org/10.1109/TVCG.2023.3239909

Madaio, M., Egede, L., Subramonyam, H., Vaughan, J., & Wallach, H. (2021). Assessing the *Fairness* of AI Systems: AI Practitioners' Processes, Challenges, and Needs for Support. *Proceedings of the ACM on Human-Computer Interaction, 6*, 1 - 26. https://doi.org/10.1145/3512899

Mazen, J., & Tong, X. (2020). Bias Correction for Replacement Samples in Longitudinal Research. *Multivariate Behavioral Research, 56*, 805 - 827. https://doi.org/10.1080/00273171.2020.1794774

McGregor, L., Murray, D., & Ng, V. (2019). International human rights law as a framework for algorithmic *accountability*. *International and Comparative Law Quarterly, 68*, 309 - 343. https://doi.org/10.1017/S0020589319000046

Mehrabi, N., Morstatter, F., Saxena, N., Lerman, K., & Galstyan, A. (2019). A Survey on Bias and *Fairness* in Machine Learning. *ACM Computing Surveys (CSUR), 54*, 1 - 35. https://doi.org/10.1145/3457607

Miroshnikov, A., Kotsiopoulos, K., Franks, R., & Kannan, A. (2020). Wasserstein-based *fairness* interpretability framework for machine learning models. *Machine Learning, 111*, 3307 - 3357. https://doi.org/10.1007/s10994-022-06213-9

Mogrovejo Andrade, J. M. (2022). Estrategias resilientes y mecanismos de las organizaciones para mitigar los efectos ocasionados por la pandemia a nivel internacional. *Región Científica, 1*(1), 202211. https://doi.org/10.58763/rc202211

Mökander, J., Juneja, P., Watson, D., & Floridi, L. (2022). The US *Algorithmic Accountability Act* of 2022 vs. The EU Artificial Intelligence Act: what can they learn from each other?. *Minds and Machines, 32*, 751 - 758. https://doi.org/10.1007/s11023-022-09612-y

Morgan, A., Chaiyachati, K., Weissman, G., & Liao, J. (2018). Eliminating Gender-Based Bias in Academic Medicine: More Than Naming the "Elephant in the Room". *Journal of General Internal Medicine, 33*, 966-968. https://doi.org/10.1007/s11606-018-4411-0

Ngxande, M., Tapamo, J., & Burke, M. (2019). Bias Remediation in Driver Drowsiness Detection Systems Using Generative Adversarial Networks. *IEEE Access, 8*, 55592-55601. https://doi.org/10.1109/ACCESS.2020.2981912

Ntoutsi, E., Fafalios, P., Gadiraju, U., Iosifidis, V., Nejdl, W., Vidal, M., Ruggieri, S., Turini, F., Papadopoulos, S., Krasanakis, E., Kompatsiaris, I., Kinder-Kurlanda, K., Wagner, C., Karimi, F., Fernández, M., Alani, H., Berendt, B., Kruegel, T., Heinze, C., Broelemann, K., Kasneci, G., Tiropanis, T., & Staab, S. (2020). Bias in data driven artificial intelligence systems—*An introductory survey*. *Wiley Interdisciplinary Reviews: Data Mining and Knowledge Discovery, 10*. https://doi.org/10.1002/widm.1356

Orozco Castillo, E. A. (2022). Experiencias en torno al emprendimiento femenino. *Región Científica, 1*(1), 20227. https://doi.org/10.58763/rc20225

Pérez Gamboa, A. J., García Acevedo, Y. y García Batán, J. (2019). Proyecto de vida y proceso formativo universitario: un estudio exploratorio en la Universidad de *Camagüey*. *Trasnsformación, 15*(3), 280-296. http://scielo.sld.cu/scielo.php?script=sci_arttext&pid=S2077-29552019000300280

Pérez-Gamboa, A. J., Gómez-Cano, C., & Sánchez-Castillo, V. (2022). Decision making in university contexts based on knowledge management systems. *Data & Metadata, 2*, 92. https://doi.org/10.56294/dm202292

Peters, U. (2022). Algorithmic Political Bias in Artificial Intelligence Systems. *Philosophy & Technology, 35*. https://doi.org/10.1007/s13347-022-00512-8

Petersen, J., Ranker, L., Barnard-Mayers, R., Maclehose, R., & Fox, M. (2021). A systematic review of quantitative bias analysis applied to epidemiological research. *International journal of epidemiology*. https://doi.org/10.1093/ije/dyab061

Petrenko, A. (2020). OECD acts as instruments of soft international law. *Law Review of Kyiv University of Law*. https://doi.org/10.36695/2219-5521.3.2020.74

Pospisil, D., & Bair, W. (2022). Accounting for Bias in the Estimation of r2 between Two Sets of Noisy Neural Responses. *The Journal of Neuroscience, 42*, 9343 - 9355. https://doi.org/10.1523/JNEUROSCI.0198-22.2022

Ricardo Jiménez, L. S. (2022). Dimensiones de emprendimiento: Relación educativa. El caso del programa cumbre. *Región Científica, 1*(1), 202210. https://doi.org/10.58763/rc202210

Ringe, W., & Ruof, C. (2020). Regulating Fintech in the EU: the Case for a Guided Sandbox. *European Journal of Risk Regulation, 11*, 604 - 629. https://doi.org/10.1017/err.2020.8

Rodríguez-Torres, E., Gómez-Cano, C., & Sánchez-Castillo, V. (2022). Management information systems and their impact on business decision making. *Data & Metadata, 1*, 21. https://doi.org/10.56294/dm202221

Royal, K. (2019). Survey research methods: A guide for creating post-stratification weights to correct for sample bias. *Education in the Health Professions, 2*, 48 - 50. https://doi.org/10.4103/EHP.EHP_8_19

Rus, C., Luppes, J., Oosterhuis, H., & Schoenmacker, G. (2022). Closing the Gender Wage Gap: Adversarial *Fairness* in Job Recommendation. *ArXiv*, abs/2209.09592. https://doi.org/10.48550/arXiv.2209.09592

Sanabria Martínez, M. J. (2022). Construir nuevos espacios sostenibles respetando la diversidad cultural desde el nivel local. *Región Científica, 1*(1), 20222. https://doi.org/10.58763/rc20222

Shen, A., Han, X., Cohn, T., Baldwin, T., & Frermann, L. (2022). Does Representational *Fairness* Imply Empirical *Fairness*? 81-95. https://doi.org/10.18653/v1/2022.findings-aacl.8

Sherman, L., Cantor, A., Milman, A., & Kiparsky, M. (2020). Examining the complex relationship between innovation and regulation through a survey of wastewater utility managers. *Journal of environmental management, 260*, 110025. https://doi.org/10.1016/j.jenvman.2019.110025

Simpson, A., & Dervin, F. (2019). Global and intercultural competences for whom? By whom? For what purpose?: an example from the Asia Society and the OECD. *Compare: A Journal of Comparative and International Education, 49*, 672 - 677. https://doi.org/10.1080/03057925.2019.1586194

Thompson, J. (2021). Mental Models and Interpretability in AI *Fairness* Tools and Code Environments. In: Stephanidis, C., *et al*. HCI International 2021 - Late Breaking Papers: Multimodality, eXtended Reality, and Artificial Intelligence. HCII 2021. *Lecture Notes in Computer Science, vol 13095*. Springer, Cham. https://doi.org/10.1007/978-3-030-90963-5_43

Vaccari, V., & Gardinier, M. (2019). Toward one world or many? A comparative analysis of OECD and UNESCO global education policy documents. *International Journal of Development Education and Global Learning*. https://doi.org/10.18546/IJDEGL.11.1.05

Wexler, J., Pushkarna, M., Bolukbasi, T., Wattenberg, M., Viégas, F., & Wilson, J. (2019). The What-If Tool: Interactive Probing of Machine Learning Models. *IEEE Transactions on Visualization and Computer Graphics, 26*, 56-65. https://doi.org/10.1109/TVCG.2019.2934619

Yam, J., & Skorburg, J. (2021). From human resources to human rights: Impact assessments for hiring algorithms. *Ethics and Information Technology, 23*, 611 - 623. https://doi.org/10.1007/s10676-021-09599-7

Yarborough, M. (2021). Moving towards less biased research. *BMJ Open Science, 5*. https://doi.org/10.1136/bmjos-2020-100116

Zahid, A., Khan, M., Khan, A., Kamiran, F., & Nasir, B. (2020). Modeling, Quantifying and Visualizing Media Bias on Twitter. *IEEE Access, 8*, 81812-81821. https://doi.org/10.1109/ACCESS.2020.2990800

Zapp, M. (2020). The authority of science and the legitimacy of international organisations: OECD, UNESCO and World Bank in global education governance. *Compare: A Journal of Comparative and International Education, 51*, 1022 - 1041. https://doi.org/10.1080/03057925.2019.1702503