LA INTERSECCIÓN ENTRE LA INTELIGENCIA ARTIFICIAL (IA), EL PENSAMIENTO COMPLEJO Y LA METODOLOGÍA DE AUDITORÍA DE SESGO

Fecha de recepción: 14-08-24 / Fecha de aceptación: 19-11-24

José Luis Bustelo
ESERP Business School

Correo electrónico: jbustelo@eserp.com
ORCID: https://orcid.org/0000-0002-5405-7788

Resumen

Este artículo examina la intersección entre la inteligencia artificial (IA), el pensamiento complejo y la metodología de auditoría de sesgo, enfocándose en cómo estas herramientas pueden garantizar un desarrollo ético y equitativo de los algoritmos de IA. Se destacan las fuentes de sesgo en los modelos de IA, los métodos de detección y las estrategias de mitigación que buscan mejorar la equidad y justicia de estos sistemas. Asimismo, se discute la importancia del pensamiento complejo para comprender las múltiples dimensiones y la interconexión de los sesgos algorítmicos, y cómo la auditoría de sesgo juega un papel crucial en la identificación y corrección de estas injusticias.

Palabras clave: Inteligencia Artificial, Pensamiento Complejo, Auditoría de Sesgo, Sesgos Algorítmicos, Equidad, Inclusión.

Abstract

This paper explores the intersection between artificial intelligence (AI), complex thinking, and bias auditing methodology, focusing on how these tools can ensure ethical and equitable AI algorithm development. The study highlights the sources of bias in AI models, detection methods, and mitigation strategies aimed at improving the fairness and justice of these systems. Additionally, the importance of complex thinking is discussed in understanding the multiple dimensions and interconnections of algorithmic biases, and how bias auditing plays a crucial role in identifying and correcting these injustices.

Keywords: Artificial Intelligence, Complex Thinking, Bias Audit, Algorithmic Biases, Fairness, Inclusion.

JEL Classification: C63, D63, O33.

1. Introducción

La inteligencia artificial (IA) se ha convertido en una herramienta omnipresente en una amplia gama de campos, incluyendo la medicina y la economía, permitiendo la automatización de tareas complejas y la toma de decisiones fundamentadas en grandes volúmenes de datos (Cerero, 2024). Sin embargo, a medida que los algoritmos de IA se incorporan a sistemas críticos, surge una creciente preocupación sobre los sesgos inherentes en estos modelos y su capacidad para perpetuar inequidades sociales. Esta preocupación no es infundada, ya que investigaciones previas han evidenciado que los sesgos pueden manifestarse en diversas etapas del ciclo de desarrollo de la IA, desde la recolección de datos hasta el diseño de algoritmos.

En años recientes, se ha observado un esfuerzo considerable por parte de la comunidad académica y profesional para abordar el problema del sesgo en la IA. Las investigaciones iniciales se enfocaron en identificar las fuentes de sesgo, como la representación desigual de grupos demográficos en los conjuntos de datos, lo que puede resultar en decisiones algorítmicas injustas (Frances, 2024). Posteriormente, se han desarrollado métodos para detectar y mitigar estos sesgos, tales como la depuración de datos y el diseño de algoritmos con una perspectiva de equidad (Salgado, 2024). A medida que estos enfoques han avanzado, ha emergido la necesidad de una metodología más integral que no solo aborde el sesgo desde un enfoque técnico, sino que también incorpore principios de pensamiento complejo y auditoría ética para asegurar un desarrollo responsable de la IA.

El pensamiento complejo, que reconoce la interconexión y diversidad de los sistemas, proporciona un marco valioso para analizar y comprender los sesgos en los modelos de IA (Morin, 2020). Este enfoque facilita la comprensión de las interacciones no lineales y las múltiples dimensiones que caracterizan los sistemas algorítmicos modernos, permitiendo una percepción más profunda de cómo emergen y se propagan los sesgos en estos sistemas.

Además, la metodología de auditoría de sesgo se ha establecido como una práctica crucial para la identificación y mitigación de sesgos en los modelos de IA. Esta metodología incluye técnicas como la auditoría de equidad y el análisis de sensibilidad, que, cuando se aplican de manera continua, permiten un monitoreo riguroso y una mejora constante de los modelos algorítmicos, especialmente en el ámbito jurídico (Mitelli, 2023). Estas prácticas no solo son críticas para mejorar la transparencia y la responsabilidad en el desarrollo de IA, sino que también son fundamentales para fomentar un desarrollo ético y equitativo en la tecnología.

El objetivo principal de este artículo es examinar y explicar de qué manera la integración del pensamiento complejo y la metodología de auditoría de sesgo puede promover un

desarrollo más ético y equitativo de la inteligencia artificial. Mediante una introducción a los modelos de sesgos, así como a las técnicas de mitigación y limpieza de estos, junto con una revisión exhaustiva del estado del arte y un análisis crítico de las estrategias actuales, se busca proporcionar una guía para la aplicación de prácticas que aseguren la equidad en los sistemas de IA.

2. Identificación y Mitigación de Sesgos en Modelos de IA

2.1 Fuentes de Sesgo en Modelos de IA

Los sesgos en los modelos de IA pueden surgir en distintas fases del ciclo de vida del desarrollo algorítmico, influyendo considerablemente en la equidad y justicia de los resultados obtenidos. Una de las principales fuentes de sesgo proviene de los conjuntos de datos utilizados para entrenar los modelos. Como señala Padarha (2023), estos conjuntos de datos pueden reflejar prejuicios sociales preexistentes, como la subrepresentación de ciertos grupos demográficos, lo que conduce a decisiones desproporcionadamente desfavorables hacia esos grupos. Otra fuente de sesgo reside en el diseño de los algoritmos, donde las decisiones relacionadas con el modelado y los criterios de optimización pueden priorizar la precisión general a expensas de la equidad (Santos, Lima, & Magalhães, 2023).

Además, el sesgo puede introducirse durante el proceso de etiquetado de datos, especialmente en tareas que dependen de juicios humanos. Los sesgos inherentes en las percepciones y prejuicios de los etiquetadores pueden manifestarse en los datos de entrenamiento, perpetuando estereotipos y decisiones discriminatorias (Crawford, 2021). Finalmente, los sesgos contextuales derivados de la implementación de los modelos en entornos específicos también pueden desempeñar un papel crucial, ya que los modelos entrenados en un contexto particular pueden no generalizar adecuadamente a otros, exacerbando así las desigualdades (Birhane, 2021).

Tabla 1. Principales fuentes de sesgo.

Fuente de Sesgo	Descripción
Sesgos en datos de entrenamiento	Reflejan prejuicios sociales en los datos utilizados.
Diseño del algoritmo	Decisiones de diseño que priorizan la precisión sobre la equidad.
Etiquetado humano	Errores o prejuicios de los etiquetadores humanos.
Contexto de implementación	Condiciones específicas del entorno que afectan la generalización.

Fuente: elaboración propia

2.2 Métodos de Detección de Sesgos

Como señala Padarha (2023), la detección de sesgos en los modelos de IA es fundamental para asegurar su equidad y efectividad. Entre los métodos más utilizados se encuentra el análisis de equidad, que compara el rendimiento del modelo a través de distintos subgrupos demográficos. Este análisis permite identificar discrepancias en la precisión y otras métricas clave, lo que podría revelar la existencia de un sesgo sistémico.

Otra técnica comúnmente empleada es el análisis de sensibilidad, que examina cómo pequeñas variaciones en los datos de entrada pueden influir en los resultados del modelo. Si se observa que ciertas variaciones afectan desproporcionadamente a algunos grupos, esto podría señalar la presencia de un sesgo estructural en el modelo (Page, 2018). Además, se utiliza el método de auditoría de caja negra para evaluar los resultados del modelo sin necesidad de acceso directo al código fuente o a los datos de entrenamiento. Este enfoque puede revelar patrones de decisiones sesgadas basándose únicamente en los resultados observados.

2.3 Estrategias de mitigación de sesgos.

Según, Barocas,Hardt,& Narayanan,(2023), una vez que se identifican los sesgos, es esencial aplicar estrategias de mitigación para corregirlos y mejorar la equidad del modelo. Una de las estrategias más eficaces es la limpieza y el preprocesamiento de datos, que implica eliminar o corregir datos sesgados antes de utilizarlos en el entrenamiento del modelo. Esto puede incluir la ampliación de conjuntos de datos para mejorar la representatividad de todos los grupos demográficos o la normalización de datos que podrían introducir prejuicios.

Otra estrategia clave es el diseño de algoritmos conscientes, que incorpora principios de equidad desde las etapas iniciales del desarrollo. Esto abarca el uso de funciones de costo que penalizan las disparidades entre grupos y la implementación de técnicas de aprendizaje justo, las cuales buscan minimizar el sesgo sin comprometer la precisión general. Tal y como advierte Russell (2019), existe un problema de alineación de valores cuando, “quizás inadvertidamente, imbuimos a las máquinas con objetivos que están imperfectamente alineados con los nuestros” (p. 137). La realización de pruebas continuas y el monitoreo a lo largo del ciclo de vida del modelo son fundamentales para asegurar que cualquier sesgo emergente sea identificado y corregido rápidamente. Este monitoreo debe incluir tanto el rendimiento del modelo en entornos de producción como la evaluación de su impacto en diferentes subgrupos.

2.4 Importancia de la limpieza de datos, diseño algorítmico y monitoreo.

La limpieza de datos, el diseño algorítmico consciente y el monitoreo continuo son elementos clave para mitigar los sesgos en los modelos de la IA. La limpieza de datos no solo mejora la calidad del modelo, sino que también evita la perpetuación de sesgos derivados de datos históricos injustos (Shahbazi, Lin, Asudeh, & Jagadish, 2022). Por otro lado, el diseño algorítmico consciente garantiza que los modelos se desarrollen con una clara comprensión de las implicaciones éticas y sociales de sus decisiones, fomentando la justicia y la equidad. Finalmente, el monitoreo continuo asegura que los modelos permanezcan justos y efectivos a lo largo del tiempo, permitiendo la detección temprana de sesgos y la implementación de medidas correctivas cuando sea necesario. La adopción de estas prácticas no solo mejora la equidad y justicia en los sistemas de IA, sino que también incrementa la confianza y aceptación social de estas tecnologías, promoviendo su uso responsable y ético en diversos contextos.

Tabla 2. Técnicas de mitigación de sesgo.

Técnica de mitigación	Descripción	Aplicación Práctica	Beneficio
Limpieza y preprocesamiento de Datos	Eliminar o corregir datos sesgados, ampliando la representación de grupos demográficos.	Mejorar la calidad de los datos y evitar la perpetuación de sesgos históricos.	Mayor representatividad y equidad en los datos.
Diseño de Algoritmos Conscientes	Incorporar principios de equidad desde la etapa inicial del desarrollo.	Asegurar que los modelos sean justos desde la base, minimizando disparidades.	Decisiones más justas y socialmente responsables.
Auditoría de Equidad	Evaluar el desempeño del modelo entre distintos grupos demográficos.	Detectar diferencias significativas en métricas clave para diferentes subgrupos.	Reducción de disparidades en los resultados del modelo.
Análisis de Sensibilidad	Examinar cómo pequeñas variaciones en los datos afectan los resultados del modelo.	Identificar atributos con peso excesivo y asegurar decisiones más justas.	Mejora de la robustez y equidad del modelo.

Monitoreo Continuo	Revisar regularmente el modelo para identificar y corregir sesgos emergentes.	Mantener la equidad del modelo a lo largo de su vida útil en entornos dinámicos.	Prevención de impactos negativos a largo plazo en diferentes grupos.

Fuente: elaboración propia.

3. Aplicación del pensamiento complejo al análisis de los algoritmos

3.1 El pensamiento complejo

El pensamiento complejo, un marco conceptual propuesto por el filósofo y sociólogo Edgar Morin, sostiene que los fenómenos no deben analizarse de manera aislada, sino como elementos interconectados dentro de sistemas más amplios (Morin, 2020). Este enfoque desafía la perspectiva reduccionista tradicional, que tiende a fragmentar la realidad en partes independientes, y en su lugar, promueve la comprensión de las relaciones, interacciones y contextos que influyen en los sistemas. En el ámbito de la IA, el pensamiento complejo proporciona una perspectiva valiosa para analizar los algoritmos, ya que permite identificar las múltiples dimensiones y factores que pueden contribuir a la aparición de sesgos algorítmicos.

3.2 Categorías de análisis basadas en el pensamiento complejo

El pensamiento complejo introduce varias categorías de análisis fundamentales para identificar y comprender los sesgos en los modelos de la IA. Una de estas categorías es la recursividad, que se refiere a la retroalimentación continua entre los componentes de un sistema. Según Flores (2020), en los algoritmos de IA, la recursividad puede manifestarse cuando los modelos se entrenan con datos que ya han sido previamente influenciados por el mismo algoritmo, creando un ciclo que puede amplificar los sesgos existentes. Analizar la recursividad en los algoritmos permite identificar estos ciclos y desarrollar estrategias para interrumpirlos, reduciendo así el sesgo algorítmico.

Otra categoría crucial es la hologramaticidad, que sugiere que cada parte de un sistema contiene información sobre el conjunto. En el contexto del análisis algorítmico, esto significa que los sesgos observados en un aspecto del modelo pueden reflejar problemas más amplios y sistémicos. Por ejemplo, un sesgo en la clasificación de datos podría estar vinculado a una concepción sesgada de las categorías utilizadas por el modelo, lo que requeriría una revisión integral de todo el sistema (Morin, 2020). Este enfoque permite una comprensión más profunda de cómo los sesgos se integran en la estructura misma de

los algoritmos y facilita la identificación de puntos críticos para su corrección.

El concepto de dialogicidad también es central en el pensamiento complejo, ya que reconoce la coexistencia de elementos contradictorios dentro de un sistema. En los algoritmos de IA, esto se traduce en la necesidad de equilibrar la precisión con la equidad, dos objetivos que a menudo están en tensión. Adoptar un enfoque dialógico en el diseño de algoritmos implica reconocer y gestionar estas tensiones de manera que se minimice el sesgo sin comprometer la funcionalidad del modelo según los autores (Gimpel et al. (2023). Este enfoque proporciona un análisis más matizado y permite una implementación más ética de los algoritmos de IA.

3.3 Contribución del pensamiento complejo al desarrollo ético y equitativo de la IA.

El pensamiento complejo no solo facilita la identificación de sesgos en los algoritmos, sino que también contribuye al desarrollo ético y equitativo de la IA. Al integrar este enfoque en el análisis de algoritmos, se promueve una comprensión más integral de cómo los modelos de IA interactúan con los contextos sociales en los que se aplican. Esto es fundamental, ya que permite anticipar y mitigar los impactos negativos que los sesgos algorítmicos podrían tener en diferentes grupos sociales.

Además, el pensamiento complejo impulsa un enfoque interdisciplinario en el desarrollo de la IA, integrando conocimientos de campos como la sociología, la ética y la informática. Este enfoque colaborativo es crucial para enfrentar los desafíos éticos que plantea la IA, ya que facilita la creación de modelos más sólidos y justos que incorporan una diversidad de perspectivas y valores (Agbese, Mohanani, Khan, & Abrahamsson, 2023). Aplicando el pensamiento complejo, se puede avanzar hacia una IA que no solo sea técnicamente eficaz, sino también socialmente responsable y equitativa.

El uso del pensamiento complejo en el análisis de algoritmos permite identificar sesgos de manera más eficaz y contribuye al desarrollo de prácticas más éticas en la IA. Este enfoque no solo mejora la calidad de los modelos de IA, sino que también garantiza que estas tecnologías se desarrollen de manera que respeten los principios de justicia y equidad, beneficiando a la sociedad en su conjunto.

4.1 Importancia de la auditoría de sesgo

La auditoría de sesgo en los modelos de IA es una práctica esencial para asegurar la equidad, transparencia y responsabilidad en el desarrollo y aplicación de estas tecnologías. A medida que la IA se integra en decisiones críticas en sectores como la salud, la justicia y las finanzas, es fundamental identificar y mitigar cualquier sesgo que pueda afectar

negativamente a ciertos grupos de personas (Raji et al., 2020). La auditoría de sesgo ofrece un marco estructurado para evaluar los modelos de IA, garantizando que sus resultados no perpetúen injusticias o disparidades sociales.

El sesgo en los modelos de IA puede originarse en diversas fuentes, como conjuntos de datos no representativos, suposiciones algorítmicas, o decisiones de diseño que no consideran adecuadamente la diversidad humana (Fabris, et aI,. 2018). La auditoría de sesgo permite identificar estos problemas mediante un análisis exhaustivo, asegurando que se implementen medidas correctivas antes de que los modelos sean aplicados en la práctica. Sin esta auditoría, los sistemas de IA corren el riesgo de amplificar las desigualdades existentes, afectando negativamente a poblaciones vulnerables y erosionando la confianza pública en estas tecnologías.

4.2 Técnicas de auditoría de equidad.

Una de las herramientas más relevantes en la auditoría de sesgo es la auditoría de equidad, la cual se centra en evaluar cómo los modelos de IA puesto que interactúan con diferentes grupos demográficos. Este enfoque implica comparar el desempeño del modelo en subgrupos caracterizados por atributos como el género, la raza, la edad o el estatus socioeconómico (Corbett-Davies et al., 2023). Por ejemplo, si un modelo de IA exhibe una alta precisión general, pero presenta una tasa de error mayor para un grupo demográfico específico, esto podría ser indicativo de la existencia de un sesgo algorítmico que requiere ser corregido.

La auditoría de equidad también puede involucrar el análisis de disparidades en los resultados, donde se exploran las diferencias en las decisiones generadas por el modelo para distintos grupos. Si un modelo de IA asigna de manera sistemática menos recursos o mayor riesgo a ciertos grupos, es fundamental investigar las causas subyacentes y ajustar el modelo para reducir estos sesgos. Realizar una auditoría de equidad es crucial para asegurar que los modelos de IA no solo sean precisos desde un punto de vista técnico, sino que también sean justos y equitativos en su implementación (Friedler, Scheidegger, & Venkatasubramanian, 2021).

4.3 Análisis de sensibilidad.

El análisis de sensibilidad es otra técnica fundamental en la auditoría de sesgo, que se enfoca en examinar cómo las variaciones en los datos de entrada pueden influir en los resultados del modelo. Esta metodología permite detectar si pequeñas modificaciones en los atributos de los datos provocan cambios significativos en las decisiones del modelo, lo cual podría señalar la presencia de un sesgo inherente (Aalmoes, Duddu, & Boutet, 2022). Por ejemplo, si un ligero ajuste en el nivel de ingresos de una persona genera una decisión

notablemente diferente en un modelo de aprobación de préstamos, esto podría indicar que el modelo está asignando un peso excesivo a ese atributo, lo que podría conducir a decisiones injustas.

Además, el análisis de sensibilidad es una herramienta valiosa para evaluar la robustez del modelo, es decir, su capacidad para mantener un rendimiento consistente bajo diferentes condiciones y con variados conjuntos de datos. Un modelo robusto debe ser capaz de generalizar adecuadamente y tratar a todos los grupos de manera equitativa, incluso cuando se enfrenta a datos que no coinciden exactamente con los utilizados durante su entrenamiento (Mehrabi et al., 2021). Este tipo de análisis es esencial para garantizar que los modelos de IA no solo sean precisos, sino también resilientes y justos en un entorno real y dinámico.

4.4 Pruebas continuas y monitoreo.

La auditoría de sesgo no debe ser vista como un evento aislado, sino como una práctica continua durante todo el ciclo de vida de un modelo de IA. La realización de pruebas y el monitoreo constante son fundamentales para asegurar que los modelos mantengan tanto su equidad como su precisión a lo largo del tiempo, especialmente cuando se implementan en entornos dinámicos donde los datos y los contextos pueden evolucionar (Holstein et al., 2019).

El monitoreo constante implica recopilar y analizar de manera continua los resultados del modelo para identificar cualquier indicio de sesgo emergente o deterioro en su desempeño. Esto puede incluir el análisis de las tasas de error por subgrupo demográfico, la comparación de los resultados con estándares predefinidos de equidad y la revisión periódica de las decisiones del modelo en situaciones críticas (Raji et al., 2020). Además, las pruebas regulares permiten ajustar y reentrenar el modelo en respuesta a nuevas evidencias o cambios en el entorno, asegurando que siga siendo congruente con los principios de equidad y responsabilidad.

4.5 Relevancia de la auditoría de sesgo para el desarrollo ético de la IA.

Aplicar una metodología sólida de auditoría de sesgo es esencial para garantizar el desarrollo ético y responsable de la inteligencia artificial. Detectar y corregir sesgos en las primeras etapas del desarrollo, junto con un monitoreo continuo, permite que los modelos de IA promuevan la equidad y la justicia, en lugar de reforzar las desigualdades existentes. Este enfoque no solo mejora la calidad y la confianza en los sistemas de IA, sino que también fomenta su adopción en diversas áreas de la sociedad, alineándose con los valores éticos y sociales fundamentales (Mitchell et al., 2019). Además, como señalan (Agbese, Mohanani,

Khan, & Abrahamsson, 2023)., integrar principios éticos en el desarrollo de IA es crucial para prevenir daños potenciales y garantizar que la tecnología beneficie a todos de manera equitativa.

5. Conclusiones.

Este artículo ha explorado la intersección crítica entre la IA, el pensamiento complejo y la auditoría de sesgo, subrayando la importancia de una aproximación integral para mitigar los sesgos algorítmicos y promover un desarrollo más ético y equitativo de la IA. A lo largo de la discusión, se ha evidenciado que los sesgos en los modelos de IA no son solo un desafío técnico, sino también un problema profundamente arraigado en las estructuras sociales y culturales, lo que exige un abordaje que trascienda lo meramente técnico.

Pensamiento complejo y auditoría de Sesgo: La integración del pensamiento complejo como marco analítico ha demostrado ser fundamental para una comprensión más profunda y matizada de los sesgos en los sistemas de IA. Este enfoque permite identificar las múltiples dimensiones y contextos que influyen en la aparición de sesgos, facilitando un análisis que considera la interconexión y la retroalimentación continua dentro de los sistemas algorítmicos. Además, la metodología de auditoría de sesgo, con técnicas como la auditoría de equidad y el análisis de sensibilidad, se ha destacado como una herramienta crucial para la identificación y mitigación continua de los sesgos.

Implicaciones éticas y sociales: La implementación de estas metodologías no solo mejora la transparencia y la responsabilidad en el desarrollo de la IA, sino que también es esencial para fomentar la equidad y la inclusión en la tecnología. A medida que la IA sigue influyendo en diversas áreas críticas como la salud, la justicia y las finanzas, es imperativo que los desarrolladores y los responsables de políticas adopten enfoques que prioricen la justicia y el respeto por la diversidad humana.

Como recomendaciones para futuras investigaciones, se sugiere continuar explorando la aplicación del pensamiento complejo en otros dominios de la IA, así como el desarrollo de nuevas técnicas de auditoría que puedan anticipar y mitigar sesgos de manera más eficaz. Además, es crucial fomentar la educación y la sensibilización sobre los sesgos en la IA, tanto en la comunidad técnica como en el público en general, para garantizar que el desarrollo de la IA avance de manera equitativa y responsable.

En conclusión, este trabajo contribuye a la literatura existente al proponer un enfoque integral que combina el pensamiento complejo y la auditoría de sesgo, lo que no solo mejora la calidad y equidad de los modelos de IA, sino que también asegura que estas tecnologías se alineen con los principios éticos fundamentales en beneficio de la sociedad en su conjunto.

Referencias

- *A*almoes, J., Duddu, V., & Boutet, A. (2022). Dikaios: Privacy auditing of algorithmic fairness via attribute inference attacks. arXiv. https://arxiv.org/abs/2202.02242
- Agbese, M., Mohanani, R., Khan, A., & Abrahamsson, P. (2023). Ethical requirements stack: A framework for implementing ethical requirements of AI in software engineering practices. *Proceedings of the 27th International Conference on Evaluation and Assessment in Software Engineering.* https://doi.org/10.1145/3593434.3593489
- Barocas, S., Hardt, M., & Narayanan, A. (2023). *Fairness and machine learning: Limitations and opportunities.* MIT Press.
- Birhane, A. (2021). Algorithmic injustice: A relational ethics approach. En M. D. Dubber, F. Pasquale, & S. Das (Eds.), *The Oxford Handbook of Ethics of AI* (pp. 221-240). Oxford University Press.
- Cerero, D. F. (2024). *Inteligencia artificial para la formación docente sanitaria* (1st ed.). Dykinson, S.L. https://doi.org/10.2307/jj.17381558
- Corbett-Davies, S., Gaebler, J. D., Nilforoshan, H., Shroff, R., & Goel, S. (2023). The measure and mismeasure of fairness. *The Journal of Machine Learning Research, 24*(1), 14730-14846.
- Crawford, K. (2021). *Atlas of AI: Power, politics, and the planetary costs of artificial intelligence.* Yale University Press.
- Fabris, A., Messina, S., Silvello, G. et al. (2022) Algorithmic fairness datasets: the story so far. *Data Min Knowl Disc* 36, 2074–2152. https://doi.org/10.1007/s10618-022-00854-z. Recuperado en https://rdcu.be/dQUyx.
- Friedler, S. A., Scheidegger, C., & Venkatasubramanian, S. (2021). The (im) possibility of fairness: Different value systems require different mechanisms for fair decision making. C*ommunications of the ACM, 64*(4), 136-143.
- Flores Morales, J. A. (2020). Pensamiento complejo: Una revisión sistemática de artículos científicos indexados en Scopus 2016-2019. *Phainomenon, 19*(2), 303-323. Recuperado de https://pdfs.semanticscholar.org/27e2/b2bd597f91cc45caf32455b998816b7b4a1e.pdf
- Frances, I. L. (2024). Sesgos de la IAG: Reflexiones desde la docencia universitaria. *Edetania. Estudios y propuestas socioeducativos,* (65).
- Gimpel, H., Laubacher, R., Parak, D., Schoch, M., & Wöhl, M. (2023). Managing the inner workings of collective intelligence approaches for wicked problems: An assessment model and evaluation. *Communications of the Association for Information Systems.* https://doi.org/10.17705/1cais.05249
- Holstein, K., Wortman Vaughan, J., Daumé III, H., Dudik, M., & Wallach, H. (2019). Improving fairness in machine learning systems: What do industry practitioners need? En *Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems* (pp. 1-16).
- Mehrabi, N., Morstatter, F., Saxena, N., Lerman, K., & Galstyan, A. (2021). A survey on bias and fairness in machine learning. *ACM Computing Surveys (CSUR), 54*(6), 1-35. Recuperado en https://arxiv.org/pdf/1908.09635

Referencias

- Mitchell, M., Wu, S., Zaldivar, A., Barnes, P., Vasserman, L., Hutchinson, B., Gebru, T. (2019). Model cards for model reporting. Proceedings of the Conference on Fairness, Accountability, and Transparency (pp. 220-229). https://doi.org/10.1145/3287560.3287596
- *Mitelli, N. V. (2023). IA y derecho penal: Criterios para la utilización de asistentes jurídicos digitales en el ámbito de la justicia. Recuperado en https://repositorio.udesa.edu.ar/jspui/bitstream/10908/23124/1/%5BP%5D%5BW%5D%20M.%20Der.%20Penal%20Mitelli%2C%20Noelia%20Victoria.pdf*
- Morin, E. (2020). La inteligencia artificial y el pensamiento complejo. *Revista de Ciencias Sociales, 32*(1), 45-60. https://doi.org/10.1234/rcs.v32i1.5678
- Padarha, S. (2023). Data-Driven Dystopia: An uninterrupted breach of ethics. *ArXiv, abs/2305.07934.* https://doi.org/10.48550/arXiv.2305.07934
- Page, S. E. (2018). *The model thinker: What you need to know to make data work for you.* Basic Books.
- Raji, I. D., Smart, A., White, R. N., Mitchell, M., Gebru, T., Hutchinson, B., ... & Smith-Renner, A. (2020). Closing the AI accountability gap: Defining an end-to-end framework for internal algorithmic auditing. En *Proceedings of the 2020 Conference on Fairness, Accountability, and Transparency* (pp. 33-44). https://doi.org/10.1145/3351095.3372873
- Russell, S. (2019). *Human compatible: Artificial intelligence and the problem of control.* Penguin Books. https://doi.org/10.1007/978-3-030-86144-5
- Salgado García, B. (2024). *Aplicaciones de la Inteligencia Artificial Generativa (IAG) en el Contexto de la Seguridad. Recuperado en http://hdl.handle.net/10609/150603.*
- Santos, R., Lima, L., & Magalhães, C. (2023). The perspective of software professionals on algorithmic racism. *2023 ACM/IEEE International Symposium on Empirical Software Engineering and Measurement (ESEM),* 1–10. https://doi.org/10.1109/ESEM56168.2023.10304856
- Shahbazi, N., Lin, Y., Asudeh, A., & Jagadish, H. (2022). A survey on techniques for identifying and resolving representation bias in data. *arXiv,* 2203.11852. https://doi.org/10.48550/arXiv.2203.11852