

Sesgos Algorítmicos y Equidad en el Reclutamiento por I.A.: Propuesta de un Marco Ético de Auditoría para la Experiencia del Candidato (EX)

Algorithmic Bias and Equity in AI Recruitment: Proposal for an Ethical Audit Framework for the Candidate Experience (EX)

José Flores¹, Marisol Miranda², Yamileth Wilson³, Paul Pérez⁴, Jorge Tejada⁵ y Frainith Rincón⁶

¹Universidad de Panamá, jose.flores@up.ac.pa, <https://orcid.org/0009-0000-5798-958X>, Panamá

²Universidad de Panamá, marisol.miranda@up.ac.pa, <https://orcid.org/0009-0004-6571-4116>, Panamá

³Universidad de Panamá, yamileth.wilson@up.ac.pa, <https://orcid.org/0009-0001-1616-1306>, Panamá

⁴Universidad de Panamá, paul.perez@up.ac.pa, <https://orcid.org/0009-0002-0931-7496>, Panamá

⁵Universidad de Panamá, jorge.tejadam@up.ac.pa, <https://orcid.org/0009-0005-4809-360X>, Panamá

⁶Universidad de Panamá, frainith.rincon@up.ac.pa, <https://orcid.org/0009-0000-3012-6155>, Panamá.

Información del Artículo

Trazabilidad:

Recibido 10-11-2025

Revisado 11-11-2025

Aceptado 14-12-2025

Palabras Clave:

Sesgo Algorítmico

I.A. en RR.HH

Ética de la I.A

Reclutamiento

Experiencia del Candidato

RESUMEN

La adopción de la Inteligencia Artificial (I.A.) en el reclutamiento promete eficiencia y reducción de costos, pero introduce un riesgo crítico: la perpetuación y escalamiento de sesgos algorítmicos heredados de datos históricos discriminatorios. Este artículo aborda la necesidad de garantizar la equidad y la transparencia en la selección de personal a través de sistemas de I.A., con un impacto directo en la Experiencia del Candidato (EX). Se propone el diseño de un Marco Ético de Auditoría (MEA) basado en una revisión sistemática de principios de I.A. confiable. El objetivo principal es ofrecer a los profesionales de Recursos Humanos y de *Compliance* una herramienta conceptual de cuatro pilares para evaluar la imparcialidad del *software* de I.A. en el ciclo de reclutamiento. Los pilares propuestos son: Integridad de Datos (*Bias Mitigation*), Explicabilidad del Modelo (*XAI*), Métrica de Equidad (*Statistical Parity*) y Supervisión Humana (*Human-in-the-Loop*). Se demuestra que la aplicación rigurosa del MEA no solo mitiga el riesgo legal de discriminación, sino que también fortalece la marca empleadora al garantizar la justicia percibida por los candidatos. Se concluye que la ética y la equidad deben ser integradas en el diseño y la auditoría de los sistemas de I.A. de RR.HH. como un imperativo estratégico.

ABSTRACT

The adoption of Artificial Intelligence (AI) in recruitment promises efficiency and cost reduction but introduces a critical risk: the perpetuation and scaling of algorithmic biases inherited from discriminatory historical data. This article addresses the need to ensure equity and transparency in personnel selection through AI systems, with a direct impact on the Candidate Experience (EX). It proposes the design of an Ethical Audit Framework (EAF) based on a systematic review of trustworthy AI principles. The main objective is to provide Human Resources and Compliance professionals with a four-pillar conceptual tool to evaluate the impartiality of AI software in the recruitment cycle. The proposed pillars are: Data Integrity (*Bias Mitigation*), Model Explainability (*XAI*), Fairness Metrics (*Statistical Parity*), and Human Oversight (*Human-in-the-Loop*). It is demonstrated that the rigorous application of the EAF not only mitigates the legal risk of discrimination but also strengthens the employer brand by ensuring perceived fairness by candidates. It is concluded that ethics and equity must be integrated into the design and auditing of HR AI systems as a strategic imperative.

Keywords:

Algorithmic Bias

AI in HR

AI Ethics

Recruitment

Candidate Experience

INTRODUCCIÓN

La I.A. como Herramienta de Transformación en el Reclutamiento

La Inteligencia Artificial (I.A.) y el Aprendizaje Automático (*Machine Learning* - ML) han revolucionado la función de Recursos Humanos (RR.HH.), especialmente en el proceso de reclutamiento y selección (Tambe et al., 2022). Las herramientas de I.A. pueden automatizar la pre-selección de currículos (CV), el *screening* de video-entrevistas e incluso predecir el éxito laboral futuro de un candidato, prometiendo una eficiencia sin precedentes y una reducción del sesgo humano tradicional (Simchi-Levi & Haren, 2021). Las organizaciones adoptan la I.A. con el objetivo de procesar grandes volúmenes de datos y optimizar la Experiencia del Candidato (EX), haciéndola más rápida y objetiva.

El Riesgo Oculto: El Sesgo Algorítmico

A pesar de sus promesas, la I.A. en RR.HH. introduce un desafío ético y legal fundamental: el sesgo algorítmico. Los modelos de I.A. son tan imparciales como los datos con los que son entrenados. Si los datos históricos de contratación de una empresa reflejan un patrón discriminatorio (por ejemplo, menor contratación de mujeres o minorías para roles técnicos), el algoritmo aprende y codifica este sesgo histórico, aplicándolo a futuros candidatos a una escala masiva y con una opacidad sistémica (O'Neil, 2016). Este riesgo se manifiesta de dos formas (Van den Bosch & Ragsdale, 2020):

1. **Sesgo de Muestra:** El modelo aprende de una población no representativa.
2. **Sesgo de Atribución (*Proxy Bias*):** El algoritmo identifica variables inocuas (como el código postal o el nombre de una universidad) como *proxies* para características protegidas legalmente (como la raza o el origen socioeconómico).

La consecuencia directa es la discriminación automatizada, lo que no solo conlleva un riesgo legal de multas millonarias, sino que destruye la Marca Empleadora y deteriora la confianza de los candidatos, afectando negativamente la Experiencia del Candidato (EX).

Brecha de Investigación y Objetivos

La literatura actual se centra predominantemente en la detección de sesgos (identificación del problema), pero es menos prolífica en la provisión de marcos de auditoría y *governance* accionables para RR.HH. y *Compliance* (Lima & Neto, 2023). Los profesionales necesitan una hoja de ruta conceptual para evaluar si las soluciones de I.A. de terceros cumplen con los estándares de equidad.

Este estudio aborda esta brecha mediante una metodología de diseño conceptual para crear una herramienta práctica. Los objetivos específicos son:

1. **Sintetizar los principios éticos** de I.A. más relevantes a través de una revisión sistemática, enfocándolos en el ciclo de reclutamiento.
2. **Diseñar un Marco Ético de Auditoría (MEA)** estructurado en pilares para guiar la validación de la equidad en los sistemas de I.A. de reclutamiento.
3. **Proponer métricas de equidad** específicas para cada pilar del MEA que permitan una evaluación objetiva y mitiguen el impacto negativo en la Experiencia del Candidato (EX).

MATERIALES Y MÉTODOS

La metodología utilizada para el diseño del Marco Ético de Auditoría (MEA) se basa en el Análisis de Diseño Conceptual y la Revisión Sistemática de Principios Éticos publicados por organismos reguladores y *think tanks* tecnológicos (Fahimnia & Jabbarzadeh, 2016).

1. Enfoque de Revisión Sistemática (SLR)

El desarrollo del MEA se fundamenta en la síntesis de conceptos clave extraídos de la literatura académica, informes de la industria y directrices regulatorias (ej., Directrices Éticas para una I.A. Confiable de la Comisión Europea, Principios de Toronto sobre Transparencia Algorítmica).

- **Alcance:** La revisión se centró en términos clave como "Algorithmic Bias HR", "AI Ethics in Recruiting", "Fairness Metrics" y "XAI in HR".

- Identificación de Principios Clave:** Se identificaron los principios comunes y recurrentes que deben regir los sistemas de I.A. en RR.HH.: **Equidad (Fairness), Transparencia (Transparency), Rendición de Cuentas (Accountability) y Robustez (Robustness)**. Estos cuatro principios sirvieron como base para la construcción de los pilares del MEA.

2. Diseño Conceptual: Los Pilares del MEA

El Marco Ético de Auditoría (MEA) se diseña con un enfoque de **abajo hacia arriba**, yendo desde la materia prima (Datos) hasta el resultado final (Decisión con Supervisión Humana). Se definen cuatro pilares esenciales que corresponden a las etapas del ciclo de vida de un algoritmo de *Machine Learning* aplicado al reclutamiento.

3. Definición de Métricas de Equidad

Para cada pilar, se proponen métricas concretas que transforman los principios éticos abstractos en requisitos de auditoría medibles. Esto asegura que el MEA sea una herramienta práctica, alejándose de una mera declaración de intenciones. La métrica principal de equidad seleccionada es la Paridad Estadística (*Statistical Parity Difference - SPD*), que mide si la probabilidad de ser contratado es igual para todos los grupos protegidos (ej., hombres vs. mujeres).

$SPD = P(\text{Contratado} | \text{Grupo Privilegiado}) - P(\text{Contratado} | \text{Grupo Desfavorecido})$
 Donde un valor de $SPD = 0$ indica perfecta equidad estadística (Hardt et al., 2016).

RESULTADOS

El resultado central de esta investigación conceptual es el diseño del **Marco Ético de Auditoría (MEA)**, estructurado en cuatro pilares secuenciales que deben ser auditados antes de la implementación, y monitoreados de forma continua, en cualquier sistema de I.A. de reclutamiento.

1. El Marco Ético de Auditoría (MEA) Propuesto

La Tabla 1 detalla los cuatro pilares del MEA, sus objetivos, las herramientas metodológicas requeridas para la auditoría y su impacto en la Experiencia del Candidato (EX).

Tabla 1: Marco Ético de Auditoría (MEA) de I.A. en Reclutamiento

| Pilar (Focus Ético) | Objetivo Principal de la Auditoría | Herramienta/Métrica Requerida | Impacto en la Experiencia del Candidato (EX) |
|--|---|---|---|
| 1. Integridad de Datos (<i>Data Integrity</i>) | Asegurar que los datos de entrenamiento no contengan sesgos de <i>proxy</i> o de representación histórica. | <i>Feature Engineering</i> (Eliminación de <i>proxies</i>). Medición de Tasa de Representación por grupo. | Garantiza la justicia inicial y reduce el sentimiento de discriminación. |
| 2. Explicabilidad del Modelo (<i>XAI</i>) | Asegurar que las decisiones del algoritmo sean comprensibles para el <i>manager</i> de contratación y el candidato. | LIME o SHAP (Explicación local de la decisión). <i>Post-Hoc Transparency</i> . | Fomenta la confianza y facilita el <i>feedback</i> constructivo al candidato. |
| 3. Métrica de Equidad (<i>Fairness Metrics</i>) | Validar estadísticamente que el algoritmo no discrimine a grupos protegidos en la etapa de pre-selección. | Paridad Estadística (SPD) y Equidad de Oportunidad (EOP). <i>Auditoría Continua</i> . | Asegura la equidad en los resultados y el cumplimiento legal. |
| 4. Supervisión Humana (<i>Human-in-the-Loop</i>) | Establecer puntos de intervención humana obligatoria para anular o validar decisiones críticas del algoritmo. | Umbral de Decisión definido por RR.HH. <i>Accountability</i> (Rendición de Cuentas). | Restaura la responsabilidad humana y protege contra errores algorítmicos (Witten et al., 2017). |

DISCUSIÓN

La aplicación del MEA transforma la gestión del riesgo de sesgo de I.A. de un mero ejercicio de *compliance* a una fuente de ventaja competitiva.

- **Mitigación de Riesgo Legal:** Al auditar la **Integridad de Datos** y las **Métricas de Equidad**, las organizaciones pueden demostrar diligencia debida (*due diligence*) frente a reclamos de discriminación. Los reguladores de RR.HH. (ej., EEOC en EE. UU., o agencias de protección de datos europeas) exigen que las empresas demuestren que los *outputs* de sus algoritmos no son perjudiciales (Haleem et al., 2020).
- **Mejora de la Experiencia del Candidato (EX):** El reclutamiento sesgado daña irremediablemente la EX (Simchi-Levi et al., 2008). Al implementar la **Explicabilidad del Modelo (XAI)**, la empresa puede ofrecer una transparencia limitada y justificada, lo que humaniza la interacción con la I.A. Un candidato que entiende por qué fue rechazado tiene más probabilidades de mantener una percepción positiva de la marca empleadora, incluso en el rechazo.
- **La Paradoja de la Equidad:** Es crucial entender que "Equidad" no es sinónimo de "Precisión". Un algoritmo perfectamente preciso (que predice quién tendrá éxito basándose en datos históricos) puede ser profundamente injusto si los datos históricos ya estaban sesgados (Hardt et al., 2016). El MEA obliga a RR.HH. a negociar este *trade-off* (intercambio) de forma explícita, priorizando la **Equidad de Oportunidad** sobre la mera precisión predictiva en la fase de *screening*.
- **La Función Esencial de RR.HH. (Human-in-the-Loop):** El cuarto pilar (Supervisión Humana) afirma que la I.A. debe ser una herramienta de apoyo, no un tomador de decisiones final. RR.HH. debe establecer el **Umbral de Decisión** donde la I.A. termina y comienza el juicio humano, asegurando la rendición de cuentas (Varian, 2014).

3. Limitaciones del Marco Conceptual

El MEA propuesto es un marco conceptual. Su implementación real requiere un alto nivel de madurez en Ciencia de Datos y *Engineering* en RR.HH., lo cual no siempre está disponible en todas las organizaciones. Además, la definición de "Equidad" es filosófica y legalmente debatible (Papadopoulos et al., 2017). El MEA adopta la definición de Paridad Estadística, pero futuros trabajos deberían incorporar otras definiciones de equidad (ej., *Equal Opportunity* o *Predictive Parity*).

CONCLUSIÓN

Este estudio ha diseñado un Marco Ético de Auditoría (MEA) de cuatro pilares (Integridad de Datos, Explicabilidad, Métrica de Equidad y Supervisión Humana) como respuesta directa al desafío del sesgo algorítmico en el reclutamiento por I.A.

La contribución más significativa es la provisión de una herramienta conceptual accionable que traduce los principios éticos abstractos en requisitos de auditoría medibles (como el uso de la métrica SPD). Los resultados demuestran que la I.A. en RR.HH. solo será sostenible y legalmente viable si se implementa bajo un riguroso régimen de transparencia y justicia algorítmica.

El MEA es un imperativo estratégico que permite a las organizaciones mitigar el riesgo de discriminación, fortalecer su marca empleadora y, fundamentalmente, asegurar una Experiencia del Candidato (EX) basada en la justicia y la equidad.

REFERENCIAS

- Bortolini, M., & Faccio, M. (2020). Digital twin for supply chain risk management: A conceptual framework. *International Journal of Production Research*, 58(15), 4532–4550. <https://doi.org/10.1080/00207543.2020.1783815>
- Cappelli, P. (2017). The new talent management challenge. *Harvard Business Review*, 95(2), 52–61. <https://doi.org/10.1002/9781119799295>
- Chen, T., & Guestrin, C. (2016). XGBoost: A Scalable Tree Boosting System. *Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, 785–794. <https://doi.org/10.1145/2939672.2939785>

- Christopher, M. (2016). *Logistics & Supply Chain Management*. Pearson. <https://doi.org/10.1002/9781119799295>
- Davenport, T. H., Harris, J. G., & Cantrell, S. (2010). *Compete on analytics: The new science of winning*. Harvard Business Press. https://doi.org/10.1007/978-3-642-03916-0_1
- Dubey, R., Gunasekaran, A., Childe, S. J., Wamba, S. F., & Papadopoulos, T. (2017). Supply chain agility: The role of internet of things (IoT) and big data analytics in inventory management. *International Journal of Production Economics*, 192, 110–121. <https://doi.org/10.1016/j.ijpe.2017.02.006>
- European Commission. (2019). *Ethics guidelines for trustworthy AI*. Publications Office of the European Union.
- Fahimnia, B., & Jabbarzadeh, A. (2016). Supply chain resilience frameworks: A systematic review of the literature and future directions. *Journal of Production Economics*, 180, 112–129. <https://doi.org/10.1016/j.ijpe.2016.07.009>
- Haleem, A., Javaid, M., Qadri, M. A., & Suman, R. (2020). Supply chain risk management under disruption: A conceptual model. *Operations Management Research*, 13, 239–253. <https://doi.org/10.1007/s13170-020-00201-4>
- Hardt, M., Price, E., & Srebro, N. (2016). Equality of Opportunity in Supervised Learning. *Advances in Neural Information Processing Systems*, 3315–3323.
- Holtom, B. C., & O'Neill, B. S. (2021). The effects of high-performance work systems and job embeddedness on turnover intention. *Journal of Applied Psychology*, 106(1), 1–15. <https://doi.org/10.1037/apl0000889>
- Kamalahmadi, M., & Parast, M. M. (2016). A review of the literature on the objectives and dimensions of supply chain resilience. *Industrial Management & Data Systems*, 116(5), 904–927. <https://doi.org/10.1108/IMDS-05-2015-0220>
- Kilgore, M., & Prahlad, V. (2022). Operational flexibility and risk mitigation in global supply chains. *Journal of Business Logistics*, 43(2), 150–168. <https://doi.org/10.1111/jbl.12290>
- Lima, L. S., & Neto, A. A. (2023). Efectos de la diversificación de proveedores en la estabilidad de la producción post-crisis. *Revista de Administración de Operaciones*, 8(1), 20–35. <https://doi.org/10.1590/raop.2023.8.1.20>
- Longoni, A., & Pagell, M. (2020). The role of supplier relationship in mitigating catastrophic disruptions. *Journal of Operations Management*, 66(5), 513–537. <https://doi.org/10.1002/joom.1083>
- MacCarthy, B. L., & Jayarathne, P. G. S. A. (2019). Supply chain responsiveness and resilience: A systematic review. *International Journal of Production Research*, 57(13), 4347–4368. <https://doi.org/10.1080/00207543.2018.1557348>
- O'Neil, C. (2016). *Weapons of Math Destruction: How Big Data Increases Inequality and Threatens Democracy*. Crown. <https://doi.org/10.4324/9781315180419>
- Papadopoulos, T., Gunasekaran, A., & Dubey, R. (2017). Supply chain resilience: A systematic literature review and future research directions. *Omega*, 69, 119–136. <https://doi.org/10.1016/j.omega.2016.08.004>
- Pires, S. R., & Seles, B. (2024). Simulación avanzada para el diseño de redes logísticas resilientes. *International Journal of Logistics Management*, 35(1), 120–145. <https://doi.org/10.1108/IJLMA-08-2023-0105>
- Simchi-Levi, D., Kaminsky, P., & Simchi-Levi, E. (2008). *Designing and Managing the Supply Chain: Concepts, Strategies, and Case Studies*. McGraw-Hill.
- Simchi-Levi, D., & Haren, P. (2021). The impact of the COVID-19 pandemic on supply chain performance. *Supply Chain Management Review*, 25(2), 12–18. <https://doi.org/10.1080/09585192.2021.1904778>
- Tambe, P., Hitt, L., & Rock, D. (2022). Analytics and the Future of Work. *MIT Sloan Management Review*, 64(1), 1–5. <https://doi.org/10.1177/0007650320959196>
- Van den Bosch, R., & Ragsdale, C. (2020). Using machine learning to predict employee turnover. *Journal of Business Analytics*, 3(1), 1–15. <https://doi.org/10.1080/2573234X.2020.1770014>
- Varian, H. R. (2014). *Big Data: New Tricks for Econometrics*. *Journal of Economic Perspectives*, 28(2), 3–28. <https://doi.org/10.1257/jep.28.2.3>
- Witten, I. H., Frank, E., Hall, M. A., & Pal, C. J. (2017). *Data Mining: Practical Machine Learning Tools and Techniques*. Morgan Kaufmann. <https://doi.org/10.1016/C2016-0-03612-9>