

Hacia una gobernanza algorítmica transparente: auditoría de sesgo. Estudio de caso

José Luis Bustelo Gracia
ESERP, Digital Business and Law School 

<https://dx.doi.org/10.5209/cgap.97604>

Recibido: 20/08/2024 • Aceptado: 26/03/2025

Resumen. Este artículo explora la necesidad de implementar auditorías de sesgo en sistemas algorítmicos, particularmente en el contexto de la gobernanza pública. Con el incremento del uso de la inteligencia artificial (IA) en la toma de decisiones gubernamentales, surge la preocupación sobre la transparencia y equidad de estos sistemas. Se analizan diversas normativas europeas, como el Reglamento de Inteligencia Artificial (AI Act), y estudios de caso que evidencian la persistencia de sesgos en algoritmos, destacando la importancia de un enfoque proactivo en la regulación y la implementación de auditorías sistemáticas. Además, se discuten las implicaciones éticas y sociales de la gobernanza algorítmica y se proponen soluciones para mitigar los riesgos asociados.

Palabras Clave: Gobernanza algorítmica, transparencia, auditoría de sesgo, inteligencia artificial.

ENG **Advancing transparent algorithmic Governance. A Case study in bias auditing**

Summary. This article examines the necessity of implementing bias audits in algorithmic systems, particularly within the context of public governance. With the increasing use of artificial intelligence (AI) in governmental decision-making, concerns about the transparency and fairness of these systems have arisen. The paper analyzes various European regulations, such as the Artificial Intelligence Act (AI Act), and case studies that highlight the persistence of biases in algorithms, emphasizing the importance of a proactive regulatory approach and systematic audits. Additionally, it discusses the ethical and social implications of algorithmic governance and proposes solutions to mitigate associated risks.

Keywords: Algorithmic governance, transparency, bias audit, artificial intelligence.

Sumario: 1. Introducción. 2. Metodología. 3. Marco teórico. 3.1. Introducción a la gobernanza algorítmica. 3.2. Ética y derechos humanos. 4. Normativa Europea en transparencia algorítmica. 5. Estudio de caso y análisis crítico. Auditoría de un algoritmo de contratación laboral. 6. Análisis crítico. Discusión. 7. Conclusiones. 8. Bibliografía.

Como citar: Bustelo Gracia, J. L. (2025). Hacia una gobernanza algorítmica transparente: auditoría de sesgo. Estudio de caso, en *Cuadernos de Gobierno y Administración Pública* 12(1), e97604. <https://dx.doi.org/10.5209/cgap.97604>

1. Introducción

La creciente adopción de sistemas de inteligencia artificial (IA) en la gobernanza pública ha planteado desafíos importantes en términos de equidad, transparencia y responsabilidad. A medida que los algoritmos se integran en la toma de decisiones, su capacidad para amplificar sesgos preexistentes y generar nuevas formas de discriminación se ha convertido en una preocupación crítica. La transparencia algorítmica y la implementación de auditorías de sesgo son, por tanto, herramientas esenciales para mitigar estos riesgos y garantizar que la inteligencia artificial se utilice de manera justa y ética.

Este artículo se inscribe en el marco de las discusiones sobre gobernanza y administración pública abordadas en Cuadernos de Gobierno y Administración Pública, particularmente en relación con el desarrollo de nuevos modelos de gobernanza, la capacidad burocrática y el impacto de la tecnología en la gestión pública. La literatura previa ha explorado diversas dimensiones de la gobernanza (Aguilar Villanueva, 2014), los distintos tipos y enfoques existentes (Arellano Gault, 2014), así como los desafíos de la sociedad moderna en la implementación de modelos efectivos (García Magariño, 2016). Más recientemente, el debate se ha ampliado para incluir la capacidad burocrática y su relación con la eficacia institucional (Guerrero García, 2023) y la integración de la inteligencia artificial en los procesos administrativos (Crespo-González et al., 2024). En este sentido, el presente trabajo busca aportar a esta línea de investigación al analizar cómo la gobernanza algorítmica redefine la toma de decisiones y el papel de la corresponsabilidad en el ámbito público.

El concepto de “gobernanza algorítmica” se refiere al uso de algoritmos para gestionar y tomar decisiones en contextos gubernamentales y corporativos. Sin embargo, estos sistemas, aunque prometen eficiencia y objetividad, a menudo carecen de transparencia, lo que dificulta la identificación y corrección de sesgos inherentes. Diversos estudios han demostrado que los algoritmos pueden perpetuar desigualdades, ya sea por la naturaleza de los datos utilizados o por decisiones de diseño tomadas por los programadores (Dekker et al., 2022).

El desarrollo de capacidades institucionales es un elemento clave en la gobernanza contemporánea, especialmente en el contexto de la digitalización y el uso de herramientas tecnológicas en la toma de decisiones. Desde una perspectiva teórico-conceptual, la gobernanza algorítmica debe entenderse como un proceso dinámico en el que las instituciones deben adaptarse a nuevas modalidades de gestión basadas en datos y automatización. Esto requiere no solo infraestructura tecnológica, sino también marcos normativos claros y estrategias de capacitación que permitan a los actores gubernamentales interpretar, supervisar y corregir las decisiones automatizadas cuando sea necesario. El fortalecimiento de estas capacidades institucionales es esencial para garantizar que la tecnología no solo optimice procesos, sino que también se alinee con principios democráticos y de equidad.

La necesidad de auditorías de sesgo en los sistemas de IA ha sido reconocida como una práctica esencial para la gobernanza responsable. Investigaciones recientes subrayan que la incorporación de auditorías independientes puede reducir significativamente el sesgo en los sistemas de IA, mejorando así la equidad en la toma de decisiones algorítmicas (Conitzer et al., 2022). Además, el caso de la implementación de la Ley Local 144 en Nueva York destaca los desafíos y beneficios de establecer regímenes de auditoría de sesgo como un estándar legal, lo que demuestra la importancia de tales medidas en la mejora de la transparencia y la confianza pública en la tecnología (Groves et al., 2024).

No obstante, para que estas auditorías sean efectivas, es fundamental comprender que el sesgo en los algoritmos no solo proviene de los datos, sino también de las decisiones tomadas en todas las etapas del desarrollo y despliegue del sistema. La legislación actual en varios países, incluida la propuesta de Reglamento de Inteligencia Artificial de la Unión Europea, se está adaptando para abordar estos problemas, proponiendo la necesidad de una regulación más amplia que abarque todos los aspectos del ciclo de vida de los algoritmos.

Es fundamental diferenciar entre algoritmos, programas y datos, ya que cada uno de estos elementos puede introducir sesgos en un sistema. Un algoritmo es una secuencia de instrucciones diseñadas para resolver un problema específico, mientras que un programa es la implementación práctica de uno o varios algoritmos en un lenguaje de programación. Los datos, por su parte, representan la información de entrada que un algoritmo procesa y pueden ser una fuente crítica de sesgo si no son representativos o contienen prejuicios inherentes. En este sentido, la auditoría de sesgos debe abordar estos tres componentes de manera diferenciada para una evaluación efectiva.

Por lo tanto, este artículo explora las diferentes dimensiones de la auditoría de sesgo en la gobernanza algorítmica, examinando tanto las normativas existentes como los estudios de caso que evidencian la urgencia de adoptar un enfoque más transparente y responsable en el desarrollo y la implementación de la IA.

2. Metodología

El presente estudio emplea un enfoque cualitativo y documental para analizar la implementación y efectividad de auditorías de sesgo en sistemas algorítmicos dentro del contexto de la gobernanza pública. Se llevaron a cabo las siguientes etapas metodológicas:

Revisión de la literatura: Se realizó una exhaustiva revisión de la literatura académica y de normativas europeas relevantes, como el Reglamento de Inteligencia Artificial (AI Act). Esta revisión permitió identificar los principales desafíos asociados a la transparencia y equidad en los sistemas algorítmicos, así como las propuestas regulatorias actuales para mitigar dichos problemas.

Estudio de caso: Se seleccionaron y analizaron estudios de caso significativos, como la implementación de auditorías de sesgo en plataformas de contratación automatizadas. Estos casos proporcionan evidencia empírica sobre los efectos de los sesgos algorítmicos en la toma de decisiones y la eficacia de las auditorías para mitigarlos.

Análisis crítico: Se aplicó un enfoque analítico para evaluar las dimensiones éticas, técnicas y legales de la gobernanza algorítmica. Este análisis incluyó la revisión de la estructura y entrenamiento de algoritmos específicos, así como la evaluación de su impacto en términos de equidad y transparencia.

Propuesta de mejoras: A partir de los hallazgos obtenidos, se propusieron mejoras regulatorias y técnicas que pueden contribuir a una gobernanza algorítmica más transparente y justa. Estas recomendaciones se fundamentan en la necesidad de una regulación proactiva y en la incorporación de auditorías sistemáticas como práctica estándar en el desarrollo y despliegue de sistemas de IA.

Este enfoque metodológico combina el análisis teórico con la evaluación práctica, permitiendo una comprensión integral de los desafíos y soluciones en la gobernanza de sistemas algorítmicos.

3. Marco teórico

3.1. Introducción a la gobernanza algorítmica

La gobernanza algorítmica se refiere a la creciente dependencia de sistemas automatizados de decisión en el ámbito de la administración pública y la toma de decisiones gubernamentales. Estos sistemas, alimentados por datos masivos y desarrollados a partir de modelos de inteligencia artificial (IA), prometen mejorar la eficiencia y la equidad en la gestión de políticas públicas. Sin embargo, su implementación también plantea serias preocupaciones éticas, particularmente en torno a la transparencia, la responsabilidad y el sesgo inherente en los algoritmos.

En las últimas décadas, los gobiernos han adoptado cada vez más herramientas algorítmicas para optimizar procesos que van desde la asignación de recursos hasta la predicción de necesidades sociales. La capacidad de estos sistemas para procesar grandes volúmenes de datos y generar decisiones rápidas y precisas ha sido aclamada como un avance revolucionario en la gobernanza pública (Dekker et al., 2022). Sin embargo, esta revolución también ha suscitado debates sobre los posibles efectos adversos de la automatización en la toma de decisiones. Las aplicaciones de algoritmos en la administración pública se utilizan en una amplia variedad de aplicaciones dentro del ámbito gubernamental. Desde la asignación de viviendas y recursos sociales hasta la vigilancia de cumplimiento de leyes, los sistemas automatizados han demostrado su capacidad para mejorar la eficiencia administrativa. Por ejemplo, en el sector de la justicia, se han implementado sistemas algorítmicos para asistir en decisiones judiciales, como la evaluación de riesgo de reincidencia en individuos condenados.

Un ejemplo notable de la gobernanza algorítmica es el uso de sistemas de predicción en la gestión de servicios sociales. Estos sistemas analizan datos históricos y actuales para predecir las necesidades futuras de la población, permitiendo a los gobiernos asignar recursos de manera más efectiva. Sin embargo, la precisión de estas predicciones depende en gran medida de la calidad de los datos y de los modelos utilizados, lo que puede llevar a errores significativos si no se gestionan adecuadamente (Groves et al., 2024). Aunque la gobernanza algorítmica ofrece ventajas evidentes, también presenta una serie de desafíos que deben ser abordados para garantizar su eficacia y equidad. Uno de los principales problemas es el sesgo algorítmico, que puede surgir en diferentes etapas del ciclo de vida del algoritmo, desde la recolección de datos hasta el desarrollo del modelo y su implementación.

El sesgo en los algoritmos puede tener graves implicaciones, especialmente cuando se aplican en contextos sensibles como la justicia o la asignación de beneficios sociales. Un estudio reciente destacó cómo los algoritmos utilizados en el sistema de justicia penal en los Estados Unidos mostraron un sesgo significativo contra minorías raciales, lo que llevó a decisiones injustas y desproporcionadas en términos de sentencias y libertad condicional (Conitzer et al., 2022). Este problema no se limita a los Estados Unidos; en todo el mundo, la falta de diversidad en los equipos de desarrollo y la dependencia de datos históricos sesgados han contribuido a la perpetuación de desigualdades a través de la tecnología.

Otro desafío importante es la falta de transparencia en los sistemas algorítmicos. Muchos de estos sistemas operan como "cajas negras", lo que significa que su funcionamiento interno no es comprensible para los usuarios o incluso para los propios desarrolladores. Esto plantea serios problemas de responsabilidad, ya que es difícil determinar quién es responsable cuando un algoritmo toma una decisión errónea o injusta (Groves et al., 2024). La opacidad de estos sistemas también dificulta la implementación de mecanismos de control y supervisión, lo que aumenta el riesgo de abuso. Por ello, la transparencia es un principio fundamental en la gobernanza pública, y su importancia se amplifica en el contexto de la gestión algorítmica. La capacidad de los ciudadanos y los organismos reguladores para comprender cómo funcionan los algoritmos y cómo se toman las decisiones es crucial para garantizar que estos sistemas se utilicen de manera justa y responsable. En este sentido, se han propuesto diversas medidas para mejorar la transparencia en la gobernanza algorítmica.

3.2. Ética y derechos humanos

Una de las medidas más prometedoras es la implementación de auditorías de algoritmos, que pueden ayudar a identificar y corregir sesgos, así como a garantizar que los sistemas cumplen con los estándares éticos y legales. Estas auditorías deben ser realizadas por entidades independientes y deben estar disponibles para el público, para que los ciudadanos puedan confiar en la imparcialidad y la justicia de los sistemas algorítmicos utilizados por el gobierno (Dekker et al., 2022). Además, la adopción de estándares abiertos y la publicación de los códigos fuente de los algoritmos son pasos importantes hacia una mayor transparencia. Esto permitiría a investigadores y organizaciones de la sociedad civil examinar los sistemas y proponer

mejoras, contribuyendo así a una gobernanza más inclusiva y participativa. Se sugiere que la gobernanza algorítmica no solo afecta la eficiencia administrativa, sino que también tiene profundas implicaciones para los derechos humanos. La adopción de sistemas algorítmicos en la gestión gubernamental puede amenazar derechos fundamentales como la privacidad, la igualdad y el acceso a la justicia. Por ejemplo, el uso de algoritmos en la vigilancia y la seguridad ha generado preocupaciones sobre la invasión de la privacidad y la discriminación racial y étnica.

Es fundamental que cualquier implementación de algoritmos en la gobernanza se realice con un enfoque basado en los derechos humanos, que considere tanto los beneficios como los riesgos potenciales de estas tecnologías. Esto implica no solo la realización de evaluaciones de impacto antes de la implementación, sino también la adopción de medidas correctivas cuando se detectan violaciones de derechos. A pesar de los desafíos, la gobernanza algorítmica tiene el potencial de transformar positivamente la administración pública. Con el desarrollo de nuevas tecnologías, como la inteligencia artificial explicable (XAI), que permite a los usuarios comprender mejor las decisiones algorítmicas, y el uso de tecnologías blockchain para garantizar la transparencia y la integridad de los datos, es posible imaginar un futuro en el que los algoritmos se utilicen de manera justa y eficiente para mejorar la vida de los ciudadanos. Es un área prometedora es la co-creación de algoritmos, donde los desarrolladores trabajan en colaboración con actores gubernamentales, organizaciones de la sociedad civil y ciudadanos para diseñar sistemas que reflejen mejor las necesidades y valores de la comunidad. Este enfoque participativo no solo mejora la transparencia y la confianza en los sistemas, sino que también ayuda a garantizar que los algoritmos se desarrollos y utilicen de manera ética y responsable (Dekker et al., 2022).

La implementación de herramientas algorítmicas en la gobernanza plantea dilemas éticos y morales que requieren un análisis detallado. La objetividad que se busca a través de la sistematización de la toma de decisiones no está exenta de riesgos, especialmente si se desestima la corresponsabilidad de los distintos actores involucrados en estos procesos. La corresponsabilidad implica que tanto los diseñadores de los sistemas como los funcionarios públicos y la ciudadanía deben participar activamente en la supervisión y control del impacto de estas herramientas. Sin un marco ético sólido, las decisiones algorítmicas podrían reforzar desigualdades existentes o generar efectos no deseados en la población afectada. Un ejemplo de ello es el uso de algoritmos en la asignación de recursos sociales: si bien pueden optimizar la distribución, también pueden omitir factores contextuales que requieren juicio humano.

La gestión algorítmica en la gobernanza representa tanto una oportunidad como un desafío significativo para los gobiernos modernos. Si bien estos sistemas pueden mejorar la eficiencia y la precisión en la toma de decisiones, también plantean serias preocupaciones sobre el sesgo, la transparencia y la protección de los derechos humanos. Es esencial que los gobiernos adopten un enfoque proactivo para abordar estos desafíos, implementando medidas que garanticen la transparencia, la responsabilidad y la equidad en el uso de algoritmos. Solo así se podrá aprovechar plenamente el potencial de la inteligencia artificial en la gobernanza pública, garantizando al mismo tiempo que se respeten los principios fundamentales de la democracia y los derechos humanos. Se añade que la creciente incorporación de sistemas algorítmicos en la toma de decisiones gubernamentales y comerciales ha traído consigo el reto de garantizar la transparencia y equidad de estas tecnologías. La necesidad de una gobernanza algorítmica centrada en la ética ha subrayado la importancia de mecanismos de auditoría para identificar y corregir posibles sesgos presentes en los modelos algorítmicos. Este enfoque no solo mejora la equidad en la toma de decisiones, sino que también promueve la confianza pública en estas tecnologías. Es cierto que uno de los principales desafíos políticos en la implementación de algoritmos en la gobernanza es la opacidad inherente a muchos de estos sistemas, lo que dificulta la identificación de sesgos y errores. La transparencia algorítmica implica hacer visibles los procesos y las decisiones tomadas por los algoritmos, permitiendo su análisis y auditoría. Según Giunchiglia et al. (2021), la diversidad es un factor clave para impulsar la transparencia, ya que permite comprender mejor y abordar los sesgos que pueden surgir en los sistemas algorítmicos. Además, la transparencia también se ve como un prerequisito para construir confianza y legitimidad en los sistemas algorítmicos, especialmente cuando afectan derechos humanos fundamentales como la privacidad y la no discriminación. La auditoría de sesgos se ha convertido en una práctica esencial para garantizar que los algoritmos no perpetúen o amplifiquen desigualdades. Lam et al. (2024) proponen un marco de auditoría basado en criterios para asegurar que los sistemas algorítmicos cumplan con estándares éticos y de gobernanza, adaptando estos principios a las inminentes normativas legislativas incipientes y las que pueden venir como por ejemplo las auditorías de sesgo en algoritmos de contratación tanto públicos como privados. Estas auditorías permiten identificar de manera sistemática los puntos críticos donde se pueden manifestar los sesgos, ofreciendo una base sólida para tomar decisiones correctivas y evitar el daño a grupos vulnerables. Explicar y diferenciar entre explicabilidad y auditabilidad en los sistemas algorítmicos se convierte en una de los mayores desafíos. Aunque ambos conceptos están relacionados con la transparencia, cumplen funciones distintas. La explicabilidad se enfoca en proporcionar a los usuarios finales una comprensión clara de cómo y por qué un algoritmo toma decisiones específicas, mientras que la auditabilidad está orientada hacia el análisis interno y externo por expertos para evaluar la precisión y la equidad de un sistema. Springer y Whittaker (2021) subrayan que estos dos elementos no pueden abordarse de manera simultánea mediante una única implementación de transparencia, ya que sirven a diferentes propósitos.

A nivel global, los gobiernos y organizaciones han comenzado a establecer políticas para regular el uso de algoritmos, con un enfoque particular en la transparencia y la auditoría de sesgos. Se destacan cómo las políticas en Estados Unidos, aunque todavía limitadas, han comenzado a incorporar requisitos de transparencia y auditoría para abordar problemas de sesgo, con iniciativas que incluyen prohibiciones temporales y

evaluaciones obligatorias. Sin embargo, queda un largo camino por recorrer para implementar una legislación más comprehensiva que pueda mitigar efectivamente los riesgos asociados con el uso de algoritmos en sectores críticos como el financiero y el de contratación laboral (Qureshi et al., 2024). En estos casos la auditoría de sistemas algorítmicos no solo es una herramienta de detección, sino también un mecanismo para corregir y mejorar continuamente los modelos. Conitzer et al. (2022) presentan un estudio que evidencia cómo la auditoría independiente puede reducir el sesgo en sistemas algorítmicos, impactando positivamente la equidad en la toma de decisiones y proporcionando una base para futuras innovaciones en gobernanza algorítmica. Este enfoque se convierte en un elemento crucial para garantizar que los sistemas operen dentro de los límites éticos y respeten los derechos fundamentales. Para avanzar hacia una gobernanza algorítmica más transparente y justa, es necesario no solo mejorar la tecnología, sino también fortalecer los marcos regulatorios y promover la colaboración entre múltiples actores. La participación activa de organizaciones de la sociedad civil, junto con evaluaciones rigurosas por parte de auditores independientes, puede garantizar que los sistemas se diseñen y operen de manera que minimicen los sesgos y maximicen la equidad y la justicia. Esto también implica la necesidad de una formación constante para los desarrolladores y responsables de políticas, asegurando que comprendan las complejidades y riesgos inherentes a los sistemas algorítmicos.

El rol del usuario en la interpretación de los resultados generados por el sistema de inteligencia artificial es crucial. Dependiendo del contexto, el usuario puede actuar como mero receptor de información o como un evaluador crítico que valida los resultados antes de su aplicación. Por ejemplo, en el ámbito de la justicia, un algoritmo de predicción de reincidencia debe ser utilizado como una herramienta de apoyo y no como una decisión definitiva sin supervisión humana. En contraste, en tareas más operativas, como la clasificación automática de correos electrónicos, la intervención humana puede no ser necesaria. Para ilustrar esto, consideremos un sistema de diagnóstico médico asistido por IA: si bien el algoritmo puede generar una probabilidad de enfermedad basada en los síntomas del paciente, la decisión final debe estar a cargo de un profesional de la salud que contextualice la información.

Las herramientas tecnológicas aplicadas a la gobernanza abarcan una amplia variedad de sistemas, desde modelos predictivos hasta plataformas de análisis de datos en tiempo real. Un aspecto clave de su funcionamiento es la capacidad de procesar grandes volúmenes de información y detectar patrones que pueden ser utilizados para optimizar la toma de decisiones. Sin embargo, estas herramientas también presentan desafíos, como la opacidad en sus criterios de decisión, la posibilidad de replicar sesgos preexistentes y la dificultad de establecer mecanismos de supervisión efectivos. Por ejemplo, en la administración de justicia, los algoritmos de evaluación de riesgo pueden influir en la determinación de sentencias o medidas cautelares, lo que plantea interrogantes sobre su imparcialidad y transparencia. Para mitigar estos riesgos, es fundamental establecer criterios claros de validación y revisión humana en su implementación.

Estas herramientas utilizadas en los procesos de gobernanza incluyen desde sistemas de inteligencia artificial para la predicción de tendencias en políticas públicas hasta plataformas de blockchain para la gestión transparente de datos y contratos inteligentes. Estas herramientas operan mediante modelos algorítmicos que analizan grandes volúmenes de información para optimizar la toma de decisiones. Un ejemplo de ello son los sistemas de análisis predictivo en la gestión urbana, los cuales permiten anticipar la demanda de servicios públicos y optimizar la asignación de recursos en tiempo real. Otro caso relevante es el uso de plataformas de participación ciudadana basadas en inteligencia colectiva, que facilitan la interacción entre la ciudadanía y los gobiernos mediante mecanismos de votación y deliberación en línea. Sin embargo, para su correcta implementación, estas tecnologías requieren una adecuada infraestructura de datos, protocolos de interoperabilidad y marcos regulatorios que garanticen su correcto funcionamiento y uso ético.

4. Normativa Europea en transparencia algorítmica

El Artificial Intelligence Act (AI Act) es una propuesta legislativa pionera dentro del marco normativo de la Unión Europea, destinada a regular el uso, desarrollo y comercialización de sistemas de IA en Europa. Actualmente, se encuentra en proceso de revisión y negociación en diversas instancias legislativas, incluyendo el Parlamento Europeo y el Consejo de la UE, donde se debaten y proponen enmiendas. Aunque el reglamento aún no ha sido adoptado definitivamente, se prevé que entre en vigor en 2025, estableciendo un marco normativo vinculante para todos los Estados miembros. Este marco busca garantizar que los sistemas de IA sean seguros, éticos y respeten derechos fundamentales como la privacidad y la no discriminación (Busuioc et al., 2022; Gstrein et al., 2024).

El AI Act se configura como la piedra angular de la estrategia regulatoria de la UE en materia de inteligencia artificial. Este reglamento clasifica los sistemas de IA en función de su nivel de riesgo, desde bajo hasta inaceptable, prohibiendo aquellos que podrían comprometer derechos fundamentales o la seguridad pública. Se presta especial atención a las aplicaciones consideradas de alto riesgo, como el reconocimiento facial en espacios públicos, los sistemas de toma de decisiones en sectores críticos como el laboral, financiero o educativo, y aquellos relacionados con infraestructuras críticas (Engelmann, 2023; Sovrano et al., 2022). En este sentido, la normativa impone obligaciones a proveedores y usuarios de IA para garantizar la transparencia mediante mecanismos como la documentación adecuada, la trazabilidad de los datos y la explicabilidad de los algoritmos empleados (Turksen et al., 2024). Estos requerimientos no solo buscan prevenir sesgos y discriminación, sino también garantizar la privacidad en la toma de decisiones automatizadas (Gyevnar et al., 2023; Calero Valdez et al., 2024).

La transparencia algorítmica se posiciona como un pilar central dentro de las normativas propuestas. Las normativas se enfocan especialmente en los sistemas de IA de alto riesgo que incluyen aplicaciones en salud, educación, empleo y administración pública, donde se exige no solo la transparencia, sino también auditorías independientes que verifiquen el cumplimiento de los estándares establecidos. Además, se imponen sanciones significativas para las organizaciones que incumplan estas normas, reforzando el compromiso de la UE con la protección de derechos y la promoción de la confianza pública. Zharova (2023) subraya que la transparencia no se limita a la accesibilidad de la información, sino que también debe garantizar la comprensibilidad para los usuarios finales, lo que es esencial para evitar sesgos y permitir decisiones informadas. En este contexto, las herramientas de explicabilidad, como los mecanismos de explicación basados en lógica, son cruciales para cumplir con los requisitos de la normativa.

Mirando hacia el futuro, el AI Act representa un esfuerzo integral para abordar los desafíos actuales y futuros de la gobernanza algorítmica en múltiples sectores. Este marco normativo no solo anticipa los riesgos éticos y sociales del despliegue masivo de la IA, sino que también propone soluciones que armonicen la innovación tecnológica con la protección de los derechos fundamentales (Wörsdörfer, 2023; Laux et al., 2023). Con la implementación del AI Act y las leyes previstas para 2025, Europa se posiciona como líder en la regulación de la IA, estableciendo un modelo de gobernanza algorítmica transparente y equitativa para el futuro.

La normativa europea en transparencia algorítmica no se limita a imponer reglas técnicas, sino que también establece principios éticos y sociales que buscan alinear el desarrollo de la inteligencia artificial con los valores fundamentales de la Unión Europea. En este sentido, el AI Act se convierte en una herramienta clave para garantizar que las tecnologías emergentes se implementen de manera responsable, considerando su impacto en la sociedad y en los derechos humanos. Este enfoque multidimensional implica no solo el cumplimiento de requisitos técnicos, sino también la creación de un marco de responsabilidad compartida entre desarrolladores, usuarios, reguladores y auditores, que refuerza la confianza en los sistemas de IA.

Además de los sistemas de alto riesgo, el AI Act también propone medidas para gestionar tecnologías emergentes cuya aplicación aún no ha sido ampliamente explorada, como la IA generativa y los sistemas autónomos. Estas tecnologías presentan desafíos únicos en términos de transparencia y trazabilidad, lo que ha llevado a la UE a incluir disposiciones específicas para garantizar que su uso se mantenga bajo control y que los riesgos se gestionen de manera adecuada. El reglamento también contempla la actualización constante de los requisitos y estándares, reconociendo la rapidez con la que evoluciona la tecnología y la necesidad de adaptar la regulación a nuevos desafíos.

Otro aspecto relevante es la interoperabilidad entre marcos regulatorios. La UE busca que el AI Act sea compatible con otras normativas internacionales y con los principios establecidos en acuerdos como el Reglamento General de Protección de Datos (GDPR). Esta alineación facilita la cooperación global en la gobernanza de la IA y permite que Europa marque una influencia significativa en la regulación mundial, promoviendo estándares que combinen innovación con protección de derechos. Esta visión integradora no solo fortalece la posición de la UE en el ámbito tecnológico, sino que también establece precedentes para el desarrollo de políticas públicas que equilibran progreso y responsabilidad.

Como aspecto negativo a destacar dentro de la normativa europea en transparencia algorítmica, aunque bien intencionada, enfrenta desafíos significativos en su implementación y efectividad. Si bien establece la necesidad de que los algoritmos sean comprensibles y auditables, la complejidad inherente de muchos modelos de inteligencia artificial dificulta que esta transparencia sea realmente alcanzable en la práctica. Además, la normativa puede imponer cargas adicionales a las empresas, especialmente a las pequeñas y medianas, que podrían no tener los recursos necesarios para cumplir con los estrictos requisitos de transparencia. Esto podría generar un entorno donde solo las grandes corporaciones puedan cumplir con las normativas, lo que podría limitar la innovación y la competencia en el sector. Por ello, la normativa no aborda de manera suficiente los desafíos relacionados con la interpretación y el uso de la información proporcionada por los algoritmos, lo que puede llevar a una falsa sensación de seguridad respecto a la fiabilidad y equidad de las decisiones automatizadas.

Finalmente, el AI Act promueve una cultura de transparencia en todos los niveles del ciclo de vida de los sistemas de IA. Desde la fase de diseño hasta la implementación y uso, la normativa exige una documentación exhaustiva que permita a las autoridades y al público comprender cómo y por qué se toman ciertas decisiones automatizadas. Este enfoque reduce la brecha entre la tecnología y los ciudadanos, empoderando a los usuarios con información clara y accesible. En resumen, la normativa europea en transparencia algorítmica no solo responde a la necesidad de regular una tecnología en rápido desarrollo, sino que sienta las bases para una gobernanza algorítmica basada en la ética, la transparencia y la inclusión, asegurando que la innovación tecnológica avance en paralelo con el respeto a los derechos fundamentales y la promoción de una sociedad más justa y equitativa.

Es importante destacar que para que un auditor pueda evaluar un algoritmo de manera efectiva, es necesario definir criterios claros y metodologías específicas. La auditoría puede realizarse a través de diversas estrategias, como el análisis de resultados en distintos subgrupos poblacionales (para detectar sesgos discriminatorios), la inspección del código fuente y la revisión del proceso de recolección y selección de datos de entrenamiento. Un enfoque recomendable es la auditoría de caja negra, en la cual el auditor analiza los insumos y salidas del sistema sin acceso al código interno, comparándolo con decisiones humanas para evaluar inconsistencias. Alternativamente, la auditoría de caja blanca permite inspeccionar la estructura interna del modelo y sus parámetros de decisión. Un ejemplo práctico sería la auditoría de un sistema de selección de personal: se podría analizar si ciertos grupos demográficos reciben sistemáticamente puntuaciones más bajas, sugiriendo un posible sesgo.

5. Estudio de caso y análisis crítico. Auditoría de un algoritmo de contratación laboral

Se presenta un estudio de caso centrado en la auditoría de un algoritmo de contratación, un área en la que los sesgos inherentes pueden tener efectos significativos en términos de igualdad de oportunidades y justicia. La auditoría algorítmica en plataformas de contratación es crucial para mitigar los sesgos inherentes que pueden perpetuar desigualdades al priorizar ciertas características sobre otras. Investigaciones han demostrado que sistemas automatizados, como los algoritmos de reclutamiento de una gran compañía tecnológica, han replicado sesgos de género al favorecer a hombres sobre mujeres en el ámbito tecnológico, contribuyendo a la subrepresentación de estas últimas en este campo. Además, estudios recientes destacan la necesidad de realizar auditorías tanto internas como externas para identificar y corregir estas disparidades en los sistemas de contratación automatizados (Ajunwa, 2019). El algoritmo analizado en este estudio se utiliza en procesos de selección para recomendar candidatos en función de sus currículums y perfiles en línea. El objetivo de la auditoría fue evaluar la equidad y transparencia del sistema, especialmente en términos de género y etnicidad. Este tipo de herramientas se han vuelto comunes en plataformas de contratación, pero investigaciones recientes han revelado que pueden perpetuar sesgos al priorizar ciertas características sobre otras.

En este caso, se descubrió que el algoritmo mostraba una preferencia sistemática por candidatos masculinos para roles técnicos, reflejando sesgos tanto en los datos de entrenamiento como en la estructura del modelo. Según Albaroudi et al. (2024), los sesgos en algoritmos de contratación pueden surgir debido a datos históricos que reflejan patrones de discriminación y estereotipos en la contratación laboral (Albaroudi et al., 2024). A pesar de los esfuerzos para eliminar este tipo de sesgos mediante técnicas como la corrección del espacio vectorial en modelos de aprendizaje profundo, el estudio evidenció que las intervenciones técnicas a menudo son insuficientes para erradicar por completo estos problemas. Adicionalmente, la auditoría identificó que el algoritmo penalizaba a los candidatos con nombres asociados a minorías étnicas, lo que resulta en una menor probabilidad de ser recomendados para entrevistas. Esta discriminación algorítmica ilustra cómo los sesgos sistémicos pueden ser amplificados cuando se automatizan decisiones a gran escala. En su investigación, Hickok et al. (2022) destacan la importancia de realizar auditorías exhaustivas que aborden no solo los resultados finales, sino también la forma en que los datos son procesados y transformados a lo largo de todo el pipeline algorítmico (Hickok et al., 2022). A partir de los hallazgos expuestos, es importante incluso reiterativo resaltar que la problemática de los sesgos en los algoritmos de contratación no solo se limita a la equidad en el acceso a oportunidades laborales, sino que también plantea preguntas fundamentales sobre la responsabilidad ética de las organizaciones que implementan estos sistemas. La creciente dependencia de herramientas automatizadas en la toma de decisiones de contratación ha generado un debate en torno a la necesidad de un marco regulatorio más robusto que aborde las implicaciones éticas y legales del uso de inteligencia artificial en procesos de selección. Como sugieren Binns (2018) y Mittelstadt et al. (2016), es fundamental que las empresas adopten prácticas de gobernanza algorítmica que incluyan auditorías continuas y la participación de diversas partes interesadas para garantizar la transparencia y la responsabilidad en el uso de estos sistemas.

Asimismo, la literatura reciente destaca la importancia de incorporar enfoques interdisciplinarios en la auditoría de algoritmos. La colaboración entre expertos en ética, tecnología y derecho es crucial para diseñar auditorías que no solo identifiquen sesgos técnicos, sino que también evalúen el impacto social y ético de los algoritmos en contextos específicos. Estos enfoques holísticos permiten una comprensión más profunda de cómo los algoritmos pueden afectar a diferentes grupos sociales, y facilitan la creación de soluciones que aborden no solo las causas inmediatas del sesgo, sino también las dinámicas estructurales que lo perpetúan.

Otra consideración clave es el papel que juega la transparencia en la mitigación de sesgos algorítmicos. La opacidad de los algoritmos de caja negra, es decir, aquellos cuyos procesos internos no son transparentes ni comprensibles para los usuarios, dificulta la detección y corrección de sesgos. Según Diakopoulos (2016), la falta de transparencia en los sistemas automatizados puede erosionar la confianza en los procesos de contratación, especialmente si los afectados no tienen acceso a los mecanismos de apelación o corrección de decisiones injustas (Diakopoulos, 2016). En este sentido, iniciativas que promuevan la explicabilidad y la accesibilidad de los algoritmos, como las defendidas por Wachter, Mittelstadt y Floridi (2017), son fundamentales para garantizar que los candidatos puedan entender y cuestionar las decisiones automatizadas que afectan su acceso a oportunidades laborales (Wachter, Mittelstadt & Floridi, 2017).

Es importante considerar que la auditoría algorítmica no debe ser vista como una solución única o definitiva, sino como parte de un enfoque más amplio de gobernanza algorítmica. Como sugieren Lepri et al. (2018), la gobernanza algorítmica efectiva debe incluir no solo auditorías técnicas, sino también la implementación de políticas organizacionales que promuevan la diversidad y la inclusión en todas las etapas del proceso de contratación, desde la recopilación de datos hasta la evaluación final de candidatos (Lepri et al., 2018). Además, es crucial que estas políticas sean respaldadas por un compromiso organizacional a largo plazo con la equidad, lo cual requiere la participación activa de todos los niveles de la organización, desde los desarrolladores de algoritmos hasta los responsables de la toma de decisiones estratégicas.

Por todo ello, la auditoría de algoritmos de contratación, como se presenta en este estudio de caso, pone de manifiesto la necesidad de enfoques multifacéticos y colaborativos para abordar los sesgos inherentes en los sistemas automatizados. Las organizaciones deben adoptar una postura proactiva, no solo en la identificación y corrección de sesgos, sino también en la creación de entornos laborales más inclusivos y equitativos que reflejen un compromiso genuino con la justicia social en la era digital.

Otro ejemplo o caso de estudio es el proporcionado por Mensah (2023), el cual analiza cómo los algoritmos de contratación pueden perpetuar sesgos en la selección de candidatos. En este estudio, se destaca que incluso con múltiples intervenciones técnicas, los sesgos de género y etnia prevalecen debido a patrones históricos y la opacidad algorítmica. La investigación resalta la importancia de las auditorías regulares para identificar estos sesgos, sugiriendo que las auditorías externas independientes son esenciales para garantizar la transparencia y la equidad en los procesos automatizados. Este caso ofrece una visión crítica sobre cómo, a pesar de los avances técnicos, la auditoría y la transparencia son componentes fundamentales para abordar la discriminación en la contratación automatizada.

Mensah (2023) proporciona un marco integral para entender los desafíos y limitaciones que enfrentan estas herramientas tecnológicas. En su estudio, se examinan los efectos de sesgos persistentes en algoritmos utilizados por empresas para la selección de candidatos. A pesar de los esfuerzos para mitigar estos sesgos mediante técnicas de corrección, como la modificación de vectores en modelos de aprendizaje profundo, los resultados muestran que los sesgos de género y etnia continúan manifestándose debido a patrones históricos reflejados en los datos.

Un aspecto central que se destaca es la necesidad de abordar la opacidad algorítmica, conocida como el “problema de la caja negra”. Este término hace referencia a la dificultad de comprender los procesos internos de decisión de los algoritmos, lo que complica la identificación y corrección de sesgos. En el caso estudiado, se observó que el algoritmo favorecía sistemáticamente a candidatos masculinos para roles técnicos, lo cual está directamente relacionado con la predominancia histórica de hombres en dichos puestos. Este sesgo estructural se ve agravado cuando los sistemas de IA replican patrones discriminatorios al priorizar ciertas características demográficas en detrimento de otras.

Además, el estudio evidencia que la discriminación algorítmica no se limita únicamente al género, sino que también afecta a candidatos de minorías étnicas. En este contexto, se identificó que los candidatos con nombres asociados a minorías tenían una menor probabilidad de ser seleccionados para entrevistas, incluso cuando sus cualificaciones eran comparables a las de otros candidatos. Esto subraya cómo los prejuicios sistémicos existentes en la sociedad pueden ser amplificados cuando se automatizan decisiones a gran escala sin un monitoreo adecuado. Mensah (2023) argumenta que la solución no solo radica en ajustes técnicos dentro del algoritmo, sino también en la implementación de auditorías regulares y exhaustivas, tanto internas como externas. Estas auditorías deben ir más allá de la revisión superficial de los resultados finales e incluir un análisis crítico de cómo los datos son procesados y transformados a lo largo del ciclo de vida algorítmico. Esto incluye la evaluación de la calidad de los datos de entrenamiento, la transparencia en la modelización, y la capacidad del sistema para proporcionar explicaciones comprensibles sobre sus decisiones.

Una conclusión importante derivada del estudio es que la transparencia, aunque esencial, no es suficiente si no se acompaña de explicabilidad y accesibilidad. La capacidad de explicar de manera clara y comprensible las decisiones automatizadas es crucial para fomentar la confianza y permitir que tanto los candidatos como los reguladores puedan entender y cuestionar los resultados. Sin estos mecanismos, la transparencia corre el riesgo de ser superficial, limitándose a la apertura de datos y códigos que son inaccesibles para la mayoría de los usuarios.

El caso de estudio también destaca la importancia de un enfoque interdisciplinario en la auditoría de algoritmos. Mensah sugiere que la colaboración entre expertos en tecnología, ética y derecho es esencial para diseñar marcos de auditoría que no solo aborden los aspectos técnicos, sino también las implicaciones sociales y éticas de los sistemas de IA. Este enfoque permite una comprensión más holística de cómo los algoritmos impactan a diferentes grupos sociales, lo que es fundamental para desarrollar soluciones que aborden las causas profundas del sesgo, más allá de los síntomas inmediatos.

El artículo de Mensah (2023) ejemplifica cómo la auditoría de algoritmos de contratación no solo debe enfocarse en la corrección de sesgos técnicos, sino también en la creación de un entorno de gobernanza algorítmica que integre la transparencia, la explicabilidad y la equidad en cada etapa del proceso. La implementación de auditorías regulares y la adopción de un enfoque holístico y multidisciplinario son esenciales para garantizar que los sistemas de IA en la contratación laboral no perpetúen ni amplifiquen las desigualdades existentes en la sociedad.

6. Análisis crítico. Discusión

El análisis de este caso revela una serie de cuestiones éticas y técnicas que deben abordarse para mejorar la transparencia y equidad de los sistemas de IA en la contratación laboral. En primer lugar, es evidente que la transparencia no se limita a la explicabilidad del algoritmo. Como subraya Mashhadi et al. (2022), la incorporación de herramientas de visualización de equidad puede ayudar a los diseñadores de sistemas y auditores a identificar patrones de sesgo que, de otro modo, podrían pasar desapercibidos. Sin embargo, la transparencia no siempre garantiza la equidad, ya que un sistema puede ser transparente en sus operaciones y aun así producir resultados sesgados.

Un aspecto crítico es el acceso limitado a la información sobre cómo se desarrollan y entrena estos algoritmos. Percy et al. (2021) argumentan que la rendición de cuentas en los sistemas de IA debe ser respaldada no solo por auditorías internas, sino también por procesos formales de acreditación externa que puedan garantizar la imparcialidad. Este nivel de escrutinio es particularmente relevante en sectores como la contratación, donde las decisiones algorítmicas pueden afectar la vida de los individuos de manera directa y a largo plazo. El caso pone de manifiesto la necesidad de diseñar modelos que no solo sean precisos, sino

que también sean justos. Según Xiang et al. (2022), es fundamental equilibrar la precisión y la equidad en los modelos predictivos, especialmente cuando se utilizan en contextos sensibles como la contratación o la educación. Esto implica no solo ajustar los modelos para reducir los sesgos, sino también repensar las métricas que se utilizan para evaluar su rendimiento.

Por otro lado, la capacidad de los sistemas de IA para corregir sesgos depende en gran medida de la calidad de los datos utilizados. Los datos de entrenamiento sesgados pueden llevar a resultados discriminatorios, y corregir estos sesgos después de que el modelo ha sido implementado es extremadamente difícil. Este problema se agrava cuando los algoritmos se entrena utilizando datos históricos que reflejan prejuicios existentes en la sociedad. Como se muestra en el estudio de Albaroudi et al. (2024), las intervenciones técnicas como la generación de datos sintéticos o la corrección de sesgos en el espacio vectorial pueden reducir el impacto de estos sesgos, pero no eliminarlos por completo. Esta auditoría destaca la importancia de un enfoque más integral en la gobernanza algorítmica. Los resultados sugieren que, además de las correcciones técnicas, se requiere una supervisión regulatoria robusta que exija transparencia y responsabilidad en cada etapa del ciclo de vida del algoritmo. Bartley et al. (2021) demuestran cómo las auditorías algorítmicas pueden ser una herramienta eficaz para identificar y corregir sesgos en plataformas digitales, pero subrayan la necesidad de implementar mecanismos de corrección continua para adaptarse a los cambios en los patrones de uso y en los datos disponibles.

El estudio de caso aquí presentado subraya la complejidad de garantizar la transparencia y la equidad en los sistemas de IA. A pesar de los avances en técnicas de auditoría y mitigación de sesgos, los sesgos algorítmicos siguen siendo un desafío persistente. Para abordar este problema, es esencial combinar soluciones técnicas con una gobernanza más sólida y enfoques regulatorios que promuevan la transparencia y la rendición de cuentas. Esto incluye no solo la implementación de auditorías exhaustivas, sino también la adaptación constante de las normativas para garantizar que los sistemas se mantengan alineados con los valores éticos y los derechos fundamentales.

El análisis de este caso revela una serie de cuestiones éticas y técnicas que deben abordarse para mejorar la transparencia y equidad de los sistemas de IA en la contratación laboral. En primer lugar, es evidente que la transparencia no se limita a la explicabilidad del algoritmo. Como subraya Mashhadi et al. (2022), la incorporación de herramientas de visualización de equidad puede ayudar a los diseñadores de sistemas y auditores a identificar patrones de sesgo que, de otro modo, podrían pasar desapercibidos. Sin embargo, la transparencia no siempre garantiza la equidad, ya que un sistema puede ser transparente en sus operaciones y aun así producir resultados sesgados.

Un aspecto crítico es el acceso limitado a la información sobre cómo se desarrollan y entrena estos algoritmos. Percy et al. (2021) argumentan que la rendición de cuentas en los sistemas de IA debe ser respaldada no solo por auditorías internas, sino también por procesos formales de acreditación externa que puedan garantizar la imparcialidad. Este nivel de escrutinio es particularmente relevante en sectores como la contratación, donde las decisiones algorítmicas pueden afectar la vida de los individuos de manera directa y a largo plazo. El caso pone de manifiesto la necesidad de diseñar modelos que no solo sean precisos, sino que también sean justos. Según Xiang et al. (2022), es fundamental equilibrar la precisión y la equidad en los modelos predictivos, especialmente cuando se utilizan en contextos sensibles como la contratación o la educación. Esto implica no solo ajustar los modelos para reducir los sesgos, sino también repensar las métricas que se utilizan para evaluar su rendimiento.

Otro punto crítico en este análisis es el papel de la ética en el desarrollo y despliegue de sistemas de IA. A pesar de los esfuerzos por mitigar el sesgo, la falta de un marco ético robusto puede llevar a soluciones parciales que no aborden las raíces del problema. Cath (2018) sostiene que la ética algorítmica debe integrarse desde las primeras etapas de desarrollo del sistema, con un enfoque que combine consideraciones técnicas y normativas para abordar de manera efectiva las cuestiones de equidad (Cath, 2018). Esta integración de la ética en el ciclo de vida del desarrollo de IA podría ayudar a mitigar los riesgos asociados con decisiones algorítmicas injustas.

La incorporación de herramientas tecnológicas en la toma de decisiones gubernamentales modifica significativamente los procesos tradicionales de gestión pública. Entre las principales implicaciones destacan la reducción de tiempos en la toma de decisiones, el acceso a información más precisa y la posibilidad de realizar simulaciones de escenarios para evaluar políticas antes de su implementación. No obstante, estos avances también plantean retos, como la dependencia excesiva de modelos automatizados, la posible pérdida de control humano sobre decisiones críticas y la necesidad de garantizar la trazabilidad de los procesos decisionales. En este contexto, es imprescindible establecer mecanismos de supervisión y validación que permitan evaluar el impacto real de estas tecnologías en la gestión gubernamental, asegurando que su uso esté alineado con los principios democráticos y los derechos ciudadanos.

Por último, el análisis pone de relieve la importancia de la adaptabilidad en los sistemas de IA para garantizar que las soluciones a los problemas de sesgo no se vuelvan obsoletas a medida que los datos y contextos sociales evolucionan. Según Holstein et al. (2019), los sistemas de IA deben diseñarse con mecanismos que permitan su ajuste continuo en respuesta a cambios en los datos o patrones de uso, lo que es esencial para mantener la equidad a largo plazo (Holstein et al., 2019). Esto sugiere que, además de la corrección de sesgos existentes, es vital que los sistemas algorítmicos sean inherentemente flexibles y capaces de evolucionar en paralelo con la sociedad a la que sirven.

7. Conclusiones

El presente artículo analiza la relevancia crucial de implementar auditorías de sesgo en el marco de la gobernanza algorítmica, abordando las dimensiones tecnológicas, regulatorias y éticas involucradas. En un contexto donde la automatización y las decisiones algorítmicas afectan directamente la vida de los ciudadanos, es esencial no solo considerar los beneficios de la inteligencia artificial (IA), sino también evaluar los riesgos y desigualdades que estos sistemas pueden perpetuar o agravar.

La creciente dependencia de algoritmos en la administración pública y en la toma de decisiones empresariales plantea desafíos que van más allá de la eficiencia técnica. La opacidad característica de muchos sistemas de IA, en particular aquellos basados en aprendizaje profundo, dificulta la trazabilidad y la detección de sesgos. Como se demuestra en este análisis, los algoritmos no son herramientas neutrales; reflejan y amplifican dinámicas sociales y prejuicios inherentes a los datos con los que son entrenados. En este sentido, las auditorías de sesgo no deben considerarse opcionales, sino imprescindibles para garantizar una IA justa y equitativa.

Un aspecto central de la gobernanza algorítmica es la necesidad de una regulación proactiva y preventiva. Normativas como el Reglamento de Inteligencia Artificial (AI Act) en Europa buscan establecer un marco integral para gestionar los riesgos asociados al uso de IA, especialmente en áreas críticas como salud, educación y justicia. Sin embargo, la implementación efectiva de estas regulaciones enfrenta desafíos considerables, especialmente debido a la heterogeneidad de los contextos nacionales en la Unión Europea, lo que complica la armonización y aplicación uniforme de las normativas.

A pesar de su importancia, las auditorías de sesgo tienen limitaciones. Como se señala en los estudios de caso presentados, su efectividad depende de la calidad de los datos, la transparencia de los algoritmos y la colaboración entre los actores involucrados. Los sesgos pueden surgir en distintas etapas del ciclo de vida del sistema, desde la recolección de datos hasta su implementación, lo que requiere un enfoque holístico que contemple tanto aspectos técnicos como sociales.

Una lección clave extraída de este análisis es que la transparencia, aunque fundamental, no es suficiente por sí sola. Para que sea efectiva, debe ir acompañada de mecanismos de explicabilidad que permitan comprender las decisiones tomadas por los sistemas de IA. Además, la transparencia debe ser accesible y procesable para el público general, lo que implica no solo abrir el acceso a los algoritmos y datos, sino también asegurar la comprensibilidad de la información. Herramientas de explicabilidad y mecanismos de participación ciudadana pueden desempeñar un rol crucial en la promoción de una gobernanza inclusiva y democrática.

Este artículo también subraya la importancia de la responsabilidad compartida entre los distintos actores en la gobernanza algorítmica. Desde desarrolladores hasta responsables políticos y organismos de auditoría, todos tienen un papel en la construcción de sistemas de IA éticos. Este enfoque colaborativo no solo fomenta la responsabilidad, sino también la innovación y la mejora continua en el diseño de algoritmos inclusivos.

En cuanto a la regulación futura, se espera que la legislación europea avance hacia la regulación de aspectos emergentes como la generación de datos sintéticos y la integración de la IA explicable (XAI). Estos desarrollos son esenciales para garantizar la auditabilidad y comprensibilidad de los sistemas. Asimismo, la regulación deberá abordar el desafío de los sistemas completamente autónomos en sectores críticos, como la banca. La adopción de estándares comunes y la promoción de la interoperabilidad entre sistemas serán clave para el éxito de la gobernanza algorítmica a nivel europeo y global.

Un aspecto crítico a destacar es la naturaleza dinámica de los sesgos algorítmicos. A medida que los datos evolucionan y los modelos se ajustan, los sesgos también cambian, lo que subraya la necesidad de auditorías continuas y mecanismos de corrección en tiempo real. Este enfoque es particularmente relevante en contextos donde las decisiones algorítmicas impactan significativamente la vida de las personas, como en la selección de personal, la administración de justicia o la concesión de créditos. La gobernanza algorítmica, por tanto, debe ser un proceso dinámico y adaptable a los cambios tecnológicos y sociales.

El estudio de caso presentado evidencia que incluso en sectores con regulaciones estrictas, como el de la contratación laboral, persisten sesgos, lo que resalta la necesidad de enfoques rigurosos y multidisciplinarios en el diseño y evaluación de algoritmos. No se trata solo de ajustar modelos, sino de reconsiderar las métricas de éxito para incluir dimensiones de equidad y justicia social. En este sentido, los desarrolladores deben adoptar una perspectiva ética desde las etapas más tempranas del desarrollo.

A nivel global, la UE tiene la oportunidad de liderar la creación de estándares internacionales en gobernanza algorítmica, influyendo en cómo otras regiones abordan la transparencia y equidad en sistemas de IA. Sin embargo, este liderazgo requerirá un diálogo inclusivo que integre a gobiernos, sociedad civil, sector privado y academia. Solo a través de una cooperación internacional sólida se podrán establecer marcos regulatorios inclusivos y adaptados a la diversidad de contextos globales.

La incorporación de herramientas tecnológicas en la toma de decisiones gubernamentales modifica significativamente los procesos tradicionales de gestión pública. Entre las principales implicaciones destacan la reducción de tiempos en la toma de decisiones, el acceso a información más precisa y la posibilidad de realizar simulaciones de escenarios para evaluar políticas antes de su implementación. No obstante, estos avances también plantean retos, como la dependencia excesiva de modelos automatizados, la posible pérdida de control humano sobre decisiones críticas y la necesidad de garantizar la trazabilidad de los procesos decisionales. En este contexto, es imprescindible establecer mecanismos de supervisión y validación que

permitan evaluar el impacto real de estas tecnologías en la gestión gubernamental, asegurando que su uso esté alineado con los principios democráticos y los derechos ciudadanos.

Por todo ello, la gobernanza algorítmica efectiva debe ir más allá de la revisión técnica y considerar un enfoque multidimensional que incorpore la evaluación continua del impacto social y contextual de las decisiones automatizadas. Este enfoque debe ser iterativo y adaptativo, respondiendo a los cambios tecnológicos y a las dinámicas sociales emergentes.

El camino hacia una gobernanza algorítmica verdaderamente transparente y equitativa requiere la implementación de auditorías rigurosas, la adopción de normativas dinámicas y la colaboración entre múltiples actores. Si bien los desafíos son significativos, también lo son las oportunidades para construir sistemas de IA que respeten los derechos fundamentales y promuevan una sociedad más justa. En un mundo cada vez más digitalizado, la equidad y la responsabilidad deben estar en el centro de cada etapa del proceso algorítmico.

8. Bibliografía

- Aguilar Villanueva L. F. (2014). Las dimensiones y los niveles de gobernanza. Cuadernos de Gobierno y Administración Pública, 1(1), 11-36. https://doi.org/10.5209/rev_CGAP.2014.v1.n1.45156
- Ajunwa, I. (2019). "The Auditing Imperative for Automated Hiring", *Harvard Journal of Law & Technology*, 34 (forthcoming 2021). <https://ssrn.com/abstract=3437631> or <http://dx.doi.org/10.2139/ssrn.3437631>.
- Albaroudi, E., Mansouri, T. and Alameer, A. (2024). "A comprehensive review of AI techniques for addressing algorithmic bias in job hiring", *AI*, 5(1), pp. 383-404. <https://doi.org/10.3390/ai5010019>.
- Arellano Gault D. (2014). ¿Uno o varios tipos de gobernanza? Más allá de la gobernanza como moda: la prueba del tránsito organizacional. Cuadernos de Gobierno y Administración Pública, 1(2), 119-137. https://doi.org/10.5209/rev_CGAP.2014.v1.n2.47538
- Bartley, N., Abeliuk, A., Ferrara, E. and Lerman, K. (2021). "Auditing algorithmic bias on Twitter", *Proceedings of the 2021 Web Conference*. DOI: 10.1145/3447535.3462491.
- Beduschi, A. (2024). "Synthetic data protection: Towards a paradigm change in data regulation?", *Big Data & Society*, 11(1), p. 20539517241231277. DOI: 10.1177/20539517241231277.
- Binns, R. (2018). "Fairness in machine learning: Lessons from political philosophy", *Proceedings of the 2018 Conference on Fairness, Accountability, and Transparency*, <https://proceedings.mlr.press/v81/binns18a/binns18a.pdf>
- Bouderhem, R. (2024). "Shaping the future of AI in healthcare through ethics and governance", *Humanities and Social Sciences Communications*, 11, p. 416. DOI: <https://doi.org/10.1057/s41599-024-02894-w> .
- Busuioc, M., Curtin, D. and Almada, M. (2023). "Reclaiming transparency: contesting the logics of secrecy within the AI Act", *European Law Open*, 2(1), pp. 79-105. DOI: 10.1017/elo.2022.47.
- Cath, C. (2018). "Governing artificial intelligence: ethical, legal and technical opportunities and challenges", *Philosophical Transactions of the Royal Society A*, 376, p. 20180080. DOI: <http://dx.doi.org/10.1098/rsta.2018.0080>.
- Calero Valdez, A., Heine, M., Franke, T., Jochems, N., Jetter, H.-C. and Schrills, T. (2024). "The European commitment to human-centered technology: The integral role of HCI in the EU AI Act's success", *arXiv*, 2402.14728. DOI: <https://dx.doi.org/10.48550/arXiv.2402.14728> .
- Conitzer, V., Hadfield, G.K. and Vallor, S. (2022). "Technical perspective: The impact of auditing for algorithmic bias", *Communications of the ACM*, 65(12), pp. 40-44. DOI: 10.1145/3571152.
- Crespo-González J., Vicente Vicente M. J. ., Saudade Baltazar M. . y Valmorisco Pizarro S. . (2024). Profesorado universitario e ia: ¿una pareja bien avenida? : Reflexiones a partir de expectativas y percepciones del estudiantado. Cuadernos de Gobierno y Administración Pública, 11(2), e95972. <https://doi.org/10.5209/cgap.95972>
- Dekker, R., Koot, P., Birbil, S. and van Embden, A.M. (2022). "Co-designing algorithms for governance: Ensuring responsible and accountable algorithmic management of refugee camp supplies", SAGE. <https://doi.org/10.1177/20539517221087855>.
- Diakopoulos, N. (2016). "Accountability in algorithmic decision making", *Communications of the ACM*, 59(2), pp. 56-62. DOI: 10.1145/2844110.
- Engelmann, A. (2023). "Algorithmic transparency as a fundamental right in the democratic rule of law", *Brazilian Journal of Law, Technology and Innovation*, 1(2), pp. 169-188. DOI: <https://dx.doi.org/10.59224/bjlti.v1i2.169-188>.
- García Magariño S. (2016). Un cuestionamiento de los supuestos que vertebran la sociedad moderna en búsquedas de nuevos modelos de gobernanza: el gobierno, la sociedad civil y la empresa nadando en un mar de complejidad. Cuadernos de Gobierno y Administración Pública, 2(2), 187-211. https://doi.org/10.5209/rev_CGAP.2015.v2.n2.51572
- Giest, S. and Grimmelikhuijsen, S. (2020). "Introduction to special issue algorithmic transparency in government: Towards a multi-level perspective", *Information Polity*, 25(4), pp. 413-416. DOI: 10.3233/IP-200010.
- Giunchiglia, F., Otterbacher, J., Kleanthous, S., Batsuren, K., Bogina, V., Kuflik, T. and Shulner Tal, A. (2021). "Towards algorithmic transparency: A diversity perspective", *arXiv preprint arXiv:2104.05658*. DOI: 10.48550/arXiv.2104.05658.
- Groves, L., Metcalf, J., Kennedy, A., Vecchione, B. and Strait, A. (2024). "Auditing work: Exploring the New York City algorithmic bias audit regime", ACM. DOI: <https://doi.org/10.1145/3630106.3658959>.

- Gstrein, O., Haleem, N. and Zwitter, A. (2024). "General-purpose AI regulation and the European Union AI Act", *Internet Policy Review*, 3, pp. 1-15. DOI: <https://dx.doi.org/10.14763/2024.3.1790>.
- Guerrero García A. L. (2023). Cuando hablamos de capacidad burocrática ¿de qué hablamos? Una revisión teórica y empírica del concepto. *Cuadernos de Gobierno y Administración Pública*, 10(2), e83296. <https://doi.org/10.5209/cgap.83296>
- Gyevnar, B., Ferguson, N. and Schafer, B. (2023). "Bridging the transparency gap: What can explainable AI learn from the AI Act?", *Frontiers in Artificial Intelligence*, 10(3), pp. 233-245. <https://dx.doi.org/10.3233/FAIA230367>.
- Hickok, M., Dorsey, C., O'Brien, T., Baur, D., Ingram, K. and Chauhan, C. (2022). "Case study: The distilling of a biased algorithmic decision system through a business lens", *Open Science Framework*. DOI: 10.31235/osf.io/t5dhu.
- Holstein, K., Wortman, V., Daumé III, H., Dudik, M., and Wallach, H. (2019). "Improving fairness in machine learning systems: What do industry practitioners need?", *Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems*, p. 600. DOI: 10.1145/3290605.3300830.
- Kharitonova, Y., Malik, N.S. and Yang, T. (2023). "The legal issue of deterrence of algorithmic control of digital platforms: The experience of China, the European Union, Russia and India", *BRICS Law Journal*, 10(1), pp. 147-170. DOI: 10.21684/2412-2343-2023-10-1-147-170.
- Lam, K., Lange, B., Blili-Hamelin, B., Davidović, J., Brown, S. and Hasan, A. (2024). "A framework for assurance audits of algorithmic systems", *Proceedings of the ACM on Human-Computer Interaction*, 7(CSCW2), pp. 1-21. DOI: 10.1145/3630106.3658957.
- Laux, J., Wachter, S. and Mittelstadt, B. (2023). "Trustworthy artificial intelligence and the European Union AI Act: On the conflation of trustworthiness and the acceptability of risk", *Regulation & Governance*, 17(2), pp. 1-21. <https://dx.doi.org/10.1111/rego.12512>.
- Lepri, B., Oliver, N., Letouzé, E., Pentland, A. and Vinck, P. (2018). "Fair, transparent, and accountable algorithmic decision-making processes", *Philosophy & Technology*, 31(4), pp. 611-627. DOI: 10.1007/s13347-017-0279-x.
- Mashhadi, A., Zolyomi, A. and Quedado, J. (2022). "A case study of integrating fairness visualization tools in machine learning education", *CHI Conference on Human Factors in Computing Systems*. DOI: 10.1145/3491101.3503568.
- Mensah, G.B. (2023). *Artificial intelligence and ethics: A comprehensive review of bias mitigation, transparency, and accountability in AI systems*. https://www.researchgate.net/publication/375744287_Artificial_Intelligence_and_Ethics_A_Comprehensive_Review_of_Bias_Mitigation_Transparency_and_Accountability_in_AI_Systems.
- Mittelstadt, B.D., Allo, P., Taddeo, M., Wachter, S. and Floridi, L. (2016). "The ethics of algorithms: Mapping the debate", *Big Data & Society*, 3(2), pp. 2053951716679679. DOI: 10.1177/2053951716679679.
- Percy, C., Dragicevic, S., Sarkar, S. and Garcez, A. (2021). "Accountability in AI: From principles to industry-specific accreditation", *AI Communications*, 34(4), pp. 431-445. DOI: 10.3233/AIC-210080.
- Qureshi, N.I., Choudhuri, S.S., Nagamani, Y., Varma, R. and Shah, R. (2024). "Ethical considerations of AI in financial services: Privacy, bias, and algorithmic transparency", *Proceedings of the 2024 IEEE International Conference on Knowledge Engineering and Computer Science (ICKECS)*. DOI: 10.1109/ICKECS61492.2024.10616483.
- Sovrano, F., Sapienza, S., Palmirani, M. and Vitali, F. (2022). "Metrics, explainability and the European AI Act proposal", *Journal of Future Studies*, 5(1), pp. 1-15. DOI: <https://dx.doi.org/10.3390/j5010010>.
- Springer, A. and Whittaker, S. (2021). "Making transparency clear: The dual importance of explainability and auditability", *Proceedings of the 24th International Conference on Intelligent User Interfaces (IUI '21)*. DOI: 10.1145/3447872.3447938.
- Turksen, U., Benson, V. and Adamyk, B. (2024). "Legal implications of automated suspicious transaction monitoring: Enhancing integrity of AI", *Journal of Banking Regulation*. DOI: 10.1057/s41261-024-00233-2.
- Wachter, S., Mittelstadt, B. and Floridi, L. (2017). "Transparent, explainable, and accountable AI for robotics", *Science Robotics*, 2(6), pp. eaan6080. DOI: 10.1126/scirobotics.aan6080.
- Wörsdörfer, M. (2023). "Mitigating the adverse effects of AI with the European Union's artificial intelligence act: Hype or hope?", *Journal of Organizational Ethics*, 10(4), pp. 1-13. <https://dx.doi.org/10.1002/joe.22238>.
- Xiang, F., Zhang, X., Cui, J., Carlin, M. and Song, Y. (2022). "Algorithmic bias in student success prediction models: Two case studies", *IEEE TALE Conference*. DOI: 10.1109/TALE54877.2022.00058.
- Zharova, A. (2023). "Achieving algorithmic transparency and managing risks of data security when making decisions without human interference: Legal approaches", *Journal of Digital Transformation and Law*. DOI: 10.21202/jdtl.2023.01.0014.

José Luis Bustelo Gracia

El Dr. José Luis Bustelo Gracia es profesor universitario especializado en Finanzas y Contabilidad, con una destacada trayectoria en la educación superior. Actualmente, imparte clases en la Universidad Isabel I, La Universidad Europea Miguel de Cervantes, y ESERP Digital Business and Law School, donde participa en programas de Máster en Asesoría Jurídica, Dirección y Planificación. Su formación académica incluye estudios en la Universidad de Barcelona, la Universidad Rey Juan Carlos y la Universidad Camilo José Cela y Universidad Oberta La Salle, con especialización en dirección financiera, banca y marketing. Su línea de investigación abarca la inteligencia artificial y su impacto en la gestión empresarial.