IA E INTEGRIDAD PÚBLICA: TRANSPARENCIA ALGORÍTMICA Y PROTECCIÓN DE DATOS EN EL COMBATE A LA CORRUPCIÓN

AI AND PUBLIC INTEGRITY: ALGORITHMIC TRANSPARENCY AND DATA PROTECTION IN THE FIGHT AGAINST CORRUPTION

Armando Hernández Cruz[1]

RESUMEN

El presente ensayo responde a la pregunta de cómo garantizar que el uso de inteligencia artificial en el combate a la corrupción fortalezca la integridad pública sin vulnerar los principios de transparencia y protección de datos personales que estructuran al Estado constitucional. A partir de una metodología cualitativa, analítico-normativa y comparada de alcance funcional, se examinan riesgos como la opacidad algorítmica, el desplazamiento de la responsabilidad y las posibles afectaciones a los derechos humanos, especialmente en materia de protección de datos personales. El análisis sostiene que la automatización no elimina la discrecionalidad, sino que la reconfigura en fases técnicas como la adquisición de datos, la formulación de variables,

ABSTRACT

This essay addresses the question of how to ensure that the use of artificial intelligence in anti-corruption policies strengthens public integrity without undermining the principles of transparency and personal data protection that structure the constitutional rule of law. Based on a qualitative, analytical-normative, and functionally comparative methodology, it examines risks such as algorithmic opacity, the displacement of responsibility, and potential impacts on human rights, particularly regarding personal data protection. The analysis argues that automation does not eliminate discretion but rather reconfigures it in technical stages such as data acquisition, variable formulation, business-rule specification, and extraction, transformation, and loading

1 Doctor en Derecho y en Ciencias Políticas y Sociales por la UNAM e investigador nacional nivel II. Sus líneas de trabajo comprenden integridad pública, derechos humanos, transparencia, combate a la corrupción e inteligencia artificial.

DOI: https://doi.org/10.5281/zenodo.21464987

RECIBIDO: 12 de mayo de 2026
ACEPTADO: 13 de julio de 2026

la definición de reglas de negocio y los procesos de extracción, transformación y carga (ETL). Como resultado, se propone un modelo de IA con integridad institucional, basado en finalidad pública expresa, transparencia verificable, auditoría algorítmica, gobernanza estricta de datos y responsabilidad humana indelegable. Se concluye que la inteligencia artificial solo puede contribuir legítimamente al combate a la corrupción si se somete a estándares reforzados de explicabilidad, control democrático, protección de datos y supervisión institucional diferenciada.

PALABRAS CLAVE: Combate a la corrupción, inteligencia artificial, integridad pública, transparencia algorítmica, protección de datos personales.

(ETL) processes. As a result, it proposes a model of "AI with institutional integrity" based on express public purpose, verifiable transparency, algorithmic auditing, strict data governance, and non-delegable human responsibility. It concludes that artificial intelligence can only legitimately contribute to the fight against corruption if it is subject to enhanced standards of explainability, democratic oversight, data protection, and differentiated institutional supervision.

KEYWORDS: Anti-corruption, artificial intelligence, public integrity, algorithmic transparency, personal data protection.

1. Introducción

La transformación digital del Estado ha dejado de ser una aspiración programática para convertirse en una realidad estructural. Ello ha redefinido la gestión pública y el ejercicio del poder mediante tecnologías de datos masivos, sistemas automatizados e inteligencia artificial (IA) (Organización para la Cooperación y el Desarrollo Económicos [OCDE], 2019). En este marco, la IA ha cobrado relevancia en la prevención y detección de la corrupción —contrataciones públicas, fiscalización automatizada y alertas tempranas— al prometer menor discrecionalidad, mayor trazabilidad y mejor rendición de cuentas (Ospina et al., 2024).

Sin embargo, su uso plantea interrogantes jurídicas e institucionales de gran calado. Aunque puede reducir la arbitrariedad, también puede generar opacidad estructural. La "caja negra" algorítmica dificulta comprender cómo se toman decisiones y bajo qué criterios (Brkan y Bonnet, 2020), lo que puede debilitar la transparencia y la responsabilidad (Pasquale, 2015). A ello se suma el uso intensivo de datos personales —incluidos datos sensibles—, que tensiona la eficacia institucional con la protección de derechos humanos, en particular el derecho a la protección de datos y el principio de proporcionalidad (Kuner, 2017).

Lo anterior configura una "sociedad del riesgo", término acuñado por Ulrich Beck

(1998) con el que se refiere a una distopía en la que el ser humano se enfrenta a una serie de riesgos derivados del avance tecnocientífico a nivel mundial. Desde esta perspectiva, el problema no es solo técnico o ético, sino constitucional e institucional: la integridad pública exige coherencia entre fines, medios y responsabilidad. El uso de herramientas no explicables o auditables puede sustituir viejas opacidades por formas tecnificadas de difícil control. En este contexto, surge la pregunta central: ¿cómo garantizar que el uso de inteligencia artificial en el combate a la corrupción fortalezca la integridad pública sin vulnerar los principios de transparencia y protección de datos personales que estructuran el Estado constitucional?

El análisis reconoce que el combate a la corrupción es un interés público de alta relevancia; no obstante, en un Estado democrático, los fines no justifican cualquier medio. Por ello, el uso de tecnologías debe regirse por principios de finalidad legítima, proporcionalidad, minimización de datos, transparencia verificable y rendición de cuentas, evitando decisiones automatizadas sin revisión efectiva.

Para responder a la pregunta de investigación, este ensayo adopta una metodología cualitativa, analítico-normativa y comparada de alcance funcional. Es analítico-normativa porque examina principios constitucionales, estándares internacionales y categorías de derecho público aplicables al uso de IA en políticas anticorrupción. Es comparada en sentido funcional, no exhaustivo, porque utiliza referentes internacionales y regionales —Organización de las Naciones Unidas para la Educación, la Ciencia y la Cultura (UNESCO, 2022), Unión Europea (2024), OCDE (2019, 2020, 2024) y Banco de Desarrollo de América Latina y el Caribe (CAF, 2024)— como parámetros para contrastar el caso mexicano. El anclaje normativo del ensayo se ubica principalmente en México, por ser el contexto institucional al que se dirige la propuesta; las referencias comparadas se emplean como criterios de contraste, no como una reconstrucción integral de derecho comparado.

Los criterios de selección de los referentes comparados fueron tres: primero, relevancia normativa para la gobernanza de IA confiable y de alto riesgo; segundo, utilidad para comprender el ciclo de datos en sistemas públicos; y tercero, pertinencia para políticas de integridad, contratación pública y rendición de cuentas. Por ello, se consideran estándares de UNESCO, OCDE y Unión Europea; referencias regionales del CAF sobre transformación digital y capacidades estatales en América Latina y experiencias empíricas vinculadas con contratación pública, detección de colusión y auditoría pública. Esta delimitación permite justificar por qué el texto mantiene un diálogo internacional, pero concentra su desarrollo normativo en el marco constitucional e institucional mexicano.

En suma, la discusión sobre IA en el sector público debe ir más allá de la eficiencia operativa e incorporarse al debate sobre integridad, legitimidad y control democrático del poder. La integridad pública en la era algorítmica no es automática, sino resultado de decisiones

normativas que subordinan la técnica a los principios constitucionales.

De tal forma que, para avanzar en el debate sobre los componentes que deben dar certeza a la eficiencia de la IA como herramienta del sector público, es necesario revisar algunos conceptos y nociones involucradas en su operación, como se hace en el siguiente apartado.

2. Marco teórico: integridad pública, corrupción y gobernanza algorítmica

El presente capítulo examina las nociones de integridad pública, la corrupción y la gobernanza algorítmica, reflexionando sobre los riesgos que acarrea el uso de inteligencia artificial en las dimensiones de transparencia y la responsabilidad institucional.

2.1 La integridad pública como categoría normativa reforzada

En la actualidad, el concepto de integridad pública se ha replanteado como una categoría normativa reforzada que exige coherencia estructural entre los fines públicos, los medios empleados y la responsabilidad institucional.

De esta forma, de acuerdo con la Oficina de las Naciones Unidas contra la Droga y el Delito (UNODC, 2019), la integridad pública se refiere al uso efectivo y honesto de los poderes y recursos confiados al sector público para fines colectivos. Esta noción se complementa con la definición de la Organización para la Cooperación y el Desarrollo Económicos (OCDE, 2017), que la describe como la alineación consistente y la adhesión a valores, principios y normas éticas compartidas para mantener y priorizar el interés público sobre los intereses privados. En este sentido, la integridad pública implica que los servidores públicos actúen como administradores temporales del poder y la autoridad conferidos por la comunidad, sustentando sus acciones en valores como la responsabilidad, la imparcialidad, la justicia y la equidad (UNODC, 2019).

En el contexto contemporáneo, la integridad pública también debe contemplar nuevos componentes como la gobernanza algorítmica y el uso de la IA en el sector público. Un modelo de integridad institucional en entornos automatizados exige que la eficacia tecnológica esté subordinada a estándares de transparencia, explicabilidad y protección de datos personales. La integridad en este ámbito implica que ninguna decisión pública con efectos jurídicos sea atribuida exclusivamente a un sistema automatizado, garantizando siempre una responsabilidad humana indelegable que pueda justificar y motivar el acto administrativo frente a la ciudadanía (Macias, 2025).

2.2 Corrupción, discrecionalidad y opacidad estructural

En el entorno institucional, la corrupción se entiende como la "Acción u omisión de un servidor público o de una persona natural o jurídica del sector privado que usa y abusa de su poder para favorecer intereses particulares a cambio de una recompensa, dañando el interés público y/o el de la entidad en la que labora" (Rivera, 2012, p. 16). Esta definición subraya que

tanto la acción como la omisión pueden desencadenar la corrupción, y que sus efectos impactan tanto al interés general como a las propias instituciones.

Por su parte, Moya y Paillama (2023) advierten que la corrupción prospera en contextos de amplia discrecionalidad, baja supervisión y opacidad decisional, donde las decisiones complejas no motivadas ni controladas generan zonas grises propicias para malas prácticas. En este contexto, los sistemas de IA prometen reducir márgenes de discrecionalidad mediante criterios automatizados de análisis y evaluación, permitiendo detectar patrones anómalos, sobreprecios o redes de colusión en ámbitos como contrataciones públicas y fiscalización.

No obstante, la sustitución parcial del juicio humano no elimina la opacidad, sino que puede transformarla. Prueba de ello es la caja negra algorítmica (Brkan y Bonnet, 2020), que introduce nuevas formas de opacidad estructural: la decisión deja de ser discrecional en sentido clásico, pero tampoco resulta plenamente comprensible.

De ello, nuevamente se deriva el riesgo central: el desplazamiento de la responsabilidad política. Aunque formalmente la autoridad decide, en la práctica la decisión se apoya en sistemas cuya lógica no es transparente, lo que desplaza la pregunta de quién decide hacia quién responde cuando interviene un algoritmo.

2.3 Gobernanza algorítmica y desplazamiento de la responsabilidad

La expansión de sistemas algorítmicos en la gestión pública ha dado lugar a lo que la literatura denomina gobernanza algorítmica: procesos de regulación, evaluación y asignación de recursos mediadas por modelos automatizados de procesamiento de datos. Es necesario subrayar que estos sistemas no son neutrales; incorporan supuestos, variables seleccionadas, ponderaciones y criterios de optimización que reflejan decisiones humanas previas.

En este escenario aparece el riesgo de una gobernanza algorítmica opaca. La literatura ha utilizado expresiones como “algoritmocracia” para describir el peligro de que sistemas automatizados adquieran un peso excesivo en decisiones políticas, sociales o económicas (Avaro y Sánchez, 2021). En el presente ensayo, el término no se emplea para afirmar que los algoritmos sustituyan actualmente y de manera generalizada a jueces, legisladores o funcionarios, sino como una categoría de advertencia: el riesgo de que la autoridad pública traslade indebidamente su responsabilidad hacia sistemas técnicos cuya operación no puede explicar ni controlar de forma suficiente.

Sin embargo, la complejidad técnica y la especialización requerida para comprender su funcionamiento generan una distancia cognitiva entre quienes diseñan los sistemas, quienes los aplican y quienes resultan afectados por sus decisiones. Esta distancia facilita un fenómeno preocupante: la tendencia a atribuir los resultados al “sistema” en lugar de a las autoridades que decidieron implementarlo.

Desde una perspectiva de ética pública, el problema no radica en que la inteligencia artificial participe en procesos decisionales, sino en que su uso genere dilución de responsabilidad. Si una decisión administrativa apoyada en IA produce un daño o afecta derechos humanos, debe existir una autoridad o mecanismo identificable que asuma la obligación de explicar, justificar y, en su caso, corregir el acto. La integridad institucional exige que la técnica permanezca subordinada a la responsabilidad humana (UNODC, 2019). Una vez abordadas las nociones de integridad pública, corrupción y gobernanza algorítmica en el entorno de la IA, se hará una aproximación a esta tecnología desde su perspectiva funcional como herramienta anticorrupción.

3. La IA en el sector público como herramienta anticorrupción: potencial y límites estructurales

La utilización de sistemas de inteligencia artificial en el sector público se ha mostrado como una innovación capaz de fortalecer las políticas de integridad institucional y reducir espacios propicios para prácticas corruptas. Ospina et al. (2024) atribuyen lo anterior a la capacidad de estas tecnologías para procesar grandes volúmenes de datos, identificar patrones anómalos y generar alertas tempranas; funciones que, sin duda, son de utilidad en áreas tradicionalmente vulnerables, como las contrataciones públicas, la fiscalización presupuestaria y la asignación de subsidios (OCDE, 2019).

Aunado a lo anterior, la IA se presenta como instrumento de prevención estructural: no solo detecta actos consumados, sino que permite identificar riesgos sistémicos. Esta naturaleza preventiva resulta particularmente relevante en modelos contemporáneos de integridad pública, que privilegian la gestión de riesgos sobre la reacción ex post (OCDE, 2020).

Sin embargo, la eficacia técnica no puede evaluarse aisladamente de su contexto normativo. La automatización de procesos no elimina la discrecionalidad; la redistribuye. El diseño del modelo, la selección de variables y la definición de umbrales de riesgo implican decisiones humanas que a su vez integran criterios valorativos y prioridades institucionales (Pasquale, 2015). La aparente neutralidad técnica puede ocultar opciones normativas que no siempre son explícitas ni deliberadas públicamente.

3.1 Automatización y reducción de discrecionalidad: ¿mito o realidad?

Uno de los argumentos más frecuentes a favor del uso de IA en el combate a la corrupción sostiene que la automatización reduce la discrecionalidad administrativa y, con ello, los incentivos para conductas indebidas (Transparencia Internacional, 2025). En efecto, la estandarización de procesos y la aplicación uniforme de criterios pueden limitar espacios de arbitrariedad.

No obstante, la literatura sobre gobernanza algorítmica advierte que la automatización no elimina la discrecionalidad, sino que la desplaza hacia etapas previas del diseño del sistema (Floridi et al., 2018). Desde una perspectiva técnico-institucional, esa transferencia ocurre en fases como la adquisición o recolección de datos, la

formulación de variables, la selección de datos de entrenamiento, la definición de categorías y la especificación de reglas de negocio aplicadas en los procesos de extracción, transformación y carga (ETL). La discrecionalidad ya no es visible en el acto administrativo concreto, pero permanece inscrita en la arquitectura de datos, en los criterios de clasificación y en los umbrales de riesgo que condicionan el resultado final. En términos operativos, un sistema anticorrupción no solo "recolecta datos" en sentido genérico. Requiere una cadena técnica de adquisición, depuración, transformación, integración, almacenamiento seguro y actualización de información. En cada fase pueden incorporarse decisiones con efectos jurídicos: qué fuentes se admiten, qué datos se excluyen, cómo se normalizan nombres de proveedores, cómo se identifican beneficiarios finales, cómo se resuelven duplicidades y qué reglas de negocio convierten un dato en una señal de riesgo. Por ello, la integridad no debe evaluarse únicamente en la salida del algoritmo, sino en todo el ciclo de vida de los datos.

En este sentido, la denominada "caja negra" encierra la dificultad de reconstruir el proceso decisional completo (Gryz y Rojszczak, 2021). Cuando una autoridad pública adopta decisiones apoyadas en sistemas que no pueden ser plenamente explicados, la motivación administrativa se ve erosionada. Como Rubio (1993) refiere, el principio de legalidad exige que las decisiones públicas sean fundadas y motivadas; si la lógica subyacente es inaccesible o incomprensible, la exigencia de justificación pierde densidad normativa.

3.2 Riesgos estructurales: opacidad, sesgo y afectación de derechos

Conviene destacar que, además de la opacidad, los sistemas algorítmicos pueden reproducir o amplificar sesgos presentes en los datos históricos. Si los datos utilizados reflejan desigualdades estructurales, prácticas discriminatorias o errores sistemáticos, el modelo tenderá a replicarlos (Barocas & Selbst, 2016). En contextos de fiscalización o asignación de beneficios, ello puede traducirse en impactos desproporcionados.

Como evidencia de ello, se puede mencionar el uso intensivo de datos personales. Calle et al. (2024) explican que los sistemas predictivos suelen requerir información detallada sobre comportamiento económico, historial administrativo o características socioeconómicas. Sin garantías adecuadas de minimización, finalidad y seguridad, el tratamiento masivo de datos puede derivar en afectaciones indebidas al derecho a la protección de datos personales y a la privacidad.

La técnica más adecuada para hacer efectivo el principio de minimización no consiste solamente en reducir ex post el volumen de datos tratados, sino en incorporar métodos de anonimización, disociación o seudonimización desde el origen de los datos. En sistemas anticorrupción, ello implica diseñar flujos que permitan calcular patrones de riesgo sin exponer innecesariamente identidades individuales, salvo cuando exista habilitación jurídica, necesidad acreditada y controles estrictos de acceso. La anonimización en la fuente, el cifrado, la separación de bases, los controles de acceso por rol, los registros

de consulta y la agregación estadística son mecanismos que permiten compatibilizar análisis predictivo con protección de datos personales.

En el ámbito del combate a la corrupción, estos riesgos adquieren una dimensión particular. La detección de riesgos o irregularidades puede implicar la elaboración de perfiles o la asignación de categorías de "alto riesgo", con consecuencias reputacionales y administrativas significativas. Si tales decisiones no cuentan con mecanismos claros de revisión y explicabilidad (Gryz y Rojszczak, 2021), se podría llegar a afectar el principio de debido proceso administrativo.

3.3 Integridad pública y responsabilidad institucional en entornos automatizados

Como se puede inferir, la incorporación de IA en políticas anticorrupción debe evaluarse a la luz del concepto de integridad pública; toda vez que la integridad no se reduce a la eficacia instrumental, sino que, como se ha visto, implica coherencia entre fines legítimos y medios compatibles con el Estado de derecho.

Desde esta perspectiva, el uso de sistemas algorítmicos para combatir la corrupción exige decisiones explicables, responsabilidad identificable, auditorías independientes y protección adecuada de datos personales. De lo contrario, puede generar opacidad y asimetrías de poder que debiliten la legitimidad democrática (Zuboff, 2019). Así, aunque la IA puede fortalecer políticas anticorrupción, requiere un marco institucional basado en transparencia, protección de datos y responsabilidad humana indelegable.

3.4 Experiencias empíricas y comparadas en contratación pública y detección de colusión

La utilidad de la IA anticorrupción se aprecia con mayor claridad cuando se examinan experiencias empíricas. En México, Aldana et al. (2022) propusieron un modelo de aprendizaje automático para identificar posibles contratos corruptos en datos de contratación pública, destacando que los predictores más relevantes no siempre se ubican en características aisladas del contrato, sino en relaciones entre compradores y proveedores. Esta evidencia es importante porque confirma que la integridad algorítmica no depende únicamente de procesar más datos, sino de construir variables relevantes, explicables y jurídicamente justificables.

En materia de colusión en licitaciones, Imhof y Wallimann (2021) han mostrado que las técnicas de aprendizaje automático pueden utilizarse para detectar patrones de *bid rigging* a partir de coaliciones de oferentes y señales estadísticas de comportamiento. Asimismo, estudios sobre detección de colusión mediante aprendizaje profundo han identificado la posibilidad de reconocer interacciones sospechosas entre participantes en concursos públicos (Huber y Imhof, 2021). Estas investigaciones muestran el potencial de la IA para fortalecer sistemas de alerta temprana; pero también evidencian el riesgo de falsos positivos y la necesidad de revisión humana antes de traducir una alerta en una consecuencia jurídica.

A nivel institucional, experiencias como los sistemas de contratación electrónica y monitoreo de compras públicas en Corea del Sur, así como el uso de tecnologías analíticas en órganos de auditoría y control en Brasil, confirman que la digitalización puede aumentar trazabilidad y reducir zonas de discrecionalidad; sin embargo, también muestran que la viabilidad del modelo depende de infraestructura de datos, calidad de registros, interoperabilidad, capacidades técnicas y controles jurídicos. Esta comparación funcional refuerza la tesis del ensayo: la IA puede fortalecer el combate a la corrupción, pero solo cuando se integra en un ecosistema institucional verificable, auditable y respetuoso de derechos.

4. La transparencia algorítmica como obligación reforzada de integridad pública

La advertencia planteada por Hans Jonas (1984) sobre la ampliación del radio de consecuencias de la acción tecnológica permite formular una regla básica para el sector público: cuanto mayor sea la capacidad de una herramienta para incidir en derechos, recursos públicos o trayectorias institucionales, mayor debe ser el estándar de responsabilidad exigible a la autoridad que la adopta. Bajo esta premisa, conviene examinar los factores de protección que permiten al Estado asegurar transparencia, control y rendición de cuentas en entornos automatizados.

4.1 De la transparencia clásica a la transparencia algorítmica

La transparencia ha sido concebida como un principio estructural del Estado democrático, orientado al acceso a la información, la publicidad de los actos y la rendición de cuentas (Diario Oficial de la Federación [DOF], 2025). En su enfoque clásico, implica dar a conocer decisiones, documentos y fundamentos que sustentan la actuación estatal.

Sin embargo, la incorporación de sistemas algorítmicos en la toma de decisiones públicas exige ampliar su alcance. Cuando intervienen modelos automatizados, la transparencia no puede limitarse al acto final; sino que debe extenderse a la lógica que guía el proceso decisional.

También está la cuestión de la "opacidad estructural" que debilita la rendición de cuentas (Pasquale, 2015). En el ámbito público, conlleva una afectación normativa; ya que el principio de motivación exige decisiones comprensibles y justificables; por ello, la transparencia algorítmica se configura como una exigencia constitucional.

Como se puede observar, las autoridades deben tener la capacidad de explicar, dentro de lo razonable, qué variables considera el sistema, cómo opera y cómo incide en la decisión final (Brkan y Bonnet, 2020), sin necesidad de revelar de manera irrestricta códigos fuente, secretos industriales o información que permita evadir controles anticorrupción. Lo relevante es garantizar un nivel de inteligibilidad compatible con el derecho a la información, la motivación administrativa y el debido proceso (Citron, 2008), lo que involucra el concepto de explicabilidad algorítmica.

4.2 Explicabilidad algorítmica, XAI y derecho a la justificación

El concepto de explicabilidad (*explainability*) ha adquirido relevancia en la literatura sobre ética de la inteligencia artificial como condición propia de una IA confiable (Floridi et al., 2018). En el contexto del sector público, la explicabilidad se vincula directamente con el derecho de las personas a comprender las decisiones que les afectan. En términos técnicos, esta exigencia puede articularse mediante enfoques de inteligencia artificial explicable o XAI (*Explainable Artificial Intelligence*), que buscan producir modelos, interfaces, registros y salidas interpretables para usuarios institucionales, órganos de control y personas afectadas.

Cuando una decisión administrativa se basa, total o parcialmente, en un sistema automatizado, la exigencia de motivación debe incorporar la explicación del papel desempeñado por dicho sistema. Si una persona es etiquetada como de "riesgo alto" en un modelo de detección de irregularidades, tiene derecho a conocer los criterios generales que sustentan esa clasificación; así como a impugnar posibles errores o sesgos.

En este escenario, la ausencia de explicabilidad o de mecanismos que la garanticen compromete no solo la transparencia, sino también el debido proceso. En el ámbito del combate a la corrupción, donde las decisiones pueden implicar investigaciones, sanciones o exclusiones de procedimientos, la falta de claridad sobre la lógica decisional puede acarrear afectaciones desproporcionadas. Por ello, la XAI no debe entenderse como un requisito meramente técnico, sino como una condición mínima de arquitectura institucional: todo sistema anticorrupción basado en IA debe prever explicaciones globales del modelo, explicaciones locales de sus resultados, registros de trazabilidad, documentación de variables y mecanismos que permitan a la autoridad revisar críticamente la recomendación algorítmica antes de convertirla en acto administrativo.

El Reglamento (UE) 2024/1689 sobre inteligencia artificial ofrece una referencia útil para esta discusión, pues exige medidas de supervisión humana, documentación, registros e información que permita interpretar los resultados de sistemas de alto riesgo. Aunque dicho Reglamento no es aplicable directamente en México, funciona como parámetro comparado para sostener que la transparencia algorítmica no puede reducirse a declaraciones abstractas de buena voluntad, sino que debe traducirse en documentación técnica, gestión de riesgos, registros de operación y mecanismos de intervención humana.

4.3 La auditoría algorítmica como instrumento de rendición de cuentas

De igual manera, la transparencia algorítmica no puede reducirse a la explicación ex post de decisiones individuales. Requiere mecanismos estructurales de auditoría que permitan evaluar el desempeño, los sesgos y la consistencia de los sistemas utilizados.

En este sentido, la implementación de evaluaciones de impacto algorítmico previas a la adopción de sistemas de IA constituye una práctica recomendada en diversas jurisdicciones comparadas. Estas

evaluaciones permiten identificar riesgos potenciales, analizar el tratamiento de datos personales y prever mecanismos de mitigación. En el contexto del combate a la corrupción, Hernández (2016) sostiene que la auditoría cumple una función doble: garantiza la eficacia del sistema y preserva la legitimidad institucional.

Para ello, la auditoría algorítmica puede adoptar diversas modalidades: interna, cuando es realizada por la propia institución; externa, cuando intervienen órganos de control independientes; o híbrida, cuando se combinan ambas (OCDE, 2019). En el ámbito de la integridad pública, la auditoría debe ser concebida como un mecanismo permanente; y no meramente ocasional.

Sin auditoría, la automatización puede convertirse en una zona de poder difícilmente controlable (Jurado y Armijo, 2022). La integridad pública exige que ningún instrumento decisional —humano o técnico— quede fuera del circuito de control.

4.4 Transparencia como condición de legitimidad anticorrupción

No entra en discusión que, el combate a la corrupción constituye un objetivo prioritario del Estado y un componente esencial del fortalecimiento institucional (Carbonell y Vázquez, 2009). Sin embargo, la legitimidad de las políticas anticorrupción no puede descansar exclusivamente en su eficacia. Debe apoyarse en la compatibilidad de los medios empleados con los principios constitucionales.

En este sentido, la transparencia algorítmica opera como condición de legitimidad. Una política de detección automatizada de irregularidades que no pueda ser explicada ni auditada corre el riesgo de generar desconfianza pública (Avaro y Sánchez, 2021), aun cuando produzca resultados cuantitativamente favorables.

En consecuencia, la transparencia algorítmica debe entenderse como una extensión del deber clásico de fundamentación y motivación de los actos administrativos. No basta con afirmar que "el sistema detectó una irregularidad"; es necesario explicar cómo y bajo qué criterios. La automatización no sustituye la obligación de justificar.

5. La protección de datos personales como límite constitucional al uso de IA en el combate a la corrupción

El derecho a la protección de datos personales representa una garantía autónoma dentro del constitucionalismo contemporáneo, vinculada a su vez a otros derechos (Araujo, 2009) y orientada a asegurar que el tratamiento de información se supedite a los principios de legalidad, finalidad específica, proporcionalidad y seguridad.

En el caso mexicano, este derecho encuentra reconocimiento expreso en el artículo 6°, apartado A, fracción II, de la Constitución Política de los Estados Unidos Mexicanos, que establece la obligación del Estado de garantizar la protección de datos personales y asegurar el ejercicio de los derechos de acceso, rectificación, cancelación y oposición (ARCO). Asimismo, el artículo

16 constitucional reconoce el derecho de toda persona a la protección de sus datos personales. Este mandato se desarrolla mediante la Ley General de Protección de Datos Personales en Posesión de Sujetos Obligados (LGPDPPSO), ordenamiento que regula el tratamiento de datos por parte de autoridades y entidades públicas bajo principios como licitud, finalidad, lealtad, consentimiento, calidad, proporcionalidad, información y responsabilidad.

En el contexto del uso de inteligencia artificial por el sector público, este derecho adquiere especial relevancia; toda vez que, en procesos anticorrupción, tales datos pueden incluir información fiscal, patrimonial, contractual o incluso características socioeconómicas de individuos y organizaciones. En este escenario cabe resaltar que, el riesgo en entornos algorítmicos radica en la tentación de incorporar el mayor volumen posible de datos bajo la premisa de que "más información produce mejores resultados".

Sin embargo, la doctrina ha advertido que el procesamiento automatizado intensivo de datos puede conllevar riesgos estructurales para los derechos humanos si no se establecen salvaguardas adecuadas (Floridi et al., 2018). En particular, el perfilamiento automatizado puede producir clasificaciones que afecten la reputación, el acceso a oportunidades o la apertura de investigaciones administrativas. Esta es la razón metodológica por la que el apartado se ancla en el marco mexicano: aunque el ensayo dialoga con estándares internacionales, su propuesta normativa se dirige a instituciones mexicanas sujetas a la Constitución y a la LGPDPPSO. Para ello, una de las principales salvaguardas debe ser el principio de finalidad.

En este punto también resulta relevante la Recomendación sobre la ética de la inteligencia artificial de la UNESCO, que reconoce como principios centrales la privacidad, la protección de datos, la transparencia, la explicabilidad, la responsabilidad, la rendición de cuentas y la supervisión humana. Su utilidad para el presente ensayo reside en que permite vincular la protección de datos personales con un enfoque de derechos humanos durante todo el ciclo de vida del sistema de IA, desde el diseño hasta la evaluación posterior al despliegue.

5.1 Finalidad y proporcionalidad en sistemas anticorrupción

El principio de finalidad exige que los datos personales sean recabados y utilizados para propósitos específicos, explícitos y legítimos (DOF, 2025). En políticas anticorrupción basadas en IA, su aplicación delimita qué información es necesaria para la detección de riesgos y evita tratamientos expansivos o indiscriminados. Además, el principio de licitud exige que la autoridad cuente con una base normativa suficiente para cada tratamiento; por tanto, reutilizar bases de datos recabadas para otros fines sin habilitación jurídica expresa puede convertir un sistema técnicamente eficaz en constitucionalmente problemático.

Por su parte, como Sánchez (2020) lo expone, el principio de proporcionalidad exige que el tratamiento de datos sea idóneo, necesario y equilibrado respecto al fin perseguido. No todo dato disponible debe ser utilizado por el sistema; solo aquél

estrictamente indispensable para cumplir la finalidad pública.

Desde la perspectiva de integridad pública, la proporcionalidad no constituye un obstáculo a la eficacia anticorrupción, sino un elemento de legitimidad institucional. Un sistema que vulnera derechos humanos compromete la confianza pública y debilita el objetivo que pretende fortalecer. Por ello, los sesgos en los datos de entrenamiento impactan directamente los principios de licitud y proporcionalidad: si el modelo utiliza variables irrelevantes, datos desactualizados o proxies discriminatorios, el tratamiento deja de ser necesario y equilibrado respecto del fin público que pretende alcanzar.

5.2 Perfilamiento, decisiones automatizadas y debido proceso

El uso de IA en el combate a la corrupción puede implicar la elaboración de perfiles de riesgo, la tipificación de sujetos como "casos prioritarios" o la generación de alertas automatizadas. Estas prácticas, si bien pueden resultar útiles para la gestión administrativa, deben ser compatibles con el derecho al debido proceso.

En este sentido, las decisiones automatizadas que producen efectos jurídicos relevantes deben estar sujetas a mecanismos de revisión humana significativa (Wachter et al., 2017). La intervención humana no puede limitarse a una validación meramente formal; debe permitir la evaluación crítica del resultado algorítmico.

En este punto converge la protección de datos personales con la responsabilidad institucional y la posibilidad de impugnar una decisión, que presupone comprender cómo se produjo. De esta forma, sin explicabilidad algorítmica y sin herramientas XAI que permitan identificar las variables relevantes, los pesos aproximados, las fuentes de datos y los criterios de clasificación, el derecho a la defensa pierde eficacia práctica.

5.3 Evaluaciones de impacto y supervisión independiente

Una herramienta cada vez más utilizada en el ámbito comparado es la evaluación de impacto en protección de datos (Comisión de Transparencia y Acceso a la Información del Estado de Nuevo León, 2018). Este recurso permite identificar riesgos antes de la implementación del sistema, establecer medidas de mitigación y documentar las decisiones por las que se opta.

En el caso de sistemas de IA utilizados con fines anticorrupción, como Morales (2020) sostiene, las evaluaciones de impacto deberían analizar los siguientes aspectos: a) Tipo de datos tratados, b) Posibles riesgos de discriminación, c) Grado de automatización decisional, d) Mecanismos de revisión y corrección y, e) Protocolos de seguridad.

La supervisión independiente (ya sea por órganos garantes de protección de datos o instancias de control interno) resulta indispensable para asegurar que la finalidad legítima del combate a la corrupción no derive en prácticas de vigilancia desproporcionadas.

No se debe perder de vista que, la protección de datos personales no es un obstáculo para

la integridad pública; es una condición de su legitimidad. La lucha contra la corrupción, aun siendo prioritaria, debesarrollarse dentro del marco de derechos compatible con el Estado constitucional; por lo que se propone el siguiente modelo dirigido a armonizar eficacia y normativa en el uso de la IA como estrategia anticorrupción.

6. Propuesta: Modelo de "IA con integridad institucional"

De acuerdo con lo expuesto a lo largo de estos párrafos, se puede sostener que, la incorporación de inteligencia artificial en el combate a la corrupción debe basarse en principios normativos claros, no solo justificarse a partir de criterios de eficiencia. De esta forma, se propone un modelo de "IA con integridad institucional" que articula tres ejes: transparencia verificable, protección de datos y responsabilidad humana indelegable, desarrollados en cinco principios o premisas.

La finalidad pública expresa exige que todo sistema se sustente en una justificación normativa clara, delimitando problema, función y límites. Esto evita expansiones indebidas ("function creep") y asegura que el uso tecnológico sea idóneo y necesario (Kuner, 2017).

La transparencia verificable implica que los sistemas sean comprensibles y auditables, mediante documentación, identificación de variables y trazabilidad. En este escenario, la explicabilidad funge como mecanismo tanto ético como jurídico, ya que permite justificar decisiones públicas (Floridi et al., 2018; Citron, 2008).

La auditoría algorítmica obligatoria debe ser continua, evaluando consistencia, sesgos, errores e impactos diferenciados. Esta supervisión garantiza tanto eficacia como legitimidad (OCDE, 2019).

La gobernanza estricta de datos requiere minimización, proporcionalidad y seguridad en el tratamiento de datos personales, evitando vigilancia excesiva y efectos discriminatorios (Barocas y Selbst, 2016).

La responsabilidad humana indelegable prohíbe decisiones completamente automatizadas en el ámbito público. Siempre debe existir una autoridad responsable, revisión humana y mecanismos de impugnación que eviten la dilución de responsabilidad (Pasquale, 2015).

Estos cinco principios no operan de forma aislada ni secuencial, sino como elementos interdependientes dentro de un esquema integral de gobernanza algorítmica. La finalidad pública expresa constituye el punto de partida porque delimita jurídicamente el objetivo legítimo del sistema y define el alcance del tratamiento de información. A partir de ello, la transparencia verificable permite conocer cómo el sistema transforma los datos en resultados y habilita condiciones de control externo e interno. La auditoría algorítmica funciona como mecanismo de supervisión permanente que verifica si el sistema continúa actuando conforme a su finalidad inicial y si genera efectos adversos o sesgos no previstos. Paralelamente, la gobernanza estricta de datos establece límites materiales sobre la información que puede utilizarse y las condiciones bajo

las cuales debe ser tratada. Finalmente, la responsabilidad humana indelegable actúa como garantía transversal que asegura que ninguna decisión con efectos jurídicos relevantes quede sustraída del control institucional y del derecho de revisión.

De esta manera, el modelo propuesto funciona como un sistema de controles recíprocos: la finalidad orienta, la transparencia permite observar, la auditoría verifica, la gobernanza de datos limita y la responsabilidad humana corrige y responde. La ausencia de cualquiera de estos componentes debilitaría la legitimidad del conjunto y aumentaría el riesgo de que la inteligencia artificial reproduzca prácticas incompatibles con el Estado constitucional de derecho.

6.1 Originalidad y valor agregado del modelo

El modelo de "IA con integridad institucional" dialoga con marcos previos como AI4People (Floridi et al., 2018), la Recomendación de la UNESCO (2022), los principios de la OCDE (2019, 2024) y el Reglamento (UE) 2024/1689. Sin embargo, su aportación no consiste en formular una nueva ética general de la inteligencia artificial, sino en traducir estos estándares al campo específico del derecho público anticorrupción. Mientras AI4People desarrolla principios amplios para una buena sociedad de IA, el modelo propuesto articula esos principios con categorías propias de la integridad pública: finalidad administrativa expresa, motivación del acto de autoridad, gobernanza de datos, auditoría algorítmica, control institucional y responsabilidad humana indelegable.

La originalidad de la propuesta radica en tres elementos. Primero, desplaza el centro del debate desde la eficiencia tecnológica hacia la legitimidad institucional. Segundo, incorpora el ciclo técnico de datos —adquisición, ETL, almacenamiento seguro, anonimización en origen y trazabilidad— como parte del estándar jurídico de integridad. Tercero, ofrece una ruta operativa para administraciones públicas mexicanas, distinguiendo entre obligaciones exigibles a órganos federales con mayor capacidad técnica y mecanismos escalonados para entidades subnacionales con menor infraestructura.

6.2 Transparencia diferenciada y dilema de transparencia adversarial

Uno de los problemas más complejos del modelo es el llamado dilema de transparencia adversarial. En políticas anticorrupción, una transparencia absoluta podría permitir que actores investigados conozcan las reglas de detección y ajusten su conducta para evadir el sistema. Por ello, la transparencia verificable no debe confundirse con publicidad irrestricta de todos los parámetros técnicos. El criterio de ponderación debe ser la transparencia diferenciada: publicidad suficiente para la ciudadanía, acceso ampliado para personas afectadas y acceso completo para órganos de control, auditoría y revisión jurisdiccional bajo deberes de confidencialidad.

Bajo esta lógica, el público debe conocer la finalidad del sistema, las categorías generales de datos tratados, la autoridad responsable, las salvaguardas y los resultados agregados de auditoría. Las personas afectadas deben recibir una explicación suficiente para ejercer defensa.

Los órganos de control deben acceder a documentación técnica completa, reglas de negocio, bitácoras, versiones del modelo, evaluaciones de impacto y pruebas de sesgo. Este diseño permite evitar la opacidad estatal sin volver "gameable" el sistema anticorrupción.

6.3 Operacionalización mínima e implementación escalonada

Para que el modelo no permanezca en el plano meramente normativo, puede traducirse en indicadores mínimos de cumplimiento. Estos indicadores no pretenden sustituir una regulación especializada, sino ofrecer una ruta de verificación institucional para órganos de transparencia, autoridades de protección de datos, órganos internos de control y sistemas anticorrupción.

La implementación debe considerar las asimetrías de capacidades institucionales existentes en México. No todas las instituciones cuentan con equipos de ciencia de datos, infraestructura segura, personal especializado o capacidad de auditoría algorítmica. Por ello, el modelo debe aplicarse de forma escalonada: en el nivel básico, mediante inventario de sistemas, responsable identificado, finalidad expresa y evaluación inicial de riesgos; en el nivel intermedio, mediante documentación técnica, anonimización en origen y auditorías periódicas; y en el nivel avanzado, mediante XAI, pruebas de sesgo, auditorías externas, trazabilidad completa y publicación de reportes agregados.

Principio	Indicador mínimo	Responsable sugerido	Periodicidad
Finalidad pública expresa	Acta o dictamen de justificación; base jurídica; delimitación de uso y exclusiones.	Área requirente y unidad jurídica.	Previo al diseño y actualización anual.
Transparencia verificable	Ficha pública del sistema; variables generales; documentación técnica reservada para control; bitácoras.	Unidad de transparencia y unidad técnica.	Publicación inicial y actualización por cambio relevante.
Auditoría algorítmica obligatoria	Pruebas de sesgo, precisión, falsos positivos, robustez y seguridad; reporte de hallazgos.	Órgano interno de control, auditor externo u órgano garante.	Al menos anual; extraordinaria ante cambios del modelo.
Gobernanza estricta de datos	Inventario de datos; anonimización/disociación en origen; controles de acceso; retención limitada.	Oficial de protección de datos o equivalente.	Previo al tratamiento y revisión semestral.
Responsabilidad humana indelegable	Autoridad firmante; protocolo de revisión humana; canal de impugnación; registro de decisiones.	Titular del área decisora y superior jerárquico.	En cada decisión con efectos jurídicos.

Fuente: elaboración propia, con base en Kuner (2017), Floridi et al. (2018), Citron (2008), OCDE (2019), Barocas y Selbst (2016) y Pasquale (2015).

En el ámbito subnacional, esta gradualidad resulta indispensable. Los gobiernos estatales y municipales pueden iniciar con registros públicos de sistemas algorítmicos, protocolos de gobernanza de datos, convenios de asistencia técnica con órganos garantes, universidades o entidades fiscalizadoras, y esquemas de capacitación básica. La falta de capacidad técnica no debe justificar el uso opaco de IA; pero sí exige estándares proporcionales, acompañamiento institucional y cooperación intergubernamental.

6.4 Responsabilidad humana indelegable y distribución de responsabilidades

La responsabilidad humana indelegable debe desarrollarse en dos planos. El primero es la responsabilidad por la decisión final: la autoridad que utiliza una recomendación algorítmica para iniciar una investigación, imponer una medida, excluir a un proveedor o generar una consecuencia administrativa debe fundar, motivar y asumir el acto como propio. El segundo es la responsabilidad por el diseño, adquisición, configuración y mantenimiento del sistema: las áreas técnicas, unidades de datos, áreas requirentes y proveedores deben responder por la calidad de datos, documentación, seguridad, pruebas, actualización y cumplimiento de condiciones contractuales.

Esta distribución puede generar responsabilidades administrativas para servidores públicos que adopten sistemas sin evaluación de impacto, omitan auditorías o deleguen indebidamente decisiones; responsabilidad patrimonial o civil cuando el funcionamiento del sistema produzca daños antijurídicos; responsabilidad contractual de proveedores por incumplimiento de especificaciones, opacidad, fallas de seguridad o uso indebido de datos; e incluso responsabilidad penal cuando exista manipulación dolosa de datos, ocultamiento de información, corrupción en la contratación tecnológica o uso del sistema para encubrir irregularidades. La clave consiste en evitar que la complejidad técnica se convierta en una zona de inmunidad institucional.

5.5 Implicaciones normativas y legislativas

De igual forma, el uso de IA en el sector público requiere un marco jurídico diferenciado, ya que sus decisiones pueden afectar derechos humanos. La regulación no debe limitarse a lineamientos éticos o administrativos, sino establecer obligaciones claras y específicas (Wachter et al., 2017).

En este contexto, la transparencia algorítmica debe convertirse en un deber legal, incluyendo documentación, publicación de criterios, registros públicos, trazabilidad de datos, reglas mínimas de explicabilidad algorítmica y estándares XAI en decisiones que afecten derechos.

Asimismo, se deben establecer evaluaciones de impacto obligatorias en protección de datos, especialmente en sistemas que procesen información sensible o generen perfiles automatizados (Kuner, 2017). Estas evaluaciones permiten anticipar riesgos y fortalecer la responsabilidad institucional, requiriendo además coordinación entre autoridades.

En materia de responsabilidad, es necesario definir qué autoridad responde por

decisiones sustentadas en IA, garantizar revisión administrativa y acceso a la justicia. La automatización no exime de ninguna manera al Estado de cumplir con el debido proceso (Citron, 2008).

Finalmente, la regulación debe integrarse en una política amplia de integridad pública, asegurando coherencia entre combate a la corrupción y respeto a derechos humanos. Sin transparencia, control y responsabilidad, el uso de IA puede erosionar la confianza ciudadana y la legitimidad democrática.

7. Conclusiones

El uso de la IA en el sector público representa un avance tecnológico dirigido a optimizar la operación del Estado. En la lucha contra la corrupción, los sistemas algorítmicos ofrecen mayor eficiencia, capacidad predictiva y detección temprana de irregularidades. Sin embargo, la eficacia tecnológica no puede erigirse como único criterio a favor del uso de la IA en el ámbito público.

Como ejemplo de lo anterior, el análisis del presente ensayo muestra que la integridad pública en entornos algorítmicos requiere estándares reforzados de transparencia, protección de datos personales y responsabilidad humana. La inteligencia artificial no elimina la discrecionalidad ni garantiza por sí misma mayor probidad; incluso puede generar nuevas formas de opacidad si no existen mecanismos claros de explicabilidad, auditoría, supervisión y trazabilidad de datos. Además, persiste el riesgo de trasladar la responsabilidad política hacia tecnologías.

A partir de este diagnóstico, se propone un modelo de "IA con integridad institucional" estructurado en cinco principios: finalidad pública expresa, transparencia verificable, auditoría algorítmica obligatoria, gobernanza estricta de datos y responsabilidad humana indelegable. Este modelo no pretende constituir un esquema cerrado; sino ofrecer un marco orientador para el diseño normativo y la formulación de políticas públicas.

Desde la perspectiva del Estado constitucional, ninguna decisión pública con efectos jurídicos relevantes puede quedar sin responsable identificable ni sin posibilidad de revisión.

Las implicaciones normativas son claras. La regulación del uso de inteligencia artificial en el sector público debe reconocer su impacto diferenciado respecto del ámbito privado y establecer obligaciones específicas de transparencia, evaluación de impacto y control institucional. La coordinación entre órganos de transparencia, autoridades de protección de datos y sistemas anticorrupción resulta indispensable para evitar fragmentación regulatoria.

En definitiva, la tecnología no sustituye la responsabilidad política, al contrario; la expande. Cuanto mayor es la capacidad del Estado para procesar datos y generar decisiones automatizadas, mayor debe ser el estándar de justificación y control democrático. De esta forma, la integridad pública en la era algorítmica no debe ser consecuencia automática de la innovación tecnológica; sino el resultado de decisiones

normativas que subordinan la tecnología a los principios constitucionales.

La inteligencia artificial puede convertirse en una aliada estratégica para el combate a la corrupción. Pero su legitimidad dependerá de su supeditación a la transparencia verificable, protección estricta de datos personales y responsabilidad humana clara. Solo bajo estas condiciones, la automatización contribuirá al fortalecimiento del Estado de derecho y no a la consolidación de nuevas formas de poder opaco que, a su vez, produzcan nuevas modalidades de asimetrías de poder.

Fuentes de información

Aldana, A., Falcón-Cortés, A., & Larralde, H. (2022). A machine learning model to identify corruption in Mexico's public procurement contracts. arXiv. https://arxiv.org/abs/2211.01478

Araujo, M. (2009). El derecho a la protección de datos personales como garantía constitucional. En M. Carbonell & R. Vázquez (coords.), Poder, derecho y corrupción. Siglo XXI.

Avaro, D., & Sánchez, C. (2021). Nuevos desafíos para la rendición de cuentas en tiempos de pandemia: populismo y algoritmocracia. Revista Mexicana de Ciencias Políticas y Sociales, 66(242), 167-187.

Barocas, S., & Selbst, A. D. (2016). Big data's disparate impact. California Law Review, 104(3), 671-732. https://doi.org/10.15779/Z38BG31

Beck, U. (1998). La sociedad del riesgo. Paidós.

Brkan, M., & Bonnet, G. (2020). Legal and technical feasibility of the GDPR's quest for explanation of algorithmic decisions: Of black boxes, white boxes and Fata Morganas. European Journal of Risk Regulation, 11(1), 18-50. https://doi.org/10.1017/err.2020.10

CAF. (2024). Inteligencia artificial y el sector público en América Latina y el Caribe. Banco de Desarrollo de América Latina y el Caribe. https://www.caf.com/

Calle, A. A., Baque, E., & Muñíz, F. (2024). El papel de la analítica predictiva en la anticipación de cambios en el entorno empresarial. Ciencia y Desarrollo, 43-55.

Carbonell, M., & Vázquez, R. (2009). Poder, derecho y corrupción. Siglo XXI.

Citron, D. K. (2008). Technological due process. Washington University Law Review, 85(6), 1249-1313.

Comisión de Transparencia y Acceso a la Información del Estado de Nuevo León. (2018). Evaluación de impacto en protección de datos personales.

Constitución Política de los Estados Unidos Mexicanos. Diario Oficial de la Federación.

Diario Oficial de la Federación. (2025). Ley General de Transparencia y Acceso a la Información Pública. DOF, 20 de marzo de 2025.

Floridi, L., Cowls, J., Beltrametti, M., Chatila, R., Chazerand, P., Dignum, V., Luetge, C., Madelin, R., Pagallo, U., Rossi, F., Schafer, B., Valcke, P., & Vayena, E. (2018). AI4People—An ethical framework for a good AI society. Minds and Machines, 28(4), 689-707. https://doi.org/10.1007/s11023-018-9482-5

Gryz, J., & Rojszczak, M. (2021). Black box algorithms and the rights of individuals: No easy solution to the "explainability" problem. Internet Policy Review, 10(2), 1-24. https://doi.org/10.14763/2021.2.1564

Hernández, O. (2016). La auditoría interna y su alcance ético empresarial. Actualidad Contable Faces, 19(33), 15-41.

Huber, M., & Imhof, D. (2021). Deep learning for detecting bid rigging: Flagging cartel participants based on convolutional neural networks. arXiv. https://arxiv.org/abs/2104.11142

Imhof, D., & Wallimann, H. (2021). Detecting bid-rigging coalitions in different countries and auction formats. arXiv. https://arxiv.org/abs/2105.00337

Jonas, H. (1984). The imperative of responsibility: In search of an ethics for the technological age. University of Chicago Press.

Jurado, D., & Armijo, J. (2022). La efectividad de la auditoría interna en el sector público: Una revisión de literatura. Administración y Desarrollo, 75-96.

Kuner, C. (2017). Transborder data flows and data privacy law. Oxford University Press.

Ley General de Protección de Datos Personales en Posesión de Sujetos Obligados. Diario Oficial de la Federación.

Macías, G. (2025). Transparencia algorítmica. R3D.

Morales, R. (2020). Evaluaciones de impacto y protección de datos personales en el sector público.

Moya, E., & Paillama, D. (2023). Corrupción en contextos de baja estatalidad: Una aproximación sociológica. Economía y Política, 10(1), 149-179.

OCDE. (2017). Recomendación del Consejo de la OCDE sobre Integridad Pública. OCDE Publishing.

OCDE. (2019). Artificial intelligence in society. OCDE Publishing. https://doi.org/10.1787/eedfee77-en

OCDE. (2020). Manual de la OCDE sobre integridad pública. OCDE Publishing. https://doi.org/10.1787/8a2fac21-es

OCDE. (2024). Recommendation of the Council on Artificial Intelligence. OECD/LEGAL/0449.

Ospina, M., Mora, R., & Maya, A. (2024). Percepción de la inteligencia artificial en la lucha contra la corrupción: Una exploración al caso del Estado de Colombia. Opera, 36, 7-45. https://doi.org/10.18601/16578651.n36.02

Pasquale, F. (2015). The black box society: The secret algorithms that control money and information. Harvard University Press.

Rivera, S. (2012). Participación ciudadana y combate a la corrupción: Cinco modelos de incidencia desde el servicio público. INDESOL.

Rubio, F. (1993). El principio de legalidad. Revista Española de Derecho Constitucional, 9-43.

Sánchez, L. (2020). Principio de proporcionalidad y protección de datos personales.

Transparencia Internacional. (2022). Artificial intelligence and corruption risks. Transparency International Secretariat.

Transparencia Internacional. (2025). Addressing corrupt uses of artificial intelligence. Transparency International Secretariat.

UNESCO. (2022). Recomendación sobre la ética de la inteligencia artificial. Organización de las Naciones Unidas para la Educación, la Ciencia y la Cultura. https://unesdoc.unesco.org/ark:/48223/pf0000381137_spa

UNODC. (2019). Integridad pública y ética. Módulo 13 de la Serie de Módulos Universitarios. Oficina de las Naciones Unidas contra la Droga y el Delito.

Unión Europea. (2024). Reglamento (UE) 2024/1689 del Parlamento Europeo y del Consejo, de 13 de junio de 2024, por el que se establecen normas armonizadas en materia de inteligencia artificial. Diario Oficial de la Unión Europea, L 2024/1689. http://data.europa.eu/eli/reg/2024/1689/oj

Wachter, S., Mittelstadt, B., & Floridi, L. (2017). Why a right to explanation of automated decision-making does not exist in the General Data Protection Regulation. International Data Privacy Law, 7(2), 76-99. https://doi.org/10.1093/idpl/ipx005

Zuboff, S. (2019). The age of surveillance capitalism. PublicAffairs.