Revista Hambatu Science

Ambato – Ecuador / ISSN 3151-815X (en línea) / julio - septiembre 2026

Volumen 1, Número 3

https://doi.org/10.63862/rhs-v1n3-511-532-2026

Auditoría de la gobernanza de inteligencia artificial en organizaciones: marco integrado de riesgos, controles, trazabilidad, supervisión humana y aseguramiento

AI governance auditing in organizations: an integrated framework for risks, controls, traceability, human oversight, and assurance

Alfonso Botero Marulanda
Universidad UTEL

Revista Hambatu Science
Periodicidad: julio - septiembre
Volumen: 1, Número: 3, Año: 2026

DOI: https://doi.org/10.63862/rhs-v1n3-511-532-2026

Auditoría de la gobernanza de inteligencia artificial en organizaciones: marco integrado de riesgos, controles, trazabilidad, supervisión humana y aseguramiento

Alfonso Botero Marulanda
Universidad UTEL, Doctorado en Gestión e Innovación Tecnológica
alfonsoboterom@gmail.com
https://orcid.org/0009-0007-8164-3081
Medellin, Colombia

Recibido: 2026-08-12

Aceptado: 2026-08-14

Published: 2026-08-19

Resumen

La expansión de la inteligencia artificial (IA) dentro de las organizaciones ha vuelto insuficiente una discusión centrada únicamente en desempeño técnico o cumplimiento formal. Cuando estos sistemas intervienen en decisiones, servicios y procesos sensibles, la cuestión pasa a ser si la organización puede explicar cómo identifica riesgos, asigna responsabilidades, controla el uso de la tecnología y conserva evidencia de lo ocurrido. Este artículo examina esa cuestión mediante una revisión integradora estructurada de literatura publicada entre 2020 y 2026, con PRISMA 2020 como referencia para la transparencia del reporte y con contraste frente a ISO/IEC 42001:2023, NIST AI RMF 1.0 y el Reglamento (UE) 2024/1689. La revisión muestra un campo conceptualmente rico, aunque fragmentado entre derecho, ética, gestión, tecnología y auditoría. A partir de esa lectura se propone un marco organizacional de cinco dimensiones —riesgos, controles, trazabilidad, supervisión humana y aseguramiento— cuya contribución reside en explicar sus relaciones y no en convertirlas en una lista de verificación. La evidencia revisada sugiere que una organización solo puede ser auditada de manera consistente cuando su propia gobernanza produce registros, responsabilidades y decisiones reconstruibles. Por ello, se sostiene que la auditabilidad debe incorporarse al diseño de la gobernanza de IA. Quedan abiertas, entre otras, la validación empírica del marco, la calidad probatoria de los registros y las condiciones de independencia del aseguramiento.

Palabras clave: gobernanza de inteligencia artificial, auditoría algorítmica, gestión de riesgos, trazabilidad, supervisión humana, aseguramiento, inteligencia artificial responsable.

Revista Hambatu Science
Periodicidad: julio - septiembre
Volumen: 1, Número: 3, Año: 2026

AI governance auditing in organizations: an integrated framework for risks, controls, traceability, human oversight, and assurance

Abstract

As artificial intelligence (AI) becomes embedded in organizational decisions, services, and operational processes, technical performance and formal compliance no longer provide a sufficient account of responsible use. A more demanding question is whether an organization can demonstrate how it identifies risks, assigns responsibility, governs system changes, enables meaningful human intervention, and preserves evidence of what occurred. This article addresses that question through a structured integrative review of literature published between 2020 and 2026. PRISMA 2020 informs reporting transparency, while ISO/IEC 42001:2023, the NIST AI RMF 1.0, and Regulation (EU) 2024/1689 provide complementary points of comparison. The literature is substantial but dispersed across legal, ethical, managerial, technical, and auditing traditions. Drawing on their areas of convergence and tension, the article develops a five-dimensional organizational framework comprising risk, controls, traceability, human oversight, and assurance. Its purpose is analytical rather than procedural: it does not offer an audit checklist or assessment instrument. The review indicates that credible auditing depends on governance arrangements that generate reconstructable decisions, responsibilities, records, and control evidence throughout the AI lifecycle. Auditability is therefore treated as a design property of AI governance rather than as a corrective exercise performed after deployment. Further research should test the framework empirically and examine evidential quality, sectoral variation, and assurance independence.

Keywords: artificial intelligence governance, algorithmic auditing, risk management, traceability, human oversight, assurance, responsible artificial intelligence.

Introducción

La inteligencia artificial ya no ocupa un lugar periférico dentro de las organizaciones. En numerosos contextos clasifica personas, prioriza expedientes, recomienda cursos de acción, detecta anomalías, distribuye recursos y condiciona la relación entre instituciones y ciudadanía. Esa expansión importa porque altera el tipo de problemas que deben gobernarse: no se trata únicamente de que un modelo acierte, sino de saber quién decidió utilizarlo, con qué datos, bajo qué límites y con qué mecanismos de respuesta cuando falla. (Salvador Serna, 2021) vincula la incorporación transversal de IA con capacidades organizativas y gobernanza de datos. (Mäntymäki et al., 2022), por su parte, la sitúan dentro de un sistema de reglas, prácticas, procesos y herramientas que debe permanecer alineado con los objetivos y valores institucionales. La revisión de (Papagiannidis et al., 2025) refuerza esta lectura al identificar prácticas estructurales, relacionales y procedimentales a lo largo del ciclo de vida. Desde el derecho, (Sánchez Acevedo, 2022) y (Landa Arroyo, 2022) añaden un límite que no puede omitirse: la capacidad tecnológica no desplaza las garantías jurídicas ni los derechos fundamentales.

Aquí aparece una distinción que atraviesa todo el artículo. Un buen resultado técnico no demuestra, por sí solo, una buena gobernanza. Es perfectamente posible encontrar un sistema con métricas aceptables que opere sobre datos mal gobernados, responsabilidades imprecisas o mecanismos de reclamación débiles. También puede ocurrir que la revisión humana exista en el procedimiento y apenas tenga efectos en la práctica. (Aguirre Sala, 2022) muestra por qué las evaluaciones de impacto resultan necesarias cuando los principios generales no bastan para anticipar daños, mientras (Flores Anarte, 2023) evidencia que la automatización puede reproducir desigualdades bajo una apariencia de neutralidad. A juicio propio, este desplazamiento —del modelo aislado hacia la organización que lo adopta y controla— es decisivo para estudiar la auditoría de IA.

La literatura, sin embargo, no ha recorrido ese desplazamiento de manera uniforme. Los estudios jurídicos han desarrollado con mayor profundidad los derechos, la discriminación y las garantías (Bernal, 2022; Landa Arroyo, 2022; Presno Linera, 2025). En gestión y sistemas de información, el interés se dirige hacia capacidades, estructuras y responsabilidades (Birkstedt et al., 2023; Minkkinen et al., 2022; Salvador Serna, 2021). Otra corriente examina evaluaciones de impacto y anticipación del daño (Aguirre Sala, 2022; Mantelero, 2024; Simón

Castellano, 2025), mientras la literatura de auditoría se ocupa de procedimientos de evaluación, independencia y control (Minkkinen et al., 2022; Mokander, 2023; Mokander & Floridi, 2021). No se ve esta diversidad como un defecto del campo. El problema surge cuando cada tradición conserva su propio objeto y vocabulario y, por esa razón, resulta difícil evaluar la gobernanza de IA como un sistema organizacional completo.

La transparencia permite observar con claridad esa dificultad. (Cotino Hueso, 2023) advierte que no existe una única transparencia algorítmica: la información debida cambia según quién la solicita, para qué la necesita y en qué contexto se utiliza el sistema. (Williams et al., 2022) llevan la discusión un paso más allá al vincular transparencia y explicabilidad con accountability. Esta diferencia es importante para una auditoría, porque informar no equivale necesariamente a dejar evidencia suficiente para verificar.

Por eso se distingue transparencia de trazabilidad. La primera puede satisfacer una obligación de información; la segunda debe permitir reconstruir hechos. Para un evaluador interesa conocer qué versión del sistema intervino, qué datos fueron relevantes, quién aprobó un cambio, qué control se ejecutó y qué evidencia quedó de esa actuación. (Hauer et al., 2023) describen mecanismos de transparencia e inspectabilidad distribuidos a lo largo del desarrollo de sistemas algorítmicos. Leída desde el aseguramiento, esa literatura conduce a una conclusión más concreta: cuando la cadena de evidencia se interrumpe, la posibilidad de formular una conclusión auditada también se reduce.

La supervisión humana plantea una dificultad semejante. (Obregón Fernández & Lazcoz Moratinos, 2021) muestran su importancia en sistemas de alto riesgo; (Laux, 2024) cuestiona los diseños que descansan en una confianza excesiva en la automatización; y (Azuaje Pirela et al., 2025) distinguen entre intervención, supervisión y control. De estos trabajos se desprende que colocar a una persona al final del proceso no basta. El revisor necesita comprender las limitaciones del sistema, recibir información pertinente y disponer de autoridad para apartarse de la recomendación automatizada. Además, debe contar con tiempo y condiciones organizacionales que hagan posible ejercer ese juicio. Cuando alguno de esos elementos falta, la intervención humana corre el riesgo de convertirse en una firma de trámite.

Estado de arte

Gobernanza de inteligencia artificial como capacidad organizacional

En este artículo se entiende la gobernanza de IA como el conjunto de arreglos mediante los cuales una organización decide quién puede hacer qué con un sistema de IA, bajo qué reglas y con qué responsabilidad. Esta formulación recoge distintos aportes.(Salvador Serna, 2021) identifica, en las administraciones públicas, componentes de estrategia, datos, organización, talento y relación con el entorno. (Minkkinen et al., 2022) ofrecen una definición organizacional que reúne reglas, prácticas, procesos y herramientas; (Papagiannidis et al., 2025) ordenan la gobernanza responsable en prácticas estructurales, relacionales y procedimentales. (Vida Fernandez, 2022) recuerda, además, que los riesgos digitales cambian con rapidez y exigen respuestas institucionales capaces de adaptarse. Lo común entre estas perspectivas no es una estructura única, sino la idea de que la gobernanza debe materializarse en decisiones y responsabilidades observables.

Esa lectura permite mantener un marco transversal sin suponer que todas las organizaciones son iguales. Un banco, una universidad, una entidad pública y una empresa industrial operan bajo obligaciones y exposiciones distintas. Aun así, cada una necesita saber quién asume el riesgo, quién autoriza el uso del sistema, cómo se escalan incidentes y ante quién se rinden cuentas. (Divino, 2024) aborda esta cuestión desde gobernanza y cumplimiento empresarial; (Salvador Serna, 2021) lo hace desde capacidades institucionales del sector público. En consecuencia, se propone dimensiones comunes, pero no una receta organizativa uniforme. Su aplicación debe variar con el sector, la jurisdicción, el tamaño, la complejidad tecnológica y, sobre todo, el nivel de riesgo.

Riesgo algorítmico y evaluación de impacto

El riesgo algorítmico comprende la posibilidad de que el diseño, entrenamiento, despliegue o uso de un sistema de IA produzca consecuencias adversas para personas, organizaciones o bienes jurídicos. (Aguirre Sala, 2022) sostiene que la evaluación de impacto debe reconocer la transversalidad de daños y riesgos y comparar modelos de evaluación; (Lazcoz Moratinos & Castillo Parrilla, 2020), a partir del caso SyRI, muestran que proporcionalidad, privacidad y garantías frente a sistemas automatizados son inseparables del análisis de riesgo; y (Mantelero,

2024) desarrolla la evaluación de impacto en derechos fundamentales como instrumento específico para sistemas de IA bajo el marco europeo. Desde una perspectiva de auditoría, (Koshiyama et al., 2024) integran riesgos legales, éticos y tecnológicos en una aproximación de auditoría algorítmica. Estas contribuciones sostienen que el riesgo no puede reducirse a error estadístico: debe comprender contexto, personas afectadas, propósito y consecuencias.

La discriminación constituye uno de los campos donde el riesgo algorítmico se hace más visible. (Ramírez-Bustamente & Páez, 2021) analizan las dificultades jurídicas para identificar discriminación en procesos automatizados de selección laboral; (Flores Anarte, 2023) estudia específicamente los sesgos de género; y (Morales Ramirez, 2023) discute la dimensión antropológica de la discriminación algorítmica. Estos trabajos convergen en una conclusión: la apariencia matemática de neutralidad no elimina decisiones normativas incorporadas en datos, variables, objetivos y umbrales.

Controles organizacionales y evidencia de aseguramiento

Hablar de control exige ir más allá de la existencia de políticas. Se considera control a toda medida deliberada que previene, detecta, corrige o reduce un riesgo identificado. Algunas medidas serán técnicas —por ejemplo, validación de datos, pruebas de robustez o seguimiento del desempeño—; otras dependerán de procedimientos, como aprobaciones, segregación de funciones o gestión de cambios. También existen controles jurídicos y organizacionales, entre ellos evaluaciones de impacto, cláusulas contractuales y mecanismos de reclamación. (Fajardo Pereira et al., 2023) muestran cómo la IA está transformando la propia práctica de auditoría; (Mota Sanchez & Herrera Exposito, 2023) examinan la auditoría algorítmica en el sector público. Para el propósito, ambos trabajos ayudan a formular la pregunta inversa: ¿qué debe poder observar un auditor para saber si los controles que rodean a la IA funcionan realmente?

La respuesta no puede obtenerse leyendo únicamente manuales y procedimientos. Un control puede estar bien redactado y fallar en la operación. Se piensa en una revisión humana formalmente obligatoria: si quien revisa no comprende el sistema, carece de contexto o no puede contradecir su resultado, el control existe en el papel, pero su eficacia es discutible. (Azuaje Pirela et al., 2025) permiten precisar esa diferencia al separar intervención, supervisión y control humano. Por ello, se distingue tres momentos analíticos: cómo fue concebido el

control, si llegó a implantarse y qué ocurrió cuando tuvo que operar. Esta distinción evita equiparar cumplimiento documental con funcionamiento efectivo.

Trazabilidad, transparencia y explicabilidad

Transparencia, explicabilidad y trazabilidad aparecen con frecuencia en una misma conversación, aunque responden a preguntas distintas. La transparencia se ocupa de qué información debe hacerse visible y para quién; la explicabilidad intenta hacer comprensible un resultado o comportamiento del sistema; la trazabilidad, en cambio, pregunta si puede reconstruirse el recorrido de una decisión. (Cotino Hueso, 2023) desarrolla la primera de estas diferencias desde las obligaciones de información. (Williams et al., 2022) y (Hauer et al., 2023) muestran que la rendición de cuentas exige mecanismos adicionales de inspectabilidad. Para la auditoría, esa distinción evita un error común: asumir que un sistema explicado es, por ese solo hecho, auditable.

La trazabilidad tampoco puede reducirse al registro de la salida final. Cuando el riesgo lo justifica, debe ser posible relacionar datos, versiones, cambios de configuración, responsables, aprobaciones, incidentes y medidas correctivas. La intensidad de ese registro no tiene por qué ser idéntica en todos los sistemas; exigirlo produciría documentación sin valor. Lo relevante es conservar aquello que permita reconstruir los hechos materiales y asociarlos con la versión realmente utilizada. En sistemas que cambian por reentrenamiento, actualización o modificación de contexto, esta continuidad temporal se vuelve especialmente importante. Sin ella, una auditoría corre el riesgo de evaluar un sistema distinto del que produjo la decisión cuestionada.

Supervisión humana significativa

La supervisión humana ocupa un lugar central en la regulación y en la literatura reciente, pero sigue siendo un concepto fácil de vaciar de contenido. (Obregón Fernández & Lazcoz Moratinos, 2021) la estudian como salvaguarda en sistemas de alto riesgo; (Laux et al., 2024) insiste en que su diseño debe permitir cuestionar la automatización, no simplemente acompañarla; (Azuaje Pirela et al., 2025) diferencian formas de intervención y control. Estas contribuciones coinciden en un punto que se considera fundamental: el valor de la supervisión depende de la capacidad efectiva de intervenir, no de la presencia nominal de una persona.

Esa capacidad se construye con varios elementos que funcionan conjuntamente. Quien supervisa necesita competencia suficiente para reconocer limitaciones y errores; también necesita información contextual que le permita interpretar la salida. La autoridad es igualmente decisiva: si no puede corregir, escalar o detener una decisión, su participación tiene poco contenido. El tiempo disponible y los incentivos de la organización completan el cuadro, porque una revisión apresurada o sistemáticamente orientada a confirmar la máquina favorece el sesgo de automatización. No se propone estos elementos como una lista cerrada de auditoría. Los utiliza para distinguir una supervisión sustantiva de otra puramente formal y para señalar qué tipo de evidencia debería poder encontrarse cuando la organización afirma que existe control humano significativo.

Auditoría de la gobernanza de IA

Se propone denominar auditoría de la gobernanza de IA al proceso sistemático mediante el cual se obtiene y valora evidencia sobre las estructuras, responsabilidades y controles que una organización utiliza para dirigir sus sistemas de IA. La definición no surge de un único antecedente. (Mokander & Floridi, 2021) desarrollan la auditoría basada en ética; (Minkkinen et al., 2022) conceptualizan la auditoría continua; (Mokander, 2023) integra aproximaciones legales, éticas y técnicas; y (Koshiyama et al., 2024) amplían el examen hacia riesgos legales, éticos y tecnológicos. Al combinar esos aportes con la literatura de gobernanza organizacional, se desplaza el foco desde el modelo aislado hacia el entorno institucional que decide cómo se desarrolla, adquiere, despliega y supervisa.

Conviene no confundir dos preguntas. Una parte de la literatura estudia cómo la IA cambia el trabajo del auditor, por ejemplo mediante automatización y análisis de grandes volúmenes de información (Fajardo Pereira et al., 2023). Otra pregunta, distinta, es cómo auditar los propios sistemas de IA y las organizaciones que los gobiernan (Koshiyama et al., 2024; Mokander, 2023). Este artículo se sitúa en la segunda. La diferencia parece sencilla, pero tiene consecuencias metodológicas: el objeto de prueba ya no es solo una transacción o un modelo, sino una red de decisiones, controles, datos, responsabilidades y evidencias.

Metodología

Se adoptó una revisión integradora estructurada de carácter conceptual, informada por PRISMA 2020 como guía de transparencia del reporte (Page et al., 2021). La elección responde al carácter interdisciplinario del objeto: la auditoría de la gobernanza de IA se distribuye entre derecho, administración pública, sistemas de información, ética aplicada, auditoría y estudios sociotécnicos. La revisión integradora permite comparar definiciones, mecanismos, relaciones y vacíos procedentes de tradiciones disciplinares diferentes sin pretender una síntesis estadística. PRISMA se utiliza para explicitar criterios de búsqueda, selección y síntesis; no se presenta esta revisión como metaanálisis ni se atribuyen cifras de flujo que no hayan sido documentadas mediante exportaciones reproducibles de las bases.

La ventana temporal se fijó entre enero de 2020 y agosto de 2026 con el propósito de capturar la fase reciente de consolidación de la gobernanza responsable, la auditoría algorítmica y los marcos regulatorios contemporáneos. Se priorizó literatura en español y del ámbito iberoamericano, complementada por un núcleo internacional de trabajos directamente relacionados con gobernanza organizacional, auditoría, accountability, riesgo y supervisión humana. Las bases objetivo para la búsqueda estructurada fueron Scopus, Web of Science, SciELO y Dialnet; adicionalmente, la verificación bibliográfica se apoyó en páginas editoriales, DOI/Crossref y repositorios institucionales cuando fue necesario confirmar metadatos.

Las cadenas de búsqueda combinaron grupos de términos en español e inglés: (“inteligencia artificial” OR “artificial intelligence” OR “sistemas algorítmicos” OR “algorithmic systems”) AND (“gobernanza” OR “governance” OR “auditoría” OR “audit*” OR “aseguramiento” OR “assurance”) AND (“riesgo” OR “risk” OR “control” OR “trazabilidad” OR “traceability” OR “transparencia” OR “transparency” OR “supervisión humana” OR “human oversight”). Se realizaron búsquedas complementarias para “auditoría algorítmica”, “evaluación de impacto”, “discriminación algorítmica”, “gobernanza de datos”, “accountability”, “inspectability” y “continuous auditing”. La estrategia buscó maximizar sensibilidad temática y después aplicar cribado conceptual por título, resumen y pertinencia respecto de las dimensiones del marco.

Los criterios de inclusión fueron: publicación entre 2020 y 2026; artículo científico o revisión académica con metadatos verificables; pertinencia directa con gobernanza, auditoría, riesgos,

controles, transparencia/trazabilidad, supervisión humana o aseguramiento; y disponibilidad de DOI cuando la publicación lo ofreciera. Se priorizaron revistas indexadas y literatura en español, sin excluir un conjunto limitado de trabajos internacionales necesarios para cubrir desarrollos que todavía tienen menor representación iberoamericana. Se excluyeron noticias, blogs, materiales comerciales, preprints cuando existía versión publicada, duplicados, documentos sin autoría trazable y textos cuyo vínculo con el problema fuera meramente tangencial. Los estándares y normas se trataron como referentes institucionales de contraste y no como evidencia científica equivalente a artículos revisados por pares.

La extracción conceptual se organizó en una matriz con autor, año, sector o ámbito de aplicación, objeto de estudio, enfoque de gobernanza, tipo de riesgo, controles, trazabilidad/transparencia, supervisión humana, auditoría/aseguramiento, hallazgo relevante y limitaciones. La codificación combinó categorías deductivas —las cinco dimensiones del marco— con categorías inductivas surgidas durante la lectura, particularmente la distinción entre transparencia y trazabilidad, entre presencia humana y supervisión significativa, y entre auditoría técnica, auditoría de procesos y evaluación de conformidad. La matriz funcionó como dispositivo de comparación y no como escala cuantitativa de calidad.

Resultados

Convergencias en la literatura

La revisión muestra una convergencia clara: el riesgo de IA ya no puede tratarse como un problema exclusivamente técnico. Los trabajos jurídicos incorporan derechos y discriminación; los de gestión añaden capacidades, responsabilidades y gobernanza de datos; las propuestas de auditoría introducen evidencia, independencia y control. Esta coincidencia aparece desde ángulos distintos en (Aguirre Sala, 2022; Mäntymäki et al., 2022; Mokander, 2023; Papagiannidis et al., 2025). Lo relevante no es que todos utilicen el mismo vocabulario —no lo hacen—, sino que desplazan la unidad de análisis hacia el sistema sociotécnico y la organización que lo sostiene.

También se encontró un acuerdo amplio alrededor de la responsabilidad organizacional. La adopción de IA obliga a distribuir funciones, definir propietarios de riesgos y establecer mecanismos de rendición de cuentas. (Salvador Serna, 2021) lo muestra desde las capacidades

del sector público; (Busuioc, 2021) examina la exigencia de hacer responsables a los sistemas algorítmicos y a las instituciones que los utilizan; (Mäntymäki et al., 2022) lo incorporan a la definición de gobernanza organizacional. Esta convergencia respalda una idea sencilla pero exigente: cuando nadie puede explicar quién decidió, quién controló y quién debía intervenir, el problema no es solo técnico; es de gobernanza.

Discrepancias y fragmentación identificada

Las discrepancias aparecen cuando se pregunta qué debe evaluarse. En derecho, el centro suele estar en la legalidad, los derechos y las garantías. La literatura de gobernanza observa estructuras y capacidades. Los trabajos técnicos privilegian desempeño, datos y robustez, mientras la auditoría exige criterios, evidencia y alcance. Ninguna perspectiva resulta suficiente por sí sola para el objeto que aquí se propone. El desafío consiste en conectarlas sin borrar sus diferencias, porque una auditoría de gobernanza necesita saber cuándo una cuestión requiere prueba técnica, cuándo exige juicio jurídico y cuándo revela una falla de diseño organizacional.

Otra ruptura se produce entre identificar riesgos y demostrar que fueron tratados. Muchos estudios describen daños potenciales con gran precisión, pero dejan menos desarrollado el recorrido que lleva desde el riesgo hasta un control verificable. En la práctica, ese recorrido debería poder reconstruirse: qué riesgo fue reconocido, qué respuesta se decidió, quién asumió la responsabilidad, qué control se implantó y qué evidencia permite valorar su funcionamiento. Esta relación es uno de los puntos donde la literatura de impacto y la de auditoría todavía dialogan menos de lo deseable.

Marco organizacional integrado propuesto

A partir de esas convergencias y tensiones se propone cinco dimensiones. El riesgo explica por qué una intervención resulta necesaria. Los controles representan la respuesta organizacional frente a esa exposición. La trazabilidad conserva la memoria verificable de decisiones y cambios. La supervisión humana introduce capacidad de juicio e intervención allí donde el riesgo lo exige. Finalmente, el aseguramiento reúne criterios y evidencia para sostener una conclusión. No son etapas rígidas ni compartimentos independientes. Su utilidad analítica aparece precisamente en las relaciones entre ellas: un control sin riesgo identificable pierde

justificación; un control sin evidencia resulta difícil de probar; y una evidencia sin trazabilidad puede perder su vínculo con el sistema realmente utilizado.

Tabla 1
Dimenciones del marco organizacional integrado propuesto

Dimensión	Pregunta de auditoría	Objeto de prueba	Evidencia esperada	Resultado conceptual
Riesgos	¿La gobernanza identifica y prioriza riesgos pertinentes?	Procesos de identificación, análisis y aceptación	Inventarios, evaluaciones de impacto, criterios y decisiones	Coherencia entre exposición y respuesta
Controles	¿Los riesgos relevantes tienen respuestas organizacionales?	Diseño e implementación de controles	Políticas, aprobaciones, pruebas, métricas e incidencias	Suficiencia y funcionamiento
Trazabilidad	¿Puede reconstruirse el ciclo de vida y las decisiones?	Cadena de registros y versiones	Logs, versiones, documentación, responsables y cambios	Auditabilidad y atribución
Supervisión humana	¿La intervención humana posee capacidad efectiva?	Roles, competencia, autoridad y escalamiento	Formación, anulaciones, excepciones, revisiones	Control humano significativo
Aseguramiento	¿La conclusión se sustenta en criterios y evidencia suficientes?	Relación criterio-evidencia-juicio	Papeles de trabajo, fuentes, pruebas y conclusiones	Confianza justificada y límites explícitos

Nota. Elaboración propia a partir de la síntesis de la literatura revisada (Aguirre Sala, 2022; Cotino Hueso, 2023; Mäntymäki et al., 2022; Mökander, 2023; Mökander & Floridi, 2021).

Alineación con estándares internacionales

ISO/IEC 42001:2023 introduce un sistema de gestión de inteligencia artificial estructurado alrededor de políticas, objetivos, procesos y mejora continua. Su relevancia para el marco propuesto reside en institucionalizar la IA como objeto de gestión organizacional y no como tecnología aislada. La dimensión de controles se relaciona con la necesidad de procesos gestionados; la trazabilidad con información documentada; y el aseguramiento con la posibilidad de evaluar el sistema de gestión. La correspondencia es conceptual y no implica equivalencia entre auditoría del estándar y auditoría integral de gobernanza.

NIST AI RMF 1.0 organiza la gestión mediante Govern, Map, Measure y Manage. La función Govern resulta especialmente compatible con la tesis de este artículo porque establece una capa transversal de políticas, responsabilidades y cultura. Map y Measure aportan lógica para comprender contexto y riesgo, mientras Manage conecta resultados con tratamiento. El marco

aquí propuesto añade una lectura de aseguramiento: pregunta cómo esas funciones producen evidencia susceptible de evaluación independiente.

Vacíos de investigación identificados

Al conectar las cinco dimensiones aparecieron cuatro grupos de preguntas que la literatura todavía no resuelve de manera satisfactoria. Se prefere hablar de vacíos relativos y no de ausencias absolutas: el corpus contiene conocimiento relevante en cada dimensión, pero ofrece menos evidencia sobre su funcionamiento conjunto dentro de organizaciones reales. Esta cautela es importante, porque una revisión no puede demostrar que un tema no exista en toda la literatura; sí puede señalar dónde el diálogo entre corrientes sigue siendo insuficiente.

El primer vacío se refiere a la validación empírica de marcos integrados. Existen desarrollos robustos sobre gobernanza organizacional, riesgo y auditoría (Birkstedt et al., 2023; Koshiyama et al., 2024; Mäntymäki et al., 2022; Papagiannidis et al., 2025) et al., 2025), pero todavía se sabe menos sobre cómo interactúan esas piezas en organizaciones concretas y si las mismas relaciones se mantienen entre sectores. El marco propuesto necesita ser sometido a contraste mediante casos múltiples, estudios Delphi, encuestas organizacionales o diseños de ciencia del diseño. Esa validación debería examinar no solo si las dimensiones son comprensibles, sino si ayudan a explicar diferencias reales en la calidad y auditabilidad de la gobernanza.

Un segundo vacío aparece en la supervisión humana y en los controles que la hacen posible. Los estudios sobre derechos y discriminación describen con claridad los daños que justifican salvaguardas (Azuaje Pirela et al., 2025; Flores Anarte, 2023; Landa Arroyo, 2022; Laux, 2024; Obregón Fernández & Lazcoz Moratinos, 2021) avanzan en la conceptualización del control humano. Falta, sin embargo, evidencia comparativa suficiente sobre qué arreglos funcionan mejor bajo diferentes niveles de riesgo. No basta con saber que existe un revisor; interesa conocer cuándo dispone de información útil, cuándo puede contradecir al sistema y qué ocurre después de una intervención. Ese tránsito desde la regla hacia el comportamiento observable merece mayor investigación.

Discusión

La lectura de los resultados lleva a separar con mayor nitidez gobernanza y auditoría. La gobernanza distribuye autoridad, fija objetivos, asigna responsabilidades y establece controles. La auditoría, en cambio, examina evidencia sobre ese sistema de dirección y control. (Minkkinen et al., 2022; Papagiannidis et al., 2025) ayudan a precisar la primera función; (Koshiyama et al., 2024; Mokander, 2023), la segunda. Esta separación evita tratar "gobernanza y auditoría" como una pareja de términos intercambiables. Lo que se propone auditar no es una IA abstracta, sino la forma en que una organización decide desarrollarla, adquirirla, desplegarla, modificarla y supervisarla.

La trazabilidad adquiere, desde esta perspectiva, un peso mayor del que suele recibir en las discusiones generales sobre transparencia. (Cotino Hueso, 2023) muestra que la información exigible depende del contexto; (Williams et al., 2022) conectan esa discusión con accountability; (Hauer et al., 2023) describen mecanismos de inspectabilidad distribuidos a lo largo del desarrollo. Para un auditor queda una pregunta adicional: ¿pueden reconstruirse los hechos relevantes? Una explicación puede ser jurídicamente útil y, aun así, resultar insuficiente para vincular una decisión con una versión concreta del sistema. De ahí que la trazabilidad funcione como puente entre la discusión normativa y la necesidad probatoria del aseguramiento.

La supervisión humana confirma que el problema no se resuelve con etiquetas. La literatura revisada obliga a preguntar qué puede hacer realmente la persona situada en el circuito decisional (Azuaje Pirela et al., 2025; Laux, 2024; Obregón Fernández & Lazcoz Moratinos, 2021). En algunos contextos bastará con revisar excepciones; en otros será necesario poder detener una decisión o escalarla. El grado adecuado dependerá del riesgo. Por eso se sostiene, que la supervisión debe evaluarse como capacidad organizacional demostrable y no como requisito formal de interfaz. Esta lectura evita tanto la idealización del juicio humano como la aceptación automática de la decisión algorítmica.

El marco tampoco sustituye las evaluaciones de impacto. (Aguirre Sala, 2022; Simón Castellano, 2025) muestran su utilidad para anticipar efectos y ordenar riesgos, especialmente cuando están comprometidos derechos. La propuesta las ubica dentro de una arquitectura más amplia. Una auditoría puede examinar si la evaluación se realizó cuando correspondía, si sus

supuestos fueron razonables, si generó medidas concretas y si esas medidas llegaron a implementarse. En ese sentido, evaluación de impacto y auditoría cumplen funciones distintas pero complementarias: una anticipa y estructura el riesgo; la otra pregunta qué evidencia existe sobre la calidad y el funcionamiento de ese proceso.

La comparación con la literatura de auditoría permite precisar dónde está la contribución. (Fajardo Pereira et al., 2023) estudian la transformación de la auditoría mediante IA. (Koshiyama et al., 2024; Minkkinen et al., 2022; Mokander, 2023; Mokander & Floridi, 2021) se ocupan de auditar sistemas y procesos de IA. Se desplazó la pregunta hacia la gobernanza que rodea esos sistemas. Ese cambio obliga a combinar evidencia técnica con responsabilidades, derechos, datos, controles y supervisión. No se propone que el auditor deba dominar por sí solo todas esas disciplinas; sí que el alcance de la auditoría reconozca cuándo necesita integrar competencias distintas.

En la práctica, el valor del marco es ofrecer un vocabulario compartido. Consejos de administración, responsables jurídicos, áreas de tecnología, riesgos, cumplimiento y auditoría suelen observar el mismo sistema desde preocupaciones diferentes. Salvador Serna (2021) muestra que la integración de IA distribuye capacidades entre estrategia, datos, organización y talento; (Divino, 2024) subraya la relación entre gobernanza y cumplimiento empresarial. Las cinco dimensiones pueden ayudar a ordenar ese diálogo sin imponer una estructura universal. Esta última cautela es esencial: un marco transversal solo resulta útil si admite variaciones por sector, jurisdicción, tamaño, complejidad y perfil de riesgo.

Conclusiones

La revisión confirma que el campo dispone de numerosos principios, estándares y aproximaciones, pero todavía articula de manera desigual gobernanza, riesgo, control, trazabilidad, supervisión humana y auditoría. Esa fragmentación explica por qué una organización puede cumplir requisitos parciales y, aun así, tener dificultades para demostrar cómo gobierna sus sistemas de IA. A lo largo del artículo se ha defendido una distinción básica: gobernar significa dirigir, asignar responsabilidades y controlar; auditar significa obtener y valorar evidencia sobre ese sistema. La literatura reciente sobre gobernanza organizacional y auditoría respalda la necesidad de conectar ambas funciones, aunque no ofrece todavía una

arquitectura única y universalmente aceptada (Birkstedt et al., 2023; Minkkinen et al., 2022; Mokander, 2023; Papagiannidis et al., 2025).

Sobre esa base se formula un marco de cinco dimensiones. El riesgo aporta la razón para intervenir; los controles expresan la respuesta de la organización; la trazabilidad conserva evidencia sobre decisiones y cambios; la supervisión humana introduce juicio e intervención cuando el contexto lo requiere; y el aseguramiento permite sostener conclusiones frente a criterios explícitos. El aporte no está en presentar estos conceptos como nuevos. Está en mostrar que su relación determina, en buena medida, si la gobernanza puede ser examinada y defendida con evidencia.

De ahí deriva la tesis principal del artículo: la auditabilidad no debería añadirse al final del ciclo de vida. Debe construirse mientras se gobierna. Si una organización no conserva versiones, decisiones, responsabilidades e incidencias, el auditor posterior encontrará un vacío que ninguna metodología puede reparar por completo. Esta idea también obliga a mantener separados tres conceptos próximos pero distintos: transparencia, explicabilidad y trazabilidad. La primera informa; la segunda ayuda a comprender; la tercera permite reconstruir. Las tres pueden ser necesarias, pero no cumplen la misma función.

La supervisión humana merece especial atención en la investigación futura. Su eficacia no depende de que una persona aparezca formalmente en el procedimiento, sino de que pueda comprender, cuestionar y modificar el curso de la decisión cuando sea necesario. Queda por determinar qué combinaciones de competencia, información, autoridad, tiempo e incentivos funcionan mejor en distintos sectores y niveles de riesgo. También será necesario estudiar cómo demostrar esa eficacia sin convertir la gobernanza en una burocracia de aprobaciones que produzca documentos, pero no control real.

Referencias

Aguirre Sala, J. F. (2022). Models and good evaluative practices to detect impacts, risks and damages of artificial intelligence. Paakat: Revista de Tecnología y Sociedad, 12(23), 1–20. https://doi.org/10.32870/Pk.a12n23.742

Azuaje Pirela, M. C., Coddou Mc Manus, A., Contreras Vásquez, P., & Viollier Bonvin, P. (2025). Intervención, supervisión y control humano en las decisiones automatizadas. Derecho PUCP, (95), 189–226. https://doi.org/10.18800/derechopucp.202502.006

Bernal, C. (2022). Derechos fundamentales e inteligencia artificial. International Journal of Constitutional Law, 20(4), 1440–1445. https://doi.org/10.1093/icon/moac099

Birkstedt, T., Minkkinen, M., Tandon, A., & Mäntymäki, M. (2023). AI governance: themes, knowledge gaps and future agendas. Internet Research, 33(7), 133–167. https://doi.org/10.1108/INTR-01-2022-0042

Busuioc, M. (2021). Accountable Artificial Intelligence: Holding Algorithms to Account. Public Administration Review, 81(5), 825–836. https://doi.org/10.1111/puar.13293

Cotino Hueso, L. (2023). Qué concreta transparencia e información de algoritmos e inteligencia artificial es la debida. Revista Española de la Transparencia, (16), 17–63. https://doi.org/10.51915/ret.272

Divino, S. (2024). Governance and Compliance Recommendations for Artificial Intelligence in Business Management. Nuevo Derecho, 20(35), 1–17. https://doi.org/10.25057/2500672X.1665

European Parliament & Council of the European Union. (2024). Regulation (EU) 2024/1689 of the European Parliament and of the Council of 13 June 2024 laying down harmonised rules on artificial intelligence (Artificial Intelligence Act). Official Journal of the European Union.

Fajardo Pereira, J., Toscano Hernández, A., García Alarcón, H., & Llanos Ayola, J. (2023). Inteligencia Artificial y Auditoría: Tendencias de la literatura científica. Panorama Económico, 31(2), 160–188. https://doi.org/10.32997/pe-2023-4575

Flores Anarte, L. (2023). Sesgos de género en la Inteligencia Artificial. Revista Internacional de Pensamiento Político, 18, 95–120. https://doi.org/10.46661/revintpensampolit.8778

Hauer, M. P., Krafft, T. D., & Zweig, K. (2023). Overview of transparency and inspectability mechanisms to achieve accountability of artificial intelligence systems. Data & Policy, 5, e36. https://doi.org/10.1017/dap.2023.30

International Organization for Standardization. (2023). Information technology—Artificial intelligence—Management system (ISO/IEC 42001:2023). International Organization for Standardization.

Koshiyama, A., Kazim, E., Treleaven, P., Rai, P., Szpruch, L., Pavey, G., Ahamat, G., Leutner, F., Goebel, R., Knight, A., Adams, J., Hitrova, C., Barnett, J., Nachev, P., Barber, D., Chamorro-Premuzic, T., Klemmer, K., Gregorovic, M., Khan, S., … Chatterjee, S. (2024). Towards algorithm auditing: managing legal, ethical and technological risks of AI, ML and associated algorithms. Royal Society Open Science, 11(5). https://doi.org/10.1098/rsos.230859

Landa Arroyo, C. (2022). CONSTITUCIÓN, DERECHOS FUNDAMENTALES, INTELIGENCIA ARTIFICIAL Y ALGORITMOS. Revista do Direito, (66), 139–158. https://doi.org/10.17058/rdunisc.vi66.17635

Laux, J. (2024). Institutionalised distrust and human oversight of artificial intelligence: towards a democratic design of AI governance under the European Union AI Act. AI & SOCIETY, 39(6), 2853–2866. https://doi.org/10.1007/s00146-023-01777-z

Laux, J., Wachter, S., & Mittelstadt, B. (2024). Trustworthy artificial intelligence and the European Union <scp>AI</scp> act: On the conflation of trustworthiness and acceptability of risk. Regulation & Governance, 18(1), 3–32. https://doi.org/10.1111/rego.12512

Lazcoz Moratinos, G., & Castillo Parrilla, J. A. (2020). Valoración algorítmica ante los derechos humanos y el Reglamento General de Protección de Datos: el caso SyRI. Revista Chilena de Derecho y Tecnología, 9(1), 207. https://doi.org/10.5354/0719-2584.2020.56843

Mantelero, A. (2024). The Fundamental Rights Impact Assessment (FRIA) in the AI Act: Roots, legal obligations and key elements for a model template. Computer Law & Security Review, 54, 106020. https://doi.org/10.1016/j.clsr.2024.106020

Mäntymäki, M., Minkkinen, M., Birkstedt, T., & Viljanen, M. (2022). Defining organizational AI governance. AI and Ethics, 2(4), 603–609. https://doi.org/10.1007/s43681-022-00143-x

Minkkinen, M., Laine, J., & Mäntymäki, M. (2022). Continuous Auditing of Artificial Intelligence: a Conceptualization and Assessment of Tools and Frameworks. Digital Society, 1(3), 21. https://doi.org/10.1007/s44206-022-00022-2

Mokander, J. (2023). Auditing of AI: Legal, ethical and technical approaches. Digital Society, 2, 49. https://doi.org/10.1007/s44206-023-00074-y

Mokander, J., & Floridi, L. (2021). Ethics-based auditing to develop trustworthy AI. Minds and Machines, 31, 323–327. https://doi.org/10.1007/s11023-021-09557-8

Morales Ramirez, G. (2023). Problematica antropologica detras de la discriminacion generada a partir de los algoritmos de la inteligencia artificial. Medicina y Etica, 34(2), 429–480. https://doi.org/10.36105/mye.2023v34n2.04

Mota Sanchez, E. M., & Herrera Exposito, E. (2023). Auditoria algoritmica en la inteligencia artificial en el sector publico. Proyecciones, 17, e025. https://doi.org/10.24215/26185474e025

Obregón Fernández, A., & Lazcoz Moratinos, G. (2021). La supervisión humana de los sistemas de inteligencia artificial de alto riesgo. Aportaciones desde el Derecho Internacional Humanitario y el Derecho de la Unión Europea. Revista Electrónica de Estudios Internacionales, (42). https://doi.org/10.17103/reei.42.08

Page, M. J., McKenzie, J. E., Bossuyt, P. M., Boutron, I., Hoffmann, T. C., Mulrow, C. D., & others. (2021). The PRISMA 2020 statement: An updated guideline for reporting systematic reviews. BMJ, 372, n71. https://doi.org/10.1136/bmj.n71

Papagiannidis, E., Mikalef, P., & Conboy, K. (2025). Responsible artificial intelligence governance: A review and research framework. The Journal of Strategic Information Systems, 34(2), 101885. https://doi.org/10.1016/j.jsis.2024.101885

Presno Linera, M. Á. (2025). El Reglamento europeo de inteligencia artificial: ¿garantía del mercado o defensa de los derechos? Revista Española de Derecho Constitucional, (135), 13–49. https://doi.org/10.18042/cepc/redc.135.01

Ramírez-Bustamente, N., & Páez, A. (2021). Análisis jurídico de la discriminación algorítmica en los procesos de selección laboral (A Legal Analysis of Algorithmic Discrimination in Hiring Processes). SSRN Electronic Journal. https://doi.org/10.2139/ssrn.3765741

Salvador Serna, M. (2021). Inteligencia artificial y gobernanza de datos en las administraciones públicas: reflexiones y evidencias para su desarrollo. Gestión y Análisis de Políticas Públicas, 20–32. https://doi.org/10.24965/gapp.i26.10855

Sánchez Acevedo, M. E. (2022). La inteligencia artificial en el sector público y su límite respecto de los derechos fundamentals. Estudios constitucionales, 20(2), 257–284. https://doi.org/10.4067/S0718-52002022000200257

Simón Castellano, P. (2025). L'avaluació d'impacte en els drets fonamentals de l'ús d'intel·ligència artificial en el sector públic: models i metodologies en perspectiva comparada. Revista Catalana de Dret Públic, 4–24. https://doi.org/10.58992/rcdp.i71.2025.4486

Tabassi, E. (2023). Artificial Intelligence Risk Management Framework (AI RMF 1.0) (NIST AI 100-1). National Institute of Standards and Technology. https://doi.org/10.6028/NIST.AI.100-1

Vida Fernandez, J. (2022). La gobernanza de los riesgos digitales: desafios y avances en la regulacion de la inteligencia artificial. Cuadernos de Derecho Transnacional, 14(1), 489–503. https://doi.org/10.20318/cdt.2022.6695

Williams, R., Cloete, R., Cobbe, J., Cottrill, C., Edwards, P., Markovic, M., Naja, I., Ryan, F., Singh, J., & Pang, W. (2022). From transparency to accountability of intelligent

systems: Moving beyond aspirations. Data & Policy, 4, e7. https://doi.org/10.1017/dap.2021.37

Declaraciones finales

Conflicto de intereses: El autor declara que no existe conflicto de interés posible.

Financiamiento: No existió asistencia financiera de partes externas al presente artículo.

Agradecimiento: A mi esposa Diana, a mis hijas Violeta y Carolina, a mi primo Jairo y en especial a mi amigo Andrés Fdo Ríos.

Uso de inteligencia artificial generativa: Durante la preparación del manuscrito se utilizó ChatGPT, herramienta de inteligencia artificial generativa desarrollada por OpenAI, como apoyo para la revisión de estilo, organización y claridad de la redacción. El autor revisó, contrastó y aprobó el contenido final, verificó las fuentes y referencias empleadas y asume íntegramente la responsabilidad por la argumentación, exactitud, integridad y conclusiones del artículo. ChatGPT no se reconoce como autor ni como fuente de evidencia científica.

Nota editorial: El artículo no deriva de una publicación anterior ni se encuentra sometido simultáneamente a otra revista.