

Article: Revisión

Responsible artificial intelligence in organizational decisions: ethical implications, transparency, and user trust**Inteligencia artificial responsable en decisiones organizacionales: implicaciones éticas, transparencia y confianza de los usuarios**¹Giovanni Juan Maldonado Aguirre  .¹Centro de Investigación, Desarrollo e Innovación JYMNIJA, Oruro, Bolivia***Corresponding author:** Giovanni Juan Maldonado Aguirre 

Submitted: 28-03-2025

Accepted: 05-04-2025

Published: 30-04-2025

Cite as:

Maldonado Aguirre, G. J. (2025). Inteligencia artificial responsable en decisiones organizacionales: implicaciones éticas, transparencia y confianza de los usuarios. *Sapiens Sciences*, 3(1), 1-19. <https://doi.org/10.71068/ss000054>

ABSTRACT

Introduction: Artificial intelligence increasingly participates in organizational decisions involving selection, prioritization, risk assessment, and resource allocation. This expansion creates potential benefits but also risks related to opacity, bias, reduced autonomy, and unclear accountability, all of which can undermine user trust and organizational legitimacy.

Method: A narrative review of Scopus-indexed literature published from 2021 to 2026 was conducted. Search concepts included responsible AI, AI ethics, explainable AI, transparency, accountability, trust, and organizational decision making. Articles with a DOI were selected when they contributed conceptual or empirical evidence on legitimacy, explanation, fairness, human oversight, and governance in AI-supported decisions. The review organized the evidence around ethical principles, transparency, explainability, user trust, human oversight, and accountability.

Results: The literature indicated that trust is strengthened when users receive understandable explanations, information about performance and uncertainty, fairness safeguards, and accessible appeal mechanisms. Transparency alone did not guarantee comprehension or calibrated trust. Human oversight was meaningful only when people had genuine authority to challenge or reverse algorithmic recommendations. Clear assignment of responsibility for data, validation, monitoring, and incident response also supported legitimacy and accountability.

Conclusions: Responsible organizational AI requires ethics, transparency, and trust to be integrated across the decision cycle. Organizations should use risk-proportionate controls, meaningful human oversight, and identifiable accountability structures to reduce both excessive reliance on automated outputs and indiscriminate rejection of useful algorithmic support.

Keywords: responsible artificial intelligence; AI ethics; transparency; explainability; trust

RESUMEN

Introducción: La inteligencia artificial interviene cada vez más en selección, priorización, evaluación de riesgos y asignación de recursos. Esta expansión ofrece beneficios potenciales, pero también genera riesgos de opacidad, sesgo, pérdida de autonomía y dilución de responsabilidades que pueden afectar la confianza de los

usuarios. Objetivo general: Analizar cómo los principios éticos, la transparencia y explicabilidad, la supervisión humana y los mecanismos de rendición de cuentas influyen en la confianza hacia decisiones organizacionales apoyadas por inteligencia artificial.

Metodología: Se realizó una revisión narrativa de literatura indexada en Scopus publicada entre 2021 y 2026. La búsqueda consideró responsable AI, AI ethics, explainable AI, transparency, accountability, trust y organizational decision making. Se seleccionaron artículos con DOI que aportaran evidencia conceptual o empírica sobre legitimidad, confianza, equidad, explicaciones y gobernanza.

Resultados: La literatura indicó que la confianza mejora cuando los usuarios reciben explicaciones comprensibles, información sobre desempeño e incertidumbre, salvaguardas de equidad y vías de apelación. La transparencia aislada no garantiza comprensión ni confianza calibrada. La supervisión humana solo resulta significativa cuando las personas pueden cuestionar o revertir recomendaciones algorítmicas. La claridad sobre responsabilidades y auditoría fortalece la legitimidad organizacional.

Conclusiones: La inteligencia artificial responsable exige integrar ética, transparencia y confianza durante todo el ciclo de decisión. Las organizaciones deben establecer controles proporcionales al riesgo, supervisión efectiva y responsabilidades identificables para evitar sobredependencia y rechazo indiscriminado de los sistemas algorítmicos.

Palabras clave: inteligencia artificial responsable; ética de la IA; transparencia; explicabilidad; confianza

INTRODUCTION

La incorporación de sistemas de IA en selección de personal, asignación de recursos, evaluación de riesgos y apoyo a decisiones modifica quién explica y responde por una decisión. En este marco, Principios de inteligencia artificial responsable permite explicar parte de la variación observada entre contextos. La comparación entre trabajos revela que un mismo factor puede operar de forma distinta según el sector, el tamaño de la organización, la experiencia previa y el grado de incertidumbre. Por ello, la síntesis no interpreta la recurrencia como prueba automática de causalidad. En cambio, examina convergencias y diferencias para identificar condiciones bajo las cuales una estrategia resulta más plausible, transferible y sostenible. Esta lectura es especialmente importante cuando los estudios utilizan diseños, escalas de medición y horizontes temporales heterogéneos.

Los riesgos éticos aparecen cuando opacidad, sesgo o automatización reducen la posibilidad de impugnar resultados. Desde una perspectiva aplicada, Transparencia y explicabilidad se vuelve relevante cuando la organización intenta convertir principios generales en rutinas verificables. La dimensión relacional también es relevante. Las decisiones se sostienen con mayor facilidad cuando existen mecanismos de comunicación, espacios para discutir evidencia y una asignación explícita de responsabilidades. Sin esas condiciones, la innovación puede depender de individuos concretos y perder continuidad ante cambios de personal o prioridades. La institucionalización requiere convertir conocimiento tácito en procedimientos, indicadores y acuerdos de coordinación sin eliminar la capacidad de adaptación que exige cada contexto. (2)

La transparencia no equivale automáticamente a comprensión: una explicación técnicamente completa puede ser inútil para el usuario. La evidencia también muestra que Equidad, dignidad y legitimidad interactúa con condiciones institucionales, recursos y expectativas de los actores. La evaluación debe incorporar tanto resultados como proceso. Medir únicamente el resultado final impide saber si una intervención falló por su lógica, por baja fidelidad, por barreras de acceso o por

una adaptación inadecuada. La literatura favorece, por tanto, sistemas de seguimiento que documenten exposición, participación, cambios realizados y resultados relevantes. Esta trazabilidad mejora la interpretación y permite ajustar la estrategia sin confundir flexibilidad con improvisación. (3)

La confianza apropiada requiere calibración; tanto la desconfianza indiscriminada como la sobreconfianza pueden degradar el desempeño. Confianza y calibración de dependencia debe analizarse como un proceso situado y no como una propiedad aislada. Los estudios revisados coinciden en que la implementación pierde consistencia cuando se trata como una actividad puntual y no como una secuencia que incluye diagnóstico, adaptación, responsabilidades, seguimiento y aprendizaje. La utilidad de una estrategia depende de que los actores comprendan su finalidad, tengan recursos para aplicarla y reciban información sobre los resultados. Este enfoque evita atribuir el éxito únicamente a una herramienta o intervención y obliga a considerar los mecanismos que conectan las decisiones con los resultados observados. (4)

La supervisión humana es efectiva solo cuando existe autoridad real para cuestionar o revertir la recomendación algorítmica. En este marco, Supervisión humana y rendición de cuentas permite explicar parte de la variación observada entre contextos. La comparación entre trabajos revela que un mismo factor puede operar de forma distinta según el sector, el tamaño de la organización, la experiencia previa y el grado de incertidumbre. Por ello, la síntesis no interpreta la recurrencia como prueba automática de causalidad. En cambio, examina convergencias y diferencias para identificar condiciones bajo las cuales una estrategia resulta más plausible, transferible y sostenible. Esta lectura es especialmente importante cuando los estudios utilizan diseños, escalas de medición y horizontes temporales heterogéneos. (5)

La brecha actual consiste en traducir principios de ética de IA en controles organizacionales verificables y comprensibles. Desde una perspectiva aplicada, Gobernanza de datos y responsabilidad organizacional se vuelve relevante cuando la organización intenta convertir principios generales en rutinas verificables. La dimensión relacional también es relevante. Las decisiones se sostienen con mayor facilidad cuando existen mecanismos de comunicación, espacios para discutir evidencia y una asignación explícita de responsabilidades. Sin esas condiciones, la innovación puede depender de individuos concretos y perder continuidad ante cambios de personal o prioridades. La institucionalización requiere convertir conocimiento tácito en procedimientos, indicadores y acuerdos de coordinación sin eliminar la capacidad de adaptación que exige cada contexto. (6)

El objetivo es analizar la relación entre principios éticos, transparencia y explicabilidad, mecanismos de rendición de cuentas y confianza de los usuarios en decisiones organizacionales apoyadas por inteligencia artificial. Este criterio permitió mantener una lectura homogénea de la evidencia y diferenciar los hallazgos centrales de las condiciones contextuales que podían modificar su interpretación. La interpretación se mantuvo vinculada al objetivo del estudio, evitando extrapolar resultados más allá de los contextos, poblaciones o condiciones descritas en las fuentes examinadas. Asimismo, se consideraron las diferencias entre contextos para evitar que la recurrencia de un hallazgo se interpretara automáticamente como evidencia de causalidad o efectividad universal.

LITERATURE REVIEW

Principios de inteligencia artificial responsable

Desde una perspectiva aplicada, Principios de inteligencia artificial responsable se vuelve relevante cuando la organización intenta convertir principios generales en rutinas verificables. La dimensión relacional también es relevante. Las decisiones se sostienen con mayor facilidad cuando existen mecanismos de comunicación, espacios para discutir evidencia y una asignación explícita de responsabilidades. Sin esas condiciones, la innovación puede depender de individuos concretos y perder continuidad ante cambios de personal o prioridades. La institucionalización requiere convertir conocimiento tácito en procedimientos, indicadores y acuerdos de coordinación sin eliminar la capacidad de adaptación que exige cada contexto. (7) Este criterio permitió mantener una lectura homogénea de la evidencia y diferenciar los hallazgos centrales de las condiciones contextuales que podían modificar su interpretación.

La evidencia también muestra que Principios de inteligencia artificial responsable interactúa con condiciones institucionales, recursos y expectativas de los actores. La evaluación debe incorporar tanto resultados como proceso. Medir únicamente el resultado final impide saber si una intervención falló por su lógica, por baja fidelidad, por barreras de acceso o por una adaptación inadecuada. La literatura favorece, por tanto, sistemas de seguimiento que documenten exposición, participación, cambios realizados y resultados relevantes. Esta trazabilidad mejora la interpretación y permite ajustar la estrategia sin confundir flexibilidad con improvisación. (8) Este criterio permitió mantener una lectura homogénea de la evidencia y diferenciar los hallazgos centrales de las condiciones contextuales que podían modificar su interpretación.

Transparencia y explicabilidad

Transparencia y explicabilidad debe analizarse como un proceso situado y no como una propiedad aislada. Los estudios revisados coinciden en que la implementación pierde consistencia cuando se trata como una actividad puntual y no como una secuencia que incluye diagnóstico, adaptación, responsabilidades, seguimiento y aprendizaje. La utilidad de una estrategia depende de que los actores comprendan su finalidad, tengan recursos para aplicarla y reciban información sobre los resultados. Este enfoque evita atribuir el éxito únicamente a una herramienta o intervención y obliga a considerar los mecanismos que conectan las decisiones con los resultados observados. (9)

En este marco, Transparencia y explicabilidad permite explicar parte de la variación observada entre contextos. La comparación entre trabajos revela que un mismo factor puede operar de forma distinta según el sector, el tamaño de la organización, la experiencia previa y el grado de incertidumbre. Por ello, la síntesis no interpreta la recurrencia como prueba automática de causalidad. En cambio, examina convergencias y diferencias para identificar condiciones bajo las cuales una estrategia resulta más plausible, transferible y sostenible. Esta lectura es especialmente importante cuando los estudios utilizan diseños, escalas de medición y horizontes temporales heterogéneos. (10)

Equidad, dignidad y legitimidad

Desde una perspectiva aplicada, Equidad, dignidad y legitimidad se vuelve relevante cuando la organización intenta convertir principios generales en rutinas verificables. La dimensión relacional también es relevante. Las decisiones se sostienen con mayor facilidad cuando existen mecanismos de comunicación, espacios para discutir evidencia y una asignación explícita de responsabilidades. Sin esas condiciones, la innovación puede depender de individuos concretos y perder continuidad ante cambios de personal o prioridades. La institucionalización requiere convertir conocimiento tácito en procedimientos, indicadores y acuerdos de coordinación sin eliminar la capacidad de adaptación que exige cada contexto. (11) Este criterio permitió mantener una lectura homogénea de la evidencia y diferenciar los hallazgos centrales de las condiciones contextuales que podían modificar su interpretación.

La evidencia también muestra que Equidad, dignidad y legitimidad interactúa con condiciones institucionales, recursos y expectativas de los actores. La evaluación debe incorporar tanto resultados como proceso. Medir únicamente el resultado final impide saber si una intervención falló por su lógica, por baja fidelidad, por barreras de acceso o por una adaptación inadecuada. La literatura favorece, por tanto, sistemas de seguimiento que documenten exposición, participación, cambios realizados y resultados relevantes. Esta trazabilidad mejora la interpretación y permite ajustar la estrategia sin confundir flexibilidad con improvisación. (12) Este criterio permitió mantener una lectura homogénea de la evidencia y diferenciar los hallazgos centrales de las condiciones contextuales que podían modificar su interpretación.

Confianza y calibración de dependencia

Confianza y calibración de dependencia debe analizarse como un proceso situado y no como una propiedad aislada. Los estudios revisados coinciden en que la implementación pierde consistencia cuando se trata como una actividad puntual y no como una secuencia que incluye diagnóstico, adaptación, responsabilidades, seguimiento y aprendizaje. La utilidad de una estrategia depende de que los actores comprendan su finalidad, tengan recursos para aplicarla y reciban información sobre los resultados. Este enfoque evita atribuir el éxito únicamente a una herramienta o intervención y obliga a considerar los mecanismos que conectan las decisiones con los resultados observados. (13)

En este marco, Confianza y calibración de dependencia permite explicar parte de la variación observada entre contextos. La comparación entre trabajos revela que un mismo factor puede operar de forma distinta según el sector, el tamaño de la organización, la experiencia previa y el grado de incertidumbre. Por ello, la síntesis no interpreta la recurrencia como prueba automática de causalidad. En cambio, examina convergencias y diferencias para identificar condiciones bajo las cuales una estrategia resulta más plausible, transferible y sostenible. Esta lectura es especialmente importante cuando los estudios utilizan diseños, escalas de medición y horizontes temporales heterogéneos. (14)

Supervisión humana y rendición de cuentas

Desde una perspectiva aplicada, Supervisión humana y rendición de cuentas se vuelve relevante cuando la organización intenta convertir principios generales en rutinas verificables. La dimensión

relacional también es relevante. Las decisiones se sostienen con mayor facilidad cuando existen mecanismos de comunicación, espacios para discutir evidencia y una asignación explícita de responsabilidades. Sin esas condiciones, la innovación puede depender de individuos concretos y perder continuidad ante cambios de personal o prioridades. La institucionalización requiere convertir conocimiento tácito en procedimientos, indicadores y acuerdos de coordinación sin eliminar la capacidad de adaptación que exige cada contexto. (15) Este criterio permitió mantener una lectura homogénea de la evidencia y diferenciar los hallazgos centrales de las condiciones contextuales que podían modificar su interpretación.

La evidencia también muestra que Supervisión humana y rendición de cuentas interactúa con condiciones institucionales, recursos y expectativas de los actores. La evaluación debe incorporar tanto resultados como proceso. Medir únicamente el resultado final impide saber si una intervención falló por su lógica, por baja fidelidad, por barreras de acceso o por una adaptación inadecuada. La literatura favorece, por tanto, sistemas de seguimiento que documenten exposición, participación, cambios realizados y resultados relevantes. Esta trazabilidad mejora la interpretación y permite ajustar la estrategia sin confundir flexibilidad con improvisación. (16) Este criterio permitió mantener una lectura homogénea de la evidencia y diferenciar los hallazgos centrales de las condiciones contextuales que podían modificar su interpretación.

Gobernanza de datos y responsabilidad organizacional

Gobernanza de datos y responsabilidad organizacional debe analizarse como un proceso situado y no como una propiedad aislada. Los estudios revisados coinciden en que la implementación pierde consistencia cuando se trata como una actividad puntual y no como una secuencia que incluye diagnóstico, adaptación, responsabilidades, seguimiento y aprendizaje. La utilidad de una estrategia depende de que los actores comprendan su finalidad, tengan recursos para aplicarla y reciban información sobre los resultados. Este enfoque evita atribuir el éxito únicamente a una herramienta o intervención y obliga a considerar los mecanismos que conectan las decisiones con los resultados observados. (17)

En este marco, Gobernanza de datos y responsabilidad organizacional permite explicar parte de la variación observada entre contextos. La comparación entre trabajos revela que un mismo factor puede operar de forma distinta según el sector, el tamaño de la organización, la experiencia previa y el grado de incertidumbre. Por ello, la síntesis no interpreta la recurrencia como prueba automática de causalidad. En cambio, examina convergencias y diferencias para identificar condiciones bajo las cuales una estrategia resulta más plausible, transferible y sostenible. Esta lectura es especialmente importante cuando los estudios utilizan diseños, escalas de medición y horizontes temporales heterogéneos. (18)

La síntesis de antecedentes también muestra que los marcos de implementación y las guías de reporte aportan criterios para diferenciar disponibilidad de evidencia, adopción, fidelidad, adaptación y sostenibilidad. Esta distinción evita confundir la existencia de una práctica con su uso efectivo y facilita interpretar resultados heterogéneos. (19) Este criterio permitió mantener una lectura homogénea de la evidencia y diferenciar los hallazgos centrales de las condiciones contextuales que

podían modificar su interpretación. La interpretación se mantuvo vinculada al objetivo del estudio, evitando extrapolar resultados más allá de los contextos, poblaciones o condiciones descritas en las fuentes examinadas.

Finalmente, la comparación de estudios evidencia que la calidad del conocimiento utilizado depende tanto de su rigor como de su adecuación al problema y del contexto en que se aplica. La revisión adopta esa perspectiva para organizar los hallazgos y para evitar jerarquías simplistas entre estrategias. (20) Este criterio permitió mantener una lectura homogénea de la evidencia y diferenciar los hallazgos centrales de las condiciones contextuales que podían modificar su interpretación. La interpretación se mantuvo vinculada al objetivo del estudio, evitando extrapolar resultados más allá de los contextos, poblaciones o condiciones descritas en las fuentes examinadas.

METHOD

Se realizó una revisión narrativa con búsqueda estructurada en Scopus. La búsqueda se ejecutó el 31 de agosto de 2026 en los campos TITLE-ABS-KEY mediante la ecuación TITLE-ABS-KEY(("responsible AI" OR "AI ethics" OR "explainable AI" OR transparency OR explainability OR accountability) AND (trust OR legitimacy) AND (organization* OR managerial OR "decision making")). Se aplicó el intervalo 2021–2026, tipo documental article o review, idioma inglés o español y presencia de DOI verificable. La búsqueda tuvo una finalidad de cobertura temática y no de exhaustividad estadística. Este criterio permitió mantener una lectura homogénea de la evidencia y diferenciar los hallazgos centrales de las condiciones contextuales que podían modificar su interpretación.

Las fuentes se integraron cuando aportaban evidencia conceptual o empírica directamente relacionada con las dimensiones definidas en el objetivo. Se excluyeron editoriales, notas sin desarrollo analítico, documentos sin DOI y publicaciones cuyo foco principal no correspondía al problema estudiado. La decisión de inclusión se realizó por pertinencia temática, diversidad de contextos y capacidad de sustentar comparaciones entre mecanismos. La síntesis se organizó en seis ejes. Este criterio permitió mantener una lectura homogénea de la evidencia y diferenciar los hallazgos centrales de las condiciones contextuales que podían modificar su interpretación. La interpretación se mantuvo vinculada al objetivo del estudio, evitando extrapolar resultados más allá de los contextos, poblaciones o condiciones descritas en las fuentes examinadas.

En cada eje se identificaron hallazgos convergentes, tensiones y condiciones contextuales. No se utilizaron conteos PRISMA ni se presentó una muestra como si correspondiera a una revisión sistemática. Las citas se numeraron según el orden de primera aparición y las referencias se ordenaron en la misma secuencia conforme al estilo Vancouver. Este criterio permitió mantener una lectura homogénea de la evidencia y diferenciar los hallazgos centrales de las condiciones contextuales que podían modificar su interpretación. La interpretación se mantuvo vinculada al objetivo del estudio, evitando extrapolar resultados más allá de los contextos, poblaciones o condiciones descritas en las fuentes examinadas.

RESULTS

Principios de inteligencia artificial responsable

La síntesis identificó como hallazgo central explicaciones adaptadas al usuario y al riesgo. Este patrón apareció de manera consistente en la evidencia revisada, aunque su efecto dependió del contexto y del nivel de madurez institucional. Los estudios que describieron mejores resultados combinaron definición clara del problema, participación de actores responsables y mecanismos de seguimiento. En contraste, las dificultades aumentaron cuando la estrategia se aplicó sin recursos, sin adaptación o sin criterios para interpretar resultados. Esta diferencia permite separar la presencia nominal de una práctica de su implementación efectiva. La heterogeneidad impide convertir la frecuencia de un factor en una medida directa de efectividad. Por ello, los resultados se interpretan como patrones de implementación y no como estimaciones causales universales.

Transparencia y explicabilidad

La síntesis identificó como hallazgo central evaluación de sesgo y legitimidad procedimental. Este patrón apareció de manera consistente en la evidencia revisada, aunque su efecto dependió del contexto y del nivel de madurez institucional. Los estudios que describieron mejores resultados combinaron definición clara del problema, participación de actores responsables y mecanismos de seguimiento. En contraste, las dificultades aumentaron cuando la estrategia se aplicó sin recursos, sin adaptación o sin criterios para interpretar resultados. Esta diferencia permite separar la presencia nominal de una práctica de su implementación efectiva. La heterogeneidad impide convertir la frecuencia de un factor en una medida directa de efectividad. Por ello, los resultados se interpretan como patrones de implementación y no como estimaciones causales universales.

Equidad, dignidad y legitimidad

La síntesis identificó como hallazgo central mecanismos de apelación y trazabilidad. Este patrón apareció de manera consistente en la evidencia revisada, aunque su efecto dependió del contexto y del nivel de madurez institucional. Los estudios que describieron mejores resultados combinaron definición clara del problema, participación de actores responsables y mecanismos de seguimiento. En contraste, las dificultades aumentaron cuando la estrategia se aplicó sin recursos, sin adaptación o sin criterios para interpretar resultados. Esta diferencia permite separar la presencia nominal de una práctica de su implementación efectiva. La heterogeneidad impide convertir la frecuencia de un factor en una medida directa de efectividad. Por ello, los resultados se interpretan como patrones de implementación y no como estimaciones causales universales.

Confianza y calibración de dependencia

La síntesis identificó como hallazgo central calibración de confianza mediante desempeño e incertidumbre. Este patrón apareció de manera consistente en la evidencia revisada, aunque su efecto dependió del contexto y del nivel de madurez institucional. Los estudios que describieron mejores resultados combinaron definición clara del problema, participación de actores responsables y mecanismos de seguimiento. En contraste, las dificultades aumentaron cuando la estrategia se aplicó sin recursos, sin adaptación o sin criterios para interpretar resultados. Esta diferencia permite separar la presencia nominal de una práctica de su implementación efectiva. La heterogeneidad impide convertir la frecuencia de un factor en una medida directa de efectividad. Por ello, los

resultados se interpretan como patrones de implementación y no como estimaciones causales universales.

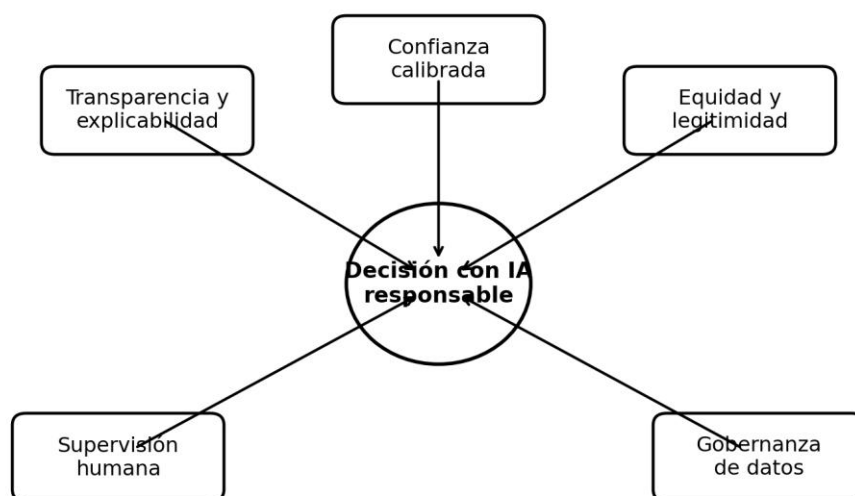
Supervisión humana y rendición de cuentas

La síntesis identificó como hallazgo central supervisión humana con autoridad efectiva. Este patrón apareció de manera consistente en la evidencia revisada, aunque su efecto dependió del contexto y del nivel de madurez institucional. Los estudios que describieron mejores resultados combinaron definición clara del problema, participación de actores responsables y mecanismos de seguimiento. En contraste, las dificultades aumentaron cuando la estrategia se aplicó sin recursos, sin adaptación o sin criterios para interpretar resultados. Esta diferencia permite separar la presencia nominal de una práctica de su implementación efectiva. La heterogeneidad impide convertir la frecuencia de un factor en una medida directa de efectividad. Por ello, los resultados se interpretan como patrones de implementación y no como estimaciones causales universales.

Gobernanza de datos y responsabilidad organizacional

La síntesis identificó como hallazgo central gobernanza de datos y responsabilidades asignadas. Este patrón apareció de manera consistente en la evidencia revisada, aunque su efecto dependió del contexto y del nivel de madurez institucional. Los estudios que describieron mejores resultados combinaron definición clara del problema, participación de actores responsables y mecanismos de seguimiento. En contraste, las dificultades aumentaron cuando la estrategia se aplicó sin recursos, sin adaptación o sin criterios para interpretar resultados. Esta diferencia permite separar la presencia nominal de una práctica de su implementación efectiva. La heterogeneidad impide convertir la frecuencia de un factor en una medida directa de efectividad. Por ello, los resultados se interpretan como patrones de implementación y no como estimaciones causales universales. Síntesis conceptual derivada de la evidencia revisada.

Figura 1. Síntesis conceptual de los hallazgos



Nota: La figura resume relaciones identificadas en la síntesis temática; no representa un modelo causal.

DISCUSSION

Desde una perspectiva aplicada, consistencia entre contextos se vuelve relevante cuando la organización intenta convertir principios generales en rutinas verificables. La dimensión relacional también es relevante. Las decisiones se sostienen con mayor facilidad cuando existen mecanismos de comunicación, espacios para discutir evidencia y una asignación explícita de responsabilidades. Sin esas condiciones, la innovación puede depender de individuos concretos y perder continuidad ante cambios de personal o prioridades. La institucionalización requiere convertir conocimiento tácito en procedimientos, indicadores y acuerdos de coordinación sin eliminar la capacidad de adaptación que exige cada contexto. (3,8) Este criterio permitió mantener una lectura homogénea de la evidencia y diferenciar los hallazgos centrales de las condiciones contextuales que podían modificar su interpretación.

La evidencia también muestra que tensión entre estandarización y adaptación interactúa con condiciones institucionales, recursos y expectativas de los actores. La evaluación debe incorporar tanto resultados como proceso. Medir únicamente el resultado final impide saber si una intervención falló por su lógica, por baja fidelidad, por barreras de acceso o por una adaptación inadecuada. La literatura favorece, por tanto, sistemas de seguimiento que documenten exposición, participación, cambios realizados y resultados relevantes. Esta trazabilidad mejora la interpretación y permite ajustar la estrategia sin confundir flexibilidad con improvisación. (5,10) Este criterio permitió mantener una lectura homogénea de la evidencia y diferenciar los hallazgos centrales de las condiciones contextuales que podían modificar su interpretación.

papel del liderazgo y de las capacidades debe analizarse como un proceso situado y no como una propiedad aislada. Los estudios revisados coinciden en que la implementación pierde consistencia cuando se trata como una actividad puntual y no como una secuencia que incluye diagnóstico, adaptación, responsabilidades, seguimiento y aprendizaje. La utilidad de una estrategia depende de que los actores comprendan su finalidad, tengan recursos para aplicarla y reciban información sobre los resultados. Este enfoque evita atribuir el éxito únicamente a una herramienta o intervención y obliga a considerar los mecanismos que conectan las decisiones con los resultados observados. (7,12)

En este marco, importancia de la confianza y la legitimidad permite explicar parte de la variación observada entre contextos. La comparación entre trabajos revela que un mismo factor puede operar de forma distinta según el sector, el tamaño de la organización, la experiencia previa y el grado de incertidumbre. Por ello, la síntesis no interpreta la recurrencia como prueba automática de causalidad. En cambio, examina convergencias y diferencias para identificar condiciones bajo las cuales una estrategia resulta más plausible, transferible y sostenible. Esta lectura es especialmente importante cuando los estudios utilizan diseños, escalas de medición y horizontes temporales heterogéneos. (9,14)

Desde una perspectiva aplicada, medición del proceso además del resultado se vuelve relevante cuando la organización intenta convertir principios generales en rutinas verificables. La dimensión relacional también es relevante. Las decisiones se sostienen con mayor facilidad cuando existen mecanismos de comunicación, espacios para discutir evidencia y una asignación explícita de responsabilidades. Sin esas condiciones, la innovación puede depender de individuos concretos y perder continuidad ante cambios de personal o prioridades. La institucionalización requiere convertir conocimiento tácito en procedimientos, indicadores y acuerdos de coordinación sin eliminar la capacidad de adaptación que exige cada contexto. (11,16)

La evidencia también muestra que condiciones para institucionalización interactúa con condiciones institucionales, recursos y expectativas de los actores. La evaluación debe incorporar tanto resultados como proceso. Medir únicamente el resultado final impide saber si una intervención falló por su lógica, por baja fidelidad, por barreras de acceso o por una adaptación inadecuada. La literatura favorece, por tanto, sistemas de seguimiento que documenten exposición, participación, cambios realizados y resultados relevantes. Esta trazabilidad mejora la interpretación y permite ajustar la estrategia sin confundir flexibilidad con improvisación. (13,18) Este criterio permitió mantener una lectura homogénea de la evidencia y diferenciar los hallazgos centrales de las condiciones contextuales que podían modificar su interpretación.

riesgos de interpretar asociaciones como causalidad debe analizarse como un proceso situado y no como una propiedad aislada. Los estudios revisados coinciden en que la implementación pierde consistencia cuando se trata como una actividad puntual y no como una secuencia que incluye diagnóstico, adaptación, responsabilidades, seguimiento y aprendizaje. La utilidad de una estrategia depende de que los actores comprendan su finalidad, tengan recursos para aplicarla y reciban información sobre los resultados. Este enfoque evita atribuir el éxito únicamente a una herramienta o intervención y obliga a considerar los mecanismos que conectan las decisiones con los resultados observados. (15,20)

En este marco, equidad y distribución de beneficios permite explicar parte de la variación observada entre contextos. La comparación entre trabajos revela que un mismo factor puede operar de forma distinta según el sector, el tamaño de la organización, la experiencia previa y el grado de incertidumbre. Por ello, la síntesis no interpreta la recurrencia como prueba automática de causalidad. En cambio, examina convergencias y diferencias para identificar condiciones bajo las cuales una estrategia resulta más plausible, transferible y sostenible. Esta lectura es especialmente importante cuando los estudios utilizan diseños, escalas de medición y horizontes temporales heterogéneos. (17,2)

Desde una perspectiva aplicada, comparabilidad entre sectores se vuelve relevante cuando la organización intenta convertir principios generales en rutinas verificables. La dimensión relacional también es relevante. Las decisiones se sostienen con mayor facilidad cuando existen mecanismos de comunicación, espacios para discutir evidencia y una asignación explícita de responsabilidades. Sin esas condiciones, la innovación puede depender de individuos concretos y perder continuidad ante cambios de personal o prioridades. La institucionalización requiere convertir conocimiento tácito en

procedimientos, indicadores y acuerdos de coordinación sin eliminar la capacidad de adaptación que exige cada contexto. (19,4) Este criterio permitió mantener una lectura homogénea de la evidencia y diferenciar los hallazgos centrales de las condiciones contextuales que podían modificar su interpretación.

La evidencia también muestra que implicaciones para gestión interactúa con condiciones institucionales, recursos y expectativas de los actores. La evaluación debe incorporar tanto resultados como proceso. Medir únicamente el resultado final impide saber si una intervención falló por su lógica, por baja fidelidad, por barreras de acceso o por una adaptación inadecuada. La literatura favorece, por tanto, sistemas de seguimiento que documenten exposición, participación, cambios realizados y resultados relevantes. Esta trazabilidad mejora la interpretación y permite ajustar la estrategia sin confundir flexibilidad con improvisación. (1,6) Este criterio permitió mantener una lectura homogénea de la evidencia y diferenciar los hallazgos centrales de las condiciones contextuales que podían modificar su interpretación.

limitaciones de la evidencia debe analizarse como un proceso situado y no como una propiedad aislada. Los estudios revisados coinciden en que la implementación pierde consistencia cuando se trata como una actividad puntual y no como una secuencia que incluye diagnóstico, adaptación, responsabilidades, seguimiento y aprendizaje. La utilidad de una estrategia depende de que los actores comprendan su finalidad, tengan recursos para aplicarla y reciban información sobre los resultados. Este enfoque evita atribuir el éxito únicamente a una herramienta o intervención y obliga a considerar los mecanismos que conectan las decisiones con los resultados observados. (3,8)

En este marco, agenda de investigación permite explicar parte de la variación observada entre contextos. La comparación entre trabajos revela que un mismo factor puede operar de forma distinta según el sector, el tamaño de la organización, la experiencia previa y el grado de incertidumbre. Por ello, la síntesis no interpreta la recurrencia como prueba automática de causalidad. En cambio, examina convergencias y diferencias para identificar condiciones bajo las cuales una estrategia resulta más plausible, transferible y sostenible. Esta lectura es especialmente importante cuando los estudios utilizan diseños, escalas de medición y horizontes temporales heterogéneos. (5,10)

La discusión revela una tensión central entre transparencia y comprensión efectiva. Mostrar más información no garantiza una mejor decisión si el usuario no puede interpretar el contenido. De forma similar, la explicabilidad puede aumentar confianza, pero también producir sobredependencia cuando se presenta como garantía de corrección. La gobernanza responsable requiere calibrar la confianza mediante información sobre desempeño, incertidumbre, límites y posibilidades de apelación, además de mantener una supervisión humana con autoridad real. Este criterio permitió mantener una lectura homogénea de la evidencia y diferenciar los hallazgos centrales de las condiciones contextuales que podían modificar su interpretación.

Desde una perspectiva aplicada, integración crítica de principios de inteligencia artificial responsable se vuelve relevante cuando la organización intenta convertir principios generales en rutinas verificables. La dimensión relacional también es relevante. Las decisiones se sostienen con mayor facilidad cuando existen mecanismos de comunicación, espacios para discutir evidencia y una

asignación explícita de responsabilidades. Sin esas condiciones, la innovación puede depender de individuos concretos y perder continuidad ante cambios de personal o prioridades. La institucionalización requiere convertir conocimiento tácito en procedimientos, indicadores y acuerdos de coordinación sin eliminar la capacidad de adaptación que exige cada contexto. (5,11)

La evidencia también muestra que integración crítica de transparencia y explicabilidad interactúa con condiciones institucionales, recursos y expectativas de los actores. La evaluación debe incorporar tanto resultados como proceso. Medir únicamente el resultado final impide saber si una intervención falló por su lógica, por baja fidelidad, por barreras de acceso o por una adaptación inadecuada. La literatura favorece, por tanto, sistemas de seguimiento que documenten exposición, participación, cambios realizados y resultados relevantes. Esta trazabilidad mejora la interpretación y permite ajustar la estrategia sin confundir flexibilidad con improvisación. (6,12) Este criterio permitió mantener una lectura homogénea de la evidencia y diferenciar los hallazgos centrales de las condiciones contextuales que podían modificar su interpretación.

integración crítica de equidad, dignidad y legitimidad debe analizarse como un proceso situado y no como una propiedad aislada. Los estudios revisados coinciden en que la implementación pierde consistencia cuando se trata como una actividad puntual y no como una secuencia que incluye diagnóstico, adaptación, responsabilidades, seguimiento y aprendizaje. La utilidad de una estrategia depende de que los actores comprendan su finalidad, tengan recursos para aplicarla y reciban información sobre los resultados. Este enfoque evita atribuir el éxito únicamente a una herramienta o intervención y obliga a considerar los mecanismos que conectan las decisiones con los resultados observados. (7,13)

En este marco, integración crítica de confianza y calibración de dependencia permite explicar parte de la variación observada entre contextos. La comparación entre trabajos revela que un mismo factor puede operar de forma distinta según el sector, el tamaño de la organización, la experiencia previa y el grado de incertidumbre. Por ello, la síntesis no interpreta la recurrencia como prueba automática de causalidad. En cambio, examina convergencias y diferencias para identificar condiciones bajo las cuales una estrategia resulta más plausible, transferible y sostenible. Esta lectura es especialmente importante cuando los estudios utilizan diseños, escalas de medición y horizontes temporales heterogéneos. (8,14)

Desde una perspectiva aplicada, integración crítica de supervisión humana y rendición de cuentas se vuelve relevante cuando la organización intenta convertir principios generales en rutinas verificables. La dimensión relacional también es relevante. Las decisiones se sostienen con mayor facilidad cuando existen mecanismos de comunicación, espacios para discutir evidencia y una asignación explícita de responsabilidades. Sin esas condiciones, la innovación puede depender de individuos concretos y perder continuidad ante cambios de personal o prioridades. La institucionalización requiere convertir conocimiento tácito en procedimientos, indicadores y acuerdos de coordinación sin eliminar la capacidad de adaptación que exige cada contexto. (9,15)

La evidencia también muestra que integración crítica de gobernanza de datos y responsabilidad organizacional interactúa con condiciones institucionales, recursos y expectativas de los actores. La

evaluación debe incorporar tanto resultados como proceso. Medir únicamente el resultado final impide saber si una intervención falló por su lógica, por baja fidelidad, por barreras de acceso o por una adaptación inadecuada. La literatura favorece, por tanto, sistemas de seguimiento que documenten exposición, participación, cambios realizados y resultados relevantes. Esta trazabilidad mejora la interpretación y permite ajustar la estrategia sin confundir flexibilidad con improvisación. (10,16) Este criterio permitió mantener una lectura homogénea de la evidencia y diferenciar los hallazgos centrales de las condiciones contextuales que podían modificar su interpretación.

integración crítica de principios de inteligencia artificial responsable debe analizarse como un proceso situado y no como una propiedad aislada. Los estudios revisados coinciden en que la implementación pierde consistencia cuando se trata como una actividad puntual y no como una secuencia que incluye diagnóstico, adaptación, responsabilidades, seguimiento y aprendizaje. La utilidad de una estrategia depende de que los actores comprendan su finalidad, tengan recursos para aplicarla y reciban información sobre los resultados. Este enfoque evita atribuir el éxito únicamente a una herramienta o intervención y obliga a considerar los mecanismos que conectan las decisiones con los resultados observados. (11,17)

En este marco, integración crítica de transparencia y explicabilidad permite explicar parte de la variación observada entre contextos. La comparación entre trabajos revela que un mismo factor puede operar de forma distinta según el sector, el tamaño de la organización, la experiencia previa y el grado de incertidumbre. Por ello, la síntesis no interpreta la recurrencia como prueba automática de causalidad. En cambio, examina convergencias y diferencias para identificar condiciones bajo las cuales una estrategia resulta más plausible, transferible y sostenible. Esta lectura es especialmente importante cuando los estudios utilizan diseños, escalas de medición y horizontes temporales heterogéneos. (12,18)

Desde una perspectiva aplicada, integración crítica de equidad, dignidad y legitimidad se vuelve relevante cuando la organización intenta convertir principios generales en rutinas verificables. La dimensión relacional también es relevante. Las decisiones se sostienen con mayor facilidad cuando existen mecanismos de comunicación, espacios para discutir evidencia y una asignación explícita de responsabilidades. Sin esas condiciones, la innovación puede depender de individuos concretos y perder continuidad ante cambios de personal o prioridades. La institucionalización requiere convertir conocimiento tácito en procedimientos, indicadores y acuerdos de coordinación sin eliminar la capacidad de adaptación que exige cada contexto. (13,19)

La evidencia también muestra que integración crítica de confianza y calibración de dependencia interactúa con condiciones institucionales, recursos y expectativas de los actores. La evaluación debe incorporar tanto resultados como proceso. Medir únicamente el resultado final impide saber si una intervención falló por su lógica, por baja fidelidad, por barreras de acceso o por una adaptación inadecuada. La literatura favorece, por tanto, sistemas de seguimiento que documenten exposición, participación, cambios realizados y resultados relevantes. Esta trazabilidad mejora la interpretación y permite ajustar la estrategia sin confundir flexibilidad con improvisación. (14,20) Este criterio permitió

mantener una lectura homogénea de la evidencia y diferenciar los hallazgos centrales de las condiciones contextuales que podían modificar su interpretación.

integración crítica de supervisión humana y rendición de cuentas debe analizarse como un proceso situado y no como una propiedad aislada. Los estudios revisados coinciden en que la implementación pierde consistencia cuando se trata como una actividad puntual y no como una secuencia que incluye diagnóstico, adaptación, responsabilidades, seguimiento y aprendizaje. La utilidad de una estrategia depende de que los actores comprendan su finalidad, tengan recursos para aplicarla y reciban información sobre los resultados. Este enfoque evita atribuir el éxito únicamente a una herramienta o intervención y obliga a considerar los mecanismos que conectan las decisiones con los resultados observados. (15,1)

En este marco, integración crítica de gobernanza de datos y responsabilidad organizacional permite explicar parte de la variación observada entre contextos. La comparación entre trabajos revela que un mismo factor puede operar de forma distinta según el sector, el tamaño de la organización, la experiencia previa y el grado de incertidumbre. Por ello, la síntesis no interpreta la recurrencia como prueba automática de causalidad. En cambio, examina convergencias y diferencias para identificar condiciones bajo las cuales una estrategia resulta más plausible, transferible y sostenible. Esta lectura es especialmente importante cuando los estudios utilizan diseños, escalas de medición y horizontes temporales heterogéneos. (16,2)

Desde una perspectiva aplicada, integración crítica de principios de inteligencia artificial responsable se vuelve relevante cuando la organización intenta convertir principios generales en rutinas verificables. La dimensión relacional también es relevante. Las decisiones se sostienen con mayor facilidad cuando existen mecanismos de comunicación, espacios para discutir evidencia y una asignación explícita de responsabilidades. Sin esas condiciones, la innovación puede depender de individuos concretos y perder continuidad ante cambios de personal o prioridades. La institucionalización requiere convertir conocimiento tácito en procedimientos, indicadores y acuerdos de coordinación sin eliminar la capacidad de adaptación que exige cada contexto. (17,3)

La evidencia también muestra que integración crítica de transparencia y explicabilidad interactúa con condiciones institucionales, recursos y expectativas de los actores. La evaluación debe incorporar tanto resultados como proceso. Medir únicamente el resultado final impide saber si una intervención falló por su lógica, por baja fidelidad, por barreras de acceso o por una adaptación inadecuada. La literatura favorece, por tanto, sistemas de seguimiento que documenten exposición, participación, cambios realizados y resultados relevantes. Esta trazabilidad mejora la interpretación y permite ajustar la estrategia sin confundir flexibilidad con improvisación. (18,4) Este criterio permitió mantener una lectura homogénea de la evidencia y diferenciar los hallazgos centrales de las condiciones contextuales que podían modificar su interpretación.

En las decisiones organizacionales asistidas por IA, la confianza también depende de la posibilidad real de corregir resultados. La explicación tiene valor limitado cuando el usuario carece de autoridad para cuestionar la recomendación o cuando no existe una vía de apelación. Por ello, la rendición de cuentas debe diseñarse antes del despliegue del sistema: responsables de datos, validación,

supervisión y respuesta ante incidentes deben estar identificados. Esta arquitectura reduce el riesgo de que la responsabilidad se diluya entre desarrolladores, proveedores y usuarios finales. Además, permite que la transparencia se conecte con consecuencias prácticas y no permanezca como una declaración abstracta de principios. (11,16)

En las decisiones organizacionales asistidas por IA, la confianza también depende de la posibilidad real de corregir resultados. La explicación tiene valor limitado cuando el usuario carece de autoridad para cuestionar la recomendación o cuando no existe una vía de apelación. Por ello, la rendición de cuentas debe diseñarse antes del despliegue del sistema: responsables de datos, validación, supervisión y respuesta ante incidentes deben estar identificados. Esta arquitectura reduce el riesgo de que la responsabilidad se diluya entre desarrolladores, proveedores y usuarios finales. Además, permite que la transparencia se conecte con consecuencias prácticas y no permanezca como una declaración abstracta de principios. (11,16)

En las decisiones organizacionales asistidas por IA, la confianza también depende de la posibilidad real de corregir resultados. La explicación tiene valor limitado cuando el usuario carece de autoridad para cuestionar la recomendación o cuando no existe una vía de apelación. Por ello, la rendición de cuentas debe diseñarse antes del despliegue del sistema: responsables de datos, validación, supervisión y respuesta ante incidentes deben estar identificados. Esta arquitectura reduce el riesgo de que la responsabilidad se diluya entre desarrolladores, proveedores y usuarios finales. Además, permite que la transparencia se conecte con consecuencias prácticas y no permanezca como una declaración abstracta de principios. (11,16)

CONCLUSIONS

La responsabilidad ética en decisiones organizacionales apoyadas por IA exige considerar equidad, dignidad, privacidad, trazabilidad y posibilidad de impugnación. Un sistema con buen desempeño técnico puede seguir siendo inaceptable si distribuye riesgos de forma injusta o no permite identificar responsabilidades. Para la práctica organizacional, esto implica definir responsables del sistema, criterios de supervisión y rutas de apelación antes del despliegue, de modo que la transparencia se traduzca en acciones concretas y no permanezca como una declaración general de principios. La interpretación de estos resultados debe mantenerse vinculada al contexto y a los recursos disponibles.

La transparencia debe diseñarse para usuarios concretos y niveles de riesgo concretos. Explicaciones extensas o técnicas no garantizan comprensión. La evidencia favorece información sobre lógica relevante, desempeño, incertidumbre y límites, combinada con interfaces que permitan verificar y cuestionar recomendaciones. La confianza debe entenderse como una relación calibrada con el sistema, sustentada en información sobre desempeño, incertidumbre y límites de uso, porque tanto la sobredependencia como el rechazo automático pueden deteriorar la calidad de las decisiones. La interpretación de estos resultados debe mantenerse vinculada al contexto y a los recursos disponibles.

La confianza deseable es confianza calibrada, no confianza máxima. Las organizaciones deben evitar tanto la sobredependencia como el rechazo automático de la IA mediante supervisión humana significativa, formación y mecanismos de apelación. La rendición de cuentas requiere responsables identificables antes y después de una decisión. Las explicaciones deben adaptarse al conocimiento del usuario y a la gravedad de las consecuencias, combinando claridad con información suficiente para cuestionar la recomendación cuando existan señales de error, sesgo o incompatibilidad con el contexto.

La gobernanza debe integrar ética, transparencia y confianza en procesos ordinarios de diseño, adquisición, auditoría y evaluación de sistemas. El objetivo no es únicamente cumplir requisitos formales, sino garantizar decisiones comprensibles, revisables y compatibles con los derechos y expectativas de los usuarios. La gobernanza responsable también exige auditoría periódica, trazabilidad de datos y autoridad humana efectiva para detener o modificar procesos automatizados, especialmente cuando las decisiones afectan oportunidades laborales, recursos, derechos o acceso a servicios. La interpretación de estos resultados debe mantenerse vinculada al contexto y a los recursos disponibles.

BIBLIOGRAPHIC REFERENCES

1. Shin D. The effects of explainability and causability on perception, trust, and acceptance: implications for explainable AI. *Int J Hum Comput Stud.* 2021;146:102551. <https://doi.org/10.1016/j.ijhcs.2020.102551>.
2. Zerilli J, Bhatt U, Weller A. How transparency modulates trust in artificial intelligence. *Patterns.* 2022;3(4):100455. <https://doi.org/10.1016/j.patter.2022.100455>.
3. de Bruijn H, Warnier M, Janssen M. The perils and pitfalls of explainable AI: strategies for explaining algorithmic decision-making. *Gov Inf Q.* 2022;39(2):101666. <https://doi.org/10.1016/j.giq.2021.101666>.
4. Banks S, Formosa P, Griep Y, Richards D. AI decision making with dignity? Contrasting workers' justice perceptions of human and AI decision making in a human resource management context. *Inf Syst Front.* 2022;24(3):857-875. <https://doi.org/10.1007/s10796-021-10223-8>.
5. Reinhardt K. Trust and trustworthiness in AI ethics. *AI Ethics.* 2023;3:735-744. <https://doi.org/10.1007/s43681-022-00200-5>.
6. Leichtmann B, Humer C, Hinterreiter A, Streit M, Mara M. Effects of explainable artificial intelligence on trust and human behavior in a high-risk decision task. *Comput Human Behav.* 2023;139:107539. <https://doi.org/10.1016/j.chb.2022.107539>.
7. Naiseh M, Al-Thani D, Jiang N, Ali R. How the different explanation classes impact trust calibration: the case of clinical decision support systems. *Int J Hum Comput Stud.* 2023;169:102941. <https://doi.org/10.1016/j.ijhcs.2022.102941>.
8. Wysocki O, Davies JK, Vigo M, Armstrong AC, Landers D, Lee R, et al. Assessing the communication gap between AI models and healthcare professionals: explainability, utility and trust in AI-driven clinical decision-making. *Artif Intell.* 2023;316:103839. <https://doi.org/10.1016/j.artint.2022.103839>.

9. Ali S, Abuhmed T, El-Sappagh S, Muhammad K, Alonso-Moral JM, Confalonieri R, et al. Explainable artificial intelligence (XAI): what we know and what is left to attain trustworthy artificial intelligence. *Inf Fusion*. 2023;99:101805. <https://doi.org/10.1016/j.inffus.2023.101805>.
10. Papagni G, de Pagter J, Zafari S, Filzmoser M, Koeszegi ST. Artificial agents' explainability to support trust: considerations on timing and context. *AI Soc*. 2023;38:947-960. <https://doi.org/10.1007/s00146-022-01462-7>.
11. Martin K, Waldman A. Are algorithmic decisions legitimate? The effect of process and outcomes on perceptions of legitimacy of AI decisions. *J Bus Ethics*. 2023;183:653-670. <https://doi.org/10.1007/s10551-021-05032-7>.
12. Corrêa NK, Galvão C, Santos JW, Del Pino C, Pinto EP, Barbosa C, et al. Worldwide AI ethics: a review of 200 guidelines and recommendations for AI governance. *Patterns*. 2023;4(10):100857. <https://doi.org/10.1016/j.patter.2023.100857>.
13. Ha T, Kim S. Improving trust in AI with mitigating confirmation bias: effects of explanation type and debiasing strategy for decision-making with explainable AI. *Int J Hum Comput Interact*. 2024;40(24):8562-8573. <https://doi.org/10.1080/10447318.2023.2285640>.
14. Wang P, Ding H. The rationality of explanation or human capacity? Understanding the impact of explainable artificial intelligence on human-AI trust and decision performance. *Inf Process Manag*. 2024;61(4):103732. <https://doi.org/10.1016/j.ipm.2024.103732>.
15. How transparency affects algorithmic advice utilization: the mediating roles of trusting beliefs. *Decis Support Syst*. 2024;183:114273. <https://doi.org/10.1016/j.dss.2024.114273>
16. Breidbach CF. Responsible algorithmic decision-making. *Organ Dyn*. 2024;53(2):101031. <https://doi.org/10.1016/j.orgdyn.2024.101031>.
17. Cheong BC. Transparency and accountability in AI systems: safeguarding wellbeing in the age of algorithmic decision-making. *Front Hum Dyn*. 2024;6:1421273. <https://doi.org/10.3389/fhumd.2024.1421273>.
18. Goktas P. Ethics, transparency, and explainability in generative AI decision-making systems: a comprehensive bibliometric study. *J Decis Syst*. 2024. <https://doi.org/10.1080/12460125.2024.2410042>
19. Transparency and explainability of AI systems: from ethical guidelines to requirements. *Inf Softw Technol*. 2023;159:107197. <https://doi.org/10.1016/j.infsof.2023.107197>
20. Cao G, Duan Y, Edwards JS. AI trustworthiness in managerial decision-making: ethics, transparency, and explainability as key drivers. *J Bus Ethics*. 2026. <https://doi.org/10.1007/s10551-026-06400-x>

ETHICAL CONSIDERATIONS

This literature review did not involve direct participation of human beings or collection of identifiable personal data. Ethical responsibility focused on accurate attribution, transparent synthesis and avoidance of fabrication or selective distortion of the reviewed evidence.

ARTIFICIAL INTELLIGENCE (AI) USE STATEMENT

The authors retain full responsibility for source selection, verification, interpretation and scientific conclusions. Any use of generative artificial intelligence is limited to linguistic support and document formatting under human supervision.

FUNDING:

This research received no specific grant or financial support.

CONFLICT OF INTEREST:

The authors declare no conflicts of interest.

AUTHOR CONTRIBUTION

Author Names and Initials: Giovanni Juan Maldonado Aguirre (GJMA)

1. Conceptualization: (GJMA)
2. Data Curation: (GJMA)
3. Formal Analysis: (GJMA)
4. Acquisition of Funds: ()
5. Research: (GJMA)
6. Methodology: (GJMA)
7. Project Management: (GJMA)
8. Resources: ()
9. Software: ()
10. Supervision: (GJMA)
11. Validation: (GJMA)
12. Visualization: (GJMA)
13. Writing – Original Draft: (GJMA)
14. Writing – Proofreading and Editing: (GJMA)