ISSN: 2665-0398

ÉTICA, REGULACIÓN Y SUPERVISIÓN HUMANA EN LA IMPLEMENTACIÓN DE IA PARA COMPLIANCE: EVIDENCIA DESDE UNA REVISIÓN SISTEMÁTICA

ETHICS, REGULATION AND HUMAN OVERSIGHT IN THE IMPLEMENTATION OF AI FOR COMPLIANCE: EVIDENCE FROM A SYSTEMATIC REVIEW

Tipo de Publicación: Articulo Científico
Recibido: 22/01/2026
Aceptado: 22/02/2026
Publicado: 29/03/2026
Código Único AV: e667
Páginas: 1(519-537)
DOI: https://doi.org/10.5281/zenodo.19323292

Autores:
José Antonio Beraún - Barrantes
Abogado
Doctor en Derecho
https://orcid.org/0000-0001-8979-2734
E-mail: jose.beraun@udh.edu.pe
Afiliación: Universidad de Huánuco
País: República del Perú

Fernando Eduardo Corcino Barrueta
Abogado
Doctor en Derecho Penal y Procesal
https://orcid.org/0000-0003-0296-4033
E-mail: fernando.corcino@udh.edu.pe
Afiliación: Universidad de Huánuco
País: República del Perú

Resumen

La incorporación de sistemas de inteligencia artificial en los programas de compliance anticorrupción ha transformado de manera significativa los mecanismos de detección, prevención y control de riesgos, al introducir modelos híbridos que combinan automatización algorítmica y supervisión humana. Esta transformación resulta especialmente relevante en contextos normativos altamente regulados, donde las decisiones asistidas por IA pueden afectar la rendición de cuentas, la protección de derechos fundamentales y la legitimidad institucional. En este marco, el objetivo del presente artículo fue identificar los desafíos éticos y regulatorios que emergen de la implementación de sistemas híbridos humano-IA en programas de compliance, a partir de la evidencia científica disponible. Para ello, se desarrolló un artículo de revisión sistemática, siguiendo las directrices PRISMA 2020, mediante la identificación, selección y análisis crítico de estudios publicados en revistas científicas con revisión por pares entre los años 2021 y 2026. Los resultados evidenciaron la presencia de desafíos recurrentes, tales como el sesgo algorítmico, la opacidad decisional, la dilución de responsabilidades, la fragmentación normativa y las limitaciones en la efectividad de la supervisión humana. Asimismo, se identificaron propuestas de gobernanza y estrategias de mitigación orientadas a la explicabilidad, la auditabilidad y la integración socio-técnica de los sistemas. En conclusión, la revisión pone de relieve que la eficacia de los sistemas híbridos humano-IA en programas de compliance depende no solo de su diseño técnico, sino de la existencia de marcos éticos y regulatorios operativos que articulen de manera coherente la automatización con el control humano significativo.

Palabras Clave Sistemas híbridos humano-IA, desafíos éticos, gobernanza regulatoria, supervisión humana, compliance anticorrupción

Abstract

The incorporation of artificial intelligence systems into anti-corruption compliance programs has significantly transformed risk detection, prevention, and control mechanisms by introducing hybrid models that combine algorithmic automation and human oversight. This transformation is especially relevant in highly regulated legal contexts, where AI-assisted decisions can affect accountability, the protection of fundamental rights, and institutional legitimacy. Within this framework, the objective of this article was to identify the ethical and regulatory challenges that arise from the implementation of hybrid human-AI systems in compliance programs, based on the available scientific evidence. To this end, a systematic review article was developed, following the PRISMA 2020 guidelines, by identifying, selecting, and critically analyzing studies published in peer-reviewed scientific journals between 2021 and 2026. The results revealed the presence of recurring challenges, such as algorithmic bias, decision opacity, dilution of accountability, regulatory fragmentation, and limitations in the effectiveness of human oversight. Furthermore, governance proposals and mitigation strategies focused on the explainability, auditability, and socio-technical integration of the systems were identified. In conclusion, the review highlights that the effectiveness of human-AI hybrid systems in compliance programs depends not only on their technical design but also on the existence of operational ethical and regulatory frameworks that coherently link automation with meaningful human control.

Keywords Human-AI hybrid systems, ethical challenges, regulatory governance, human oversight, anti-corruption compliance

Esta obra está bajo una Licencia Creative Commons Atribución No Comercial-Sin Derivar 4.0 Internacional

ISSN: 2665-0398

Introducción

La incorporación de la inteligencia artificial (IA) en los programas de compliance anticorrupción fue entendida como una transformación relevante en la gestión del cumplimiento normativo corporativo. En este contexto, los sistemas híbridos humano-IA, caracterizados por la combinación de capacidades algorítmicas y supervisión humana, fueron descritos como una alternativa funcional para la detección y prevención de conductas corruptas, en la medida en que permiten complementar la eficiencia computacional con el juicio humano informado, tal como señaló Sagar (2025). No obstante, esta convergencia tecnológica también fue asociada a desafíos éticos y regulatorios de considerable complejidad, los cuales demandaron un abordaje analítico sistemático.

Desde la literatura reciente se identificaron tensiones estructurales entre los procesos de automatización algorítmica y la atribución de responsabilidad humana en escenarios de cumplimiento normativo. Al respecto, Celsi & Zomaya (2025) advirtieron la presencia de preocupaciones vinculadas a sesgos algorítmicos, transparencia, rendición de cuentas y a la relación problemática entre automatización y supervisión humana. Dichas tensiones adquirieron mayor relevancia en entornos altamente regulados, donde las decisiones mediadas por sistemas automatizados pueden incidir sobre derechos fundamentales y obligaciones legales.

En relación con los marcos de gobernanza, diversos estudios evidenciaron limitaciones estructurales para afrontar la complejidad de los sistemas híbridos. Chintoh et al., (2024) señalaron que, en el contexto estadounidense, persistió la ausencia de un marco integral de gobernanza para la IA, particularmente en materia de transparencia, equidad y compatibilidad con la normativa de protección de datos personales. Esta insuficiencia regulatoria fue asociada a escenarios de incertidumbre respecto de la delimitación de responsabilidades entre agentes humanos y componentes algorítmicos.

De manera complementaria, la operacionalización de principios éticos en sistemas híbridos fue identificada como un desafío adicional. En este sentido, Pasupuleti (2025) propuso cuatro pilares para una gobernanza ética de la IA: alineación legal y política, principios de diseño ético, auditabilidad técnica y participación de múltiples actores, destacando que los modelos institucionales que priorizaron la transparencia y la rendición de cuentas tendieron a fortalecer la confianza pública.

La producción científica reciente abordó de forma progresiva los desafíos derivados de la implementación de sistemas híbridos humano-IA en contextos de cumplimiento normativo, aportando

Revista Aula Virtual, ISSN: 2665-0398; Periodicidad: Continua
Volumen: 7, Número: 14, Año: 2026 (Enero 2026 - Junio 2026)
Esta obra está bajo una Licencia Creative Commons Atribución No Comercial-Sin Derivar 4.0 Internacional

evidencia relevante para la comprensión de esta problemática multidimensional. En particular, Birkstedt et al., (2023), a partir de una revisión sistemática, identificaron vacíos críticos en la gobernanza de la IA y propusieron cuatro agendas de investigación futura: técnica, de actores y contexto, regulatoria y de procesos. Sus hallazgos mostraron que la aplicación efectiva de principios éticos en organizaciones permaneció insuficientemente desarrollada, especialmente en lo relativo a la asignación de responsabilidades entre componentes humanos y algorítmicos.

De manera similar, Giarmoleo et al., (2024) realizaron una revisión sistemática de 309 estudios centrados en preocupaciones éticas de la IA, concluyendo que muchas de las soluciones planteadas carecieron de mecanismos operativos concretos. Esta brecha entre principios éticos abstractos y su implementación práctica fue considerada especialmente relevante para los sistemas de compliance, donde la supervisión humana de decisiones algorítmicas constituye un elemento crítico.

En el ámbito financiero, Ridzuan et al., (2024) examinaron aplicaciones de IA vinculadas al cumplimiento normativo, identificando desafíos relacionados con transparencia, rendición de cuentas y gobernanza. Sus resultados sugirieron que la integración de IA en funciones de compliance exigió un equilibrio entre innovación tecnológica y responsabilidad ética, lo que reforzó la necesidad de marcos regulatorios adaptativos orientados a la colaboración humano-máquina.

Por su parte, Akinrinola et al., (2024) propusieron estrategias orientadas a mitigar dilemas éticos en el desarrollo de IA, entre ellas la adopción de modelos de IA explicable, esquemas de supervisión humana en el bucle y mecanismos de gobernanza ética. Dichas propuestas ofrecieron un marco analítico pertinente para comprender cómo los principios de transparencia, equidad y rendición de cuentas pueden ser operacionalizados en sistemas híbridos aplicados al compliance anticorrupción.

A pesar de estos avances, la literatura evidenció vacíos relevantes que justificaron la realización de una revisión sistemática focalizada en sistemas híbridos humano-IA en programas de cumplimiento normativo anticorrupción. Un primer vacío se vinculó con la ausencia de marcos específicos para la asignación de responsabilidades en sistemas híbridos. Manan et al., (2025) observaron que las tecnologías de vigilancia basadas en IA plantearon dilemas éticos asociados a sesgos algorítmicos y afectaciones a la privacidad, sin que se definieran con claridad los mecanismos de distribución de responsabilidades en contextos normativos.

Un segundo vacío estuvo relacionado con la limitada integración de estándares de explicabilidad

Esta obra está bajo una Licencia Creative Commons Atribución No Comercial-Sin Derivar 4.0 Internacional

ISSN: 2665-0398

en entornos regulados. Bhardwaj (2025), en su revisión sistemática sobre IA explicable en el sector financiero, identificó problemas persistentes de aceptación institucional derivados de la ausencia de estándares consolidados, la dependencia de métodos post hoc y la escasa incorporación de teorías conductuales sobre confianza, particularmente en procesos de auditoría regulatoria.

El tercer vacío se asoció a la fragmentación del conocimiento sobre gobernanza ética de la IA en el sector público. Adepoju & Chinonyerem (2025) documentaron experiencias de implementación de IA en supervisión gubernamental, incluyendo algoritmos de detección de fraude y técnicas de procesamiento del lenguaje natural, señalando la necesidad de marcos integrales que articulen de manera coherente la automatización con la supervisión humana.

Finalmente, Patel (2025) destacó la relevancia de desarrollar investigaciones comparativas sobre modelos de IA explicable aplicados a la evaluación de riesgos financieros, subrayando las limitaciones existentes para su adopción en entornos altamente regulados y evidenciando la escasa atención específica al cumplimiento normativo en materia anticorrupción.

A partir de los vacíos identificados, el presente artículo de revisión sistemática tuvo como objetivo identificar y analizar los desafíos éticos y regulatorios asociados a la implementación de sistemas híbridos humano-IA en programas de cumplimiento, con base en la evidencia científica disponible.

Metodología

Para el presente trabajo se empleó la metodología PRISMA 2020 (Preferred Reporting Items for Systematic Reviews and Meta-Analyses). La estrategia de búsqueda en la base de datos Scopus se estructuró mediante la siguiente fórmula booleana: *(("artificial intelligence" OR "AI" OR "machine learning" OR "algorithmic") AND ("human-in-the-loop" OR "human oversight" OR "hybrid governance" OR "human-AI collaboration") AND ("compliance" OR "regulatory" OR "anti-corruption" OR "ethics" OR "governance") AND ("challenges" OR "risks" OR "accountability" OR "transparency" OR "bias"))*

Para guiar la revisión sistemática, se formularon las siguientes preguntas: a) ¿Cuáles son los principales desafíos éticos documentados en la literatura científica respecto a la implementación de sistemas híbridos humano-IA en programas de compliance anticorrupción?, b) ¿Qué marcos regulatorios y de gobernanza han sido propuestos para abordar la integración de supervisión humana con automatización algorítmica en contextos de cumplimiento normativo? c) ¿Qué estrategias y mecanismos de mitigación se han identificado para superar los desafíos éticos y regulatorios en sistemas híbridos humano-IA aplicados al compliance?

Revista Aula Virtual, ISSN: 2665-0398; Periodicidad: Continua
Volumen: 7, Número: 14, Año: 2026 (Enero 2026 - Junio 2026)
Esta obra está bajo una Licencia Creative Commons Atribución No Comercial-Sin Derivar 4.0 Internacional

Las búsquedas bibliográficas fueron realizadas en las bases de datos Scopus, siguiendo enfoques metodológicos previamente empleados en revisiones sistemáticas orientadas al análisis de la gobernanza de la inteligencia artificial, conforme a lo documentado en estudios previos. Para tal fin, se emplearon palabras clave seleccionadas en idioma inglés, entre las que se incluyeron *artificial intelligence, human-in-the-loop, hybrid governance, compliance, anti-corruption, ethical challenges, regulatory frameworks, algorithmic accountability, transparency y algorithmic bias*.

En relación con los criterios de inclusión, se consideraron elegibles aquellos artículos publicados entre los años 2021 y 2026, difundidos en revistas científicas con revisión por pares, que abordaran de manera explícita los desafíos éticos o regulatorios asociados al uso de la inteligencia artificial en contextos de compliance o gobernanza. Asimismo, se incluyeron investigaciones que examinaran la interacción humano-IA en sistemas de cumplimiento normativo, siempre que los textos se encontraran disponibles en idioma inglés o español.

De forma complementaria, se establecieron criterios de exclusión orientados a delimitar el corpus analítico. En este sentido, fueron excluidos artículos de opinión, editoriales y cartas al editor carentes de sustento empírico, así como estudios que abordaran aplicaciones de la inteligencia artificial sin vinculación con el compliance, la gobernanza o los marcos regulatorios.

También se descartaron publicaciones duplicadas o versiones preliminares de investigaciones ya consideradas, trabajos centrados exclusivamente en aspectos técnicos sin análisis éticos o regulatorios, documentos sin acceso al texto completo y literatura gris, incluyendo informes técnicos no sometidos a revisión por pares y documentos de trabajo.

La aplicación sistemática de estos criterios permitió delimitar un conjunto de estudios caracterizado por su coherencia temática y rigor académico, lo que facilitó una síntesis analítica de la evidencia disponible sobre los desafíos éticos y regulatorios asociados a los sistemas híbridos humano-IA en programas de compliance anticorrupción (Ver Figura 1).

Esta obra está bajo una Licencia Creative Commons Atribución No Comercial-Sin Derivar 4.0 Internacional

ISSN: 2665-0398

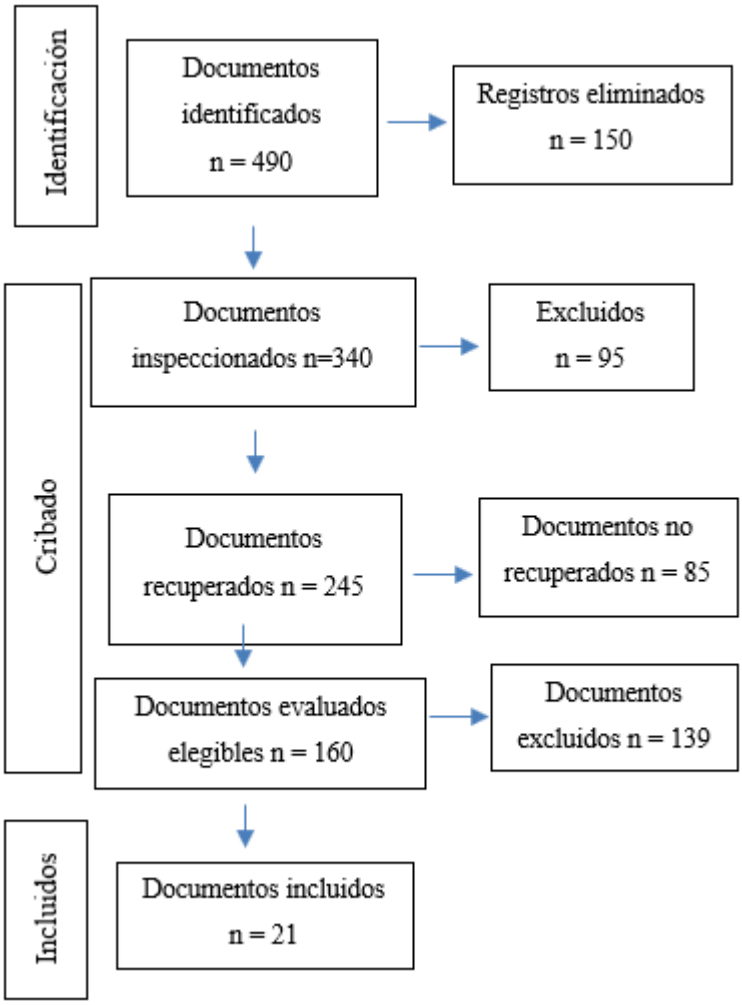

Figura 1. Identificación de estudios que utilizan el método prismático

Revista Aula Virtual, ISSN: 2665-0398; Periodicidad: Continua
Volumen: 7, Número: 14, Año: 2026 (Enero 2026 - Junio 2026)
Esta obra está bajo una Licencia Creative Commons Atribución No Comercial-Sin Derivar 4.0 Internacional

ISSN: 2665-0398

Resultados

Autor	Contexto / sector	Configuración humana–IA	Principales desafíos éticos	Principales desafíos regulatorios
Albaroudi et al., (2026)	Toma de decisiones organizacionales (público–privado)	Human-in-the-loop con mitigación de sesgos	Sesgo algorítmico; opacidad decisional; sobreconfianza en salidas de IA	Ausencia de auditorías obligatorias; falta de exigencias legales de transparencia
Zaidan et al., (2026)	Gobernanza global de IA	Sistemas híbridos bajo supervisión institucional	Dilución de la rendición de cuentas; ética declarativa no operativa	Fragmentación regulatoria; ausencia de estándares vinculantes
Mozzarelli & Schneider (2026)	Gobernanza corporativa (UE)	IA en decisiones estratégicas con control humano formal	Control humano ilusorio; automatización indebida de discrecionalidad	Insuficiencia del AI Act para procesos decisionales híbridos
Floridi & Ascani (2026)	Instituciones públicas (Italia)	Sistemas deliberativos con human-on-the-loop	Dependencia cognitiva; *deskilling* institucional; desigualdad digital	Falta de estándares éticos operativos; necesidad de gobernanza anticipatoria
Ji et al., (2025)	Sistemas de IA de alto impacto (global)	IA avanzada con supervisión humana limitada	Pérdida de control humano efectivo; desalineación ética; riesgos sistémicos	Gobernanza fragmentada; debilidad regulatoria frente a sistemas complejos
Saeed & Prybutok (2026)	Organizaciones privadas	Delegación de tareas a IA con autonomía variable	Transferencia indebida de responsabilidad; conflicto utilidad–ética	Vacíos normativos sobre responsabilidad por delegación
Lança & Rocha (2026)	Gestión editorial jurídica (Brasil)	IA editorial con supervisión humana directa	Opacidad algorítmica; automatización acrítica; afectación de autonomía intelectual	Ausencia de directrices normativas específicas para IA editorial
Frid et al., (2025)	Institución sanitaria pública (UE)	GenAI con human-in-the-loop obligatorio	Sesgos; alucinaciones; tensión eficiencia–seguridad	Exigencias de GDPR y EU AI Act; DPIA y trazabilidad
Bui (2025)	Gobernanza de plataformas digitales	Moderación legalmente informada con HITL y apelación	Dilución de accountability; externalidades por falsos positivos	Falta de debido proceso algorítmico exigible; proporcionalidad

Tabla 1. Desafíos éticos y regulatorios de los sistemas híbridos humano–IA en contextos de compliance

Esta obra está bajo una Licencia Creative Commons Atribución No Comercial-Sin Derivar 4.0 Internacional

ISSN: 2665-0398

Autor	Contexto / sector	Marco regulatorio o de gobernanza identificado	Naturaleza del marco	Modelo de supervisión humana	Mecanismos clave de integración humano–IA
Floridi & Ascani (2026)	Sector público / gobernanza de IA	Gobernanza anticipatoria y ética institucional de IA	Soft law ético-institucional	Supervisión humana distribuida	Evaluación de impacto, accountability ex ante, gobernanza anticipatoria
Kortukova (2024)	Unión Europea / derechos fundamentales	Regulación basada en riesgos y derechos (EU AI Act)	Hard law	Supervisión humana graduada por nivel de riesgo	Clasificación de riesgos, controles ex ante y ex post
Zaidan et al., (2026)	Gobernanza tecnológica global	Modelo híbrido ética–regulación para IA	Marco híbrido (soft → hard)	Supervisión humana normativa	Traducción de principios éticos en reglas organizacionales
Saup et al., (2026)	Toma de decisiones estratégicas	Modelo socio-técnico de gobernanza con "admission gates"	Marco organizacional aplicado	Human-in-the-loop estructural	Puertas de admisión: calidad, procedencia, explicabilidad, responsabilidad
Mozzarelli & Schneider (2026)	Gobernanza corporativa	Gobernanza corporativa de decisiones asistidas por IA	Hard law interpretado + gobierno corporativo	Supervisión humana jerárquica	Comités, control organizacional, asignación de responsabilidad
Wang & Tobias (2025)	Finanzas / sistemas de decisión	Gobernanza de IA basada en Human-in-the-Loop	Marco técnico-regulatorio	Human-in-the-loop operativo	Auditoría, monitoreo continuo, control de deriva
Kioskli et al., (2025)	Infraestructuras críticas / confianza	Modelo de madurez de supervisión humana (TrustSense)	Marco técnico de medición	Human-on-the-loop graduado	Indicadores de madurez, evaluación periódica
Frid et al., (2025)	IA generativa / protección de datos	Gobernanza híbrida GDPR–AI Act para GenAI	Hard law + soft law	Supervisión humana con escalamiento	DPIA, control humano reforzado, explicabilidad
Alotaibi (2025)	Auditoría y finanzas corporativas	Corporate Finance and Regulatory Governance Framework (CFRGF)	Marco regulatorio-organizacional	Human-in-the-loop institucional	Asignación de roles, reporting, controles financieros
Park et at., (2025)	Sistemas autónomos / regulación sectorial	Preparación regulatoria basada en competencias humanas	Marco regulatorio-competencial	Supervisión humana basada en capacidades	Capacitación, reglas claras de intervención
Albaroudi et al., (2026)	Reclutamiento automatizado	Marco de gobernanza ética, equitativa y sostenible	Marco aplicado de gobernanza	Human-in-the-loop con métricas	Explicabilidad, métricas de equidad, control humano
Bui (2025)	Plataformas digitales / cumplimiento legal	Marco legalmente consciente con cadena ley-política-acción	Hard law operacionalizado	Human-in-the-loop obligatorio	Evidencia mínima, acciones graduadas, apelación
Goffin (2025)	IA de alto riesgo / salud	Gobernanza institucional de la supervisión humana (AI Act)	Hard law + gobernanza institucional	Human-in-command	Comités multidisciplinarios, responsabilidad clara

Tabla 2. Gobernanza y supervisión humana en sistemas humano-IA de compliance

Revista Aula Virtual, ISSN: 2665-0398; Periodicidad: Continua
Volumen: 7, Número: 14, Año: 2026 (Enero 2026 - Junio 2026)
Esta obra está bajo una Licencia Creative Commons Atribución No Comercial-Sin Derivar 4.0 Internacional

Autor	Dominio / contexto	Desafío ético/regulatorio central	Estrategia / mecanismo de mitigación (tipología)	Operacionalización (cómo se implementó)	Evidencia / resultados reportados
Albaroudi et al., (2026)	Reclutamiento / gobernanza IA	Sesgo algorítmico, baja transparencia; ausencia de métricas ambientales	Gobernanza técnica dual (equidad + sostenibilidad)	Debiasing; métricas de equidad (p. ej., SPD/DI); XAI (SHAP); medición de emisiones (CodeCarbon)	Mejoras en equidad; reducción sustantiva de emisiones; aumento de confianza percibida
Moreno-Sánchez (2026)	Salud (marco replicable)	Principios éticos/regulatorios poco operativizados	Trustworthy-by-design (traducción normativa a requisitos)	Conversión de principios en requisitos técnicos, métricas y roles de supervisión humana	Coherencia diseño-ética-regulación; tensiones residuales entre principios
Curcin (2026)	Sistemas complejos regulados (LHS)	Deriva del modelo, opacidad, vacíos de responsabilidad	Gobernanza dinámica (ciclo continuo de aprendizaje/auditoría)	Monitoreo continuo; retroalimentación humana; recalibración; trazabilidad y métricas	Se argumentó mitigación de sesgos/errores en el tiempo; dependencia de madurez institucional
Floridi & Ascani (2026)	Parlamento / gobernanza pública	Riesgo de automatización excesiva; pérdida de deliberación; legitimidad	Gobernanza anticipatoria (control humano reforzado + diseño pro-ético)	Supervisión humana directa; transparencia; privacidad; etiquetado/controles ex ante alineados a AI Act	Se reportó refuerzo de legitimidad; persistencia de riesgo de deskilling
Mozzarelli & Schneider (2026)	Decisión estratégica corporativa	Incompletitud regulatoria para decisiones discrecionales; responsabilidad	Arquitectura socio-técnica (IA como soporte, humano como decisor responsable)	Redistribución funcional: IA para escenarios/analítica; humano conserva juicio y responsabilidad	Se sostuvo reducción de automatización acrítica; requerimiento de madurez organizacional
Galan-Cubillo & Saez-Soro (2025)	Gestión académica (modelo transferible)	Trazabilidad insuficiente; fiabilidad de outputs	Rol humano especializado (AI editor / validador experto)	Validación humana continua; control de fuentes; detección/corrección de sesgos y errores	Mejora de calidad y confianza; costo en capacidades humanas
Alotaibi (2025)	Gobernanza financiera / reporting	Baja auditabilidad y trazabilidad; riesgo de no cumplimiento	Gobernanza por pilares (controles + XAI + segregación + evidencia)	Controles continuos; explicabilidad; segregación de funciones; evidencia para auditoría	Se reportó aumento de transparencia y cumplimiento; complejidad de implementación
Saeed & Prybutok (2026)	Sistemas agentivos organizacionales	Trade-off utilidad/ética; aceptación y legitimidad	Delegación ética mixta (valores + supervisión humana)	Integración explícita de fairness/transparencia; control humano; evaluación por stakeholders	Mayor aceptación/legitimidad; persistencia de trade-offs

Esta obra está bajo una Licencia Creative Commons Atribución No Comercial-Sin Derivar 4.0 Internacional

ISSN: 2665-0398

Autor	Dominio / contexto	Desafío ético/regulatorio central	Estrategia / mecanismo de mitigación (tipología)	Operacionalización (cómo se implementó)	Evidencia / resultados reportados
Bergomi (2025)	Salud (DSS clínico)	Sesgo de automatización; sobreconfianza en explicaciones	XAI orientada a acción + capacitación	Comparación de explicadores (p. ej., SHAP); evaluación de entendibilidad/accionabilidad; entrenamiento	Mayor comprensión con explicaciones preferidas; automatización no desapareció
Saup et al. (2026)	Decisión estratégica con GenAI	Falta de "admisibilidad" organizacional de outputs	Gates de gobernanza (criterios para uso formal)	Controles de procedencia, responsabilidad, explicabilidad y validación humana antes de adopción	Facilitó paso de pilotos a decisiones; dependencia de alineación interna
Bui (2025)	Moderación legalmente consciente (Vietnam)	Desalineación entre etiquetas ML y categorías legales; debido proceso	Pipeline legal-auditable (ley→política→etiqueta→evidencia→acción→apelación)	Evidencia mínima; acciones graduadas; racionales obligatorios; HITL; apelación y SLA	Acciones graduadas superaron remoción binaria en frontera daño-costo; menor tasa de revocación
Yan (2025)	Seguridad laboral (minería)	Opacidad de modelos en contextos de riesgo	AutoML interpretable + HITL	AutoGluon + SHAP/LIME; revisión humana; chequeos éticos/bias	Alta precisión con explicaciones; datos sintéticos
Park et al., (2025)	Navegación autónoma (MASS)	Vacíos normativos: rol humano, responsabilidad y competencia	Regulación basada en escenarios + reconocimiento del rol humano	STPA (escenarios de pérdida) + Delphi (priorización); paquetes de acción; inclusión de operador remoto en normas	Consenso experto alto; top-risks y bundles de acción
Kioskli et al., (2025)	Gobernanza organizacional (trustSense)	Dificultad para medir "supervisión humana significativa" (AI Act)	Evaluación de madurez de oversight (diagnóstico-mejora)	Cuestionarios por rol; scoring de madurez; feedback guiado; diseño privacy-preserving	Validación piloto; utilidad para identificar brechas
Frid et al., (2025)	Hackathon hospitalario GenAI	Sesgo, alucinaciones, privacidad y cumplimiento (GDPR/AI Act)	Gobernanza experimental controlada (pilotos responsables)	Datos anonimizados; entorno seguro; HITL; evaluación; mitigación de sesgos (muestreo, prompt tuning, disclaimers)	Prototipos con desempeño variable; enfoque fuerte en cumplimiento y evaluación
Wang &Tobias (2025)	Finanzas digitales (AML)	Trade-offs costo-riesgo-cumplimiento; coordinación humano-IA	Colaboración adaptativa + optimización multi-objetivo	Arquitectura en capas; RL; Pareto frontier; puntos de revisión/override humano	+256% eficiencia AML; AUC-ROC 0.923; reducción de tiempos manteniendo cumplimiento

Tabla 3. Estrategias de mitigación ético-regulatorias en sistemas híbridos humano-IA

Revista Aula Virtual, ISSN: 2665-0398; Periodicidad: Continua
Volumen: 7, Número: 14, Año: 2026 (Enero 2026 - Junio 2026)
Esta obra está bajo una Licencia Creative Commons Atribución No Comercial-Sin Derivar 4.0 Internacional

Discusión de resultados

El presente artículo de revisión sistemática tuvo como objetivo identificar y analizar los desafíos éticos y regulatorios que emergen de la implementación de sistemas híbridos humano–IA en programas de compliance, a partir de la evidencia científica disponible. Los resultados obtenidos permiten una comprensión integrada de esta problemática, al evidenciar patrones recurrentes en la literatura, así como divergencias relevantes entre enfoques normativos, organizacionales y tecnológicos.

Un primer resultado central del estudio fue la identificación de desafíos éticos estructurales, entre los que destacaron el sesgo algorítmico, la opacidad decisional, la dilución de la rendición de cuentas y la sobreconfianza en las salidas de los sistemas automatizados. Estos hallazgos convergen con lo reportado por Birkstedt et al., (2023), quienes, a partir de una revisión sistemática, señalaron que la falta de mecanismos claros de asignación de responsabilidad constituye uno de los vacíos más críticos en la gobernanza de la inteligencia artificial. De manera similar, Giarmoleo et al., (2024) identificaron que, si bien la literatura reconoce ampliamente estos desafíos, las propuestas de solución tienden a permanecer en un nivel declarativo, sin traducirse en prácticas operativas consistentes.

En relación con la tensión entre automatización algorítmica y supervisión humana, los resultados del presente estudio mostraron que los modelos de supervisión humana frecuentemente adoptados en contextos de compliance como human-in-the-loop o human-on-the-loop no garantizan necesariamente un control humano efectivo. Esta conclusión es consistente con los planteamientos de Mozzarelli & Schneider (2026), quienes advirtieron que la supervisión humana puede convertirse en un mecanismo meramente formal o ilusorio cuando no se acompaña de capacidades institucionales reales y de una delimitación normativa precisa de responsabilidades. En contraste, algunos estudios revisados, como el de Floridi & Ascani (2026), sostuvieron que la adopción de enfoques de gobernanza anticipatoria puede fortalecer la agencia humana, siempre que se integren evaluaciones de impacto y mecanismos ex ante de rendición de cuentas.

Desde la dimensión regulatoria, los resultados evidenciaron una fragmentación normativa significativa, particularmente en jurisdicciones donde coexisten marcos de *soft law* y *hard law* sin una articulación clara. Este hallazgo converge con lo señalado por Kortukova (2024), quien analizó el enfoque basado en riesgos del Reglamento Europeo de IA (EU AI Act) y destacó sus avances en materia de derechos fundamentales, aunque también sus limitaciones para abordar la complejidad de los

Esta obra está bajo una Licencia Creative Commons Atribución No Comercial-Sin Derivar 4.0 Internacional

ISSN: 2665-0398

sistemas híbridos. No obstante, algunos estudios incluidos en la revisión, como el de Zaidan et al., (2026), sugirieron que los modelos híbridos de gobernanza, que traducen principios éticos en reglas organizacionales vinculantes, podrían ofrecer una vía intermedia para reducir dicha fragmentación.

Asimismo, el análisis comparativo mostró que, en contextos de *compliance* anticorrupción y financiero, la exigencia de explicabilidad y auditabilidad se presenta como un desafío transversal. Este resultado coincide con la revisión sistemática de Bhardwaj (2025), quien identificó que la ausencia de estándares consolidados de IA explicable limita la aceptación institucional de estos sistemas en procesos regulatorios. Sin embargo, algunos estudios empíricos revisados, como el de Albaroudi et al., (2026), reportaron avances en la operacionalización de métricas de equidad y explicabilidad, lo que sugiere que las divergencias observadas en la literatura podrían explicarse por diferencias en el grado de madurez organizacional y en los contextos sectoriales analizados.

Finalmente, en lo relativo a las estrategias de mitigación ético-regulatorias, los resultados del presente estudio mostraron una convergencia creciente hacia enfoques sociotécnicos, que conciben a la IA como un sistema integrado de componentes técnicos, humanos y normativos. Este enfoque es coherente con lo planteado por Saeed & Prybutok (2026), quienes destacaron la necesidad de modelos de delegación ética mixta que equilibren utilidad, legitimidad y supervisión humana. No obstante, la evidencia también reveló que dichas estrategias suelen depender fuertemente de capacidades institucionales avanzadas, lo que limita su transferibilidad a organizaciones con menores recursos.

A pesar del rigor metodológico aplicado, este estudio presentó ciertas limitaciones que deben ser consideradas al interpretar los resultados. En primer lugar, la revisión se circunscribió a artículos publicados en revistas con revisión por pares y disponibles en idioma inglés o español, lo que pudo excluir literatura relevante en otros idiomas o informes técnicos de organismos reguladores que no cumplen con estos criterios. En segundo lugar, la heterogeneidad de los estudios incluidos, tanto en términos de contextos sectoriales como de enfoques metodológicos, dificultó la realización de comparaciones completamente homogéneas entre los hallazgos.

Asimismo, la inclusión de publicaciones recientes y de artículos en modalidad de publicación anticipada pudo introducir variaciones en la solidez empírica de algunos resultados, dado que no todos los estudios contaban con validaciones longitudinales. Finalmente, al tratarse de una revisión sistemática, los hallazgos dependen de la calidad y el alcance de la evidencia disponible, lo

Revista Aula Virtual, ISSN: 2665-0398; Periodicidad: Continua
Volumen: 7, Número: 14, Año: 2026 (Enero 2026 - Junio 2026)
Esta obra está bajo una Licencia Creative Commons Atribución No Comercial-Sin Derivar 4.0 Internacional

que implica que ciertos desafíos emergentes podrían encontrarse subrepresentados en la literatura actual.

A partir de los resultados obtenidos y de las limitaciones identificadas, se sugieren varias líneas para futuras investigaciones. En primer lugar, resulta necesario desarrollar estudios empíricos comparativos que analicen la implementación de sistemas híbridos humano–IA en programas de compliance anticorrupción en diferentes jurisdicciones, con el fin de evaluar el impacto real de los marcos regulatorios existentes.

En segundo lugar, futuras investigaciones podrían profundizar en la operacionalización de la supervisión humana significativa, explorando métricas, indicadores y modelos organizacionales que permitan evaluar su efectividad más allá de su formulación normativa.

Adicionalmente, se recomienda avanzar en estudios que integren enfoques interdisciplinarios, combinando derecho, ética, ingeniería y ciencias organizacionales, para abordar de manera holística los desafíos identificados. Finalmente, sería pertinente explorar el rol de la cultura organizacional y de las capacidades institucionales en la adopción de estrategias de mitigación ético-regulatorias, particularmente en contextos de economías emergentes, donde la evidencia disponible sigue siendo limitada.

Conclusiones

Los resultados de esta revisión sistemática evidenciaron que la implementación de sistemas híbridos humano–IA en programas de compliance se encuentra atravesada por un conjunto de desafíos éticos y regulatorios recurrentes y estructurales. Entre los hallazgos más relevantes se identificaron el sesgo algorítmico, la opacidad en los procesos decisionales automatizados, la dilución de la rendición de cuentas y las tensiones persistentes entre eficiencia tecnológica y control humano efectivo.

Asimismo, la evidencia analizada mostró que los marcos normativos y de gobernanza existentes resultan, en muchos casos, fragmentados o insuficientes para abordar la complejidad de estos sistemas híbridos, lo que limita su aplicación coherente en contextos de cumplimiento normativo, particularmente en materia anticorrupción. Estos hallazgos contribuyen al campo de estudio al ofrecer una síntesis integrada y comparativa de la literatura reciente, permitiendo comprender de manera sistemática los riesgos y vacíos que acompañan la adopción de inteligencia artificial en funciones críticas de compliance.

En relación con el objetivo de investigación, el estudio permitió identificar de manera clara y fundamentada los principales desafíos éticos y regulatorios que emergen de la implementación de sistemas híbridos humano–IA en programas de

Esta obra está bajo una Licencia Creative Commons Atribución No Comercial-Sin Derivar 4.0 Internacional

ISSN: 2665-0398

compliance, según la evidencia científica disponible. La revisión mostró que dichos desafíos no se limitan a aspectos técnicos de diseño algorítmico, sino que se extienden a dimensiones organizacionales, jurídicas e institucionales, como la asignación de responsabilidades, la exigencia de explicabilidad, la protección de derechos fundamentales y la necesidad de mecanismos efectivos de supervisión humana.

En este sentido, los resultados confirman que la eficacia de los sistemas híbridos en contextos de compliance depende tanto de su arquitectura técnica como de la existencia de marcos regulatorios claros y de capacidades institucionales adecuadas para su implementación y control.

El presente trabajo se desarrolló como un artículo de revisión sistemática, siguiendo las directrices establecidas por PRISMA 2020, lo que permitió garantizar un proceso metodológico riguroso, transparente y replicable en la identificación, selección y análisis de la literatura científica.

Esta aproximación metodológica resultó particularmente adecuada para abordar un campo de estudio caracterizado por la diversidad de enfoques teóricos, normativos y sectoriales, facilitando la integración de evidencias provenientes de distintas disciplinas y contextos. La naturaleza sistemática del estudio refuerza la validez de las conclusiones alcanzadas, en la medida en que estas se sustentan en patrones consistentes identificados a lo largo del corpus analizado.

Finalmente, desde una perspectiva más amplia, los resultados de esta investigación ponen de manifiesto la necesidad de avanzar hacia modelos de gobernanza ético-regulatoria más integrales y operativos, que permitan articular de manera efectiva la automatización algorítmica con la supervisión humana en programas de compliance.

En este sentido, futuras investigaciones podrían profundizar en el análisis empírico de casos de implementación concreta de sistemas híbridos humano–IA, evaluar comparativamente el impacto de distintos marcos regulatorios y explorar mecanismos innovadores para medir la efectividad de la supervisión humana significativa. Asimismo, resulta pertinente ampliar la investigación hacia contextos institucionales y geográficos menos explorados, con el fin de fortalecer la aplicabilidad y generalización del conocimiento producido en este campo emergente.

Referencias

Adepoju, A., & Chinonyerem, C. (2025). Advancing good governance through AI-powered oversight in the United States: Risks and opportunities for public institutions. *MEJHLAR, 9*(6). Documento en línea. Disponible https://doi.org/10.70382/mejhlar.v9i6.069

Akinrinola, O., Okoye, C., Ofodile, O., & Ugochukwu, C. (2024). Navigating and reviewing ethical dilemmas in AI development: Strategies for transparency, fairness, and

Revista Aula Virtual, ISSN: 2665-0398; Periodicidad: Continua
Volumen: 7, Número: 14, Año: 2026 (Enero 2026 - Junio 2026)
Esta obra está bajo una Licencia Creative Commons Atribución No Comercial-Sin Derivar 4.0 Internacional

accountability. *GSC Advanced Research and Reviews, 18*(3), 50–58. Documento en línea. Disponible https://doi.org/10.30574/gscarr.2024.18.3.0088

Albaroudi, E., Mansouri, T., Hatamleh, M., & Alameer, A. (2026). HitHire: The future of ethical, fair, and sustainable AI recruitment: A governance framework. *Array, 29*, 100592. Documento en línea. Disponible https://doi.org/10.1016/j.array.2025.100592

Alotaibi, K. O. (2025). Developing a comprehensive financial reporting governance framework using AI techniques. *Engineering, Technology & Applied Science Research, 15*(6), 29202–29207. Documento en línea. Disponible https://doi.org/10.48084/etasr.14202

Bergomi, L., Nicora, G., Orlowska, M. A., Podrecca, C., Bellazzi, R., Fregosi, C., … Parimbelli, E. (2025). Which explanations do clinicians prefer? A comparative evaluation of XAI understandability and actionability in predicting the need for hospitalization. *BMC Medical Informatics and Decision Making, 25*, 269. Documento en línea. Disponible https://doi.org/10.1186/s12911-025-03045-0

Bhardwaj, A. (2025). Systematic literature review on explainable AI in finance: Methods, applications, and research gaps. *International Journal for Multidisciplinary Research, 7*(5). Documento en línea. Disponible https://doi.org/10.36948/ijfmr.2025.v07i05.58963

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

Bui, L. V. (2025). Legally-aware AI moderation in Vietnam: A cybersecurity law-compliant framework for platform governance. *Journal of Economic Criminology, 10*, Article 100193. Documento en línea. Disponible https://doi.org/10.1016/j.jeconc.2025.100193

Celsi, L., & Zomaya, A. (2025). Perspectives on managing AI ethics in the digital age. *Information, 16*(4), 318. Documento en línea. Disponible https://doi.org/10.3390/info16040318

Chintoh, G., Segun-Falade, O., Odionu, C., & Ekeh, A. (2024). Legal and ethical challenges in AI governance: A conceptual approach to developing ethical compliance models in the United States. *International Journal of Social Science Exceptional Research, 3*(1), 103–109. Documento en línea. Disponible https://doi.org/10.54660/ijsser.2024.3.1.103-109

Curcin, V., Delaney, B., Alkhatib, A., Cockburn, N., Dann, O., Kostopoulou, O., … Friedman, C. (2026). Learning health systems provide a glide path to safe landing for AI in health. *Artificial Intelligence in Medicine, 173*, 103346. Documento en línea. Disponible https://doi.org/10.1016/j.artmed.2025.103346

Floridi, L., & Ascani, A. (2026). Governing artificial intelligence in parliamentary institutions: An anticipatory ethics approach. *AI & Society*. Advance online publication. Documento en línea. Disponible https://doi.org/10.1007/s11023-025-09743-y

Frid, S., Bassegoda, O., Camacho Mahamud, M. A., Sanjuan, G., Armengol de la Hoz, M. Á., Celi, L., Cano Franco, I., Anmella, G., Cuñat López, T., Arellano, A. L., Leguízamo-Martínez, L. M., Mezquita, L., Peñafiel Macías, P. A., Gallardo-Pizarro, A., González Colom, R., Renú Jornet, A., Bracons Cucó, G., & Borrat Frigola, X. (2025). Bridging generative AI and healthcare practice: Insights from the GenAI Health Hackathon at Hospital Clínic de Barcelona. *BMJ Health & Care Informatics, 32*, e101640. Documento en línea. Disponible https://doi.org/10.1136/bmjhci-2025-101640

Galan-Cubillo, E., & Saez-Soro, E. (2025). Enhancing academic conferences with AI: Defining the role of the human–AI editor. *IAES International Journal of Artificial Intelligence, 14*(6), 4484–4493. Documento en línea.

Esta obra está bajo una Licencia Creative Commons Atribución No Comercial-Sin Derivar 4.0 Internacional

Disponible https://doi.org/10.11591/ijai.v14.i6.pp4484-4493

Giarmoleo, F., Ferrero, I., Rocchi, M., & Pellegrini, M. (2024). What ethics can say on artificial intelligence: Insights from a systematic literature review. *Business and Society Review, 129*(2), 258–292. Documento en línea. Disponible https://doi.org/10.1111/basr.12336

Goffin, T. (2025). Does AI in healthcare need an editor-in-chief? A leading example of true human oversight. *European Journal of Health Law, 32*(5), 579–602. Documento en línea. Disponible https://doi.org/10.1163/15718093-12423581

Ji, J., Qiu, T., Chen, B., Zhou, J., Zhang, B., Hong, D., Lou, H., Wang, K., Duan, Y., He, Z., Vierling, L., Zhang, Z., Zeng, F., Dai, J., Pan, X., Xu, H., O'Gara, A., Ng, K., Tse, B., … Gao, W. (2025). *AI alignment: A contemporary survey*. *ACM Computing Surveys, 58*(5), Article 132. Documento en línea. Disponible https://doi.org/10.1145/3770749

Kioskli, K., Kavakli, E., & Vrochidis, S. (2025). TrustSense: Measuring human oversight maturity for trustworthy AI systems. *Computers, 14*(11), 483. Documento en línea. Disponible https://doi.org/10.3390/computers14110483

Kortukova, T., Dei, M., Kudin, V. I., Onyshchenko, A., & Kravchuk, P. (2025). Legal challenges of artificial intelligence in the European Union's digital economy. *International Journal of Informatics and Communication Technology, 14*(3), 960–971. Documento en línea. Disponible https://doi.org/10.11591/ijict.v14i3.pp960-971

Lança, T. A., & Rocha, E. S. S. (2026). Inteligência artificial na gestão editorial de periódicos científicos: Uma análise aplicada à Revista Digital de Direito Administrativo (RDDA). *Revista Digital de Biblioteconomia e Ciência da Informação, 24*, e026017. Documento en línea. Disponible https://doi.org/10.20396/rdbci.v24i00.8680441

Manan, L., & Zakir, M. (2025). Digital surveillance, migration control, and human rights: Ethical dilemmas in the use of technology to govern human mobility. *Inverge Journal of Social Sciences, 4*(3), 319–336. Documento en línea. Disponible https://doi.org/10.63544/ijss.v4i3.170

Moreno-Sánchez, P. A., Del Ser, J., van Gils, M., & Hernesniemi, J. (2026). A design framework for operationalizing trustworthy artificial intelligence in healthcare. *Information Fusion, 127*, 103812. Documento en línea. Disponible https://doi.org/10.1016/j.inffus.2025.103812

Mozzarelli, M., & Schneider, G. (2026). AI and strategic decisions: Facing the incompleteness. *European Business Organization Law Review*. Advance online publication. Documento en línea. Disponible https://doi.org/10.1007/s40804-025-00356-7

Park, H., Kim, J., Jung, M., Kang, S.-Y., Kim, D., Kim, C., & Jang, U. (2025). Risk management challenges in maritime autonomous surface ships (MASSs): Training and regulatory readiness. *Applied Sciences, 15*, 10993. Documento en línea. Disponible https://doi.org/10.3390/app152010993

Pasupuleti, M. (2025). Governance frameworks for ethical AI deployment in public sector services. *IJAIRI, 5*(5), 561–573. Documento en línea. Disponible https://doi.org/10.62311/nesx/rphcr17

Patel, D. (2025). Towards transparent artificial intelligence: A comparative study of explainable AI models for decision-making in financial risk assessment. *International Scientific Journal of Engineering and Management, 4*(6), 1–9. Documento en línea. Disponible https://doi.org/10.55041/isjem04570

Ridzuan, N., Masri, M., Anshari, M., Fitriyani, N., & Syafrudin, M. (2024). AI in the financial sector: The line between innovation, regulation, and ethical responsibility. *Information, 15*(8),

Esta obra está bajo una Licencia Creative Commons Atribución No Comercial-Sin Derivar 4.0 Internacional

432. Documento en línea. Disponible https://doi.org/10.3390/info15080432

Saeed, K., & Prybutok, V. R. (2026). When utility meets ethics: A stakeholder perspective on agentic information systems delegation. *International Journal of Information Management, 86*, 102976. Documento en línea. Disponible https://doi.org/10.1016/j.ijinfomgt.2025.102976

Sagar, V. (2025). Regulatory human–agent teams: Co-creating ethical oversight. *European Modern Studies Journal, 9*(5), 562–583. Documento en línea. Disponible https://doi.org/10.59573/emsj.9(5).2025.51

Saup, T.-O., Asghar, J., Kanbach, D. K., & Kraus, S. (2026). From pilots to decision systems: Embedding generative AI into strategic decision-making through a socio-technical and governance lens. *Journal of Decision Systems, 35*(1), Article 2597835. Documento en línea. Disponible https://doi.org/10.1080/12460125.2025.2597835

Wang, T., & Tobias, G. R. (2025). Research on intelligent optimization mechanisms of financial process modules through machine learning-enhanced collaborative systems in digital finance platforms. *Future Technology, 4*(4), 240–254. Documento en línea. Disponible https://doi.org/10.55670/fpll.futech.4.4.20

Yan, Y., & Li, J. (2025). Interpretable AutoML for predicting unsafe miner behaviors via psychological-contract signals. *AI, 6*, 314. Documento en línea. Disponible https://doi.org/10.3390/ai6120314

Zaidan, E., Al-Shehabi, O., & Khamis, A. (2026). From ethical principles to regulatory governance of artificial intelligence. *Technology in Society, 85*, 103159. Documento en línea. Disponible https://doi.org/10.1016/j.techsoc.2025.103159

Esta obra está bajo una Licencia Creative Commons Atribución No Comercial-Sin Derivar 4.0 Internacional