Universidad CENFOTEC

Escuela de Ciberseguridad

Maestría en Ciberseguridad y Seguridad de la Información

Tema:

Propuesta de marco de gobernanza para el uso responsable de plataformas de inteligencia artificial en Pymes de Costa Rica

Elaborado por:

Oscar Avila Madriz

Fecha: Abril, 2026

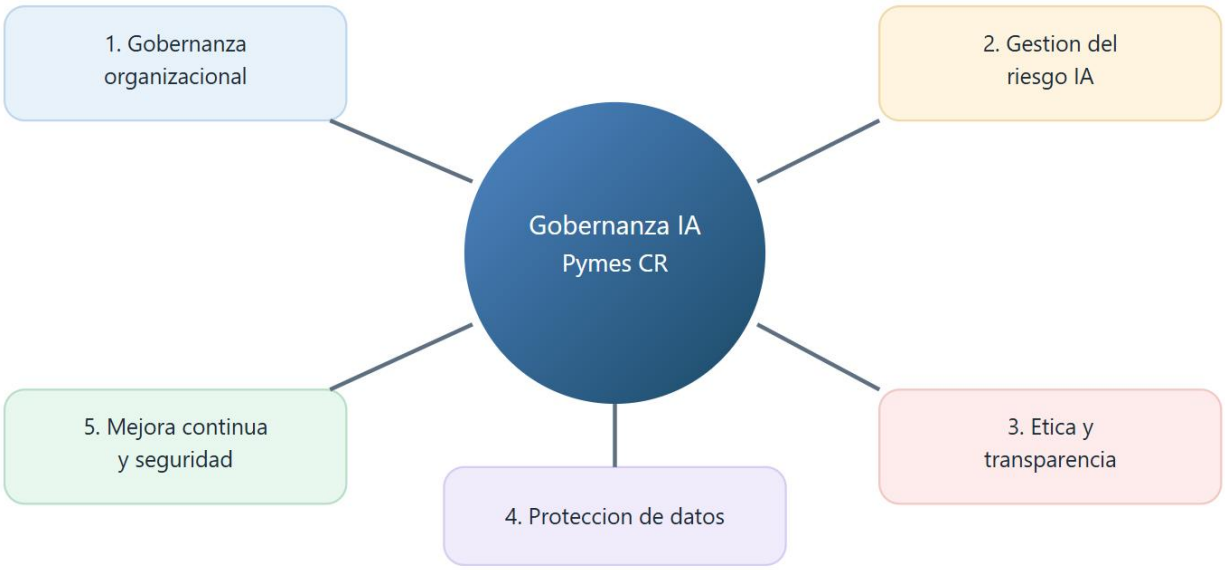
Resumen

La adopción de inteligencia artificial (IA) en las pequeñas y medianas empresas (Pymes) costarricenses ha crecido de forma acelerada, particularmente en procesos de atención al cliente, automatización operativa y análisis de información. Sin embargo, este crecimiento no ha sido acompañado por estructuras formales de gobernanza que orienten su implementación responsable. En consecuencia, muchas organizaciones operan con brechas en materia de seguridad de la información, privacidad, gestión del riesgo y control ético de los sistemas utilizados. El objetivo de esta investigación consistió en diseñar un marco de gobernanza para el uso responsable de plataformas de IA en Pymes de Costa Rica, alineado con estándares internacionales y ajustado a sus capacidades técnicas, organizacionales y financieras.

Metodológicamente, se desarrolló una investigación aplicada de enfoque cualitativo, con diseño no experimental y documental. Se utilizó revisión sistemática de literatura con base en el proceso de Biolchini et al. (2005), integrando fuentes académicas y normativas clave entre 2018 y 2025.

Entre los marcos analizados destacan NIST AI RMF 1.0, ISO/IEC 42001:2023, AI Act de la Unión Europea, lineamientos éticos de la OCDE y la Estrategia Nacional de IA de Costa Rica 2024-2027. Los hallazgos evidencian que, aunque existen referentes robustos a nivel internacional, su aplicación directa en Pymes resulta limitada por su complejidad operativa y costo de adopción. Asimismo, se identificaron riesgos críticos para el contexto nacional, incluyendo filtración de datos sensibles, uso no autorizado de herramientas de IA (shadow IT), sesgos algorítmicos, dependencia de proveedores externos e incumplimiento normativo.

A partir de estos resultados, se diseñó un marco de gobernanza estructurado en cinco componentes: gobernanza organizacional, gestión del riesgo de IA, ética y transparencia, protección de datos y seguridad de la información, y supervisión con mejora continua. Además, se incorporaron herramientas prácticas de implementación: modelo de madurez progresivo, matriz de riesgos, lista de autoevaluación, hoja de ruta por fases, definición de recursos necesarios y criterios de evaluación del marco.



En síntesis, el estudio concluye que las Pymes costarricenses requieren un modelo de gobernanza progresivo, contextualizado y económicamente viable para adoptar IA de forma segura, ética y sostenible. La propuesta desarrollada constituye una base aplicable para fortalecer la ciberseguridad organizacional, la confianza digital y la competitividad del sector.

Palabras Clave: Gobernanza de IA, Pymes, ciberseguridad, riesgo algorítmico, ética digital, protección de datos, cumplimiento normativo, mejora continua.

Abstract

The adoption of artificial intelligence (AI) among Costa Rican small and medium-sized enterprises (SMEs) has increased rapidly, especially in customer service, process automation, and data analysis. However, this growth has not been matched by formal governance structures to guide responsible implementation. As a result, many organizations face gaps in information security, privacy, risk management, and ethical control of AI-enabled processes. The objective of this research was to design an AI governance framework for the responsible use of AI platforms in Costa Rican SMEs, aligned with international standards and adapted to their technical, organizational, and financial capacities. Methodologically, the study followed an applied research approach with a qualitative, non-experimental, and documentary design. A systematic literature

review based on Biolchini et al. (2005) was conducted, incorporating key academic and institutional sources published between 2018 and 2025.

The review included NIST AI RMF 1.0, ISO/IEC 42001:2023, the European Union AI Act, OECD ethical guidelines, and Costa Rica's National AI Strategy 2024-2027. Findings show that, although international frameworks are robust, direct implementation in SMEs is often limited by complexity and resource constraints. Critical risks identified for the Costa Rican context include sensitive data leakage, unauthorized AI usage (shadow IT), algorithmic bias, excessive dependence on external providers, and regulatory non-compliance.

Based on these findings, a five-component governance framework was designed: organizational governance, AI risk management, ethics and transparency, data protection and information security, and supervision with continuous improvement. Practical implementation tools were also incorporated: a progressive maturity model, a risk matrix, a self-assessment checklist, a phased roadmap, required resources, and evaluation criteria. In conclusion, Costa Rican SMEs need a progressive, contextualized, and economically viable governance model to adopt AI in a secure, ethical, and sustainable way. The proposed framework provides an actionable foundation to strengthen organizational cybersecurity, digital trust, and sector competitiveness.

Keywords: AI governance, SMEs, cybersecurity, algorithmic risk, digital ethics, data protection, regulatory compliance, continuous improvement.

INTRODUCCIÓN

La inteligencia artificial se ha consolidado como una tecnología de alto impacto en la transformación digital empresarial. En Costa Rica, su adopción en Pymes avanza de manera significativa; sin embargo, este proceso ocurre en un entorno donde persisten vacíos de gobernanza y lineamientos operativos específicos para organizaciones de menor escala. La ausencia de políticas internas, roles definidos y controles básicos incrementa la exposición a riesgos de seguridad, privacidad y sesgos en decisiones asistidas por IA. En paralelo, la presión competitiva impulsa la incorporación de herramientas de terceros sin procesos formales de evaluación y supervisión. Este escenario justifica la necesidad de una propuesta de gobernanza que traduzca marcos internacionales a una realidad Pyme, bajo criterios de proporcionalidad, simplicidad y

aplicabilidad práctica. La investigación se orienta, por tanto, al diseño de un marco que permita implementar IA con responsabilidad institucional, alineación ética y control progresivo del riesgo.

TRABAJOS RELACIONADOS

El diseño de un marco de gobernanza de inteligencia artificial requiere la convergencia de marcos de gestión de riesgos, regulaciones y directrices éticas. En esta sección se analiza la literatura seleccionada mediante la metodología de Biolchini et al. (2005), contrastando los estándares internacionales con las condiciones del ecosistema Pymes costarricense.

Estándares internacionales de gobernanza y riesgo algorítmico

La gobernanza formal de la IA a nivel global se sustenta en dos pilares técnico-normativos principales. El NIST AI RMF 1.0 (2023) organiza la gestión del riesgo en cuatro funciones: Gobernar, Mapear, Medir y Gestionar. Su diseño es flexible y agnóstico respecto al tipo de sistema de IA, lo que lo convierte en el marco con mayor potencial de adaptación para Pymes. No obstante, su adopción plena requiere capacidades de medición y monitoreo continuo que exceden los recursos típicos de organizaciones sin personal especializado en ciencia de datos o auditoría algorítmica.

La norma ISO/IEC 42001:2023 establece los requisitos para un Sistema de Gestión de Inteligencia Artificial (SGIA), con un enfoque basado en el ciclo Plan-Do-Check-Act y controles definidos en su Anexo A. Aunque proporciona una estructura robusta de gobernanza corporativa, el costo asociado a la certificación, la complejidad documental y la necesidad de auditoría externa generan una barrera de entrada prohibitiva para organizaciones con estructuras de TI limitadas.

El AI Act de la Unión Europea (2024) introduce la primera regulación vinculante basada en niveles de riesgo (inaceptable, alto, limitado y mínimo). Su relevancia es conceptual para esta investigación, pero su aplicación directa resulta inviable para Pymes regionales debido a la complejidad jurídica y las sanciones financieras diseñadas para corporaciones de escala europea.

A nivel de principios éticos, Jobin, Ienca y Vayena (2019) analizaron 84 guías éticas de IA a nivel mundial, identificando once principios recurrentes, entre los cuales destacan la transparencia, la justicia, la responsabilidad, la privacidad y la no

maleficencia. Estos principios definen el "qué" de la gobernanza ética, pero la literatura no ofrece el "cómo" para entornos de recursos severamente limitados.

Contexto normativo nacional

A nivel local, la Estrategia Nacional de Inteligencia Artificial de Costa Rica (ENIA-CR) 2024-2027, promulgada por el MICITT, traza la hoja de ruta para la adopción tecnológica, la ética digital y la competitividad. Sin embargo, el análisis de su contenido revela una desconexión operativa: define objetivos macroeconómicos y principios alineados con la OCDE, pero omite herramientas prácticas o guías específicas para que las Pymes puedan implementar dichos principios. Las Pymes costarricenses se encuentran en un escenario donde el 50% ya utiliza algún tipo de IA (Microsoft, 2025), pero ninguna cuenta con un marco de gobernanza adaptado a la legislación local, particularmente la Ley 8968 de Protección de la Persona frente al Tratamiento de sus Datos Personales.

Identificación de la brecha

La siguiente tabla sintetiza las limitaciones de los marcos vigentes al ser aplicados en el contexto Pyme costarricense:

Tabla 1

Limitaciones de marcos de gobernanza de IA

Marco	Enfoque principal	Fortalezas	Barrera para Pymes CR
NIST AI RMF 1.0	Gestión de riesgos sociotécnicos	Flexible, escalable, aplicable a cualquier sistema	Exige métricas complejas y personal dedicado de monitoreo continuo
ISO/IEC 2001:2023	Sistema de gestión auditable	Gobernanza corporativa robusta, controles de ciberseguridad	Alta burocracia documental, costos de certificación y auditoría externa

AI Act (UE)	Regulación por niveles de riesgo	Primer marco legal vinculante, protección de derechos	Complejidad jurídica extraterritorial, sanciones inviables para Pymes
ENIA-CR 2024-2027	Política pública nacional	Alineación gubernamental, fomento de ética digital	Enfoque macro y conceptual, sin herramientas prácticas para el empresario Pyme

Nota. Elaboración propia a partir de NIST (2023), ISO/IEC (2023), Unión Europea (2024) y MICITT (2024).

Los marcos internacionales asumen que las organizaciones poseen departamentos de cumplimiento, presupuestos independientes de ciberseguridad y arquitecturas de datos maduras. Es en este vacío donde se posiciona la presente investigación, proponiendo un marco progresivo y modular que traduce las exigencias del NIST AI RMF y la ISO/IEC 42001 en controles de ciberseguridad de bajo costo y alta aplicabilidad para las Pymes de Costa Rica.

MÉTODO

El estudio se desarrolló como investigación aplicada, con alcance descriptivo y propositivo, bajo un enfoque cualitativo. El diseño fue no experimental y documental, dado que se analizaron fuentes secundarias sin manipulación de variables. La revisión sistemática de literatura siguió el proceso metodológico de Biolchini et al. (2005): definición del problema, formulación de pregunta de investigación, planificación de búsqueda, criterios de inclusión/exclusión, selección de estudios y síntesis de resultados.

La pregunta de investigación formulada fue: ¿qué marcos, lineamientos o buenas prácticas de gobernanza de inteligencia artificial pueden adaptarse al contexto de las pequeñas y medianas empresas en Costa Rica, considerando sus capacidades técnicas, organizacionales y financieras? El muestreo fue no probabilístico de tipo teórico, dado que las fuentes se seleccionaron con base en su relevancia conceptual y metodológica

para la construcción del marco, y no mediante criterios estadísticos. Este enfoque permitió priorizar marcos internacionales con reconocimiento institucional y cobertura temática complementaria entre sí.

Se utilizaron términos en español e inglés vinculados con gobernanza de IA, gestión de riesgo algorítmico, ética, marcos normativos y aplicación en Pymes. La ventana temporal de análisis se delimitó entre 2018 y 2025 para incorporar marcos vigentes y evidencia reciente. Se seleccionaron seis fuentes estructurales para el diseño del marco: ENIA-CR 2024-2027, NIST AI RMF 1.0, ISO/IEC 42001:2023, AI Act de la Unión Europea, OECD Framework for the Classification of AI Systems y el estudio de Jobin, Ienca y Vayena (2019) sobre principios éticos globales de IA.

Como técnica de análisis se empleó análisis de contenido, complementado con triangulación conceptual entre marcos internacionales y condiciones del contexto costarricense. La validación de pertinencia y aplicabilidad del marco se prevé mediante juicio de expertos como etapa posterior de contrastación, habiéndose definido para ello criterios de evaluación específicos con indicadores medibles de cobertura, cumplimiento, reducción de riesgos y avance en el nivel de madurez.

RESULTADOS

El análisis comparativo confirmó que los marcos internacionales ofrecen bases sólidas, pero presentan barreras de adopción directa para Pymes debido a exigencias de madurez organizacional, disponibilidad de talento y costo de implementación.

Se identificaron como riesgos de mayor criticidad la filtración de datos sensibles y el uso no autorizado de plataformas de IA por parte de colaboradores.

También se evidenciaron riesgos moderados recurrentes asociados a sesgo algorítmico, pérdida de integridad de información por resultados no verificados, dependencia de proveedores y cumplimiento legal insuficiente. A partir de estos hallazgos, se diseñó un marco de gobernanza con cinco componentes interrelacionados: (1) gobernanza organizacional, para definir roles, políticas y responsabilidades; (2) gestión del riesgo de IA, para identificar, evaluar y mitigar amenazas prioritarias; (3) ética y transparencia, para promover trazabilidad y decisiones explicables; (4) protección de datos y seguridad de la información, para controlar el tratamiento de

información y reducir exposición; y (5) supervisión y mejora continua, para mantener actualizado el marco frente a cambios normativos y tecnológicos.

Como soporte operativo, el marco integra un conjunto de herramientas prácticas diseñadas bajo el principio de proporcionalidad, que permiten transformar los principios generales en acciones concretas para organizaciones con recursos limitados.

El modelo de madurez progresivo organiza la gobernanza en cuatro niveles. El Nivel 1 (Inicial) refleja una organización sin políticas ni controles, donde las herramientas de IA se utilizan de forma informal y sin supervisión. El Nivel 2 (Básico) corresponde a organizaciones que han establecido políticas iniciales y designado responsabilidades, aunque los controles aún son incipientes. El Nivel 3 (Gestionado) describe a organizaciones con procesos formales de gestión del riesgo, controles documentados y mecanismos de supervisión periódica. El Nivel 4 (Avanzado) representa la integración plena de la gobernanza de IA en la cultura organizacional, con un ciclo activo de mejora continua y capacidad de adaptación proactiva ante cambios normativos y tecnológicos. Este modelo permite a cada Pyme identificar su punto de partida y avanzar de forma gradual sin exigirse alcanzar el nivel más alto de manera inmediata.

La matriz de riesgos identifica y prioriza ocho amenazas asociadas al uso de IA en Pymes, valoradas según su probabilidad e impacto en una escala de 1 a 9. Los dos riesgos clasificados como críticos (nivel 9) son la filtración de datos sensibles por ingreso de información confidencial en plataformas externas y el uso no autorizado de herramientas de IA por parte de colaboradores (shadow IT). Como riesgos moderados se identificaron el sesgo algorítmico en decisiones organizacionales, la dependencia excesiva de proveedores externos, el incumplimiento normativo, la pérdida de integridad de información por alucinaciones del modelo, y la falta de documentación del uso de IA. Para cada riesgo se definen medidas de mitigación específicas, tales como políticas de clasificación de datos, prohibición de ingreso de información sensible, inventarios de herramientas, supervisión humana de decisiones críticas y revisión de contratos con proveedores.

La lista de chequeo de autoevaluación consta de 25 preguntas distribuidas a través de los cinco componentes del marco, con formato de respuesta dicotómica (Sí/No). Este instrumento permite a cualquier Pyme realizar un diagnóstico rápido de su

estado actual de gobernanza de IA sin requerir conocimiento técnico especializado, y su resultado se vincula directamente con los niveles del modelo de madurez.

Complementariamente, la propuesta incluye una hoja de ruta de implementación organizada en tres fases: diagnóstico y sensibilización (meses 1-2), implementación de controles básicos (meses 3-5) y seguimiento con mejora continua (mes 6 en adelante). Se definen además los recursos necesarios para la implementación, los cuales se limitan principalmente al tiempo del responsable designado y del equipo directivo, sin requerir inversiones en software especializado ni contratación de personal adicional. Finalmente, se establecen seis criterios de evaluación medibles con indicadores, metas y frecuencias de revisión que permiten verificar la efectividad del marco a lo largo del tiempo.

Con el propósito de demostrar la aplicabilidad del marco en un escenario representativo, se desarrolló un caso de uso basado en una Pyme hipotética del sector contable. ContaDigital S.R.L. es una firma de servicios contables y asesoría tributaria con 32 empleados, ubicada en el Gran Área Metropolitana de Costa Rica, que utiliza ChatGPT para redacción de respuestas a clientes, resumen de documentos legales y generación de borradores de informes tributarios.

En la fase de diagnóstico, la aplicación del instrumento de autoevaluación ubicó a la organización en el Nivel 1 (Inicial) del modelo de madurez. Se identificaron condiciones críticas: ausencia total de políticas de uso de IA, inexistencia de un responsable de gobernanza, ningún proceso de identificación de riesgos, y la práctica confirmada de colaboradores ingresando datos sensibles de clientes (números de cédula, estados financieros y detalle de obligaciones tributarias) directamente en la plataforma sin ningún tipo de control.

Durante la fase de implementación de controles básicos, se designó un responsable de gobernanza, se elaboró una política de uso aceptable, se aplicó la matriz de riesgos identificando la filtración de datos financieros como riesgo crítico (nivel 9), se estableció la prohibición de ingresar datos personales o financieros reales en plataformas de IA, se clasificó la información en tres niveles (pública, interna y confidencial), y se implementaron controles de acceso que restringen el uso de ChatGPT exclusivamente a colaboradores autorizados.

Seis meses después de la implementación, una segunda aplicación del checklist de autoevaluación ubicó a la empresa en el Nivel 3 (Gestionado). Los indicadores de desempeño registraron los siguientes resultados: 100% del personal capacitado en uso responsable de IA, cero incidentes de ingreso de datos sensibles en plataformas externas desde la implementación de la política, y 100% de las herramientas de IA documentadas en el inventario organizacional. La transición del Nivel 1 al Nivel 3 se logró sin inversiones en software especializado, sin contratación de personal adicional y utilizando únicamente los recursos internos existentes de la organización.

A nivel de controles técnicos, la investigación identificó que las políticas de gobernanza deben acompañarse de mecanismos que las hagan cumplir en la práctica. Para ello, se propone la implementación progresiva de *gateways* de inteligencia artificial: soluciones técnicas que se ubican entre los usuarios de la organización y las plataformas de IA externas (como ChatGPT, Copilot o Gemini), funcionando de manera análoga a un proxy de seguridad especializado.

Las funciones principales de un *gateway* de IA incluyen la inspección de contenido saliente mediante análisis de *prompts* antes de su transmisión a la plataforma, detectando patrones que correspondan a datos sensibles; el bloqueo o redacción automática de información confidencial identificada; el registro centralizado de auditoría de todos los *prompts* enviados y respuestas recibidas; y el control de acceso por plataforma, que permite definir qué herramientas de IA están autorizadas y bloquear el acceso a plataformas no aprobadas, contribuyendo directamente a la gestión del *shadow IT*.

La implementación de estas soluciones se propone en tres niveles según la capacidad de la Pyme: un nivel básico de costo mínimo mediante políticas de grupo (GPO) o administración de dispositivos móviles (MDM) para restringir acceso a sitios no autorizados; un nivel intermedio con proxies web que permiten registrar tráfico hacia plataformas de IA y aplicar reglas básicas de prevención de pérdida de datos (DLP); y un nivel avanzado con plataformas especializadas que ofrecen inspección profunda de *prompts*, redacción automática de datos sensibles y paneles de cumplimiento. Para que el *gateway* pueda inspeccionar el contenido cifrado, se requiere implementar

inspección TLS (*TLS inspection*), lo cual implica la instalación de certificados raíz internos en los dispositivos gestionados.

Este componente técnico refuerza directamente la gestión del riesgo al mitigar la exfiltración de datos, la protección de información al automatizar reglas de clasificación, y la supervisión continua al proporcionar registros que alimentan los indicadores de desempeño del marco. Su incorporación progresiva permite que la gobernanza trascienda el plano documental y se materialice en la infraestructura tecnológica de la organización.

CONCLUSIONES

La investigación permitió establecer que la adopción de IA en Pymes costarricenses avanza más rápido que su capacidad de gobernanza, lo que incrementa riesgos de ciberseguridad, privacidad y cumplimiento. También se concluye que no existe, en el contexto nacional, una guía operativa específica para este segmento empresarial que traduzca los marcos internacionales en prácticas simples y aplicables.

El marco diseñado responde a ese vacío mediante una estructura modular de cinco componentes y un conjunto de herramientas operativas que permiten su adopción progresiva. El caso de uso demostró que una Pyme puede transitar del Nivel 1 (Inicial) al Nivel 3 (Gestionado) en seis meses, eliminando riesgos críticos de filtración de datos y alcanzando cobertura total en capacitación e inventario de herramientas, sin requerir inversiones en software especializado ni contratación de personal adicional.

Se concluye además que la gobernanza de IA en Pymes no puede limitarse al plano documental. La propuesta de controles técnicos como los *gateways* de inteligencia artificial demuestra que existen mecanismos accesibles para materializar las políticas en la infraestructura tecnológica, reforzando la prevención de exfiltración de datos y la gestión del *shadow IT* desde una perspectiva de seguridad de la información.

En términos de aporte, el trabajo integra fundamentos internacionales con necesidades locales, generando una base de referencia aplicable para la implementación institucional y como punto de partida para su implementación en organizaciones del sector Pyme costarricense.

BIBLIOGRAFIA

Microsoft. (2025, mayo 12). *50 % de las pymes en Costa Rica utilizan algún tipo de IA, según encuesta*. Microsoft News Center Latinoamérica. https://news.microsoft.com/es-xl/50-de-las-pymes-en-costa-rica-utilizan-algun-tipo-de-ia

MICITT. (2024). *Estrategia Nacional de Inteligencia Artificial de Costa Rica 2024-2027*. Ministerio de Ciencia, Innovación, Tecnología y Telecomunicaciones. https://www.camtic.org/wp-content/uploads/2024/04/Estrategia-Nacional-de-Inteligencia-Artificial-Version-21.03.24-Para-consulta-pu%CC%81blica.pdf

Unión Europea. (2023). *Proposal for an Artificial Intelligence Act*. European Commission.

Instituto de Normas Internacionales de Estandarización. (2023). ISO/IEC 42001:2023 – Information technology — Artificial intelligence — Management system. Ginebra, Suiza: ISO.

Instituto Nacional de Estándares y Tecnología. (2023). *Artificial Intelligence Risk Management Framework (AI RMF 1.0)*. U.S. Department of Commerce. https://doi.org/10.6028/NIST.AI.100-1

Biolchini, J., Gomes, P., Cruz, A., & Horta, F. (2005). *Systematic Review in Software Engineering*. Technical Report RT-ES 679/05. COPPE/UFRJ.

Organización para la Cooperación y el Desarrollo Económico. (2022). *OECD Framework for the Classification of AI Systems.* OCDE.

Ray, S. (2023, mayo 2). Samsung bans ChatGPT among employees after sensitive code leak. *Mashable*. https://www.forbes.com/sites/siladityaray/2023/05/02/samsung-bans-chatgpt-and-other-chatbots-for-employees-after-sensitive-code-leak/

Jobin, A., Ienca, M., & Vayena, E. (2019). The global landscape of AI ethics guidelines. *Nature Machine Intelligence*, 1(9), 389–399. https://www.nature.com/articles/s42256-019-0088-2