

HERRAMIENTA DE EVALUACIÓN Y RECOMENDACIÓN PARA LA PROMOCIÓN DE USO RESPONSABLE DE IA EN PYMES LATINOAMERICANAS

David José Daza Jaimes
Diana Marcela Daza Jaimes
Juan Camilo Medina Moreno
Luis Carlos Ordoñez Montenegro
Ángela Pinilla Parra

Resumen

Las MiPymes representan el 99,5 % de las unidades productivas de América Latina y el Caribe, pero adoptan la IA generativa de forma acelerada y sin salvaguardas mínimas, en un contexto marcado por la informalidad y la dependencia de proveedores externos. Los estándares globales (NIST AI RMF e ISO/IEC 42001) resultan impracticables a esta escala, y las iniciativas regionales existentes funcionan como listas de verificación estáticas. Este trabajo propone una herramienta tipo SaaS que democratiza la gobernanza ética de la IA al promover la aplicación de normas internacionales y principios fundamentales dentro del sector privado. El instrumento traduce la densidad técnica del NIST AI RMF, la norma ISO/IEC 42001 y un marco ético de cinco principios a un árbol adaptativo de preguntas redactadas en lenguaje llano, y, sobre esa base, despliega tres capas encadenadas (diagnóstico, recomendaciones y ejecución) que culminan en planes de acción para ayudar a nutrir la gobernanza ética de la IA en la empresa.

Al tratarse de un entregable de diseño ilustrado mediante un caso teórico, sus resultados se argumentan en términos de cobertura, robustez y trazabilidad, no de desempeño estadístico. La principal conclusión es que la gobernanza ética de la IA en el Sur Global no debe ser un privilegio corporativo, sino una herramienta de gestión accesible, incluso para las empresas más pequeñas.

1. INTRODUCCIÓN

Para 2025, las micro, pequeñas y medianas empresas representaron el 90 % de las empresas a nivel mundial, fueron responsables del 70 % del empleo total y aportaron aproximadamente el 50 % del PIB global (Bhalla y Madrekar, 2025). América Latina y el Caribe (ALC) no es ajena a este panorama, en donde este tipo de organizaciones representan el 99,5 % del total de las unidades productivas, dentro del cual las microempresas constituyen la gran mayoría con un 88,4 %. No obstante, este predominio numérico no se traduce proporcionalmente en su contribución económica, pues son únicamente responsables de una cuarta parte del PIB regional (CAF, 2024).

Esta brecha productiva se encuentra vinculada a factores transversales como las limitadas capacidades de gestión, la especialización en la exportación de materias primas de bajo valor agregado y a elevados índices de informalidad laboral y empresarial (CAF, 2024). De acuerdo con datos de la Organización Internacional del Trabajo (OIT), aproximadamente el 50 % de las MiPymes operan en la informalidad, enfrentando restricciones críticas para acceder a recursos humanos calificados, infraestructura básica y tecnologías avanzadas (OIT, 2022).

A pesar de la priorización de políticas de fomento y modernización de MiPymes en las agendas públicas regionales, persisten brechas de autonomía, equidad económica y, cada vez con más fuerza, de innovación, entendido como la generación o introducción de ideas para abordar distintos tipos de

retos (OCDE/CAF, 2019). «Las empresas pueden innovar adoptando innovaciones existentes que son nuevas para la empresa pero que ya han sido aplicadas por otros (difusión tecnológica) o introduciendo nuevas innovaciones en el mercado (desarrollo tecnológico). La innovación es un motor fundamental de la productividad y la competitividad, pero muchas empresas —en particular las pequeñas y jóvenes— se enfrentan a obstáculos para la adopción y el desarrollo de nuevas tecnologías» (CAF, 2024 citando a OCDE, 2023).

En materia tecnológica, la implementación de la Inteligencia Artificial (IA) refleja también una asimetría, no solo entre PyMES y grandes empresas (CAF, 2024), sino en el rol de la región en la cadena de valor global de la IA. Como señala la CEPAL (2025): «La región consume intensamente soluciones finales [de IA], pero integra y produce menos (...) la región corre el riesgo de consolidarse como consumidora neta de IA, con baja capacidad de adaptación a contextos locales y dependencia de proveedores externos».

Esta dependencia tecnológica a proveedores externos debilita la capacidad de los países latinoamericanos para desarrollar soluciones locales y supervisarlas de manera independiente. La falta de capacidades internas para auditar, modificar o supervisar de manera independiente los sistemas de IA adquiridos genera un escenario donde el despliegue de soluciones automatizadas se realiza sin las salvaguardas mínimas o estructuras de gobernanza necesarias para garantizar un uso seguro y ético (CENIA, 2025). Las MiPymes de ALC, impulsadas por la promesa de mejorar su eficiencia operativa, las bajas barreras de entrada e interfaces simples de la IA Generativa integran estas herramientas de manera apresurada, frecuentemente omitiendo la evaluación de riesgos sistémicos (CENIA, 2025) debido a la ausencia de metodologías accesibles y adecuadas a su escala operativa (CCIT, 2025).

Tecnología, ética y derechos humanos

La adopción de la IA en la mayoría de los sectores como herramienta de productividad implica la transformación de muchas facetas de la vida. Desde la creación de contenido sintético, pasando por solicitudes de empleo, evaluaciones de soluciones financieras o regulación de acceso a centros educativos, hasta la predicción de delitos, tasas de reincidencia o detección de fraudes. Los sistemas de IA participan cada vez más en la elaboración de pronósticos, recomendaciones o en la toma de decisiones con efectos sobre las personas, bien sea con consecuencias negativas o que pueden llegar a afectar derechos fundamentales; siendo muchas de estas decisiones tomadas por entidades privadas (Algorithm Watch, 2025). Este contexto sitúa a las empresas como intermediarios críticos en el ejercicio de los derechos humanos (efecto horizontal de los derechos fundamentales o la teoría del *Drittwirkung der Grundrechte* (Borowski, 2020)).

En este nuevo ecosistema, la injerencia de los principios constitucionales en las relaciones comerciales e industriales repercute de manera directa en las responsabilidades legales y éticas de cualquier organización, y la ausencia de una regulación específica o de marcos sancionatorios explícitos en materia de inteligencia artificial en diversas jurisdicciones de ALC no puede constituirse como una excusa para la inacción o para operar al margen de los estándares éticos vigentes en las sociedades democráticas. Las responsabilidades en materia de derechos humanos son preexistentes a las leyes de gobernanza tecnológica; por lo tanto, cualquier empresa que despliegue sistemas algorítmicos de impacto social está sujeta al deber de diligencia debida, viéndose obligada a mitigar activamente riesgos asociados a la opacidad y la injusticia distributiva de sus herramientas, entre otros.

Para hacer frente a los riesgos del despliegue algorítmico, durante los últimos años, han surgido varias iniciativas de gobernanza de la IA a nivel internacional, regional y nacional orientados a establecer principios, estándares, marcos y lineamientos para el uso de esta tecnología en distintos sectores. Aunque la mayoría de los principios se solapan entre sí, es posible, siguiendo al filósofo italiano Luciano Floridi (2024), agruparlos en cinco: beneficencia, no maleficencia, autonomía, justicia y explicabilidad. Mientras los principios conservan su carácter general y universal, la aplicación práctica de la gobernanza exige trazar una distinción clara y funcional dentro de cada principio: qué corresponde exigir y verificar al proveedor tecnológico (responsabilidad originaria) y qué corresponde al uso responsable por parte del usuario u organización que lo emplea (responsabilidad operativa).

La herramienta propuesta

Para abordar de manera estructural las asimetrías descritas, mitigar la condición de ALC como consumidora neta de tecnología y facilitar el cumplimiento ético de forma transversal en el Sur Global, el presente documento presenta una propuesta de intervención tecnológica basada en el modelo de Software as a Service (SaaS). Esta iniciativa surge ante la insuficiencia operativa de los estándares globales tradicionales —tales como el marco NIST AI RMF o la norma internacional ISO/IEC 42001—, los cuales, a pesar de su rigor, imponen exigencias burocráticas, de costos y de especialización técnica orientadas a grandes corporaciones, resultando impracticables para el ecosistema de las MiPymes informales de la región.

El objetivo central del SaaS propuesto es democratizar el acceso a la gobernanza de la IA, permitiendo a las organizaciones con recursos limitados evaluar y gestionar sus riesgos algorítmicos mediante un entorno automatizado y verificable. La principal innovación metodológica del sistema radica en su capacidad para actuar como un motor de traducción simplificada. Al ingresar a la plataforma, el software genera un diagnóstico personalizado del panorama actual de la organización, parametrizado no en función de abstracciones normativas, sino según el tipo de uso de IA que la organización declare (desarrollo de productos para terceros, automatización de procesos internos, o ambos) y el nivel de riesgo identificado en sus respuestas. El SaaS procesa internamente la densidad técnica del NIST AI RMF y la norma ISO/IEC 42001, traduciéndola a un cuestionario adaptativo, diseñado y congelado por el equipo, redactado en un lenguaje llano.

Una vez procesada la información del entorno, la plataforma contrasta las prácticas de la empresa en materia de IA, con la arquitectura simplificada de los cinco principios éticos mencionados, con el NIST AI Risk Management Framework (AI RMF) como marco de referencia para la gestión de riesgos de IA, y con la norma ISO/IEC 42001 como estándar internacional para la implementación y evaluación de sistemas de gestión de inteligencia artificial (AIMS). Como resultado de esta evaluación, el SaaS no se limita a emitir una puntuación estática, sino que genera de forma automatizada un plan de acción incremental estructurado en dos horizontes temporales críticos para la supervivencia y formalización de la MiPyme:

Plan de recomendaciones básicas a 30 días: Prioriza las brechas de mayor riesgo y menor esfuerzo detectadas por el motor. Esto suele incluir, según el caso, controles sobre datos personales, supervisión humana, inventario de sistemas y canales de atención, pero la secuencia exacta depende del perfil de la organización.

Plan de recomendaciones de consolidación a 90 días: Orientado a la estructuración de capacidades internas sostenibles. Este bloque guía a la MiPyme en la capacitación básica de sus operarios, la documentación del ciclo de vida de las herramientas integradas y la alineación de sus metas de productividad con salvaguardas éticas que sirvan como factor de diferenciación competitiva y confianza reputacional en el mercado digital.

Con esta propuesta, el presente estudio busca demostrar que la gobernanza ética de la IA en el Sur Global no debe ser un privilegio corporativo exclusivo, sino una capacidad operativa accesible que proteja los derechos fundamentales y fortalezca la equidad competitiva en la economía impulsada por los datos.

2. TRABAJOS RELACIONADOS

El desarrollo de metodologías de gobernanza y evaluación ética de IA en ALC cuenta con antecedentes en el ámbito multilateral. El esfuerzo previo más similar y representativo del estado del arte actual es la iniciativa fAIr LAC, impulsada por el Banco Interamericano de Desarrollo (BID), la cual materializó la herramienta de Autoevaluación Ética de IA para Actores del Ecosistema Emprendedor. Este instrumento provee un marco conceptual y espacios de entrenamiento para que las empresas emergentes mitiguen riesgos asociados a la privacidad, la equidad y la transparencia a través de un cuestionario cualitativo de autoverificación. A nivel gremial, la Cámara Colombiana de Informática y Telecomunicaciones (CCIT) ha propuesto una hoja de ruta para una Plataforma Regional que busca articular al sector privado y proveer herramientas (toolkits) de IA específicas para MiPymes.

A pesar de su relevancia, estas iniciativas presentan una brecha operativa crítica frente a la informalidad estructural de la región. Tanto fAIr LAC como la propuesta de la CCIT funcionan como listas de verificación (checklists) estáticas o directrices macroestructurales. Su aplicación presupone que las organizaciones poseen madurez administrativa y personal especializado capacitado para asimilar terminología compleja y traducir principios abstractos en procesos de ingeniería. Esto genera un vacío de ejecución en las microempresas informales, las cuales quedan excluidas de los beneficios de la seguridad algorítmica por carecer de la capacidad para asimilar marcos de alta densidad técnica. El sistema propuesto en este trabajo busca cerrar ese vacío de ejecución al generar, además del diagnóstico, los artefactos de cierre que la organización necesita para implementar las recomendaciones.

Por otro lado, en lo relativo al aspecto ético, en la región ha habido propuestas de distintas organizaciones y organismos estatales, como la UNESCO, la OCDE, el gobierno colombiano o peruano, entre otros, que proponen una serie de principios que, tomados en su totalidad, serían más de 100, dificultando el despliegue de gobernanza orientado a la construcción de marcos y lineamientos para el uso de esta tecnología en distintos sectores. Frente a esta problemática, adoptamos la propuesta de Luciano Floridi (2024), quien sostiene la tesis que el alto número de principios es causado no por un número idéntico a problemáticas o soluciones, sino a la terminología utilizada, propone un marco unificado que condensa las decenas de directrices globales en cinco pilares esenciales que se derivan de la bioética: beneficencia, no maleficencia, autonomía, justicia y explicabilidad (Floridi, 2024). Si bien la tesis de Floridi dota de robustez y síntesis conceptual a la ética algorítmica, su aplicación en el diseño de software sigue siendo predominantemente teórica en la academia y abstracta para el entorno empresarial ordinario, haciendo necesaria su aplicabilidad en herramientas prácticas.

Como marco ético adoptado para el aplicativo, se adoptaron los principios enunciados en las recomendaciones de uso de la IA de la UNESCO y OCDE, ampliamente aceptadas por países latinoamericanos, y que tienen una totalidad de 15 distintos principios, que puede ser traducidos o comprendidos por los cinco principios enunciados:

Tabla 1: Traductibilidad de los principios enunciados en UNESCO y OCDE a la propuesta de Floridi (2024)

	Principios incluidos en el marco	Principios propuestos por Floridi (2024)
UNESCO	Proporcionalidad e inocuidad	Beneficencia + No Maleficencia
	Seguridad y protección	Beneficencia + Explicabilidad
	Equidad y No discriminación	Justicia
	Sostenibilidad	Justicia
	Derecho a la intimidad y protección de datos	Autonomía + No maleficencia
	Supervisión y decisión humanas	Autonomía + Explicabilidad
	Transparencia y Explicabilidad	Explicabilidad
	Responsabilidad y rendición de cuentas	Explicabilidad
	Sensibilización y educación	Justicia
	Gobernanza y colaboración adaptativas	Justicia + Beneficencia
OCDE	Crecimiento inclusivo, desarrollo sostenible y bienestar	Beneficencia + Autonomía + Justicia
	Respetar el Estado de derecho, los derechos humanos y los valores democráticos	Beneficencia + Autonomía + Justicia + Explicabilidad
	Transparencia y explicabilidad	Explicabilidad + Autonomía
	Solidez, seguridad y protección	No maleficencia
	Responsabilidad	Explicabilidad + Beneficencia

3. METODOLOGÍA

Idea central del diseño

El sistema se construyó separando dos funciones que en muchas herramientas de IA se mezclan sin querer: lo que decide y lo que redacta. Todo lo que tiene una consecuencia normativa (qué tan grave es una brecha, qué controles le faltan a la empresa, qué tan madura es su gobernanza de IA) lo calcula un motor de reglas fijo, sin intervención de ningún modelo de lenguaje. El modelo de lenguaje entra recién al final, y solo para una tarea: redactar la política de IA a la medida de cada empresa, con base en lo que el motor ya determinó.

Esta separación es la respuesta directa al problema que motiva el proyecto. Muchas herramientas de evaluación de IA terminan siendo «cumplimiento cosmético» porque si un modelo de lenguaje decide tanto el diagnóstico como la redacción, dos empresas con las mismas respuestas podrían recibir resultados distintos cada vez que se ejecuta. Con el diseño elegido, eso no puede pasar: las mismas respuestas siempre dan el mismo diagnóstico.

Motor de diagnóstico

La evaluación se basa en un árbol de 34 nodos en total: 15 preguntas raíz más 19 subpreguntas condicionales. De las 15 raíz, 11 son comunes a cualquier organización y 4 se activan según el tipo de uso de IA declarado (desarrollo de productos para terceros, automatización interna, o ambos). Las subpreguntas aparecen solo si la respuesta anterior lo amerita; por ejemplo, si la empresa declara que usa datos personales, se le pregunta si tiene consentimiento, y si no los usa, esa pregunta no aparece. De los 34 nodos, 30 son puntuables y contribuyen al diagnóstico, 4 son campos de justificación en texto libre que enriquecen el diagnóstico sin puntuar. Ninguna empresa responde más de 20 preguntas en un recorrido.

Cada respuesta tiene un puntaje fijo definido de antemano. Para combinar preguntas con escalas distintas, el motor normaliza cada respuesta dividiéndola entre su puntaje máximo (escala 0 a 1). Las respuestas marcadas como «No Aplica» se excluyen del cálculo en lugar de tratarse como cero.

Por otra parte, cada pregunta está conectada a tres marcos de referencia: los cinco principios éticos (Beneficencia, No maleficencia, Autonomía, Justicia, Explicabilidad), ISO/IEC 42001, y NIST AI RMF. El sistema calcula tres puntajes finales mediante un promedio ponderado:

$$\text{Puntaje por eje} = 100 \cdot \frac{\sum_i a_i \cdot w_{i,\text{eje}}}{\sum_i w_{i,\text{eje}}} \quad (1)$$

donde a_i es el puntaje normalizado de la respuesta i (0 a 1) y $w_{i,\text{eje}}$ es el peso de esa pregunta en ese eje específico. La suma recorre únicamente las preguntas aplicables a la ruta de bifurcación de la organización y efectivamente puntuadas, lo que mantiene comparable a una PYME que respondió una trayectoria corta frente a otra que respondió una trayectoria larga. El resultado es un vector ($x = \text{Ético}$, $y = \text{ISO}$, $z = \text{NIST}$) en un espacio de tres dimensiones, cada eje normalizado entre 0 y 100, que ubica visualmente a la organización evaluada.

Los 90 pesos (30 preguntas \times 3 ejes) se generaron mediante un modelo de lenguaje con acceso a los textos oficiales de NIST e ISO, que propuso cada peso junto con la justificación normativa que lo respalda. Estos pesos fueron revisados por el equipo y se encontraron casos donde el modelo había citado un control normativo incorrecto, lo cual se corrigió manualmente antes de congelarlos. Solo se auditó una muestra, no el conjunto completo. Una vez congelados, el modelo de lenguaje no vuelve a participar en el cálculo del diagnóstico.

Generación de la política (RAG)

Una vez que el motor identifica las brechas, el sistema necesita encontrar la normativa correcta antes de redactar cualquier cosa. El problema es que las preguntas del árbol están escritas en lenguaje simple

para PYMEs, pero los estándares usan jerga técnica y legal. Una sola forma de buscar no basta: si se busca solo por palabras exactas, no encuentra nada cuando el usuario dice «datos de clientes» y la norma dice «tratamiento de datos personales del titular»; si se busca solo por significado, puede fallar cuando lo que se necesita es localizar un código de control específico como «A.8.4».

Por eso el sistema combina tres formas de búsqueda simultáneas: una que localiza directamente el control normativo que cada pregunta ya trae asociado en su mapeo, otra que busca por coincidencia de palabras y códigos exactos, y una tercera que busca por significado, aunque las palabras no coincidan. Las tres producen un ranking de resultados cada una, y esos tres rankings se combinan en uno solo mediante una fórmula que da más peso a los fragmentos que aparecen bien posicionados en varias búsquedas a la vez:

$$\text{score} = \sum \frac{1}{k + \text{posición en cada ranking}} \quad (2)$$

donde $k = 60$ es una constante que controla cuánto se penaliza a los resultados que aparecen más abajo en cada ranking. Un fragmento que aparece primero en las tres búsquedas obtiene un puntaje mucho mayor que uno que solo aparece en una. El conjunto fusionado se reordena una última vez evaluando cada fragmento junto con la consulta original, para quedarse con los más relevantes.

Si, después de todo este proceso, ningún fragmento supera un umbral mínimo de relevancia (0.35 sobre 1.0), el sistema no genera nada. Prefiere abstenerse antes que redactar una política sin respaldo normativo suficiente.

Para evitar que el modelo de lenguaje invente contenido al redactar la política, el sistema aplica tres filtros sucesivos, cada uno cubriendo lo que el anterior podría dejar pasar. El primero es la abstención por falta de fuentes que se acaba de describir. El segundo verifica automáticamente que toda cita normativa mencionada en la política (artículos, controles ISO, leyes) exista realmente en la base normativa, sin necesidad de otro modelo, como quien busca un código en una lista. El tercero usa un segundo modelo de lenguaje para comprobar que cada afirmación de la política esté realmente respaldada por las fuentes citadas, no solo que la cita exista, sino que diga lo que la política afirma que dice, marcando cada afirmación como respaldada, no respaldada o parcial.

4. RESULTADOS

Esta sección presenta los resultados de diseño del instrumento y de su ilustración mediante un caso teórico, para lo cual conviene fijar el alcance desde el inicio. El entregable es una arquitectura de software acompañada del repositorio con el código fuente trabajado, no un estudio empírico. Por restricciones de tiempo no fue posible probar la herramienta en campo, de modo que no existen datos reales de PyMEs y el sistema se presenta como un modelo teórico cuyo funcionamiento se ilustra con un caso construido. En consecuencia, los resultados no son de desempeño estadístico sino de cobertura, consistencia estructural y trazabilidad. La sección 4.6 argumenta la robustez de las afirmaciones en esos términos y delimita lo que el diseño no permite afirmar.

4.1. Estructura del instrumento

El instrumento quedó conformado por un árbol adaptativo de 15 preguntas raíz (q1 a q13 más q3b y variantes de rama), precedido por una pregunta de bifurcación que clasifica a la organización en una de tres rutas según su uso de IA: desarrollo de productos o servicios para terceros, automatización de procesos internos, o ambos. Sobre las preguntas raíz se despliegan subpreguntas condicionadas que solo aparecen cuando la respuesta previa lo justifica. Por ejemplo, las preguntas de documentación e investigación de incidentes (q12a, q12b) solo se activan si la organización declara haber tenido un incidente (q12). El árbol totaliza 34 nodos, de los cuales 30 son puntuables, es decir contribuyen al diagnóstico mediante un mapeo a control y un puntaje por opción, y 4 son campos de justificación en texto libre que enriquecen el diagnóstico sin puntuar.

Observación: el número de preguntas que efectivamente ve una organización es menor que el total de nodos, porque la mayoría son condicionales. *Interpretación:* la profundidad del cuestionario es proporcional a la exposición declarada. Una organización con pocos sistemas y sin incidentes recorre un camino corto, mientras que una con sistemas de alto riesgo activa las ramas de evaluación de impacto y supervisión humana. Esto operacionaliza el diseño consciente de recursos, central para una PyME sin equipo dedicado.

4.2. Cobertura de las cuatro fuentes

Cada nodo puntuable mapea de forma simultánea a una o más funciones del NIST AI RMF, a uno o más controles del Anexo A de ISO/IEC 42001, y a uno o más de los cinco principios éticos de Floridi (2024) que consolidan los marcos de la OCDE y la UNESCO (beneficencia, no maleficencia, autonomía, justicia, explicabilidad). La regla de diseño, según la cual ningún nodo puntuable entra sin al menos un control de destino, se verificó sobre los 30 nodos sin excepciones. La Tabla 2 resume la distribución.

Tabla 2: Cobertura del instrumento sobre las cuatro fuentes. Cada cifra indica el número de nodos puntuables (de 30) que mapean a esa categoría. Un nodo puede mapear a más de una, por lo que las columnas no suman 30.

Fuente	Categoría	Nodos que la cubren
NIST AI RMF	GOVERN	12
	MAP	7
	MEASURE	10
	MANAGE	7
ISO 42001 (Anexo A)	12 controles distintos (A.2.2 a A.10.3)	30 nodos sobre 12 controles
Principios éticos (Floridi, 2024) recogiendo OCDE/UNESCO	Explicabilidad	14
	Autonomía	9
	No maleficencia	8
	Justicia	6
	Beneficencia	2

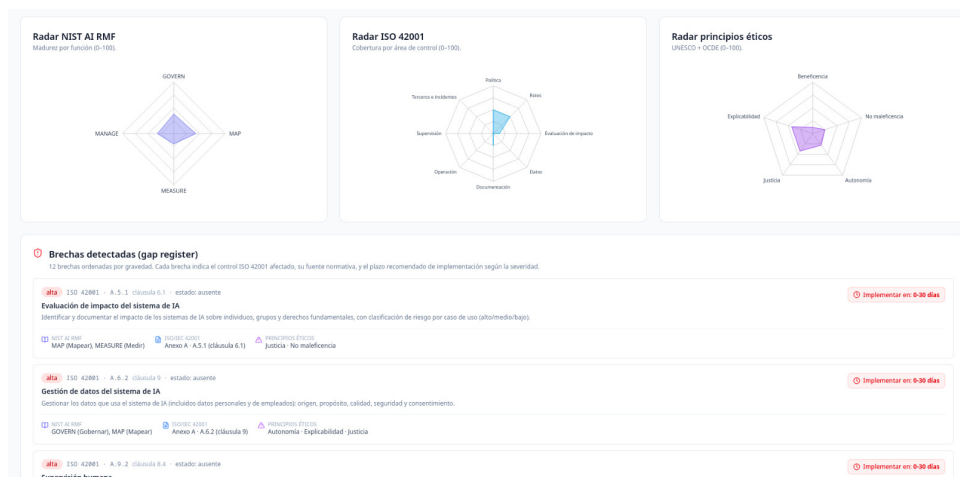


Figura 1: Cobertura del instrumento sobre NIST AI RMF, ISO/IEC 42001 y principios éticos.

Observación: las cuatro funciones del NIST quedan cubiertas, con mayor densidad en GOVERN y MEASURE. Entre los principios éticos, explicabilidad es transversal (14 nodos) y beneficencia es el menos representado (2 nodos). *Interpretación:* la concentración en explicabilidad es coherente con su rol de principio habilitante de los otros cuatro, según el marco de Floridi adoptado en el diseño. La baja

densidad de beneficencia es una característica observable del instrumento, no necesariamente un defecto. La beneficencia, entendida como que el bien sea real y verificable, se evalúa mejor caso por caso que mediante una pregunta de autodiagnóstico, y la sección 5 la retoma como limitación.

4.3. Mecánica de puntaje y contraste percepción contra diagnóstico

Cada opción de respuesta lleva un puntaje normalizado entre 0,0 y 1,0, y cada nodo lleva tres pesos independientes (ético, ISO, NIST) que permiten proyectar el mismo conjunto de respuestas sobre los tres ejes de salida sin recalculer el cuestionario. Las opciones marcadas como «No aplica» reciben puntaje nulo y se excluyen del promedio, de modo que un control irrelevante para el contexto de la organización no la penaliza. Una decisión de diseño deliberada se observa en la pregunta de incidentes (q12): tanto declarar que hubo un incidente como declarar que no lo hubo reciben puntaje pleno, mientras que «No sé» recibe puntaje parcial. El reconocimiento de un incidente se trata como señal de madurez en la detección, no como falla. Lo que penaliza el instrumento es la ausencia de visibilidad, no la existencia del incidente.

El instrumento incorpora además una pregunta de autoevaluación Likert (q13), en la que la organización califica de 1 a 5 su propio nivel de gobernanza antes de ver el resultado del sistema. *Observación:* esto produce dos puntajes comparables, el percibido y el calculado, sobre la misma dimensión. *Interpretación:* la brecha entre ambos es, en sí misma, un hallazgo diagnóstico. Una distancia amplia entre una autoevaluación alta y un puntaje calculado bajo es el indicador más directo del problema que el sistema busca atacar, que es la confianza no fundamentada en evidencia.

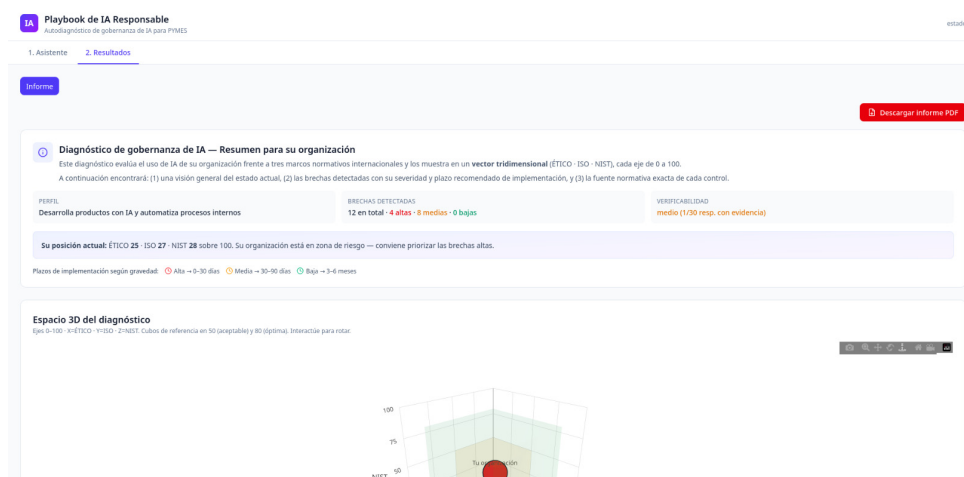


Figura 2: Resumen del diagnóstico generado por el sistema para el caso teórico.

4.4. Despliegue de las tres capas a partir de las respuestas

Al completar el cuestionario, el sistema procesa el conjunto de respuestas y genera de forma encadenada las tres capas de salida, que se presentan a la organización en secuencia. La Tabla 3 resume qué produce cada capa y de qué insumo se deriva.

Tabla 3: Salidas de cada capa y su insumo. Las tres se generan a partir del mismo conjunto de respuestas, y cada capa consume la anterior.

Capa	Salidas	Insumo
1, Diagnóstico	Perfil y alcance, inventario de casos de uso con su nivel de riesgo, radar de madurez por función NIST, cobertura ISO 42001, registro de brechas priorizado, indicador de verificabilidad por respuesta	Respuestas, puntajes y mapeo a control
2, Recomendaciones	Una recomendación por brecha con referencia cruzada NIST/ISO, priorización riesgo contra esfuerzo, justificación normativa y ética, rol responsable sugerido, secuenciación 0 a 30 a 90 días, criterio de cierre verificable	Registro de brechas de la Capa 1
3, Ejecución	Política de IA generada con LLM, plan de implementación con responsable y fecha, constancia verificable	Recomendaciones de la Capa 2

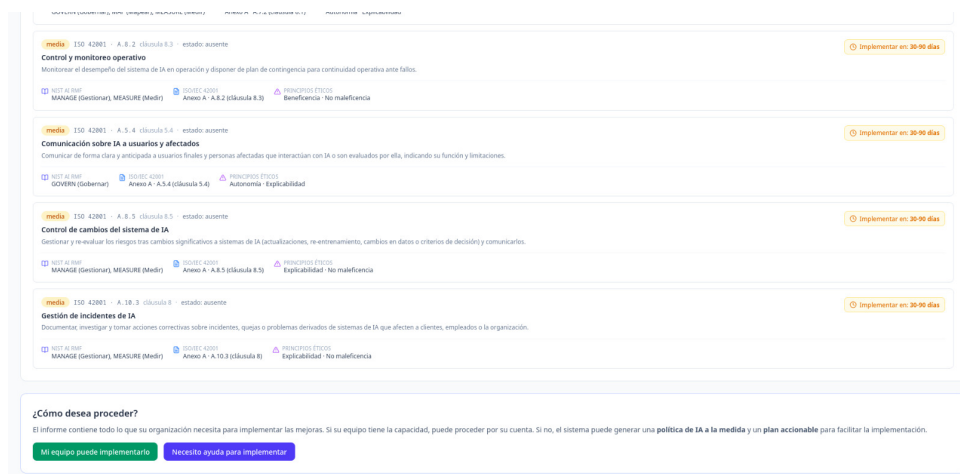


Figura 3: Registro de brechas priorizado generado por la Capa 1.

Observación: ninguna capa se genera de forma aislada, ya que cada una toma como entrada la salida de la anterior. *Interpretación:* el despliegue secuencial es lo que convierte el cuestionario en un sistema de extremo a extremo y no en un test que arroja un puntaje. La Capa 3 materializa la diferencia operativa central del enfoque, que es producir el artefacto de cierre y no solo recomendarlo, y la sección 4.5 lo ilustra siguiendo una brecha individual a través de las tres. Las salidas descritas corresponden a lo que la arquitectura específica y a lo que el código del repositorio implementa, no a salidas obtenidas de organizaciones reales.

4.5. Recorrido del caso teórico sobre las tres capas

Para ilustrar el funcionamiento de extremo a extremo se construyó un caso teórico, dado que no se dispone de datos reales de PyMES. El caso corresponde a una organización ficticia de la ruta «Ambos», elegida porque ejercita el árbol completo. La Figura 4 ubica el caso en el espacio tridimensional del diagnóstico, compuesto por los ejes ético, ISO y NIST. La Tabla 4 reconstruye la cadena de trazabilidad de una brecha a lo largo de las tres capas.

Tabla 4: Cadena de trazabilidad de una brecha a través de las tres capas, para el caso teórico. Ilustra la propiedad de extremo a extremo, en la que cada salida conserva el identificador de la anterior.

Eslabón	Contenido del caso teórico
Respuesta (Capa 1)	q4 con valor «alto» en un sistema y q7 (supervisión humana) baja, es decir un sistema de alto riesgo con revisión humana débil
Brecha y control	Falta de supervisión sobre sistema de alto riesgo. NIST MEASURE y GOVERN, ISO A.7.2, principios de autonomía y no maleficencia
Recomendación (Capa 2)	Instaurar revisión humana obligatoria antes de decisiones de alto impacto. Prioridad alta, esfuerzo medio, responsable la Dirección
Artefacto (Capa 3)	Política de IA generada con LLM que incluye la obligación de supervisión humana antes de decisiones de alto impacto, ajustada al sector y a la ruta de la organización, acompañada del plan de implementación con responsable y fecha
Criterio de cierre	Existencia del procedimiento documentado y de registro de revisiones. El estado de la respuesta pasa a «verificado» al adjuntar evidencia

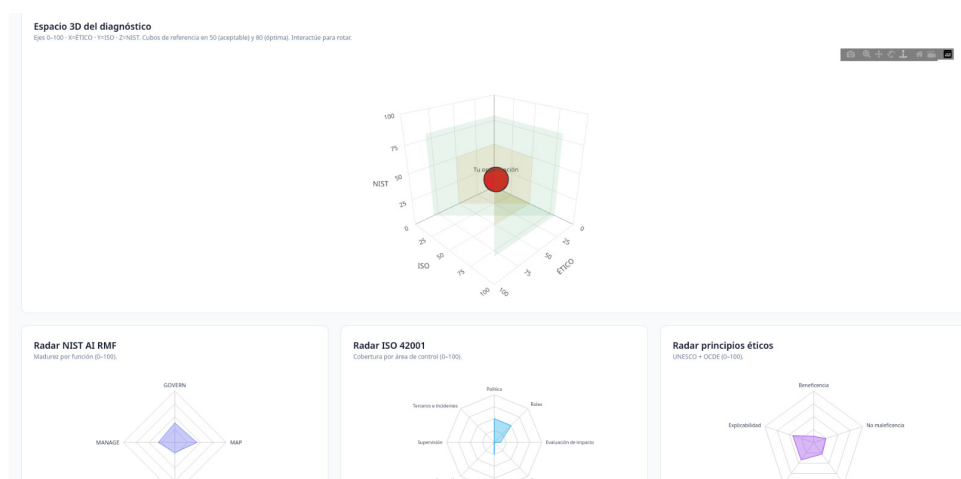


Figura 4: Ubicación del caso teórico en el espacio tridimensional ético, ISO y NIST.

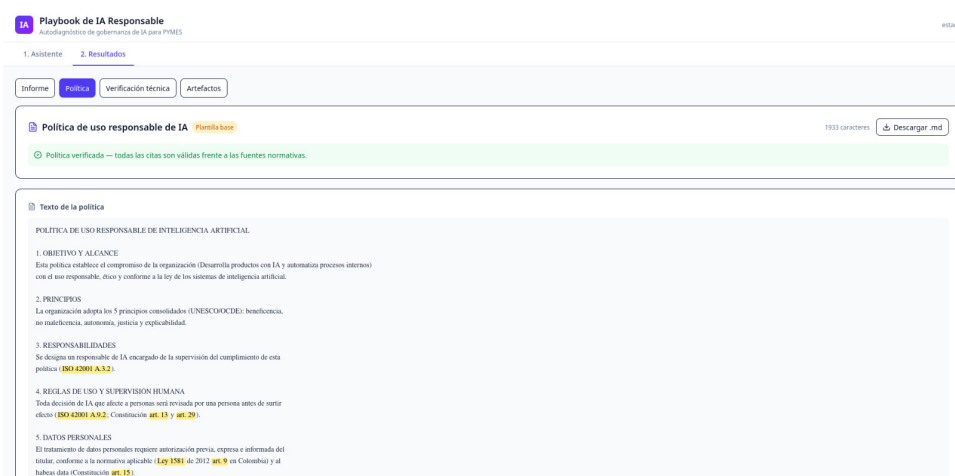


Figura 5: Artefacto de política generado a partir del diagnóstico y las brechas del caso teórico.

Observación: a partir de un único conjunto de respuestas, el recorrido produce un perfil de madurez (Capa 1), una recomendación priorizada con responsable y esfuerzo (Capa 2) y un artefacto de cierre con

su criterio de verificación (Capa 3). *Interpretación:* la cadena ilustra que el sistema no se detiene en el diagnóstico ni en la recomendación, sino que genera el artefacto, y que cada salida es rastreable hasta la respuesta que la originó. Al tratarse de un caso construido, este recorrido muestra el comportamiento previsto del modelo, no un resultado medido en una organización real.

4.6. Robustez de las afirmaciones

Las afirmaciones de este trabajo no son de desempeño estadístico. No hay muestra, grupo de comparación ni datos de campo, por lo que las nociones de tamaño muestral y significancia no aplican, y el trabajo se presenta como un modelo teórico. La validación disponible descansa en dos elementos: las propiedades estructurales verificables del instrumento y el código del repositorio que implementa la lógica descrita. La robustez se argumenta en tres planos.

Primero, completitud de la cobertura. La regla de que ningún nodo puntuable existe sin un control de destino se verificó sobre los 30 nodos, y la regla simétrica de la Capa 2, según la cual ninguna brecha genera recomendación sin un control de cierre, cierra el circuito en ambos sentidos. Es una propiedad del diseño, comprobable por inspección del cuestionario y del código, no sujeta a variación muestral.

Segundo, estabilidad del puntaje frente a sensibilidad del enrutamiento. El sistema es deliberadamente sensible a dos entradas: la respuesta de bifurcación, que altera el subconjunto de preguntas aplicables, y la clasificación de riesgo por sistema (q4), que activa las ramas de impacto y supervisión. Esa sensibilidad es el comportamiento buscado. En cambio, dado un mismo conjunto de respuestas, el puntaje de madurez es determinista, ya que descansa sobre un mapeo fijo de nodos a funciones y pesos, de modo que la misma entrada produce siempre el mismo radar. La variabilidad está en el camino recorrido, no en el cálculo.

Tercero, control del dato faltante. Las opciones «No sé» reciben puntaje cero y se tratan como brecha, mientras que «No aplica» se excluye del promedio. *Interpretación:* esta distinción evita dos sesgos opuestos, que la ignorancia se premie como cumplimiento y que la irrelevancia se castigue como deficiencia.

Limitaciones que acotan estas afirmaciones: el mapeo de nodos a controles y los pesos fueron asignados por el equipo y no se sometieron a verificación independiente de concordancia entre evaluadores; el instrumento no se probó en campo con PYMES reales, por lo que no existen datos de uso; y el caso presentado es teórico y único, de modo que ilustra el funcionamiento previsto, pero no establece validez externa. Estas limitaciones se desarrollan en la sección 5.

5. DISCUSIÓN Y LIMITACIONES

Dentro de las limitaciones identificadas en la construcción del SaaS, identificamos cuatro limitaciones generadas por diversos factores, entre ellos por tiempo y condiciones preexistentes en la región:

1. Heterogeneidad normativa y opción metodológica por un marco principalista supranacional. Una primera limitación está relacionada con la fragmentación regulatoria del entorno latinoamericano. A excepción de Perú, que cuenta con una ley específica de inteligencia artificial, la mayoría de las jurisdicciones de ALC carecen de un marco legal integral y vinculante en la materia, configurando un mosaico normativo disperso y de distinto grado de madurez. Ante esta ausencia, y dadas las restricciones de tiempo del ejercicio, se optó por anclar la arquitectura ética de la herramienta exclusivamente en las recomendaciones de la UNESCO (2021) y la OCDE (2019), ampliamente aceptadas y suscritas por los países de la región, en lugar de comparar e incluir uno a uno los distintos marcos normativos nacionales.

2. Riesgos inherentes al sustrato tecnológico: sesgos y alucinaciones. Una segunda limitación deriva de la naturaleza misma de la tecnología sobre la que se construye el SaaS. Al apoyarse en modelos de lenguaje de gran escala para operar como motor de traducción simplificada y para la generación de políticas y planes de acción, la herramienta no está exenta de los riesgos estructurales asociados a estos sistemas, particularmente la presencia de sesgos y la producción de alucinaciones. La mitigación de esta limitación exige salvaguardas adicionales, tales como la verificación humana de los planes de acción

emitidos y la trazabilidad de las recomendaciones respecto del NIST AI RMF y la norma ISO/IEC 42001 que las sustentan.

3. Idoneidad y fiabilidad de la información de entrada. Una tercera limitación se relaciona con la calidad de los insumos que el usuario aporta al sistema. La utilidad del diagnóstico depende de manera directa de la veracidad y la precisión de las respuestas que la organización registra; sin embargo, dada la elevada informalidad administrativa que caracteriza al ecosistema de MiPymes de la región, no puede presuponerse que quien responde cuenta con la alfabetización digital o conocimiento técnico suficiente para diligenciar los cuestionarios de forma confiable.

4. Relevancia diferencial de los reactivos según el sector empresarial. Una cuarta limitación obedece a la tensión entre la estandarización del instrumento y la diversidad de los contextos a los que se aplica. El cuestionario fue concebido como una arquitectura general orientada al universo amplio de las MiPymes, por lo que organizaciones de sectores específicos pueden requerir concepto o aproximaciones especializadas que no se encuentran contemplados y, a la inversa, encontrar entre las preguntas planteadas algunas de escasa pertinencia para su operación particular. Sin una prueba de campo, abarcar estas limitaciones se vuelve imposible.

Trabajo a futuro

El trabajo futuro se estructurará en dos frentes. Por un lado, el fortalecimiento normativo de la herramienta mediante la definición de una política propia de gobernanza de IA y el desarrollo de descargos de responsabilidad (legal disclaimers) robustos para delimitar jurídicamente los diagnósticos y proteger los datos procesados. Por el otro, el escalamiento aplicado a través del despliegue de programas piloto con MiPymes para evaluar la usabilidad del SaaS y la efectividad de los planes a 30 y 90 días en entornos informales. Finalmente, se buscaría consolidar alianzas estratégicas con Cámaras de Comercio, incubadoras de empresas y Secretarías de Desarrollo Económico para masificar el acceso a la plataforma, promoviendo como un estándar práctico de cumplimiento ético y competitividad, inicialmente en Colombia y a futuro en la región.

6. CONCLUSIÓN

Este trabajo presentó el diseño de una herramienta tipo SaaS orientada a democratizar la gobernanza ética de la inteligencia artificial en las MiPymes de América Latina y el Caribe. El instrumento traduce la densidad técnica del NIST AI RMF y la norma ISO/IEC 42001 a un árbol adaptativo de preguntas redactadas en lenguaje llano, y despliega tres capas encadenadas que van del diagnóstico a la generación de artefactos de cierre verificables, sin detenerse en la recomendación abstracta. La principal contribución no es el diagnóstico en sí mismo, sino la capacidad del sistema para producir los documentos que una PYME necesita para actuar sobre sus brechas, cerrando el vacío de ejecución que caracteriza a las herramientas existentes en la región.

Al tratarse de un modelo teórico sin validación en campo, el alcance de las afirmaciones se limita a las propiedades estructurales verificables del instrumento: cobertura completa de las cuatro fuentes normativas, determinismo del puntaje frente a un mismo conjunto de respuestas, y trazabilidad de extremo a extremo desde la respuesta hasta el artefacto generado. La validación empírica mediante pilotos con MiPymes reales es el paso necesario para determinar si estas propiedades de diseño se traducen en utilidad operativa efectiva. Con todo, el ejercicio demuestra que es técnicamente viable construir un puente entre los estándares internacionales de gobernanza de IA y las capacidades reales de las empresas más pequeñas del Sur Global, y que ese puente no requiere simplificar los principios éticos sino hacerlos ejecutables.

CÓDIGO Y DATOS

Repositorio de código: <https://github.com/DavidDaza2906/ai-saas-playbook>

Datos: el corpus normativo tiene una arquitectura de dos capas (incluido en el repositorio, `app/backend/data/corpus/`). El núcleo universal, activo para cualquier país, contiene NIST AI RMF 1.0, ISO/IEC 42001:2023 y los 5 principios éticos consolidados (UNESCO + OCDE; Floridi, 2024), en fragmentos etiquetados con metadatos que alimentan el vector de diagnóstico y la recuperación del RAG. El paquete por país se carga según el país declarado; para la demostración con Colombia incluye la Constitución Política y la Ley 1581 de 2012, utilizados para la justificación normativa y el contexto legal específico. El sistema es portable: agregar otro país solo requiere añadir su Constitución y legislación estatutaria; el núcleo universal no se modifica.

El repositorio incluye también el conjunto de evaluación de referencia (`golden_set.json`) con 10 consultas anotadas con controles ISO y referencias legales esperadas; el árbol de preguntas (`questions.json`) con 15 preguntas base más 19 subpreguntas con mapeos NIST/ISO/principios y pesos congelados; y los resultados de evaluación (`eval_result.json`) con métricas RAGAS sobre el conjunto de referencia (`recall@6=0.85`, precisión de citas=1.0, fidelidad=0.884).

Otros artefactos: demostración en vivo en <http://localhost:5173> (instrucciones de despliegue en el README del repositorio). Informe técnico completo (`INFORME_TECNICO.md` en la raíz del repositorio) con arquitectura, motor determinista, pipeline RAG, defensa anti-alucinación en 3 capas y métricas. Guion de demostración (`app/SHOWCASE.md` en el repositorio) con recorrido paso a paso para jueces. Reporte PDF descargable generado por el sistema (WeasyPrint) con diagnóstico, registro de brechas, recomendaciones, plan de acción y constancia verificable.

REFERENCIAS

Algorithm Watch. (2025). *Automated Decision-Making Systems and Fundamental Rights: The Horizontal Effect in the Private Sector*. Algorithm Watch.

Banco de Desarrollo de América Latina (CAF). (2024). *Heterogeneidad productiva y brechas de competitividad en las MiPymes de América Latina: Reporte de Economía y Desarrollo*. CAF.

Banco Interamericano de Desarrollo (BID). (2025). *Autoevaluación ética de IA para actores del ecosistema emprendedor: Guía de aplicación práctica*. fAIR LAC. <https://fairlac.iadb.org/en/emprendimiento/autoevaluacion>

Bhalla, A., & Madrekar, S. (2025). The Global State of MSMEs: Driving Economic Resilience and Employment in Developing Economies. *Journal of Innovation and Entrepreneurship*, 14(1), 45–62.

Borowski, M. (2020). La Drittwirkung ante el trasfondo de la transformación de los derechos morales en derechos fundamentales. *Revista Derecho del Estado*, núm. 45. <https://www.redalyc.org/journal/3376/337662821001/337662821001.pdf>

Cámara Colombiana de Informática y Telecomunicaciones (CCIT). (2025). *Hoja de ruta para la Plataforma Regional del sector privado sobre inteligencia artificial (IA)*. CCIT.

Centro Nacional de Inteligencia Artificial (CENIA). (2025). *Índice Latinoamericano de Inteligencia Artificial (ILIA) 2025*. CENIA.

Comisión Económica para América Latina y el Caribe (CEPAL). (2025). *Uso de soluciones de inteligencia artificial en América Latina y el Caribe 2025: Oportunidades y desafíos para la transformación productiva*. Naciones Unidas.

Floridi, L. (2024). *Ética de la Inteligencia Artificial*. Herder.

Instituto Nacional de Estándares y Tecnología. (2023). *Marco de Gestión de Riesgos de Inteligencia Artificial (AI RMF 1.0)* (Informe Técnico NIST AI 100-1). Departamento de Comercio de los Estados Unidos. <https://doi.org/10.6028/NIST.IR.8474>

OCDE/CAF. (2019). *Índice de Políticas para PyMEs: América Latina y el Caribe 2019*. OECD Publishing.

Organización de las Naciones Unidas para la Educación, la Ciencia y la Cultura (UNESCO). (2021). *Recomendación sobre la Ética de la Inteligencia Artificial*. UNESCO.

Organización Internacional de Normalización & Comisión Electrotécnica Internacional. (2023). *Tecnología de la información — Inteligencia artificial — Sistema de gestión* (Norma ISO/IEC 42001:2023). <https://www.iso.org/standard/81230.html>

Organización Internacional del Trabajo (OIT). (2022). *Panorama Laboral 2022: América Latina y el Caribe*. Oficina Regional de la OIT.

Organización para la Cooperación y el Desarrollo Económicos (OCDE). (2019). *Principios de la OCDE sobre la Inteligencia Artificial*. OCDE.

APÉNDICE

Se introduce la definición robusta de los cinco principios éticos que alimentarán el SaaS, con las respectivas responsabilidades derivadas a usuarios de herramientas de IA:

Beneficencia. El uso de un sistema de IA debe producir un bien real y verificable, no una mera promesa de beneficio (no sólo para quien lo emplea, sino para la comunidad afectada) con especial atención a grupos vulnerables. Corresponde al proveedor precisar para qué sirve la herramienta, en qué condiciones y a quién beneficia; y corresponde a quien la emplea comprobar que ese bien es efectivo en su contexto, absteniéndose de usarla cuando no lo sea. *Responsabilidades del usuario:* juzgar si lo que la herramienta entrega es efectivamente útil, correcto y apropiado para el contexto real de aplicación, y descartar su uso cuando no lo sea.

No maleficencia. Los daños deben anticiparse y prevenirse antes que repararse. Debe haber una vigilancia anticipatoria sobre daños individuales, colectivos, intergeneracionales y sistémicos. Corresponde al proveedor evaluar los riesgos antes del despliegue, en proporción a su gravedad posible, y acreditar las salvaguardas correspondientes; y corresponde a quien la emplea conocer la herramienta lo suficiente para advertir cuándo su resultado es erróneo o sesgado y para no introducir en ella datos privados que no deban utilizarse. *Responsabilidades del usuario:* conocer suficientemente la herramienta para reconocer cuándo su resultado es erróneo, sesgado o inapropiado, y abstenerse de introducir en ella datos privados o que no deban utilizarse.

Autonomía. Las decisiones que afectan a las personas deben permanecer bajo control humano: las personas que puedan verse afectadas deben poder conservar el control sobre las decisiones que les atañen y la capacidad de un juicio informado sobre cuándo y cómo delegar en el sistema. *Responsabilidades del usuario:* ejercer ese control de forma efectiva, decidiendo de manera informada el grado de delegación y reservando la decisión humana en los usos de mayor riesgo.

Justicia. Los beneficios y las cargas del uso de la IA deben distribuirse de forma equitativa, sin discriminación por sexo, raza, religión, discapacidad, edad u orientación sexual. La responsabilidad sea trazable: corresponde al proveedor acreditar que el sistema no discrimina y cómo controla los sesgos; y corresponde a quien la emplea documentar su uso, de modo que pueda reconstruirse qué ocurrió y a quién corresponde responder. *Responsabilidades del usuario:* documentar y hacer transparente el propio uso, de modo que la cadena de responsabilidad sea trazable.

Explicabilidad. Condición que habilita a los otros cuatro principios y se articula en dos preguntas: la inteligibilidad (¿cómo funciona?) y la rendición de cuentas (¿quién responde por cómo funciona?). Sin ella, la beneficencia no distingue el bien real de un proxy, la no maleficencia no traza la causa del daño, la autonomía no funda una delegación informada y la justicia se vuelve inauditable. *Responsabilidades del usuario:* una obligación que excede lo individual y se vuelve institucional: crear y sostener las condiciones para que quienes usan la herramienta puedan comprenderla, cuestionarla y emplearla responsablemente.

DECLARACIÓN DE USO DE LLMs

Gemini fue utilizado para la redacción y sintaxis del documento. Fue corregido y estructurado de manera manual.

NotebookLM para la administración y cotejo de citas de referencias utilizadas.

Claude fue utilizado para la generación de prompt de contexto final (verificado de manera manual) que recoge el brainstorm e idea final del producto, construcción de preguntas basados en los estándares ISO 42001 y NIST AI RMF; creación del SaaS y creación del RAG.