Ética de la inteligencia artificial, marcos globales y sus limitaciones

Duban Fernando Cuervo Martines[1]
Harrizon Alexander Soler Galindo[2]

Resumen

Actualmente, la gobernanza mundial está experimentando un desafío inusual: la evolución creciente de los Modelos Lingüísticos de Gran Escala (MLGE). Los logros de estos sistemas han avanzado desde una concepción teórica hasta convertirse en factores vitales del ámbito político global e institucional. Actualmente, la mayoría de las empresas más prestigiosas utilizan estas herramientas. La reglamentación relacionada con valores y principios, basada en nuestra legislación existente, resulta difícil de aplicar. Este estudio desarrolla una evaluación crítica de la literatura relacionada con este problema, en conformidad con las directrices PRISMA, para analizar cómo corporaciones como OpenAI, Anthropic, Meta, Google y la empresa de inteligencia Intel controlan su comportamiento interno y cumplen con sus responsabilidades ante las otras empresas en sus plataformas comerciales.

La recolección de datos se efectuó utilizando las bases de datos Scopus e IEEE Xplore, utilizando diez diferentes métodos. Después de un análisis inicial de 959 documentos, fueron eliminados 871 de los cuales quedaron 88 trabajos. Según este estudio final, los informes éticos de las empresas difieren drásticamente de su comportamiento real. Al respecto, el Artículo Europeo sobre el Derecho de la Inteligencia Artificial es una regulación que carece de defectos estructurales cuando intenta regular prácticas intrínsecas a la opacidad e impredecibilidad. En cuanto al Sesgo literario, se observa que la información disponible se basa predominantemente en la parte norte del mundo, lo que resulta inquietante porque la tecnología está tomando impulso en América Latina, donde existe gran oportunidad para su implementación. El artículo concluye con recomendaciones prácticas para robustecer la ética corporativa en el desarrollo de LLM, aterrizando el análisis en las realidades regulatorias de Colombia y la región.

Palabras clave: ética en inteligencia artificial, modelos de lenguaje LLM, gobernanza de tecnología, empresas y responsabilidad, revisión sistemática con PRISMA, leyes de inteligencia artificial.

AI Ethics, Global Frameworks and Their Limitations

Abstract

International governance is undergoing an extraordinary shift as large language models (LLMs) transition from simple theoretical concepts to crucial geopolitical and regulatory players. At present, a small group of dominant corporations leads the global rollout, highlighting the shortcomings of current regulatory systems in addressing underlying ethical concerns. Utilizing a systematic literature review under the PRISMA protocol, this study provides a critical examination

[1] Duban Fernando Cuervo Martinez, Ingeniería de Sistemas, Universidad Santo Tomás, Tunja, Colombia. Duban.cuervo@usantotomas.edu.co

[2] Harrizon Alexander Soler Galindo, Facultad de Ingeniería de Sistemas, Universidad Santo Tomás, Tunja, Colombia. Harrizon.soler@usantotomas.edu.co.

of how dominant firms OpenAI, Anthropic, Meta, Google, and Microsoft manage transparency, internal governance, and corporate accountability within their commercial systems. Data collection spanned the specialized Scopus and IEEE Xplore databases across eight structured search strategies; after screening 959 initial records, a definitive corpus of 69 articles published between 2017 and 2026 was curated. The primary findings expose a profound disconnect between corporate ethical rhetoric and actual operational practices. Moreover, ambitious legal initiatives such as the European Union AI Act exhibit structural blind spots when attempting to oversee models that are inherently opaque and probabilistic. A critical bias also emerges in current literature: the systematic marginalization of the Global South, a deeply concerning trend given the swift, unmonitored integration of these tools across Latin America. This paper concludes with actionable recommendations to anchor corporate ethics firmly within LLM development, tailoring the analysis to the distinct regulatory landscapes of Colombia and the broader region.

Keywords: AI ethics frameworks, large language models, AI governance, corporate accountability, PRISMA systematic review, artificial intelligence regulation.

1. Introducción

Hoy en día las inteligencias artificiales generativas como GPT-4, Gemini, Claude, Llama o Copilot están metidas en todos lados. Prácticamente todo el mundo las está usando y en muy poco tiempo cambiaron por completo cómo se trabaja, cómo nos comunicamos y la forma en que se toman las decisiones en las empresas grandes y chicas [36].

La realidad es que este cambio no fue ni gradual ni intencional. Ocurrió de forma abrupta, y la economía global integró estas herramientas a un ritmo asombroso, dejando obsoletos los métodos de trabajo tradicionales en cuestión de meses. Pero claro, como todo lo que crece rápido, esto trajo encima un montón de problemas éticos, legales y corporativos; son dilemas críticos que las leyes actuales jamás anticiparon y por eso no sirven para controlarlos bien [37].

Reflexionando sobre ello, el debate en torno a la ética en la inteligencia artificial no es reciente; ha cobrado cada vez más importancia desde principios de la década de 2010.En ese tiempo, organizaciones internacionales como la OCDE, la UNESCO y la Comisión Europea sacaron directrices generales para intentar que los algoritmos fueran transparentes, auditables y respetaran los derechos humanos [3]. El problema con las fórmulas convencionales es que fueron creadas para sistemas obsoletos que se regían por directrices rígidas. Resultan ineficaces para las complejidades generativas a las que nos enfrentamos hoy en día. Aquellas instrucciones anteriores presuponían programas lógicos, en lugar de entidades que aprenden únicamente mediante estadísticas, analizando grandes cantidades de texto generado por humanos en internet. Por este motivo, cuando surgieron los grandes modelos de lenguaje (LLM, por sus siglas en inglés) desarrollados con vastos conjuntos de datos y miles de millones de parámetros, que demostraban la capacidad de escribir de forma similar a los humanos, se hizo evidente que la falta de claridad en los algoritmos, los sesgos presentes en los datos y la concentración de poder en un pequeño número de empresas plantean desafíos que las regulaciones tradicionales no podían abordar adecuadamente [28].

Existe una gran discrepancia entre lo que las empresas expresan en sus discursos atractivos sobre la responsabilidad y lo que realmente llevan a cabo en el mundo real, y esto es un tema ampliamente abordado en la literatura científica contemporánea [47]. No se trata de un incidente ocasional; los investigadores consideran que es algo común y estructural en la industria tecnológica. Por ejemplo, OpenAI ha recibido bastante crítica por su negativa a revelar qué conjuntos de datos utiliza para entrenar sus modelos más recientes [6]. Meta fue sancionada con

una multa histórica de 1.200 millones de euros por compartir datos sin autorización, violando el Reglamento General de Protección de Datos (GDPR) [7], y Google ha enfrentado enormes multas antimonopolio que alcanzaron los 4.340 millones de euros [8]. Ante esta situación, resulta evidente que permitir que las empresas se autorregulen utilizando directrices éticas basadas en buenas intenciones no será suficiente para supervisar eficazmente los programas de máster en derecho (LLM).

Para empezar hay un problema de diseño, como ocurre en el caso de los primeros sistemas de IA que usaban reglas de lógica o aprendizaje supervisado, utilizando menos datos y mejor controlados. Las tecnologías modernas de IA emplean aprendizaje profundo e imitan al cerebro humano por medio de grandes conjuntos de datos como la web, los libros y código fuente. Esta cantidad de información produce problemas éticos: los prejuicios y sesgos no salen de una programación consciente, sino que surgen de las estadísticas del aprendizaje, lo cual es difícil de filtrar. La legislación actúa igualmente desafiante, pues las responsabilidades legales se dispersan entre quien proporciona el hardware, quién desarrolla el programa de software final y quien usa el software. De manera que actualmente las leyes no son aptas ni justas en resolver este conflicto.

Las grandes compañías como Google, Anthropic y OpenAI establecieron directivas internas propias, como límites del uso de la IA o innovaciones como una "Constitución de IA", para responder a las críticas. Aunque esto no es suficiente para convencer al resto, académicos y universidades no están completamente satisfechos con este camino. Whittaker [9] ha afirmado firmemente que estas directivas no van más allá que "un lavado de imagen ético", argumentando que redactan estos documentos atractivos para mejorar su reputación y evitar leyes gubernamentales restrictivas, en lugar de reformar realmente sus procesos habituales de investigación. El estudio sugiere que la falta de consistencia entre la publicidad de las empresas y sus prácticas cotidianas es bastante frecuente; sin embargo, algunos autores son más severos e indican que algunas compañías hacen realidades sus promesas [28].

El sistema legal es igualmente complejo, pues aunque la Unión Europea aprobó su histórica Ley de IA en 2024 la primera ley obligatoria de este tipo en el mundo, presenta deficiencias significativas en lo que respecta a la IA de propósito general. Las investigaciones demuestran que la categorización de riesgos rígida e inmutable resulta ineficaz al tratar con modelos que exhiben un comportamiento impredecible y cambian [48]. El enfoque por niveles ignora el hecho de que el nivel de riesgo de un modelo depende del entorno en el que se emplea; además, la apertura de los datos de entrenamiento es prácticamente inexistente, y existen lagunas si el modelo se desarrolla fuera de Europa.

La situación es caótica fuera de Europa, como pasa en Estados Unidos donde no hay una legislación federal única para regular la IA; pero existen marcos voluntarios y normativas sectoriales, por otro lado China ha regulado estrictamente la IA generativa, con el registro de algoritmos ante el gobierno y la moderación de contenidos. En cuanto a las políticas de inteligencia artificial en América Latina y el Sur Global, por otro lado, están en pañales, y, casi siempre, se limitan a copiar marcos internacionales sin adaptarlos a sus problemas económicos o sociales concretos.

Todo esto reviste una urgencia geopolítica muy significativa para América Latina. A pesar de la rápida adopción de estas tecnologías por parte de hospitales, tribunales e instituciones académicas, no existen ni legislación local ni capacidades técnicas para examinar de forma independiente el funcionamiento interno de estos algoritmos. Como resultado de esta disparidad, dependemos cada vez más de corporaciones extranjeras. Corremos el riesgo de perder autonomía regulatoria, agravar la desigualdad ya existente e internalizar los prejuicios culturales de las

naciones ricas donde se concibieron inicialmente estos modelos si los adoptamos sin la debida autorización local.

Por eso, este trabajo busca revisar de forma crítica cómo manejan la ética las empresas tecnológicas líderes en el mundo, haciendo una revisión sistemática de literatura entre 2017 y 2026 usando las bases de datos Scopus e IEEE Xplore. Queremos comparar temas de transparencia, responsabilidad corporativa y gobernanza para ver qué vacíos hay y dar recomendaciones para mejorar esto en las empresas, pensando sobre todo en lo que esto implica para los entornos emergentes de América Latina.

El resto del documento se divide así: la sección 2 explica la metodología con el protocolo PRISMA y las búsquedas; la sección 3 muestra los resultados; la sección 4 es la discusión comparando los hallazgos con otra literatura; y la sección 5 trae las conclusiones y límites del estudio.

2. Metodología

Dado que no existe un repositorio único de normativas éticas corporativas sobre IA, el camino metodológico más adecuado para este tipo de pregunta es la revisión sistemática de literatura bajo el protocolo PRISMA [17; 69]. Se seleccionó esta estructura porque requiere que tengamos un registro transparente de cada rechazo y selección, lo que permite que cualquiera pueda duplicar el proceso o cuestionar los estándares de manera imparcial. No se trataba únicamente de compilar artículos, sino de examinar las tendencias, los conflictos y las verdaderas deficiencias que la academia ha ignorado al estudiar la ética en las organizaciones que brindan maestrías en derecho (LLM). Asimismo, el estudio tiene como objetivo entender los discursos normativos que las empresas tecnológicas más grandes afirman seguir y examinar la legitimidad de estas prácticas en comparación con la publicidad corporativa. Esto proporciona un fundamento firme para entender la manera en que Latinoamérica es capaz de hacer uso de estas enseñanzas globales y empezar a edificar su propia estrategia regulatoria, en vez de únicamente acatar sin cuestionar las regulaciones del Norte Global.

2.1. Bases de datos y estrategias de búsqueda

Las búsquedas se llevaron a cabo en dos bases de datos que se complementan entre sí: Debido a su cobertura multidisciplinar en ingeniería, derecho y ciencias sociales, Scopus; e IEEE Xplore; por su especialización en tecnología y ciencias de la computación. Se elaboraron ocho cadenas de búsqueda estructuradas (Q1-Q8) con el objetivo de recopilar la bibliografía acerca de marcos éticos, regulación de LLM e IA generativa, gobernanza y responsabilidad corporativa.

Tabla 1. Estrategias de búsqueda sistemática Cadenas de consulta y focos temáticos

N°	Base de datos	Cadena de búsqueda	Foco temático
Q1	Scopus / IEEE	TITLE-ABS-KEY ((“ethical framework*" OR "ethics framework*”) AND (“large language model*" OR "generative AI" OR "foundation model*”) AND (“governance" OR "accountability" OR "responsible AI”))	Marcos éticos y gobernanza para LLM
Q2	Scopus / IEEE	TITLE-ABS-KEY ((“corporate governance" OR "AI governance”) AND (“AI ethics" OR "artificial intelligence ethics" OR "responsible AI”) AND (“accountability" OR "transparency”) AND (“large	Gobernanza corporativa y responsabilidad en IA

N°	Base de datos	Cadena de búsqueda	Foco temático
		language model*" OR "generative AI" OR "foundation model*"))	
Q3	Scopus / IEEE	TITLE-ABS-KEY ("corporate social responsibility" AND ("large language model*" OR "generative AI" OR "artificial intelligence") AND ("ethics" OR "bias" OR "accountability" OR "transparency"))	RSE e IA generativa
Q4	Scopus / IEEE	TITLE-ABS-KEY ("training data" AND ("ethics" OR "ethical concern*" OR "ethical implication*") AND ("large language model*" OR "foundation model*" OR "generative AI") AND ("bias" OR "transparency" OR "consent" OR "copyright" OR "governance"))	Datos de entrenamiento: ética y sesgo
Q5	Scopus / IEEE	TITLE-ABS-KEY ("algorithmic bias" AND ("large language model*" OR "LLM" OR "generative AI") AND ("governance" OR "accountability" OR "regulation" OR "corporate") AND ("ethics" OR "fairness" OR "transparency"))	Sesgo algorítmico en LLM
Q6	Scopus / IEEE	TITLE-ABS-KEY ("algorithmic transparency" AND ("artificial intelligence" OR "machine learning" OR "large language model*") AND ("corporate" OR "big tech" OR "tech company*" OR "platform*") AND ("accountability" OR "ethics" OR "governance"))	Transparencia algorítmica y Big Tech
Q7	Scopus / IEEE	TITLE-ABS-KEY (("AI policy" OR "AI governance" OR "artificial intelligence governance") AND ("ethical framework*" OR "AI ethics") AND ("comparative" OR "cross-national" OR "cross-country" OR "comparative analysis") AND ("large language model*" OR "generative AI" OR "foundation model*" OR "artificial intelligence"))	Política de IA comparativa y marcos éticos
Q8	Scopus	(AI regulation" OR "AI policy" OR "AI governance") AND ("large language models" OR LLM OR "generative AI" OR "foundation models") AND ("Big Tech" OR OpenAI OR Google OR Meta OR Anthropic)	Regulación IA y empresas tecnológicas líderes

Fuente: elaboración propia.

Las búsquedas fueron realizadas durante el primer trimestre de 2026 y cubren publicaciones del período 2017-2026. La restricción temporal al año 2017 se justifica por ser el momento en que la literatura sobre LLM comienza a consolidarse a raíz de la publicación del trabajo seminal de Vaswani et al. [70] sobre la arquitectura Transformer (Transformador), que sienta las bases técnicas de los sistemas actuales. El límite superior de 2026 incluye publicaciones en prensa o de acceso anticipado (early access) disponibles al momento de la búsqueda.

2.2. Criterios de inclusión y exclusión

Se establecieron los siguientes criterios de inclusión: (a) artículos publicados entre 2017 y 2026 en revistas indexadas o memorias de conferencias arbitradas con revisión por pares, (b) trabajos que aborden explícitamente la ética de la IA, marcos éticos o regulatorios aplicados a

LLM o IA generativa, (c) estudios que mencionen al menos uno de los criterios de análisis, los cuales son transparencia, responsabilidad corporativa, gobernanza, sesgo algorítmico o regulación y (d) publicaciones disponibles en texto completo en las bases de datos consultadas.

Los criterios de exclusión contemplaron: artículos de opinión sin respaldo empírico o teórico sólido, documentos de divulgación sin proceso de revisión por pares, trabajos centrados exclusivamente en aspectos técnicos sin dimensión ética o regulatoria, artículos duplicados identificados en ambas bases de datos y publicaciones en idiomas distintos al español e inglés, dada la capacidad de revisión del investigador.

La aplicación de estos criterios fue realizada por el investigador en dos rondas. En la primera ronda (R1) se aplicaron los criterios sobre títulos y resúmenes, en la segunda ronda (V2) se revisó el texto completo de los artículos que superaron el primer filtro. Este proceso garantizó la consistencia y reproducibilidad de las decisiones de inclusión.

2.3. *Proceso de selección PRISMA*

El embudo de selección funcionó de la siguiente manera. Las ocho cadenas de búsqueda arrojaron 959 registros. De entrada, herramientas como Rayyan.ai nos permitieron detectar y eliminar duplicados un paso más tedioso de lo esperado, pues algunos artículos aparecían hasta en tres entradas distintas con variaciones de título.

Tabla 2. Proceso de selección PRISMA Revisión sistemática

Fase PRISMA	Actividad	Resultado
Identificación	Búsqueda en Scopus e IEEE Xplore (8 queries: Q1-Q8)	959 registros
Screening 1	Revisión de criterios mediante rayyan.ai	138 artículos para revisión exhaustiva
Cribado 1	Revisión de títulos, palabras claves	86 artículos aprobados
Cribado 2	Revisión exhaustiva de documentos	69 artículos aprobados
Inclusión total	Corpus final para análisis	69 artículos incluidos

Fuente: elaboración propia a partir del proceso de búsqueda y selección sistemática.

Tras el filtro inicial de títulos y resúmenes logramos rescatar 138 textos candidatos, pero la lectura a fondo de esos documentos redujo el corpus definitivo a solo 69 artículos. Incluir el 7,2% que hemos encontrado inicialmente no fue una coincidencia; además, esto coincide con lo que ocurre en las revisiones donde el tema es muy específico. Esto también nos permite apreciar algo intrigante: un importante porcentaje del contenido reciente de IA se refiere únicamente a las mejoras en el programa y no menciona ningún debate sobre leyes ni normas reales. La gran mayoría de los descartes severos tuvieron lugar después de identificar la primera capa filtradora porque aparecíamos ante cientos de propuestas técnicas que no contaban con un componente ético alguno o analizaban sistemas antiguos que no tenían nada que ver con los LLM actuales. Por tanto, gran parte de estos 69 artículos está disponible en Scopus e IEEE Xplore, y la mayoría data del periodo comprendido entre 2024 y 2026. Esto tiene sentido si consideramos cómo el interés académico aumentó exponencialmente después de que herramientas como ChatGPT se hicieron populares entre los años 2022 y 2023. La Tabla 2 sintetiza los hallazgos de cada una de estas etapas.

2.4. *Instrumentos de análisis*

Para organizar y interpretar todos los artículos seleccionados, hemos empleado dos herramientas simples que diseñamos nosotros mismos. El primero fue un cuestionario cuyos datos han sido agrupados en cinco grandes categorías temáticas. Este procedimiento consistió en recolectar información relacionada con numerosas áreas tales como las limitaciones inherentes al derecho internacional, fallos en los procesos de algunas empresas tecnológicas más importantes, etc. También hicimos un análisis comparativo mediante un esquema, en el que empleamos cinco distintas medidas para evaluar cada articulo: la gestión de los datos, la transparencia de la política de negocios, la aplicación de la legislación vigente, el deber de empresa y el seguimiento del consentimiento informado.

El análisis textual fusiono pensamientos previamente establecidos con nuevos descubrimientos que surgieron de manera inesperada mientras se leía. Esto proporcionó un marco que nos permitió seguir usando las categorías preexistentes. Aunque asumimos estos conceptos, también nos dio libertad para abordar temas significativos sin haberlos incluido originalmente como parte de nuestras estrategias iniciales, tales como el entorno político detrás de los sesgos de los algoritmos o la compleja red jurídica que surgió cuando buscábamos identificar las áreas responsables en la configuración de la cadena de suministro de una red inteligente neuronal.

3. Resultados

Tras la evaluación de los 69 artículos, sus hallazgos fueron categorizados bajo seis temas principales. En primer lugar, se enumeraron los documentos disponibles en la colección. Luego, se identificaron las distintas escuelas éticas presentes en la literatura disponible hasta la fecha. Después, se compararon las acciones realizadas por las cinco corporaciones informáticas en relación con el contenido que ofrecían. Se centraron luego en la atención prestada a los problemas asociados con el procesamiento de datos de entrenamiento. También se abordaron las condiciones legales actuales en el globo entero. Finalmente, se analizaron las actualizaciones recientes sobre la materia publicadas en publicaciones recientes. Cada uno de estos aspectos mostró fenómenos y circunstancias discordantes respecto a lo indicado inicialmente en el estudio.

3.1. Caracterización del corpus

De la revisión inicial del corpus se desprende que aproximadamente un tercio de los artículos (33%) son trabajos teórico-normativos que intentan establecer qué formas deberían tomar las regulaciones, en lugar de describir hechos concretos. Esta proporción, sumada a la presencia de artículos empírico-tecnológicos del 27%, indica que el campo se encuentra en etapa temprana de formulación y definición conceptual. Por ello, la escasez de estudios empíricos cuantitativos no es indicativa de una falta metodológica en la literatura sino del fenómeno reciente que es este tema dentro de la investigación académica.

Por su parte, en lo que concierne a la región geográfica del objeto de estudio, la Unión Europea ocupa el primer lugar (41%), seguido por Estados Unidos (33%), el contexto global sin foco geográfico (18%), Asia (5%) y América Latina (3%, principalmente México y Ecuador). Estos números dicen algo que no debería pasar desapercibido, el Sur Global sigue ausente del debate que define cómo se regulará la tecnología que ya está usando, ese número es llamativo si consideramos que varias de las tecnologías analizadas se adoptan con rapidez en los países latinoamericanos.

En cuanto al período de publicación, el 68% de los artículos del corpus fueron publicados entre 2024 y 2026, con un pico pronunciado en 2024-2025. Esta concentración temporal confirma la emergencia del tema como área de investigación activa, directamente vinculada a la masificación de los sistemas LLM a partir del lanzamiento público de ChatGPT en noviembre de 2022 y la posterior proliferación de modelos competidores. El período 2017-2021 está representado por apenas el 8% del corpus, mientras que el 24% restante corresponde a publicaciones de 2022-2023.

Los modelos más mencionados en el corpus son ChatGPT/GPT-4 de OpenAI (presente en el 61% de los artículos), seguido por Gemini de Google (34%), Claude de Anthropic (22%), Llama de Meta (18%) y los modelos de Microsoft/Azure (15%). Los marcos regulatorios más referenciados son el GDPR europeo (67%), el EU AI Act (58%), el NIST AI RMF (21%) y los Principios de IA de la OCDE (18%). Los journals con mayor representación en el corpus son IEEE Transactions on Technology and Society, AI & Society, Philosophy & Technology e IEEE Access. La Tabla 5 sintetiza la frecuencia de aparición de modelos LLM y marcos regulatorios en el corpus analizado.

Tabla 3. Frecuencia de modelos LLM y marcos regulatorios en el corpus (n = 69 artículos)

Modelo LLM	% artículos	Marco regulatorio	Referente	% artículos
ChatGPT/GPT-4(OpenAI)	**61%**	GDPR (UE)	Reglamento Europeo Protecc. Datos	**67%**
Gemini (Google)	34%	EU AI Act (UE)	Ley de IA de la Unión Europea	58%
Claude (Anthropic)	22%	NIST AI RMF (EE. UU.)	Marco de gestión de riesgos de IA	21%
Llama (Meta)	18%	Principios OCDE	Principios de IA de la OCDE (2019/2024)	18%
Copilot / Azure (Microsoft)	15%	UNESCO IA Ethics	Recomendación UNESCO sobre Ética IA	12%

Fuente: elaboración propia a partir del análisis del corpus de revisión sistemática.

En cuanto a los tipos de metodología empleada en los artículos del corpus, se identificaron análisis documental y comparativo de marcos normativos (38%), estudios de caso corporativos (22%), análisis de texto automatizado sobre outputs de LLM (18%), experimentos controlados para detección de sesgos (14%) y revisiones de literatura (8%). La diversidad metodológica es una fortaleza del campo, aunque también dificulta la síntesis y comparación directa de resultados.

3.2. *Marcos éticos identificados en la literatura*

El análisis temático del corpus permite identificar cuatro categorías principales de marcos éticos presentes en la literatura, con características, alcances y limitaciones claramente diferenciados entre sí.

3.2.1. Marcos corporativos voluntarios

Los códigos de conducta que las mismas empresas se autoimponen son el recurso más común en la industria, pero también el que recibe los golpes más duros por parte de los

investigadores. Casos conocidos como las guías éticas de Google, la IA Constitucional que propone Anthropic o las políticas de OpenAI sirven para ilustrar esto. La crítica principal de la literatura se enfoca en que estos compromisos quedan sin efecto debido a la falta de auditorías internas sólidas y metodologías claras para evaluar si las empresas los cumplen. Casi todos los libros mencionan metas idealistas como seguridad, justicia, transparencia y privacidad, pero pocos ofrecen maneras precisas de comprobar si las empresas las incumplen en sus operaciones diarias.

De acuerdo con estudios revisados, existen grandes diferencias entre lo que las empresas prometen a la sociedad y lo que realmente hacen en sus departamentos de desarrollo. Esta distorsión está principalmente atribuida a que se complica enormemente convertir conceptos abstractos como la "equidad" en algoritmos de aprendizaje profundo; estas últimas pueden ser incomprensibles. A pesar de eso, algunos autores sostienen que muchas veces ese hueco existe deliberadamente, actuando más como estrategia de relaciones públicas para mejorar la imagen de la empresa y evitando las posibles leyes gubernamentales que podrían surgir, y no como un verdadero modelo que funcione para los ingenieros.

3.2.2. Marcos regulatorios vinculantes

En el ámbito de las leyes obligatorias, el paso significativo a nivel internacional se ha manifestado en la Ley de IA de la UE aprobada en 2024, la cual organiza los sistemas de inteligencia artificial en cuatro niveles según el potencial perjudicial. Aunque la propuesta aparentemente es sistemática, en realidad resulta difícil cumplirla en la práctica; el riesgo de un grande modelo de IA puede alterarse drásticamente en función de la intención del usuario. Por ejemplo, el mismo sistema puede emplearse para la optimización de correos corporativos o para difundir falso contenido médico. La rigidez de estas categorías estáticas resulta repetida en la literatura académica [48]. En el caso de modelos destinados a un fin determinado, la ley pretende establecer respuestas claras sobre la procedencia de los datos de entrenamiento, la evaluación de riesgos a escala sistémica y el respeto a la propiedad intelectual.

Se muestra la academia muy crítica frente a la amplidez de este reglamento europeo en el caso específico de los LLM. Trabajos como la de Pleskach [48] evidencian deficiencias legales preocupantes en aspectos como la moderación del contenido sintetizado y la cadena de responsabilidades en caso de que múltiples organizaciones interactúen con un mismo modelo. También, desde una perspectiva similar, Jonnala y sus colaboradores [28] observan que la legislación europea fue concebida con el objetivo de regular las aplicaciones informáticas preexistentes que siguen reglas definidas; por tanto, se considera insuficiente al tratar de supervisar sistemas probabilísticos cuya respuesta fluctúa y se ve influenciada por el contexto específico en el que se utilicen.

3.2.3. Marcos técnicos y estándares

Desde el punto de vista de la ingeniería del software práctico, herramientas como la plataforma AIF360 de IBM o la suite de evaluación TrustLLM se convierten en instrumentos útiles para explorar y analizar cuántos sesgos hay en los algoritmos y si estas respuestas son estable. En el ámbito de las normativas, la ISO/IEC 42001 promueve la gestión organizativa de tales avances. La ventaja de estos métodos técnicos radica en que proporcionan datos numéricos y medidas precisas que pueden compararse entre diferentes modelos. La literatura expresa, sin embargo, la dificultad que existen por la ausencia de una definición universal de la equidad matemática. Además, las mesas de establecimiento de estándares proyectan textos para estandarizar la tecnología con gran lentitud, más rápida que la evolución que se produzca en el campo del

software. Por poner un ejemplo, la propia norma ISO/IEC 42001 ya se considera antigua y precisa de reforma antes de que muchas regiones latinas puedan aplicarla satisfactoriamente.

3.2.4. Marcos normativos internacionales

Por último, las propuestas de instituciones multilaterales como la Recomendación sobre la Ética de la IA de la UNESCO del 2021 o los Principios de la OCDE actualizados en 2024 gozan de un gran respaldo político global. Su verdadero mérito, a diferencia de leyes regionales como el GDPR, es que nacieron pensando en problemas que van más allá del continente europeo. La contrapartida es que nadie está obligado a seguirlos: su implementación depende de que los gobiernos nacionales los conviertan en ley, y esa voluntad política brilla por su ausencia en buena parte del Sur Global.

3.3. Comparativa global de enfoques corporativos

La Tabla 4 sintetiza los patrones de comportamiento ético documentados para las principales empresas analizadas en el corpus, a partir de la comparación de cinco criterios de evaluación identificados en la literatura revisada.

Tabla 4. Comparativa de enfoques éticos en empresas líderes de IA generativa

Empresa	Transparencia	Responsabilidad corporativa	Gobernanza datos	Cumplimiento normativo	Modelo predominante
OpenAI	Baja no divulga datasets	Alta en papers, cuestionada en práctica	Opaca en entrenamiento	Resistencia a regulación	Autorregulación + lobby
Anthropic	Media IA Constitucional publicada	Alta principios HHH documentados	Selectiva y controlada	Colaboración regulatoria	Autorregulación ética
Google / DeepMind	Media reportes de impacto	Alta declarativa, cuestionada	GDPR compliance parcial	Multada €4.34B (antimonopolio)	Híbrido (soft+hard law)
Meta	Baja modelos abiertos sin datos	Media incidentes documentados	Multada €1.2B GDPR	Resistencia activa	Código abierto + evasión regulatoria
Microsoft	Media-Alta Azure enterprise	Media cloud compliance	Compliance enterprise (GDPR, HIPAA)	Colaboración con reguladores	Corporativo / enterprise

Fuente: elaboración propia a partir del análisis del corpus de revisión sistemática.

La comparativa revela diferencias estructurales significativas entre las empresas analizadas. OpenAI ha sido consistentemente señalada como el actor con mayor resistencia a la transparencia en datos de entrenamiento [6], con una estrategia que combina altos niveles de publicación científica con opacidad sobre los aspectos más sensibles del proceso de desarrollo. La transición de OpenAI de organización sin fines de lucro a entidad con fines de lucro ha sido analizada en varios artículos del corpus como un caso paradigmático de tensión entre misión ética fundacional y presiones comerciales.

Anthropic destaca en el corpus por haber publicado documentación explícita sobre su marco ético interno la IA Constitucional (Constitutional AI) y los principios HHH (Helpful,

Honest, Harmless) incluyendo papers técnicos que describen los mecanismos de alineación empleados. Sin embargo, la efectividad de este marco sin mecanismos externos de rendición de cuentas ha sido cuestionada [19], y la empresa enfrenta el mismo desafío estructural que sus competidores, la dificultad de verificar externamente el cumplimiento de sus compromisos declarados.

Google/DeepMind presentan el caso más complejo, puesto que es una empresa con una historia documentada de investigación ética en IA incluyendo la pionera creación del equipo de Ética de IA de Google, pero también con incidentes documentados de despido de investigadores que señalaron problemas éticos en productos de la empresa, y con el mayor volumen de sanciones regulatorias entre las empresas analizadas. Este patrón sugiere que la existencia de equipos de ética internos no garantiza, por sí sola, comportamiento ético corporativo verificable.

Meta presenta un perfil particular dado su apuesta estratégica por los modelos de código abierto la familia Llama como alternativa al desarrollo propietario. La literatura está dividida sobre las implicaciones éticas de esta estrategia, por un lado, los modelos abiertos permiten mayor escrutinio externo y democratización del acceso; por otro, eliminan los mecanismos de control de acceso que los modelos propietarios pueden implementar, facilitando el uso malicioso. La ausencia de divulgación sobre los datos de entrenamiento de los modelos Llama, pese a su carácter "abierto", es señalada como una contradicción en el corpus.

Microsoft ocupa una posición diferente en el ecosistema, de modo que actúa fundamentalmente como integrador y desplegador de modelos desarrollados por terceros (principalmente OpenAI). Esta condición traslada una parte significativa de la responsabilidad ética del diseño de base hacia sus proveedores tecnológicos. La empresa ha desarrollado el marco de IA Responsable más detallado entre las analizadas en términos de herramientas de implementación práctica para desarrolladores, incluyendo directrices de contenido, herramientas de red-teaming y mecanismos de reporte de incidentes.

3.4. Tensiones éticas en datos de entrenamiento

En el corpus identifique tres ejes principales de conflicto ético en la adquisición y uso de datos de entrenamiento para LLM, presentes en al menos 15 de los 69 artículos analizados.

3.4.1. Consentimiento y propiedad intelectual

Múltiples estudios documentan el uso masivo de datos extraídos de internet sin consentimiento de los creadores originales. McIntyre [39] desarrolla una argumentación filosófica sobre el derecho moral de los creadores a restringir el entrenamiento de IA con sus obras, argumentando que este derecho se fundamenta en principios de autonomía personal y control sobre la propia expresión creativa, independientemente de los marcos legales de derechos de autor actualmente vigentes.

Sobre este mismo problema, el trabajo de Haidemariam y Gran [27] revisa a fondo el famoso paquete de datos Books3 y demuestra que contiene un montón de libros protegidos que los autores jamás autorizaron distribuir. Pero más allá del robo de contenido, estos investigadores alertan sobre un fuerte sesgo lingüístico: como casi todo está en inglés, los modelos terminan dándole prioridad a ese idioma, lo que deja en clara desventaja y con peores herramientas a quienes usan los LLM en otras lenguas.

En la parte legal, Al-Kaswan e Izadi [9] ponen la lupa sobre cómo se explota el código abierto para entrenar herramientas comerciales como GitHub Copilot. El punto crítico que ellos defienden es que, aunque ese código fuente esté libre en internet, meterlo a la fuerza en sistemas

que luego van a cobrar por su uso puede romper de frente las licencias originales y sus condiciones de uso. El corpus evidencia que el debate legal sobre los límites del fair use en el contexto del entrenamiento de LLM está activamente en litigio en múltiples jurisdicciones y no ha sido resuelto definitivamente en ninguna de ellas.

3.4.2. Opacidad corporativa sobre datos

La falta de transparencia sobre los datos de entrenamiento utilizados es documentada sistemáticamente en el corpus. Estudios como el de Buchicchio et al. [20] señalan que incluso los ingenieros que implementan sistemas basados en LLM carecen de acceso a información detallada sobre los conjuntos de datos de entrenamiento de los modelos fundacionales que utilizan. Esta opacidad dificulta notablemente la detección temprana de sesgos algorítmicos, entorpece la auditoría de seguridad y obstaculiza el cumplimiento de obligaciones legales críticas, tales como la notificación de brechas bajo el GDPR o la verificación rigurosa de los derechos de propiedad intelectual. Esta opacidad es estructural los datos de entrenamiento de los modelos fundacionales se consideran propiedad intelectual sensible y genera lo que algunos autores denominan una cadena de irresponsabilidad en donde cada actor de la cadena de suministro puede argumentar que la responsabilidad recae en otro nivel.

Zhu et al. [67] proponen un método técnico de auditoría para detectar el uso parcial de datasets específicos en LLM mediante agregación difusa de pertenencia, ofreciendo una herramienta que podría contribuir a mayor transparencia en este aspecto. Sin embargo, los autores reconocen que la aplicabilidad práctica de estas técnicas requiere niveles de acceso a los modelos que los proveedores frecuentemente no otorgan a investigadores externos.

3.4.3. Sesgo algorítmico estructural

Un tema muy revisado en los artículos es cómo los sesgos se meten en los modelos desde la etapa de preentrenamiento. Sobre esto, el experimento práctico que hizo Choudhary [22] nos da cifras bastante claras: puso a prueba a ChatGPT-4, Gemini, Perplexity y Claude con preguntas sobre temas políticos y sociales bastante complejos, y descubrió que todos tendían a dar respuestas inclinadas hacia visiones de corte liberal y progresista. Esto hace evidente que las prácticas empresariales para «harmonizar» la tecnología aún no garantizan que el lenguaje mantenga su neutralidad.

Desde una perspectiva teórica más avanzada, Ferrara [24] observa este problema y postula que el lenguaje de las personas trae consigo una carga prejuiciosa derivada de la historia. Por eso, opina que intentar lograr una neutralidad total en el lenguaje de LLM es insostenible por razones lógicas: por ello, el desafío principal debe ser la transparencia; las compañías deberían ser clara sobre los valores y posiciones que adoptan sus sistemas.

Ahmad y su equipo [4] sugieren combinar métodos matemáticos con análisis sociales para detectar y restringir estos prejuicios, dado que no alcanzan las soluciones puramente técnicas porque ignoran la relevancia del entorno cultural en la que funcionan las LLMs. Así como Narayan [44] expone una idea alternativa frente a lo preocupante predominio de usar modelos dominados por el inglés y europeos, debería diseñar inteligencias artificiales orientadas específicamente para cada cultura y entrenadas con información y realidad local. Esto ayudaría a evitar las limitaciones inherentes a las LLMs.

3.5. *Perspectiva regulatoria global y brechas identificadas*

El estudio comparativo de regulaciones globales, soportado por los artículos de mayor relevancia en el corpus, permite identificar cinco modelos regulatorios predominantes con características, fortalezas y limitaciones diferenciadas. La Tabla 5 sintetiza estas dimensiones.

Tabla 5. Modelos regulatorios globales para IA Comparativa

Modelo	Región / Referente	Tipo de regulación	Fortaleza principal	Limitación principal
Europeo	UE (EU AI Act + GDPR)	Vinculante, basado en riesgo	Mayor protección de derechos	Inhibición de innovación, complejidad de implementación
Estadounidense	EE.UU. (NIST AI RMF, HIPAA, CCPA)	Sectorial + marcos voluntarios	Flexibilidad e innovación	Incertidumbre regulatoria, incapacidad de contener Big Tech
Chino	China (CAC, regulación LLM 2023)	Estatal dirigido	Control y refuerzo robusto	Instrumentalización política, restricción de libertades
Asiático flexible	Singapur, Corea del Sur	Voluntario con refuerzo sectorial	Equilibrio innovación-seguridad	Fragmentación regulatoria, baja exigibilidad
Latinoamericano emergente	Ecuador, México, Colombia	Reactivo y fragmentado	Adaptabilidad a capacidades locales	Ausencia de estrategia nacional integrada, bajo refuerzo

Fuente: elaboración propia a partir del análisis del corpus de revisión sistemática.

Las brechas normativas más frecuentemente documentadas en el corpus incluyen: (a) la incompatibilidad de los marcos determinísticos de protección de datos con la naturaleza probabilística de los LLM, (b) la ausencia de regulación en las etapas iniciales de la recopilación de los datos de entrenamiento en la mayoría de jurisdicciones, (c) la falta de mecanismos claros de responsabilidad compartida en las cadenas de suministro de IA, (d) la escasez de capacidades técnicas de refuerzo en los organismos reguladores y (e) la subrepresentación del Sur Global incluyendo Latinoamérica en los debates regulatorios internacionales.

Kulothungan y Gupta [34] argumentan que ninguno de los modelos regulatorios existentes es globalmente replicable en su forma actual, dada la dependencia de contextos institucionales, legales y tecnológicos específicos. Proponen en cambio un modelo adaptativo de gobernanza que combina principios compartidos a nivel internacional con mecanismos de implementación ajustables a las capacidades y prioridades regulatorias de cada jurisdicción.

Desde este ángulo descentralizado, encaja muy bien con la realidad de América Latina, donde las normas locales en vigor se topan frecuentemente con la falta de recursos o de formación en las instituciones del Estado. Revisando casos particulares de la región, el estudio de Meza y su grupo [40] dirigido a Ecuador presentó una visión de gobernanza fragmentada que reacciona solo cuando los problemas se han instalado; aunque hay disposiciones para salvaguardar datos, el país no cuenta con un plan nacional integral integrado que incluya desarrollo tecnológico, leyes y capacitación profesional. Por otro lado, Aguilar Antonio [2] señala que México está atrapado en el mayor retraso de la región norteamericana por desarrollar políticas de IA, lo cual lleva a una escasa coordinación entre sus entidades y poco desempeño en mesas de debates internacionales.

Las escenas colombianas reflejan estas características regionales. El país cuenta con importantes instrumentos legales como la Ley 1581 de 2012 sobre protección de datos personales, el Conpes 3975 de 2019 sobre digitalización y transparencia, y otras normas éticas de inteligencia artificial que hoy se encuentran en el Congreso en 2024. El mayor inconveniente radica en que aún no existe una norma legal obligatoria y específica para manejar los LLM o la Inteligencia Artificial generativa. Igualmente el Conpes 3975 da lineamientos sobre transparencia e igualdad en el ámbito público, pero no habla de auditorias externas ni de sancionar a las grandes compañías tecnológicas internacionales que operan en el territorio nacional.

Mientras tanto, el debate judicial sigue estancado, mientras programas como los LLM se infiltran con rapidez en campos extremadamente delicados como las universidades, los hospitales públicos y las oficinas judiciales. Se produce en un vacío jurídico absoluto que deja a la población vulnerable a la proliferación de sesgos y la disminución de derechos digitales, aspectos todavía prácticamente desconocidos por las autoridades. Con relación a esto, las universidades colombianas, tales como la Universidad Santo Tomás, tienen una tarea crucial de liderazgo para iniciar investigaciones y proponer normativas éticas que nos ayuden a desarrollar reglas en línea con nuestra propia situación, colaborando para llenar esa brecha significativa que separa el rápido desarrollo tecnológico de la toma de acción del estado en esta materia.

3.6. *Tendencias de publicación y análisis bibliométrico*

El comportamiento de las publicaciones dentro del corpus ayuda a entender cómo ha venido cambiando el interés por este tema. Al revisar los números, salta a la vista un aumento gigante en la cantidad de documentos publicados justo después de que ChatGPT se volviera masivo a finales de 2022. Para ser exactos, la literatura se multiplicó por seis al comparar el periodo de 2022 con el de 2024. Esto demuestra claramente que la comunidad científica reaccionó con prisa ante lo que pasaba en el mercado, en lugar de adelantarse con estudios previos a la llegada de estas herramientas. Este crecimiento se dio con mucha más fuerza en las plataformas de IEEE que en las revistas dedicadas a las ciencias sociales o las humanidades, una diferencia que da a entender que el debate lo están dominando los ingenieros y los expertos en computación, dejando un espacio mucho menor para las opiniones de filósofos, abogados o sociólogos.

En cuanto a las redes de trabajo entre autores, encontramos que los artículos muestran muy poca colaboración entre diferentes instituciones o países; la gran mayoría de los textos fueron escritos por equipos del mismo entorno geográfico y de la misma carrera. Llama la atención lo raro que es encontrar papers firmados en conjunto por investigadores del Norte y del Sur Global, lo que confirma que hacen falta puentes de cooperación para que las realidades y necesidades de los países emergentes tengan peso en las discusiones sobre las reglas éticas del planeta.

Para cerrar, el rastreo de las palabras clave nos permitió identificar varios temas que vienen ganando fuerza en el último tiempo. Por ejemplo, la preocupación por alinear los valores dentro de los LLM pasó de estar en apenas un 4% de los textos antes de 2023 a superar el 23% en los artículos publicados entre 2025 y 2026. Del mismo modo, las discusiones sobre los derechos de autor con respecto a los grandes modelos explotaron a partir de 2023, seguidas muy de cerca desde 2024 por análisis enfocados en regular la IA generativa dentro de sectores específicos como la medicina, las escuelas y los juzgados. Por último, empieza a asomar con timidez pero con mucha proyección el debate sobre cómo se comportan estos modelos en idiomas poco representados o en entornos que cuentan con muy bajos recursos tecnológicos.

4. Discusión

Lo que encontramos en esta revisión no solo respalda sino que profundiza en esa desconexión ya conocida entre los discursos éticos que las tecnológicas publican y la forma en que operan realmente. Los estudios coinciden en destacar la importancia de la autocorrección de las compañías en promover una cultura de responsabilidad interna, pero esta es demasiado restrictiva cuando se trata de limitar tecnologías con el alcance general de las actuales LLM. El hecho de que los problemas persistan a pesar de la denuncia del medio y de la investigación científica es un signo inequívoco de que solucionarlo no significa escribir normas morales mejor o más sofisticadas, sino un problema vinculado a incentivos económicos básicamente puros. Las empresas que desarrollan estas herramientas entran en competencia en un escenario donde ganar, mantener, usar el dato y expandir el sistema determinan a quien se mantiene. Si comprometemos una ética que supone retrasar las innovaciones, perder fuentes de información o aumentar los procesos de revisión, va en contra al interés financiero de las empresas. Abordar esta contradicción exige más que los lineamientos voluntarios; hay que confiar en controles externos reales con efectividad legal.

El ejemplo más claro que saltó en la literatura tiene que ver con el secreto que rodea a los datos de entrenamiento. Mientras casi todas las empresas hacen campañas públicas defendiendo la transparencia, en la práctica ni los propios ingenieros que diseñan herramientas sobre esos modelos saben con exactitud qué información se usó para alimentarlos [20]. Perera y su equipo [47] describen esto de forma muy acertada como una "desconexión significativa" entre las políticas de la empresa y lo que le cuentan al público. A partir de los textos analizados, se puede deducir que la transparencia se maneja más como un escudo reputacional que como una política real, una diferencia clave si se busca diseñar normativas que sirvan. Este comportamiento encaja con lo que advierten Tran y sus colaboradores [58], quienes señalan que la mayoría de los proveedores de LLM están incumpliendo las exigencias básicas de la Ley de IA de la Unión Europea.

Al contrastar las estrategias regulatorias a nivel global, se hacen evidentes varios sacrificios complejos que repercuten en el debate sobre qué rumbo deberían tomar las regiones emergentes. Por un lado, el enfoque de la Unión Europea es el que más protege los derechos de los usuarios y el que tiene mayor peso legal, pero la academia lo critica por sus trabas burocráticas que podrían frenar la innovación y por ser un laberinto inviable para las pequeñas empresas [34]. En la otra orilla, el modelo de Estados Unidos le da total libertad a la innovación y flexibilidad al mercado, pero el corpus demuestra que esa laxitud genera incertidumbre y ha sido incapaz de frenar los abusos de las grandes tecnológicas [52]. Finalmente, el camino que tomó China destaca por su efectividad de control, pero despierta alertas severas debido a que instrumentaliza la regulación algorítmica para la vigilancia social y el control político del Estado [60].

Otra es la estrategia de código abierto que ha defendido Meta y otros problemas éticos en cuanto a la difusión del código fuente. La academia ha tenido debates complejos sobre esta cuestión: la liberación del código permite que los investigadores independientes auditen la tecnología y detecten vulnerabilidades, pero también puede desactivar los mecanismos de defensa contra el malware al permitir a hackers armados con estas herramientas.

La literatura revisada encuentra un área insuficientemente regulada que no se ha considerado un control específico de exportaciones o un tipo de licencia de utilización que garantice que los códigos libres no sean objeto de manipulaciones. En términos prácticos, este modelo se plantea en toda Latinoamérica, aunque los estudios académicos en esta dirección son relativamente bajos. En su análisis, Meza [40], desde Ecuador, y Aguilar Antonio [2], desde México, presentan casos de fragmentación de gobernanza que probablemente estén compartidos

por buena parte del continente. Esta mezcla de alta adopción tecnológica con baja gobernanza es preocupante porque las ventajas de los LLM (como los aumentos de productividad o el acceso rápido a información) se quedan en los sectores de la población que ya están en una posición ventajosa, mientras que los golpes duros (como las noticias falsas, los sesgos y el riesgo de desplazamiento laboral) se reparten de manera generalizada.

La captura desigual de beneficios y riesgos puede profundizar desigualdades ya existentes en contextos donde las brechas socioeconómicas son pronunciadas. Los hallazgos sobre sesgo político de los LLM [22] plantean interrogantes adicionales para los contextos latinoamericanos. El trabajo de Choudhary [22] sobre sesgo político en ChatGPT-4, Gemini, Perplexity y Claude nos parece especialmente preocupante para el contexto latinoamericano. Si estos modelos tienen inclinaciones ideológicas detectables como sugiere la evidencia y millones de personas los usan para informarse sobre temas políticos sin saber que esas inclinaciones existen, estamos hablando de un riesgo para la deliberación democrática que todavía no ha sido tomado suficientemente en serio por los reguladores de la región. Y eso en un momento en que varios países latinoamericanos viven procesos electorales o de construcción institucional especialmente frágiles. Este riesgo es particularmente significativo en el caso de situaciones donde la alfabetización digital y mediática hace que los ciudadanos no puedan detectar ni compensar los sesgos de los sistemas de inteligencia artificial que consultan.

Desde una perspectiva crítica, cabe señalar que prácticamente todos estos marcos (orgánicos o corporativos) fueron pensados en el Norte Global, teniendo en cuenta sus propias características culturales, legales e institucionales. Solicitar principios de privacidad fundamentados en el GDPR europeo a comunidades donde la visión colectiva sobre la comunidad supera a la individual y buscar aplicar el consentimiento informado en áreas donde se encuentran serias carencias en cuestión de alfabetización digital, constituyen una tarea ardua para traducir y adaptar lo cual la academia aún no ha empezado a tratar. Formular principios éticos que sean plenamente interculturales sigue siendo uno de los mayores obstáculos para el avance de esta disciplina.

El corpus revisado muestra una extraordinaria diversidad de procedimientos, abarcando temas como debates filosóficos y teóricos, análisis jurídico e informes sobre casos. La complejidad del estudio hace que no pueda efectuar mediciones estadísticas o formular afirmaciones que respalden conclusiones basadas en hechos concretos y claramente definidos. El análisis no está limitado únicamente a esos aspectos. Por esta razón, resulta fundamental proseguir investigando con el objetivo de diseñar instrumentos y herramientas originales capaces de evaluar ambas disciplinas simultáneamente: la tecnología y la gestión lingüística y administrativa asociada a la inteligencia artificial.

5. Conclusiones

Los estudios iniciales plantearon una pregunta, pero al seguir investigando nos encontramos con una respuesta compleja y hasta incómoda: las grandes compañías de tecnología no coinciden con los principios éticos que reconocen y difunden al público. Todos los 69 textos revisados expresaron que hay importantes contrastes entre el comunicado oficial de esas empresas y las realidades diarias de su industria.

Específicamente, estos estudios señalan que hay una enorme contradicción en el seno del ámbito tecnológico actual. Fue sorprendente observar empresas pioneras en la literatura sobre inteligencia artificial ética, presentar hechos y figuras que han acumulado multas importantes durante años, ocultar la originales de la información empleada para entrenar sus sistemas y compartir culpas entre ellas mismas. De hecho, esta situación puede interpretarse como resultado

de una falta de integridad, aunque la realidad muestra un obstáculo económico y empresarial en juego. En un entorno tan competitivo, donde la eficacia se mediría por la velocidad de desarrollo y la gestión de grandes volúmenes de datos, cualquier barrera ética asumida se vería abrumada por los objetivos económicos y empresariales de la empresa. Como consecuencia, la autorregulación interna funciona como medio para identificar los principios éticos (como las banderas), pero es totalmente inadecuada sin un control regulatorio externo con suficiente fuerza criminal.

Otra regulación del mundo, el EU AI Act, tampoco ha alcanzado su objetivo principal. Esto se debe a que la normativa europea fue concebida con el propósito de regular un sistema predecible, tradicionalmente conocido como "sistemas determinísticos", mientras que los LLM actuales operan de manera probabilística, contextual e intrínsicamente oculta. La clasificación por niveles de riesgo asume que podemos predecir el impacto de un sistema de IA: esa asunción se rompe cuando el mismo modelo puede usarse para redactar poesía o para fabricar desinformación. Lo que se necesita no es parchear el marco existente, sino repensar la unidad de análisis: regular cadenas de suministro y capacidades, no productos finales.

La tercera conclusión es la que más nos interesa desde nuestra posición en la Universidad Santo Tomás de Tunja. América Latina representa apenas el 3 % del corpus que revisamos. No es que no haya adopción de LLMs en la región al contrario, la adopción está siendo acelerada y acrítica; es que la región casi no produce conocimiento sobre los marcos éticos que deberían regularla. Mientras eso no cambie, los estándares que terminaremos aplicando habrán sido diseñados sin considerar nuestras realidades lingüísticas, institucionales, socioeconómicas. Cambiar eso no es solo un asunto académico: es una decisión política sobre soberanía tecnológica.

5.1. Recomendaciones orientadoras

A partir de estos hallazgos, se proponen las siguientes recomendaciones orientadoras para el fortalecimiento ético del desarrollo de LLM en el ámbito corporativo y regulatorio:

- Implementar regulación en la fase inicial de datos de entrenamiento, en donde se exija transparencia sobre las fuentes de datos utilizadas, los mecanismos de consentimiento aplicados y los procedimientos de auditoría de sesgo, antes del entrenamiento del modelo. Esta regulación debe diseñarse de modo que sea tecnológicamente neutral y pueda adaptarse a la evolución rápida de las arquitecturas de LLM.
- Adoptar un modelo de responsabilidad diferenciada por cadena de suministro, que distribuya obligaciones entre proveedores de modelos fundacionales, afinadores (fine-tuners) y desplegadores, siguiendo el esquema establecido por el EU AI Act pero con mecanismos de responsabilidad solidaria en caso de daño documentado.
- Desarrollar marcos de gobernanza adaptativos para contextos del Sur Global, basados en principios de regulación por riesgo, capacidades institucionales existentes y armonización gradual con estándares internacionales (UNESCO, OCDE, ISO/IEC 42001). La regulación debe ser proporcional a las capacidades de refuerzos disponibles, para evitar marcos normativos que existan solo en papel.

Desde un punto de vista práctico, creemos necesaria una expansión de la labor normativa abriendencia de auditorías auditivas independientes y obligatorias del sesgo. Para ello deberían evaluarse principalmente dimensiones políticas, culturales e idiomáticas de las regiones, garantizando que los datos y sus resultados sean compartidos totalmente entre investigadores y instituciones del sector público y la sociedad civil, evitando que solo se ofrezcan exclusivamente a organismos estatales.

Al respecto, resulta crucial amplificar la participación latinoamericana en las instituciones internacionales claves en la política tecnológica global. Se deben crear espacios de representación y convertirlo en una voz que traduzca esa representación mediante las estrategias nacionales con recursos reales. Finalmente, para dar forma a este propósito, las condiciones más indispensables radican en apoyar la realización de investigaciones interdisciplinarias sobre ética en IA desde la perspectiva latinoamericana. El impulso educacional debe fusionar la ingeniería con filosofía, derecho y sociología para asegurar que la voz de comunidades que ya sufre los efectos directos del despliegue de esos modelos en áreas tan sensibles como la salud, la educación y el sistema judicial, se haga escuchar.

Tabla 6. Resumen de propuestas clave y lineamientos éticos para modelos LLM.

Recomendación	Actor responsable	Nivel	Urgencia
R1. Regulación de datos de entrenamiento: obligar a que se publiquen las fuentes, se pida permiso a los autores y se analicen los prejuicios antes de entrenar el sistema.	Reguladores nacionales e internacionales	Regulatorio	**Alta**
R2. Responsabilidad diferenciada por cadena de suministro: dividir las tareas y obligaciones legales según el rol de proveedores, desarrolladores o usuarios comerciales.	Reguladores / empresas tecnológicas	Regulatorio / Corporativo	**Alta**
R3. Marcos de gobernanza adaptativos para el Sur Global: crear reglas realistas que se ajusten al presupuesto de cada institución local y se alineen con los estándares de la UNESCO, OCDE e ISO.	Gobiernos nacionales / organismos internacionales	Regulatorio / Internacional	Media-Alta
R4. Auditoría independiente y obligatoria de sesgo para LLM masivos, abriendo el acceso de los resultados a científicos y organizaciones ciudadanas.	Reguladores / organizaciones de auditoría	Regulatorio / Técnico	Media-Alta
R5. Fortalecer presencia latinoamericana en OCDE, UNESCO, ITU, ISO y desarrollar estrategias nacionales de IA con mandato institucional.	Gobiernos regionales / cancillerías	Internacional / Político	Media
R6. Promover investigación de carácter mixto en la ética de la IA, dándole un enfoque regional que incluya directamente a las poblaciones afectadas.	Universidades / centros de investigación	Académico	Media

Fuente: elaboración propia a partir de los hallazgos de la revisión sistemática.

5.2. *Limitaciones del estudio*

En cuanto a las limitaciones del estudio, se reconocen las siguientes: (1) la búsqueda se limitó a Scopus e IEEE Xplore, excluyendo otras bases de datos relevantes como Web of Science, ACM Digital Library, PubMed o Google Scholar, lo que puede introducir sesgos de cobertura, (2) la mayoría del corpus está en inglés, lo que puede subrepresentar perspectivas no anglófonas, particularmente relevante dado el foco en el contexto latinoamericano, (3) el período de búsqueda (2017-2026) captura el fenómeno en su fase más reciente, pero la velocidad de cambio del campo implica que algunos hallazgos pueden quedar desactualizados rápidamente y (4) la revisión fue realizada por un investigador individual con revisión del director, sin proceso de doble codificación independiente que permitiera calcular índices de concordancia entre evaluadores.

Líneas de trabajo futuro

Como líneas de trabajo futuro se proponen: (1) ampliar la revisión sistemática a bases de datos especializadas en derecho, ciencias sociales y humanidades, (2) desarrollar estudios de caso específicos sobre la adopción de LLM en sectores críticos de Colombia y Latinoamérica, como salud, educación, justicia y banca, con atención a los impactos diferenciados sobre grupos vulnerables, (3) diseñar y validar instrumentos de evaluación ética de LLM adaptados al contexto latinoamericano, que incorporen dimensiones de diversidad lingüística, pluralismo cultural y equidad socioeconómica, (4) analizar la coherencia entre los marcos regulatorios nacionales emergentes en la región y los estándares internacionales de referencia, identificando las brechas de implementación más urgentes y (5) explorar el potencial de enfoques participativos y comunitarios para el diseño de marcos de gobernanza de IA en contextos del Sur Global, superando el modelo experto-descendente predominante en la literatura revisada.

Referencias

[1] Abueid, A. I., & Jazzar, M. M. (2025). Generative AI and the Ethics of Innovation: Towards Inclusive and Sustainable Futures. International Conference on Computer and Applications, ICCA 2025 - Proceedings. https://doi.org/10.1109/ICCA66035.2025.11430745

[2] Aguilar Antonio, J. M. (2024). Path and governance model of artificial intelligence (AI) policies in North American countries[Trayectoria y modelo de gobernanza de las políticas de inteligencia artificial (IA) de los países de América del Norte]. Justicia (Barranquilla), 29(45). https://doi.org/10.17081/just.29.45.7162

[3] Ahi, K., & Valizadeh, S. (2025). Large Language Models (LLMs) and Generative AI in Cybersecurity and Privacy: A Survey of Dual-Use Risks, AI-Generated Malware, Explainability, and Defensive Strategies. 2025 Silicon Valley Cybersecurity Conference, SVCC 2025. https://doi.org/10.1109/SVCC65277.2025.11133642

[4] Ahmad, A., Vallès, Y., & Idaghdour, Y. (2026). Bias in AI systems: integrating formal and socio-technical approaches. Frontiers in Big Data, 8. https://doi.org/10.3389/fdata.2025.1686452

[5] Alaswad, S., Kalganova, T., & Awad, W. (2025). Trustworthiness of Legal Considerations for the Use of LLMs in Education. 2025 International Conference on Decision Aid Sciences and Applications (DASA), 1-8. https://doi.org/10.1109/DASA68193.2025.11498739

[6] Albaroudi, E., Mansouri, T., Hatamleh, M., & Alameer, A. (2026). HitHire: The future of ethical, fair, and sustainable AI recruitment - A governance framework. Array, 29. https://doi.org/10.1016/j.array.2025.100592

[7] Al-Billeh, T., Hmaidan, R., Al-Hammouri, A., & al Makhmari, M. (2024a). The Risks of Using Artificial Intelligence on Privacy and Human Rights: Unifying Global Standards. Jurnal Media Hukum, 31(2), 333-350. https://doi.org/10.18196/jmh.v31i2.23480

[8] Ali, E., McOwen, B. K., Ashraf, A., & Kim, D. (2025). Navigating the AI Ethics Frontier: A Cross-national Comparison of AI Policy Documents for Developing Responsible AI Workforce. ASEE Annual Conference and Exposition, Conference Proceedings. https://doi.org/10.18260/1-2--56995

[9] Al-Kaswan, A., & Izadi, M. (2023). The (ab)use of Open Source Code to Train Large Language Models. Proceedings - 2023 IEEE/ACM 2nd International Workshop on Natural Language-Based Software Engineering, NLBSE 2023, 9-10. https://doi.org/10.1109/NLBSE59153.2023.00008

[10] Alyami, N. M., Badawood, S., Alomari, K. M., Al-Ababneh, H. A., al Adwan, M. N., Khader, J. A., & Zawaideh, F. H. (2026). ETHICAL AND LEGAL ASPECTS OF DIGITAL MARKETING, INFORMATION TECHNOLOGY, AND THEIR INFLUENCE ON CORPORATE GOVERNANCE AND SUSTAINABLE DEVELOPMENT. Corporate Law and Governance Review, 8(2), 63. https://doi.org/10.22495/clgrv8i2p6

[11] Anand, S., Gopinath, A., & Periyasami, K. (2026). Evaluating Bias Detection and Mitigation Approaches Across Classical and Large Language Models. IEEE Access. https://doi.org/10.1109/ACCESS.2026.3669650

[12] Baldassarre, M. T., Caivano, D., Fernández Nieto, B., & Ragone, A. (2025). Ethics-Driven Incentives: Supporting Government Policies for Responsible Artificial Intelligence Innovation. IEEE Intelligent Systems, 40(4), 55-63. https://doi.org/10.1109/MIS.2025.3583222

[13] Bang, J., Lee, B. T., & Park, P. (2023). Examination of Ethical Principles for LLM-Based Recommendations in Conversational AI. 2023 International Conference on Platform Technology and Service, PlatCon 2023 - Proceedings, 109-113. https://doi.org/10.1109/PlatCon60102.2023.10255221

[14] Bawamohiddin, A. B., & Arshad, N. I. (2025). Contextualized Hybrid Model for AI Ethics Governance in Malaysia's TVET Sector. Proceedings of 2025 IEEE International Conference on Data and Software Engineering, ICoDSE 2025, 544-549. https://doi.org/10.1109/ICoDSE68111.2025.11351672

[15] Beg, M. H. R., & Mehndiratta, V. (2024a). The Ethics of Generative AI: Analyzing ChatGPT's Impact on Bias, Fairness, Privacy, and Accountability. Proceedings of the 2024 International Conference on Innovative Computing, Intelligent Communication and Smart Electrical Systems, ICSES 2024. https://doi.org/10.1109/ICSES63760.2024.10910483

[16] Benraouane, S. A. (2024). AI Management System Certification According to the ISO/IEC 42001 Standard: How to Audit, Certify, and Build Responsible AI Systems. In AI Management System Certification According to the ISO/IEC 42001 Standard: How to Audit, Certify, and Build Responsible AI Systems. Taylor and Francis. https://doi.org/10.4324/9781003463979

[17] Berengueres, J. (2024). How to Regulate Large Language Models for Responsible AI. IEEE Transactions on Technology and Society, 5(2), 191-197. https://doi.org/10.1109/TTS.2024.3403681

[18] Bhardwaj, N., Bhardwaj, A., & Garg, L. (2025). Controlling Bias in Generative AI: Techniques for Fair and Equitable Data Generation in Socially Sensitive Applications. 3rd International Conference on Integrated Circuits and Communication Systems, ICICACS 2025. https://doi.org/10.1109/ICICACS65178.2025.10967861

[19] Boutsikaris, L., & Polykalas, S. (2025). Ethical Analysis of Large Language Models: A Comparison of GPT, DeepSeek, Claude, Gemini, LLaMA, and Qwen. International Symposium on Technology and Society, Proceedings, 1-7. https://doi.org/10.1109/ISTAS65609.2025.11269621

[20] Buchicchio, E., Angelis, A. de, Moschitta, A., Santoni, F., Marco, L. S., & Carbone, P. (2024). Design, Validation, and Risk Assessment of LLM-Based Generative AI Systems Operating in the Legal Sector. ISSE 2024 - 10th IEEE International Symposium on Systems Engineering, Proceedings. https://doi.org/10.1109/ISSE63315.2024.10741134

[21] Bunzel, N. (2025). Compliance Made Practical: Translating the EU AI Act into Implementable Security Actions. Proceedings - 2025 IEEE/ACM International Workshop on Responsible AI Engineering, RAIE 2025, 69-73. https://doi.org/10.1109/RAIE66699.2025.00016

[22] Choudhary, T. (2025). Political Bias in Large Language Models: A Comparative Analysis of ChatGPT-4, Perplexity, Google Gemini, and Claude. IEEE Access, 13, 11341-11379. https://doi.org/10.1109/ACCESS.2024.3523764

[23] Coen Collins, A. (2026). Garbage in, garbage out? How the monster of AI art reflects human fault, bias, and capitalism in contemporary culture. AI and Society. https://doi.org/10.1007/s00146-025-02799-5

[24] Ferrara, E. (2023). Should ChatGPT be biased? Challenges and risks of bias in large language models. First Monday, 28(11). https://doi.org/10.5210/fm.v28i11.13346

[25] Ferreira, A. H., Trevisan, A. C., Baptista, C. M., Ramos-Antón, R., Comin, Á. A., Carvalho, H. F., Vendrell, S., & Sá, V. O. (2026). Meta-Identity and Algorithmic Mediation on Digital Platforms: A Comparative Analysis of AI-Human Content Categorization. Societies, 16(4). https://doi.org/10.3390/soc16040132

[26] Gaurav, Khan, H., Pandey, S., Kumar, B. V., Akram, S. V., & Pal, P. (2024). Artificial Intelligence (AI) in the Field of Corporate Social Responsibility. Proceedings - 2024 3rd International Conference on Sentiment Analysis and Deep Learning, ICSADL 2024, 164-168. https://doi.org/10.1109/ICSADL61749.2024.00033

[27] Haidemariam, T., & Gran, A. B. (2026). On the problems of training generative AI: towards a hybrid approach combining technical and non-technical alignment strategies. AI and Society. https://doi.org/10.1007/s00146-025-02445-0

[28] Jonnala, S., Parida, P. K., & Thomas, N. M. (2025). EU AI Act Underrepresented and Insufficient to Address the Risk and Vulnerabilities of Generative AI. International Journal of Business Analytics, 12(1). https://doi.org/10.4018/IJBAN.388757

[29] Joshi, H., Hassani, S., Gandhi, D., & Hartman, L. (2025). Approaches to Responsible Governance of GenAI in Organizations. International Symposium on Technology and Society, Proceedings, 1-15. https://doi.org/10.1109/ISTAS65609.2025.11269657

[30] Kataria, V., Khan, S. A., & Gupta, S. (2025). Leveraging AI for Governance: Navigating Ethical Challenges and Enhancing Democratic Values in Public Services. 2025 International Conference on Intelligent and Secure Engineering Solutions, CISES 2025, 13-18. https://doi.org/10.1109/CISES66934.2025.11265150

[31] Khalifa, M., & Sabry, M. (2024). The Role of AI in Shaping Modern Legal Frameworks: A Political Science Perspective. 2024 International Conference on Decision Aid Sciences and Applications, DASA 2024. https://doi.org/10.1109/DASA63652.2024.10836624

[32] Kim, S. Y., Lim, J. H., Lee, D. W., Min, S. H., Kim, I. J., Jo, M. il, Shin, J. Y., & Kim, W. G. (2025). Selective LLM Unlearning via SAE-Based Token Importance Score. 1-6. https://doi.org/10.1109/isncc66965.2025.11250473

[33] Kovari, A., Ghafourian, Y., Hegedus, C., Naim, B. A., Mezei, K., Varga, P., & Tauber, M. (2025). Let's Have a Chat with the EU AI Act. Proceedings of IEEE/IFIP Network Operations and Management Symposium 2025, NOMS 2025. https://doi.org/10.1109/NOMS57970.2025.11073655

[34] Kulothungan, V., & Gupta, D. (2025). Towards Adaptive AI Governance: Comparative Insights from the U.S., EU, and Asia. Proceedings - 2025 IEEE 11th Conference on Big Data Security on Cloud, BigDataSecurity 2025, 32-38. https://doi.org/10.1109/BigDataSecurity66063.2025.00018

[35] Kumar, R., Kumar, H., & Shalini, K. (2025). Detecting and Mitigating Bias in LLMs through Knowledge Graph-Augmented Training. 2025 International Conference on Artificial Intelligence and Data Engineering, AIDE 2025 - Proceedings, 608-613. https://doi.org/10.1109/AIDE64228.2025.10987418

[36] Lee, I. (2025). Generative Artificial Intelligence for Enterprises: Ecosystem, Typology of Applications, and Challenges. IT Professional, 27(2), 28-34. https://doi.org/10.1109/MITP.2025.3544261

[37] Liu, J. Y. (2025). Generative Artificial Intelligence: Ethical Challenges and Trust Mechanisms. IT Professional, 27(6), 25-31. https://doi.org/10.1109/MITP.2025.3589384

[38] Mavani, V. K. (2026). An Autonomous Governance Framework for Generative AI: Real-Time PII Redaction and Compliance in LLM-Driven Data Pipelines. 2026 IEEE 5th International Conference on AI in Cybersecurity (ICAIC), 1-4. https://doi.org/10.1109/ICAIC67076.2026.11395823

[39] McIntyre, J. H. (2026). The Right to Restrict AI Training. Philosophy and Technology, 39(1). https://doi.org/10.1007/s13347-025-01032-x

[40] Meza, J., Saltos, M. F. L., Campoverde, E. U. V., Cejas, M. C. N., & Castro, V. C. A. (2025). AI Regulation for Ecuador. International Conference on eDemocracy and eGovernment, ICEDEG, (2025), 311-316. https://doi.org/10.1109/ICEDEG65568.2025.11081572

[41] Mhatre, A. (2023). Detecting the presence of social bias in GPT-3.5 using association tests. Proceedings of 3rd International Conference on Advanced Computing Technologies and Applications, ICACTA 2023. https://doi.org/10.1109/ICACTA58201.2023.10392776

[42] Mitsunaga, T. (2023). Heuristic Analysis for Security, Privacy and Bias of Text Generative AI: GhatGPT-3.5 case as of June 2023. 2023 IEEE International Conference on Computing, ICOCO 2023, 301-305. https://doi.org/10.1109/ICOCO59262.2023.10397858

[43] Murugesan, G. S., & Viswanathan, S. A. (2024). Enhancing Human-Machine Interaction: A Study on Deployment of LLM and Gen AI Hybrid Models in Responsible Humanoids for Human Assistance. 2024 15th International Conference on Computing Communication and Networking Technologies, ICCCNT 2024. https://doi.org/10.1109/ICCCNT61001.2024.10724819

[44] Narayan, M., Pasmore, J., Sampaio, E., Raghavan, V., Maity, S., Waters, G., & Howard, A. M. (2025). Mitigating Bias in Large Language Models Through Culturally-Relevant LLMs. ETHICS 2025 - 2025 IEEE International Symposium on Ethics in Engineering, Science, and Technology: Emerging Technologies, Ethics, and Social Justice. https://doi.org/10.1109/ETHICS65148.2025.11098204

[45] Obeng, C. P., Narasinghe, N. M., Striker, R., & Vazquez, E. A. (2025). Security Framework for Agentic Home AI in Preventive Healthcare: Cyber Threats Worth Noting. 2025

Cyber Awareness and Research Symposium, CARS 2025. https://doi.org/10.1109/CARS67163.2025.11337832

[46] Pawana, I. W. A. J., Purnama, F., Linawati, Kim, Y. J., & You, I. (2025). Generative AI for Cybersecurity: LLM-Driven Attack Dataset Augmentation in Web API Detection. 2025 1st International Conference on Consumer Technology, ICCT-Pacific 2025. https://doi.org/10.1109/ICCT-Pacific63901.2025.11012893

[47] Perera, H., Lee, S. U., Liu, Y., Xia, B., Lu, Q., Zhu, L., Cairns, J., & Nottage, M. (2026). Achieving Responsible AI through ESG: Insights and Recommendations from Industry Engagement. IEEE Software, 1-10. https://doi.org/10.1109/MS.2026.3689573

[48] Pleskach, M. (2025). Legal Loopholes in the EU Artificial Intelligence Act: Addressing Disinformation and Human Rights Protection - The Need for Further Refinement. Proceedings - International Conference on Advanced Computer Information Technologies, ACIT, 852-858. https://doi.org/10.1109/ACIT65614.2025.11185881

[49] Sachan, S., Miller, T., & Nguyen, M. P. (2025). Responsible LLM Deployment for High-Stake Decisions by Decentralized Technologies and Human-AI Interactions. ICHMS 2025 - 5th IEEE International Conference on Human-Machine Systems: AI and Large Language Models: Transforming Human-Machine Interactions, 99-104. https://doi.org/10.1109/ICHMS65439.2025.11154208

[50] Samadhiya, V., Sagar, R., Sinha, M., & Menon, S. (2025). Findings from Building AI Risk Mitigation Platform for Enterprise-Grade Solutions. International Conference on Electrical, Computer, and Energy Technologies, ICECET 2025. https://doi.org/10.1109/ICECET63943.2025.11472138

[51] Sankaran, S. (2025). Enhancing Trust Through Standards: A Comparative Risk-Impact Framework for Aligning ISO AI Standards with Global Ethical and Regulatory Contexts. International Conference on Artificial Intelligence, Computer, Data Sciences, and Applications, ACDSA 2025. https://doi.org/10.1109/ACDSA65407.2025.11166403

[52] Saqib, H. M., & Amin, H. (2026). Comparative analysis of AI regulation for fintech cybersecurity and privacy in the European Union and Qatar. Discover Artificial Intelligence. https://doi.org/10.1007/s44163-025-00736-5

[53] Sas, M. (2024). Unleashing Generative Non-Player Characters in Video Games: An AI Act Perspective. 2024 IEEE Gaming, Entertainment, and Media Conference, GEM 2024. https://doi.org/10.1109/GEM61861.2024.10585442

[54] Shah, S. B., Thapa, S., Acharya, A., Rauniyar, K., Poudel, S., Jain, S., Masood, A., & Naseem, U. (2025). Navigating the Web of Disinformation and Misinformation: Large Language Models as Double-Edged Swords. IEEE Access, 13, 169262-169282. https://doi.org/10.1109/ACCESS.2024.3406644

[55] Shenoy, P., & Saarela, M. (2026). AI Literacy and Governance in Indian Education: Mapping Policy Approaches and Institutional Initiatives. International Conference on Artificial Intelligence, Computer, Data Sciences, and Applications, ACDSA 2026. https://doi.org/10.1109/ACDSA67686.2026.11467788

[56] State, G., & Olar, D. (2025). Bias Detection in AI Recruitment. Transparency, Accountability and Empirical Assessment Using Synthetic CVs and Rapid Testing. Proceedings - RoEduNet IEEE International Conference. https://doi.org/10.1109/RoEduNet68395.2025.11208334

[57] Tabassum, A., Elmahjub, E., Padela, A. I., Zwitter, A., & Qadir, J. (2025). Generative AI and the Metaverse: A Scoping Review of Ethical and Legal Challenges. IEEE Open Journal of the Computer Society, 6, 348-359. https://doi.org/10.1109/OJCS.2025.3536082

[58] Tran, Q., Salg, J., Shpileuskaya, K., Wang, Q., Putzar, L., & Blankenburg, S. (2025). Bridging AI and Regulation: Large Language Models for Documentation Compliance Check. 1-10. https://doi.org/10.1109/ijcnn64981.2025.11229064

[59] Turaga, V. S. A. I. P., Pahi, T., Tjoa, S., Siami-Namini, S., & Namin, A. S. (2025). CO2 (Co-Compliance Officer): An LLM-Based Ontology-Driven Methodology for Generating Knowledge Graphs and AI Compliance Checking. IEEE Access, 13, 210576-210606. https://doi.org/10.1109/ACCESS.2025.3639228

[60] Tuzov, V., & Lin, F. (2024). Two paths of balancing technology and ethics: A comparative study on AI governance in China and Germany. Telecommunications Policy, 48(10). https://doi.org/10.1016/j.telpol.2024.102850

[61] Vagnoni, S., & Palmirani, M. (2025). Detecting Transitional Provisions in EU Legislation with Hybrid AI for Better Regulation. International Conference on eDemocracy and eGovernment, ICEDEG, (2025), 190-197. https://doi.org/10.1109/ICEDEG65568.2025.11081693

[62] Vanam, L., Kundurthy, O. H., & Ghadiyaram, R. (2025). A Framework for Quantifying Ethical and Regulatory Risks in Big Data Analytics and AI-Assisted Banking Software Development. 1-6. https://doi.org/10.1109/asiancon66527.2025.11280739

[63] Voria, G., Scala, B., Todisco, L., Venditto, C., Giordano, G., Catolino, G., & Palomba, F. (2026). Fair and square? Evaluating fairness of LLM-generated synthetic datasets. Information and Software Technology, 191, 107980. https://doi.org/10.1016/j.infsof.2025.107980

[64] Xavier, S., Yogarajah, P., Incorvaia, G., & Chaurasia, P. (2025). Towards Responsible AI: Evaluating Large Language Models Across Trustworthy Dimensions. Irish Signals and Systems Conference: Signalling Our Strength, ISSC 2025, 1-6. https://doi.org/10.1109/ISSC67739.2025.11290002

[65] Yamani, A., Baslyman, M., & Ahmed, M. (2025). Are We Aligned? A Preliminary Investigation of the Alignment of Responsible AI Values between LLMs and Human Judgment.

Proceedings - 2025 2nd IEEE/ACM International Conference on AI-Powered Software, AIware 2025, 133-142. https://doi.org/10.1109/AIware69974.2025.00022

[66] Yin, R., & Liu, X. (2025). From Technological Alienation to Value Regression: Ethical Regulation of Algorithmic Bias in the Post-Truth Era. Proceedings - 2025 IEEE 10th International Conference on Smart Cloud, SmartCloud 2025, 32-37. https://doi.org/10.1109/SmartCloud66068.2025.00008

[67] Zhu, H., Liang, S., Chen, B., Wang, S. L., Zhang, Z., & Ding, W. (2026). Auditing Partial Dataset Usage in Large Language Models via Fuzzy Membership Aggregation. IEEE Transactions on Fuzzy Systems. https://doi.org/10.1109/TFUZZ.2026.3665818

[68] Zhuang, X., & Wu, Y. (2025). Large Language Models in Corporate Investment: An Analysis of ChatGPT's Transformative Applications, Strategic Impacts, and Future Opportunities. IT Professional, 27(2), 21-27. https://doi.org/10.1109/MITP.2025.3543555

[69] Page, M. J., McKenzie, J. E., Bossuyt, P. M., Boutron, I., Hoffmann, T. C., Mulrow, C. D., Shamseer, L., Tetzlaff, J. M., Akl, E. A., Brennan, S. E., Chou, R., Gluud, C., Mayo-Wilson, E., McDonald, S., McGuinness, L. A., Stewart, L. A., Thomas, J., Tricco, A. C., Welch, V. A., ... Moher, D. (2021). The PRISMA 2020 statement: An updated guideline for reporting systematic reviews. BMJ, 372, n71. https://doi.org/10.1136/bmj.n71

[70] Vaswani, A., Shazeer, N., Parmar, N., Uszkoreit, J., Jones, L., Gomez, A. N., Kaiser, Ł., & Polosukhin, I. (2017). Attention is all you need. Advances in Neural Information Processing Systems, 30. https://arxiv.org/abs/1706.03762