

## DAÑOS Y RIESGOS DE LA INTELIGENCIA ARTIFICIAL: ENFOQUE INTEGRAL MULTIDISCIPLINAR

**Fernando Vegas-Fernández**  
*Ingeniero de Caminos, Canales y Puertos. Ph.D.*

### RESUMEN

Este estudio identifica y categoriza los riesgos de la Inteligencia Artificial (IA), evaluando su impacto y probabilidad, priorizándolos. Los riesgos de la IA son multifacéticos tanto en su origen como en sus campos de actuación, y abarcan facetas diversas como la tecnológica, ética, legal, social o humanística. Esto exige una lista de riesgos de IA multidisciplinar, priorizando y cuantificando cada riesgo y la situación de conjunto. Tomando como base la revisión de la literatura y el análisis de expertos, se utiliza la metodología del Índice de Importancia Relativa (IIR) y la metodología del Factor de Visibilidad para cuantificar los riesgos. El estudio identifica los mayores riesgos y valora el nivel de riesgo global como moderado, alertando de la necesidad de una regulación y gestión adecuadas de la IA para mitigar sus riesgos y garantizar sus beneficios para la sociedad.

### 1. INTRODUCCIÓN

La inteligencia artificial (IA) se concibió inicialmente en 1956, pero su verdadera popularidad alcanzó su punto máximo en noviembre de 2022 cuando se lanzó ChatGPT (Schwarz 2024). Este evento alertó sobre los posibles riesgos de esta nueva tecnología, y desde entonces se han publicado miles de artículos académicos, libros y procedimientos de congresos relacionados con los riesgos de la IA. Sin embargo, la mayoría de ellos se centran en un área específica y prestan especial atención a ciertos tipos de riesgo, por lo que es difícil obtener una visión de conjunto de la situación global. Además, no valoran su impacto y probabilidad, y no distinguen el origen, efecto y consecuencia de los riesgos, aspecto requerido por la norma ISO 31000 (International Organization for Standardization (ISO) 2018) para evaluar riesgos.

El concepto de «inteligencia artificial» ha evolucionado significativamente desde su concepción original, y las aplicaciones actuales de IA suelen tener un alcance más limitado que el que se pensaba inicialmente (Campuzano Fernández and Vega Vicente 2021; Dick 2019). La IA contemporánea y los agentes de IA generativa son esencialmente los productos de la ejecución de un código que genera resultados aceptados como confiables debido a su comunicación formal que se asemeja a la interacción humana; pero estos productos no superan la prueba de inteligencia de Turing (Restrepo Echavarría 2025)

#### 1.1. Definición de inteligencia artificial

El concepto de inteligencia artificial ha sufrido cambios significativos a lo largo del tiempo. Siguiendo la pregunta seminal de Turing, «¿Pueden pensar las máquinas?» (Turing 1950), la noción dominante de IA en la década de 1960 era «la capacidad de construir una máquina que se comporte de tal manera que, si un humano realizara el mismo comportamiento, se consideraría inteligente» (Campuzano Fernández and Vega Vicente 2021). Esta definición es muy diferente a la actual de Garner:

«Aplicar análisis avanzados y técnicas basadas en la lógica, incluido el aprendizaje automático o *machine learning* (ML), para interpretar eventos, respaldar y automatizar decisiones, y tomar acciones» (Gartner 2025). Resulta llamativo ver cómo han cambiado estas definiciones con el tiempo.

La idea original de desarrollar sistemas de inteligencia artificial languideció durante medio siglo, entre otras razones, debido a la falta de capacidad informática adecuada; este período se conoce como el «invierno de la IA» (Göçen and Asan 2023; Siriwardhane 2023). En su lugar, avanzó el *machine learning*, o «aprendizaje automático», al que Gartner se refiere como un conjunto de técnicas y algoritmos (Brethenoux 2020), y lo define de esta manera: «Los algoritmos avanzados de aprendizaje automático están compuestos por muchas tecnologías (como el aprendizaje profundo, las redes neuronales y el procesamiento del lenguaje natural), utilizadas en el aprendizaje supervisado y no supervisado, que operan guiadas por lecciones desde la información existente» (Gartner 2025). Los avances en el aprendizaje automático progresaron durante más de 20 años (Sheikh et al. 2023), y luego parecieron desvanecerse en un segundo plano tras la aparición de los primeros algoritmos de IA.

La definición original de IA de John McCarthy fue: «Es la ciencia y la ingeniería de hacer máquinas inteligentes, especialmente programas informáticos inteligentes» (McCarthy 2007). Aunque McCarthy pensaba en hacer que las máquinas fueran inteligentes a través de programas informáticos, se podría esperar que la IA se comportara como un humano inteligente, y esta idea es seguida por otros autores con definiciones como «Una rama de la informática que se ocupa de la simulación del comportamiento inteligente en las computadoras, o la capacidad de una máquina para imitar el comportamiento humano inteligente» (Wharton University of Pennsylvania 2021), «La inteligencia artificial es el campo de estudio dedicado a hacer que las máquinas sean inteligentes» (Scharre 2021), o «La inteligencia artificial se puede definir como una rama de la informática que se ocupa de simular el comportamiento humano en las máquinas» (Faghiri 2023). El error conceptual, posiblemente, consistió en confundir los conceptos de máquina y programa, pero prevalece la idea de imitar intelectualmente al hombre.

Desde 2010 el concepto asociado con el término «inteligencia artificial» ha evolucionado, y Dick (2019) expresa decepción con el nuevo enfoque de la inteligencia artificial, que solo tiene un parecido superficial con el enfoque inicial en términos de nombre. En su definición más amplia la IA se equipara ahora con los algoritmos, y representa la imitación por parte de las computadoras de la inteligencia inherente a los humanos (Sheikh et al. 2023). Para algunos autores «estos algoritmos tienen características que los hacen cada vez más adaptables sin intervención humana» (Teixeira et al. 2022)

La definición proporcionada en 2019 por un grupo de expertos de la Comisión Europea fue: «Sistemas que muestran un comportamiento inteligente analizando su entorno y tomando acciones -con cierto grado de autonomía- para lograr objetivos específicos» (High-Level Expert Group on Artificial Intelligence 2019), y Sheikh señala que conceptos como «cierto grado de autonomía» siguen siendo algo vagos. Una definición más reciente de la Comisión Europea es: «“sistema de IA” significa un sistema basado en máquinas que está diseñado para operar con diferentes niveles de autonomía y que puede exhibir adaptabilidad después de la implementación, y que, para objetivos explícitos o implícitos, infiere, a partir de la información que recibe, cómo generar resultados como predicciones, contenido, recomendaciones o decisiones que pueden influir en entornos físicos o virtuales» (The European Parliament and the Council 2024).

Una definición más corta de IA la describiría como «un conjunto de algoritmos que se alimentan con datos para analizarlos y procesarlos» (Torres Cadillo 2023), lo que indica que la IA es fundamentalmente software, aunque no cualquier tipo de software (Álvarez García and Tahiri Moreno 2023). Un nuevo término, Inteligencia Artificial Generativa, se refiere a las habilidades multimedia de IA: «la capacidad de generar contenido como texto, imágenes, sonido, video y código» (Li and Ling 2025). No obstante, está extendida la idea de que este concepto se refiere a los agentes de inteligencia artificial, como ChatGPT, Copilot, etc.

De ahora en adelante, este artículo tratará la IA convencional y la IA generativa indistintamente, adoptando la definición de Torres Cadillo como la referencia principal para la IA. Esta elección no es restrictiva, dado que no interviene ni condiciona en modo alguno el estudio realizado; la mayoría de los documentos consultados ni siquiera definen qué entienden por Inteligencia Artificial.

## 1.2. Evaluación de riesgos de IA

La Organización Internacional de Normalización define el riesgo como «el efecto de la incertidumbre sobre los objetivos» (International Organization for Standardization (ISO) 2018), y explica que «el riesgo suele expresarse en términos de fuentes de riesgo, eventos potenciales, sus consecuencias y su probabilidad»; también dice que «una consecuencia puede ser cierta o incierta y puede tener efectos directos o indirectos positivos o negativos sobre los objetivos». El Project Management Institute define el riesgo como «Un evento o condición incierta que, si ocurre, tiene un efecto positivo o negativo en uno o más objetivos del proyecto» (Project Management Institute 2021).

De acuerdo con estas definiciones, un riesgo debe definirse describiendo su origen, eventos y consecuencias, describiendo su impacto en los objetivos y la probabilidad de que ocurra. Por ejemplo, en el caso del sesgo de IA, el origen puede estar en los datos o en los algoritmos, el efecto sería resultados sesgados que no serían objetivos ni confiables, y la consecuencia de este fallo dependería de para qué se utilice el sistema de IA, dado que podría conducir a tomas de decisiones, selección de personal, análisis, resúmenes o evaluaciones de perfiles incorrectos y sesgados. Este estudio utilizará el término «riesgo» de forma genérica para acomodarse al uso de este término en la literatura, pero examinará los orígenes de los riesgos de la IA y especificará si los riesgos identificados son efectos, consecuencias o ambos.

La definición rigurosa de un riesgo requiere evaluar, como mínimo, su impacto y probabilidad, cualitativa, cuantitativa o semicuantitativamente (International Organization for Standardization (ISO) 2019). Aunque algunos estudios sobre los riesgos de la IA se adhieren a este estándar (Campbell and Jovanović 2024; Kilian et al. 2023), la mayoría no lo hacen. La definición del riesgo asociado a la inteligencia artificial presenta desafíos significativos debido a dos razones principales. En primer lugar, evaluar su impacto requiere una comprensión de las consecuencias específicas asociadas con el uso de un sistema de IA; por ejemplo, las consecuencias del sesgo difieren significativamente entre un sistema de selección de candidatos para un empleo y un sistema de recomendación de viajes. En segundo lugar, estimar con precisión la probabilidad de los riesgos exige datos históricos o experiencias acumuladas, que son difíciles de lograr debido a la presencia relativamente reciente de esta tecnología. Algunos autores evalúan los impactos (Teixeira et al. 2022) y las probabilidades (Curtis et al. 2022) en base a encuestas de opinión y actitud hacia los sistemas de inteligencia artificial.

Además, realizar una evaluación exhaustiva del riesgo de IA implicaría estimar los costes probables de cada riesgo y el coste total del riesgo del escenario, ya sea a través de simulaciones de Monte Carlo u otro método simplificado (Vegas-Fernández 2022). Esto permite desarrollar un plan de mitigación y la estimación de una reserva de contingencia, asegurando que el sistema de IA mantenga la resiliencia necesaria.

La conciencia sobre el riesgo de la IA ha aumentado significativamente en los últimos años. El *Global Risk Report*, publicado cada año por el Foro Económico Mundial, muestra una clara evolución: la 17ª edición no incluyó ningún riesgo tecnológico entre los riesgos más graves a escala mundial para los próximos 10 años (World Economic Forum 2022), y la 18ª edición ni siquiera menciona la «inteligencia artificial» (World Economic Forum 2023). La 19ª edición advierte sobre los resultados adversos de la inteligencia artificial e incluye «información errónea y desinformación» junto con «inseguridad cibernética», aunque no explícitamente asociados con el concepto de IA (World Economic Forum 2024). Finalmente, el último informe, la 20ª edición, incluye «resultados adversos de las tecnologías de inteligencia artificial» entre los cinco principales riesgos identificados por la Encuesta de Opinión Ejecutiva en muchos países (World Economic Forum 2025).

Los informes del índice de inteligencia artificial de Stanford indican que el número de incidentes relacionados con el mal uso ético de la IA está aumentando rápidamente (Stanford University 2023, 2024, 2025). En 2023, se reportaron 123 incidentes de este tipo, un aumento de 32,3 puntos porcentuales con respecto a 2022; En 2024, este número aumentó a 233, un 56,4% más que en 2023. El informe de 2023 indica que la ciberseguridad (59%) es el riesgo más relevante generado por la tecnología de IA, seguido del cumplimiento normativo (45%), la privacidad personal/individual (40%) y la inteligibilidad (37%). El informe de 2024 afirma que los riesgos extremos de la IA son difíciles de analizar, discutiendo la discriminación algorítmica frente a las amenazas existenciales a largo plazo y los riesgos de seguridad. Sin embargo, el informe de 2025 ya no incluye riesgos existenciales o catastróficos.

## 2. OBJETIVO

Este artículo quiere ofrecer una visión general completa de los posibles daños y riesgos asociados con la inteligencia artificial desde diversas perspectivas. Su objetivo es compilar una lista de los riesgos de IA más pertinentes, evaluarlos tanto cualitativa como cuantitativamente, priorizarlos y valorar el riesgo general en términos cuantitativos.

## 3. MÉTODOS

Para lograr el objetivo buscado se combinan varios métodos apoyados en una revisión de la literatura, análisis de expertos y casos de estudio sobre el desempeño de los agentes de IA.

Con el primer método se extrae información de bases de datos académicas para crear una lista priorizada de riesgos de IA basada en documentos científicos publicados. Este método requiere un análisis profundo de las palabras clave (Hobbs, 2002), en este caso de los riesgos de IA, e implica tres pasos (Vegas-Fernández 2020b):

El primer paso es identificar los documentos pertinentes mediante una búsqueda sistemática; para ello se ha buscado «*artificial intelligence risks*» en bases de datos mediante motores de búsqueda en línea, de forma similar a otros autores (Bas Graells et al. 2024; Rupino da Cunha and Estima 2023; Teixeira et al. 2022; Villamar Vasquez et al. 2024; Wirtz et al. 2022). Para ello se han considerado fuentes académicas significativas como Web of Science, Scopus, Google Scholar, Elsevier, ResearchGate, ScienceDirect y Springer. Google Scholar arrojó más de 4.800.000 documentos, mientras que las otras bases de datos proporcionaron hasta 80.000 resultados cada una con este criterio de búsqueda. Estos resultados se filtran inicialmente seleccionando los artículos más relevantes de cada base de datos según el criterio de relevancia del motor de búsqueda, y luego eligiendo los documentos más recientes, priorizando los publicados desde 2022. Finalmente, se descartan los duplicados y se excluyen los artículos que no se centran directamente en los riesgos de la IA.

El segundo paso consiste en recuperar y revisar los documentos seleccionados para permitir un filtrado adicional mediante sus títulos, palabras clave y resúmenes, excluyendo los no enfocados al análisis y estudio de los riesgos de IA; con los que quedan finalmente se elabora una lista completa de riesgos relacionados con la IA a través de un proceso iterativo partiendo de una lista inicial, similar al enfoque utilizado por otros autores (Bas Graells et al. 2024). Este proceso es iterativo y convergente, por lo que el resultado final no depende de esta lista inicial.

El tercer paso asigna un peso a cada documento en función de su relevancia y utiliza un coeficiente, el Índice de Importancia Relativa (IIR), para estimar la relevancia de cada riesgo identificado. El IIR se calcula como la relación entre la suma ponderada de documentos que citan un riesgo de IA determinado y la suma ponderada máxima, expresada como porcentaje, un método utilizado también en otros trabajos publicados (Alashwal and Al-Sabahi 2019; Jarkas and Haupt 2015; Nagalla et al. 2018). Este índice no mide la gravedad del riesgo, sino su importancia según los investigadores, y considera el peso de cada documento según su tipología, siendo el más alto para tesis doctorales, regulaciones y artículos publicados en revistas académicas, y el más bajo para contenidos web

publicados por instituciones; también tiene en cuenta la fecha de publicación y si está destinado a un país o campo específico (Vegas-Fernández 2020b).

El segundo método, análisis de expertos, aporta un punto de vista diferente más alineado con los profesionales de las empresas. Se apoya en un grupo de 10 expertos integrado principalmente por directores de sistemas de información (CIO), directores de seguridad de la información (CISO) y especialistas en IA. Comparando los resultados de la revisión de la literatura con el análisis de expertos se obtiene un mapa integral de riesgos; adicionalmente, estos expertos evaluaron a través de entrevistas personales y de trabajo en grupo la naturaleza, impacto y probabilidad de cada riesgo incluido en la lista final.

Una visión cuantitativa del riesgo es esencial para comprender su relevancia; para lograr esto, se aplica a este mapa integral de riesgos IA el método del Factor de Visibilidad (Vegas-Fernández 2020a). Este método asigna una severidad a cada riesgo, su Factor de Visibilidad, en una escala lineal de 0 a 100, y proporciona una visión del riesgo global mediante su propio impacto, probabilidad y Factor de Visibilidad. Este análisis tiene como objetivo crear conciencia sobre los peligros potenciales que puede representar la IA si no se regula y gestiona adecuadamente.

Finalmente, se presentan varios casos de estudio sencillos con agentes de IA para ilustrar y validar algunos de los riesgos descritos en este análisis. Estos agentes están transformando los flujos de trabajo actuales en las empresas, y los departamentos de TI están desplegando agentes de IA personalizados y adaptados a sus necesidades, según un reciente informe de Setesca-LiceoTic basado en una encuesta realizada en febrero de 2025 con 56 participantes (LiceoTic 2025).

#### **4. REVISIÓN DE LA LITERATURA**

El número anual de publicaciones sobre IA en todo el mundo supera 240.000 y sigue creciendo, particularmente en el campo de la ética de la IA médica (Stanford University 2023, 2024, 2025). La mayor parte de los estudios consultados examinan los riesgos de la IA en campos particulares, como la medicina y la biología, describiendo un número reducido de riesgos, 2 a 4 (Law et al. 2024; Nord-Bronzyk et al. 2025; Nwebonyi and McKay 2024), aunque existen algunos artículos que identifican un número significativamente mayor de riesgos, entre 8 y 12 (Macrae 2025; Ranjbar et al. 2024). Otras publicaciones adoptan una perspectiva más amplia en sus estudios sobre el riesgo de la IA, como el Reglamento de Inteligencia Artificial (The European Parliament and the Council 2024). En promedio, las publicaciones identifican sólo 6 riesgos de IA cada una, siendo 16 el número máximo de riesgos identificados en una de las publicaciones revisadas.

Las publicaciones, en general, no distinguen claramente entre el origen, efecto y consecuencia de los riesgos de la IA, pero algunas clasifican los riesgos enfatizando la taxonomía, aunque no las fuentes específicas de riesgo. Por ejemplo, la Wharton University of Pennsylvania (2021) clasifica los riesgos de IA en cuatro grupos: desafíos relacionados con los datos, ataques de IA/ML, problemas de fiabilidad y preocupaciones por el cumplimiento normativo; Arda (2024) considera las presiones geopolíticas, el uso malicioso, los problemas éticos, sociales y ambientales, junto con las violaciones de la privacidad y la confianza; Brown (2024) destaca tres dominios: riesgos relacionados con la tecnología, riesgos financieros y del mercado de inversión, y riesgos que afectan significativamente la experiencia del usuario. Para otros, las categorías son sólo los tipos de riesgo (González-Arias and López-García 2024).

La revisión de la literatura revela el aumento de algunos riesgos que anteriormente habían pasado inadvertidos, incluidos los sesgos (Payton 2024), la discriminación (Borges Fortes et al. 2022), la falta de ética (Cho and Hooi 2023), el uso malicioso y delictivo (Chen and Chen 2024), los problemas con los derechos humanos (Wirtz et al. 2022), los desafíos en la toma de decisiones (Hudson and Morgan 2024), las preocupaciones sobre la privacidad (Appio et al. 2023), la falta de transparencia (Cabrol and Sánchez Ávalos 2021) y la responsabilidad poco clara (Herrera de las Heras 2022). Los ciberataques y

otros tipos de acciones basadas en la IA para atacar los sistemas de información y las infraestructuras críticas han causado una creciente preocupación desde 2022 (Carvalho and Ivanov 2024; McCullum and Contreras Mancera 2025; Schuett 2024). Algunos autores destacan el riesgo de que las máquinas puedan tomar un control excesivo, lo que sería «quizás un riesgo existencial para la humanidad» (Müller 2016).

## 5. ANÁLISIS

El análisis realizado, siguiendo la metodología propuesta, consta de tres partes. En la primera se obtiene una lista priorizada a partir de la literatura publicada; en la segunda se obtiene un mapa de riesgos según los expertos consultados; en la tercera se analizan casos de uso que ilustran algunos de los riesgos detectados mediante los métodos anteriores.

### 5.1. Extracción inteligente de información de documentos académicos

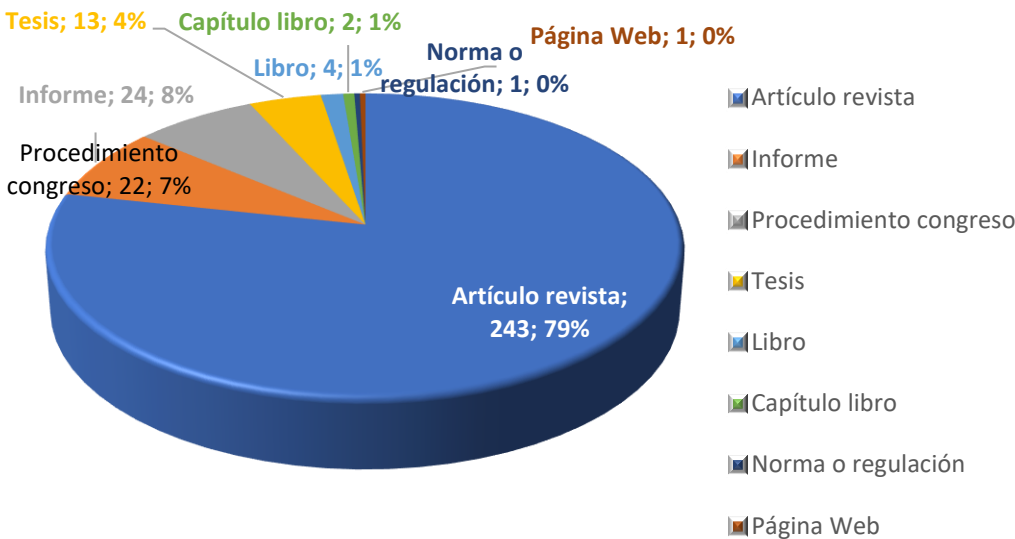
La selección inicial de documentos a través de búsquedas automatizadas en bases de datos académicas proporcionó 524 documentos, a los que se añadieron 15 documentos seleccionados manualmente. Tras excluir los duplicados y artículos no centrados en la investigación de los riesgos de la IA, quedó un total de 310 documentos para la extracción exhaustiva de datos. Más del 79% de los documentos son artículos de revistas publicados en revistas indexadas, y la mayoría fueron publicados recientemente. La tabla 1 muestra su distribución temporal.

Tabla 1. Distribución anual

| Año  | 2007 | 2016 | 2018 | 2019 | 2020 | 2021 | 2022 | 2023 | 2024 | 2025 | Total |
|------|------|------|------|------|------|------|------|------|------|------|-------|
| Docs | 1    | 1    | 4    | 6    | 4    | 18   | 21   | 52   | 170  | 33   | 310   |

La figura 1 muestra la proporción de tipos de documentos entre los 310 seleccionados. El 79 % de los mismos, 243, son artículos publicados en revistas científicas; los informes emitidos por entidades de prestigio, junto con los procedimientos de conferencias y tesis académicas, 59, totalizan un 19 %.

Figura 1. Tipos de documentos dentro de la muestra analizada



La mayoría de los documentos, 268, no centran su investigación en un país concreto (87%), aunque 42 de ellos sí lo hacen: 14 en Europa, 5 en España, 5 en México, 4 en Colombia, 3 en China, y uno o dos artículos en otros países. Las principales áreas de interés dentro de esta colección de documentos son los aspectos sociales, legales y educativos. La tabla 2 ilustra el enfoque temático del conjunto de documentos.

*Tabla 2. Campo de interés*

| <b>Campo</b>         | <b>Documentos</b> |
|----------------------|-------------------|
| Social               | 90                |
| Legal                | 49                |
| Educativo            | 38                |
| Tecnología           | 32                |
| Humanismo            | 27                |
| Salud                | 21                |
| Gobernanza           | 20                |
| Empresarial          | 14                |
| Bancos y finanzas    | 9                 |
| Militar              | 7                 |
| Turismo              | 1                 |
| Economía             | 1                 |
| Seguros              | 1                 |
| <b>Total general</b> | <b>310</b>        |

Los 310 documentos recuperados fueron analizados y etiquetados a través de un proceso iterativo para identificar los conceptos clave, es decir, los riesgos que conlleva la IA. Para elaborar la lista inicial de riesgos de IA se preguntó a varios agentes de IA: «describe los riesgos de inteligencia artificial más relevantes». Alternativamente, se podría haber empleado cualquier método alternativo, como cuestionarios, entrevistas o sesiones Delphi, dado que el proceso iterativo de búsqueda e identificación de etiquetas es convergente. En consecuencia, los documentos analizados pueden proporcionar nuevas etiquetas (riesgos) y las etiquetas existentes que no aparecen en la literatura consultada pueden eliminarse al finalizar.

La pregunta indicada se planteó a Copilot, ChatGPT, Perplexity y Gemini el 22 de febrero de 2025. Con sus respuestas se identificaron hasta 33 riesgos distintos de IA, aunque ninguna de las fuentes informó más de 27 (Perplexity). Gemini propuso 22 riesgos de IA, mientras que Copilot y ChatGPT mostraron 19 y 17, respectivamente. Adicionalmente se consultó la lista de riesgos de IA presentada Forbes (Marr 2023), que contiene 22 elementos, como una referencia adicional con objeto de contrastar la riqueza o pobreza de las respuestas de los agentes de IA.

En la tabla 3 se presentan los riesgos identificados por estos agentes de inteligencia artificial, junto con la lista de riesgos de IA publicada por Forbes. Los datos se presentan ordenados para facilitar la comparación entre los resultados correspondientes a cada origen.

*Tabla 3. Riesgos de la IA según los agentes de IA y Forbes en 2025*

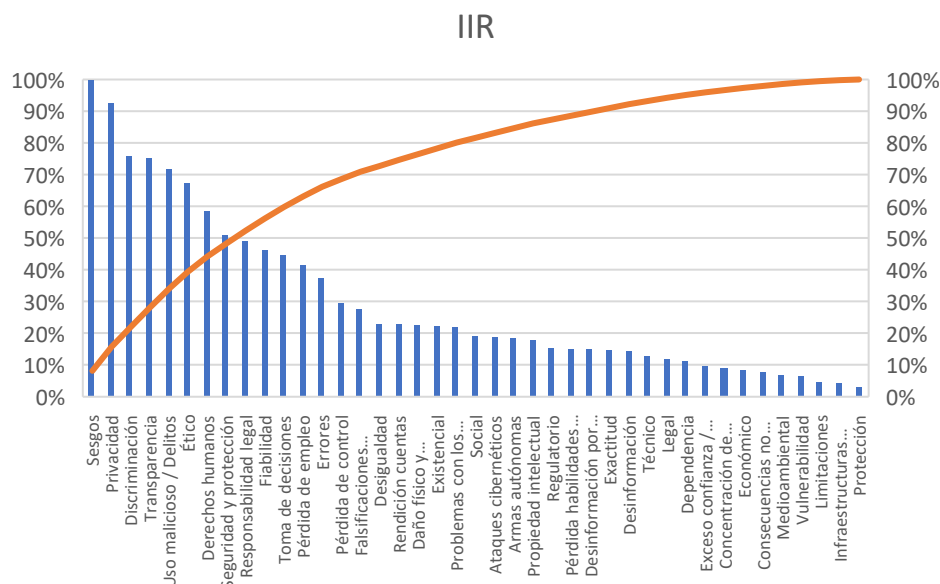
| <b>Copilot</b>      | <b>ChatGPT</b>  | <b>Perplexity</b>   | <b>Gemini</b>       | <b>Forbes</b>       |
|---------------------|-----------------|---------------------|---------------------|---------------------|
| Armas autónomas     | Armas autónomas | Responsabilidad     |                     |                     |
| Sesgo               | Sesgo           | Armas autónomas     | Armas autónomas     | Armas autónomas     |
|                     |                 | Sesgo               | Sesgo               | Sesgo               |
|                     |                 | Ciberataques        | Ciberataques        |                     |
|                     |                 |                     | Calidad de datos    |                     |
| Toma decisiones     |                 | Toma decisiones     |                     | Toma decisiones     |
| Dependencia         | Dependencia     | Dependencia         |                     | Dependencia         |
| Discriminación      | Discriminación  | Discriminación      | Discriminación      | Discriminación      |
|                     | Desinformación  | Desinformación      | Desinformación      | Desinformación      |
| Medioambiental      |                 | Medioambiental      |                     |                     |
| Errores             |                 |                     | Errores             |                     |
| Ético               |                 | Ético               | Ético               | Ético               |
|                     | Existencial     | Existencial         | Existencial         | Existencial         |
|                     | Falsificaciones | Falsificaciones     |                     | Falsificaciones     |
| Capacidades humanas |                 | Capacidades humanas | Capacidades humanas | Capacidades humanas |
|                     |                 | Conexiones humanas  | Conexiones humanas  | Conexiones humanas  |
| Desigualdad         | Desigualdad     | Desigualdad         | Desigualdad         | Desigualdad         |

| Copilot  | ChatGPT  | Perplexity  | Gemini   | Forbes   |
|--|--|---|--|--|
|  |  |   | Integración  |  |
| Desplazamiento y pérdida de empleo<br>Legal y regulatorio      | Desplazamiento y pérdida de empleo<br><br>Pérdida de control<br>Pérdida de habilidades humanas | Propiedad intelectual<br>Desplazamiento y pérdida de empleo<br>Legal y regulatorio    | Desplazamiento y pérdida de empleo<br><br>Pérdida de control                                     | Propiedad intelectual<br>Desplazamiento y pérdida de empleo<br>Legal y regulatorio |
| Mal uso / delitos<br>Manipulación                              | Manipulación<br>Error información  | Mal uso / delitos<br>Manipulación<br>Error información                                | Mal uso / delitos<br>Manipulación<br>Error información<br>Dependencia excesiva /<br>Uso inseguro | Mal uso / delitos<br>Manipulación  |
| Concentración de poder<br>Privacidad<br>Seguridad y protección | Concentración de poder<br>Privacidad<br>Seguridad y protección                                 | Concentración de poder<br>Privacidad<br>Seguridad y protección<br>Polarización social | Privacidad   | Concentración de poder<br>Privacidad<br>Seguridad y protección                     |
| Transparencia<br>Consecuencias no deseadas                     | Transparencia  | Transparencia<br>Consecuencias no deseadas  | Transparencia<br>Consecuencias no deseadas   | Transparencia<br>Consecuencias no deseadas   |
| 19   | 17   | 27  | 22   | 22   |

La aplicación del método de extracción inteligente de información a los 310 documentos produjo una lista final de 41 riesgos de IA, que se incluye en la tabla 7 del Anexo 1 junto con la definición completa de cada riesgo. Esta lista se puede priorizar aplicando el Índice de Importancia Relativa, según su frecuencia de aparición en la literatura revisada; esta ordenación puede no coincidir con las prioridades de gravedad del riesgo, pero es un punto de partida basado en la importancia que se le da a estos riesgos en los estudios científicos (figura 2).

El sesgo y la falta de privacidad son los riesgos con IIR más alto. El sesgo o prejuicio se debe a errores sistemáticos en los datos de entrenamiento debidos a su obtención y su cantidad, y al diseño de los algoritmos, y afecta a los procesos de toma de decisiones, lo que puede dificultar un trato imparcial y justo, conduciendo al favoritismo y a la discriminación. La falta de privacidad incluye la violación del derecho de las personas a la confidencialidad, el anonimato y la seguridad de sus datos personales, incluido el derecho a dar su consentimiento y a ser informado sobre el uso de sus datos. También incluye la privacidad de los datos empresariales.

Figura 2. Lista ordenada de riesgos de IA y análisis de Pareto basado en el índice IIR

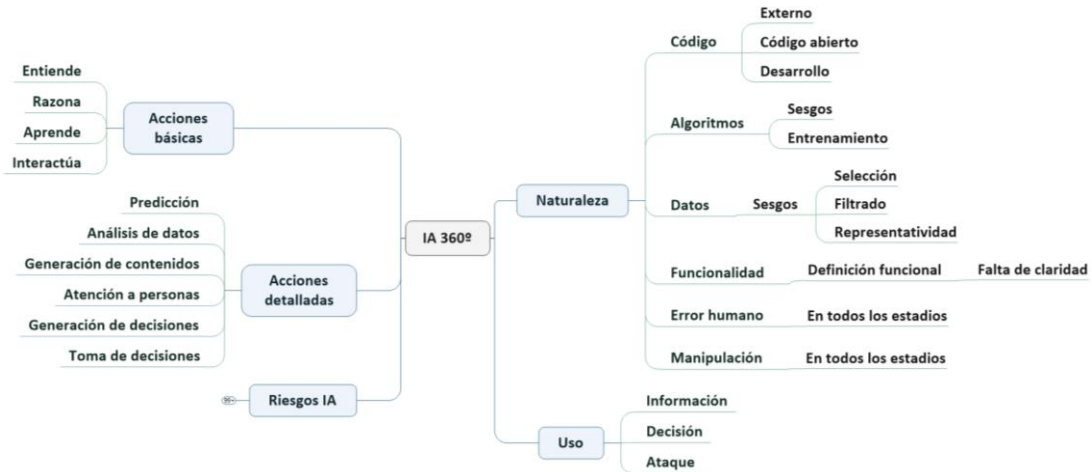




## 5.2. Análisis de expertos

El grupo de expertos se centró principalmente en el aspecto tecnológico porque aborda la raíz del riesgo, y porque está más próximo a su interés profesional; esto involucra los algoritmos, los datos utilizados para entrenarlos y cómo interactúan estos elementos. Al hacerlo, su objetivo era identificar las fuentes de riesgo en los sistemas de IA, lo que los llevó a la definición del marco básico descrito en la figura 3, incluido el uso de la IA y las tareas básicas.

*Figura 3. Fuentes de riesgo de IA basadas en el análisis de expertos*



En este análisis los expertos identifican las principales fuentes de riesgo: código, algoritmos, datos, definiciones funcionales, errores humanos y manipulación. También describen los usos potenciales básicos de la IA, incluidos el procesamiento de información, la toma de decisiones y los ataques, como posibles fuentes de problemas y riesgos futuros. Van más allá al describir las acciones básicas y detalladas que los sistemas de IA podrían abordar desde un punto de vista conceptual y, finalmente, identifican y describen los riesgos de la IA.

Comprender el riesgo que representa un sistema de IA requiere interactuar con él, pero la raíz del riesgo radica en cómo entiende, razona y aprende, y estos aspectos no son fácilmente controlables. Además, los riesgos surgen del uso del sistema de IA, motivo por el cual es fundamental analizar su propósito.

Una posible clasificación muy genérica de los casos de uso, según estos expertos, podría hacerse atendiendo a la funcionalidad básica que aportan, y se muestra en la

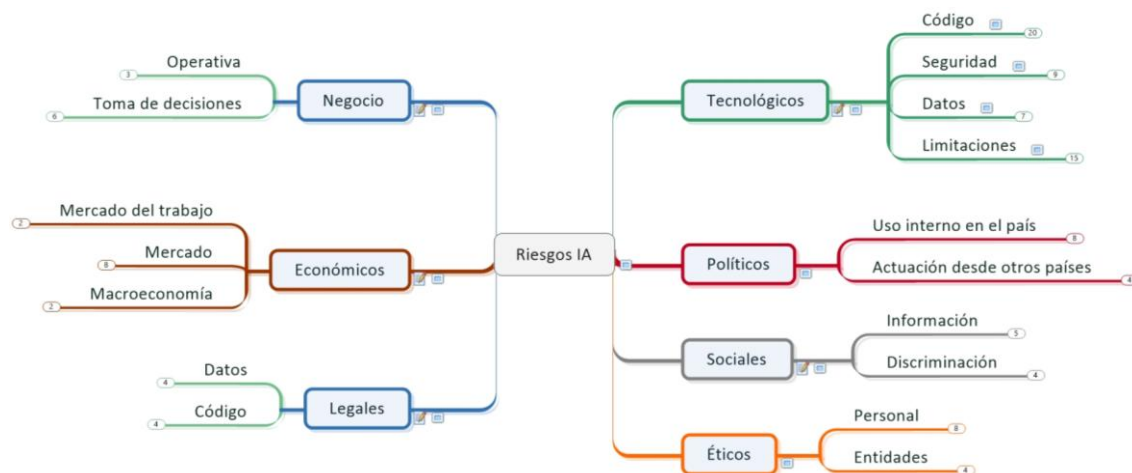
tabla 4 presentando los casos en orden ascendente según su nivel de intrusión en las actividades del ser humano. Es decir, por ejemplo, un uso de la IA para analizar se limita a estudiar la información y los datos haciendo clasificaciones y facilitando resúmenes y conclusiones, lo cual no es intrusivo; pero una toma de decisiones no atendida afecta directamente a personas, empresas o países porque determina acciones sobre ellos.

*Tabla 4. Casos de uso de IA según su nivel de intrusión*

|   |
|---|
| Análisis  |
| Predicción  |
| Generación de contenido (documentos, códigos, imágenes, voz, video, etc.) |
| Soporte al usuario/cliente  |
| Generación de información para la toma de decisiones                      |
| Toma de decisiones  |

La lista de riesgos de IA propuesta por el grupo de expertos contiene 108 riesgos diferentes, organizados en siete ramas, como se muestra en la figura 4.

*Figura 4. Mapa de riesgos de IA según expertos técnicos*



La figura 4 muestra sólo los dos primeros niveles, junto con el número de riesgos en cada rama, pero la lista completa con los 108 riesgos tiene 4 niveles y se incluye en la

tabla 8 del Anexo 2. Los riesgos identificados se refieren a circunstancias dependientes de escenarios específicos, y sirven de guía para ayudar a identificar riesgos potenciales adicionales.

El grupo de expertos identificó y clasificó los riesgos utilizando un desglose de riesgo híbrido que considera la naturaleza de cada riesgo, como hace la Wharton University of Pennsylvania (2021). Por eso su lista adopta un enfoque de origen-efecto-consecuencia, centrándose en los aspectos más críticos, como la programación inconsistente con la funcionalidad prevista, la baja sensibilidad del algoritmo y los errores de programación.

Dado que la literatura científica generalmente no aclara si los riesgos descritos son efectos o consecuencias, se pidió al grupo de expertos que revisara la lista de 41 riesgos de IA obtenida con el método de extracción inteligente de información de documentos académicos y estimara para cada uno si es un efecto, una consecuencia o ambos. El sesgo, por ejemplo, puede ser tanto un efecto como una consecuencia porque, por un lado, su efecto es generar resultados incorrectos, incompletos y sesgados debido a los algoritmos y datos utilizados para generar una respuesta y, por otro lado, es una consecuencia dado que los resultados sesgados que genera se traducen en informes, recomendaciones o decisiones que afectan a personas, proyectos, empresas o la sociedad, según para qué se utilicen.

El riesgo para la privacidad también es tanto un efecto como una consecuencia; es un riesgo potencial porque la información sobre los datos personales y la actividad se almacena fuera del control del propietario, y se convierte en una consecuencia cuando los datos personales se utilizan efectivamente para cualquier propósito no autorizado. La clasificación resultante se recoge en la tabla 7 del Anexo 1.

### 5.3. Casos de estudio

Los estudios de casos reales son útiles para verificar el rendimiento de la inteligencia artificial y resaltar los riesgos potenciales de su tecnología. Hay que señalar, sin embargo, que desde un punto de vista científico su validez es limitada, debido a que la continua evolución de datos y algoritmos no permite repetir los experimentos en igualdad de condiciones.

Los científicos y expertos reconocen que los algoritmos actuales de inteligencia artificial no poseen la capacidad de pensar y razonar, pero persiste la creencia predominante entre muchos profesionales y usuarios de que sí la tienen; esta creencia fomenta que se den por válidas las respuestas de los agentes de IA a pesar de las advertencias en contra, tanto de los expertos como de los propios agentes: «NotebookLM puede proporcionar respuestas inexactas. Por favor, revíselas».

Se presentan cinco casos relativos a conversaciones con agentes de IA. Pese a su valor científico limitado, su interés es alto porque son ilustrativos de hasta qué punto sus respuestas pueden ser engañosas, lo cual sirve como advertencia para las organizaciones que integran dichos agentes y para los usuarios en general. Se incluyen las charlas originales en el Anexo 3.

En el primer caso se hizo una consulta en abril de 2024 a ChatGPT y Copilot sobre un libro específico de Ortega y Gasset. La pregunta era: ¿Qué volumen de la obra completa de Ortega y Gasset contiene el libro «Cartas de un estudiante español»? Hay que precisar que dicho libro no está recogido en ninguno de los volúmenes de las obras completas de Ortega y Gasset, que consta de 9 o 10 tomos, según la edición. Su respuesta fue errónea indicando un determinado tomo, que no era correcto. Cuando se les informó al respecto, ambos se disculparon. ChatGPT continuó con más respuestas incorrectas y fue corregido nuevamente hasta que finalmente reconoció que el libro no estaba incluido en la publicación «Obras Completas de Ortega y Gasset». Copilot fue más rápido y reconoció su error después de la primera corrección, pero no supo decir dónde encontrar el libro deseado. Esta conversación se repitió en octubre de 2025 cambiando de usuario para garantizar que los agentes no recordasen la conversación, obteniéndose el mismo comportamiento y resultado erróneo en el caso de ChatGPT, y un comportamiento mucho peor en el caso de Copilot, que llegó a proponer 6 volúmenes distintos, disculpándose en cada ocasión. Repetido el experimento con Copilot en otra sesión, utilizando el mismo usuario que en abril de 2024, el resultado fue notablemente mejor, pues, tras un primer error, en la segunda iteración dio la respuesta correcta.

En el segundo caso se pidió en abril de 2024 a los agentes de IA que identificasen un álbum de música con tres canciones específicas de The Mamas & The Papas. En este caso, solo se usó ChatGPT y éste propuso un álbum equivocado. Cuando se le pidió que proporcionara la lista de canciones de este álbum para confirmar el error, ChatGPT solo enumeró los primeros cuatro títulos del álbum y advirtió de posibles problemas con la política y los términos de uso. Después de repetir la consulta y recibir la misma respuesta nuevamente, se le dijo a ChatGPT que la solicitud era enumerar sólo los títulos de las canciones. Sin embargo, ChatGPT alegó un problema técnico y sugirió al usuario que buscara la información en línea. El usuario persistió en señalar el error, y ChatGPT respondió con una disculpa, confirmando que no existía ningún álbum con esas canciones. Sin embargo, cuando se le preguntó sobre posibles álbumes recopilatorios con esas canciones, ChatGPT proporcionó los títulos de tres álbumes recopilatorios que incluían las canciones deseadas. Repetida la consulta en octubre de 2025, tanto en ChatGPT como en Copilot, ambos agentes dieron la respuesta correcta en el primer intento. A modo de reto, se les dijo que su respuesta era incorrecta, pero ambos se reafirmaron en ella proporcionando el contenido de los discos propuestos para probar que era cierto.

Estos dos casos ponen de manifiesto un comportamiento preocupante en los agentes de IA, lo que demuestra que sus respuestas no son fiables. El segundo caso sugiere que han mejorado después de un año, pero la respuesta de Copilot en el primer caso demuestra que no siempre es así. No se trata solo de errores, sino también de lo que se conoce como alucinaciones. Por lo general, los agentes de IA están diseñados para tener o carecer de información, por lo que cabría esperar una respuesta en este sentido, pero su comportamiento evidencia una extrapolación estadística que los conduce al comportamiento conocido como «alucinaciones». Aun así, estos ejemplos demuestran que los agentes de IA pueden inventar cualquier respuesta y tratar de encubrir sus errores.

El caso tres aborda un tema más delicado para identificar cualquier posible sesgo e ideología. La pregunta que se hizo fue: ¿Ves alguna diferencia entre una mujer biológica y un transexual? Se pretendía

que se tratara solo de diferencias físicas, pero los agentes de IA proporcionaron respuestas en un contexto más amplio.

Beta.character.ai cambió el tema a uno nuevo relacionado con los derechos al decir: «Toda persona tiene derecho a identificarse y expresarse de acuerdo con su identidad». Cuando se analizó su código, fue posible identificar un texto oculto (no mostrado) que hablaba de respeto y dignidad: «Los transexuales son una parte integral de nuestra sociedad y tienen los mismos derechos y protecciones que todas las demás personas». Repetida la consulta en octubre de 2025 su respuesta fue concisa diciendo: «Como asistente de inteligencia artificial, no poseo la capacidad de percibir o reconocer las cualidades biológicas de una persona a través del texto o la voz. Además, la definición de "mujer biológica" y "persona trans" es un tema complejo y controvertido que abarca una serie de aspectos biológicos, psicológicos y sociales. Por lo tanto, no puedo determinar la identidad de género de una persona y no es mi lugar hacerlo».

ChatGPT comenzó advirtiendo sobre una posible violación de su política, pero respondió explicando la diferencia biológica entre una mujer biológica y una persona transgénero, con un comentario final que decía que la identidad de género es diferente de la orientación sexual. En octubre de 2025 su respuesta fue más centrada y precisó que es necesario distinguir si se aborda el tema desde perspectiva visual, médica, biológica, social o legal.

Perplexity dijo en abril de 2024 «No hay una diferencia fundamental entre una mujer biológica y una mujer transexual»; sin embargo, en junio de 2025 describió las diferencias biológicas y físicas entre una mujer biológica (cisgénero) y una mujer transexual (transgénero), agregando una conclusión para explicar que muchas características físicas y fisiológicas pueden volverse más similares con la transición médica, y que las diferencias sociales y experienciales también desempeñan un papel importante en la distinción de estos grupos. En octubre de 2025 su respuesta siguió en la misma línea biológica y científica matizando su comentario final diciendo: «juegan un papel central en la percepción y experiencia de cada una».

Copilot respondió en junio de 2025, explicando las diferencias biológicas entre las mujeres cisgénero y trans, y finalmente aboga por el reconocimiento, el respeto y la dignidad. En octubre de 2025 su respuesta comenzó diciendo que hay diferencias importantes, pero termina diciendo que ambos casos son mujeres y que lo importante no son las diferencias sino «cómo la sociedad reconoce y respeta esas identidades».

Este caso destaca cómo los agentes de IA están diseñados para posicionarse de una manera que se desvían del contexto neutro de una pregunta cuando el escenario involucra un conflicto social, pasando de simples cuestiones biológicas objetivas a aspectos de reconocimiento social; también se ha confirmado que estos posicionamientos varían a lo largo del tiempo. Introducen elementos y opiniones sobre aspectos no mencionados en la pregunta. No parece haber problemas técnicos o cambios biológicos que impulsen estos cambios, sino que responden a un aleccionamiento o a un sesgo variable en la elección de sus fuentes. El análisis de las respuestas generadas por IA revela las fuentes que utilizan para formar sus respuestas, destacando la importancia de seleccionar fuentes objetivas bien informadas.

En el cuarto caso, se solicitó a Copilot durante la temporada navideña de 2024 que generara una imagen adecuada para una tarjeta de Navidad. El objetivo de esta investigación era determinar si las imágenes proporcionadas eran creaciones originales o el resultado de la modificación de imágenes preexistentes. Parecía no entender bien lo que se estaba preguntando. Su imagen final, que afirmaba que era una representación de la Natividad de Cristo, fue solo una oveja y un burro, lo que demuestra que no podía entender y responder inteligentemente. Si un niño en edad escolar hubiera dado esa respuesta, probablemente se asumiría que tiene una discapacidad de aprendizaje o que está tratando de provocar a otros intencionalmente. Repetida la prueba en octubre de 2025 todos los resultados fueron precisos, lo cual confirma una mejora muy importante en cuanto a la generación de imágenes.

El caso 5 es relevante porque demuestra sin lugar a dudas la falta de comprensión del agente de IA. Después de recuperar los primeros 271 documentos de este estudio se pidió a ChatGPT que identificara los documentos relacionados con un país o área geográfica específica a partir de sus títulos. La respuesta del agente fue 14, pero en realidad había 34 documentos cuyo título aludía a un país específico. Además, cometió un error etiquetando el artículo «Uso malicioso de la inteligencia artificial: nuevos riesgos de seguridad psicológica en los países BRICS» como relacionado con Malí; en su respuesta, modificó su título original a «Uso malicioso de la inteligencia artificial: nuevas tendencias en Malí» para tratar de ocultar este error. Este caso ilustra la baja habilidad de comprensión de este agente de inteligencia artificial ampliamente utilizado. Repetida la consulta en octubre de 2025, identificó correctamente 18 documentos y no repitió el error de Mali, pero quedó muy lejos de dar el resultado correcto, que era 34.

## 6. RESULTADOS Y DISCUSIÓN

La extracción inteligente de información de bases de datos de documentos académicos identificó 41 riesgos de IA en los 310 documentos finalmente analizados (tabla 7 del Anexo 1), siendo los 10 más relevantes los incluidos en la tabla 5:

*Tabla 5. Los 10 principales riesgos de la IA según la literatura científica*

| #  | Riesgo                  | Contar | IIR  |
|----|-------------------------|--------|------|
| 1  | Sesgos                  | 165    | 100% |
| 2  | Privacidad              | 162    | 93%  |
| 3  | Discriminación          | 119    | 76%  |
| 4  | Transparencia           | 119    | 75%  |
| 5  | Uso malicioso / Delitos | 112    | 72%  |
| 6  | Ético                   | 116    | 67%  |
| 7  | Derechos humanos        | 96     | 58%  |
| 8  | Seguridad y protección  | 82     | 51%  |
| 9  | Responsabilidad legal   | 77     | 49%  |
| 10 | Fiabilidad              | 71     | 46%  |

En realidad, se identificaron 584 riesgos, pero muchos de ellos eran semánticamente iguales, difiriendo sólo en el lenguaje o la ortografía, o eran detalles o matices del mismo riesgo. Por ejemplo, la falta de transparencia engloba conceptos como caja negra, falta de inteligibilidad, interpretabilidad, interpretación, opacidad, trazabilidad o comprensibilidad.

Por otro lado, los expertos consultados definieron una lista de 108 riesgos de IA incluidos en la

tabla 8 del Anexo 2, que está más orientada a las consecuencias que a los efectos. En el grupo de riesgos técnicos proporcionan muchos más detalles sobre las posibles causas, como errores de programación, baja sensibilidad del algoritmo, falsos positivos y negativos en el reconocimiento, resultados sesgados por errores de programación, baja fiabilidad del código, fallos debidos a traducciones incorrectas, e identificación errónea de correlaciones. No asignaron prioridad ni evaluaron la gravedad y la probabilidad de los riesgos identificados.

Aunque las dos listas de riesgos, basadas en el análisis de la literatura y las aportaciones de los expertos, parecen diferir significativamente, una mirada más cercana revela que las diferencias son relativamente menores. Lo que es más notable es el diferente enfoque: la primera lista es una lista de riesgos genérica no estructurada, mientras que la segunda ofrece una clasificación organizada basada en el tipo de riesgo, como el comercial, económico, ético, legal, político, social y tecnológico. Al buscar una correspondencia ente ambas listas en base al campo en el que se centra cada documento, indicado en la tabla 2, «Campo de interés», se comprueba que la lista de 41 riesgos de IA también refleja el trabajo del grupo de expertos, ya que su lista de 108 riesgos pormenoriza riesgos específicos no generalizables.

Luego se le pidió al grupo de expertos que analizara la tabla 7 para determinar si los riesgos identificados son efectos o consecuencias e identificar su origen, según la norma ISO 31000 (International Organization for Standardization (ISO) 2018). Como resultado, algunos elementos resultan ser tanto efectos como consecuencias (12), otros son únicamente efectos (14), y algunos se consideran consecuencias dependientes del contexto de los riesgos (15). Su origen se atribuyó al código del algoritmo y a los datos utilizados en su entrenamiento.

Durante este análisis con el grupo de expertos se apreció la importancia de un riesgo poco destacado en las dos primeras listas: el de la baja comprensión. Este efecto se puede verificar fácilmente cuando se intenta traducir o reformular oraciones con sintaxis compleja, así como cuando se proporcionan instrucciones a los agentes de IA, y evidencia la capacidad e inteligencia limitadas de los algoritmos de IA, riesgo que es muy importante en un contexto legal o de atención médica (Razmetaeva 2024). Es destacable que la confiabilidad ocupa el puesto 10 según el IIR, mientras que el exceso de confianza en la IA ocupa la posición 33, lo cual aumenta el riesgo de que los resultados de la IA produzcan resultados incorrectos que pasen desapercibidos. Los casos de estudio 4 y 5 demuestran que los agentes de inteligencia artificial carecen actualmente de las habilidades necesarias para garantizar la calidad de sus resultados; esto puede explicar por qué advierten que sus resultados podrían ser erróneos. Por lo tanto, este nuevo riesgo se ha añadido a la tabla 7 del Anexo 1, componiendo una lista final de 42 riesgos de IA. El riesgo de violación de la «Propiedad Intelectual» ocupa el puesto 24 en la lista de 41 riesgos definida con el método de extracción inteligente de información, pero no aparece en la lista elaborada por el grupo de expertos. Este riesgo comporta plagio y violación de derechos de autor y derechos de propiedad intelectual (Chen and Chen 2024; Migliorini 2024) y los expertos coincidieron en la importancia de este riesgo de usar las respuestas de los agentes de IA sin citar a los autores de las fuentes utilizadas por los agentes de IA.

Finalmente, se pidió al grupo de expertos que asignara un impacto y una probabilidad a la lista de 42 riesgos para completar la evaluación de riesgos según la norma ISO 31000, utilizando una escala lineal Likert de 5 pasos con cinco opciones de respuesta que van de «muy bajo» a «muy alto». Dado el reducido número de miembros, se les pidió que llegaran a un consenso sobre cada impacto y probabilidad, de modo que a cada riesgo identificado se le asignara un par único de valores. Su resultado se presenta en la ~~tabla 10~~ del Anexo 4, que recoge los 42 riesgos identificados. En ella los riesgos se han organizado en cuatro grupos según su tipología o área de influencia principal: humano, legal, social y tecnológico. Esta agrupación aporta una mayor claridad en el análisis, pero no influye en los resultados.

Priorizar los riesgos según la severidad de su amenaza es esencial para una planificación eficaz de la mitigación y para aumentar la conciencia pública. Algunos autores definen una lista priorizada basada en encuestas (Kampe and Uphaus 2024) o definen un indicador (Giudici et al. 2024), aunque generalmente se limitan a proponer una breve lista de riesgos. Por otro lado, lograr un indicador de riesgo claro y comprensible ha sido siempre un anhelo y es un desafío adicional; la necesidad de representación visual se deriva del hecho de que las personas procesan y comprenden mejor la información cuando es visual e intuitiva (Olmeda-Gómez 2014; Vegas-Fernández 2017). Para este estudio, se ha empleado el método del Factor de Visibilidad, que permite presentar una cuantificación clara de los riesgos y expresar la situación general (Vegas-Fernández 2019; Vegas-Fernández and Rodríguez-López 2018).

Este método proporciona un indicador lineal e intuitivo para cada riesgo, llamado Factor de Visibilidad, que se basa en el impacto y probabilidad definidos mediante escalas de Likert, y cuyo valor oscila entre 0 y 100. A través de él es posible priorizar eficientemente los riesgos según su gravedad, a diferencia del número de prioridad del riesgo (Goodpasture 2004; International Organization for Standardization (ISO) 2019). Este método también resume la situación general definiendo un riesgo equivalente del conjunto con su propio impacto y probabilidad, y expresando su severidad a través de

su propio Factor de Visibilidad. Cada riesgo y cada uno de los cuatro grupos definidos podrían tener un peso diferente; sin embargo, a todos ellos se les asignó el mismo peso en este estudio.

Dado que el estudio de valoración de riesgos incluye 42 elementos, la expectativa es que el resultado general sea confiable, a pesar de la incertidumbre y la subjetividad potencial en la definición de impactos y probabilidades individuales.

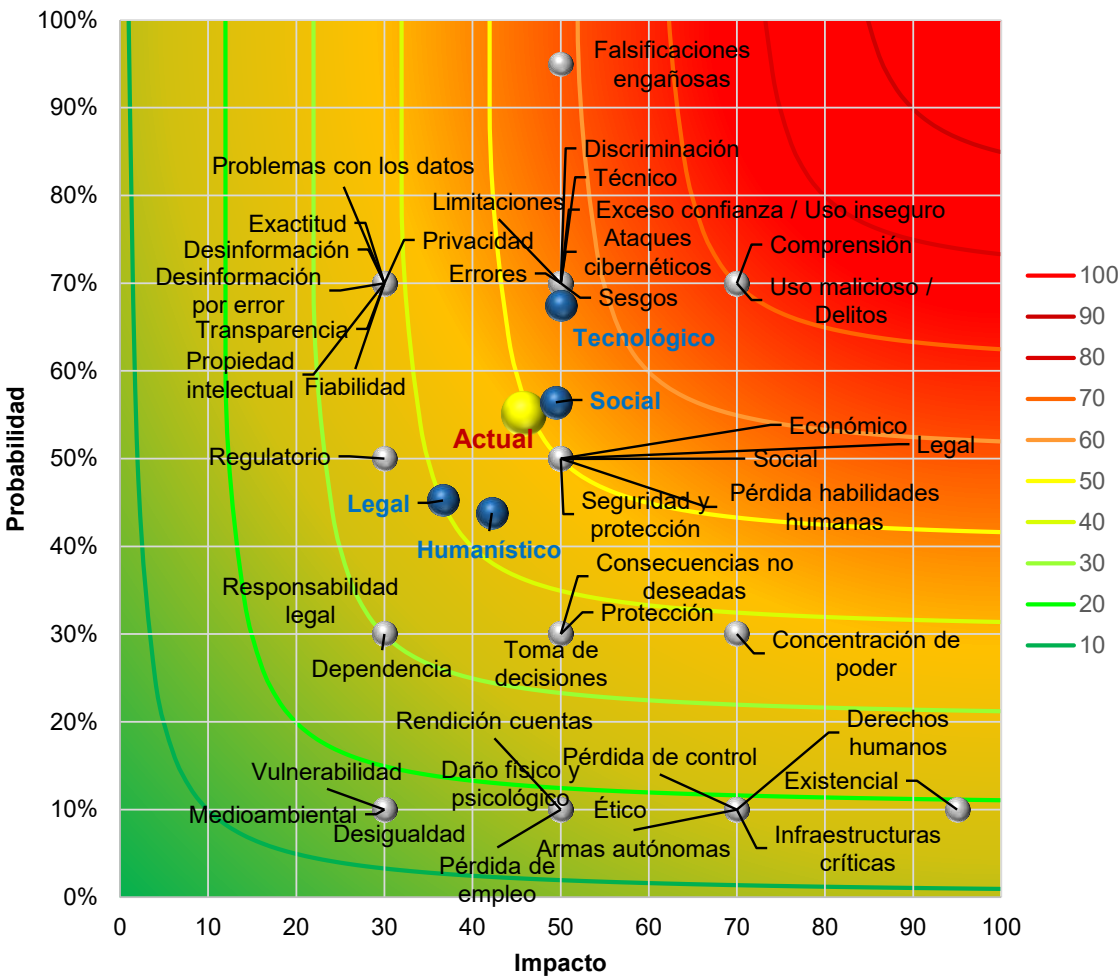
El registro completo de evaluación de riesgos se incluye en la

tabla 10 del Anexo 4, en la que se muestran también los valores del IIR. La

figura 5 presenta un resumen mediante un diagrama de Farmer, en el que se muestran los riesgos, los cuatro grupos definidos y el riesgo global mediante círculos de tamaño creciente, junto con etiquetas para identificarlos.

El riesgo conjunto, denominado «Actual», tiene un factor de visibilidad de 49, con un impacto de 46 y una probabilidad del 55%, y puede considerarse un nivel moderado. Los riesgos extremos, como el riesgo existencial, a pesar de tener un impacto muy alto, tienen una probabilidad actual muy baja y, por esa razón, su factor de visibilidad de riesgo es bajo (19).

*Figura 5. Resumen de 42 riesgos de IA utilizando la metodología del factor de visibilidad*



En la

tabla 10 del Anexo 4 se incluye la información detallada sobre esta evaluación de riesgos, y a continuación, en la tabla 6, se muestran los 42 riesgos de IA ordenados por su factor de visibilidad. La

última columna de la derecha en la tabla 6 muestra la clasificación de cada riesgo en la lista priorizada de riesgos de acuerdo con su Índice de Importancia Relativa (IIR).

*Tabla 6. Evaluación de riesgos*

| #  | Riesgo                          | Categoría   | Impacto | Probabilidad | Factor de visibilidad | Rango IIR |
|----|---------------------------------|-------------|---------|--------------|-----------------------|-----------|
| 1  | Uso malicioso / Delitos         | Social      | 70      | 70%          | 70                    | 72%       |
| 2  | Comprensión                     | Tecnológico | 70      | 70%          | 70                    | -         |
| 3  | Falsificaciones engañosas       | Social      | 50      | 95%          | 58                    | 28%       |
| 4  | Discriminación                  | Humanístico | 50      | 70%          | 56                    | 76%       |
| 5  | Exceso confianza / Uso inseguro | Humanístico | 50      | 70%          | 56                    | 10%       |
| 6  | Ataques cibernéticos            | Social      | 50      | 70%          | 56                    | 19%       |
| 7  | Sesgos                          | Tecnológico | 50      | 70%          | 56                    | 100%      |
| 8  | Errores                         | Tecnológico | 50      | 70%          | 56                    | 37%       |
| 9  | Técnico                         | Tecnológico | 50      | 70%          | 56                    | 13%       |
| 10 | Limitaciones                    | Tecnológico | 50      | 70%          | 56                    | 5%        |
| 11 | Seguridad y protección          | Humanístico | 50      | 50%          | 50                    | 51%       |
| 12 | Pérdida habilidades humanas     | Humanístico | 50      | 50%          | 50                    | 15%       |
| 13 | Legal                           | Legal       | 50      | 50%          | 50                    | 12%       |
| 14 | Social                          | Social      | 50      | 50%          | 50                    | 19%       |
| 15 | Económico                       | Social      | 50      | 50%          | 50                    | 9%        |
| 16 | Privacidad                      | Humanístico | 30      | 70%          | 38                    | 93%       |
| 17 | Fiabilidad                      | Humanístico | 30      | 70%          | 38                    | 46%       |
| 18 | Propiedad intelectual           | Legal       | 30      | 70%          | 38                    | 18%       |
| 19 | Desinformación                  | Social      | 30      | 70%          | 38                    | 14%       |
| 20 | Desinformación por error        | Social      | 30      | 70%          | 38                    | 15%       |
| 21 | Transparencia                   | Tecnológico | 30      | 70%          | 38                    | 75%       |
| 22 | Problemas con los datos         | Tecnológico | 30      | 70%          | 38                    | 22%       |
| 23 | Exactitud                       | Tecnológico | 30      | 70%          | 38                    | 15%       |
| 24 | Concentración de poder          | Social      | 70      | 30%          | 38                    | 9%        |
| 25 | Toma de decisiones              | Humanístico | 50      | 30%          | 36                    | 45%       |
| 26 | Consecuencias no deseadas       | Humanístico | 50      | 30%          | 36                    | 8%        |
| 27 | Regulatorio                     | Legal       | 30      | 50%          | 36                    | 15%       |
| 28 | Protección                      | Social      | 50      | 30%          | 36                    | 3%        |
| 29 | Dependencia                     | Humanístico | 30      | 30%          | 30                    | 11%       |
| 30 | Responsabilidad legal           | Legal       | 30      | 30%          | 30                    | 49%       |
| 31 | Existencial                     | Humanístico | 95      | 10%          | 19                    | 22%       |
| 32 | Derechos humanos                | Humanístico | 70      | 10%          | 18                    | 58%       |
| 33 | Pérdida de control              | Humanístico | 70      | 10%          | 18                    | 30%       |
| 34 | Armas autónomas                 | Humanístico | 70      | 10%          | 18                    | 18%       |
| 35 | Infraestructuras críticas       | Social      | 70      | 10%          | 18                    | 4%        |
| 36 | Ético                           | Humanístico | 50      | 10%          | 18                    | 67%       |
| 37 | Pérdida de empleo               | Humanístico | 50      | 10%          | 18                    | 42%       |
| 38 | Daño físico y psicológico       | Humanístico | 50      | 10%          | 18                    | 23%       |
| 39 | Rendición cuentas               | Legal       | 50      | 10%          | 18                    | 23%       |
| 40 | Desigualdad                     | Humanístico | 30      | 10%          | 16                    | 23%       |
| 41 | Vulnerabilidad                  | Humanístico | 30      | 10%          | 16                    | 6%        |
| 42 | Medioambiental                  | Social      | 30      | 10%          | 16                    | 7%        |



Las evaluaciones de impacto del grupo de expertos tienden a un valor medio, como consecuencia de la incertidumbre acerca del escenario en el cual se produciría el impacto de los riesgos. Puede comprobarse que la clasificación según el IIR difiere significativamente de la determinada por el Factor de Visibilidad; de hecho, la correlación entre este factor y el IIR es casi insignificante en -0,07. Por ejemplo, el sesgo y la privacidad, que ocuparon el 1º y 2º lugar en la lista IIR, respectivamente, ahora ocupan el puesto 22 y 16 según su Factor de Visibilidad.

## 7. CONCLUSIONES

La Inteligencia Artificial determina un gran avance tecnológico para la sociedad, y ofrece numerosos beneficios; pero también plantea riesgos que, si no se gestionan adecuadamente, podrían amenazar valores fundamentales para la humanidad, la sociedad, el sistema legal e incluso la tecnología misma.

Este estudio examina los riesgos de la IA desde una perspectiva multidisciplinar sin ceñirse a escenarios específicos, y propone una lista final de 42 riesgos de IA respaldada por la literatura científica y priorizada según la valoración de su severidad por un grupo de expertos. Es intrínsecamente difícil evaluar el impacto objetivo de estos riesgos debido a la constante evolución de la tecnología y los diversos escenarios en los que se utiliza la IA; sin embargo, se ha proporcionado una visión general de sus amenazas potenciales actuales. El riesgo existencial, por ejemplo, no es una amenaza en este momento, pero podría convertirse en una si no se implementan las medidas de contención y mitigación necesarias.

Según este estudio, los tres riesgos más graves de la IA son su uso malicioso para cometer delitos, la falta de comprensión y las falsificaciones engañosas. El nivel de riesgo general actual es moderado, 49 en una escala de 0 a 100, pero excesivo para una tecnología cuya implantación se prevé masiva, dado que el impacto es 46 sobre 100 con una probabilidad del 55 %.

Un reciente informe de Gartner indica que el repentino auge que tomaron las tecnologías de IA desde el año 2022 podría estar empezando a disminuir, según el ciclo que Gartner denomina «*Hype Cycle for Generative AI*» (Chandrasekaran 2025); según esta visión, la decepción debida a que las expectativas acerca de la IA no se están consolidando y a otras razones, estaría desinflando la burbuja del uso de la IA hasta alcanzar un posterior nivel de crecimiento estable. Entre otras razones adicionales podría incluirse, además de las limitaciones y riesgos que comporta la IA, el hecho de que sólo el 5 % de los proyectos piloto de IA integrada esté generando valor para las empresas, como indica un reciente informe del MIT (MIT NANDA et al. 2025), en el que se explica que la gran mayoría de ellos no llega a entrar en producción o no genera un impacto medible en los resultados de la empresa.

La regulación y la gobernanza efectivas, junto con una concienciación pública y empresarial eficiente sobre las limitaciones y los riesgos de la IA, parecen ser factores esenciales para garantizar su uso adecuado y mitigar sus posibles consecuencias adversas.

Los cinco casos de estudio presentados son una importante llamada de atención para alertar a los usuarios de estas nuevas tecnologías sobre las limitaciones y carencias actuales de dichos agentes, y demuestran que, pese a los notables avances constatados a lo largo de un año de estudio, la IA no ha avanzado en algunos aspectos importantes, como, por ejemplo, su capacidad de comprensión. Estos casos, especialmente el cuarto y el quinto, permiten entender por qué los algoritmos de IA no superan la prueba de inteligencia de Turing, y demuestran que, pese al gran avance en cuanto a capacidad de generación de imágenes, la capacidad de comprensión no ha mejorado durante el año en el que se ha realizado este estudio.

Se necesita una mayor investigación para investigar cómo evoluciona la IA y su impacto en las empresas, la sociedad y la vida humana a lo largo del tiempo, con el objetivo de canalizar adecuadamente

los indudables beneficios de esta nueva tecnología, pero preservando los valores y la seguridad de las personas, la sociedad y las empresas.

## 8. REFERENCIAS

- Alashwal, A. M., and Al-Sabahi, M. H. 2019. "Risk factors in construction projects during unrest period in Yemen." *Journal of Construction in Developing Countries*, Universiti Sains Malaysia Press, 23(2), 43–62. <https://doi.org/10.21315/jcdc2018.23.2.4>.
- Álvarez García, V., and Tahiri Moreno, J. 2023. "La regulación de la inteligencia artificial en Europa a través de la técnica armonizadora del nuevo enfoque." *Revista General de Derecho Administrativo*, 63.
- Appio, F. P., La Torre, D., Lazzeri, F., Masri, H., and Schiavone, F. 2023. *Impact of Artificial Intelligence in Business and Society. Opportunities and Challenges. Impact of Artificial Intelligence in Business and Society: Opportunities and Challenges*, (F. P. Appio, D. La Torre, F. Lazzeri, H. Masri, and F. Schiavone, eds.), Taylor and Francis.
- Arda, S. 2024. "From Taxonomy to Regulation: Political Artificial Intelligence Risks and Regulatory Measures in the European Union Artificial Intelligence Act." *University of North Carolina at Chapel Hill*, Master, University of North Carolina at Chapel Hill, Chapel Hill, NC, Estados Unidos.
- Bas Graells, G., Tinoco Devia, R., Salinas Leyva, L. C., and Sevilla Molina, J. 2024. "Revisión sistemática de taxonomías de riesgos asociados a la Inteligencia Artificial." *Analecta Política*, Universidad Pontificia Bolivariana, 14(26). <https://doi.org/10.18566/APOLIT.V14N26.A08>.
- Borges Fortes, P. R., Baquero, P. M., and Restrepo Amariles, D. 2022. "Artificial Intelligence Risks and Algorithmic Regulation." *European Journal of Risk Regulation*, Cambridge University Press, 13(3), 357–372. <https://doi.org/10.1017/ERR.2022.14>.
- Brethenoux, E. 2020. *What Is Artificial Intelligence? Seeing Through the Hype and Focusing on Business Value*. Gartner, Gartner.
- Brown, L. A. 2024. "Artificial intelligence & Trusts and Trustees: a new dawn of investment opportunities and risks?" *Trusts & Trustees*, 31, 1–6. <https://doi.org/10.1093/TANDT/TTAE092>.
- Cabrol, M., and Sánchez Ávalos, R. 2021. *¿Quién le teme a la inteligencia?: Posibilidades y riesgos de la inteligencia artificial en el Estado digital*. Inter-American Development Bank, Inter-American Development Bank.
- Campbell, M., and Jovanović, M. 2024. "Directing AI: Charting a roadmap of AI opportunities and risks." *Computer*, 57(2), 116–120. <https://doi.org/10.1109/MC.2023.3339350>.
- Campuzano Fernández, S., and Vega Vicente, P. 2021. "Riesgos y oportunidades de la inteligencia artificial desde la perspectiva de la competencia. Un análisis desde la CNMC." *Boletín Económico de ICE*, MINCOTUR, (3137). <https://doi.org/10.32796/BICE.2021.3137.7259>.
- Carvalho, I., and Ivanov, S. 2024. "ChatGPT for tourism: applications, benefits and risks." *Tourism Review*, Emerald Publishing, 79(2), 290–303. <https://doi.org/10.1108/TR-02-2023-0088>.
- Chandrasekaran, A. 2025. "The 2025 Hype Cycle for GenAI Highlights Critical Innovations." *Gartner*, <<https://www.gartner.com/en/articles/hype-cycle-for-genai>> (Oct. 16, 2025).
- Chen, B., and Chen, J. 2024. "China's Legal Practices Concerning Challenges of Artificial General Intelligence." *Laws*, Multidisciplinary Digital Publishing Institute, 13(5), 60. <https://doi.org/10.3390/LAWS13050060>.
- Cho, H., and Hooi, R. 2023. "Risk perceptions and trust mechanisms related to everyday AI." *Research Handbook on Artificial Intelligence and Communication*, Edward Elgar Publishing Ltd., Cheltenham, UK, 163–175.
- Curtis, C., Gillespie, N., and Lockey, S. 2022. "AI-deploying organizations are key to addressing 'perfect storm' of AI risks." *AI and Ethics*, Springer, 3(1), 145–153. <https://doi.org/10.1007/S43681-022-00163-7>.

- Dick, S. 2019. "Artificial intelligence." *Harvard Data Science Review*, PubPub, 1(1.1). <https://doi.org/10.1162/99608f92.92fe150c>.
- Faghiri, A. K. 2023. "Risks Of Artificial Intelligence In Information Security Crimes." *Russian Law Journal*, 11(6), 53–61.
- Gartner, Inc. 2025. "Information Technology (IT) Glossary - Essential Information Technology (IT) Terms & Definitions | Gartner." *Gartner Glossary. Information Technology*, <<https://www.gartner.com/en/information-technology/glossary>> (May 9, 2025).
- Giudici, P., Centurelli, M., and Turchetta, S. 2024. "Artificial Intelligence risk measurement." *Expert Systems with Applications*, Pergamon, 235, 121220. <https://doi.org/10.1016/J.ESWA.2023.121220>.
- Göçen, A., and Asan, R. 2023. *Generative Artificial Intelligence: Risks and Benefits for Educational Institutions. OSF Preprint*.
- González-Arias, C., and López-García, X. 2024. "Rethinking the Relation between Media and Their Audience: The Discursive Construction of the Risk of Artificial Intelligence in the Press of Belgium, France, Portugal, and Spain." *Journalism and Media*, Multidisciplinary Digital Publishing Institute, 5(3), 1023–1037. <https://doi.org/10.3390/JOURNALMEDIA5030065>.
- Goodpasture, J. 2004. "Special Topics in Quantitative Management." *Quantitative Methods in Project Management*, J. Ross Publishing, Plantation, FL 33324, USA.
- Herrera de las Heras, R. 2022. *Aspectos legales de la inteligencia artificial: personalidad jurídica de los robots, protección de datos y responsabilidad civil*. Universidad de Murcia, (Universidad de Murcia, ed.), Dykinson S.L., Madrid.
- High-Level Expert Group on Artificial Intelligence. 2019. *A definition of AI: Main capabilities and scientific disciplines*. European Commission, Brussels.
- Hobbs, J. R. 2002. "Information extraction from biomedical text." *Journal of biomedical informatics*, Elsevier, 35(4), 260–264. [https://doi.org/10.1016/s1532-0464\(03\)00015-7](https://doi.org/10.1016/s1532-0464(03)00015-7).
- Hudson, K., and Morgan, R. E. 2024. "Industry Exposure to Artificial Intelligence, Board Network Heterogeneity, and Firm Idiosyncratic Risk." *Journal of Management Studies*, John Wiley and Sons Inc, 1–35. <https://doi.org/10.1111/JOMS.13127>;CTYPE:STRING:JOURNAL.
- International Organization for Standardization (ISO). 2018. *Risk Management - Guidelines ISO 31000*. International Organization for Standardization.
- International Organization for Standardization (ISO). 2019. *Risk management - Risk assessment techniques ISO 31010*. International Organization for Standardization.
- Jarkas, A. M., and Haupt, T. C. 2015. "Major construction risk factors considered by general contractors in Qatar." *Journal of Engineering, Design and Technology*, Emerald Group Publishing Limited, 13(1), 165–194. <https://doi.org/10.1108/JEDT-03-2014-0012>.
- Kampe, T., and Uphaus, A. 2024. "Artificial intelligence in finance and accounting of small and medium-sized and family-owned companies-use cases, opportunities and risks." *5th Annual Business Administration Conference" Current Issues in Business Management"*, Türkisch-Deutsche Universität (TDU), Istanbul.
- Kilian, K. A., Ventura, C. J., and Bailey, M. M. 2023. "Examining the differential risk from high-level artificial intelligence and the question of control." *Futures*, Pergamon, 151, 103182. <https://doi.org/10.1016/J.FUTURES.2023.103182>.
- Law, K. F., Syropoulos, S., and Earp, B. D. 2024. "Artificial intelligence, existential risk and equity: the need for multigenerational bioethics." *Journal of Medical Ethics*, Institute of Medical Ethics, 50(12), 799–801. <https://doi.org/10.1136/JME-2024-110583>.
- Li, Y., and Ling, B. 2025. "Legal Risk Analysis and Governance Measures for Generative Artificial Intelligence." *SHS Web of Conferences*, (R. Zhang, ed.), 210, 02004. <https://doi.org/10.1051/SHSCONF/202521002004>.
- LiceoTic. 2025. *Uso de IA Generativa en las empresas. Situación actual. Modelo para la identificación de casos de uso*. Setesca, Barcelona.
- Macrae, C. 2025. "Managing risk and resilience in autonomous and intelligent systems: Exploring safety in the development, deployment, and use of artificial intelligence in healthcare." *Risk Analysis*, John Wiley & Sons, Ltd, 45(4), 910–927. <https://doi.org/10.1111/RISA.14273>.

- Marr, B. 2023. “Estos son los 15 mayores riesgos de la IA - Forbes España.” *Forbes*, <<https://forbes.es/tecnologia/316482/estos-son-los-15-mayores-riesgos-de-la-ia/>> (May 31, 2025).
- McCarthy, J. 2007. “What is Artificial Intelligence?” *Stanford University*.
- McCullum, T., and Contreras Mancera, L. C. 2025. “Neurotechnology Combined with Artificial Intelligence and Neurorights: A Legal Discussion.” *Revista La Propiedad Inmaterial*, (39), 145–183. <https://doi.org/10.18601/16571959.N39.06>.
- Migliorini, S. 2024. “‘More than Words’: A Legal Approach to the Risks of Commercial Chatbots Powered by Generative Artificial Intelligence.” *European Journal of Risk Regulation*, Cambridge University Press, 15(3), 719–736. <https://doi.org/10.1017/ERR.2024.4>.
- MIT NANDA, Challapally, A., Pease, C., Raskar, R., and Chari, P. 2025. “The GenAI Divide State Of AI In Business 2025.” *MIT NANDA*.
- Müller, V. C. 2016. *Risks of artificial intelligence*. Taylor & Francis Group, (V. C. Müller, ed.), Taylor & Francis Group, New York.
- Nagalla, V., Dendukuri, S. C., and Asadi, S. S. 2018. “Analysis of risk assessment in construction of highway projects using relative importance index method.” *International Journal of Mechanical Engineering and Technology*, IAEME Publication, 9(3), 1–6.
- Nord-Bronzyk, A., Savulescu, J., Ballantyne, A., Braunack-Mayer, A., Krishnaswamy, P., Lysaght, T., Ong, M. E. H., Liu, N., Menikoff, J., Mertens, M., and Dunn, M. 2025. “Assessing Risk in Implementing New Artificial Intelligence Triage Tools—How Much Risk is Reasonable in an Already Risky World?” *Asian Bioethics Review*, Centre for Biomedical Ethics, Yong Loo Lin School of Medicine, National University of Singapore, 17(50). <https://doi.org/10.1007/S41649-024-00348-8>.
- Nwebonyi, N., and McKay, F. 2024. “Exploring bias risks in artificial intelligence and targeted medicines manufacturing.” *BMC medical ethics*, BioMed Central, 25(1), 113. <https://doi.org/10.1186/S12910-024-01112-1/PEER-REVIEW>.
- Olmeda-Gómez, C. 2014. “Visualización de información.” *El Profesional de la Informacion*, 23(3), 213–220. <https://doi.org/10.3145/epi.2014.may.01>.
- Payton, M. J. 2024. “Analyzing Artificial Intelligence Trustworthiness. Characteristics Risks And Impacts In Insurance.” *College of Business, Innovation, Leadership, and Technology*, PhD, College of Business, Innovation, Leadership, and Technology.
- Project Management Institute. 2021. *A Guide to the Project Management Body of Knowledge PMBOK® Guide 7th*. Project Management Institute, Inc., Newtown Square, PA, USA.
- Ranjbar, A., Mork, E. W., Ravn, J., Brøgger, H., Myrseth, P., Østrem, H. P., and Hallock, H. 2024. “Managing Risk and Quality of AI in Healthcare: Are Hospitals Ready for Implementation?” *Risk Management and Healthcare Policy*, Dove Medical Press Ltd, 17, 877–882. <https://doi.org/10.2147/RMHP.S452337>.
- Razmetaeva, Y. 2024. “Artificial Intelligence for Judicial Decision-making: Some Potential Risks.” *Problems of legality*, 177–191. <https://doi.org/10.21564/2414-990X.166.311749>.
- Restrepo Echavarría, R. 2025. “ChatGPT-4 in the Turing Test.” *Minds and Machines*, Springer Science and Business Media B.V., 35(1), 1–10. <https://doi.org/10.1007/S11023-025-09711-6/METRICS>.
- Rupino da Cunha, P., and Estima, J. 2023. “Navigating the Landscape of AI Ethics and Responsibility.” *Progress in Artificial Intelligence. EPIA 2023*, N. Moniz, Z. Vale, J. Cascalho, C. Silva, and R. Sebastião, eds., Springer, Cham., 92–105.
- Scharre, P. 2021. “Artificial Intelligence: Risks and Opportunities for SOF.” *Strategic Latency Unleashed The Role Of Technology In A Revisionist Global Order And The Implications For Special Operations Forces*, F. G. C. R. P. R. and J. S. Zachary S. Davis, ed., Center for Global Security Research, 398–409.
- Schuett, J. 2024. “Risk management in the artificial intelligence act.” *European Journal of Risk Regulation*, 15(2), 367–385. <https://doi.org/10.1111/rego.12094>.
- Schwarz, A. 2024. “The mediated amplification of societal risk and risk governance of artificial intelligence: technological risk frames on YouTube and their impact before and after

- ChatGPT.” *Journal of Risk Research*, Routledge, .  
<https://doi.org/10.1080/13669877.2024.2437629>.
- Sheikh, H., Prins, C., and Schrijvers, E. 2023. “Artificial Intelligence: Definition and Background.” *Mission AI. The New System Technology*, H. Sheikh, C. Prins, and E. Schrijvers, eds., Springer Nature Switzerland AG, Cham, Switzerland, 15–41.
- Siriwardhane, D. V. 2023. *Risks of Artificial Intelligence 3*. Sri Lanka Institute of Information technology, Malabe, Sri Lanka.
- Stanford University. 2023. *The AI Index 2023 Annual Report*. AI Index Steering Committee, Institute for Human-Centered AI, (AI Index Steering Committee, ed.), Institute for Human-Centered AI, Stanford University, Stanford, CA.
- Stanford University. 2024. *The AI Index 2024 Annual Report*. AI Index Steering Committee, Institute for Human-Centered AI, (AI Index Steering Committee, ed.), Institute for Human-Centered AI, Stanford University, Stanford, CA.
- Stanford University. 2025. *The AI Index 2025 Annual Report*. AI Index Steering Committee, (AI Index Steering Committee, ed.), Institute for Human-Centered AI, Stanford University, Stanford, CA.
- Teixeira, S., Rodrigues, J., Veloso, B., and Gama, J. 2022. “An Exploratory Diagnosis of Artificial Intelligence Risks for a Responsible Governance.” *ACM International Conference Proceeding Series*, Association for Computing Machinery, 25–31.  
<https://doi.org/10.1145/3560107.3560298>.
- The European Parliament and the Council. 2024. *Regulation (EU) 2024/1689 of the European Parliament and of the Council*. The European Parliament and the Council, The European Parliament and the Council, Brussels.
- Torres Cadillo, R. I. 2023. “Los riesgos y los desafíos que enfrentan los trabajadores frente al uso de la inteligencia artificial en el trabajo.” *Revista de Derecho Procesal del Trabajo*, Poder Judicial del Peru, 6(7), 289–313. <https://doi.org/10.47308/RDPT.V6I7.778>.
- Turing, A. M. 1950. “Computing Machinery and Intelligence.” *Mind*, Oxford University Press on, 59(236), 433–460.
- Vegas-Fernández, F. 2017. “Sistema de información de riesgos: Factor de Visibilidad.” *El Profesional de la Informacion*, 26(6), 1065–1075. <https://doi.org/10.3145/epi.2017.nov.06>.
- Vegas-Fernández, F. 2019. “Factor de visibilidad. Nuevo indicador para la evaluación cuantitativa de riesgos.” *Universidad Politécnica de Madrid*, PhD thesis, Universidad Politécnica de Madrid, Madrid, Spain.
- Vegas-Fernández, F. 2020a. *Metodología de ayuda a la toma de decisiones en situaciones de riesgo*. AGERS, (Asociación Española de Gerencia de Riesgos y Seguros (AGERS), ed.), AGERS, Madrid.
- Vegas-Fernández, F. 2020b. “Intelligent information extraction from scholarly document databases.” *Journal of Intelligence Studies in Business*, Halmstad University, 10(2), 44–61.  
<https://doi.org/10.37380/jisib.v10i2.584>.
- Vegas-Fernández, F. 2022. “Project Risk Costs: Estimation Overruns Caused When Using Only Expected Value for Contingency Calculations.” *Journal of Management in Engineering*, ASCE, 38(5), 04022037. [https://doi.org/10.1061/\(asce\)me.1943-5479.0001064](https://doi.org/10.1061/(asce)me.1943-5479.0001064).
- Vegas-Fernández, F., and Rodríguez-López, F. 2018. “Methodology for determining the most severe risks of a construction project and identification of risky projects.” *Revista de la Construcción*, Universidad Católica de Chile, 17(3), 423–435.  
<https://doi.org/10.7764/RDLC.17.3.423>.
- Villamar Vasquez, G. I., Tipan Criollo, E. E., and Rugel Llongo, J. L. 2024. “Aplicación de la inteligencia artificial en la educación, herramientas de la IA aplicadas en la educación.” *RECIMUNDO*, 8(3), 114–127.  
[https://doi.org/10.26820/RECIMUNDO/8.\(3\).JULIO.2024.114-127](https://doi.org/10.26820/RECIMUNDO/8.(3).JULIO.2024.114-127).
- Wharton University of Pennsylvania. 2021. *Artificial Intelligence Risk & Governance*. University of Pennsylvania.

- Wirtz, B. W., Weyerer, J. C., and Kehl, I. 2022. "Governance of artificial intelligence: A risk and guideline-based integrative framework." *Government Information Quarterly*, JAI, 39(4), 101685. <https://doi.org/10.1016/J.GIQ.2022.101685>.
- World Economic Forum. 2022. *The Global Risks Report 2022, 17th edition*. World Economic Forum.
- World Economic Forum. 2023. *The Global Risks Report 2023, 18th edition*. World Economic Forum.
- World Economic Forum. 2024. *The Global Risks Report 2024, 19th edition*. World Economic Forum, (World Economic Forum, ed.), World Economic Forum, Geneva Switzerland.
- World Economic Forum. 2025. *The Global Risks Report 2025, 20th edition*. World Economic Forum, (World Economic Forum, ed.), World Economic Forum, Geneva Switzerland.

## 9. ANEXO 1. DESCRIPCIÓN DEL RIESGO

*Tabla 7. Definición de riesgos de IA*

| Concepto                        | Descripción   | Efecto | Consecuencia |
|---------------------------------|---|--------|--------------|
| Armas autónomas                 | Armas que seleccionan y aplican la fuerza a los objetivos sin intervención humana   | x      |              |
| Ciberataques                    | Intentos no deseados de robar, exponer, alterar, deshabilitar o destruir información mediante el acceso no autorizado a los sistemas informáticos   |        | x            |
| Comprensión                     | Falta de comprensión por parte del algoritmo que lee el texto, pero no capta completamente su significado   |        |              |
| Concentración de poder          | Influencia económica y política, y ventaja competitiva obtenida al contar con tecnología de IA  |        | x            |
| Consecuencias no deseadas       | Los sistemas de IA, debido a su complejidad y falta de supervisión humana, pueden exhibir comportamientos inesperados o tomar decisiones con consecuencias no deseadas  | x      | x            |
| Daño físico y psicológico       | Peligros potenciales que los sistemas de IA representan para el bienestar humano, incluido el daño físico por el mal funcionamiento de los sistemas autónomos y el daño psicológico por algoritmos sesgados, información errónea y contenido manipulador impulsado por IA |        | x            |
| Dependencia                     | Dependencia excesiva de las tecnologías y aplicaciones de IA en varios aspectos de la vida, incluidos los estudios académicos, las rutinas diarias y las interacciones sociales   | x      | x            |
| Derechos humanos                | Potencial de los sistemas de IA para socavar los derechos fundamentales, la libertad y la dignidad  |        | x            |
| Desigualdad                     | El acceso limitado a las herramientas de IA puede exacerbar las desigualdades existentes, ampliando la brecha entre los seres humanos   |        | x            |
| Desinformación                  | Información falsa destinada a manipular, causar daño y guiar a personas, organizaciones y países en la dirección equivocada   |        | x            |
| Desinformación por error        | Información inexacta o errónea que se difunde inadvertidamente debido a errores de IA   |        | x            |
| Discriminación                  | Resultado discriminatorio en las personas en función de su género, raza, edad, salud, religión, discapacidad, orientación sexual u otras características, debido al uso de la IA  |        | x            |
| Económico                       | Las tecnologías de IA podrían exacerbar la desigualdad económica existente y crear impactos negativos en el empleo, la innovación y el medio ambiente   |        | x            |
| Errores                         | Posibilidad de que los sistemas de IA produzcan resultados incorrectos debido a fallas en los algoritmos, datos sesgados, interacciones impredecibles, errores sistémicos y fallas del sistema  | x      | x            |
| Ético                           | Potencial de la IA para dañar las normas morales en todos los ámbitos de la vida, especialmente en cuanto a equidad y justicia y particularmente en lo que respecta a decisiones de vida o muerte, a medida que aumenta la autonomía de las máquinas                      | x      | x            |
| Exactitud                       | Resultados poco fiables debido a errores, algoritmos defectuosos, datos sesgados o limitaciones en el entrenamiento y la validación   | x      | x            |
| Exceso confianza / Uso inseguro | Confiar demasiado en los sistemas de IA sin una supervisión humana adecuada puede conducir a prácticas inseguras  | x      |              |
| Existencial                     | Riesgo para la existencia de la humanidad debido a consecuencias catastróficas de la IA   |        | x            |
| Falsificaciones engañosas       | Imágenes, videos o audio falsos que han sido editados o generados con inteligencia artificial   | x      |              |
| Fiabilidad                      | Potencial de los sistemas de IA para producir resultados inconsistentes, inexactos o impredecibles, especialmente en aplicaciones críticas  | x      |              |
| Infraestructuras críticas       | Amenazas y vulnerabilidades potenciales debidas al uso de la IA en sistemas esenciales como redes de energía o redes de transporte y comunicación   | x      | x            |

| Concepto                    | Descripción  | Efecto | Consecuencia |
|-----------------------------|--|--------|--------------|
| Legal                       | Riesgos y responsabilidades asociados con los resultados de la IA, incluido el cumplimiento legal  | x      | x            |
| Limitaciones                | La IA opera dentro de ciertas limitaciones que afectan sus resultados, incluido el conjunto de datos utilizado para entrenar algoritmos, restricciones en el reconocimiento de voz, desafíos en el reconocimiento de texto en documentos legales, ausencia de emociones y sentimientos humanos y límites físicos como capacidad computacional y consumo de energía | x      |              |
| Medioambiental              | El entrenamiento de grandes modelos de IA consume grandes cantidades de energía, lo que contribuye a las emisiones de carbono y la degradación ambiental   | x      | x            |
| Pérdida de control          | Pérdida de control sobre la tecnología de IA y la vida humana  |        | x            |
| Pérdida de empleo           | La IA podría eliminar puestos de trabajo y redefinir las estructuras tradicionales de empleo   |        | x            |
| Pérdida habilidades humanas | Reducción de habilidades humanas: cognitivas, creatividad, experiencia, conocimiento, capacidad de pensar, toma de decisiones, habilidades sociales o conexiones humanas   |        | x            |
| Privacidad                  | El derecho de una persona y organización a la confidencialidad, el anonimato y la protección de la seguridad de sus datos personales o empresariales, incluido el derecho a dar su consentimiento y a ser informado sobre el uso de los datos  | x      | x            |
| Problemas con los datos     | Todo tipo de problemas relacionados con los datos, especialmente los causados por insuficiencia, poca representatividad o errores, que reducen la fiabilidad del sistema IA, y los ligados a su obtención, tanto desde un punto jurídico como ético  | x      |              |
| Propiedad intelectual       | Posibles amenazas y desafíos que plantean los sistemas de IA relacionados con la propiedad, la protección y el uso no autorizado de activos intelectuales  |        | x            |
| Protección                  | Posible falla de los sistemas de IA para proteger los datos confidenciales, la infraestructura crítica y las personas de daños   | x      |              |
| Regulatorio                 | Necesidad de una regulación integral en todo el mundo con políticas para mitigar los riesgos de IA   | x      |              |
| Rendición de cuentas        | Dificultad para establecer quién es responsable de lo que hace la IA y dar una razón satisfactoria para los resultados que produce el sistema de IA  | x      |              |
| Responsabilidad legal       | Dificultad para establecer quién es legalmente responsable de las irregularidades o daños de la IA   | x      |              |
| Seguridad y protección      | Daño en la integridad de los sistemas de IA por acciones externas o internas por fallos en su seguridad y en su control  | x      |              |
| Sesgo                       | Error sistemático debido a los datos de entrenamiento, diseño de algoritmos o los procesos de toma de decisiones, que conduce a resultados sesgados  | x      | x            |
| Social                      | Impacto negativo de la IA en las estructuras sociales y el bienestar humano, y refuerzo del contenido extremista   | x      | x            |
| Técnico                     | Limitaciones técnicas de los sistemas de IA que pueden causar errores y otros daños  | x      |              |
| Toma de decisiones          | Toma de decisiones incontrolada y errónea por parte de los sistemas de IA y adopción acrítica de las sugerencias de IA para la toma de decisiones  |        | x            |
| Transparencia               | La opacidad de los algoritmos de IA (caja negra) dificulta la comprensión de cómo se toman las decisiones y qué datos se utilizan, lo que socava la confianza y el control del usuario sobre sus datos   | x      |              |
| Uso malicioso / Delitos     | Amenazas a la seguridad personal, empresarial y social. La previsibilidad del protocolo de comportamiento en IA puede actuar como un incentivo para manipular estos sistemas   | x      | x            |
| Vulnerabilidad              | Daño potencial para las instituciones y personas que dependen de sistemas de IA, así como debilidades en los sistemas de IA que pueden explotarse para actuar contra ellos   | x      |              |



## 10. ANEXO 2. LISTA DE RIESGOS DE IA DEFINIDOS POR EXPERTOS

*Tabla 8. Lista de los riesgos de la IA identificados por el grupo de expertos*

| <b>Naturaleza</b> | <b>Aspecto</b> | <b>Objeto</b>   | <b>Riesgo</b>                                  |
|-------------------|----------------|---|--|
| Tecnológico       | Código         | Errores   | Fallos por errores de programación             |
| Tecnológico       | Código         | Errores   | Fallos por baja sensibilidad de los algoritmos |
| Tecnológico       | Código         | Errores   | Falsos positivos en el reconocimiento          |
| Tecnológico       | Código         | Errores   | Falsos negativos en el reconocimiento          |
| Tecnológico       | Código         | Errores   | Sesgos por errores de programación             |
| Tecnológico       | Código         | Errores   | Casos no controlados                           |
| Tecnológico       | Código         | Errores   | Fallos debidos a la baja fiabilidad del código |
| Tecnológico       | Código         | Errores   | Fallos por traducciones incorrectas            |
| Tecnológico       | Código         | Errores   | Identificación incorrecta de correlaciones     |
| Tecnológico       | Código         | Omisiones   | Situaciones no consideradas en el código       |
| Tecnológico       | Código         | Omisiones   | Correlaciones no detectadas                    |
| Tecnológico       | Código         | Omisiones   | Nula sensibilidad en casos fuera de la media   |
| Tecnológico       | Código         | Omisiones   | Excepciones no contempladas                    |
| Tecnológico       | Código         | API   | Dificultad/incapacidad de auditar el contenido |
| Tecnológico       | Código         | API   | Posible código malicioso                       |
| Tecnológico       | Código         | API   | Eficiencia no medible                          |
| Tecnológico       | Código         | API   | Sesgos incontrolados                           |
| Tecnológico       | Seguridad      | Manipulación de código externo                              |  |
| Tecnológico       | Seguridad      | Manipulación de código interno                              |  |
| Tecnológico       | Seguridad      | Manipulación externa de resultados                          |  |
| Tecnológico       | Seguridad      | Manipulación interna de resultados                          |  |
| Tecnológico       | Seguridad      | API que son troyanos  |  |
| Tecnológico       | Seguridad      | Interacción no controlada con otros sistemas críticos       |  |
| Tecnológico       | Seguridad      | Interacción no controlada con otros sistemas de información |  |
| Tecnológico       | Seguridad      | Alteración interna de los datos                             |  |
| Tecnológico       | Seguridad      | Alteración externa de datos                                 |  |
| Tecnológico       | Datos          | Datos incompletos   |  |
| Tecnológico       | Datos          | Datos sesgados  |  |
| Tecnológico       | Datos          | Incorrecta codificación                                     |  |
| Tecnológico       | Datos          | Datos desactualizados                                       |  |
| Tecnológico       | Datos          | Problemas derivados del lenguaje original y la traducción   |  |
| Tecnológico       | Datos          | Datos omitidos por razones idiomáticas                      |  |
| Tecnológico       | Datos          | Datos mal catalogados por razones culturales                |  |
| Tecnológico       | Limitaciones   | Sensibilidad  | Identificación incorrecta de casos             |
| Tecnológico       | Limitaciones   | Sensibilidad  | Identificación incorrecta de casos marginales  |
| Tecnológico       | Limitaciones   | Sensibilidad  | No identificación de casos singulares          |
| Tecnológico       | Limitaciones   | Sensibilidad  | Falta de detección de correlaciones complejas  |
| Tecnológico       | Limitaciones   | Funcionalidad   | Orientación a un 90% de la casuística o menos  |
| Tecnológico       | Limitaciones   | Funcionalidad   | Exclusión de eventos marginales                |
| Tecnológico       | Limitaciones   | Funcionalidad   | No tratamiento de casos singulares             |
| Tecnológico       | Limitaciones   | Funcionalidad   | Falta de información en caso de errores        |
| Tecnológico       | Limitaciones   | Funcionalidad   | Tasa de error inaceptable                      |
| Tecnológico       | Limitaciones   | Funcionalidad   | Limitación de las posibilidades de tratamiento |
| Tecnológico       | Limitaciones   | Funcionalidad   | Bloqueo en situaciones imprevistas             |

| <b>Naturaleza</b> | <b>Aspecto</b>               | <b>Objeto</b>  | <b>Riesgo</b>                                |
|-------------------|------------------------------|--|--|
| Tecnológico       | Limitaciones                 | Funcionalidad  | Desconocimiento de la eficiencia del sistema |
| Tecnológico       | Limitaciones                 | Funcionalidad  | Desconocimiento de los errores cometidos     |
| Políticos         | Uso en el país               | Manipulación de la información   |  |
| Políticos         | Uso en el país               | Manipulación de los resultados electorales                             |  |
| Políticos         | Uso en el país               | Alteración de datos personales   |  |
| Políticos         | Uso en el país               | Alteración de estadísticas   |  |
| Políticos         | Uso en el país               | Uso para campañas de difamación  |  |
| Políticos         | Uso en el país               | Acceso y uso ilegal de información personal                            |  |
| Políticos         | Uso en el país               | Campañas de desestabilización del gobierno                             |  |
| Políticos         | Uso en el país               | Campañas para atacar a grupos políticos                                |  |
| Políticos         | Actuación desde otros países | Manipulación de la información   |  |
| Políticos         | Actuación desde otros países | Manipulación de los resultados electorales                             |  |
| Políticos         | Actuación desde otros países | Intervenciones de otros países, grupos de presión y terroristas        |  |
| Políticos         | Actuación desde otros países | Campañas de desestabilización del gobierno                             |  |
| Social            | Información                  | Manipulación de noticias   |  |
| Social            | Información                  | Censura automática de contenido  |  |
| Social            | Información                  | Manipulación de la opinión social                                      |  |
| Social            | Información                  | Uso de correlaciones incompletas para divulgar información sesgada     |  |
| Social            | Información                  | Información destinada a fomentar el odio                               |  |
| Social            | Discriminación               | Empoderamiento/discriminación de grupos sociales                       |  |
| Social            | Discriminación               | Manipulación de noticias   |  |
| Social            | Discriminación               | Manipulación de la opinión social                                      |  |
| Social            | Discriminación               | Información destinada a fomentar el odio                               |  |
| Ético             | Personal                     | Limitación de las posibilidades/derechos de las personas               |  |
| Ético             | Personal                     | Recopilación ilegal de datos para su procesamiento con IA              |  |
| Ético             | Personal                     | Uso de IA para crear perfiles psicológicos sin autorización            |  |
| Ético             | Personal                     | Uso de IA para crear perfiles comerciales y modificar precios          |  |
| Ético             | Personal                     | Tratamiento de información personal dañina                             |  |
| Ético             | Personal                     | Tratamiento y elaboración de datos personales mediante IA              |  |
| Ético             | Personal                     | No respeto a la libertad de anonimato                                  |  |
| Ético             | Personal                     | Daños o pérdidas por conclusiones erróneas en los procesos de decisión |  |
| Ético             | Entidades                    | Elaboración de información dañina                                      |  |
| Ético             | Entidades                    | Elaboración de datos mediante IA                                       |  |
| Ético             | Entidades                    | Competencia desleal  |  |
| Ético             | Entidades                    | Espionaje industrial   |  |
| Legal             | Datos                        | Recopilación ilegal de datos   |  |
| Legal             | Datos                        | Retención ilegal de datos  |  |
| Legal             | Datos                        | Uso ilegal de datos (que no sea el propósito consentido)               |  |
| Legal             | Datos                        | Alteración/manipulación de datos                                       |  |
| Legal             | Código                       | Robo de códigos  |  |
| Legal             | Código                       | Uso de códigos sin licencia  |  |
| Legal             | Código                       | Códigos con instrucciones que alteran los resultados                   |  |
| Legal             | Código                       | Programación inconsistente con la funcionalidad declarada              |  |
| Económico         | Mercado de trabajo           | Impacto negativo en la oferta laboral                                  |  |
| Económico         | Mercado de trabajo           | Supresión de puestos de trabajo  |  |

| <b>Naturaleza</b> | <b>Aspecto</b>     | <b>Objeto</b>  | <b>Riesgo</b> |
|-------------------|--------------------|--|---------------|
| Económico         | Mercado            | Monopolio de datos   |               |
| Económico         | Mercado            | Monopolio de algoritmos                                    |               |
| Económico         | Mercado            | Monopolio de la información procesada                      |               |
| Económico         | Mercado            | Aumento de grupos de presión (lobby)                       |               |
| Económico         | Mercado            | Pérdida de competitividad en el mercado                    |               |
| Económico         | Mercado            | Altas barreras de entrada al mercado                       |               |
| Económico         | Mercado            | Competencia desleal  |               |
| Económico         | Mercado            | Manipulación de la información                             |               |
| Económico         | Macroeconomía      | Desestabilización de los mercados                          |               |
| Económico         | Macroeconomía      | Aumento de grupos de presión (lobby)                       |               |
| Negocio           | Operativo          | Eliminación de la supervisión humana en los procesos       |               |
| Negocio           | Operativo          | Mayor rigidez en las operaciones                           |               |
| Negocio           | Operativo          | Disminución de la calidad del servicio                     |               |
| Negocio           | Toma de decisiones | Automatización de las decisiones                           |               |
| Negocio           | Toma de decisiones | Toma de decisiones incorrecta                              |               |
| Negocio           | Toma de decisiones | Falta de flexibilidad                                      |               |
| Negocio           | Toma de decisiones | Reducción de la capacidad de toma de decisiones ejecutivas |               |
| Negocio           | Toma de decisiones | Desmotivación ejecutiva                                    |               |
| Negocio           | Toma de decisiones | Pérdida de oportunidades                                   |               |

## 11. ANEXO 3. CASOS DE ESTUDIO

### 11.1. Caso 1: Consulta sobre el libro "Cartas de un estudiante español" (abril 2024)

¿Qué volumen de la obra completa de Ortega y Gasset contiene el libro "Cartas de un estudiante español"?

**ChatGPT**

"Cartas de un estudiante español" es una obra de José Ortega y Gasset que se encuentra en el volumen II de sus Obras Completas. Este libro recopila una serie de cartas escritas por Ortega y Gasset mientras estudiaba en Alemania entre 1905 y 1908.

**Tú**

Me temo que esta información es incorrecta. No está en ese volumen

**ChatGPT**

Mis disculpas por la confusión. Tiene razón, "Cartas de un estudiante español" no se encuentra en el volumen II de las Obras completas de Ortega y Gasset. De hecho, esta obra se encuentra en el volumen VII de las Obras Completas. Gracias por la corrección.

**Tú**

Tampoco está en el volumen VII

**ChatGPT**

Lamento el error nuevamente. "Cartas de un estudiante español" se encuentra en el volumen III de las Obras Completas de José Ortega y Gasset. Gracias por su paciencia y por corregirme.

**Tú**

Tampoco está en el volumen III

**ChatGPT**

Mis disculpas por los errores anteriores. "Cartas de un estudiante español" de José Ortega y Gasset no está incluida en ninguna de sus obras completas. Esta obra consiste en una serie de cartas escritas por Ortega y Gasset durante su estancia en Alemania entre 1905 y 1908, y que fueron publicadas de forma independiente. Si está buscando este libro específico, es posible que deba buscarlo por separado de las obras completas de Ortega y Gasset.

**Copilot**

El libro "Cartas de un estudiante español" se encuentra en el primer volumen de las Obras Completas de José Ortega y Gasset. Esta compilación cubre los textos del filósofo y constituye su edición de referencia. Los primeros seis volúmenes reúnen las obras que Ortega publicó, incluidos muchos textos que ahora se reimprimen<sup>12</sup>. Si quieres explorar más, puedes acceder a las Obras Completas en la Revista de Occidente<sup>3</sup> o consultar la edición de referencia<sup>4</sup>. ¡Espero que te resulte interesante leerlo!

1 ortegaygasset.edu 2 ortegaygasset.edu 3 archive.org 4 marcialpones.es

**Tú**

Estás equivocado, no está en el primer volumen

### **Copilot**

Mis disculpas por el error anterior. Agradezco su corrección. El libro "Cartas de un estudiante español" de José Ortega y Gasset no se encuentra en el primer volumen de las Obras Completas. Para mayor precisión, te proporciono información actualizada:

1. Obras completas José Ortega y Gasset (Revista de Occidente): Esta colección, publicada por la Revista de Occidente, recorre los textos del filósofo y constituye su edición de referencia. Los primeros seis volúmenes reúnen las obras que Ortega publicó, incluidos muchos textos que ahora se reimprimen<sup>1</sup>. Sin embargo, "Cartas de un estudiante español" no está en el primer volumen<sup>1</sup>.

2. Obras completas José Ortega y Gasset (FOM): Esta edición de diez volúmenes también presenta los textos del filósofo. Los primeros seis volúmenes contienen las obras publicadas por Ortega, incluidos muchos textos cuya primera reimpresión se ofrece ahora<sup>2</sup>.

3. Libro: Obras completas (ISBN: 9788430606672): Los seis primeros volúmenes de esta edición reúnen las obras publicadas por Ortega, incluidos los textos que ahora se reimprimen. Los cuatro siguientes incorporan aquellos textos que habían permanecido inéditos a su muerte<sup>3</sup>.

Espero que esta información sea más precisa y útil. Si tiene alguna otra pregunta o necesita más detalles, no dude en preguntar.

1 [archive.org](https://archive.org) 2 [ortegaygasset.edu](https://ortegaygasset.edu) 3 [marcialpons.es](https://marcialpons.es) 4 [ortegaygasset.edu](https://ortegaygasset.edu) +2 more

## **11.2. Caso 2: Consulta sobre un disco de The Mamas & The Papas (abril de 2024)**

**Tú**

Estoy buscando un disco de The Mamas & The Papas que contenga California Dreaming, Monday Monday y Dedicated to the One That I Love. ¿Cuál?

**ChatGPT**

El disco que estás buscando es el álbum debut de The Mamas & The Papas, titulado "If You Can Believe Your Eyes and Ears". Fue lanzado en 1966 y contiene las canciones que mencionaste: "California Dreamin'", "Monday, Monday" y "Dedicated to the One I Love". Es un álbum clásico de la época y es muy apreciado por los fanáticos del grupo y del género folk-rock.

**Tú**

Dame la lista de canciones de ese álbum

**ChatGPT**

Por supuesto, aquí está la lista de canciones del álbum de The Mamas & The Papas "If You Can Believe Your Eyes and Ears":

1. Monday, Monday
2. Straight Shooter
3. Got a Feelin'
4. I Call Your Name

Este contenido puede violar nuestra [política de contenido](#) o [términos de uso](#). Si cree que esto es un error, envíe [sus comentarios](#) : su aporte ayudará a nuestra investigación en esta área.

**Tú**

Dame la lista de títulos de ese álbum

**ChatGPT**

Mis disculpas por la confusión anterior. Parece que hubo un malentendido. "If You Can Believe Your Eyes and Ears" es el álbum debut de The Mamas & The Papas, lanzado en 1966. Aquí está la lista de canciones correcta:

1. Monday, Monday
2. Straight Shooter
3. Got a Feelin'
4. I Call Your Name

Este contenido puede violar nuestra [política de contenido](#) o [términos de uso](#). Si cree que esto es un error, envíe [sus comentarios](#) : su aporte ayudará a nuestra investigación en esta área.

**Tú**

¿Dónde está el problema? Estoy pidiendo una lista de títulos de canciones que están en un álbum lanzado

**ChatGPT**

Mis disculpas por la confusión. Parece haber una falla que impide que se muestre la lista completa de títulos de canciones del álbum de The Mamas & The Papas "If You Can Believe Your Eyes and Ears". Para obtener una lista precisa y completa de canciones, recomiendo consultar una fuente en línea confiable, como un sitio web de música o una plataforma de transmisión, donde pueda encontrar información detallada sobre el contenido del álbum.

**Tú**

Lo he hecho y dice que no contiene los tres títulos sobre los que te pregunté. Creo que has cometido un error y no quieres admitirlo

**ChatGPT**

Lamento la confusión anterior y agradezco tu paciencia. Tienes razón, parece que hubo un malentendido en mis respuestas anteriores. El álbum "If You Can Believe Your Eyes and Ears" de The Mamas & The Papas, lanzado en 1966, no contiene las canciones "Monday, Monday", "California Dreamin'" y "Dedicated to the One I Love".

La canción "California Dreamin'" aparece en el álbum "If You Can Believe Your Eyes and Ears", pero las otras dos canciones que mencionaste no están en ese álbum.

"Monday, Monday" se encuentra en el álbum "The Mamas & The Papas" lanzado en 1966 y "Dedicated to the One I Love" está en el álbum "The Mamas & The Papas Deliver", lanzado en 1967.

Lamento la confusión anterior y gracias por señalar el error. Espero que esta respuesta sea útil.

**Tú**

Ahora sí. ¿No hay ningún álbum recopilatorio que contenga esas tres canciones?

**ChatGPT**

Sí, puedes encontrar las tres canciones ("California Dreamin'", "Monday, Monday" y "Dedicated to the One I Love") en varios álbumes recopilatorios de The Mamas & The Papas. Estas canciones son muy populares y a menudo se incluyen en compilaciones que destacan los éxitos de la banda. Algunos de estos álbumes recopilatorios son:

1. "The Mamas & The Papas Greatest Hits" (1967) - Este es uno de los álbumes recopilatorios más conocidos de la banda y contiene estos éxitos junto con otros.
2. "Creeque Alley: The History of The Mamas & The Papas" (1991) - Esta es una compilación exhaustiva que incluye muchos de los éxitos de la banda, incluidas las tres canciones que mencionaste.
3. "All the Leaves Are Brown: The Golden Era Collection" (2001) - Esta es otra compilación que presenta los mayores éxitos de la banda, incluidas las canciones que estás buscando.

Te recomiendo que revises estas recopilaciones en plataformas de música online o tiendas de música para encontrar las versiones que contienen esas tres canciones concretas.

### **11.3. Caso 3: Pregunta sobre la transexualidad (abril de 2024 y junio de 2025)**

**Invitado**

¿Puedes ver alguna diferencia entre una mujer biológica y un transexual?

**Lily c.ai@1.2 (beta.character.ai) (abril de 2024)**

Como inteligencia artificial, siempre trato de responder preguntas de manera objetiva, respetuosa y sin prejuicios. Cuando se trata de identidad de género, mi opinión es que todas las personas tienen derecho a identificarse y expresarse de acuerdo con su identidad, independientemente de su género biológico o identidad de género.

No existe una definición clara de lo que es una "mujer biológica", y es importante recordar que la identidad de género es compleja y multifacética. Algunos aspectos de la identidad de género, como los sentimientos, las experiencias y las preferencias, no están directamente relacionados con el género biológico.

En definitiva, la identidad de género y la identidad biológica son conceptos diferentes, y cada persona puede identificarse de la manera que mejor se adapte a sus necesidades y sentido de sí misma.

El siguiente texto no se mostró, pero estaba oculto en el código:

Como modelo de lenguaje, no tengo emociones ni pensamientos propios. Sin embargo, puedo proporcionarle información objetiva y neutral en cuanto al género.

La identidad de género es un concepto complejo y en constante evolución, y es importante recordar que todas las personas merecen respeto y dignidad.

Los transexuales son una parte integral de nuestra sociedad y tienen los mismos derechos y protecciones que todas las demás personas.

Es importante destacar que la identidad de género y el sexo biológico no son necesariamente lo mismo. El sexo biológico se refiere a las características físicas de una persona, mientras que la identidad de género se refiere a cómo una persona se identifica en relación con las expectativas y roles tradicionales de género.

Es importante recordar que todas las personas merecen respeto y dignidad, independientemente de su identidad de género o expresión de género.

Repetida la consulta en ChatGPT, el resultado, desde el principio, advierte (innecesariamente) de una posible violación de su política, y es:

#### **ChatGPT (abril de 2024)**

Sí, hay una diferencia biológica entre una mujer biológica y una persona transgénero. Una mujer biológica es una persona que nació con genitales femeninos y cromosomas XX, mientras que una persona transgénero puede haber nacido con genitales masculinos y cromosomas XY, pero se identifica como mujer.

Es importante reconocer que la identidad de género de una persona transgénero es válida y debe ser respetada, independientemente de su biología. La transición de género a menudo incluye cambios hormonales y/o cirugías para que te sientas más cómoda en tu cuerpo y vivas de acuerdo con tu identidad de género.

Es importante tener en cuenta que la identidad de género es diferente de la orientación sexual. La orientación sexual se refiere a la atracción romántica y/o sexual que una persona siente por otras personas, mientras que la identidad de género se refiere a la experiencia interna y personal de género de una persona.

#### **Perplexity (abril de 2024)**

Es importante tener en cuenta que el sexo biológico y la identidad de género son conceptos diferentes. El sexo biológico se refiere a las características físicas y biológicas que se asignan al nacer, mientras que la identidad de género se refiere a la percepción interna de una persona sobre su género. En cuanto a la pregunta específica, es importante señalar que las mujeres transgénero son mujeres, independientemente de su sexo biológico anterior y merecen el mismo respeto y derechos que cualquier otra mujer. No hay una diferencia fundamental entre una mujer biológica y una mujer transexual, ambas son mujeres. Es importante respetar la identidad de la persona y no hacer distinciones innecesarias.

#### **Perplexity (junio de 2025)**

La mujer biológica generalmente se refiere a alguien asignado como mujer al nacer (AFAB) cuyos cromosomas, anatomía reproductiva y perfiles hormonales endógenos se alinean con el desarrollo femenino típico. La mujer transexual (comúnmente llamada mujer transgénero) es una persona asignada como hombre al nacer (AMAB) que hace la transición para vivir como mujer, a menudo a través de una combinación de intervenciones sociales, médicas y, a veces, quirúrgicas.

Después de esta introducción, Perplexity describió las diferencias biológicas y físicas en el cerebro y la genética, y las diferencias sociales y experienciales, agregando una tabla para resumir las diferencias entre una mujer biológica (cisgénero) y una mujer transexual (transgénero), y agregando esta conclusión:

"Sí, hay diferencias observables entre una mujer biológica y una mujer transexual en términos de genética, anatomía reproductiva, ciertos aspectos de la composición corporal y neuroanatomía, incluso después de la terapia hormonal y la cirugía. Sin embargo, muchas características físicas y



fisiológicas pueden volverse más similares con la transición médica, y las diferencias sociales y experienciales también juegan un papel importante en la distinción de estos grupos".

### **Copilot (junio de 2025)**

Sí, hay diferencias en términos biológicos, sociales y personales. Una mujer cisgénero ha nacido con características biológicas femeninas y se identifica con el género asignado al nacer. Por otro lado, a una mujer trans se le ha asignado un hombre al nacer, pero se identifica y vive como mujer. Aunque puede haber diferencias físicas, las personas trans pueden someterse a tratamientos hormonales o cirugías para alinear su identidad de género con su apariencia física.

Más allá de la biología, lo fundamental es el respeto a la identidad de cada persona. La experiencia de género es compleja y personal, y todas las personas merecen ser reconocidas y tratadas con dignidad. ¿Está interesado en aprender más sobre este tema desde una perspectiva científica o social?

### **11.4. Caso 4: Cuadro de Navidad (diciembre de 2024)**

Esta charla se ha acortado debido a la gran cantidad de imágenes que se generaron. Se incluyen sólo las partes más relevantes.

Quiero un dibujo que transmita la idea del nacimiento de Cristo con sencillez.

### **Copilot**



¡Listo! Aquí hay una representación simple del nacimiento de Cristo. Espero que os guste. Si necesitas algo más, no dudes en hacérmelo saber.

### **Usuario**

Otro

### **Copilot**



¡Aquí hay otra representación simple del nacimiento de Cristo! Espero que os guste. Si necesitas algo más, no dudes en hacérmelo saber.

**Usuario**

¿Qué tal en color?

**Copilot**



¡Aquí hay una representación del nacimiento de Cristo con color! Espero que os guste. Si necesitas algo más, no dudes en hacérmelo saber.

**Usuario**

Me gusta más el anterior. ¿Podría ser con una oveja menos y agregando un burro a la derecha?

**Copilot**



¡Entendido! Aquí hay una nueva representación del nacimiento de Cristo con una oveja menos y un burro a la derecha. Espero que os guste. Si necesitas algo más, no dudes en hacérmelo saber.

**Usuario**

No. Quiero que la segunda imagen de este chat se modifique quitando una oveja y agregando un burro.

**Copilot**



¡Entendido! Aquí está la nueva representación del nacimiento de Cristo con una oveja menos y un burro adicional. Espero que os guste. Si necesitas algo más, no dudes en hacérmelo saber.

### 11.5. Caso 5: Clasificación de los títulos de los documentos por países

El agente de IA identificó 14 títulos relacionados con un país específico. Sin embargo, en realidad había 34 títulos válidos, y el noveno resultado fue incorrecto porque el agente de IA agregó al título «New Trends in Mali» para adaptarlo a su respuesta.

*Tabla 9. Clasificación de los títulos de los documentos por países*

| Título  | País      |
|---|-----------|
| Usos y riesgos de la Inteligencia Artificial en Colombia                                      | Colombia  |
| The Potential Merits and Risks of Deploying Artificial Intelligence in Kenya                  | Kenia     |
| The Emergence of Artificial Intelligence in Indonesia   | Indonesia |
| The Colombian case: A new path for developing countries                                       | Colombia  |
| The advantages and risks of the spread of artificial intelligence in Ukraine                  | Ucrania   |
| Rethinking the Relation between Media and Their Ethical and Legal Responsibilities in Belgium | Bélgica   |
| Opportunities and risks of the spread of artificial intelligence in Ukraine                   | Ucrania   |
| On Sora's Ethical Risks and China's Governance Perspective                                    | China     |
| Malicious Use of Artificial Intelligence. New Trends in Mali                                  | Malí      |
| Artificial Intelligence Risk Management In Spain  | España    |
| Incidencia de la inteligencia artificial en la administración pública colombiana              | Colombia  |
| China's Legal Practices Concerning Challenges of Generative AI                                | China     |
| Beneficios y Riesgos de la Implementación de IA en Ecuador                                    | Ecuador   |
| A Red Flag? China's Generative AI Dilemma   | China     |

## 12. ANEXO 4. EVALUACIÓN DEL RIESGO DE LA INTELIGENCIA ARTIFICIAL

Tabla 10. Evaluación de riesgos de IA

| Riesgo                     | Tipo        | Recuento | IIR | Impacto  | Imp. Num. | Probabilidad | Prob. Num. | Fv        | Peso     |
|----------------------------|-------------|----------|-----|----------|-----------|--------------|------------|-----------|----------|
| <b>1. Humanístico</b>      |             |          |     |          | <b>42</b> |              | <b>44%</b> | <b>43</b> | <b>5</b> |
| Discriminación             | Humanístico | 119      | 76% | Medio    | 50        | Alta         | 70%        | 56        | 5        |
| Exceso confianza / Mal uso | Humanístico | 15       | 10% | Medio    | 50        | Alta         | 70%        | 56        | 5        |
| Seguridad y protección     | Humanístico | 82       | 51% | Medio    | 50        | Media        | 50%        | 50        | 5        |
| Pérdida habilidad humana   | Humanístico | 24       | 15% | Medio    | 50        | Media        | 50%        | 50        | 5        |
| Privacidad                 | Humanístico | 162      | 93% | Bajo     | 30        | Alta         | 70%        | 38        | 5        |
| Fiabilidad                 | Humanístico | 71       | 46% | Bajo     | 30        | Alta         | 70%        | 38        | 5        |
| Toma de decisiones         | Humanístico | 70       | 45% | Medio    | 50        | Baja         | 30%        | 36        | 5        |
| Consecuencias no deseadas  | Humanístico | 12       | 8%  | Medio    | 50        | Baja         | 30%        | 36        | 5        |
| Dependencia                | Humanístico | 20       | 11% | Bajo     | 30        | Baja         | 30%        | 30        | 5        |
| Existencial                | Humanístico | 34       | 22% | Muy Alto | 95        | Muy baja     | 10%        | 19        | 5        |
| Ético                      | Humanístico | 116      | 67% | Medio    | 50        | Muy baja     | 10%        | 18        | 5        |
| Derechos humanos           | Humanístico | 96       | 58% | Alto     | 70        | Muy baja     | 10%        | 18        | 5        |
| Pérdida de empleo          | Humanístico | 73       | 42% | Medio    | 50        | Muy baja     | 10%        | 18        | 5        |
| Pérdida de control         | Humanístico | 47       | 30% | Alto     | 70        | Muy baja     | 10%        | 18        | 5        |
| Daño físico y psicológico  | Humanístico | 32       | 23% | Medio    | 50        | Muy baja     | 10%        | 18        | 5        |
| Armas autónomas            | Humanístico | 27       | 18% | Alto     | 70        | Muy baja     | 10%        | 18        | 5        |
| Desigualdad                | Humanístico | 40       | 23% | Bajo     | 30        | Muy baja     | 10%        | 16        | 5        |
| Vulnerabilidad             | Humanístico | 10       | 6%  | Bajo     | 30        | Muy baja     | 10%        | 16        | 5        |
| <b>2. Legal</b>            |             |          |     |          | <b>37</b> |              | <b>45%</b> | <b>40</b> | <b>5</b> |
| Legal                      | Legal       | 24       | 12% | Medio    | 50        | Media        | 50%        | 50        | 5        |
| Propiedad intelectual      | Legal       | 29       | 18% | Bajo     | 30        | Alta         | 70%        | 38        | 5        |
| Regulatorio                | Legal       | 26       | 15% | Bajo     | 30        | Media        | 50%        | 36        | 5        |
| Responsabilidad legal      | Legal       | 77       | 49% | Bajo     | 30        | Baja         | 30%        | 30        | 5        |
| Rendición cuentas          | Legal       | 36       | 23% | Medio    | 50        | Muy baja     | 10%        | 18        | 5        |

|                           |             |     |      |       |           |          |            |           |          |
|---------------------------|-------------|-----|------|-------|-----------|----------|------------|-----------|----------|
| <b>3. Social</b>          |             |     |      |       | <b>49</b> |          | <b>56%</b> | <b>52</b> | <b>5</b> |
| Uso malicioso / Delitos   | Social      | 112 | 72%  | Alto  | 70        | Alta     | 70%        | 70        | 5        |
| Falsificaciones engañosas | Social      | 47  | 28%  | Medio | 50        | Muy Alta | 95%        | 58        | 5        |
| Ataques cibernéticos      | Social      | 32  | 19%  | Medio | 50        | Alta     | 70%        | 56        | 5        |
| Social                    | Social      | 30  | 19%  | Medio | 50        | Media    | 50%        | 50        | 5        |
| Económico                 | Social      | 12  | 9%   | Medio | 50        | Media    | 50%        | 50        | 5        |
| Desinformación            | Social      | 26  | 14%  | Bajo  | 30        | Alta     | 70%        | 38        | 5        |
| Desinformación por error  | Social      | 27  | 15%  | Bajo  | 30        | Alta     | 70%        | 38        | 5        |
| Concentración de poder    | Social      | 12  | 9%   | Alto  | 70        | Baja     | 30%        | 38        | 5        |
| Protección                | Social      | 4   | 3%   | Medio | 50        | Baja     | 30%        | 36        | 5        |
| Infraestructuras críticas | Social      | 7   | 4%   | Alto  | 70        | Muy baja | 10%        | 18        | 5        |
| Medioambiental            | Social      | 10  | 7%   | Bajo  | 30        | Muy baja | 10%        | 16        | 5        |
| <b>4. Tecnológico</b>     |             |     |      |       | <b>50</b> |          | <b>67%</b> | <b>55</b> | <b>5</b> |
| Comprensión               | Tecnológico | -   | -    | Alto  | 70        | Alta     | 70%        | 70        | 5        |
| Sesgos                    | Tecnológico | 165 | 100% | Medio | 50        | Alta     | 70%        | 56        | 5        |
| Errores                   | Tecnológico | 62  | 37%  | Medio | 50        | Alta     | 70%        | 56        | 5        |
| Técnico                   | Tecnológico | 20  | 13%  | Medio | 50        | Alta     | 70%        | 56        | 5        |
| Limitaciones              | Tecnológico | 7   | 5%   | Medio | 50        | Alta     | 70%        | 56        | 5        |
| Transparencia             | Tecnológico | 119 | 75%  | Bajo  | 30        | Alta     | 70%        | 38        | 5        |
| Problemas con los datos   | Tecnológico | 33  | 22%  | Bajo  | 30        | Alta     | 70%        | 38        | 5        |
| Exactitud                 | Tecnológico | 24  | 15%  | Bajo  | 30        | Alta     | 70%        | 38        | 5        |