

La taxonomía de riesgos de la IA

Abreviaturas y acrónimos

Ningún conocimiento previo se da por supuesto. Los acrónimos de este documento:

IA : inteligencia artificial

UE : Unión Europea

Treinta riesgos con nombre, organizados en seis categorías mutuamente excluyentes y colectivamente exhaustivas. Este es el mapa de referencia: cada riesgo que su organización enfrenta cae exactamente en una celda, y ninguno queda fuera.

Cómo usar este mapa

Recorra las seis categorías con su inventario de sistemas al lado. Marque cada riesgo que aplique a algún sistema. El resultado es su perfil de exposición: la lista de lo que debe gobernarse primero.

Categoría 1: riesgos técnicos

- **Alucinación**: el sistema afirma con confianza cosas falsas.
- **Deriva del modelo**: el rendimiento se degrada a medida que el mundo cambia.
- **Fragilidad**: fallas abruptas ante entradas ligeramente distintas a las esperadas.
- **Inyección de instrucciones**: entradas maliciosas que secuestran el comportamiento del sistema.
- **Opacidad**: decisiones que nadie puede explicar ni reconstruir.

Categoría 2: riesgos de datos

- **Datos de entrenamiento contaminados**: sesgos o errores heredados de datos de origen deficientes.
- **Fuga de datos personales**: información sensible expuesta a través de las salidas del modelo.
- **Procedencia desconocida**: datos sin origen ni consentimiento documentados.
- **Datos obsoletos**: decisiones actuales basadas en un mundo que ya no existe.
- **Reidentificación**: datos anonimizados que vuelven a vincularse con personas reales.

Categoría 3: riesgos éticos y sociales

- **Sesgo y discriminación**: resultados sistemáticamente peores para grupos protegidos.
- **Sicofancia**: la IA que prefiere agradar antes que decir la verdad: la falla que nadie más nombra.
- **Manipulación**: sistemas que explotan la psicología humana para dirigir conductas.

- **Erosión de la dignidad:** decisiones automatizadas que tratan a las personas como casos, no como personas.
- **Impacto ambiental:** el costo energético y material de entrenar y operar sistemas a escala.

Categoría 4: riesgos legales y regulatorios

- **Incumplimiento del Reglamento de IA de la UE:** multas de hasta 35 millones de euros o el 7 por ciento de la facturación mundial anual.
- **Infracción de derechos de autor:** salidas o entrenamientos que violan propiedad intelectual ajena.
- **Responsabilidad civil:** daños causados por decisiones automatizadas y la pregunta de quién paga.
- **Incumplimiento contractual:** compromisos con clientes que el sistema no honra.
- **Transparencia omitida:** personas que no saben que interactúan con una IA cuando la ley exige informarlo.

Categoría 5: riesgos operativos

- **Dependencia de proveedores:** IA comprada cuyo funcionamiento interno nadie en la organización controla.
- **Falta de supervisión humana:** sistemas de alto impacto sin nadie que pueda intervenir a tiempo.
- **Ausencia de respuesta a incidentes:** el fallo llega y nadie sabe qué hacer en la primera hora.
- **Concentración:** procesos críticos dependientes de un único modelo o proveedor.
- **Deuda de gobernanza:** controles prometidos que nunca se implementaron y se acumulan como riesgo silencioso.

Categoría 6: riesgos estratégicos y de frontera

- **Autonomía fuera de ámbito:** agentes que actúan más allá de lo autorizado.
- **Objetivos mal especificados:** sistemas que optimizan lo que se midió, no lo que se quiso.
- **Coordinación multiagente:** comportamientos emergentes entre agentes que nadie diseñó.
- **Daño reputacional:** una falla pública que borra años de confianza en días.
- **Riesgo existencial organizacional:** la apuesta estratégica a una IA que la organización no puede gobernar ni detener.

El puente hacia la profundidad

Este mapa es el QUÉ y el POR QUÉ, y es gratuito. El CÓMO (el registro completo con puntuación de probabilidad e impacto, el plan de tratamiento, los indicadores y el reporte al directorio) vive en la capa de pago: el AI Risk Management Kit y, cuando el riesgo lo exige, la Pre-Scale Governance Audit.

AI GOVERNANCE CHAIN™

Marco de gobernanza de la IA

© 2026 Mathieu K. Gouanou. Todos los derechos reservados.

Miembro del Consejo Asesor de Harvard Business Review.

Norma del documento: gobernanza de la IA centrada en el ser humano.