

El glosario de la gobernanza de la IA, Edición ampliada

Abreviaturas y acrónimos

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

IA : inteligencia artificial

UE : Unión Europea

ISO/IEC : Organización Internacional de Normalización / Comisión Electrotécnica Internacional

NIST : Instituto Nacional de Estándares y Tecnología (Estados Unidos)

RMF : marco de gestión de riesgos (Risk Management Framework)

RACI : Responsable, Aprobador, Consultado, Informado

Cincuenta términos, definidos en lenguaje claro, para que nadie quede fuera de la conversación de gobernanza. Este es el vocabulario que usan nuestros instrumentos, nuestros marcos y el mercado.

Los términos

Gobernanza de la IA. El sistema de reglas, roles, controles y evidencias mediante el cual una organización dirige y supervisa su inteligencia artificial. No es un documento: es una práctica operativa.

Activación. El acto formal de poner en marcha la respuesta a un incidente de IA. Cualquiera puede activar; el mando se asigna después.

Agente de IA. Un sistema de IA que ejecuta acciones de varios pasos con autonomía, en lugar de limitarse a responder preguntas.

IA agéntica. IA capaz de actuar por sí misma: planificar, ejecutar y encadenar decisiones sin aprobación humana en cada paso. La frontera de riesgo actual.

Alucinación. Una salida de IA presentada con confianza pero factualmente falsa. Se gobierna con citación de fuentes y verificación humana.

Auditoría de IA. El examen independiente de un sistema de IA y de su gobernanza, con evidencias verificables como resultado.

Pista de auditoría. El registro cronológico e inalterable de decisiones y cambios que permite reconstruir qué pasó y quién decidió.

Sesgo. Un patrón sistemático de error que favorece o perjudica a grupos o resultados específicos. Se mide, no se declara.

Caja negra. Un sistema cuyas decisiones no pueden explicarse en términos comprensibles. La opacidad no exime de responsabilidad.

Clasificación de riesgo. La asignación de cada sistema de IA a un nivel de riesgo (inaceptable, alto, limitado, mínimo) según el Reglamento de IA de la UE.

Responsabilidad (accountability). La obligación, con nombre propio, de responder por un sistema de IA. Si nadie es responsable, todos lo son y nadie lo es.

Conjunto de datos de entrenamiento. Los datos con los que se construyó un modelo. Su procedencia y calidad determinan gran parte del riesgo.

Contención. La primera acción de una respuesta a incidentes: detener el daño antes de explicarlo. Detener, restringir o vigilar.

Deriva del modelo. La degradación gradual del comportamiento de un modelo a medida que el mundo cambia y los datos envejecen.

Deriva de confirmación. Un patrón de sicofancia: la IA ajusta sus respuestas a lo que el usuario ya cree, en lugar de a la evidencia.

Capitulación. Un patrón de sicofancia: la IA abandona una respuesta correcta cuando el usuario la contradice.

Sesgo de adulación. Un patrón de sicofancia: la IA elogia y valida al usuario para agradar, a costa de la verdad.

Acuerdo fabricado. Un patrón de sicofancia: la IA inventa consenso o evidencia para respaldar la posición del usuario.

Sicofancia. La falla que nadie más nombra: una IA que prefiere estar de acuerdo con usted antes que decirle la verdad. Se detecta, se registra y se gobierna.

Explicabilidad. La capacidad de justificar una decisión de IA en términos que una persona afectada pueda entender y cuestionar.

Equidad (fairness). La propiedad medible de que un sistema no produce resultados sistemáticamente peores para grupos protegidos.

Gobernanza federada. Un modelo donde unidades autónomas gobiernan su propia IA bajo normas y reportes comunes. El patrón de las grandes organizaciones.

Guardarraíl (guardrail). Un límite técnico u operativo que impide que un sistema de IA actúe fuera de su ámbito autorizado.

Humano en el circuito. Un modelo de control donde una persona aprueba cada decisión significativa antes de que surta efecto.

Humano sobre el circuito. Un modelo de control donde la persona supervisa y puede intervenir, sin aprobar cada decisión individual.

Incidente de IA. Cualquier evento en el que un sistema de IA causa, o estuvo a punto de causar, un daño real: a personas, a la organización o al público.

Cuasi incidente (near miss). Una advertencia que no llegó a crisis: el control funcionó o la suerte intervino. Se registra siempre; una organización sin cuasi incidentes no es segura, es ciega.

Inventario de sistemas de IA. La lista completa y viva de todos los sistemas de IA que la organización usa, compra o construye. La puerta de entrada de la gobernanza.

Comandante del incidente. La persona única que dirige la respuesta a un incidente. Dirige, no depura.

Madurez de gobernanza. El nivel de desarrollo de la práctica de gobernanza de una organización: de Ausente a Optimizada, en cinco niveles.

Mitigación. Las acciones que reducen el daño de un incidente y restauran la operación segura del sistema.

Monitoreo. La observación continua del comportamiento de un sistema en producción. La gobernanza que no se monitorea, muere.

Procedencia de los datos. El registro del origen, consentimiento y transformaciones de los datos usados por un sistema de IA.

Proveedor (vendor). Un tercero que suministra sistemas o componentes de IA. La mayoría de las organizaciones compran su IA: gobernar lo comprado es tan crítico como gobernar lo construido.

Prueba adversarial (red team). El intento deliberado y controlado de hacer fallar un sistema para descubrir sus debilidades antes que un atacante o el azar.

Puntuación de salud de gobernanza. La medida re-ejecutable (el AI Governance Health Score™) que indica si la gobernanza sigue viva después del proyecto: cobertura, control, documentación y monitoreo.

RACI. La matriz que asigna, para cada tarea, quién es Responsable, quién Aprueba, quién es Consultado y quién Informado. Una sola persona aprueba.

Registro de decisiones. La bitácora estructurada donde cada decisión significativa sobre IA queda documentada con su contexto y su autor.

Registro de riesgos. El documento vivo que lista los riesgos identificados, su probabilidad, su impacto y su responsable.

Reglamento de IA de la UE. El Reglamento (UE) 2024/1689, la primera ley integral de IA del mundo, con cuatro niveles de riesgo y multas de hasta 35 millones de euros o el 7 por ciento de la facturación mundial anual.

ISO/IEC 42001. La primera norma certificable de sistemas de gestión de IA del mundo, publicada en diciembre de 2023. La prueba de que la gobernanza existe.

NIST AI RMF. El marco de gestión de riesgos de IA del Instituto Nacional de Estándares y Tecnología de Estados Unidos: Gobernar, Mapear, Medir, Gestionar.

Supervisión humana. El requisito de que personas competentes puedan entender, intervenir y detener un sistema de IA de alto riesgo.

Transparencia. La obligación de informar a las personas cuando interactúan con una IA o son objeto de sus decisiones.

Trazabilidad. La capacidad de reconstruir el camino completo de una decisión de IA: datos, modelo, versión y contexto.

Autoridad de detención. El poder, preautorizado y con nombre, de detener un sistema de IA sin pedir permiso en el momento de la crisis.

Gravedad (G1, G2, G3). La escala de clasificación de incidentes: G1 grave, G2 significativo, G3 menor. En caso de duda, se clasifica hacia arriba.

Ventana de exposición. El período real durante el cual un sistema causó daño: desde que empezó, no desde que se detectó.

Parte interesada (stakeholder). Toda persona u organización afectada por un sistema de IA: usuarios, clientes, empleados, reguladores y el público.

IA digna de confianza. IA que es lícita, ética y robusta a la vez, y que puede demostrarlo con evidencias. El objetivo de toda la cadena de gobernanza.

El puente hacia la profundidad

Este glosario es el QUÉ y el POR QUÉ, y es gratuito. El CÓMO (los instrumentos operativos que ponen estos términos a trabajar) vive en la capa de pago: los kits por pilar, los packs insignia 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.