

Los principios de una IA responsable

Abreviaturas y acrónimos

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

IA : inteligencia artificial

Siete principios escritos para usarse bajo presión, no para enmarcarse en la pared. Cada uno cabe en una decisión real.

Los siete principios

1. El ser humano primero. Ningún objetivo del sistema vale más que la persona a la que afecta. Cuando el desempeño y la dignidad chocan, gana la dignidad. 2. La equidad como deber. Tratar distinto a grupos distintos sin justificación no es un error estadístico: es una falla de gobernanza con nombre y dueño. 3. La verdad antes que el agrado. Un sistema de inteligencia artificial (IA) que le da la razón para complacerlo es un riesgo, no una virtud. Exija sistemas capaces de sostener un desacuerdo fundado. 4. La transparencia. Las personas saben cuándo interactúan con una IA y cuándo un contenido fue generado. Sin letra pequeña. 5. La rendición de cuentas. Cada sistema tiene un responsable humano con nombre. La frase «lo decidió el algoritmo» no existe en una organización gobernada. 6. La contestabilidad. Toda persona afectada por una decisión automatizada puede entenderla, cuestionarla y obtener una revisión humana. 7. La proporción y la contención. Se despliega la capacidad necesaria, no toda la disponible; y todo sistema puede detenerse, por diseño, por una persona autorizada.

Cómo se usan

En cada decisión de despliegue, recorra los siete en cinco minutos. El principio que no pueda afirmar con evidencia es su próxima tarea de gobernanza. Los principios sin evidencia son publicidad.

El puente hacia la profundidad

Los principios son el POR QUÉ, y son gratuitos. El CÓMO operativo (el estatuto del comité de ética, las plantillas de decisión, el registro de riesgos éticos y el reporte al directorio) es el AI Ethics Operating Kit de la capa de pago.

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.