

El chequeo rápido de riesgos de frontera

Abreviaturas y acrónimos

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

IA : inteligencia artificial

Los sistemas autónomos actúan sin pedir permiso. Diez preguntas, cinco minutos, para saber si su organización gobierna sus agentes de IA o solo los observa.

Las diez preguntas

- ¿Algún sistema de su organización ejecuta acciones de varios pasos sin aprobación humana en cada paso?
- ¿Algún agente puede gastar dinero, enviar comunicaciones o modificar datos por sí mismo?
- ¿Sabe usted exactamente qué herramientas y accesos tiene cada agente?
- ¿Existe un límite técnico (no solo una instrucción) que impida al agente actuar fuera de su ámbito?
- ¿Puede usted detener cada agente en menos de cinco minutos, y lo ha probado?
- ¿Algún agente interactúa con otros agentes, propios o de terceros?
- ¿Registra cada acción del agente en una pista de auditoría que un humano pueda leer?
- ¿Alguien revisa esos registros con una cadencia definida?
- ¿El objetivo que optimiza el agente fue revisado por alguien distinto de quien lo escribió?
- ¿Existe un plan de respuesta específico para el día en que un agente actúe fuera de ámbito?

Cómo leer su resultado

Cuente sus respuestas negativas. **Cero a dos**: su frontera está razonablemente gobernada; mantenga la cadencia. **Tres a cinco**: zona de vigilancia; los vacíos que marcó son su plan de trabajo del próximo trimestre. **Seis o más**: zona roja; su organización opera agentes que no puede detener ni explicar, y eso exige acción antes de escalar.

Un caso anónimo

Una empresa logística grande respondió que no a la pregunta cinco: nadie había probado nunca detener el agente de enrutamiento. La prueba se hizo un viernes por la noche en un entorno controlado. Tardó cuarenta minutos, no cinco. Hoy tarda noventa segundos, porque la autoridad de detención y el procedimiento existen.

El puente hacia la profundidad

Este chequeo es el QUÉ y el POR QUÉ, y es gratuito. El CÓMO (la arquitectura de guardarraíles, los patrones multiagente y la evaluación de preparación para el despliegue) vive en la capa de pago: el Frontier and AGI Governance 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.