Plantilla

Marco de Responsabilidad para Agentes de IA

Cómo usar esta plantilla

Responde cada pregunta con honestidad. No hay respuestas correctas. El valor está en identificar las brechas actuales de tu organización. Puedes completarla de manera individual o usarla como insumo para una sesión con tu equipo directivo.

Diagnóstico rápido

Marca con una X el estado actual de tu organización en cada dimensión.

Dimensión	Sin implementar	En progreso	Implementado
Control de acceso (RBAC)	○	○	○
Observabilidad y logs	○	○	○
Human-in-the-loop definido	○	○	○
Soberanía de datos	○	○	○

Nivel 1

Control de acceso (RBAC)

Principio de mínimo privilegio: el agente solo accede a lo que necesita para su tarea específica.

1 **¿Qué agentes de IA están actualmente en producción o en piloto en tu organización?**

Nombre del agente, tarea que ejecuta y sistema al que tiene acceso.

2 **¿Tienes un inventario documentado de los permisos de cada agente?**

¿Sabes exactamente a qué bases de datos, APIs o sistemas puede acceder cada uno?

3 **¿Existe algún agente con acceso a sistemas que van más allá de su tarea?**
Ejemplo: un agente de cotizaciones con acceso a nómina o datos sensibles de clientes.

Nivel 2 **Observabilidad y logs**
Si no puedes ver lo que hace el agente en tiempo real, no puedes gobernarlo.

4 **¿Tienes logs de las acciones que ejecuta cada agente (con timestamp y contexto)?**
¿Quién tiene acceso a esos logs y con qué frecuencia los revisan?

5 **¿Tienes alertas automáticas cuando un agente opera fuera de rangos esperados?**
Ejemplo: volumen inusual de transacciones, errores repetidos, accesos a sistemas no habituales.

6 **Si un agente tomó una decisión incorrecta hoy, ¿en cuánto tiempo lo detectarías?**
Sé honesto: ¿minutos, horas, días? ¿Qué necesitarías cambiar para reducir ese tiempo?

Nivel 3

Human-in-the-loop (HITL)

No es que un humano revise todo — es definir en qué umbral de riesgo se activa la revisión humana.

7 **¿Tienes definido qué nivel de impacto activa una revisión humana antes de que el agente ejecute?**

Ejemplo: cualquier transacción mayor a $X, cualquier modificación a datos de más de N clientes.

8 **¿Quién es el responsable designado de aprobar o detener la acción de un agente cuando se activa ese umbral?**

Nombre del rol o persona. La responsabilidad debe recaer en alguien concreto, no en un sistema.

9 **¿Tus agentes de alto riesgo cumplen con EU AI Act Art. 14 o NIST AI RMF en materia de supervisión humana?**

Si operas en sectores regulados (finanzas, salud, energía), esta respuesta tiene implicaciones legales.

Nivel 4 — Soberanía de datos

¿Dónde se procesan los datos que usa el agente? ¿Quién más puede acceder a ellos?

10 ¿Sabes en qué región o infraestructura se procesan los datos que usan tus agentes?

¿Están en la nube de un proveedor externo? ¿En servidores propios? ¿En un país con regulación compatible?

11 ¿Los datos que procesa el agente están cubiertos por tu política de privacidad y las leyes locales vigentes?

En México: LFPDPPP. En Brasil: LGPD. En Colombia: Ley 1581.

12 ¿Existe riesgo de que el agente procese datos confidenciales en plataformas no autorizadas?

Relacionado con Shadow AI: ¿tus colaboradores usan herramientas externas con datos de la empresa?

Las 3 preguntas del liderazgo

Si no puedes responder estas tres con claridad, ahí está tu trabajo

1 **¿Tienes un registro de qué decisiones puede tomar cada agente de manera autónoma y cuáles requieren aprobación humana?**

Documéntalo como política, no como suposición.

2 **¿Sabes exactamente qué datos está procesando cada agente y bajo qué política de acceso?**

Si la respuesta es 'más o menos', no la sabes.

3 **Si un agente comete un error con consecuencias reales ahora mismo, ¿tienes claro a quién llamas y qué protocolo activas?**

Nombre, rol y pasos concretos — no 'escalo con el equipo de IT'.

Compromisos de acción

Con base en esta reflexión, las 3 acciones concretas que tomaré en los próximos 30 días:

1	**Acción:** _____
	Responsable: _____ **Fecha:** _____
2	**Acción:** _____
	Responsable: _____ **Fecha:** _____
3	**Acción:** _____
	Responsable: _____ **Fecha:** _____

Acerca de Klustomer
Av. Insurgentes sur 601, Col. Nápoles, 03810, CDMX

Tel. +52 (55) 1319-1808

contacto@klustomer.com

www.klustomer.com