

Gobernanza de IA para instituciones y empresas

Los principios que aplicamos en proyectos públicos y privados para que la inteligencia artificial genere valor sin exponer a su organización.

La gobernanza no es burocracia: es lo que permite que la IA genere valor sin exponer a la organización. Estos son los siete principios que aplicamos, con pasos concretos y señales de alerta para cada uno.

01 Empiece por el problema, no por la herramienta

La IA no es un objetivo; es un medio. Antes de elegir modelo o plataforma, defina el problema de negocio, quién se beneficia y cómo medirá el resultado. Un caso de uso sin métrica es un experimento caro. Priorice 2–3 casos con impacto claro y comience por el de mayor retorno y menor riesgo.

EN LA PRÁCTICA

- Escriba cada caso de uso en una frase: "para [quién], que [problema], con [métrica]".
- Estime el retorno esperado (horas, dinero o riesgo evitado) antes de invertir.
- **Señal de alerta:** "queremos usar IA" sin un problema concreto detrás.

02 Conozca sus datos y qué es sensible

La IA rinde con datos confiables y falla con datos dispersos. Clasifique su información en pública, interna y sensible; defina quién accede a qué; y decida qué datos **nunca** deben salir de su entorno. Para casos con requisitos de soberanía, considere despliegues on-premise o en su propia nube. La calidad importa tanto como la disponibilidad.

EN LA PRÁCTICA

- Haga un inventario simple: qué datos tiene, dónde viven y qué tan sensibles son.
- Defina una regla clara para los datos que no pueden salir de su perímetro.
- **Señal de alerta:** información crítica que solo vive en hojas de cálculo y correos.

03 Ponga límites claros de alcance

Un asistente debe saber qué puede y qué no puede hacer. Defina el alcance por escrito: temas permitidos, acciones permitidas y qué hacer cuando no sabe la respuesta —decir "no está en mi base de conocimiento", en lugar de inventar—. Un asistente que improvisa respuestas destruye la confianza; uno que reconoce sus límites la construye.

EN LA PRÁCTICA

- Documente el alcance como un contrato: dentro / fuera de alcance, y el mensaje para lo que no sabe.
- Fundamente las respuestas en su base de conocimiento (RAG), con las fuentes citadas.
- **Señal de alerta:** el asistente responde con seguridad sobre temas para los que no tiene datos.

04 Mantenga supervisión humana donde importa

La IA propone; la persona decide. Para acciones con consecuencias (enviar, cobrar, publicar, borrar), use un flujo **propone** → **aprueba** → **ejecuta**. Automatice lo repetitivo y escale a un humano lo crítico. La supervisión no es un freno: es lo que hace la automatización segura y lo que permite escalar sin perder el control.

EN LA PRÁCTICA

- Clasifique las acciones por consecuencia: reversibles (auto) vs. irreversibles (requieren aprobación).
- Deje siempre visible qué hizo la IA y qué queda pendiente de aprobar.
- **Señal de alerta:** la IA ejecuta acciones con impacto sin que nadie las revise.

05 Exija trazabilidad y auditoría

Todo lo que la IA hace debe quedar registrado: qué se pidió, qué respondió, con qué fuentes y quién aprobó. Sin bitácora no hay confianza, ni aprendizaje, ni cumplimiento. La trazabilidad también le permite mejorar el sistema con evidencia real y responder ante una auditoría o un incidente.

EN LA PRÁCTICA

- Registre cada interacción y acción con fecha, usuario, fuentes y resultado.
- Revise periódicamente los registros para detectar errores y oportunidades de mejora.
- **Señal de alerta:** no puede reconstruir por qué la IA hizo lo que hizo.

06 Proteja la privacidad por diseño

La privacidad no es un añadido posterior: se diseña desde el inicio. Cifrado en reposo y tránsito, aislamiento de datos por área o cliente, control de accesos por rol y un proceso claro para las solicitudes de titulares de datos (ARCO / DSR). En el sector público, documente todo y prevea la trazabilidad que exigirá cualquier revisión.

EN LA PRÁCTICA

- Aplique el mínimo privilegio: cada persona y cada agente ve solo lo que necesita.
- Tenga listo el procedimiento para solicitudes de acceso, rectificación y eliminación.
- **Señal de alerta:** "la privacidad la resolvemos después de lanzar".

07 Evalúe a sus proveedores con criterio

No todos los proveedores son iguales. Antes de firmar, pregunte: ¿dónde viven mis datos? ¿se usan para entrenar modelos? ¿puedo usar mi propio modelo y mis propias llaves (BYOK)? ¿hay trazabilidad y control de accesos? ¿qué pasa con mis datos si decido irme (portabilidad)? Las respuestas revelan el nivel de gobernanza real —no el del folleto—.

EN LA PRÁCTICA

- Pida por escrito: residencia de datos, uso para entrenamiento, portabilidad y certificaciones.
- Prefiera arquitecturas que le permitan cambiar de modelo sin rehacer todo (BYOK).
- **Señal de alerta:** respuestas vagas sobre dónde y cómo se procesan sus datos.

Checklist final

- ✓ Cada caso de uso tiene un objetivo de negocio y una métrica.
- ✓ Los datos sensibles están identificados y protegidos.
- ✓ El alcance de cada asistente está definido por escrito.
- ✓ Las acciones con consecuencias pasan por aprobación humana.
- ✓ Existe bitácora de cada interacción y acción.
- ✓ Hay control de accesos por rol.
- ✓ La privacidad está en el diseño, no como parche.
- ✓ Se evaluó al proveedor (datos, entrenamiento, portabilidad).
- ✓ Hay un responsable de la gobernanza de IA.
- ✓ Se revisa y ajusta periódicamente.

Ruta de adopción en 90 días

Un camino realista para pasar de la intención a un primer resultado con control. No intente todo a la vez: avance por fases y decida con evidencia.

DÍAS 1-30 · CIMIENTOS

Ordenar la base antes de construir

- Inventaríe y clasifique sus datos (pública / interna / sensible) y defina qué no puede salir de su perímetro.
- Elija 2-3 casos de uso con problema claro y métrica; descarte los que no tengan forma de medirse.
- Nombre a un responsable de la gobernanza de IA y acuerde reglas mínimas de uso.

DÍAS 31-60 · PILOTO CON CONTROL

Un solo caso, bien hecho

- Implemente un caso de uso con supervisión humana en las acciones de consecuencia (propone → aprueba → ejecuta).
- Active la bitácora desde el primer día: qué se pidió, qué respondió, con qué fuentes y quién aprobó.
- Mida contra la métrica definida y recoja el criterio de las personas que lo usan.

DÍAS 61-90 · ESCALAR CON EVIDENCIA

Decidir go / no-go y repetir

- Evalúe resultados vs. la métrica y el costo real; documente lo aprendido.
- Ajuste alcance, controles y accesos; corrija lo que falló en el piloto.
- Decida si escalar ese caso y planifique el siguiente con el mismo método.

Glosario esencial de gobernanza de IA

Los términos que aparecen una y otra vez en un proyecto de IA con gobernanza, en lenguaje claro.

Gobernanza de IA

Conjunto de reglas, roles y controles que definen cómo se usa la IA de forma segura, trazable y alineada con el negocio.

RAG (generación aumentada por recuperación)

Técnica que fundamenta las respuestas de la IA en su propia base de conocimiento, citando fuentes en lugar de inventar.

Human-in-the-loop (supervisión humana)

Flujo donde la IA propone y una persona aprueba antes de ejecutar acciones con consecuencias.

Trazabilidad

Registro de qué se pidió, qué respondió la IA, con qué fuentes y quién aprobó; base de la confianza y del cumplimiento.

BYOK (traiga su propia llave)

Arquitectura que le permite usar su propio modelo y sus propias credenciales, sin quedar atado a un proveedor.

On-premise / nube propia

Despliegue dentro de su infraestructura para datos con requisitos de soberanía o alta sensibilidad.

ARCO / DSR

Solicitudes de titulares de datos: acceso, rectificación, cancelación y oposición (o data subject requests).

Alucinación

Respuesta que suena convincente pero es incorrecta o inventada; se mitiga con RAG, límites de alcance y supervisión.

Mínimo privilegio

Cada persona y cada agente accede solo a lo estrictamente necesario para su función.

Residencia de datos

Dónde se almacenan y procesan físicamente sus datos; clave para soberanía y cumplimiento.

¿Quiere aplicar esto en su organización?

Acompañamos a instituciones y empresas a implementar IA con gobernanza, supervisión humana y resultados medibles.

Escríbanos a hola@globalagenttic.com · globalagenttic.com