DOCUMENTO DE ANÁLISIS

Datos y privacidad en la inteligencia artificial

Lo que las condiciones de uso no dicen en voz alta

Cómo los principales motores de IA procesan los datos de sus usuarios, qué varía según el plan contratado y qué buenas prácticas pueden adoptar las organizaciones

Nick Perot

Abogado | Consultor en transformación digital | Docente

www.nickperot.com Marzo 2026

1. Lo que la mayoría de los usuarios no lee

Cada vez que una persona interactúa con un motor de inteligencia artificial generativa, envía datos. A veces es un texto. A veces es un archivo. A veces es información personal, financiera o confidencial. Lo que ocurre con esos datos después de enviarse depende de algo que la mayoría de los usuarios no lee: las condiciones de uso y las políticas de privacidad del proveedor.

Un estudio publicado en 2025 por el Stanford Institute for Human-Centered Artificial Intelligence (HAI) analizó las políticas de privacidad de seis desarrolladores líderes de modelos de lenguaje en Estados Unidos. La conclusión principal fue directa: las seis compañías utilizan los datos de las conversaciones de sus usuarios para entrenar sus modelos por defecto, y algunas conservan esa información en sus sistemas de manera indefinida.

> *King, J. et al., "User Privacy and Large Language Models: An Analysis of Frontier Developers' Privacy Policies," Stanford Institute for Human-Centered AI, 2025.*

Este documento analiza qué hacen los principales motores de IA con los datos de sus usuarios, cómo varían las condiciones según el plan contratado, qué ocurre cuando se envía retroalimentación, y qué buenas prácticas pueden adoptar personas y organizaciones para proteger su información.

2. Entrenamiento con datos del usuario: la regla, no la excepción

El hallazgo central del estudio de Stanford es que el uso de datos de conversación para entrenamiento de modelos no es una práctica excepcional. Es el estándar de la industria. Los seis proveedores analizados alimentan sus modelos con los datos de las interacciones de sus usuarios por defecto, a menos que el usuario tome medidas activas para impedirlo.

Esto significa que cuando una persona pega un contrato, comparte datos de salud, sube un archivo con información financiera o simplemente hace una pregunta que revela preferencias personales, esa información puede ser incorporada al proceso de entrenamiento del modelo. El estudio de Stanford describe un escenario ilustrativo: un usuario que pide recetas bajas en azúcar o aptas para problemas cardíacos está revelando información de salud que el sistema puede clasificar y que, a través del ecosistema del proveedor, podría terminar influyendo en la publicidad o en los perfiles comerciales que recibe.

> *King et al. (2025), Stanford HAI.*

> Si la información ingresada a un motor de IA no desaparece cuando se cierra la aplicación, la pregunta operativa no es si se usan los datos. Es en qué condiciones y por cuánto tiempo.

3. Lo que cambia según el plan contratado

Una de las variables más relevantes y menos visibles es que las condiciones de privacidad no son uniformes dentro de un mismo proveedor. Varían significativamente según el tipo de plan contratado. El patrón es consistente en la industria.

Planes gratuitos y de consumo individual

En la mayoría de los proveedores, los planes gratuitos y los planes individuales pagos operan bajo las mismas condiciones de privacidad. Las conversaciones se utilizan para entrenamiento por defecto. La retención de datos puede extenderse durante meses o incluso de manera indefinida. Algunos proveedores permiten que revisores humanos lean transcripciones de conversaciones como parte de sus procesos de seguridad y mejora. En ciertos casos, incluso el pago de una suscripción mensual no cambia la política de entrenamiento: el usuario sigue necesitando desactivar la opción manualmente.

Planes empresariales y acceso por API

Los planes empresariales y el acceso a través de API suelen tener condiciones sustancialmente diferentes. En general, los proveedores declaran que no utilizan datos de clientes empresariales para entrenar modelos por defecto. La retención de datos tiende a ser más corta. Los controles de acceso son más granulares. La documentación de procesamiento de datos suele estar disponible.

La diferencia crítica

La distinción entre plan de consumo y plan empresarial no es un detalle comercial. Es una decisión de privacidad. Una organización que permite que sus equipos usen herramientas de IA en planes gratuitos o individuales para procesar información institucional está enviando datos a un entorno donde esos datos pueden ser incorporados al entrenamiento de modelos, revisados por personas y retenidos por períodos prolongados.

> El precio del plan no es solo una decisión de presupuesto. Es una decisión de privacidad y de riesgo institucional.

4. Retroalimentación, usos en segundo plano y la letra chica

El envío de retroalimentación como envío de datos

Cuando un usuario marca una respuesta como útil o inútil, o corrige un resultado, está enviando retroalimentación al proveedor. Esa retroalimentación, en la mayoría de los casos, se utiliza para mejorar el modelo. Lo que muchos usuarios no advierten es que la retroalimentación no viaja sola: viaja con el contexto de la conversación. Marcar un pulgar hacia abajo en una respuesta sobre un contrato confidencial es, en la práctica, enviar el contenido de ese contrato al sistema de mejora del proveedor.

Usos en segundo plano y ecosistemas integrados

En los proveedores que operan ecosistemas de múltiples productos (buscadores, correo electrónico, redes sociales, asistentes de voz, plataformas de productividad), las interacciones con el motor de IA pueden combinarse con información proveniente de otros servicios del mismo proveedor. El estudio de Stanford identifica esta práctica en proveedores multiproducto: las consultas al motor de IA se fusionan con historiales de búsqueda, compras, interacciones en redes sociales y otros datos de la plataforma.

> *King et al. (2025), Stanford HAI: "User interactions routinely get merged with information gleaned from other products consumers use on those platforms."*

Esto significa que el motor de IA no opera de manera aislada. Opera dentro de un ecosistema de datos donde la información fluye entre servicios. Para el usuario, puede parecer una conversación privada. Para el proveedor, es una pieza más en un perfil integrado.

Lo que dicen las condiciones sobre eliminación de datos

La mayoría de los proveedores ofrecen la opción de eliminar conversaciones. Pero eliminar no siempre significa borrar. Algunos proveedores retienen datos eliminados por períodos de 30 días. Otros pueden estar obligados por orden judicial a retener conversaciones incluso después de que el usuario las haya borrado. Y en todos los casos, si los datos ya fueron incorporados a un ciclo de entrenamiento, la eliminación de la conversación original no revierte la influencia que esos datos ejercieron sobre el modelo.

Datos de menores

El estudio de Stanford encontró que las prácticas respecto a datos de menores varían entre proveedores, pero la mayoría no toma medidas activas para excluir los datos de menores de sus procesos de entrenamiento. Algunos proveedores declaran que no permiten cuentas a menores de

18 años pero no requieren verificación de edad. Otros permiten explícitamente la recolección de datos de menores bajo ciertas condiciones.

> *King et al. (2025): "Developers' practices vary, but most are not taking steps to remove children's input from their data collection and model training processes."*

5. El marco regulatorio: dónde hay protección y dónde no

La protección efectiva de los datos que se envían a motores de IA depende, en gran medida, de la jurisdicción bajo la cual opera el proveedor y el usuario.

En la Unión Europea, el Reglamento General de Protección de Datos (GDPR) establece un piso de protección que incluye transparencia, minimización de datos, limitación de finalidad y el derecho del usuario a solicitar la eliminación de sus datos. En 2024, la autoridad italiana de protección de datos sancionó a uno de los principales proveedores de IA con una multa de 15 millones de euros por falta de transparencia en la recolección y el uso de datos a través de su motor de IA generativa.

> *Garante per la protezione dei dati personali (Italia), resolución contra proveedor de IA por infracción de GDPR, 2024; Reglamento (UE) 2016/679 (GDPR).*

En Estados Unidos, la situación es diferente. No existe una ley federal integral de privacidad de datos. La protección es fragmentaria, depende de regulaciones estatales y, en la práctica, deja a los usuarios con menos garantías legales frente a los proveedores de IA. El estudio de Stanford lo resume con claridad: hay cientos de millones de personas interactuando con motores de IA que recolectan datos personales para entrenamiento, y prácticamente no se ha realizado investigación para examinar las prácticas de privacidad de estas herramientas.

En América Latina, como se analizó en documentos anteriores de esta serie, la regulación específica de IA es incipiente. Los usuarios y organizaciones de la región operan, en la mayoría de los casos, bajo las condiciones que los proveedores definen unilateralmente en sus términos de servicio, sin un marco regulatorio local que establezca estándares mínimos de protección para el uso de datos en sistemas de IA.

6. Qué pueden hacer las organizaciones

Elegir el plan y el canal con criterio de privacidad

La primera decisión de privacidad no es técnica. Es contractual. Antes de usar un motor de IA para procesar información institucional, la organización necesita saber si el plan contratado utiliza los datos para entrenamiento, cuánto tiempo los retiene, quién puede revisarlos y bajo qué jurisdicción

se procesan. Si la respuesta a cualquiera de estas preguntas no está clara, el motor no está listo para recibir información sensible.

Desactivar el entrenamiento con datos del usuario

En la mayoría de los proveedores, la opción de exclusión del entrenamiento existe pero no está activada por defecto. La organización debe verificar y documentar que la opción está desactivada en cada cuenta que utilice herramientas de IA. Esto incluye tanto las cuentas institucionales como las cuentas personales que los empleados puedan usar para tareas laborales.

No ingresar información sensible en planes de consumo

Datos personales de terceros, información financiera, documentos legales, datos de salud, credenciales de acceso, código propietario y estrategias confidenciales no deberían procesarse a través de planes gratuitos o individuales de motores de IA. Si la organización necesita procesar este tipo de información con IA, debería utilizar planes empresariales, acceso por API con retención mínima o, en casos críticos, modelos de ejecución local que no transmitan datos a servidores externos.

Anonimizar antes de enviar

Cuando no es posible evitar el uso de motores de IA en la nube, la información debería anonimizarse antes de enviarse. Reemplazar nombres reales por genéricos, empresas por "Compañía X", fechas específicas por referencias temporales relativas. Esto no elimina el riesgo, pero lo reduce significativamente.

Revisar periódicamente las condiciones de uso

Las políticas de privacidad de los proveedores de IA cambian. En 2025, al menos un proveedor que se posicionaba como la opción de mayor privacidad modificó sus términos para pasar de un modelo de consentimiento explícito (opt-in) a un modelo de exclusión voluntaria (opt-out), donde el silencio del usuario se interpreta como aceptación. Las organizaciones deberían revisar las condiciones de uso de sus herramientas de IA al menos dos veces al año.

> *Sobre el cambio de modelo opt-in a opt-out en 2025: Stanford HAI, King et al. (2025); Psychiatric Times, "Chatbot Privacy Is an Oxymoron," marzo de 2026.*

Documentar y capacitar

Todo equipo que use herramientas de IA debería conocer las condiciones de privacidad del motor que utiliza, saber qué datos puede y no puede ingresar, y entender qué ocurre con la información después de enviarla. Esto requiere capacitación específica, no solo un documento de política interna.

7. Síntesis: lo que hay que verificar antes de usar un motor de IA

Checklist de privacidad para organizaciones

Variable	Pregunta que la organización debe responder
Entrenamiento	¿El proveedor usa los datos de las conversaciones para entrenar modelos? ¿Es opt-in u opt-out? ¿Está desactivado en todas las cuentas?
Retención	¿Cuánto tiempo retiene el proveedor los datos después de la interacción? ¿Qué ocurre cuando el usuario elimina una conversación?
Revisión humana	¿Pueden empleados o contratistas del proveedor leer transcripciones de conversaciones?
Ecosistema	¿Se combinan los datos de la IA con información de otros productos del mismo proveedor?
Jurisdicción	¿Bajo qué legislación se procesan los datos? ¿Aplica GDPR u otra normativa de protección?
Plan contratado	¿Las condiciones de privacidad del plan utilizado son adecuadas para el tipo de información que se procesa?
Menores	¿Se procesan datos de menores? ¿Qué medidas tiene el proveedor al respecto?
Retroalimentación	¿La retroalimentación enviada al proveedor incluye el contenido de la conversación? ¿Puede desactivarse?

Elaboración propia a partir de King et al. (2025), Stanford HAI; NIST AI RMF 1.0; Reglamento (UE) 2016/679 (GDPR).

8. Consideraciones finales

Los motores de inteligencia artificial generativa operan, en su configuración por defecto, bajo condiciones de privacidad que favorecen al proveedor. Los datos de las conversaciones se utilizan para entrenamiento. La retención puede ser prolongada. La información puede cruzarse con otros servicios del mismo ecosistema. Y las condiciones cambian, a veces sin aviso prominente.

La protección efectiva de los datos no depende del motor en sí. Depende del plan contratado, de la configuración elegida, de la jurisdicción aplicable y, sobre todo, de las decisiones que toma la organización antes de enviar información a un sistema que no controla.

El estudio de Stanford concluye con una recomendación que funciona como regla de mínima: los usuarios deberían pensar dos veces antes de compartir información en conversaciones con motores de IA y, siempre que sea posible, optar activamente por la exclusión del entrenamiento. Para una organización, esa recomendación individual se convierte en una política institucional.

> La privacidad en la IA no se protege con confianza. Se protege con configuración, con contratos leídos y con la decisión de no enviar a un sistema externo lo que no se está dispuesto a compartir.

Referencias

Garante per la protezione dei dati personali (2024). *Resolución sancionatoria contra proveedor de IA por infracción de GDPR*. Italia, 2024. Multa de 15 millones de euros.

King, J. et al. (2025). *User Privacy and Large Language Models: An Analysis of Frontier Developers' Privacy Policies*. Stanford Institute for Human-Centered Artificial Intelligence (HAI).

MIT Technology Review (2025). *The State of AI: Chatbot companions and the future of our privacy*. Noviembre de 2025.

NIST (2023). *Artificial Intelligence Risk Management Framework (AI RMF 1.0)*. NIST AI 100-1. Gaithersburg: National Institute of Standards and Technology.

Psychiatric Times (2026). *Chatbot Privacy Is an Oxymoron: Assume Your Data Is Always At Risk*. Marzo de 2026.

Unión Europea (2016). *Reglamento (UE) 2016/679 — Reglamento General de Protección de Datos (GDPR)*. Diario Oficial de la UE.

Unión Europea (2024). *Reglamento (UE) 2024/1689 — Ley de Inteligencia Artificial (AI Act)*. Diario Oficial de la UE, 13 de junio de 2024.