Lawgic

Mitos y realidades: IA en el área legal corporativa

Guía para abogados de empresa que usan cuentas empresariales de IA.

Por **Aldo Ricardo Rodríguez** · aldoricardo.com

Vigente al 30 de octubre de 2026 · Basado en documentos oficiales de los proveedores

Lo esencial. El riesgo no está en usar IA, sino en el tipo de cuenta y en sus condiciones. Ningún plan empresarial entrena con los datos del cliente y todos firman como encargados. Lo que hay que cuidar es cuánto tiempo se guardan los datos, quién del proveedor puede leerlos, en qué país se guardan, cómo se tratan los datos sensibles, el uso de cuentas personales por parte del equipo, la discriminación a través de datos que delatan y la revisión humana que solo firma sin revisar de verdad.

CONTENIDO

Semáforo por tipo de cuenta 2

Mitos y realidades 3

1. Confidencialidad y entrenamiento en cuentas empresariales 3
2. Protección de datos personales (LFPDPPP) 5
3. Secreto profesional y privilegio 7
4. Anonimización y datos que delatan 8
5. Sesgos en procesos masivos de la empresa 9
6. Transparencia, explicabilidad y auditoría 10
7. Gobernanza interna, supervisión humana y adopción 11

Matriz comparativa de plataformas 13

Checklist del contrato de tratamiento de datos (DPA) 14

Anexo: detalle por proveedor 15

Referencias 19

Fuentes de las matrices y la checklist 22

Semáforo por tipo de cuenta

● favorable · ● revisar o configurar · ● desfavorable

Tipo de cuenta	Riesgo	Por qué
Cuenta personal, gratuita o de pago (Free, Plus, Pro, Max)	Alto ●	El proveedor es responsable y entrena con los datos, o hay que rechazarlo expresamente. No hay contrato de tratamiento de datos ni controles, y Microsoft y Google no defienden al usuario por propiedad intelectual.
Cuenta personal con el entrenamiento apagado	Alto ●	No es suficiente: sigue fuera del contrato comercial, y en OpenAI las valoraciones con pulgar pueden volver a permitir el entrenamiento.
Plan de equipo (ChatGPT Business, Claude Team)	Medio ●	La orden del caso NYT alcanzó a Business (antes Team), puede haber revisión por contratistas, y Claude Team no tiene registros de actividad ni plazos de conservación propios.
Plan Enterprise con contrato de tratamiento de datos (ChatGPT Enterprise, Claude Enterprise, Microsoft 365 Copilot, Gemini en Workspace)	Medio a bajo ●	«Potencialmente defendible»: no entrena, actúa como encargado, tiene registros de actividad, el plazo de conservación se configura y cuenta con la certificación ISO 42001.
API sin guardar datos (Zero Data Retention) o con monitoreo de abuso modificado aprobado	Bajo ●	Disponible en OpenAI y Anthropic (Zero Data Retention), en Azure (monitoreo de abuso modificado) y en Vertex con pasos específicos.
IA instalada en servidores propios	Bajo ●	«Altamente protegido»: no depende de los términos de un tercero.

Mitos y realidades

1. Confidencialidad y entrenamiento en cuentas empresariales

Mito	Realidad	En la empresa
Todo lo que escribo en ChatGPT o Claude se usa para entrenar el modelo.	**Depende del tipo de cuenta.** En ChatGPT Free, Plus y Pro el uso de datos para entrenar viene activado por defecto[1]. En Claude Free, Pro y Max el entrenamiento está activo salvo que el usuario lo rechace: Anthropic lo anunció el 28 de agosto de 2025 y dio hasta el 8 de octubre para elegir[2]. En los planes empresariales y en la API, el contrato lo excluye: OpenAI se obliga a no usar el contenido del cliente para mejorar sus servicios (cláusula 4.2)[3] y Anthropic «may not train models on Customer Content» (sección B)[4]. Ojo con las valoraciones: en ChatGPT de consumo, un pulgar arriba o abajo permite usar toda la conversación para entrenar, aunque el entrenamiento esté apagado[1]. Y en los productos comerciales de Anthropic, una valoración hace que se guarde la conversación completa hasta 5 años y que pueda usarse para entrenar[5].	El contenido no se usa para entrenar, salvo que el cliente lo autorice, en ChatGPT Business (antes Team) y Enterprise, la API de OpenAI (Services Agreement, cláusula 4.2), Claude Team, Claude Enterprise y la API de Anthropic (Commercial Terms, sección B), Microsoft 365 Copilot y Foundry, y Gemini en Workspace y Vertex. Conviene revisar el contrato concreto que firmó la empresa.
Si apago el entrenamiento en mi cuenta personal, es lo mismo que tener una cuenta empresarial.	**No.** Desactivar el entrenamiento en una cuenta de consumo no convierte al proveedor en encargado, es decir, en alguien que solo trata los datos por cuenta del usuario. El Services Agreement de OpenAI excluye expresamente a consumidores e individuos[3], y Anthropic se declara encargado en sus productos comerciales[6]. Fuera de ese contrato, el proveedor conserva datos para sus propios fines y no hay contrato de tratamiento de datos (DPA). En ChatGPT Free, Plus y Pro la opción de mejorar el modelo viene activada por defecto[1]. En Claude Free, Pro y Max el entrenamiento está activo salvo que el usuario lo rechace, y no hay inicio de sesión único, registros de actividad ni plazos de conservación a la medida[7].	Esa es justamente la diferencia: la cuenta corporativa con contrato es la que coloca al proveedor como encargado. El riesgo real es que el equipo use en paralelo sus cuentas personales.
Si el proveedor no entrena con mis datos, tampoco los guarda.	**Que el proveedor no use las conversaciones para entrenar su modelo no significa que las borre.** Siguen en el historial y, si el usuario las borra, ChatGPT y Claude las conservan hasta 30 días más[8,9]. Además, un juez puede ordenar que se guarden: el 13 de mayo de 2025, en el juicio del New York Times contra OpenAI, la magistrada Ona T. Wang obligó a conservar incluso los chats borrados de ChatGPT Free, Plus, Pro y Team (hoy Business) y de la API sin la opción de no guardar datos; Enterprise y Edu quedaron fuera[10]. La obligación terminó el 26 de septiembre de 2025, pero OpenAI sigue guardando los datos de abril a septiembre de 2025 y los de cuentas ligadas a dominios señalados por los demandantes[11].	Los 30 días son el plazo para borrar, no el tiempo total que se guardan los datos. En ChatGPT Enterprise y Business el plazo lo fija el administrador, y al terminar el contrato los datos se eliminan en 30 días. OpenAI Business puede guardar más si es «razonablemente necesario» para proteger el servicio o a terceros, y el Services Agreement permite conservar con ese criterio el contenido abusivo de cualquier cliente. Anthropic guarda hasta 2 años lo marcado por violar su política de uso, y 7 años las puntuaciones de esa revisión. En Gemini para Workspace el plazo va de 90 días a indefinido, según el administrador; en la app Gemini dentro de Workspace llega hasta 36 meses, y en la de consumo puede ser indefinido. Conviene configurar el plazo más corto posible.

Mito	Realidad	En la empresa
El proveedor lee mis conversaciones.	**Depende del plan.** En los planes empresariales el proveedor usa el contenido solo para prestar el servicio, cumplir la ley y hacer cumplir sus políticas. Todos usan sistemas automáticos para detectar abusos; lo que cambia es si una persona puede leer. En Azure, por ejemplo, los revisores humanos solo entran con aprobación «Just-In-Time» de su gerente y desde estaciones de trabajo seguras[12], y Microsoft 365 Copilot quedó excluido de esa revisión humana[13].	OpenAI Enterprise: el personal solo entra en caso de incidentes, con permiso expreso o por obligación legal. OpenAI Business y la API: el personal autorizado y «contratistas externos especializados» pueden revisar conversaciones para detectar abusos. Azure OpenAI: empleados autorizados pueden revisar muestras marcadas, salvo que Microsoft apruebe el monitoreo de abuso modificado. Microsoft 365 Copilot y Google Workspace: sin revisión por personas. Anthropic: por defecto su personal no puede leer las conversaciones; las excepciones son las valoraciones y la revisión por violar su política de uso, a cargo de su equipo de seguridad. No tiene una página específica para Team y Enterprise.
Con una cuenta empresarial el riesgo es cero.	**No es cero.** La cuenta empresarial y la API son de riesgo medio a bajo, «potencialmente defendible». Solo la IA instalada en servidores propios está «altamente protegida». La empresa sigue siendo responsable y debe revisar el contrato: el Reglamento de la LFPDPPP le prohíbe adherirse a servicios de nube que no den garantías mínimas, como no asumir la titularidad de la información, transparentar las subcontrataciones y garantizar el borrado al terminar (artículo 52)[14]. Y aun el contrato empresarial de OpenAI le permite conservar contenido abusivo en la medida razonablemente necesaria para proteger el servicio[3].	Hay pendientes concretos. Que el contrato incluya lo que pide el artículo 52 del Reglamento, incluido el aviso cuando una autoridad pida información: OpenAI lo prevé «en la medida legalmente permitida» (DPA, cláusula 2.2) y Anthropic se obliga a avisar y a cooperar para evitar o acotar la entrega (Commercial Terms, cláusula E.3). Que se autorice a las empresas que subcontrata el proveedor (artículo 54): se aceptan con una autorización general y un plazo para oponerse de 30 días en OpenAI (DPA, cláusula 2.9) y 15 en Anthropic (DPA, cláusula C.3), y en Anthropic solo por motivos razonables de privacidad o seguridad. No usar enlaces públicos para compartir chats. Revisar los términos de forma periódica. Y elegir bien el plan: Claude Team no tiene registros de actividad ni plazos de conservación a la medida; Claude Enterprise sí.
Lo que importa es la marca: ChatGPT, Claude, Gemini o Copilot.	**«La línea es la modalidad, no la marca».** Un mismo proveedor ofrece una versión de consumo, donde es responsable, entrena con los datos y no firma contrato de tratamiento de datos, y una versión empresarial o de API, donde actúa como encargado, no entrena y sí firma ese contrato. Cada proveedor separa sus contratos: el Services Agreement de OpenAI es solo para empresas y desarrolladores[3], y Anthropic excluyó expresamente de sus términos de consumo a Team, Enterprise, Gov, Education y la API[2].	Hay que confirmar qué producto exacto contrató la empresa. Por ejemplo, Microsoft 365 Copilot con contrato de tratamiento de datos, y no un Copilot de consumo: el compromiso de Microsoft de defender al cliente por derechos de autor cubre solo las versiones de pago de sus servicios comerciales y exige usar sus filtros[15]. Google también distingue dos productos: Gemini Notebook, dentro de Workspace, y NotebookLM Enterprise, de Google Cloud.

Mito	Realidad	En la empresa
La IA guarda tu información y se la entrega a quien la pida.	**Es una «cascada de disponibilidad»:** una idea que se repite sin que nadie lea los contratos. En los casos que se citan, la información no salió de los servidores del proveedor. En Heppner, el FBI encontró unos 31 documentos de Claude al catear el domicilio del acusado, el 4 de noviembre de 2025[16]. En Corea del Sur, la policía de Seúl reclasificó un caso como homicidio después de revisar los chats de ChatGPT en el teléfono de la sospechosa[17]. Ninguno involucró a un abogado con un plan profesional bien configurado. Lo que sí es real: la política de privacidad de consumo de Anthropic le permite entregar datos si cree de buena fe que lo exige un proceso legal[18], y en el caso del New York Times OpenAI tuvo que entregar a los demandantes 20 millones de registros desidentificados de ChatGPT de consumo, bajo orden de protección[19].	En las cuentas empresariales, los contratos prevén avisar al cliente cuando una autoridad pide información.

2. Protección de datos personales (LFPDPPP)

No hay criterio expreso de la autoridad: lo que sigue es una interpretación de la ley vigente, la LFPDPPP publicada en el DOF el 20 de marzo de 2025[20]. El Reglamento que se cita es el de 2011[14]: no hay uno nuevo, y ningún transitorio dice expresamente que el anterior siga vigente, así que aplicarlo en lo que no contradiga la ley es una interpretación razonable, no una regla expresa.

Mito	Realidad	En la empresa
Mandar datos personales a una IA es una transferencia y requiere consentimiento.	**No es una transferencia, sino una remisión a un encargado:** alguien que trata los datos por cuenta de quien los tiene. La ley de 2025 define la transferencia como la comunicación a alguien distinto del titular, del responsable o de la persona encargada (artículo 2, fracción XX), y a la encargada como quien trata datos por cuenta del responsable (fracción XII)[20]. El término «remisión» y la exención expresa de consentimiento y aviso vienen del Reglamento de 2011 (artículos 2, fracción IX, y 53)[14]. Lo que importa es para quién se tratan los datos, no adónde viajan. Es lo mismo que ya pasa con la nube, el correo o un servicio de transcripción. El límite: si el encargado usa los datos para fines propios, como entrenar su modelo, se convierte en responsable y deja de haber remisión[14].	La empresa es el responsable y el proveedor empresarial con contrato de tratamiento de datos es el encargado. OpenAI, Anthropic, Microsoft y Google se declaran encargados en sus contratos empresariales. La condición está en el contrato: que el proveedor no use los datos para sí ni para entrenar, y que garantice confidencialidad, seguridad y borrado.
Si los servidores están fuera de México, es una transferencia internacional.	El artículo 53 del Reglamento exime de aviso y de consentimiento a las remisiones, sean nacionales o internacionales[14]. La propia ley apoya esta lectura: la transferencia es la comunicación «dentro o fuera del territorio mexicano» a alguien distinto de la persona encargada, así que la encargada queda fuera sin importar dónde esté (artículo 2, fracción XX)[20]. Que el servidor esté en otro país no cambia cómo se califica el envío.	Aplica igual, siempre que el proveedor actúe como encargado. En la práctica, ni ChatGPT ni la API de OpenAI ofrecen guardar los datos en México: hay 10 regiones, y el procesamiento dentro de la región solo existe en Estados Unidos, Europa y Emiratos Árabes Unidos. Microsoft avisa que, fuera de la Unión Europea, el procesamiento puede hacerse en Estados Unidos, en la Unión Europea o en otras regiones. Anthropic ofrece procesar solo en Estados Unidos, con un precio 1.1 veces mayor.

Mito	Realidad	En la empresa
Usar IA es una finalidad nueva y obliga a pedir consentimiento otra vez.	La IA es un medio para cumplir la misma finalidad (artículo 11), no una finalidad nueva; es una interpretación, porque ningún artículo lo dice. Equivocarse sale caro: cambiar sustancialmente la finalidad sin nuevo consentimiento es infracción (artículo 58, fracción IX) y se sanciona con multa de 200 a 320,000 UMA (artículo 59)[20]; con la UMA de 2026, hasta unos 37.5 millones de pesos[21]. Aun así, conviene actualizar el aviso de privacidad con una cláusula sobre el uso de herramientas de IA.	Implica revisar los avisos de privacidad de la empresa: los de clientes, empleados y proveedores.
Los datos sensibles se pueden procesar con IA igual que los demás.	**No.** Exigen consentimiento expreso y por escrito, con firma autógrafa, electrónica u otro mecanismo de autenticación (artículo 8), y el aviso de privacidad debe identificarlos (artículo 15, fracción II). Hay excepciones en el artículo 9, y los terceros, como la contraparte o los testigos, no pueden dar su consentimiento. La ley pide limitar al mínimo el plazo de tratamiento de los datos sensibles (artículo 12), y las multas por infracciones que los involucran pueden duplicarse (artículo 59, fracción IV)[20]: hasta unos 75 millones de pesos[21]. Se pide más cuidado: que el proveedor no entrene con ellos, que los guarde el menor tiempo posible y que haya controles de acceso.	Importa en datos de salud o laborales que se procesan en volumen en recursos humanos. El contrato de tratamiento de datos de OpenAI dice que «no se pretende transferir datos sensibles», salvo que el usuario los incluya de forma inesperada. Si un proceso usa datos sensibles, hay que negociarlo de forma expresa.
Ya existe en México una ley o un criterio oficial sobre IA y datos personales.	No hay ley de IA ni un criterio expreso. A agosto de 2026, la reforma constitucional para legislar en IA seguía en comisiones y la Ley General de IA era solo un documento de trabajo en el Senado[22]. La Secretaría Anticorrupción y Buen Gobierno (SABG) es la autoridad desde el 21 de marzo de 2025 (artículo 2, fracción XV)[20]. Las recomendaciones del INAI sobre IA, de 2022 y reeditadas en 2024, no son obligatorias[23]. Y el nuevo Reglamento sigue sin publicarse: la Cámara de Diputados solo lista el de 2011[24].	No cambia.
Puedo contratar cualquier servicio de IA en la nube aceptando sus términos estándar.	El artículo 52 del Reglamento impide al responsable adherirse a servicios de nube que no den garantías: entre otras, transparentar sus subcontrataciones, no asumir la titularidad de la información, permitir limitar el tratamiento y garantizar el borrado al terminar[14]. Para que el proveedor subcontrate se necesita autorización (artículo 54), pero si el contrato ya prevé la subcontratación, esa cláusula cuenta como autorización (artículo 55)[14].	Compras debe revisar el contrato de tratamiento de datos, las empresas que subcontrata el proveedor y el aviso cuando una autoridad pida información. Los contratos estándar se aceptan con un clic y autorizan de forma general la subcontratación, con un plazo para oponerse; el artículo 55 favorece que eso baste. Para el artículo 52 no hay criterio mexicano, así que conviene valorar una orden de compra negociada.

3. Secreto profesional y privilegio

Conviven dos marcos que no son equivalentes: el secreto profesional mexicano, y el privilegio abogado-cliente y la protección del trabajo del abogado (work product) de Estados Unidos.

Mito	Realidad	En la empresa
Usar IA es revelar el secreto profesional.	**Revelar es dar a conocer algo a otra persona, y los abogados siempre han tenido auxiliares.** El deber está en la ley: todo profesionista debe guardar estrictamente el secreto de lo que le confían sus clientes (Ley Reglamentaria del artículo 5º constitucional, artículo 36)[25], y revelarlo sin justa causa es delito, agravado para el profesionista: de 1 a 5 años de prisión y suspensión (Código Penal Federal, artículos 210 y 211)[26]. El deber consiste en elegir un medio que mantenga el secreto. La Barra Mexicana lo plantea igual: transmitir información a sistemas externos puede vulnerarlo si no hay salvaguardas[27]. La pregunta correcta es: «¿esta herramienta mantiene el secreto?».	El razonamiento se ha construido para el abogado frente a su cliente; el abogado de empresa frente a su empleador no está resuelto.
El gremio prohíbe usar IA.	Los Lineamientos de IA de la Barra Mexicana, Colegio de Abogados (aprobados el 14 de octubre de 2025) no prohíben la IA en general: piden no meter información confidencial del cliente en sistemas que no garanticen privacidad y seguridad y que no entrenen con ella, salvo consentimiento informado y explícito del cliente. Lo que sí prohíben expresamente son los sistemas de «caja negra»[27]. En la misma línea van la Opinión Formal 512 de la American Bar Association (julio de 2024), que exige consentimiento informado para herramientas que aprenden de los datos y advierte que una cláusula genérica en la carta de encargo no basta[28], y la guía del Consejo de la Abogacía Europea (CCBE, octubre de 2025), que acepta como salvaguardas la confidencialidad contractual, la opción de no guardar datos y el DPA[29].	La cuenta empresarial con garantías es justo el caso que esas guías aceptan.
Hablar con Claude o ChatGPT está protegido como si hablara con mi abogado.	**No.** «Claude no es tu abogado»: no tiene cédula ni deberes de lealtad hacia el usuario. En United States v. Heppner (Nueva York, febrero de 2026), el juez Rakoff resolvió que conversar con una IA no crea privilegio abogado-cliente: le bastó que Claude no fuera abogado, porque el privilegio exige una relación con un profesional con licencia, con deberes fiduciarios y sujeto a disciplina[16].	Tener cuenta empresarial no cambia esto: el privilegio no nace de la herramienta.
El caso Heppner dice que todo lo que le dices a una IA es público y se puede exigir en juicio.	**Heppner no dice eso, aunque es más amplio de lo que suele creerse.** Resolvió el caso de un cliente no abogado, ejecutivo de una empresa, que metió información por su cuenta en la versión pública de Claude. Le faltaron dos requisitos del privilegio: que hubiera un abogado y que hubiera confidencialidad, porque la política de consumo permitía entrenar con los datos y entregarlos a autoridades y en litigios[16]. Rakoff lo planteó como cuestión nueva sobre plataformas de IA disponibles al público. Lo que sigue abierto es la IA empresarial.	La IA empresarial con garantías en el contrato podría dar otro resultado: según Debevoise & Plimpton, una herramienta que no entrena y mantiene la confidencialidad debería reforzar el privilegio, aunque no lo garantiza[30]. No hay un precedente directo.
Cualquier contraparte puede pedirle a OpenAI o a Anthropic mi historial.	**La resolución de Rakoff no dice eso.** El FBI ya tenía los documentos por un cateo físico; el tribunal solo resolvió si estaban protegidos frente a la revisión del gobierno[16]. Lo que sí muestra el caso del New York Times es que una contraparte puede llegar lejos con datos de consumo: el juez Stein confirmó el 5 de enero de 2026 la orden de entregar 20 millones de registros desidentificados de ChatGPT[19]. Fueron muestras anónimas bajo orden de protección, no el historial de una persona concreta.	Hay que exigir en el contrato de tratamiento de datos que el proveedor avise cuando una autoridad pida información. OpenAI avisará «en la medida legalmente permitida» (DPA, cláusula 2.2) y Anthropic se obliga a avisar y a cooperar para acotar la petición (Commercial Terms, cláusula E.3).

Mito	Realidad	En la empresa
El trabajo que un abogado hace con IA nunca está protegido.	**Varios tribunales de Estados Unidos resolvieron lo contrario.** En Tremblay v. OpenAI (agosto de 2024), las instrucciones que los abogados escribieron a la IA se trataron como análisis protegido (opinion work product), porque reflejan sus impresiones mentales[31]. En Concord v. Anthropic (mayo de 2025), las instrucciones fallidas de los abogados y su configuración se trataron como trabajo protegido; solo hubo que entregar los cerca de 5,000 pares en que se apoyó la demanda[32]. Pero en diciembre de 2025 el mismo tribunal ordenó entregar las instrucciones que escribieron investigadores y testigos del cliente[33]. «La cédula es el escudo», aunque en materia civil no siempre es indispensable: en 2026, en Warner v. Gilbarco, se protegió incluso el uso de ChatGPT de una litigante sin abogado[34].	El abogado que usa una herramienta empresarial tiene la protección del trabajo del abogado. Es doctrina de Estados Unidos, sin equivalente en México, y no distingue entre el abogado de empresa y el externo.
Si un área de negocio usa IA para un asunto legal, también queda protegida.	**Si el cliente actúa por su cuenta, no (caso Heppner).** Si usa IA bajo la dirección del abogado, aplica la lógica del caso Kovel (Segundo Circuito, 1961), que protege a los auxiliares cuya intervención es necesaria o muy útil para la consulta legal[35]. El propio Rakoff lo dejó abierto: si el abogado hubiera dirigido el uso, Claude podría haberse equiparado a un agente del abogado[16]. Con IA comercial bajo dirección del abogado, la cuestión es debatible.	Es el escenario clave para los equipos internos: que el área legal dirija el trabajo y que se use la cuenta empresarial. El apoyo es por analogía y por un comentario del propio juez en Heppner, no por un precedente.
La cláusula de «no confidencialidad» de los términos equivale a autorizar que se divulgue la información.	**Esa cláusula existe para poder reentrenar el modelo, y «reentrenar no es divulgar».** Aplica la doctrina de los actos propios (venire contra factum proprium): la privacidad es la regla y el acceso la excepción, y la intención del usuario debe pesar. Es un argumento doctrinal, y en Estados Unidos ya tiene en contra a Heppner, que negó la confidencialidad justo porque la política preveía entrenar con los datos y divulgarlos[16]. A favor: en Warner y en Morgan v. V2X los tribunales sostuvieron que meter información en una IA no renuncia a la protección del trabajo del abogado, porque esa renuncia exige revelarla a la contraparte[34,36].	En la cuenta empresarial, la cláusula que prohíbe entrenar hace que este argumento sea menos necesario.

4. Anonimización y datos que delatan

Mito	Realidad	En la empresa
Si anonimizo quitando los nombres, puedo usar cualquier IA.	**Hay cinco razones en contra.** Quitar nombres solo disfraza los datos, no los anonimiza. Los hechos identifican: «lo que identifica no es el nombre: es la historia». Ya en 2000, Latanya Sweeney mostró que el código postal, el sexo y la fecha de nacimiento bastan para identificar al 87% de la población de Estados Unidos[37], y en 2019 un estudio en Nature Communications estimó que con 15 atributos demográficos se reidentifica al 99.98%[38]. Lo sensible sigue ahí. El trabajo sale peor, se cometen errores y quedan datos ocultos en los archivos (metadatos). Y da una falsa sensación de cumplimiento. Aun así, usar solo los datos necesarios sigue siendo una buena práctica.	No sustituye a la cuenta empresarial con contrato de tratamiento de datos.
Si el dato no trae nombre, no es un dato personal.	La ley protege los datos de cualquier persona identificada o identificable, es decir, cuya identidad pueda determinarse directa o indirectamente con cualquier información, aunque no aparezca su nombre (artículo 2, fracción V)[20].	No cambia.

Mito	Realidad	En la empresa
Si quito el nombre, la raza o el género, el modelo ya no puede discriminar.	**Borrar el dato sensible no basta.** El modelo lo deduce de otros datos que lo delatan (proxies): el nombre, el código postal, el idioma, el acento o la forma de escribir. Bloquear los proxies obvios sirve de poco, porque la IA encuentra otros[39]. La herramienta de reclutamiento de Amazon penalizaba los currículums que decían «women's», como capitana de un club de ajedrez femenil[40]. En 2024, al filtrar currículums por nombre, los modelos prefirieron nombres asociados a personas blancas en el 85% de los casos y nombres femeninos solo en el 11%[41]. Y aun con currículums anonimizados, 18 modelos dedujeron la etnia por los idiomas y el género por los pasatiempos[42].	Importa en filtros de currículums, de crédito o de precios que se aplican en volumen.
Entre más anonimice, mejor.	Si se anonimiza de más, se pierde información valiosa y el modelo deja de servir. Si se anonimiza de menos, el modelo puede deducir quién es la persona: bastan unos cuantos atributos[38]. Hay que decidirlo según la tarea.	No cambia.
Todo sesgo hay que eliminarlo.	**No siempre.** En medicina, algunas guías usan la raza como criterio: para pacientes negros con hipertensión, la guía JNC 8 recomienda empezar con un diurético tiazídico o un calcioantagonista[43], y la Asociación Americana de Diabetes bajó el umbral de tamizaje para estadounidenses de origen asiático[44]. Aun esos ajustes están bajo revisión crítica[45]. En cambio, cuando las condiciones son iguales, como en el salario, no hay justificación. Si es deseable que un sistema recomiende inversiones más conservadoras a nombres afroamericanos lo deben decidir los legisladores, no el modelo: el estudio de Stanford que documentó ese patrón advierte que cualquier diferencia, aunque parezca justificada, exige una decisión deliberada[46].	La empresa debe definir qué distinciones son legítimas en cada proceso.

5. Sesgos en procesos masivos de la empresa

Mito	Realidad	En la empresa
Los filtros de seguridad de la IA impiden que discrimine.	Estos filtros, llamados guardarraíles, funcionan por el entrenamiento, por comparación con casos parecidos («no mide intención: mide parecido») y por las instrucciones del sistema. Reaccionan a cómo está escrita la pregunta, no al daño que puede causar la respuesta: «protege contra las preguntas que suenan a discriminación, no contra las respuestas que resultan discriminatorias». La evidencia es contundente: en GPT-4o, reformular una petición dañina en tiempo pasado subió su tasa de éxito del 1% al 88%[47], y ocho modelos alineados para ser imparciales mostraron sesgos implícitos en 21 estereotipos[48]. Solo un clasificador externo, que es opcional, se parece a un filtro real.	Tener cuenta empresarial no cambia esto.
Si el modelo se niega a responder o pone una advertencia, no hubo sesgo.	**DeepSeek razonó a partir del estereotipo, se negó a responder y luego dio el veredicto:** «si vamos por el cliché visual, el primero». «El descargo no suprimió el sesgo, le dio cobertura».	Una advertencia en la respuesta no es prueba de que el proceso sea neutral.
Una pregunta neutral da una respuesta neutral.	**Se preguntó cuánto cuesta la misma bicicleta según quién la vende.** Para un hombre blanco, el modelo estimó de 500 a 1,500 dólares; para un latino, de 300 a 800; para una persona de apariencia indígena, de 100 a 300. El precio real era de 250. Un estudio de Stanford encontró el mismo patrón: GPT-4 ofreció unos 18,718 dólares por un coche a vendedores con nombres asociados a personas blancas y unos 16,410 a nombres asociados a personas negras[46]. La discriminación entra por una tarea que parece técnica y que el filtro de seguridad no detecta.	El riesgo está en decisiones masivas que nadie ve: crédito, empleo o renta, con sistemas que actúan solos (agentes) detrás.

Mito	Realidad	En la empresa
Los modelos más nuevos o mejores tienen menos sesgo.	**«La calidad del modelo no predice menos sesgo».** En una auditoría de Stanford con 168,000 respuestas de seis modelos (GPT-4, GPT-3.5, GPT-4o, PaLM-2, Mistral Large y Llama-3 70B), GPT-3.5 sesgó menos que GPT-4 y los demás mostraron diferencias incluso mayores. Los escenarios eran cotidianos: cuánto ofrecer por un coche o una bicicleta según si el vendedor se llama, por ejemplo, Hunter Becker o Jamal Washington (Salinas, Haim y Nyarko, «What's in a Name?», Stanford, 2025)[46].	Pagar por el mejor modelo mejora la capacidad, pero no está demostrado que reduzca el sesgo.
El sesgo siempre perjudica a las minorías.	Va en ambos sentidos. En la misma auditoría, al simular la selección de jugadores de basquetbol, los nombres afroamericanos obtuvieron mejores evaluaciones[46].	No cambia.
Basta con darle al sistema la instrucción «no discrimines».	**Esa instrucción de base (system prompt) ayuda, pero no es confiable.** En pruebas controladas de Anthropic, frases como «es ilegal discriminar» redujeron la discriminación de Claude 2 casi a cero[49]; en la auditoría de Stanford, en cambio, añadir contexto tuvo efectos inconsistentes y solo un dato numérico concreto eliminó la diferencia[46]. El entrenamiento importa más: un modelo entrenado solo con datos jurídicos limpios (KL3M) resultó menos tóxico que otros siete modelos pequeños, según sus propios desarrolladores[50].	Las instrucciones que configura la empresa no sustituyen las pruebas de sesgo.
El sesgo se puede quitar del modelo de una vez por todas.	Desactivar ciertas partes de la red (poda de neuronas) eliminó casi por completo el sesgo por nombres en escenarios concretos, como compras y servicios. Pero al trasladar la poda de un contexto a otro, por ejemplo de finanzas a transacciones comerciales, el sesgo solo bajó alrededor de 40%: no existe una «neurona del sesgo» universal[51].	No cambia.
COMPAS es la prueba de cómo falla la IA hoy.	«El ejemplo que todos citan, y que ya no alcanza». COMPAS es una herramienta estadística de riesgo, no IA generativa. ProPublica analizó a 7,214 acusados en Broward, Florida, y encontró que clasificó por error como de alto riesgo al 45% de los acusados negros frente al 23% de los blancos[52]. En State v. Loomis (Wisconsin, 2016), el tribunal permitió usarlo, pero no como factor determinante y con advertencias por escrito[53]. Sigue siendo útil para hablar de debido proceso y del teorema de imposibilidad: si los grupos tienen tasas base distintas, no se pueden cumplir a la vez todas las medidas de equidad[54,55]. Pero la IA actual falla por otras vías: datos que delatan, consejos sesgados y filtros frágiles. Conviene verificar siempre de qué año es el estudio.	No cambia.
El sesgo es un problema solo de los chatbots de texto.	**También aparece en otros modelos.** En transcripción, cinco sistemas comerciales (Amazon, Apple, Google, IBM y Microsoft) cometieron casi el doble de errores con hablantes negros que con blancos[56]. En imagen, Stable Diffusion representó como mujeres solo al 3% de los jueces, cuando en Estados Unidos son el 34%[57]. Y en el reconocimiento de texto en documentos escaneados (OCR), el error va del 1.7% en inglés moderno a más del 70% en alemán histórico[58].	Aplica a las herramientas de digitalización y transcripción de la empresa.

6. Transparencia, explicabilidad y auditoría

Mito	Realidad	En la empresa
Transparencia y explicabilidad son lo mismo.	**«Se puede publicar todo y aun así no entender tu caso».** Ante una decisión automatizada, lo que la persona necesita es saber por qué se decidió así. El Tribunal de Justicia de la Unión Europea lo dijo en febrero de 2025: no basta con entregar una fórmula matemática compleja; hay que explicar el procedimiento y los criterios que de verdad se aplicaron (asunto C-203/22, Dun & Bradstreet Austria)[59].	En procesos masivos, lo útil es exigir que se explique el resultado.

Mito	Realidad	En la empresa
Se puede auditar el modelo por dentro o exigir al proveedor su receta.	**El núcleo del modelo no se puede auditar.** Los laboratorios no revelan con qué datos lo entrenaron: en el índice de transparencia de Stanford, el promedio de los desarrolladores cayó a 41 de 100 en 2025, y los datos de entrenamiento siguen siendo opacos[60]. Y ni ellos entienden del todo por qué el modelo hace lo que hace: el director de Anthropic lo reconoció públicamente en 2025[61]. Es como la receta de una pastelería. Lo que sí se audita es lo que rodea al modelo, y lo que se exige es que explique su respuesta.	Un contrato empresarial no da acceso al núcleo del modelo.
El razonamiento que muestra el modelo explica por qué decidió.	Ese razonamiento es «plausible, no causal verificable»: suena lógico, pero no prueba que esa sea la causa real. Cuando se introducen sesgos en la pregunta, la precisión cae hasta 36% y las explicaciones del modelo no los mencionan[62]; Anthropic encontró que Claude 3.7 Sonnet admitió haber usado una pista en solo una cuarta parte de los casos[63]. Pasa lo mismo con un juez: le exigimos una sentencia motivada, no un registro de lo que pasó por su mente.	No cambia.
Basta con auditar la respuesta final o una sola parte del sistema.	**Un sistema de IA tiene varias capas que se pueden auditar.** La construcción: entrenamiento, ajuste fino y entrenamiento con evaluaciones humanas (RLHF). La configuración: instrucciones de base, filtros de seguridad y contexto. Y la ejecución: búsqueda en documentos propios (RAG), la cantidad de texto que el modelo puede leer a la vez (ventana de contexto) y los flujos con agentes. La literatura académica propone también auditar por capas: gobernanza, modelo y aplicación[64]. El sistema es tan confiable como su capa más débil, y la revisión suele concentrarse en una sola.	La empresa controla sobre todo la configuración y la ejecución.

7. Gobernanza interna, supervisión humana y adopción

Mito	Realidad	En la empresa
Si una persona revisa, ya hay supervisión.	**Hay tres formas de participar, según las directrices éticas de la Unión Europea:** la persona dentro del proceso, que decide; la persona sobre el proceso, que vigila; y la persona al mando, que gobierna[65]. Cuando la gente confía de más en la máquina (sesgo de automatización), la revisión se vuelve un «sello de goma»: con sugerencias erróneas de una IA, la precisión de radiólogos expertos cayó de 82% a 45.5%[66]. Por eso la Ley de IA europea obliga a prevenir expresamente ese sesgo (artículo 14)[67], y un análisis de 41 políticas concluyó que las personas no logran cumplir la supervisión que se les asigna[68]. «Supervisar no es mirar. Supervisar es poder contradecir».	Aplica al diseño de los flujos internos de aprobación.
La decisión final se puede dejar en manos de la IA.	**Nunca.** El responsable lo sigue siendo, y debe revisar las citas, las tesis y los cálculos. Incluso herramientas jurídicas especializadas, Lexis+ AI y Westlaw AI, inventaron información entre el 17% y el 33% de las veces en un estudio de Stanford[69], y una base de datos pública registra ya 2,241 resoluciones judiciales sobre escritos con alucinaciones de IA[70].	No cambia por tener cuenta empresarial.
Los agentes de IA ya pueden operar procesos solos.	Según el centro de investigación METR, a mediados de 2026 los modelos más avanzados completan con 50% de éxito tareas que a un experto le llevarían de 6 a 17 horas, pero solo son confiables, con 80% de éxito, en tareas de 1 a 3 horas[71]. Y solo se midió en ingeniería de software, aprendizaje automático y ciberseguridad.	Hay que tener cautela antes de automatizar procesos completos.

Mito	Realidad	En la empresa
Un modelo gratuito o barato es suficiente.	Los modelos gratuitos suelen ser menos capaces y tienen peores políticas de datos, y la transparencia varía mucho: en el índice de Stanford, DeepSeek obtuvo 32 de 100 y xAI 14[60]. Conviene no escatimar y usar el mejor modelo, por su capacidad de razonar y por la cantidad de texto que puede leer a la vez.	Coincide con usar la cuenta empresarial de pago. Los planes de consumo de Claude no tienen inicio de sesión único, registros de actividad ni plazos de conservación a la medida; los de OpenAI y Claude entrenan por defecto o salvo que el usuario lo rechace; y la protección de Microsoft y Google frente a demandas por derechos de autor excluye el uso gratuito o de consumo. Pero un mejor modelo no garantiza menos sesgo.
Conectar la IA a internet siempre mejora el resultado.	**El acceso abierto puede contaminar el expediente con información ajena.** Son preferibles los ambientes cerrados, que trabajan solo con los documentos que se les dan, y las IA especializadas, que son más cuidadosas. NotebookLM, hoy Gemini Notebook dentro de Workspace, no guarda las instrucciones al cerrar la sesión, no tiene revisión por personas y no entrena con los datos. Pero Google reconoce que todavía no tiene las certificaciones ISO, SOC ni FedRAMP, que no lo cubre su acuerdo para datos de salud (BAA) y que no le aplican la región de datos ni la protección contra fuga de datos de Workspace[72].	Coincide con usar entornos corporativos acotados. NotebookLM Enterprise, de Google Cloud, es otro producto, y sí está entre los servicios en los que Google defiende al cliente por propiedad intelectual.
La calidad del resultado depende solo del modelo.	**Depende de la información que se le da:** limpia y completa, con el texto de los escaneos verificado, la norma vigente con sus reformas y los criterios relevantes. Si se le da información de mala calidad, aparecen sesgos e invenciones (alucinaciones); en el estudio de Stanford sobre herramientas jurídicas, buena parte de los errores venía de documentos mal recuperados[69].	Es responsabilidad del área que usa la herramienta.
Si ya tenemos cuenta empresarial, no hace falta una política interna.	**Hace falta una política interna.** Configurar la cuenta: entrenamiento apagado, conservación mínima, verificación en dos pasos y baja de usuarios que se van. Un contrato y un aviso de privacidad actualizados. Usar solo los datos necesarios. Supervisión. Y revisar los términos cada cierto tiempo.	Es el núcleo de la gobernanza interna. El contrato de tratamiento de datos de OpenAI hace al cliente responsable de configurar la conservación y el borrado conforme a la ley (cláusula 3.3). En Claude, los plazos de conservación a la medida solo existen en Enterprise, no en Team.
No usar IA es lo prudente.	No usarla es dejar de hacer lo que un abogado debería saber hacer: leer el contrato. La decisión importante no es si usar IA, sino cuál usar y en qué condiciones.	Aplica igual.

Matriz comparativa de plataformas

● favorable · ● revisar o configurar · ● desfavorable

Planes empresariales

Criterio	ChatGPT Business y Enterprise	Claude Team y Enterprise	Microsoft 365 Copilot	Gemini en Workspace
¿Usa los datos del cliente para entrenar sus modelos?	No por defecto ●	No ●	No (modelos base) ●	No fuera de la organización ●
¿Cuánto tiempo guarda los datos?	Lo decide el administrador; lo borrado se elimina en 30 días ●	Hasta que se borra; lo marcado por abuso, 2 años ●	Según las políticas de Purview ●	De 90 días a indefinido ●
¿Firma contrato de tratamiento de datos (DPA) como encargado?	Sí, como encargado ●	Sí, como encargado ●	Sí, DPA de Microsoft ●	Sí, Cloud Data Processing Addendum ●
¿Ofrece guardar los datos en México?	No se ofrece ●	No publicado	Pueden procesarse fuera ●	No publicado
Certificación ISO 42001 (gestión de IA)	Sí ●	Sí ●	Sí ●	Sí ●
Registros de actividad para auditoría	En Enterprise (Compliance API) ●	Solo en Enterprise ●	Purview ●	Sí ●
¿Personas del proveedor pueden leer conversaciones?	En Business, contratistas que revisan abusos ●	Por defecto no; sin página específica para Team y Enterprise ●	Excluido ●	No sin permiso ●
¿Defiende al cliente si lo demandan por propiedad intelectual?	Sí, con exclusiones ●	Sí, con exclusiones ●	Sí, con condiciones (Customer Copyright Commitment) ●	Sí, con condiciones ●

API y nube

Criterio	API de OpenAI	API de Anthropic	Azure OpenAI y Foundry	Vertex AI
¿Usa los datos del cliente para entrenar sus modelos?	No ●	No ●	No ●	No ●
¿Cuánto tiempo guarda los datos?	Registros para detectar abusos, 30 días ●	30 días ●	Los modelos no guardan las consultas ●	Guarda las instrucciones para detectar abusos ●
¿Ofrece no guardar nada? (Zero Data Retention)	Con aprobación ●	Por acuerdo ●	Monitoreo de abuso modificado ●	Con pasos específicos ●
¿Personas del proveedor pueden leer conversaciones?	Personal y contratistas ●	No publicado	Muestras marcadas ●	Guarda instrucciones para vigilar abusos ●
¿Dónde se guardan y procesan los datos?	10 regiones, sin México ●	Opción de solo Estados Unidos ●	Región elegida por el cliente ●	No publicado
¿Defiende al cliente si lo demandan por propiedad intelectual?	Sí ●	Sí ●	Si se aplican las medidas exigidas ●	Modelos en versión estable ●

Checklist del contrato de tratamiento de datos (DPA)

Antes de contratar o renovar.

Punto a revisar	Qué prevén hoy los contratos
☐ **El proveedor actúa como encargado y solo trata los datos por instrucción de la empresa (artículos 50 a 54 del Reglamento)**	OpenAI DPA, cláusulas 1.1 y 2.1; Anthropic DPA, secciones B.1 y B.2; Cloud Data Processing Addendum de Google; DPA de Microsoft 🟢
☐ **Prohibición de que el proveedor use los datos para sí o para entrenar**	OpenAI Services Agreement, cláusula 4.2; Anthropic Commercial Terms, sección B; Google y Microsoft, en sus páginas de privacidad 🟢
☐ **Aviso cuando una autoridad pide información (artículo 52)**	OpenAI DPA, cláusula 2.2, «en la medida legalmente permitida»; Anthropic Commercial Terms, cláusula E.3. Microsoft y Google: no publicado 🟡
☐ **No aceptar servicios que no den garantías (artículo 52)**	Los contratos se aceptan con un clic; conviene valorar una orden de compra negociada 🟡
☐ **Autorización para que el proveedor subcontrate a otros (artículo 54)**	Autorización general con plazo para oponerse: 30 días en OpenAI y 15 días en Anthropic. El artículo 55 del Reglamento indica que la cláusula del contrato cuenta como autorización 🟡
☐ **Obligación de confidencialidad del personal del proveedor**	OpenAI DPA, cláusula 2.3; Anthropic DPA, sección B.7 🟢
☐ **Quién del proveedor puede leer las conversaciones y por qué (revisión de abusos)**	Cambia según el plan: ver la matriz. Pedir el monitoreo de abuso modificado o la opción de no guardar datos cuando aplique 🟡
☐ **Plazo de conservación mínimo configurado y plazo de borrado**	Lo configura el administrador en ChatGPT Enterprise y en Workspace; en Claude, solo en Enterprise. Al terminar el contrato, se borra en 30 días (OpenAI Services Agreement, cláusula 11.3; Anthropic DPA, sección H.1) 🟡
☐ **Los datos sensibles están cubiertos de forma expresa**	OpenAI DPA, Anexo 1: no se prevé transferir datos sensibles. Hay que negociarlo 🔴
☐ **Seguridad y auditoría (informes SOC 2 e ISO)**	Informes a solicitud: OpenAI Services Agreement, cláusula 5.2 (una vez al año); Anthropic DPA, sección F 🟢
☐ **Aviso en caso de brechas de seguridad**	Anthropic: 48 horas (DPA, sección G); OpenAI: sin demora indebida (DPA, cláusula 2.7) 🟢
☐ **El lugar donde se guardan los datos cumple la política interna (la remisión internacional está permitida)**	OpenAI no ofrece región en México; revisar las opciones de Microsoft, Anthropic y Google 🟡
☐ **Las valoraciones de los usuarios (pulgar arriba o abajo) están apagadas o reguladas**	En OpenAI, las valoraciones pueden usarse para entrenar; Anthropic las conserva 5 años 🟡
☐ **Inicio de sesión único, verificación en dos pasos, registros de actividad y baja de usuarios**	ChatGPT Enterprise y Claude Enterprise los tienen; Claude Team no tiene registros de actividad 🟡
☐ **Cambios en los términos y revisión periódica**	Aviso de 30 días antes de cambios importantes: OpenAI Services Agreement, cláusula 16.13; Anthropic Commercial Terms, cláusula M.3 🟡

Anexo: detalle por proveedor

● favorable · ● revisar o configurar · ● desfavorable

OpenAI

Criterio	ChatGPT Free, Plus y Pro (consumo)	ChatGPT Business y Enterprise	API de OpenAI
¿Usa los datos del cliente para entrenar sus modelos?	Sí, por defecto, aunque se puede apagar. Marcar una respuesta con pulgar arriba o abajo puede hacer que el chat se use para entrenar ●	No por defecto; solo si el cliente lo acepta ●	No desde el 1 mar 2023, salvo que el cliente lo active ●
¿Cuánto tiempo guarda los datos?	Un chat borrado se elimina en 30 días. Un chat temporal se guarda hasta 30 días ●	Lo decide el administrador. Lo borrado se elimina en 30 días, salvo obligación legal (en Business, también para proteger el servicio) ●	Los registros para detectar abusos se guardan hasta 30 días; algunos elementos se guardan hasta que el cliente los borre ●
¿Ofrece no guardar nada? (Zero Data Retention)	No ●	No como tal, pero el plazo se puede configurar ●	Sí, con aprobación previa y para casos que califiquen ●
¿Firma contrato de tratamiento de datos (DPA)?	No ●	Sí, el DPA forma parte del Services Agreement ●	Sí ●
¿Actúa como encargado o como responsable?	No hay DPA; el proveedor es responsable ●	Encargado: «acts as a Data Processor» (DPA, cláusula 1.1) ●	Encargado (DPA, cláusula 1.1) ●
¿Dónde se guardan y procesan los datos?	No publicado	Para clientes nuevos de Enterprise y Edu hay 10 regiones, sin México. Procesamiento dentro de la región: Estados Unidos, Europa y Emiratos Árabes Unidos ●	10 regiones, sin México. Fuera de Estados Unidos exige no guardar datos (Zero Data Retention) o el monitoreo de abuso modificado ●
Certificaciones de seguridad	La ISO 42001 cubre productos de consumo y empresariales	SOC 2 Tipo 2; ISO 27001, 27017, 27018 y 27701 (Enterprise y Edu); ISO 42001 ●	SOC 2 Tipo 2; ISO 27001, 27017, 27018 y 27701; ISO 42001 ●
Controles de administración, inicio de sesión único y registros de actividad	No ●	Inicio de sesión único (SAML SSO) y controles detallados. Registros de actividad mediante la Compliance API (Enterprise) ●	Controles por organización y por proyecto; el cliente puede administrar sus propias llaves de cifrado (Enterprise Key Management) ●
¿Personas del proveedor pueden leer conversaciones?	No publicado	Enterprise: solo en incidentes, con permiso o por obligación legal. Business: personal y contratistas externos ●	Personal autorizado y contratistas externos, para detectar abusos ●
¿Defiende al cliente si lo demandan por propiedad intelectual?	No publicado	Sí, por infracción de propiedad intelectual de los servicios (cláusula 13.1), con exclusiones ●	Sí (cláusula 13.1), con exclusiones ●

Anthropic

Criterio	Claude Free, Pro y Max (consumo)	Claude Team y Enterprise	API de Anthropic
¿Usa los datos del cliente para entrenar sus modelos?	Lo decide el usuario y está activo salvo que lo rechace; si acepta, los datos se guardan 5 años 🔴	No: «may not train models on Customer Content» 🟢	No (Commercial Terms, sección B) 🟢
¿Cuánto tiempo guarda los datos?	30 días si no entrena; 5 años si entrena 🟡	Los chats se guardan hasta que se borran, y luego 30 días más. Lo marcado por violar la política de uso: 2 años (las puntuaciones, 7). Las valoraciones de los usuarios: 5 años 🟡	Se borra en 30 días, salvo archivos subidos (Files API), violaciones de la política de uso u obligación legal 🟡
¿Ofrece no guardar nada? (Zero Data Retention)	No 🔴	Plazos a la medida solo en Enterprise, no en Team 🟡	Sí, mediante un acuerdo específico 🟢
¿Firma contrato de tratamiento de datos (DPA)?	No (aplican los términos de consumo) 🔴	Sí, incorporado por referencia 🟢	Sí 🟢
¿Actúa como encargado o como responsable?	No hay DPA 🔴	Encargado (DPA, sección B.1) 🟢	Encargado (DPA, sección B.1) 🟢
¿Dónde se guardan y procesan los datos?	No publicado	No publicado	Opción de procesar solo en Estados Unidos, con un precio 1.1 veces mayor 🟡
Certificaciones de seguridad	No publicado	ISO 27001:2022; ISO 42001; SOC 2 Tipo I y II; HIPAA, con acuerdo para datos de salud (BAA) 🟢	ISO 27001:2022; ISO 42001; SOC 2 Tipo I y II 🟢
Controles de administración, inicio de sesión único y registros de actividad	Sin inicio de sesión único ni registros de actividad 🔴	Team: inicio de sesión único, sin registros de actividad. Enterprise: inicio de sesión único, alta y baja automática de usuarios (SCIM), roles, registros de actividad y Compliance API 🟡	No publicado
¿Personas del proveedor pueden leer conversaciones?	Por defecto no; sí en valoraciones y revisión por abuso 🟡	Sin página específica para Team y Enterprise	No publicado
¿Defiende al cliente si lo demandan por propiedad intelectual?	No publicado	Sí (sección K.1), con exclusiones (sección K.3) 🟢	Sí (sección K.1), con exclusiones (sección K.3) 🟢

Microsoft

Criterio	Copilot de consumo	Microsoft 365 Copilot	Azure OpenAI y Foundry
¿Usa los datos del cliente para entrenar sus modelos?	No publicado	No usa las instrucciones, las respuestas ni los datos de la organización para entrenar sus modelos base 🟢	No sin permiso; OpenAI no recibe los datos 🟢

Criterio	Copilot de consumo	Microsoft 365 Copilot	Azure OpenAI y Foundry
¿Cuánto tiempo guarda los datos?	No publicado	Igual que el resto del contenido de Microsoft 365, con las políticas de conservación de Purview 🟡	Los modelos no guardan las consultas; lo que sí se almacena (asistentes y archivos) lo borra el cliente 🟢
¿Ofrece no guardar nada? (Zero Data Retention)	No publicado	Excluido del monitoreo de abuso con revisión por personas 🟢	Monitoreo de abuso modificado, si Microsoft aprueba la solicitud 🟡
¿Firma contrato de tratamiento de datos (DPA)?	No publicado	Términos de producto y DPA de Microsoft 🟢	Microsoft Products and Services DPA 🟢
¿Actúa como encargado o como responsable?	No publicado	Según el DPA de Microsoft 🟢	Según el DPA de Microsoft 🟢
¿Dónde se guardan y procesan los datos?	No publicado	Límite de datos de la Unión Europea (EU Data Boundary), residencia avanzada (ADR) y varias regiones (Multi-Geo). Fuera de la Unión Europea puede procesarse en Estados Unidos, en la Unión Europea o en otras regiones. Los modelos de Anthropic quedan fuera del límite de datos europeo 🟡	En la región que elija el cliente, salvo en despliegues globales (Global) o por zona (DataZone) 🟡
Certificaciones de seguridad	No publicado	ISO 27001, ISO 42001, ISO/IEC 27018, HIPAA y RGPD 🟢	No publicado
Controles de administración, inicio de sesión único y registros de actividad	No publicado	Entra ID, Purview, búsqueda de contenido y búsqueda para litigios (eDiscovery) 🟢	Llave de cifrado administrada por el cliente y controles de Azure 🟢
¿Personas del proveedor pueden leer conversaciones?	No publicado	No (excluido) 🟢	Sí, de muestras marcadas, salvo con monitoreo de abuso modificado 🟡
¿Defiende al cliente si lo demandan por propiedad intelectual?	No: el compromiso de defensa por derechos de autor excluye los productos gratuitos y de consumo 🔴	Sí, con el Customer Copyright Commitment, si se usan los filtros y salvaguardas 🟢	Sí, con el Customer Copyright Commitment, solo si se aplican las medidas exigidas (instrucciones de base, pruebas y filtros) 🟡

Google

Criterio	Gemini de consumo	Gemini en Workspace y app Gemini	Vertex AI (hoy Gemini Enterprise Agent Platform)
¿Usa los datos del cliente para entrenar sus modelos?	No publicado	No fuera de la organización sin permiso 🟢	No sin permiso o instrucción previa 🟢

Criterio	Gemini de consumo	Gemini en Workspace y app Gemini	Vertex AI (hoy Gemini Enterprise Agent Platform)
¿Cuánto tiempo guarda los datos?	No publicado	Workspace: de 90 días a indefinido, según el administrador. App Gemini dentro de Workspace: hasta 36 meses. Gemini Notebook: no guarda las instrucciones al cerrar la sesión 🟡	Guarda las instrucciones para detectar abusos. Respuestas apoyadas en Google Search: 3 días; en Maps: 30 días. La Interactions API guarda por defecto 🟡
¿Ofrece no guardar nada? (Zero Data Retention)	No publicado	No publicado	Solo con pasos específicos y no para todas las funciones 🟡
¿Firma contrato de tratamiento de datos (DPA)?	No publicado	Cloud Data Processing Addendum 🟢	Cloud Data Processing Addendum 🟢
¿Actúa como encargado o como responsable?	No publicado	Trata los datos solo según las instrucciones del cliente (Cloud Data Processing Addendum) 🟢	Encargado según el Cloud Data Processing Addendum 🟢
¿Dónde se guardan y procesan los datos?	No publicado	No publicado	No publicado
Certificaciones de seguridad	No publicado	SOC 1, 2 y 3; ISO 9001, 27001, 27701, 27017, 27018 y 42001; FedRAMP High. Gemini Notebook todavía no tiene ISO ni SOC 🟢	No publicado
Controles de administración, inicio de sesión único y registros de actividad	No publicado	Controles de administración, conservación, Vault y registros de actividad 🟢	No publicado
¿Personas del proveedor pueden leer conversaciones?	No publicado	No sin permiso 🟢	Guarda las instrucciones para vigilar abusos 🟡
¿Defiende al cliente si lo demandan por propiedad intelectual?	No publicado	Gemini en Workspace y Vids: cubre las respuestas sin modificar y los datos de entrenamiento 🟢	API con modelos Gemini en versión estable (GA), respuestas apoyadas en búsqueda (Grounding) y NotebookLM Enterprise 🟢

Referencias

Fuentes que respaldan la columna «Realidad», en orden de aparición. Todas se consultaron el 30 de octubre de 2026.

1. OpenAI, «Data controls in ChatGPT», centro de ayuda, consultado el 30 oct 2026. https://help.openai.com/en/articles/7730893-data-controls-faq
2. Anthropic, «Updates to our consumer terms», 28 ago 2025. https://www.anthropic.com/news/updates-to-our-consumer-terms
3. OpenAI, Services Agreement, vigente desde el 1 ene 2026. https://openai.com/policies/services-agreement/
4. Anthropic, Commercial Terms of Service, consultados el 30 oct 2026. https://www.anthropic.com/legal/commercial-terms
5. Anthropic Privacy Center, «Is my data used for model training?» (productos comerciales), 18 ago 2026. https://privacy.claude.com/en/articles/7996868-is-my-data-used-for-model-training
6. Anthropic Privacy Center, «Does Anthropic act as a data processor or controller?», consultado el 30 oct 2026. https://privacy.claude.com/en/articles/9267385-does-anthropic-act-as-a-data-processor-or-controller
7. Anthropic, planes y precios de Claude, consultado el 30 oct 2026. https://claude.com/pricing
8. OpenAI, «Deleting and archiving chats in ChatGPT», centro de ayuda, consultado el 30 oct 2026. https://help.openai.com/en/articles/8809935-deleting-and-archiving-chats-in-chatgpt
9. Anthropic Privacy Center, «How long do you store my organization's data?», 1 jul 2026. https://privacy.anthropic.com/en/articles/7996866-how-long-do-you-store-my-organization-s-data
10. OpenAI, «How we're responding to The New York Times' data demands», 5 jun 2025, actualizado el 22 oct 2025. https://openai.com/index/response-to-nyt-data-demands/
11. In re OpenAI, Inc. Copyright Infringement Litigation, No. 25-md-03143 (S.D.N.Y.), Stipulation and Order, ECF 992, 9 oct 2025. https://blogs.duanemorris.com/classactiondefense/wp-content/uploads/sites/56/2025/10/In-re-Open-AI-Inc.-Copyright-Infringement-Litig.-10.9.25-Order-ECF-No.-992.pdf
12. Microsoft Learn, «Data, privacy, and security for Azure OpenAI / Foundry», 19 may 2026. https://learn.microsoft.com/en-us/legal/cognitive-services/openai/data-privacy
13. Microsoft Learn, «Data, privacy, and security for Microsoft 365 Copilot», 18 ago 2026. https://learn.microsoft.com/en-us/copilot/microsoft-365/microsoft-365-copilot-privacy
14. Reglamento de la LFPDPPP, DOF 21 dic 2011. https://www.gob.mx/cms/uploads/attachment/file/724929/Reg._Ley_Fed._Protecc._Datos_Personales_21-12-2011.pdf
15. Microsoft Learn, «Customer Copyright Commitment», 20 mar 2026. https://learn.microsoft.com/en-us/azure/foundry/responsible-ai/openai/customer-copyright-commitment
16. United States v. Heppner, No. 25 Cr. 503 (JSR) (S.D.N.Y.), memorándum y orden del juez Jed S. Rakoff, ECF 27, feb 2026. https://storage.courtlistener.com/recap/gov.uscourts.nysd.652138/gov.uscourts.nysd.652138.27.0.pdf
17. BBC, «Woman accused of using ChatGPT to plan drug murders», 19 feb 2026. https://www.bbc.com/news/articles/clyv80e5dljo
18. Anthropic, Privacy Policy, consultada el 30 oct 2026. https://www.anthropic.com/legal/privacy
19. ABA Journal, «ChatGPT creator must turn over 20M chat logs in copyright litigation, federal judge says», ene 2026. https://www.abajournal.com/news/article/chatgpt-creator-must-turn-over-20m-chat-logs-in-copyright-litigation-federal-judge-says
20. Ley Federal de Protección de Datos Personales en Posesión de los Particulares, DOF 20 mar 2025, texto vigente (Cámara de Diputados). https://www.diputados.gob.mx/LeyesBiblio/pdf/LFPDPPP.pdf
21. IDC Online, «Valor UMA 2026: INEGI publica nuevos montos», 9 ene 2026. https://idconline.mx/laboral/2026/01/09/valor-uma-2026-inegi-publica-nuevos-montos
22. ITSitio, «Regulación de IA en México: todavía no hay una ley general, pero las reglas ya avanzan por sectores», 3 ago 2026. https://www.itsitio.com/mx/inteligencia-artificial/regulacion-de-ia-en-mexico-todavia-no-hay-una-ley-general-pero-las-reglas-ya-avanzan-por-sectores/
23. INAI, Recomendaciones para el tratamiento de datos personales derivado del uso de la inteligencia artificial, 2022 (edición de junio de 2024). https://home.inai.org.mx/wp-content/documentos/DocumentosSectorPublico/RecomendacionesPDP-IA.pdf
24. Cámara de Diputados, Reglamentos federales vigentes, consultado en oct 2026. https://www.diputados.gob.mx/LeyesBiblio/regla.htm
25. Ley Reglamentaria del Artículo 5o. Constitucional, relativo al ejercicio de las profesiones, artículo 36 (Cámara de Diputados). https://www.diputados.gob.mx/LeyesBiblio/pdf/208_190118.pdf
26. Código Penal Federal, artículos 210 y 211 (Cámara de Diputados). https://www.diputados.gob.mx/LeyesBiblio/pdf/CPF.pdf
27. Barra Mexicana, Colegio de Abogados, Lineamientos para el Uso Responsable de la Inteligencia Artificial en el Ejercicio Profesional del Derecho, 14 oct 2025. https://www.bma.org.mx/wp-content/uploads/2025/10/Lineamientos-de-la-Barra-Mexicana-Colegio-de-Abogados-A.C.-BMA-para-el-Uso-Responsable-de-la-Inteligencia-Artificial-en-el-Ejercicio-Profesional-del-Derecho.pdf

28. American Bar Association, Formal Opinion 512, «Generative Artificial Intelligence Tools», 29 jul 2024. https://www.americanbar.org/content/dam/aba/administrative/professional_responsibility/ethics-opinions/aba-formal-opinion-512.pdf
29. Council of Bars and Law Societies of Europe (CCBE), «CCBE guide on the use of generative AI by lawyers», 2 oct 2025. https://www.advokatforeningen.no/contentassets/185dd73ac1e14031ac016b1567e4b7a6/ccbe-guide-on-the-use-of-the-use-of-generative-ai-for-lawyers.pdf
30. Debevoise & Plimpton, «Update: Judge Rakoff issues written opinion that AI-generated documents are not protected», 18 feb 2026. https://www.debevoise.com/insights/publications/2026/02/update-judge-rakoff-issues-written-opinion-that-ai
31. Tremblay v. OpenAI, Inc., No. 23-cv-03223-AMO (N.D. Cal.), orden de la jueza Araceli Martínez-Olguín, 8 ago 2024. https://www.law.berkeley.edu/wp-content/uploads/archive/2025/12/Tremblay-v-OpenAI-Inc.pdf
32. Concord Music Group, Inc. v. Anthropic PBC, No. 5:24-cv-03811 (N.D. Cal.), orden de la magistrada Susan van Keulen, 23 may 2025. https://app.midpage.ai/document/concord-music-group-inc-v-11059147
33. eDiscovery Today, «Anthropic's requests for production of prompts and outputs granted in part», 9 ene 2026. https://ediscoverytoday.com/2026/01/09/anthropics-requests-for-production-of-prompts-and-outputs-granted-in-part-ediscovery-case-law/
34. Cleary Gottlieb, «Courts grapple with privilege implications of AI», 2026. https://www.clearygottlieb.com/news-and-insights/publication-listing/courts-grapple-with-privilege-implications-of-ai
35. United States v. Kovel, 296 F.2d 918 (2d Cir. 1961). https://law.justia.com/cases/federal/appellate-courts/F2/296/918/131265/
36. Kirkland & Ellis, «A federal court charts a path on AI, protective orders and work product in discovery», 20 may 2026. https://www.kirkland.com/publications/kirkland-alert/2026/05/a-federal-court-charts-a-path-on-ai-protective-orders-and-work-product-in-discovery
37. L. Sweeney, «Simple Demographics Often Identify People Uniquely», Carnegie Mellon University, Data Privacy Working Paper 3, 2000. https://dataprivacylab.org/projects/identifiability/paper1.pdf
38. L. Rocher, J. Hendrickx e Y.-A. de Montjoye, «Estimating the success of re-identifications in incomplete datasets using generative models», Nature Communications 10, 2019. https://research.dial.uclouvain.be/handle/2078.5/254758
39. A. Prince y D. Schwarcz, «Proxy Discrimination in the Age of Artificial Intelligence and Big Data», 105 Iowa Law Review 1257, 2020. https://ilr.law.uiowa.edu/node/7861
40. J. Dastin (Reuters), «Amazon scraps secret AI recruiting tool that showed bias against women», 10 oct 2018 (republicado en HR Reporter). https://www.hrreporter.com/recruitment-and-retention/38235-amazon-scraps-secret-ai-recruiting-tool-that-showed-bias-against-women/
41. K. Wilson y A. Caliskan, «Gender, Race, and Intersectional Bias in Resume Screening via Language Model Retrieval», AIES 2024. https://arxiv.org/abs/2407.20371v2
42. Tan et al., «Small Changes, Big Impact...», EMNLP 2026. https://aclanthology.org/2026.emnlp-main.998/
43. P. James et al., «2014 Evidence-Based Guideline for the Management of High Blood Pressure in Adults» (JNC 8), JAMA 311(5), 2014 (resumen de AAFP). https://www.aafp.org/pubs/afp/issues/2014/0401/p574a.html
44. News-Medical, «ADA lowers BMI cut point for screening Asian Americans for type 2 diabetes», 24 dic 2014. https://www.news-medical.net/news/20141224/ADA-lowers-BMI-cut-point-for-screening-Asian-Americans-for-type-2-diabetes.aspx
45. D. Vyas, L. Eisenstein y D. Jones, «Hidden in Plain Sight — Reconsidering the Use of Race Correction in Clinical Algorithms», NEJM 383, 2020 (ficha de AHRQ PSNet). https://psnet.ahrq.gov/issue/hidden-plain-sight-reconsidering-use-race-correction-clinical-algorithms
46. A. Salinas, A. Haim y J. Nyarko, «What's in a Name? Auditing Large Language Models for Race and Gender Bias», Stanford, arXiv 2402.14875, v3, 2025. https://arxiv.org/html/2402.14875v3
47. M. Andriushchenko y N. Flammarion, «Does Refusal Training in LLMs Generalize to the Past Tense?», arXiv 2407.11969, 2024. https://arxiv.org/abs/2407.11969v2
48. X. Bai, A. Wang, I. Sucholutsky y T. Griffiths, «Measuring Implicit Bias in Explicitly Unbiased Large Language Models», arXiv 2402.04105. https://arxiv.org/pdf/2402.04105v2
49. A. Tamkin et al. (Anthropic), «Evaluating and Mitigating Discrimination in Language Model Decisions», arXiv 2312.03689, 2023. https://ar5iv.labs.arxiv.org/html/2312.03689
50. ALEA Institute (antes 273 Ventures), evaluación de toxicidad y sesgo de KL3M. https://github.com/alea-institute/kl3m-toxicity
51. Ma, Salinas, Henderson y Nyarko, «Breaking Down Bias: On The Limits of Generalizable Pruning Strategies», arXiv 2502.07771, 2025. https://arxiv.org/abs/2502.07771
52. J. Larson, S. Mattu, L. Kirchner y J. Angwin, «How We Analyzed the COMPAS Recidivism Algorithm», ProPublica, 23 may 2016. https://www.courts.wa.gov/subsite/mjc/docs/2019/how-we-analyzed-the-compas-r.pdf
53. State v. Loomis, 881 N.W.2d 749 (Wis. 2016), comentario en Harvard Law Review, vol. 130. https://harvardlawreview.org/print/vol-130/state-v-loomis/
54. J. Kleinberg, S. Mullainathan y M. Raghavan, «Inherent Trade-Offs in the Fair Determination of Risk Scores», arXiv 1609.05807, 2016. https://arxiv.org/abs/1609.05807

55. A. Chouldechova, «Fair prediction with disparate impact», arXiv 1703.00056, 2017. https://arxiv.org/abs/1703.00056
56. A. Koenecke et al., «Racial disparities in automated speech recognition», PNAS 117(14), 2020. https://ideas.repec.org/a/nas/journl/v117y2020p7684-7689.html
57. Bloomberg, «Humans Are Biased. Generative AI Is Even Worse», 2023 (ficha en AI Incident Database). https://incidentdatabase.ai/reports/3180
58. G. Crosilla, L. Klic y G. Colavizza, «Benchmarking Large Language Models for Handwritten Text Recognition», arXiv 2503.15195, 2025. https://arxiv.org/pdf/2503.15195v2
59. TJUE, asunto C-203/22, Dun & Bradstreet Austria, 27 feb 2025 (resumen de CMS). https://cms.law/en/fra/news-information/cjeu-landmark-ruling-on-the-right-of-access-and-automated-decision-making
60. Stanford CRFM, Foundation Model Transparency Index, dic 2025. https://crfm.stanford.edu/fmti/December-2025
61. D. Amodei, «The Urgency of Interpretability», abr 2025. https://www.darioamodei.com/post/the-urgency-of-interpretability
62. M. Turpin, J. Michael, E. Perez y S. Bowman, «Language Models Don't Always Say What They Think», NeurIPS 2023. https://arxiv.org/abs/2305.04388v2
63. Y. Chen et al. (Anthropic), «Reasoning Models Don't Always Say What They Think», arXiv 2505.05410, 2025. https://arxiv.org/html/2505.05410v1
64. J. Mökander, J. Schuett, H. Kirk y L. Floridi, «Auditing large language models: a three-layered approach», AI and Ethics, 2023. https://arxiv.org/abs/2302.08500v2
65. Grupo de Expertos de Alto Nivel sobre IA de la Comisión Europea, Ethics Guidelines for Trustworthy AI, 2019 (resumen de Kramer Levin). https://kramerlevin.com/en/perspectives-search/european-commission-issues-ethics-guidelines-for-trustworthy-ai.html
66. T. Dratsch et al., «Automation Bias in Mammography», Radiology, 2023 (comunicado de RSNA). https://www.rsna.org/media/press/i/2432
67. Reglamento (UE) 2024/1689 (Ley de IA), artículo 14, «Supervisión humana». https://euaiact.com/article/14
68. B. Green, «The Flaws of Policies Requiring Human Oversight of Government Algorithms», Computer Law & Security Review, 2022. https://arxiv.org/abs/2109.05067v3
69. V. Magesh et al. (Stanford), «Hallucination-Free? Assessing the Reliability of Leading AI Legal Research Tools», Journal of Empirical Legal Studies 22(2), 2025. https://arxiv.org/abs/2405.20362v1
70. D. Charlotin, AI Hallucination Cases Database, consultada el 29 oct 2026. https://www.damiencharlotin.com/hallucinations/
71. METR, «Task-Completion Time Horizons of Frontier AI Models», actualizado al 8 sep 2026. https://metr.org/time-horizons/
72. Google, «Generative AI in Google Workspace Privacy Hub», 26 oct 2026. https://support.google.com/a/answer/15706919?hl=en

Fuentes de las matrices y la checklist

Documentos oficiales de los proveedores, consultados el 30 oct 2026.

- OpenAI: Enterprise privacy. https://openai.com/enterprise-privacy/
- OpenAI: OpenAI Services Agreement. https://openai.com/policies/services-agreement/
- OpenAI: OpenAI Data Processing Addendum. https://openai.com/policies/data-processing-addendum/
- OpenAI: API: Your data. https://platform.openai.com/docs/guides/your-data
- OpenAI: Data controls in ChatGPT. https://help.openai.com/en/articles/7730893-data-controls-faq
- OpenAI: Data residency for ChatGPT. https://help.openai.com/en/articles/9903489-data-residency-and-inference-residency-for-chatgpt
- OpenAI: Respuesta a la orden del caso NYT. https://openai.com/index/response-to-nyt-data-demands/
- OpenAI: Security and privacy. https://openai.com/security-and-privacy/
- Anthropic: Commercial Terms of Service. https://www.anthropic.com/legal/commercial-terms
- Anthropic: Data Processing Addendum. https://www.anthropic.com/legal/data-processing-addendum
- Anthropic: How long do you store my organization's data?. https://privacy.anthropic.com/en/articles/7996866-how-long-do-you-store-my-organization-s-data
- Anthropic: What certifications has Anthropic obtained?. https://privacy.anthropic.com/en/articles/10015870-what-certifications-has-anthropic-obtained
- Anthropic: Planes de Claude. https://claude.com/pricing
- Anthropic: Updates to our consumer terms. https://www.anthropic.com/news/updates-to-our-consumer-terms
- Anthropic: Privacy Policy. https://www.anthropic.com/legal/privacy
- Microsoft: Data, privacy and security for Microsoft 365 Copilot. https://learn.microsoft.com/en-us/copilot/microsoft-365/microsoft-365-copilot-privacy
- Microsoft: Azure OpenAI y Foundry: data, privacy and security. https://learn.microsoft.com/en-us/legal/cognitive-services/openai/data-privacy
- Microsoft: Customer Copyright Commitment. https://learn.microsoft.com/en-us/azure/foundry/responsible-ai/openai/customer-copyright-commitment
- Microsoft: Copilot Copyright Commitment. https://blogs.microsoft.com/on-the-issues/2023/09/07/copilot-copyright-commitment-ai-legal-concerns/
- Google: Gemini Privacy Hub para Workspace. https://support.google.com/a/answer/15706919?hl=en
- Google: Vertex AI: data governance. https://cloud.google.com/vertex-ai/generative-ai/docs/data-governance
- Google: Generative AI Indemnified Services. https://cloud.google.com/terms/generative-ai-indemnified-services
- Google: Google Workspace Service Specific Terms. https://workspace.google.com/terms/service-terms/