



NARSIL

INTELLIGENCE

ESTUDIOS NARSIL · 05

Una IA que lee tus documentos antes de contestar

Cómo funciona el RAG, la técnica que hace que un asistente responda con la información de tu empresa y cite de dónde la saca, y qué hace falta para que funcione de verdad

OCTUBRE DE 2026

Resumen

1. **Un modelo de lenguaje no conoce tu empresa.** Sabe mucho del mundo, pero nada de tus contratos, manuales o procedimientos, y cuando no sabe algo tiende a rellenar el hueco con algo verosímil.
2. **Lo que inventa un asistente lo responde la empresa.** Un tribunal canadiense condenó en 2024 a Air Canada por la información falsa que dio su chatbot.
3. **El RAG busca antes de responder.** El sistema localiza en tus documentos los fragmentos relevantes y el modelo contesta a partir de ellos, citando la fuente.
4. **Reduce los errores, pero no los elimina.** Stanford midió en 2024 entre un 17 % y un 33 % de consultas con información inventada en herramientas jurídicas comerciales basadas en RAG.
5. **Los permisos son parte del diseño.** Si el buscador no los respeta, la IA enseña a cada usuario lo que no debería ver.
6. **Se puede montar entero en local.** Documentos, índice, modelo y registro de consultas pueden quedarse en la infraestructura propia.
7. **Se empieza en pequeño y midiendo.** Un conjunto de documentos acotado, limpio y con un juego de preguntas de referencia vale más que todo el archivo de golpe.

01

Un modelo que no conoce tu empresa

Los grandes modelos de lenguaje han aprendido de una cantidad enorme de texto público. Por eso saben redactar, resumir y razonar sobre casi cualquier tema. Pero no han leído nunca los contratos de tu empresa, su manual de calidad, sus actas o el historial de incidencias de sus clientes. Para todo lo que es propio de una organización, un modelo recién instalado parte de cero.

El problema no es solo que no sepa. Es lo que hace cuando no sabe. Un modelo de lenguaje genera el texto más probable a partir de lo que le piden, y lo más probable no siempre es lo cierto. Ante una pregunta sobre la política interna de devoluciones, puede construir una respuesta impecable en la forma y equivocada en el fondo, con el mismo tono de seguridad que cuando acierta. Es lo que se conoce como **alucinación**, y es el principal obstáculo para usar la IA con información que importa.

02

Cuando inventa, paga la empresa

En noviembre de 2022 un cliente de Air Canada consultó al asistente de la web de la aerolínea si podía acogerse a la tarifa reducida por fallecimiento de un familiar. El asistente le dijo que podía comprar el billete y pedir la reducción después, en los 90 días siguientes. No era cierto: la política de la compañía, publicada en otra página de la misma web, exigía solicitarla antes de viajar.

Cuando la aerolínea le negó el reembolso, el cliente reclamó. En febrero de 2024 el Tribunal de Resolución Civil de la Columbia Británica le dio la razón (*Moffatt contra Air Canada*, 2024 BCCRT 149). La compañía llegó a argumentar que el chatbot era una entidad responsable de sus propios actos; el tribunal rechazó la idea y concluyó que Air Canada responde de toda la información de su web, salga de una página fija o del chatbot. La indemnización fue pequeña. El precedente, no.

La lección para cualquier organización es directa: **un asistente de IA de cara al público habla en nombre de la empresa**, y lo que diga tiene que estar respaldado por la información oficial de la empresa.

03

Qué es el RAG

La técnica que más se usa para resolver este problema se llama **generación aumentada por recuperación**, conocida por sus siglas en inglés, RAG. La propuso en 2020 un equipo de investigadores de Facebook AI encabezado por Patrick Lewis, en un trabajo presentado en la conferencia NeurIPS, con una motivación que sigue siendo la central: que las respuestas de un modelo tuvieran una procedencia que se pudiera inspeccionar y que el conocimiento se pudiera actualizar sin reentrenarlo.

La idea es sencilla. En lugar de pedirle al modelo que responda de memoria, el sistema primero **busca** en una colección de documentos los fragmentos relacionados con la pregunta y después le pide al modelo que **responda a partir de esos fragmentos**. El modelo deja de ser la fuente del conocimiento y pasa a ser quien lo lee y lo explica.

UNA COMPARACIÓN ÚTIL

Es la diferencia entre preguntarle algo a un experto que contesta de memoria y pedirle que antes consulte el expediente. El segundo tarda un poco más, pero puede decir en qué página lo ha visto.

04

Cómo funciona por dentro

Preparar los documentos

El primer paso ocurre antes de que nadie pregunte nada. Los documentos de la organización se convierten a texto, se dividen en **fragmentos** de un tamaño manejable (unos párrafos, una cláusula, una sección del manual) y se guardan en un índice junto con sus datos: de qué documento salen, de qué fecha son y quién tiene permiso para verlos.

Para que la búsqueda funcione por significado y no solo por palabras exactas, cada fragmento se transforma en una representación numérica que captura su contenido. Así, una pregunta sobre «baja por paternidad» puede encontrar un fragmento que hable de «permiso por nacimiento». En la práctica, los buenos sistemas combinan esta búsqueda semántica con la búsqueda clásica por palabras, que sigue siendo insustituible para nombres propios, referencias y códigos.

Buscar lo relevante

Cuando llega una pregunta, el sistema busca en el índice los fragmentos más relacionados, suele reordenarlos con un segundo filtro más fino y se queda con unos pocos. Esa selección es crítica: si el fragmento correcto no está entre los elegidos, el modelo no podrá usarlo, por bueno que sea.

Responder citando

Por último, el modelo recibe la pregunta junto con los fragmentos seleccionados y la instrucción de responder solo con esa información, indicando de dónde sale cada afirmación. Un sistema bien diseñado tiene que saber decir «no consta en la documentación» cuando la respuesta no está, en lugar de completar el hueco por su cuenta.

05

Lo que el RAG no resuelve

El RAG mejora mucho la fiabilidad, pero no la garantiza. El estudio más citado al respecto lo publicó en 2024 el laboratorio RegLab de la Universidad de Stanford. Sus autores sometieron a varias herramientas comerciales de investigación jurídica basadas en RAG a 202 consultas revisadas por juristas, y encontraron que **entre el 17 % y el 33 % de las respuestas contenían información inventada**, pese a que alguno de esos productos se anunciaba como libre de alucinaciones.

Los fallos tienen causas conocidas. A veces la búsqueda trae el fragmento equivocado, o uno parecido pero de otro contrato. A veces los documentos están desactualizados, duplicados o se contradicen entre sí, y el sistema no tiene forma de saber cuál vale. Hay preguntas que no se responden con un

fragmento sino razonando sobre todo el archivo («¿cuántos contratos vencen este año?»), y para las que la búsqueda de fragmentos no es la herramienta adecuada. Y el modelo, incluso con el texto correcto delante, puede interpretarlo mal o añadir algo de su cosecha.

La conclusión no es renunciar al RAG, sino diseñarlo sabiendo dónde falla: con citas que el usuario pueda comprobar con un clic, con documentación ordenada y con una evaluación continua.

06

Permisos y seguridad

Un buscador corporativo que ignora los permisos es una fuga de información esperando a ocurrir. Si todos los documentos de la empresa entran en el mismo índice sin control, cualquier usuario podrá preguntar por las nóminas de dirección, el expediente de un empleado o el contrato confidencial de un cliente, y el asistente se lo resumirá con amabilidad. **Los permisos del documento original tienen que acompañar a cada fragmento**, y la búsqueda solo debe devolver lo que la persona que pregunta podría abrir por sí misma.

Hay un segundo riesgo menos evidente. Un documento puede contener instrucciones escondidas dirigidas a la IA, en texto blanco sobre blanco o en una nota al margen, y el modelo puede obedecerlas al leerlo. Es la inyección de instrucciones, que analizamos en el estudio 02 de esta serie. En un RAG, todo documento que llega de fuera debe tratarse como no fiable, y el asistente no debería poder hacer nada más que responder.

07

Cómo se sabe si funciona

La forma seria de evaluar un RAG es la misma que se usaría con una persona nueva en el puesto: preguntarle cosas cuya respuesta ya se conoce. Antes de poner el sistema en manos de nadie, conviene preparar un **juego de preguntas de referencia** (cincuenta es un buen comienzo) redactadas por quienes conocen la documentación, con la respuesta correcta y el documento donde aparece.

Con ese juego se miden tres cosas: si la respuesta es correcta, si la cita la respalda de verdad y si el sistema sabe decir que no lo sabe cuando la información no está. Ese conjunto se vuelve a pasar cada vez que cambian los documentos, el modelo o la configuración, y se amplía con las preguntas reales que fallen en el uso diario.

08

En la nube o en casa

Un RAG tiene tres piezas que manejan información sensible: los documentos, el índice y el modelo que lee los fragmentos. Si el modelo está en la nube, cada pregunta viaja con los fragmentos de tus documentos a los servidores del proveedor, con las condiciones que analizamos en el estudio 01.

La alternativa es montarlo entero en la infraestructura propia: un modelo abierto ejecutado en local, un índice en servidores de la organización, los permisos heredados del sistema documental y un registro de quién pregunta qué. Con el hardware actual es viable para la mayoría de las organizaciones, como explicamos en el estudio 03. Para documentación interna sin datos personales, un proveedor en la nube con plan de empresa puede ser suficiente; para expedientes, contratos o datos de clientes, lo prudente es que nada salga de casa.

09

Dónde aporta más

El RAG da su mejor resultado allí donde hay mucha documentación, preguntas repetidas y un coste alto por equivocarse. Un despacho puede consultar sus propios dictámenes y contratos tipo. Una empresa industrial, sus manuales técnicos y el historial de averías. Un departamento de recursos humanos, el convenio y las políticas internas. Un servicio de atención al cliente, su base de conocimiento, con la garantía de que el asistente no promete lo que la empresa no ofrece. En sanidad, protocolos y guías clínicas; las historias de pacientes, solo con todas las garantías y en local.

10

Por dónde empezar

1. **Elegir un conjunto acotado de documentos** con valor claro y un grupo de usuarios definido: un manual, un archivo de contratos, una base de conocimiento.
2. **Ordenarlo antes de indexarlo:** fuera duplicados, borradores y versiones caducadas. Un RAG nunca será mejor que la documentación que lee.
3. **Definir los permisos** desde el primer día, heredados del sistema donde viven los documentos.
4. **Escribir el juego de preguntas de referencia** y medir aciertos, citas y silencios antes de abrirlo.
5. **Decidir dónde se ejecuta** según la sensibilidad de la información, y ampliar por áreas.

Un asistente que responde con la información de la empresa, cita de dónde la saca y reconoce cuándo no la tiene es probablemente el uso de la IA con más retorno para la mayoría de las organizaciones. No por la tecnología, que ya está madura, sino por el orden documental y el criterio con que se implanta.

Anexo

Fuentes

Páginas consultadas el 2 de octubre de 2026, salvo las noticias, fechadas en su publicación.

1. Lewis, P. et al. *Retrieval-Augmented Generation for Knowledge-Intensive NLP Tasks*. NeurIPS 2020.
arxiv.org/abs/2005.11401
2. Magesh, V. et al. *Hallucination-Free? Assessing the Reliability of Leading AI Legal Research Tools*. Stanford RegLab, 2024.
arxiv.org/abs/2405.20362
3. *Moffatt v. Air Canada*, 2024 BCCRT 149. Civil Resolution Tribunal, Columbia Británica, febrero de 2024. Resúmenes en americanbar.org y mccarthy.ca.
4. OWASP. *LLM01:2025 Prompt Injection*. genai.owasp.org/llmrisk/llm01-prompt-injection
5. NARSIL Intelligence. *Estudios 01, 02 y 03*. narsilintelligence.com/estudios



NARSIL INTELLIGENCE

Diseñamos, construimos e implantamos sistemas de inteligencia artificial en la infraestructura de cada organización, y formamos a los equipos que los usan.

NARSILINTELLIGENCE.COM