

Una base de conocimiento para tu empresa

Cómo convertir documentos dispersos en un segundo cerebro al que puedes preguntar

Presentación conceptual · Vector Brain

Contenido

- 1. Introducción: el problema
- 2. La idea: un segundo cerebro para tu empresa
- 3. Qué hay dentro de una base de conocimiento
- 4. Cómo llegas a la base de conocimiento
- 5. Lo que el agente de IA hace por ti
- 6. Protección de datos y confianza: la respuesta honesta
- 7. ¿Tiene futuro la base de conocimiento?
- **Anexo: el equipo de escaneo, o cómo nace la base de conocimiento**

1 · Introducción: el problema

Antes de ver cómo funciona una base de conocimiento (knowledge base), una pregunta previa: ¿por qué merece la pena?

El conocimiento está ahí, pero nunca a mano

Da igual si es una gestoría, una correduría de seguros o una oficina técnica: el conocimiento de una empresa está en sus documentos. En correos, ofertas, contratos, actas, notas y papeles escaneados. Todo está ahí. Pero cuando lo *necesitas*, casi nunca está donde acabas de mirar.

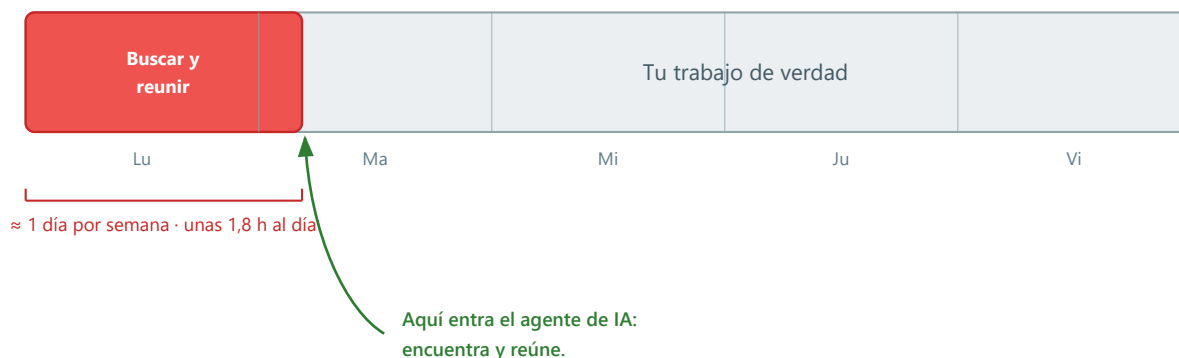
Un caso típico: llega una pregunta de una cliente, de la administración o de un compañero. La respuesta existe, en algún sitio. Y «en algún sitio» quiere decir: en un correo de hace tres meses, en una carpeta del servidor, en un expediente antiguo o en la cabeza de una compañera que hoy no está. Así que toca buscar y juntar piezas.

Lo que cuesta

Buscar y juntar información no es un asunto menor. Se come una parte sorprendente de la jornada. Un estudio muy citado del McKinsey Global Institute lo pone en cifras: quienes trabajan con información dedican de media **1,8 horas al día** a buscarla y reunirla. Es **casi una cuarta parte de la jornada**. Dicho de otra forma: **casi un día entero de trabajo, cada semana**.

Tu semana laboral

Cinco días de trabajo y adónde se va buena parte



Casi una cuarta parte, semana tras semana

Y no pasa una vez, sino **cada semana**: tiempo que no va al trabajo en sí, sino a volver a encontrar y juntar algo que hace tiempo que está en la empresa.

Las carpetas tienen un límite

Casi todas las empresas tienen un orden: números, carpetas, subcarpetas, y cada cual sabe dónde va cada cosa. Y funciona: conoces el expediente, lo abres y ahí están los documentos.

Pero una estructura de carpetas solo responde a una pregunta: ¿dónde está? Qué dice el contenido lo tiene que leer y unir alguien en su cabeza. Además, ordena por expedientes: cada documento está en una sola carpeta. En cuanto la pregunta va más allá, no hay carpeta que valga: ¿ya tuvimos un caso así? ¿Dónde más nos pasó esto? ¿Qué respondimos entonces?

Y la carpeta bien ordenada es el caso bueno. El correo de hace tres meses, la nota suelta, lo que sabe la compañera que hoy no está: todo eso se queda fuera.

El esfuerzo de verdad casi nunca es leer un documento, sino **reunir información de varias fuentes**: juntar el correo adecuado y el contrato y el acta y la nota de entonces en una sola imagen. Ese paso es el que más tiempo se lleva, y con carpetas solas no desaparece.

Justo ahí brilla un agente de IA

Y aquí está la clave: buscar y reunir es justo lo que un agente de IA hace **rápido y sin esfuerzo**, en dos pasos:

- **Encuentra rápido.** En lugar de ir clic a clic por carpetas y buzones, haces tu pregunta con tus palabras y en segundos tienes los pasajes que importan, de todos los documentos.
- **Reúne de muchas fuentes.** El agente de IA junta lo relevante de distintos documentos y te da **un resultado terminado**: un resumen, un borrador, una respuesta a un correo. No diez hallazgos que tienes que ordenar tú, sino algo completo.

Así, lo que hoy se come casi una cuarta parte del tiempo pasa a ser cosa de segundos. Por eso merece la pena una base de conocimiento.

Lo que hace falta: la base de conocimiento

Un agente de IA solo puede reunir lo que conoce, y por sí solo no conoce tu empresa. Necesita un sitio donde tenga ese conocimiento a mano. Cómo encaja todo lo cuenta la siguiente parte: **La idea: un segundo cerebro para tu empresa.**

Fuentes

- **McKinsey Global Institute:** *The social economy: Unlocking value and productivity through social technologies*. Julio de 2012. — Resultado principal: quienes trabajan con información dedican de media **1,8 h/día (≈ 9,3 h/semana)** a buscar y reunir información; según el informe, «nearly 20 percent» de la semana laboral corresponde solo a la búsqueda de información interna (más un 28 % al correo electrónico).
- **Para situar la cifra:** otros estudios confirman el rango «de casi una cuarta parte a algo menos de un tercio»: Interact/APQC ≈ 19,8 % del tiempo de trabajo; IDC ≈ 30 %. La formulación del texto, «casi una cuarta parte / casi un día por semana», es por tanto prudente.

2 · La idea: un segundo cerebro para tu empresa

El problema está claro. Antes de hablar de tecnología, esta es la idea a la que apunta todo.

Pensar y recordar

Un cerebro hace dos cosas: piensa y recuerda.

Pensar, los agentes de IA de hoy lo hacen bien. Redactan, comparan, resumen, sacan conclusiones. Lo que les falta es recordar. Un agente de IA empieza cada conversación de cero. No conoce tus expedientes, ni tus clientes, ni cómo se trabaja en tu empresa. Es listo, pero no sabe nada de tu caso.

Esa otra mitad la pone una base de conocimiento (knowledge base). Es la memoria: ahí está el conocimiento de tu empresa, preparado para que un agente de IA pueda leerlo. Juntos forman un segundo cerebro para tu empresa. El agente de IA pone el pensar; tu base de conocimiento, el recordar.

Una buena memoria está ordenada, se mantiene al día y, ante una pregunta, saca los pocos pasajes que de verdad vienen al caso.

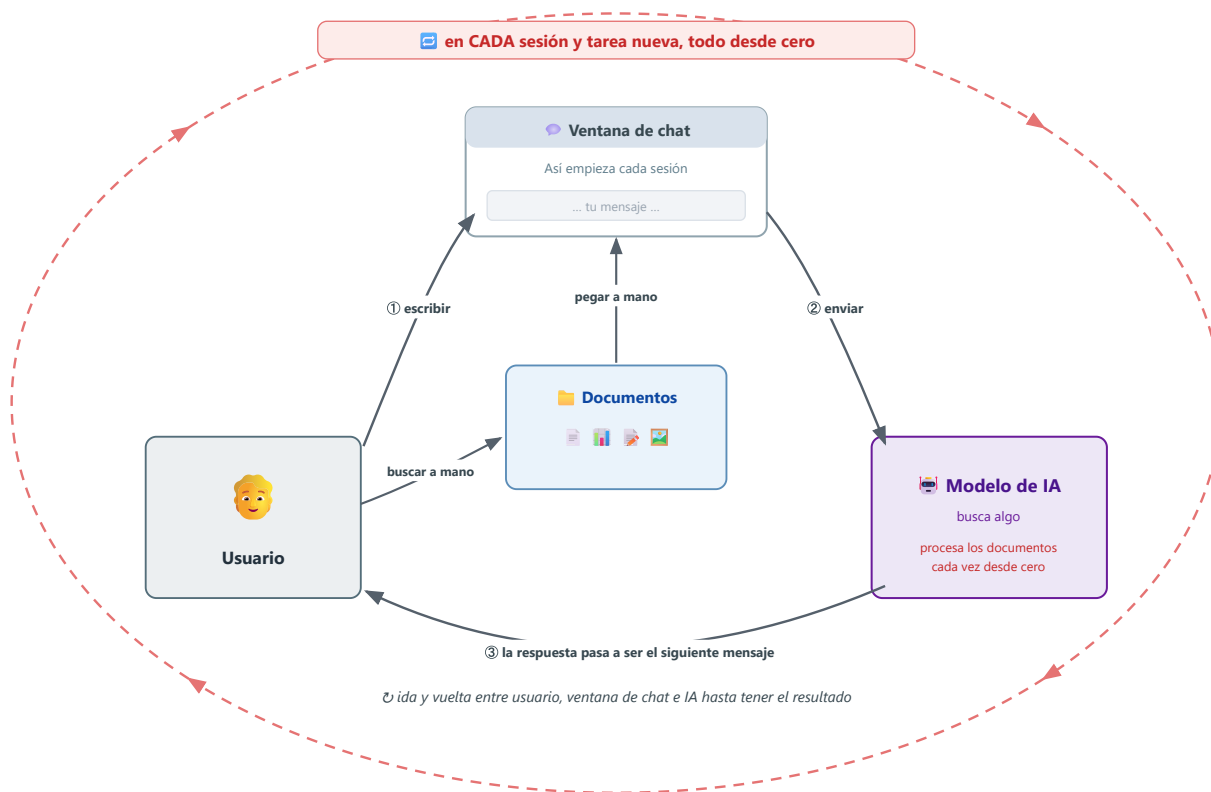
Así se trabaja con un chatbot de IA

Si ya usas IA, conoces el patrón. Abres una ventana de chat y primero buscas los documentos que hacen falta: el contrato, el correo antiguo, el acta. Lo pegas todo a mano. Y solo entonces empieza el trabajo de verdad.

Es exactamente la búsqueda del capítulo anterior. El chatbot no te la ahorra: la da por hecha.

Y hay algo más que muchos notan tarde: cierras la ventana y todo desaparece. Con el siguiente expediente vuelves a empezar de cero y a juntar los mismos documentos.

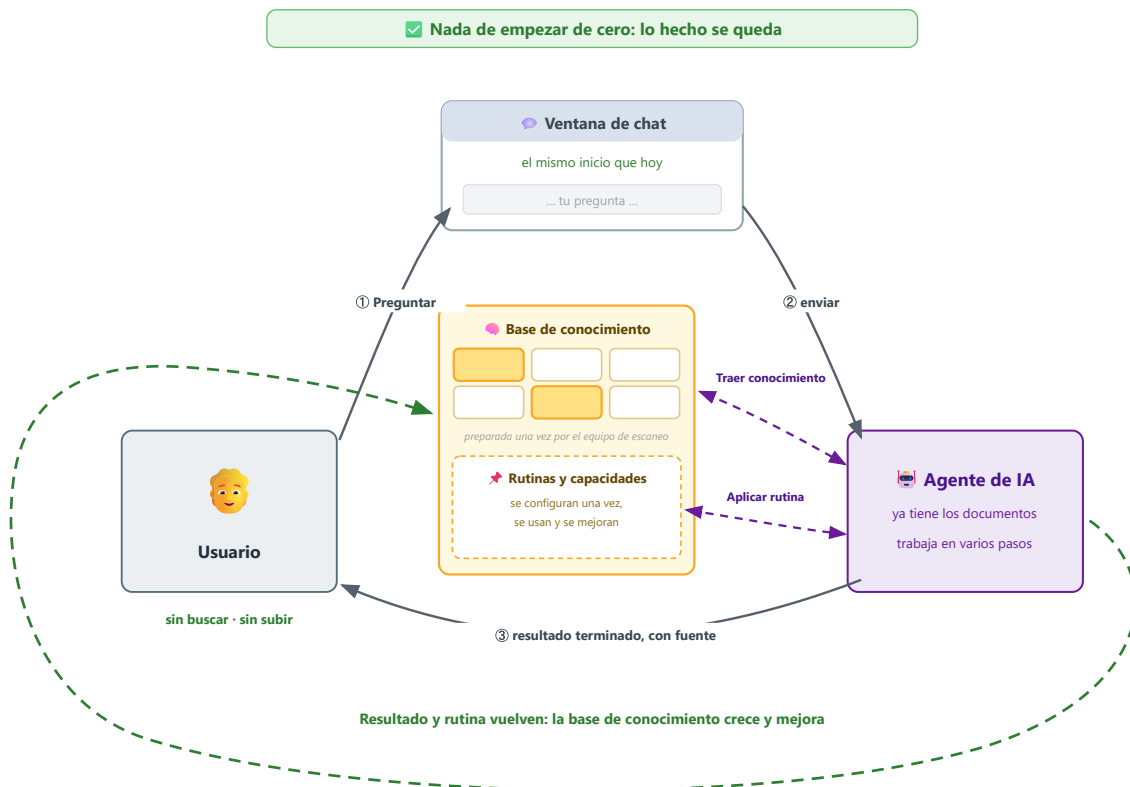
Cómo se trabaja hoy con IA: el mismo ciclo en cada sesión



Así se trabaja con un agente de IA

Con una base de conocimiento: cada sesión se apoya en la anterior

La otra cara de «Cómo se trabaja hoy con IA»: mismo inicio, otro mundo



A primera vista, todo se parece a lo de hoy. Haces tu pregunta y recibes una respuesta. La diferencia está en lo que desaparece por el camino.

Ya no buscas ni subes nada. El agente de IA va él mismo a la base de conocimiento y toma los pasajes que encajan con tu pregunta.

Y cuando terminas un expediente, el trabajo no se pierde. El resultado vuelve a la base de conocimiento y la próxima vez ya está ahí.

Así desaparecen los dos costes del primer capítulo. La búsqueda, porque los documentos ya están disponibles. El eterno empezar de cero, porque el conocimiento se queda. Lo que construyes una vez nunca se pierde.

Instrucciones, capacidades, rutinas

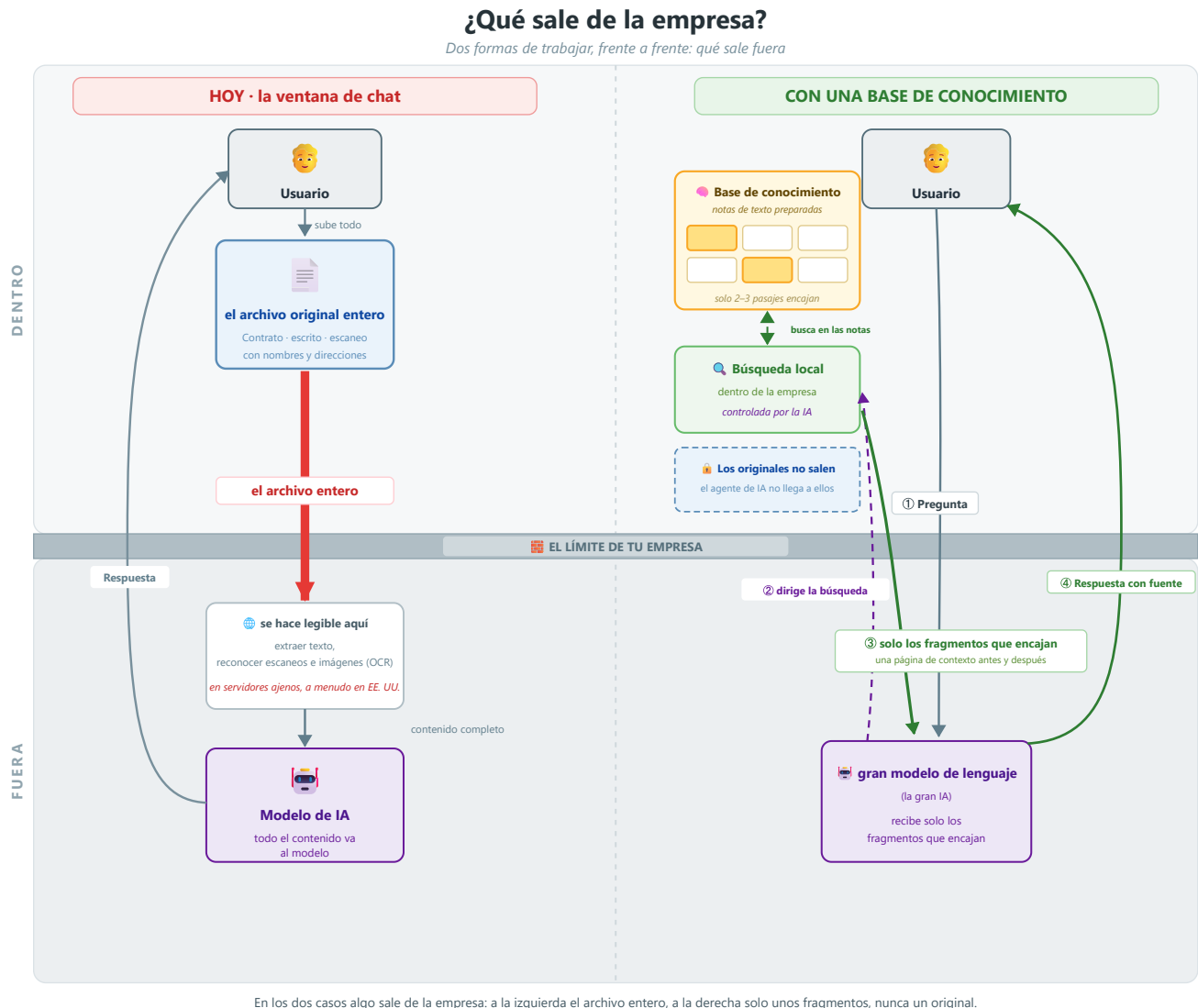
Con un agente de IA pasa como con alguien nuevo en la empresa: primero hay que explicarle cómo se hacen las cosas.

Una vez que dejas claro cómo se hace un trámite en tu empresa, queda registrado. En el siguiente caso lo vuelve a aplicar, con otro cliente o en otro proyecto.

Con el uso, los procesos se van afinando y las instrucciones mejoradas vuelven a la base de conocimiento, hasta que el agente hace la tarea de forma fiable y como tú quieres.

Qué datos salen de la empresa

Las dos formas de trabajar también se diferencian en lo que de verdad sale de tu empresa.



Hoy, con la ventana de chat, subes el archivo entero: el contrato, la carta, el acta, completos y con todo lo que llevan dentro. El archivo va al proveedor, muchas veces en Estados Unidos, y solo allí se convierte en texto legible: se extrae el texto y se reconocen imágenes y escaneos, todo en sus servidores. El contenido completo acaba en el modelo.

Con una base de conocimiento es al revés. El agente de IA lee tu pregunta, lanza una búsqueda y esa búsqueda se ejecuta dentro de tu empresa, sobre las notas de texto preparadas de tu archivo. No sobre los originales: a esos ni siquiera llega. De cada resultado se lleva solo lo que hay alrededor, más o menos una página antes y otra después. Solo esos fragmentos salen fuera, al gran modelo de lenguaje, que con ellos redacta la respuesta, con la fuente.

O sea, también aquí sale algo de la empresa (eso es parte de la respuesta honesta), pero son unos pocos párrafos en lugar de documentos enteros. Qué significa esto para la protección de datos lo explica más adelante un capítulo propio.

3 · Qué hay dentro de una base de conocimiento

La idea está clara. Falta ver qué contiene una base de conocimiento y en qué se convierten tus documentos.

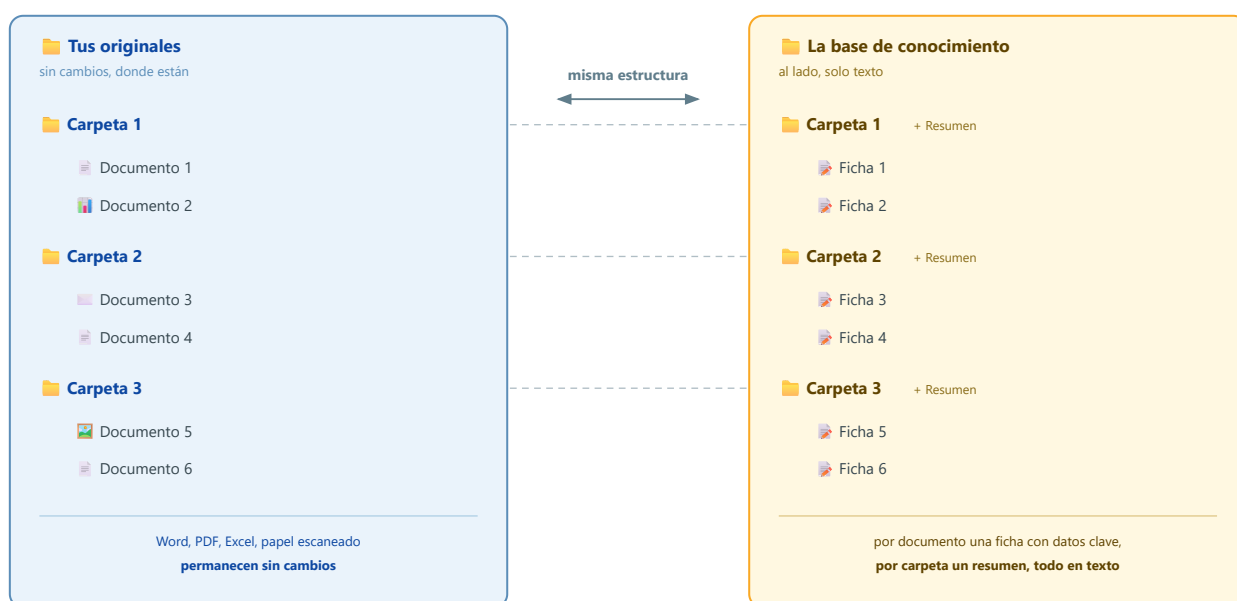
El mismo orden que ya tienes

La base de conocimiento no es un sistema extraño en el que tu equipo tenga que aprender a moverse. Sigue el orden que ya tienes en la empresa: las mismas carpetas, la misma estructura, el mismo orden que en tu servidor.

Cómo es ese orden lo decide tu empresa. En una son clientes de una gestoría, en otra clientes a secas, en otra expedientes, casos, asuntos o proyectos. La base de conocimiento no lo discute: refleja lo que hay.

La estructura de carpetas se mantiene

A la izquierda tus originales, a la derecha la base de conocimiento: misma estructura, otro contenido



Da igual si tus carpetas se llaman clientes, expedientes, casos, asuntos o proyectos: la base de conocimiento adopta la estructura que ya tienes.

Si alguien se maneja en el servidor, se maneja en la base de conocimiento. No hay que reordenar, renombrar ni mover nada.

Qué guarda de cada expediente

De cada expediente hay un resumen: de qué va, quién participa y en qué punto está. Cada documento tiene además una ficha breve: qué tipo de documento es, qué datos clave contiene (nombres, fechas, importes o plazos, por ejemplo) y dónde está el original.

Todo es texto sencillo que puedes abrir y leer tú mismo cuando quieras. Nada está en un formato cerrado.

Cómo es una de estas entradas en detalle lo verás en el anexo «El equipo de escaneo».

Tus originales se quedan como están

La base de conocimiento se crea **al lado**, como un fondo aparte. En tus originales no se toca nada, y el agente de IA no llega a ellos ni para leer ni para escribir.

Lo que se crea no es una segunda copia de tus archivos, sino una descripción resumida de ellos. No es ahorrar donde no toca, es justo la idea: si le das todo a la vez a un agente de IA, las respuestas empeoran, no mejoran. Demasiado material lo confunde, igual que a una persona. Lo que necesita son los pocos pasajes que encajan con la pregunta.

Cuánto cuesta el cambio

No hay traslado ni migración. Tu archivo se queda intacto donde está, y la base de conocimiento crece al lado.

Y crece sola. Como tu equipo sigue guardando los documentos como siempre, la base de conocimiento no es un proyecto que alguien tenga que mantener aparte: avanza con el trabajo normal.

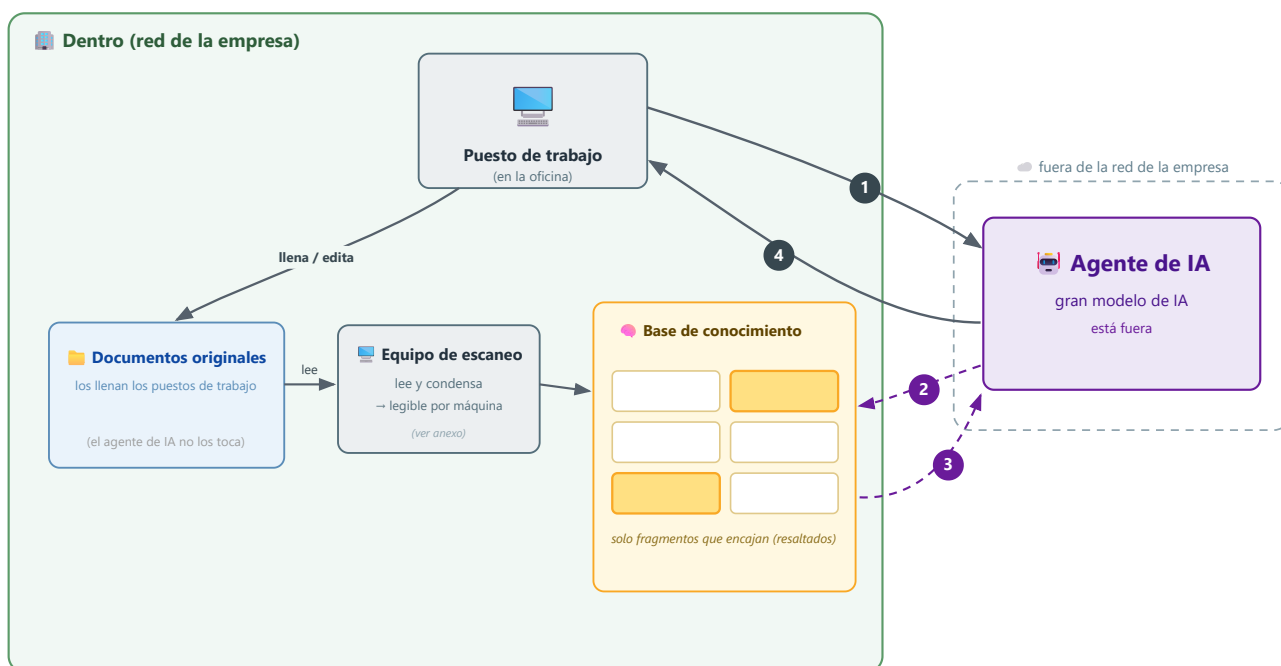
Además, casi toda esta preparación es automatización programada: recorrer archivos, convertir en texto los papeles escaneados, detectar duplicados, mirar dentro de los archivos comprimidos. La IA de verdad solo entra donde hay que entender y resumir contenido. Si te interesa el reparto en detalle, lo tienes en el anexo «El equipo de escaneo».

4 · Cómo llegas a la base de conocimiento

La base de conocimiento está lista. Queda la pregunta práctica: ¿cómo llegas a ella, en la oficina y fuera?

Hablar con el agente de IA desde la oficina

Dentro de la empresa: crear y preguntar



Leyenda: una pregunta al agente de IA, en cuatro pasos

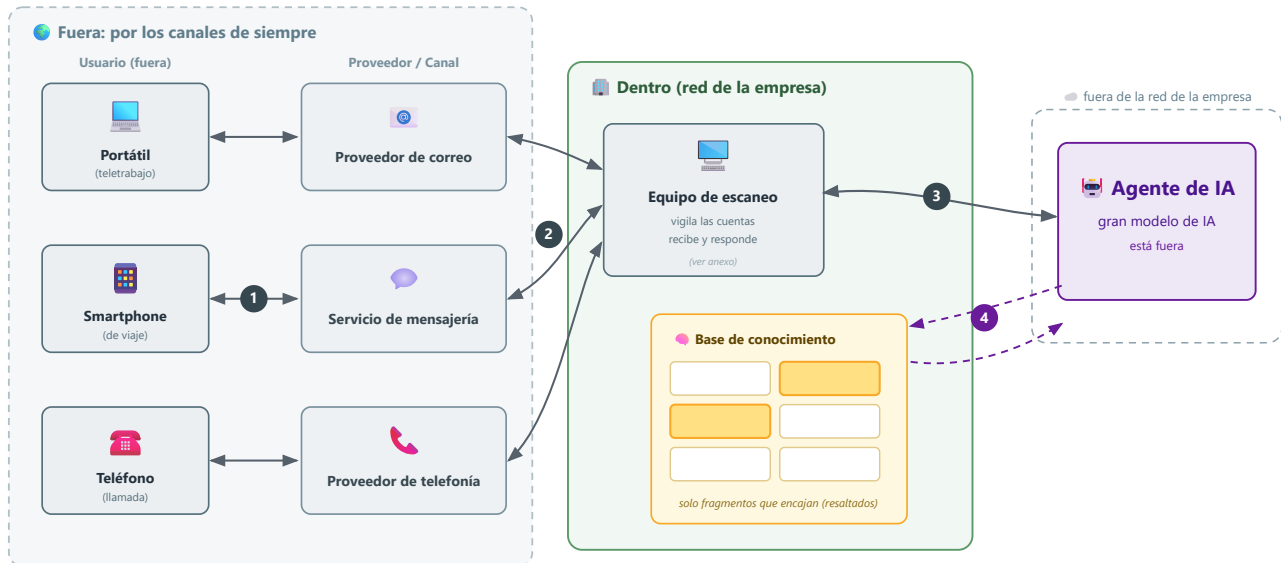
- 1 El puesto de trabajo hace una pregunta al agente de IA.
- 2 Solo entonces entra un momento en la base de conocimiento, **sin acceso permanente**.
- 3 Vuelven **solo los fragmentos que encajan**, no todo el archivo.
- 4 Con ellos, el agente de IA redacta la respuesta, con la fuente.

En la oficina, casi nada cambia: escribes tu pregunta en la ventana de chat y te llega la respuesta. Te ayuda una función de navegación ya preparada que encuentra los datos rápido y sin fallos. Basta con escribir, por ejemplo, «expediente 123» o «cliente Heinz Meier», y el agente de IA sabe dónde mirar. Tu equipo no tiene que aprender nada más.

El agente de IA **no tiene acceso permanente** a tu archivo. Solo entra cuando alguien pregunta, toma los fragmentos que encajan y redacta con ellos la respuesta. Después, el acceso se cierra.

Hablar con el agente de IA desde fuera

Fuera de la oficina: por los canales de siempre



Leyenda: el camino de una pregunta desde fuera (los canales van en los dos sentidos)

- 1 Preguntas por el **canal de siempre** (correo, mensajería, teléfono): **sin VPN y sin iniciar sesión**.
- 2 El **equipo de escaneo** de la oficina vigila esas cuentas, recibe la pregunta y devuelve la respuesta **por la misma vía**.
- 3 Le pide al **agente de IA** que redacte la respuesta.
- 4 Este toma **solo los fragmentos que encajan** de la base de conocimiento, **sin acceso permanente**.

Desde fuera solo entra un canal controlado: la red de la empresa sigue cerrada.

Desde fuera no necesitas VPN ni iniciar sesión. Escribes un correo, mandas un mensaje o llamas por teléfono, igual que harías con un compañero.

Dentro de la empresa, el equipo de escaneo vigila esas cuentas. Recibe tu pregunta, le pide la respuesta al agente de IA y te la devuelve por el mismo canal. Para ti es una respuesta normal a un mensaje normal. Todo lo demás que hace este equipo está en el anexo.

La seguridad viene de regalo

Lo habitual para dar acceso desde fuera es una VPN. En el peor de los casos, eso abre media red de la empresa a un dispositivo que anda por ahí.

Aquí es al revés. Hacia fuera solo hay un canal controlado, y la red de la empresa sigue cerrada. El acceso es más cómodo y, a la vez, hay menos por donde atacar. No pasa a menudo, y es uno de los efectos más agradables de construirlo así.

Dentro y fuera, el mismo patrón

En la oficina o fuera, el agente de IA nunca está conectado de forma fija a tu archivo. Solo entra para cada pregunta concreta y después vuelve a quedarse fuera.

5 · Lo que el agente de IA hace por ti

La base de conocimiento está lista y el acceso, resuelto. Queda la pregunta que de verdad importa: ¿qué ganas tú en el día a día?

Preguntar o encargar: la diferencia

Lo que casi todo el mundo conoce de la IA es un chatbot: preguntas, responde, sigues con lo tuyo. Es un interlocutor que ha leído mucho, pero no hace nada.

Un agente de IA funciona de otra manera. Hace solo tareas de varios pasos: lee, busca, reúne, redacta y archiva. No le consultas, le encargas.

Eso es posible gracias a la base de conocimiento. Un chatbot conoce internet, pero no tu empresa. Un agente de IA que trabaja sobre tu propio archivo conoce tus expedientes, tus clientes y todo lo que ha pasado antes.

Chatbot o agente de IA: la diferencia

A la izquierda, lo que casi todos conocen. A la derecha, lo que es posible con tu propia base de conocimiento.



La diferencia no está en el modelo, sino en a qué puede acceder.

Lo que la IA sabe hacer

Casi toda la utilidad está en cosas nada espectaculares que hoy te quitan tiempo:

Encontrar. Preguntas con tus palabras en lugar de ir clic a clic por carpetas y buzones. ¿En qué expediente salió este tema? ¿Qué acordamos entonces?

Resumir. Un contrato largo, un informe, toda una documentación: reducido a lo esencial y con la extensión que necesites.

Comparar. Dos versiones de un documento, una al lado de la otra, con las diferencias explicadas en claro. ¿Qué ha cambiado entre la versión antigua y la nueva?

Reunir. Aquí es donde hoy se pierde más tiempo: juntar las muchas piezas de un expediente en una imagen completa. El agente de IA hace justo eso y te da un resultado terminado, no diez hallazgos que tienes que ordenar tú.

Y siempre igual: la respuesta llega **con la fuente**. Ves de qué documentos sale y puedes consultarlos cuando quieras. Esa es la diferencia entre una herramienta de la que te puedes fiar y una que podría estar adivinando.

Resultados terminados, no conversaciones largas

Además de responder, el agente de IA te prepara cosas que puedes usar directamente:

Redacta **borradores** de cartas, ofertas o informes, y en el estilo que ya usa tu empresa, porque lo conoce de los documentos anteriores.

Rellena **formularios y plantillas que se repiten**, porque los datos ya están en la base de conocimiento.

Y te entrega resultados **estructurados** cuando los necesitas: un listado de varios expedientes en forma de tabla, con el que puedes seguir calculando al momento.

Rutinas con horario

Hasta aquí siempre preguntabas tú. Lo más interesante empieza cuando ya no hace falta preguntar.

Defines una vez lo que te interesa: ciertos expedientes, ciertos indicadores, temas pendientes. A partir de ahí, el resumen te llega solo, por ejemplo cada mañana: qué está resuelto, qué está pendiente, qué ha cambiado desde ayer y dónde hay atascos.

Lo mismo con **plazos y seguimientos**. El agente de IA los tiene presentes y avisa a tiempo, en lugar de esperar a que alguien se acuerde de mirarlos.

Así, perseguir a la gente por teléfono se convierte en un aviso que llega solo. Ese es el gran salto: un chatbot espera preguntas; un agente trabaja en segundo plano y te trae lo importante sin que se lo pidas.

Hace el trabajo pesado, no la decisión

Todo lo anterior es buscar, copiar, reunir y preparar borradores. Justo el trabajo que se come buena parte del día.

Lo que el agente de IA no hace es decidir por ti. Te deja un borrador, no una decisión tomada. Te dice lo que pone en los documentos y te da la fuente. Qué hacer con eso lo sigues decidiendo tú con tu equipo, solo que con el trabajo previo hecho.

¿Y los datos?

Para hacer ese trabajo, el agente de IA tiene que poder leer los fragmentos que importan de tu base de conocimiento. Qué significa esto para la protección de datos, y por qué la respuesta honesta no es «todo se queda en la empresa», lo explica el siguiente capítulo.

6 · Protección de datos y confianza: la respuesta honesta

Aquí muchas ofertas ponen una frase tranquilizadora que no es verdad. Nosotros lo hacemos de otra forma.

Primero, la respuesta honesta

Muchos proveedores te dirán que tus datos nunca salen de tu empresa. Para lo que aquí se describe, eso sería mentira.

La razón es sencilla. Las tareas exigentes (entender, resumir, redactar) hoy solo las hace bien un gran modelo de IA. Un equipo capaz de hacerlo todo en tu propio sótano costaría más que el resto del proyecto, y aun así sería más flojo. Para la preparación en segundo plano, tu propia infraestructura sobra; para pensar, no.

Así que sí, salen datos. La pregunta no es si salen, sino **cuántos** y **a quién**.

Qué sale de verdad de la empresa

Aquí se ve la diferencia. No sale tu archivo, y tampoco salen tus documentos.

Salen los **fragmentos que encajan con una pregunta concreta**. Solo esos, y solo en el momento en que alguien pregunta. No hay ninguna conexión permanente por la que se vaya tu archivo, ni ningún proveedor tiene una copia de tus documentos.

Tus **documentos originales** quedan fuera desde el principio. El agente de IA no llega a ellos: trabaja solo con la base de conocimiento.

Comparado con lo habitual hoy, es una reducción enorme: en lugar de documentos enteros salen párrafos sueltos, y solo los que hacen falta.

Adónde va y qué puedes elegir

Queda la cuestión de quién lo recibe. Hay tres caminos viables, y puedes elegir el que mejor encaje con tu empresa.

Un proveedor europeo. En Europa hay proveedores de IA potentes sujetos al **Reglamento General de Protección de Datos (RGPD)**; **Mistral AI**, de Francia, es el ejemplo más conocido. Es el camino más sencillo, porque así la exportación de datos ni se plantea.

Un gran proveedor que se compromete por contrato a no entrenar con tus datos. También con los grandes proveedores estadounidenses puedes dejar fijado que tus contenidos **no se usen para entrenar los modelos**. En las cuentas para empresas es lo normal, no la excepción. Tu consulta se procesa en el centro de datos, la respuesta vuelve, y el contenido ni entra en el modelo ni se guarda de forma permanente. Si quieres descartar del todo el almacenamiento, en las cuentas para empresas puedes acordar además que, más allá del procesamiento inmediato, **no se guarde nada**.

Otra cosa es la exportación de datos a Estados Unidos en sí, un tema distinto que no conviene mezclar con el del entrenamiento. Hoy también se puede reducir o evitar del todo: los grandes proveedores ya ofrecen **procesamiento dentro de la UE**, unas veces con almacenamiento y procesamiento en Europa, otras porque sus modelos funcionan en los centros de datos europeos de las grandes plataformas en la nube. Así el procesamiento se queda en Europa; y donde sí van datos a Estados Unidos, hay bases contractuales establecidas para ello.

Un modelo abierto con el operador que tú elijas. Hoy existen modelos de libre uso cuyo funcionamiento está a la vista. No funcionan en manos de quien los desarrolló, sino donde tú decidas, por ejemplo en un

centro de datos europeo. Quien desarrolló el modelo no ve nada de tus datos, porque no participa en el procesamiento.

La objeción que antes se oía ha perdido casi todo su peso. Estos modelos abiertos han mejorado tanto en los últimos años que, para las tareas de las que hablamos aquí, la distancia con los modelos punteros de los grandes proveedores apenas importa. Ganar control ya no significa pagar con resultados claramente peores. Así, una duda abierta sobre protección de datos se convierte en un asunto que se resuelve como cualquier otro contrato con un proveedor. Con cualquiera de los tres caminos se puede trabajar **conforme al RGPD**, siempre que uses la cuenta para empresas con los contratos adecuados y no el servicio gratuito para particulares.

¿Y tus secretos empresariales?

La segunda gran objeción suele ser: «pero así sale fuera lo que sabemos como empresa».

Veamos de qué se trata en realidad. En la inmensa mayoría de los casos son **procesos normales**: cómo se calcula una oferta, cómo se tramita una reclamación, qué documentos van en un expediente, cómo se organiza una carpeta. Esos procesos existen en Alemania, en Estados Unidos, en China, en Argentina y en el Congo. Están descritos millones de veces y los modelos los conocen hace tiempo, de todo el mundo. Eso es oficio, no secreto.

Claro que también hay secretos empresariales de verdad: un resultado de investigación, un procedimiento, una fórmula, un diseño que nadie más tiene. Pero eso afecta a muy pocas empresas, y donde existe, es una parte pequeña y fácil de delimitar de los documentos.

Se revisa una vez si hay zonas así y, si las hay, se dejan fuera. El resto, que es con diferencia la mayor parte de tu archivo, no supone ningún problema.

La comparación que casi nadie hace

Y ahora lo que suele faltar en esta discusión: ¿cómo están las cosas hoy en tu empresa?

En muchas empresas, la gente lleva tiempo subiendo documentos a mano a servicios de IA gratuitos, porque les facilita el trabajo. Contratos enteros, cartas enteras, muchas veces con nombres y direcciones. Sin contrato, sin control y sin que nadie lo sepa.

Desde el punto de vista de la protección de datos, eso normalmente no está permitido. Pero nadie se da cuenta, porque no se ve.

Así que no se trata de elegir entre «cumple la protección de datos» y «no la cumple». Se trata de elegir entre una **situación sin control que nadie conoce del todo** y un **camino regulado**, en el que está claro qué sale, adónde va y con qué contrato.

Verificable, y tuya

Hay dos cosas más, que tienen menos que ver con la ley y más con la confianza.

El agente de IA responde **con la fuente**. Ves de qué documentos sale cada afirmación y puedes consultarlos. No tienes que creerle: puedes comprobarlo.

Y la base de conocimiento es **texto sencillo en tus propios sistemas**. Ningún formato que solo pueda leer un proveedor concreto. Si cambias de proveedor o terminas el proyecto, el archivo preparado se queda contigo. El modelo lo alquilas; la base de conocimiento es tuya.

Algo obvio, para terminar

Cómo se configura todo esto en cada caso depende de tu sector y de tus datos. Una asesoría fiscal tiene otras necesidades que una oficina técnica, y datos especiales como los de salud requieren un cuidado aparte. Esta página no sustituye la revisión de tu responsable de protección de datos: muestra que el tema tiene solución y cómo.

7 · ¿Tiene futuro la base de conocimiento?

Una reflexión para terminar.

Quizá pienses: «Es una tecnología muy nueva y va tan rápido... quién sabe qué habrá el año que viene».

Es verdad, y la objeción tiene sentido. Pero no afecta a la parte de la que hablamos aquí. Lo que cambia rápido son los modelos de IA: cada pocos meses sale uno mejor, y cuál uses mañana es cuestión de semanas, no de años. El conocimiento ordenado de tu empresa no cambia por eso. Un expediente preparado hoy seguirá describiendo lo mismo dentro de cinco años.

Y una cosa no quita la otra; al contrario, una es la condición de la otra. Los agentes de IA de los próximos años, sean los que sean, solo podrán trabajar con lo que tu empresa les ponga delante en forma de conocimiento preparado y rutinas asentadas. Una base de conocimiento ordenada no es, por tanto, una idea de IA más, sino la base sobre la que se apoyan las demás. Sin ella, hasta la mejor herramienta se queda sin el contexto que necesita para trabajar.

Que la tecnología ya está lista hoy no admite dudas. Incluso modelos más pequeños y de libre uso son ya tan capaces que hacen buena parte del trabajo diario de forma fiable y rápida. Lo que hoy falta casi nunca es el modelo: es la base ordenada sobre la que pueda trabajar en tu empresa.

Y hay una segunda reflexión que, de entrada, no tiene nada que ver con la inteligencia artificial. En muchas empresas, lo más valioso del conocimiento está en unas pocas cabezas: en las personas que llevan años y saben por qué algo se hace así y no de otra manera. Si una de ellas se jubila o cambia de trabajo, se lleva un trozo de la empresa. Una base de conocimiento guarda lo que, si no, solo se transmite de palabra, como memoria ordenada de tu empresa que sobrevive a los cambios de personal. Así que vale la pena incluso si nunca escribes una línea con una IA.

Anexo: el equipo de escaneo, o cómo nace la base de conocimiento

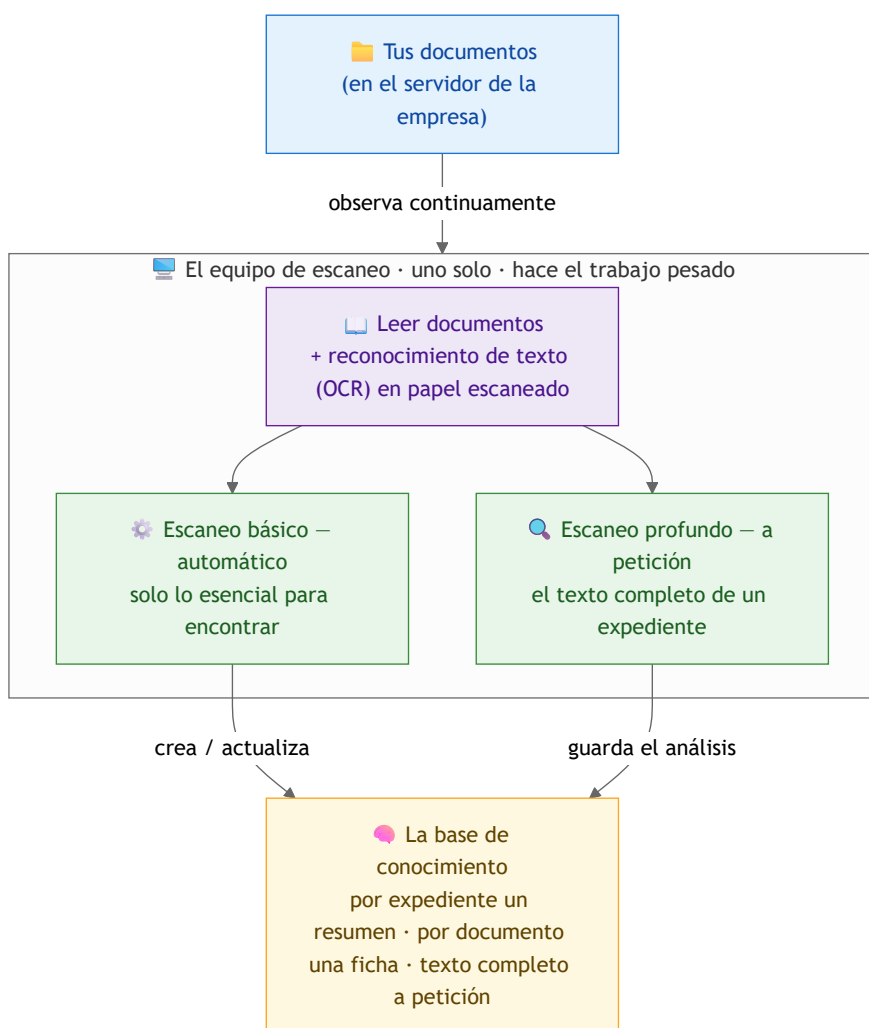
La parte principal muestra qué hace una base de conocimiento. Este anexo muestra cómo se crea. Es para quien quiera más detalle; para entender la parte principal no hace falta.

El motor en segundo plano

Para convertir miles de documentos dispersos en una base de conocimiento en la que se pueda buscar, primero alguien tiene que leerlos, clasificarlos y resumirlos. De eso se encarga un único equipo, instalado solo para esto: el **equipo de escaneo**.

Es el motor de todo y, a la vez, lo primero que se instala. En cuanto este equipo está en marcha, la base de conocimiento se llena sola.

El proceso de un vistazo



Cómo leer el gráfico

Arriba están tus documentos en el servidor. El equipo de escaneo, en el centro, los **vigila** de forma continua, los **lee** (los papeles escaneados, con reconocimiento de texto) y trabaja en **dos modos**: el escaneo básico automático y el escaneo profundo a petición. Los dos llenan el mismo destino: la **base de conocimiento**, abajo.

Por qué un solo equipo

Las herramientas pesadas están **solo en este equipo**: abrir documentos de cualquier formato, convertir en texto los papeles escaneados, resumir contenidos. Eso tiene ventajas claras:

- **Los demás puestos de trabajo siguen ligeros.** Tu equipo no necesita ningún software especial.
- **Mantenimiento en un solo sitio.** Actualizaciones, paquetes de idioma para el reconocimiento de texto, ajustes: todo en un único lugar.
- **Nada se hace dos veces.** Como solo escanea un equipo, el sistema no se estorba a sí mismo; cada expediente se procesa una sola vez y bien.

Hay una condición: este equipo tiene que estar encendido para que se escanee. Si está apagado, el trabajo espera a que vuelva a arrancar. No se pierde nada.

Qué hace el programa y dónde trabaja de verdad la IA

Aquí merece la pena mirar con detalle, porque hoy casi todo se vende como «IA». En realidad, la mayor parte de este trabajo es **lógica de programa**, que hace siempre lo mismo y de forma comprobable:

- Recorrer carpetas y detectar qué es nuevo o ha cambiado
- Abrir formatos de archivo y extraer el texto (PDF, Word, Excel, correo electrónico, ZIP)
- Lanzar el reconocimiento de texto cuando una página es solo una imagen
- Detectar duplicados comparando el **contenido** de los archivos, no el nombre
- Mirar dentro de los archivos comprimidos y comprobar si su contenido ya está descomprimido
- Copiar la estructura de carpetas existente y guardar los resultados

Eso es contabilidad, no inteligencia. Funciona de forma fiable, casi no cuesta nada y, con los mismos datos de entrada, da siempre el mismo resultado.

La **IA de verdad** solo entra donde hay que **entender** algo:

- ¿De qué va realmente este documento?
- Escribir un resumen breve
- Extraer los datos clave: participantes, fechas, importes, plazos
- Montar, a partir de los documentos sueltos, una visión completa del expediente

Este reparto es intencionado. Mantiene los costes bajos, porque el modelo de lenguaje, que es caro, solo se usa en la pequeña parte donde de verdad ayuda. Y hace el resultado más fácil de comprobar: donde ha trabajado un programa, puedes consultar la regla.

El reconocimiento de texto tampoco es magia

El reconocimiento de texto (OCR) es un procedimiento entrenado, sí, pero muy especializado: convierte píxeles en letras y no entiende nada del contenido. Es parte de la caja de herramientas, no de la parte que piensa.

Dos modos de trabajo

El escaneo básico: automático. El equipo revisa por su cuenta, cada cierto tiempo, si hay algo nuevo o cambiado en las carpetas: durante el día, un vistazo rápido más a menudo; por la noche, una pasada completa. Recoge **a propósito solo lo esencial para encontrar las cosas**: qué documentos hay, de qué tratan más o menos y los datos clave. No lee cada documento de principio a fin. Para encontrar y clasificar, basta con eso.

El escaneo profundo: a petición. Si alguien necesita el **texto completo** de un expediente, pide un escaneo profundo. Entonces el equipo de escaneo lee todas las páginas enteras y guarda el resultado como un análisis consultable junto al documento original. El escaneo profundo **complementa** al básico, no lo sustituye.

Cuándo hace falta de verdad. Para el día a día basta el escaneo básico: dónde está el expediente, de qué iba, quién participó, en qué punto está. Un escaneo profundo merece la pena cuando importa el texto exacto: por ejemplo, cuando hay que revisar una cláusula concreta de un contrato de sesenta páginas, cuando se analiza un informe largo o cuando un expediente pasa a otra persona que tiene que entenderlo entero. Por eso se pide para expedientes concretos, en lugar de escanear todo el archivo completo por si acaso.

Cómo es una entrada

Todo lo que genera el equipo de escaneo es texto sencillo, con dos tipos de entradas.

Por cada documento, una ficha breve. Recoge en pocas líneas qué tipo de documento es, de qué trata, qué datos clave contiene y dónde está el original. Por ejemplo:

Tipo: carta, 3 páginas, escaneada (léida con reconocimiento de texto)

De qué trata: consulta sobre una partida de la oferta del mes pasado; piden una aclaración antes de fin de mes.

Datos clave: participantes · fecha · partida afectada · plazo mencionado

Original: nombre del archivo y ubicación

Nota: está también en una segunda carpeta (duplicado, no se registra otra vez)

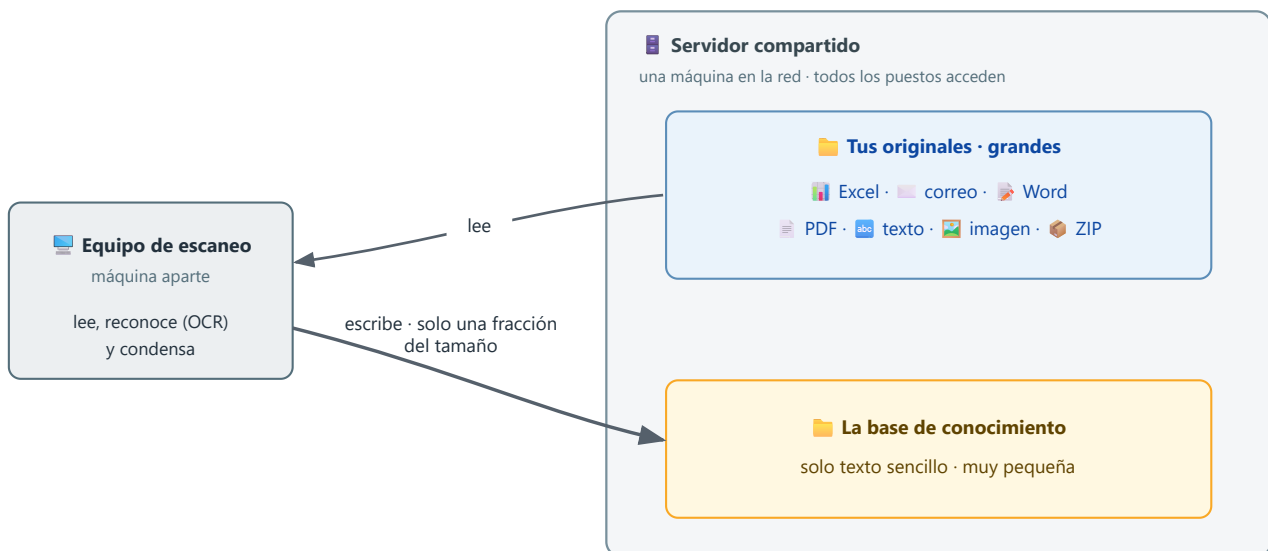
Por cada expediente, un resumen. Recoge de qué trata el expediente en conjunto, qué documentos incluye, cómo se ordenan en el tiempo y en qué punto está.

Las dos cosas las puede leer una persona y comprobarlas de un vistazo. Precisamente por eso la base de conocimiento conserva su valor aunque cambie la herramienta que hay detrás.

Misma estructura, una fracción del tamaño

La base de conocimiento **copia tu estructura de carpetas tal cual**: las mismas carpetas, los mismos expedientes, en el mismo orden. La diferencia está en el contenido: donde en el original hay un escaneo pesado de un contrato o una carpeta entera de fotos, al lado solo hay una nota de texto breve.

Las dos cosas están en el mismo sitio. Los documentos originales **y** la base de conocimiento están en un **servidor compartido en la red**, al que acceden todos los puestos de trabajo. El equipo de escaneo es una máquina aparte: lee los documentos de ese servidor y guarda la base de conocimiento allí mismo, al lado.



De archivos grandes, texto breve

La base de conocimiento ocupa solo una fracción de los originales: no contiene imágenes, escaneos ni maquetación, solo su contenido en texto. Por eso se consulta en segundos y se hace copia de seguridad sin esfuerzo, mientras los originales grandes siguen intactos donde están.

Nada dos veces: duplicados y archivos comprimidos

Los archivos que han ido creciendo con los años casi nunca están ordenados. El mismo documento suele estar varias veces, y los archivos comprimidos se quedan ahí después de descomprimirlos.

Duplicados: cada documento, una sola vez. El equipo de escaneo detecta las repeticiones comparando el **contenido** de los archivos, no solo el nombre, y registra cada documento una sola **vez**. Las demás copias se anotan, pero no se crea ninguna entrada duplicada.

Archivos ZIP: mirar dentro en lugar de contar doble. Un caso frecuente: alguien guarda un archivo comprimido, lo descomprime y lo deja ahí. El equipo de escaneo **mira dentro**: si el contenido ya está descomprimido, no lo registra otra vez. Solo se añade lo que está únicamente dentro del archivo comprimido.

Límites a propósito

No todo tiene sentido como texto, y es intencionado:

- **Los documentos puramente gráficos**, como planos grandes, dibujos o mapas, no se pasan a texto completo; apenas tienen texto útil. Se anotan, y los detalles se consultan en el original.
- **Las tablas de datos muy grandes** (largas series de mediciones, registros o exportaciones) no se transcriben, sino que se anotan como «datos masivos». Es mejor analizarlas directamente en la fuente.

Esta contención es una señal de calidad. La base de conocimiento sigue ligera y rápida, en lugar de llenarse de lastre.

Lo que el equipo de escaneo no hace

El equipo de escaneo **solo lee**. Nunca modifica, mueve ni borra un documento original, y tampoco podría: no tiene permiso de escritura sobre los originales. Se quedan intactos en su sitio de siempre; al lado solo se

crean notas y referencias.

Y el escaneo en sí ocurre **entero dentro de la empresa**, en tus propios equipos. Mientras se preparan los datos, tus documentos no salen de la empresa.

No confundir con el uso diario

Esto vale para **el escaneo**. Cuando después alguien trabaja con el agente de IA, los fragmentos que encajan con la pregunta sí salen fuera, a un gran modelo de lenguaje. Qué sale exactamente, adónde y cómo se puede regular está en la parte principal, en el capítulo «Protección de datos y confianza».