

Die Wissensdatenbank für Ihre Firma

Wie aus verstreuten Unterlagen ein durchsuchbares zweites Gehirn wird

Konzept-Präsentation · Vector Brain

Inhalt

- 1. Einleitung — Das Problem
- 2. Die Idee — ein zweites Gehirn für Ihre Firma
- 3. Wie eine Wissensdatenbank aufgebaut ist
- 4. Kommunikationswege mit der Wissensdatenbank
- 5. Was der KI-Agent für Sie tun kann
- 6. Datenschutz und Vertrauen — die ehrliche Antwort
- 7. Wissensdatenbank zukunftssicher?
- **Anhang: Der Scan-Rechner — wie die Wissensdatenbank entsteht**

1 · Einleitung — Das Problem

Bevor wir zeigen, wie eine Wissensdatenbank (Knowledge Base) funktioniert, die Frage davor: Warum lohnt sich der Aufwand überhaupt?

Das Wissen ist da – nur nicht griffbereit

Ob Kanzlei, Versicherungsbüro oder Planungsbüro: Das Wissen eines Betriebs steckt in seinen Unterlagen — in E-Mails, Angeboten, Verträgen, Protokollen, Notizen, eingescanntem Papier. Es ist alles vorhanden. Nur: Wenn man es *braucht*, ist es selten genau dort, wo man gerade nachschaut.

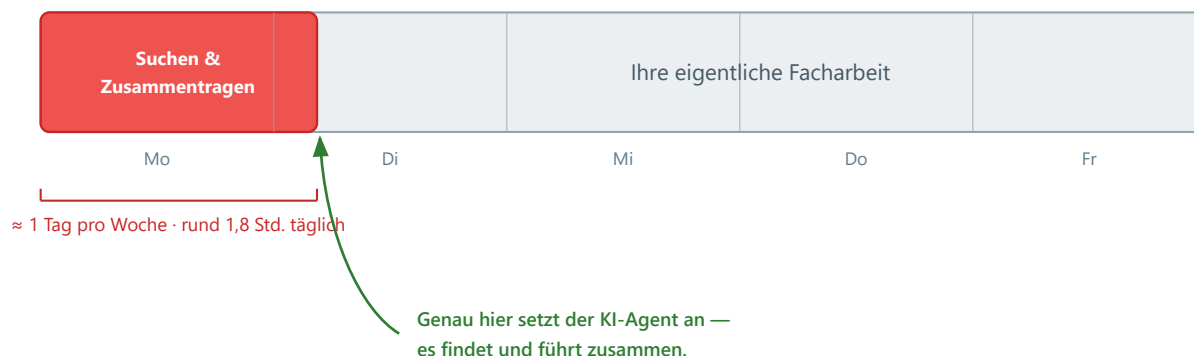
Ein typischer Moment: Eine Frage kommt herein — von einer Kundin, einer Behörde, einem Kollegen. Die Antwort steht irgendwo. Aber „irgendwo“ heißt in der Praxis: in einer E-Mail von vor drei Monaten, in einem Ordner auf dem Server, in einem alten Vorgang, vielleicht im Kopf einer Kollegin, die gerade nicht da ist. Also fängt man an zu suchen — und zusammenzutragen.

Was das kostet

Dieses Suchen und Zusammentragen ist kein Randthema. Es frisst einen erstaunlich großen Teil des Arbeitstags. Eine viel zitierte Untersuchung des McKinsey Global Institute beziffert es so: Wissensarbeiter verbringen im Schnitt **1,8 Stunden pro Tag** damit, Informationen zu suchen und zusammenzutragen. Das sind **fast ein Viertel der Arbeitszeit** — oder anschaulicher: **rund ein ganzer Arbeitstag, jede Woche.**

Ihre Arbeitswoche

Fünf Arbeitstage — und wohin ein großer Teil der Zeit fließt



Fast ein Viertel — Woche für Woche

Das passiert nicht einmalig, sondern **jede Woche aufs Neue**: Zeit, die nicht in die eigentliche Facharbeit fließt, sondern ins Wiederfinden und Zusammensuchen von etwas, das längst im Haus vorhanden ist.

Ordnerstruktur und ihre Grenzen

Nun hat fast jeder Betrieb eine Ordnung. Nummern, Hauptordner, Unterordner, und jeder weiß, wo was hingehört. Das ist auch kein Selbstbetrug, das funktioniert: Man kennt den Vorgang, geht hinein, findet die Unterlagen.

Nur beantwortet eine Ordnerstruktur immer nur die eine Frage: Wo liegt es? Was darin steht, muss weiterhin jemand lesen und im Kopf zusammensetzen. Und sie ordnet nach Vorgängen — ein Dokument liegt in genau einem Ordner. Sobald eine Frage quer dazu läuft, gibt es keinen Ordner, in den man dafür gehen könnte: Hatten wir so einen Fall schon einmal? Wo ist uns das sonst noch begegnet? Was haben wir damals geantwortet?

Dazu kommt: Der saubere Ordner ist der gute Fall. Die E-Mail von vor drei Monaten, die Notiz, das Wissen der Kollegin, die gerade nicht da ist — das liegt daneben.

Der eigentliche Aufwand ist dabei oft gar nicht das Lesen eines einzelnen Dokuments, sondern das **Zusammentragen aus mehreren Quellen**: die passende E-Mail *und* den Vertrag *und* das Protokoll *und* die Notiz von damals zu einem stimmigen Bild zusammenzusetzen. Genau dieser Schritt kostet die meiste Zeit — und mit Ordnerstruktur allein lässt er sich nicht wegräumen.

Genau hier ist ein KI-Agent stark

Und damit sind wir beim Kern der Sache: Suchen und Zusammentragen ist genau das, was ein KI-Agent **schnell und mühelos** erledigt — in zwei Schritten:

- **Schnell finden.** Statt sich durch Ordner und Postfächer zu klicken, stellt man seine Frage in normaler Sprache — und bekommt in Sekunden die passenden Stellen, quer über alle Unterlagen hinweg.
- **Aus vielen Quellen zusammenführen.** Ein KI-Agent zieht die relevanten Informationen aus verschiedenen Dokumenten zusammen und formt daraus **ein fertiges Ergebnis** — eine Übersicht, einen Entwurf, eine E-Mail-Antwort. Nicht zehn Fundstellen zum Selber-Sortieren, sondern ein zusammenhängendes Ganzes.

Aus fast einem Viertel Arbeitszeit fürs Suchen und Zusammentragen wird so eine Sache von Sekunden. Das ist der Grund, warum sich eine Wissensdatenbank lohnt.

Voraussetzung: die Wissensdatenbank

Ein KI-Agent kann nur zusammenführen, was er auch kennt — und von sich aus kennt er Ihre Firma nicht. Er braucht eine Stelle, an der Ihr Wissen für ihn abrufbar liegt. Wie dieses Zusammenspiel aussieht, zeigt der nächste Teil: **Die Idee — ein zweites Gehirn für Ihre Firma.**

Quellen

- **McKinsey Global Institute:** *The social economy: Unlocking value and productivity through social technologies*. Juli 2012. — Kernbefund: Wissensarbeiter verbringen im Schnitt **1,8 Std./Tag (≈ 9,3 Std./Woche)** mit dem Suchen und Zusammentragen von Informationen; laut Report entfallen „nearly 20 percent“ der Arbeitswoche allein auf das Suchen nach internen Informationen (plus 28 % auf E-Mail).
- **Einordnung der Größenordnung:** Korrespondierende Studien stützen den Bereich „fast ein Viertel bis knapp ein Drittel“ — Interact/APQC ≈ 19,8 % der Arbeitszeit; IDC ≈ 30 %. Die im Text verwendete Formulierung „fast ein Viertel / rund ein Tag pro Woche“ ist damit belastbar konservativ gewählt.

2 · Die Idee — ein zweites Gehirn für Ihre Firma

Das Problem ist benannt. Bevor es um Technik geht, hier das Bild, auf das alles zuläuft.

Denken und Erinnern

Ein Gehirn kann zwei Dinge. Es denkt, und es erinnert sich.

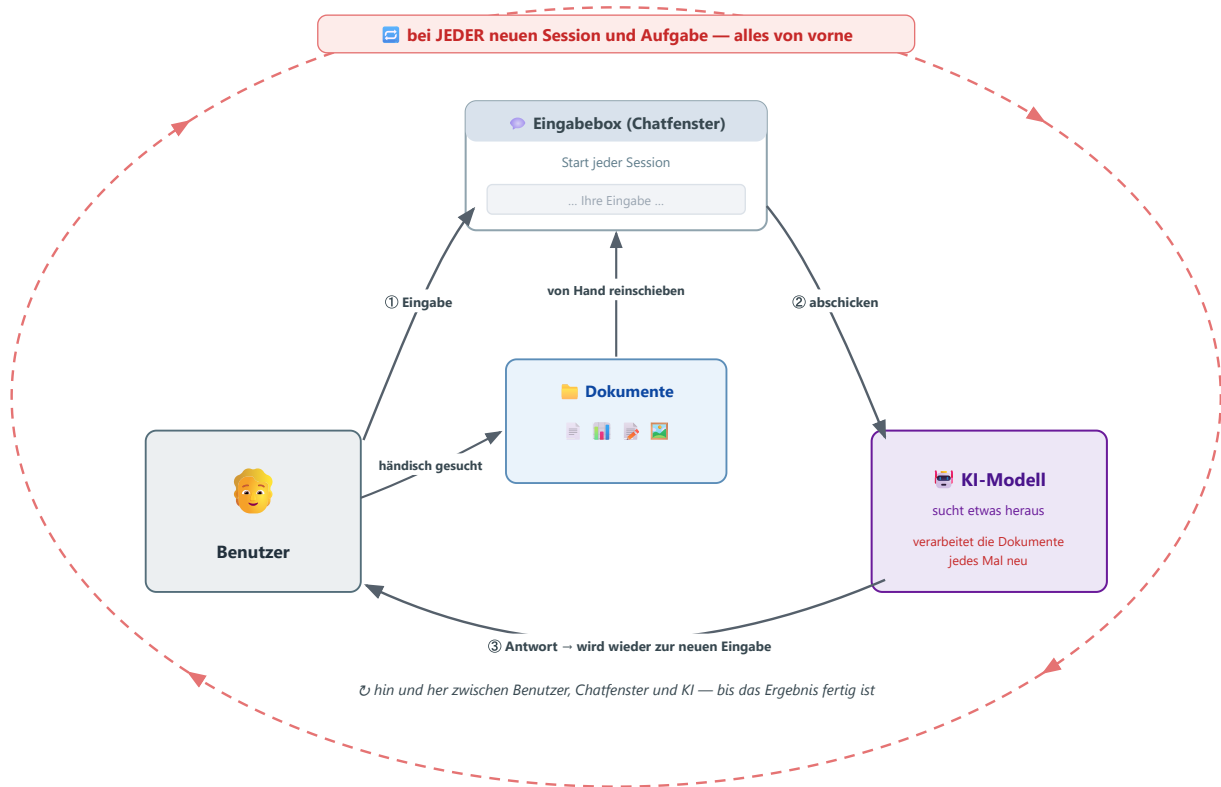
Das Denken beherrschen die heutigen KI-Agenten gut. Sie formulieren, vergleichen, fassen zusammen, ziehen Schlüsse. Woran es fehlt, ist das Erinnern. Ein KI-Agent startet jedes Gespräch bei null. Er kennt Ihre Vorgänge nicht, Ihre Kunden nicht, Ihre Gepflogenheiten nicht. Klug ist er, aber ahnungslos, was Ihren Fall angeht.

Diese zweite Hälfte liefert eine Wissensdatenbank (Knowledge Base). Sie ist das Gedächtnis: Darin liegt das Wissen Ihres Betriebs, aufbereitet so, dass ein KI-Agent es lesen kann. Zusammen ergibt das ein zweites Gehirn für Ihre Firma. Das Denken kommt vom KI-Agenten, das Erinnern aus Ihrer Wissensdatenbank.

Ein Gedächtnis ist geordnet, es wird laufend aktuell gehalten, und es gibt auf eine Frage die paar Stellen heraus, die wirklich zur Sache gehören.

Arbeitsweise mit KI-Chatbots

Wie man heute mit KI arbeitet — ein Kreislauf, jede Session neu



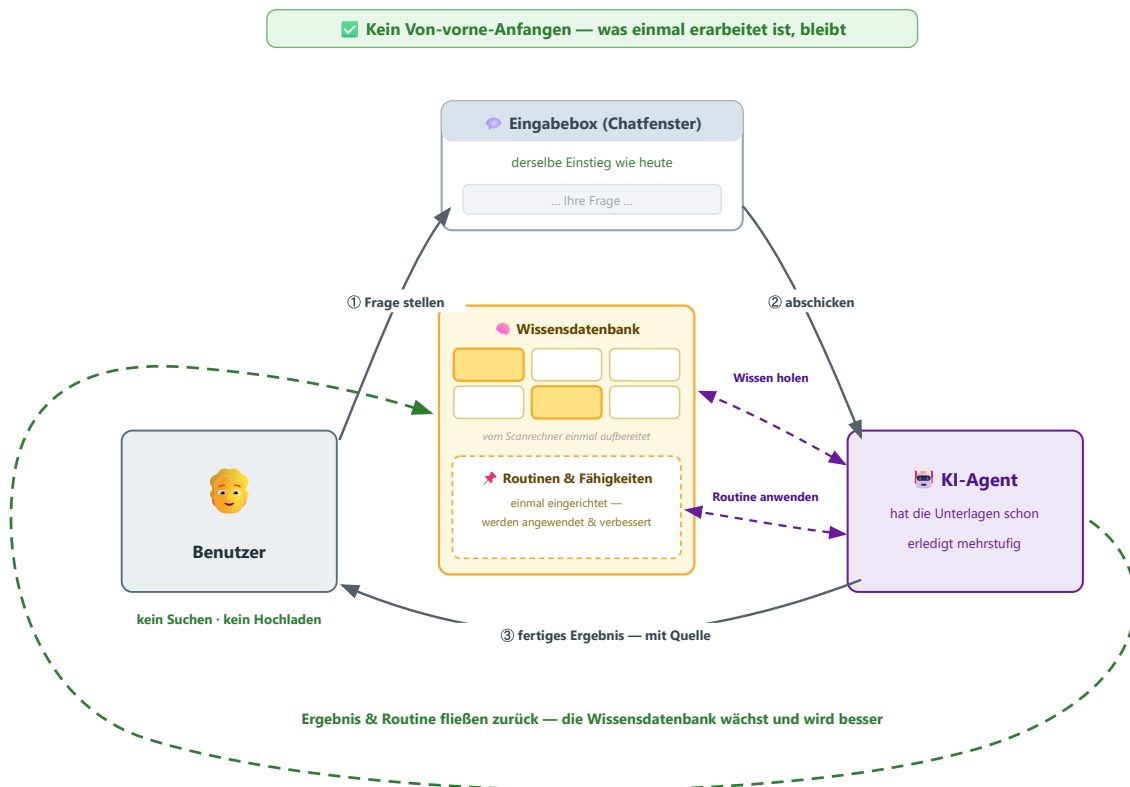
Wer heute schon mit KI arbeitet, kennt das Muster. Man öffnet ein Chatfenster und sucht erst einmal die passenden Unterlagen zusammen. Der Vertrag, die alte E-Mail, das Protokoll. Man fügt alles von Hand ein. Und dann, erst dann, beginnt die eigentliche Arbeit.

Das ist genau die Suche aus dem letzten Kapitel. Ein Chatbot nimmt sie Ihnen nicht ab, er setzt sie voraus. Dazu kommt etwas, das viele erst später bemerken: Ist das Fenster geschlossen, ist alles weg. Beim nächsten Vorgang fangen Sie wieder von vorn an und suchen dieselben Unterlagen noch einmal zusammen.

Arbeitsweise mit KI-Agenten

Wie es mit einer Wissensdatenbank läuft — jede Session baut auf der vorherigen auf

Gegenstück zu „Wie man heute mit KI arbeitet“ — gleicher Einstieg, andere Welt



Auf den ersten Blick sieht der Ablauf aus wie heute. Sie stellen Ihre Frage, Sie bekommen eine Antwort. Der Unterschied liegt in dem, was dazwischen wegfällt.

Sie suchen nichts mehr zusammen und laden nichts mehr hoch. Der KI-Agent greift selbst auf die Wissensdatenbank zu und holt sich die Stellen, die zu Ihrer Frage passen.

Und wenn der Vorgang erledigt ist, endet er nicht im Nichts. Das Ergebnis fließt zurück in die Wissensdatenbank. Beim nächsten Mal ist es schon da.

Damit sind beide Kosten aus dem ersten Kapitel weg. Die Suche entfällt, weil die Unterlagen bereitliegen. Der ewige Neuanfang entfällt, weil das Wissen bleibt. Was Sie dort einmal aufgebaut haben, ist nie umsonst.

Anweisungen / Skills / Routinen

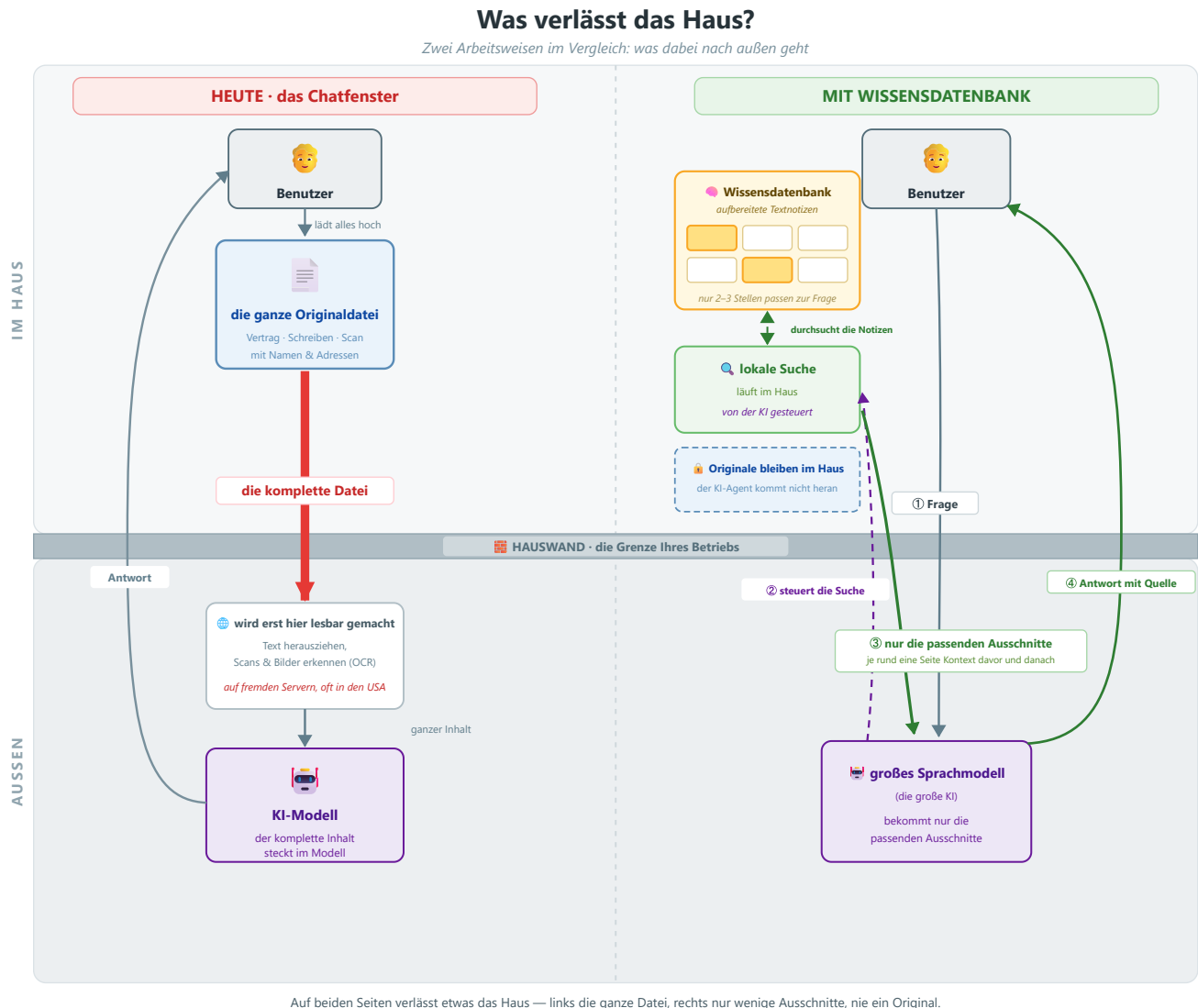
Wie bei einem neuen Mitarbeiter, dem man erst die Prozesse in der Firma erklären muss, verhält es sich mit einem KI-Agenten.

Wenn Sie einmal festgelegt haben, wie bei Ihnen ein bestimmter Vorgang abläuft, ist dieser Ablauf hinterlegt. Beim nächsten Fall wenden Sie ihn wieder an, auf einen anderen Kunden, ein anderes Projekt.

Dabei werden die Abläufe verfeinert, und verbesserte Anweisungen fließen zurück in die Wissensdatenbank — bis der Agent die Aufgabe zuverlässig so erledigt, wie Sie sie haben wollen.

Welche Daten verlassen die Firma

Die beiden Wege unterscheiden sich auch darin, was tatsächlich Ihr Haus verlässt.



Heute, über das Chatfenster, laden Sie die ganze Datei hoch: den Vertrag, das Schreiben, das Protokoll, vollständig und mit allem, was darin steht. Diese Datei geht zum externen Anbieter, oft in die USA, und wird erst dort lesbar gemacht. Der Text wird herausgezogen, Bilder und Scans werden erkannt, alles auf dessen Servern. Der komplette Inhalt landet im Modell.

Mit einer Wissensdatenbank läuft es umgekehrt. Der KI-Agent liest Ihre Frage, überlegt sich dazu eine Suche und lässt sie im Haus über die aufbereiteten Textnotizen Ihres Vorgangs laufen, nicht über die Originale, an die er gar nicht herankommt. Von jeder Fundstelle nimmt er nur die nähere Umgebung mit, etwa eine Seite davor und eine danach. Erst diese Ausschnitte gehen nach außen an das große Sprachmodell, das daraus die Antwort formuliert, mit Quelle.

Auch hier verlässt also etwas das Haus, das gehört zur ehrlichen Antwort, nur eben ein paar Absätze statt ganzer Dokumente. Was das für den Datenschutz bedeutet, steht später in einem eigenen Kapitel.

3 · Wie eine Wissensdatenbank aufgebaut ist

Die Idee ist klar. Bleibt die Frage, was in so einer Wissensdatenbank eigentlich drinsteht und was dabei aus Ihren Unterlagen wird.

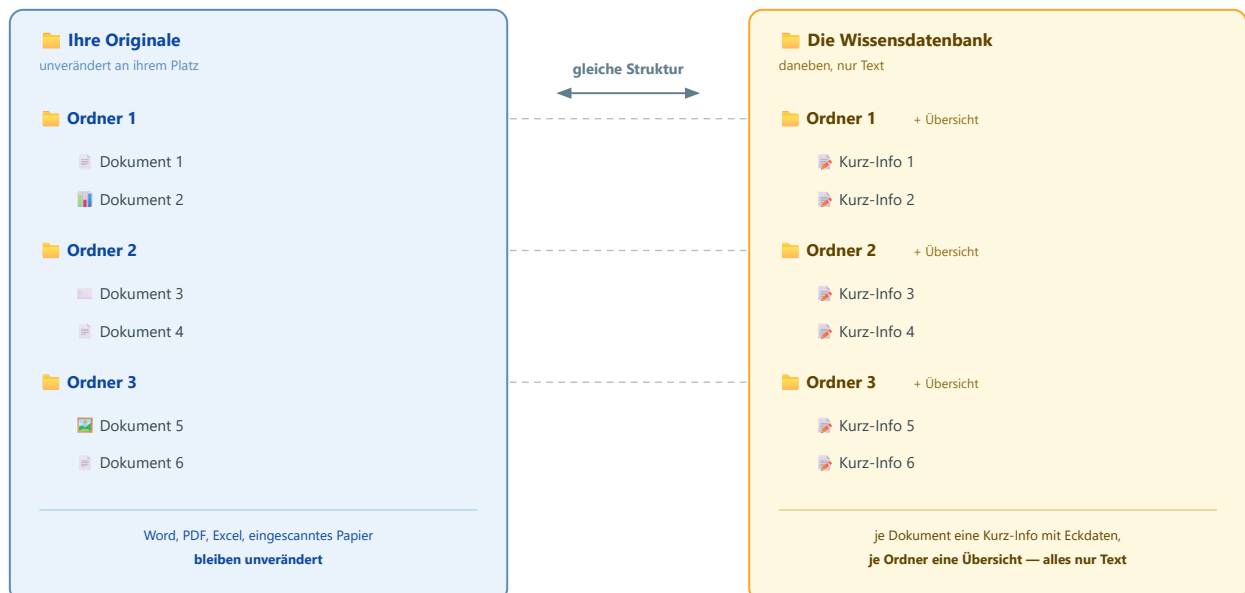
Dieselbe Ordnung, die Sie schon haben

Die Wissensdatenbank ist kein fremdes System, in dem sich Ihre Leute erst zurechtfinden müssten. Sie übernimmt die Ordnung, die bei Ihnen ohnehin gilt: dieselben Ordner, dieselbe Gliederung, dieselbe Reihenfolge wie auf Ihrem Server.

Wie diese Ordnung aussieht, gibt Ihr Betrieb vor. In der einen Firma sind es Mandanten, in der nächsten Kunden, anderswo Akten, Fälle, Vorgänge oder Projekte. Die Wissensdatenbank fragt nicht danach, sie bildet ab, was da ist.

Die Ordnerstruktur bleibt dieselbe

Links Ihre Originale, rechts die Wissensdatenbank — gleiche Gliederung, anderer Inhalt



Ob Ihre Ordner Mandanten, Kunden, Akten, Fälle, Vorgänge oder Projekte heißen, spielt keine Rolle — die Wissensdatenbank übernimmt die Gliederung, die Sie bereits haben.

Wer sich auf dem Server auskennt, kennt sich damit auch in der Wissensdatenbank aus. Es muss nichts umsortiert, umbenannt oder umgezogen werden.

Was zu jedem Vorgang darin steht

Zu jedem Vorgang gibt es eine Übersicht: worum es geht, wer beteiligt ist, wie der Stand ist. Zu jedem einzelnen Dokument kommt eine Kurz-Info dazu. Sie hält fest, was für ein Schriftstück das ist, welche Eckdaten darin vorkommen, etwa Namen, Daten, Beträge oder Fristen, und wo das Original liegt.

Das Ganze ist einfacher Text, den Sie jederzeit selbst öffnen und lesen können — nichts steckt in einem verschlossenen Format.

Wie so ein Eintrag im Einzelnen aufgebaut ist, steht im Anhang „Der Scan-Rechner“.

Ihre Originale bleiben, wie sie sind

Die Wissensdatenbank entsteht **daneben**, als eigener, zweiter Bestand. An Ihren Originalen wird nichts verändert — und der KI-Agent kommt an sie gar nicht heran, weder lesend noch schreibend.

Was entsteht, ist keine zweite Kopie Ihrer Dateien, sondern eine verdichtete Beschreibung davon. Das ist kein Sparen am falschen Ende, sondern der Sinn der Sache: Wer einem KI-Agenten alles auf einmal vorlegt, bekommt schlechtere Antworten, nicht bessere — zu viel Material verwirrt ihn genauso wie einen Menschen. Gebraucht werden die paar Stellen, die zur Frage gehören.

Aufwand der Umstellung auf eine Wissensdatenbank

Es gibt keinen Umzug und keine Migration. Ihr Bestand bleibt unangetastet an seinem Platz, während die Wissensdatenbank daneben wächst.

Sie wächst dabei von selbst. Weil Ihre Leute ihre Dokumente ablegen wie immer, ist die Wissensdatenbank kein Projekt, das jemand nebenher pflegen muss — sie kommt mit der normalen Arbeit mit.

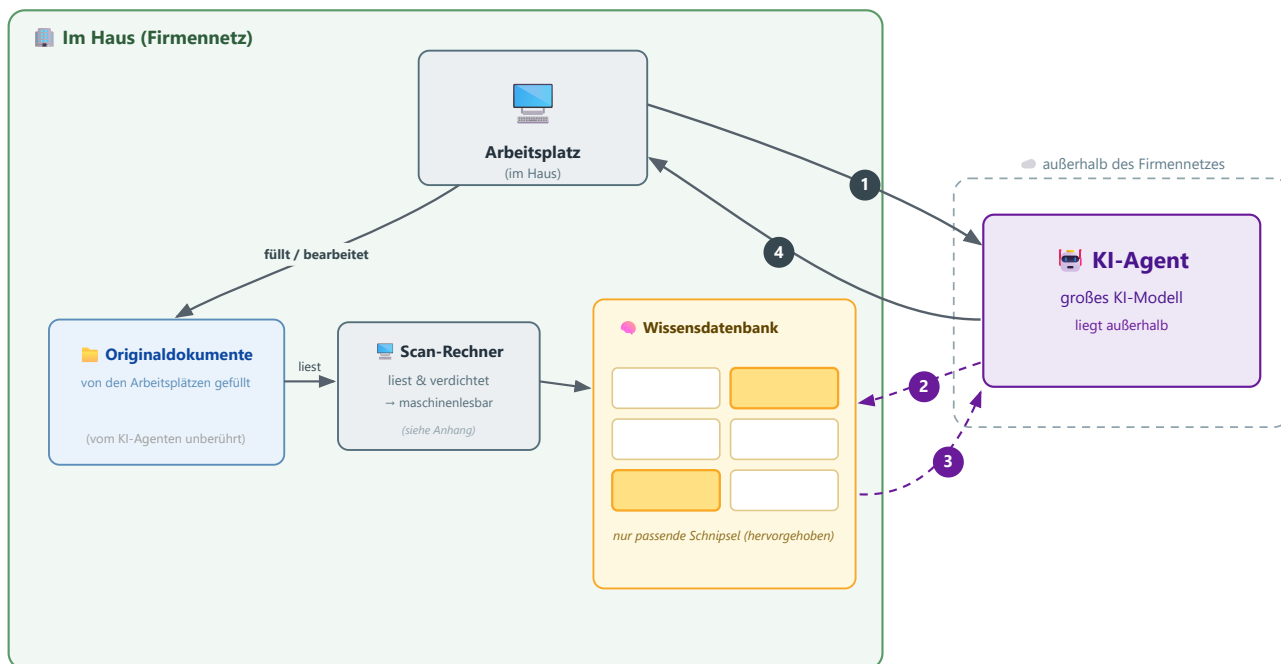
Der größere Teil dieser Aufbereitung ist übrigens programmierte Automatik: Dateien durchgehen, eingescanntes Papier lesbar machen, Doppelungen erkennen, in Archive hineinschauen. Echte KI kommt erst dort ins Spiel, wo Inhalte verstanden und zusammengefasst werden müssen. Wen die Aufteilung im Einzelnen interessiert, findet sie im Anhang „Der Scan-Rechner“.

4 · Kommunikationswege mit der Wissensdatenbank

Die Wissensdatenbank steht. Bleibt die praktische Frage: Wie kommt man eigentlich an sie heran, im Büro und unterwegs?

Kommunikation mit dem KI-Agenten in der Firma

Im Haus — Aufbau & Frage



Legende — eine Frage an den KI-Agenten (die vier Schritte)

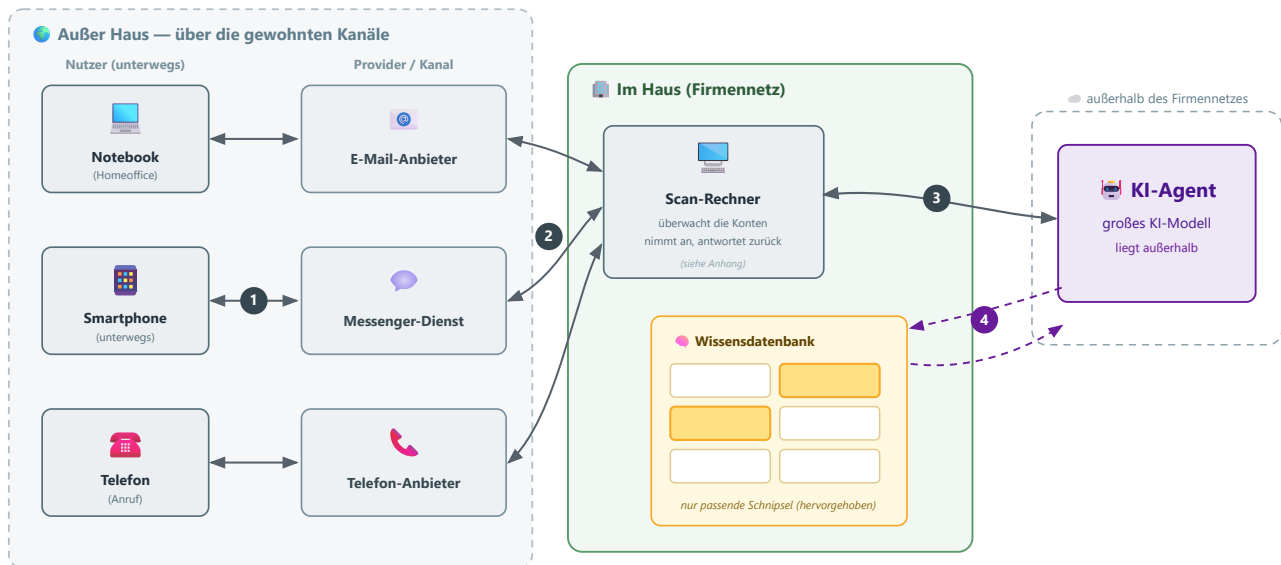
- 1 Der Arbeitsplatz stellt dem KI-Agenten eine Frage.
- 2 Erst dadurch ausgelöst greift er kurz in die Wissensdatenbank — **kein Dauerzugriff**.
- 3 Zurück kommen **nur die passenden Schnipsel** — nicht der ganze Bestand.
- 4 Der KI-Agent formt daraus die Antwort — mit Quelle.

Im Büro ändert sich am Ablauf wenig: Sie stellen Ihre Frage im Chatfenster, die Antwort kommt zurück. Dabei hilft Ihnen die vorbereitete Navigations-Fähigkeit, die Daten schnell und zuverlässig zu finden. Sie brauchen nur z. B. „Vorgang 123“ oder „Mandant Heinz Meier“ einzugeben, und der KI-Agent weiß, wo er nachschauen muss. Mehr müssen Ihre Leute nicht lernen.

Der KI-Agent bekommt **keinen Dauerzugriff** auf Ihren Bestand. Er greift erst zu, wenn eine Frage gestellt wird, holt sich die dazu passenden Ausschnitte und formt daraus die Antwort. Danach ist der Zugriff wieder beendet.

Kommunikation mit dem KI-Agenten außerhalb der Firma

Von unterwegs — über die gewohnten Kanäle



Legende — der Weg einer Frage von unterwegs (Kanäle laufen in beide Richtungen)

- 1 Sie fragen über den **gewohnten Kanal** (E-Mail, Messenger, Telefon) — **kein VPN, kein Login-Ritual.**
- 2 Der **Scan-Rechner** im Haus überwacht die Konten, nimmt die Frage auf — und schickt die Antwort **denselben Weg zurück.**
- 3 Er lässt den **KI-Agenten** die Antwort formulieren.
- 4 Der zieht dafür **nur die passenden Schnipsel** aus der Wissensdatenbank — **kein Dauerzugriff.**

Von außen führt nur ein einziger kontrollierter Kanal hinein — das Firmennetz bleibt geschlossen.

Von unterwegs brauchen Sie weder VPN-Einwahl noch Anmeldung. Sie schreiben eine E-Mail, schicken eine Nachricht im Messenger oder rufen an, so wie Sie es mit einem Kollegen auch tun würden.

Im Haus behält der Scan-Rechner diese Konten im Blick. Er nimmt Ihre Frage entgegen, lässt den KI-Agenten antworten und schickt das Ergebnis denselben Weg zurück. Für Sie sieht es aus wie eine normale Antwort auf Ihre normale Nachricht. Was dieser Rechner sonst noch alles tut, steht im Anhang.

Sicherheit ist dabei ein Nebeneffekt

Der übliche Weg, Mitarbeitern von außen Zugang zu geben, ist ein VPN-Zugang. Damit öffnet man im Zweifel das halbe Firmennetz für ein Gerät, das irgendwo unterwegs ist.

Hier ist es umgekehrt. Nach außen zeigt nur ein einziger, kontrollierter Kanal, und das Firmennetz selbst bleibt geschlossen. Der Zugang wird bequemer und die Angriffsfläche gleichzeitig kleiner. Das ist selten der Fall und einer der angenehmeren Nebeneffekte dieser Bauweise.

Innen wie außen dasselbe Muster

Ob im Büro oder von unterwegs — der KI-Agent ist zu keinem Zeitpunkt fest mit Ihrem Bestand verbunden. Er greift nur auf die einzelne Frage hin zu und ist danach wieder außen vor.

5 · Was der KI-Agent für Sie tun kann

Die Wissensdatenbank steht, der Zugang ist geklärt. Bleibt die Frage, die am Ende zählt: Was haben Sie davon im Alltag?

Der Unterschied zwischen Fragen und definierten, selbstständigen Routinen

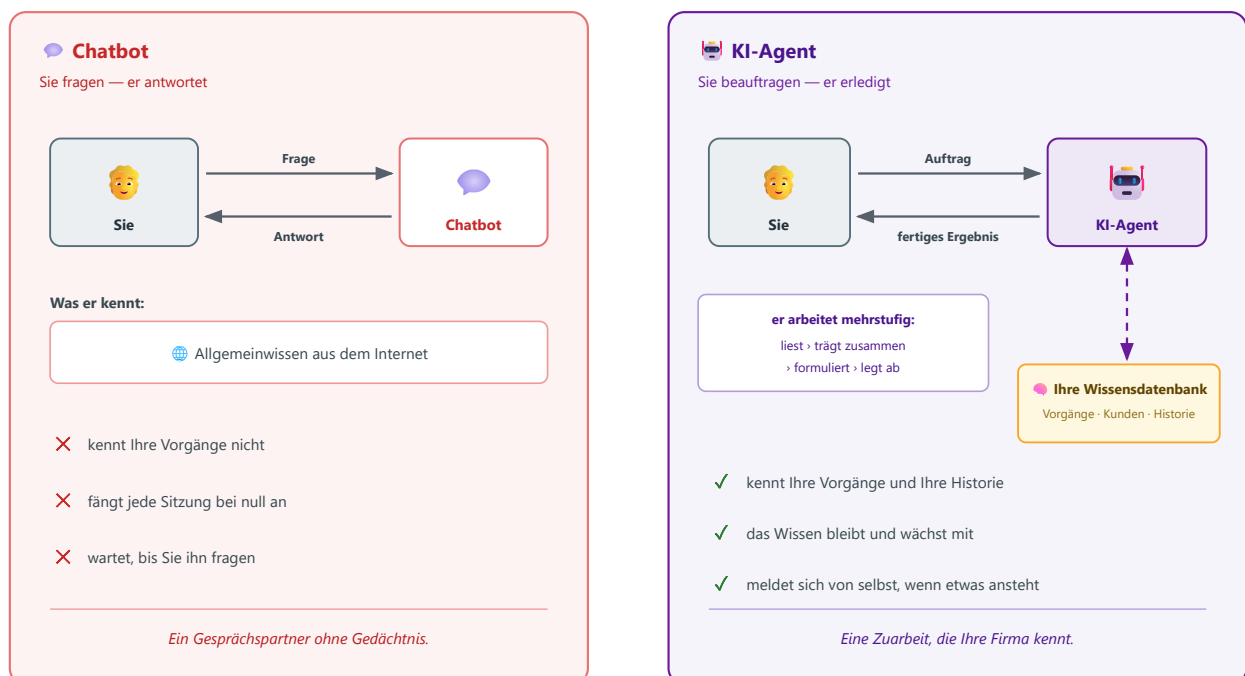
Was die meisten von KI kennen, ist ein Chatbot: Sie fragen, er antwortet, Sie machen weiter. Er ist ein sehr belebter Gesprächspartner, aber er tut nichts.

Ein KI-Agent arbeitet anders. Er erledigt mehrstufige Aufgaben selbstständig: Er liest, sucht heraus, trägt zusammen, formuliert und legt ab. Sie beauftragen ihn, statt ihn zu befragen.

Dass das überhaupt geht, liegt an der Wissensdatenbank. Ein Chatbot kennt das Internet, aber nicht Ihre Firma. Ein KI-Agent auf Ihrem eigenen Bestand kennt Ihre Vorgänge, Ihre Kunden und Ihre Vorgeschichte.

Chatbot oder KI-Agent — der Unterschied

Links das, was die meisten kennen. Rechts das, was auf Ihrer eigenen Wissensdatenbank möglich ist.



Der Unterschied liegt nicht am Modell, sondern daran, worauf es zugreifen kann.

Was die KI kann

Der größte Teil des Nutzens steckt in unspektakulären Dingen, die heute Zeit kosten:

Finden. Sie fragen in normaler Sprache, statt sich durch Ordner und Postfächer zu klicken. Wo liegt der Vorgang, in dem es um diese Sache ging? Was haben wir damals vereinbart?

Zusammenfassen. Ein langer Vertrag, ein Gutachten, eine ganze Aktenlage — auf das Wesentliche gebracht, in dem Umfang, den Sie brauchen.

Vergleichen. Zwei Fassungen eines Dokuments nebeneinandergelegt, mit den Unterschieden im Klartext. Was hat sich zwischen dem alten und dem neuen Stand geändert?

Zusammentragen. Das ist der Punkt, an dem heute die meiste Zeit verloren geht: die vielen Einzelteile eines Vorgangs zu einem stimmigen Bild zusammensetzen. Der KI-Agent macht genau das und liefert ein fertiges

Ergebnis, nicht zehn Fundstellen zum Selbersortieren.

Und immer gilt: Die Antwort kommt **mit Quellenangabe**. Sie sehen, aus welchen Unterlagen sie stammt, und können jederzeit nachschlagen. Das ist der Unterschied zwischen einem Werkzeug, dem man vertrauen kann, und einem, das raten könnte.

Fertige Ergebnisse statt langwieriger Kommunikation

Über das Beantworten hinaus erarbeitet der KI-Agent Ihnen Dinge, die Sie direkt weiterverwenden:

Er schreibt **Entwürfe** für Briefe, Angebote oder Stellungnahmen, und zwar in dem Stil, den Ihr Haus ohnehin verwendet, weil er ihn aus Ihren früheren Schreiben kennt.

Er füllt **wiederkehrende Formulare und Vorlagen** aus, weil die nötigen Angaben in der Wissensdatenbank stehen.

Und er gibt Ergebnisse **strukturiert** aus, wenn Sie das brauchen: eine Aufstellung über mehrere Vorgänge hinweg, als Tabelle, mit der Sie sofort weiterrechnen können.

Zeitgesteuerte Routinen

Bis hierhin haben immer Sie gefragt. Der interessantere Teil ist der, bei dem Sie nicht mehr fragen müssen.

Sie legen einmal fest, was Sie interessiert: bestimmte Vorgänge, bestimmte Kennzahlen, offene Punkte. Danach kommt der Überblick von selbst, zum Beispiel jeden Morgen. Was ist erledigt, was ist offen, was hat sich seit gestern geändert, wo hakt es.

Dasselbe gilt für **Fristen und Wiedervorlagen**. Der KI-Agent behält sie im Blick und meldet sich rechtzeitig, statt darauf zu warten, dass jemand nachsieht.

Aus Hinterhertelefonieren wird damit eine Meldung, die von selbst eintrifft. Das ist der eigentliche Sprung: Ein Chatbot wartet auf Fragen, ein Agent arbeitet im Hintergrund und bringt Ihnen das Wichtige aktiv.

Er übernimmt die aufwendige Arbeit, nicht die Entscheidung

Alles, was hier steht, betrifft das Suchen, Kopieren, Zusammentragen und Vorformulieren. Also genau die Arbeit, die einen großen Teil des Tages frisst.

Was der KI-Agent nicht übernimmt, ist Ihr Urteil. Er legt Ihnen einen Entwurf hin, keine fertige Entscheidung. Er sagt Ihnen, was in den Unterlagen steht, und nennt die Quelle dazu. Was daraus folgt, entscheiden weiterhin Sie und Ihre Leute — nur eben mit der Vorarbeit schon erledigt.

Und die Daten?

Damit der KI-Agent das leisten kann, muss er die passenden Ausschnitte aus Ihrer Wissensdatenbank auch lesen dürfen. Was das für den Datenschutz bedeutet, und warum die ehrliche Antwort darauf nicht „alles bleibt im Haus“ lautet, klärt das nächste Kapitel.

6 · Datenschutz und Vertrauen — die ehrliche Antwort

An dieser Stelle steht in vielen Angeboten ein beruhigender Satz, der nicht stimmt. Wir machen es anders.

Die ehrliche Antwort zuerst

Sie werden von Anbietern oft hören, Ihre Daten verließen niemals Ihr Haus. Für das, was hier beschrieben wird, wäre das gelogen.

Der Grund ist unspektakulär. Die anspruchsvollen Aufgaben, also verstehen, zusammenfassen, formulieren, kann heute nur ein großes KI-Modell. Eine Anlage, die so etwas vollständig bei Ihnen im Keller leistet, kostet mehr als das gesamte übrige Vorhaben und wäre trotzdem schwächer. Für die Aufbereitung im Hintergrund reicht Ihre eigene Technik völlig aus, für das Denken nicht.

Also gehen Daten hinaus. Die Frage ist nicht, ob das passiert, sondern **wie viel** hinausgeht und **zu wem**.

Was tatsächlich das Haus verlässt

Hier wird der Unterschied deutlich. Es geht nicht Ihr Bestand hinaus und es gehen auch nicht Ihre Dokumente hinaus.

Hinaus gehen die **Ausschnitte, die zu einer konkreten Frage passen**. Nur diese, und nur in dem Moment, in dem jemand tatsächlich fragt. Es gibt keine dauerhafte Verbindung, über die Ihr Bestand fließt, und es gibt keine Kopie Ihrer Unterlagen bei einem Anbieter.

Ihre **Originaldokumente** bleiben dabei von vornherein außen vor. Der KI-Agent kommt an sie gar nicht heran, er arbeitet ausschließlich mit der Wissensdatenbank.

Verglichen mit dem, was heute üblich ist, ist das eine drastische Verkleinerung: Statt ganzer Dokumente wandern einzelne Absätze hinaus, und zwar nur die, die gebraucht werden.

Wohin es geht — und welche Wahl Sie haben

Bleibt die Frage nach dem Empfänger. Dafür gibt es drei gangbare Wege, und Sie können sich aussuchen, welcher zu Ihrem Haus passt.

Ein europäischer Anbieter. Es gibt leistungsfähige KI-Anbieter in Europa, die der **Datenschutz-Grundverordnung (DSGVO)** unterliegen; **Mistral AI** aus Frankreich ist das bekannteste Beispiel. Das ist der einfachste Weg, weil die Frage des Datenexports damit gar nicht erst entsteht.

Ein großer Anbieter mit vertraglichem Trainings-Ausschluss. Auch bei den großen amerikanischen Anbietern lässt sich festlegen, dass Ihre Inhalte **nicht zum Training der Modelle** verwendet werden. Über die geschäftlichen Zugänge ist das ohnehin die Regel, nicht die Ausnahme. Ihre Anfrage wird im Rechenzentrum verarbeitet, die Antwort geht zurück, und der Inhalt wird weder ins Modell eingearbeitet noch dauerhaft aufbewahrt. Wer die Speicherung ganz ausschließen will, kann bei den geschäftlichen Zugängen zusätzlich vereinbaren, dass über die unmittelbare Verarbeitung hinaus **gar nichts gespeichert** wird.

Getrennt davon zu klären ist der Datenexport in die USA an sich — ein eigener Punkt, den man nicht mit der Trainingsfrage verwechseln sollte. Auch er lässt sich inzwischen entschärfen oder ganz vermeiden: Die großen Anbieter bieten mittlerweile **Verarbeitung innerhalb der EU** an, teils unmittelbar mit europäischem Speicher- und Verarbeitungsort, teils dadurch, dass ihre Modelle in den europäischen Rechenzentren der

großen Cloud-Plattformen laufen. Dann bleibt die Verarbeitung in Europa; und wo doch Daten in die USA gehen, gibt es dafür etablierte vertragliche Grundlagen.

Ein offenes Modell bei einem Betreiber Ihrer Wahl. Es gibt inzwischen frei verfügbare Modelle, deren Bauplan offenliegt. Sie laufen nicht beim ursprünglichen Entwickler, sondern dort, wo Sie sie betreiben lassen, etwa in einem europäischen Rechenzentrum. Der Entwickler des Modells sieht von Ihren Daten nichts, weil er an der Verarbeitung gar nicht beteiligt ist.

Der Einwand, den man dazu früher hören konnte, hat sich weitgehend erledigt. Diese offenen Modelle sind in den vergangenen Jahren so stark geworden, dass der Abstand zu den Spitzenmodellen der großen Anbieter für die Aufgaben, um die es hier geht, kaum noch ins Gewicht fällt. Sie zahlen für mehr Kontrolle also nicht mehr mit spürbar schlechteren Ergebnissen.

Damit wird aus einer offenen Datenschutzfrage ein Punkt, den man abarbeitet wie jeden anderen Dienstleistervertrag auch. Auf jedem dieser drei Wege ist ein **DSGVO-konformer Betrieb möglich** — vorausgesetzt, man wählt den geschäftlichen Zugang mit den passenden Verträgen, nicht den kostenlosen Endkunden-Dienst.

Und Ihre Firmengeheimnisse?

Der zweite große Einwand lautet: Da geht doch unser Betriebswissen hinaus.

Dann sehen wir uns einmal an, worum es dabei tatsächlich geht. In den allermeisten Fällen sind es **normale Abläufe**. Wie ein Angebot kalkuliert wird, wie man eine Reklamation bearbeitet, welche Unterlagen zu einem Vorgang gehören, wie eine Akte aufgebaut ist. Solche Abläufe gibt es in Deutschland, in den USA, in China, in Argentinien und im Kongo. Sie sind millionenfach beschrieben, und die Modelle kennen sie längst aus aller Welt. Das ist Handwerk, kein Geheimnis.

Echte Betriebsgeheimnisse gibt es natürlich auch: ein Forschungsergebnis, ein Verfahren, eine Rezeptur, eine Konstruktion, die sonst niemand hat. Nur trifft das auf die wenigsten Betriebe überhaupt zu, und wo es zutrifft, betrifft es einen kleinen, klar abgrenzbaren Teil der Unterlagen.

Man prüft einmal, ob es solche Bereiche gibt, und nimmt sie gegebenenfalls aus. Der Rest, also der weitaus größte Teil Ihres Bestands, ist unproblematisch.

Der Vergleich, den selten jemand zieht

Und jetzt der Punkt, der in dieser Diskussion meistens fehlt: Wie sieht es denn heute bei Ihnen aus?

In vielen Betrieben laden Mitarbeiter längst Unterlagen von Hand in frei zugängliche KI-Dienste hoch, weil es die Arbeit erleichtert. Ganze Verträge, ganze Schreiben, oft mit Namen und Adressen darin. Das geschieht ohne Vertrag, ohne Prüfung und ohne dass jemand davon weiß.

Datenschutzrechtlich ist das in aller Regel nicht zulässig. Es fällt nur niemandem auf, weil es unsichtbar passiert.

Die Wahl steht also nicht zwischen „datenschutzkonform“ und „nicht datenschutzkonform“. Sie steht zwischen einem **unkontrollierten Zustand, den niemand überblickt**, und einem **geregelten Weg**, bei dem festgelegt ist, was hinausgeht, wohin es geht und unter welchem Vertrag.

Nachprüfbar, und es gehört Ihnen

Zwei Dinge kommen hinzu, die weniger mit Recht und mehr mit Vertrauen zu tun haben.

Der KI-Agent antwortet **mit Quellenangabe**. Sie sehen, aus welchen Unterlagen eine Aussage stammt, und können nachschlagen. Sie müssen ihm nicht glauben, Sie können ihn prüfen.

Und die Wissensdatenbank selbst ist **einfacher Text auf Ihren eigenen Systemen**. Kein Format, das nur ein bestimmter Anbieter lesen kann. Wenn Sie den Anbieter wechseln oder das Vorhaben beenden, bleibt der aufbereitete Bestand bei Ihnen. Das Modell mieten Sie, die Wissensdatenbank gehört Ihnen.

Eine Selbstverständlichkeit zum Schluss

Wie das im Einzelfall auszugestalten ist, hängt von Ihrer Branche und Ihren Daten ab. Ein Steuerberater hat andere Auflagen als ein Planungsbüro, und besondere Kategorien wie Gesundheitsdaten verlangen eigene Sorgfalt. Diese Seite ersetzt keine Prüfung durch Ihre Datenschutzbeauftragten, sie zeigt, dass die Frage lösbar ist und wie.

7 · Wissensdatenbank zukunftssicher?

Ein Gedanke zum Schluss.

Man könnte jetzt sagen: „Das ist noch so eine neue Technologie, und sie entwickelt sich so schnell weiter — wer weiß, was es nächstes Jahr gibt.“

Das stimmt, und der Einwand ist berechtigt. Er trifft nur nicht den Teil, um den es hier geht. Was sich schnell ändert, sind die KI-Modelle selbst: Alle paar Monate kommt ein besseres, und welches Sie morgen einsetzen, ist eine Frage von Wochen, nicht von Jahren. Ihr geordnetes Firmenwissen ändert sich davon nicht. Ein heute aufbereiteter Vorgang beschreibt in fünf Jahren immer noch dieselbe Sache.

Beides steht auch nicht im Gegensatz zueinander — im Gegenteil, das eine ist die Voraussetzung für das andere. Auch die KI-Agenten der nächsten Jahre, welche auch immer das sein werden, können nur mit dem arbeiten, was Ihre Firma an aufbereitetem Wissen und eingespielten Routinen bereitstellt. Eine geordnete Wissensdatenbank ist damit nicht eine von vielen KI-Ideen, sondern die Grundlage, auf der die anderen erst stehen. Ohne sie bleibt auch das beste Werkzeug ohne den Kontext, den es für Ihre Arbeit braucht.

Dass die Technik dafür heute schon taugt, steht dabei außer Frage. Selbst kleinere, frei verfügbare Modelle sind inzwischen so leistungsfähig, dass sie einen großen Teil der täglichen Arbeit zuverlässig und schnell erledigen. Was heute noch fehlt, ist selten das Modell — es ist die geordnete Grundlage, auf der es in Ihrem Betrieb arbeiten kann.

Dazu kommt ein zweiter Gedanke, der mit künstlicher Intelligenz zunächst gar nichts zu tun hat. In vielen Betrieben steckt der wertvollste Teil des Wissens in wenigen Köpfen — bei den Leuten, die seit Jahren dabei sind und wissen, warum etwas so gemacht wird und nicht anders. Geht einer von ihnen in Rente oder wechselt den Arbeitgeber, geht ein Stück Betrieb mit. Eine Wissensdatenbank hält fest, was sonst nur mündlich weitergegeben wird — als geordnetes Gedächtnis Ihres Hauses, das einen Personalwechsel übersteht. Sie ist damit auch dann ein Gewinn, wenn Sie nie eine Zeile mit einer KI schreiben.

Anhang: Der Scan-Rechner — wie die Wissensdatenbank entsteht

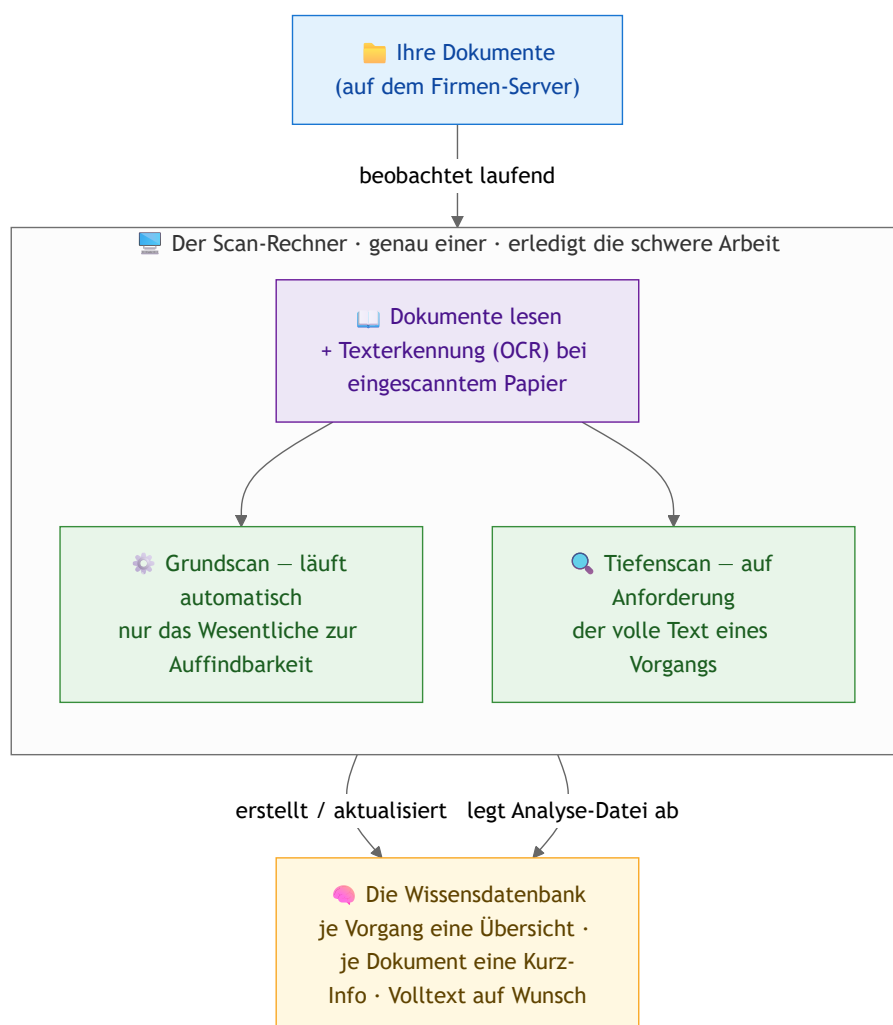
Der Hauptteil zeigt, was eine Wissensdatenbank leistet. Dieser Anhang zeigt, wie sie entsteht. Für alle, die es genauer wissen wollen; zum Verständnis des Hauptteils ist er nicht nötig.

Der Motor im Hintergrund

Damit aus tausenden verstreuten Dokumenten eine durchsuchbare Wissensdatenbank wird, muss sie zuerst jemand lesen, einordnen und verdichten. Diese Arbeit übernimmt ein einziger, fest dafür eingerichteter Rechner: der **Scan-Rechner**.

Er ist der Motor des Ganzen und zugleich der erste Baustein, den man einrichtet. Steht dieser Rechner, füllt sich die Wissensdatenbank von selbst.

Der Weg auf einen Blick



So liest man das Bild

Oben liegen Ihre Unterlagen auf dem Server. Der Scan-Rechner in der Mitte **beobachtet** sie laufend, **liest** sie (bei eingescanntem Papier mit Texterkennung) und arbeitet in **zwei Betriebsarten** — automatischer Grundscan und angeforderter Tiefenscan. Beide füllen dasselbe Ziel: die **Wissensdatenbank** unten.

Warum genau ein Rechner

Die aufwendigen Werkzeuge liegen **nur auf diesem einen Rechner**: Dokumente aller Formate öffnen, eingescanntes Papier per Texterkennung lesbar machen, Inhalte auf das Wesentliche eindampfen. Das hat handfeste Vorteile:

- **Alle anderen Arbeitsplätze bleiben schlank.** Ihre Mitarbeiterinnen und Mitarbeiter brauchen keine Spezialsoftware.
- **Wartung an einer einzigen Stelle.** Aktualisierungen, Sprachpakete für die Texterkennung, Einstellungen — alles an genau einem Ort.
- **Keine doppelte Arbeit.** Weil nur ein Rechner scannt, kann sich das System nicht selbst ins Gehege kommen; jeder Vorgang wird sauber einmal verarbeitet.

Eine Voraussetzung gibt es: Dieser Rechner muss laufen, damit Scans passieren. Ist er aus, wartet die Arbeit, bis er wieder läuft. Verloren geht dabei nichts.

Was Programm ist und wo wirklich KI arbeitet

Hier lohnt sich der genaue Blick, weil in der öffentlichen Wahrnehmung meist alles als „KI“ verbucht wird. Tatsächlich ist der größere Teil dieser Arbeit **stumpfe Programm-Logik**, die jederzeit nachvollziehbar dasselbe tut:

- Ordner durchgehen und erkennen, was neu oder geändert ist
- Dateiformate öffnen und den Text herausholen (PDF, Word, Excel, E-Mail, ZIP)
- Die Texterkennung anstoßen, wenn eine Seite nur ein Bild ist
- Dubletten erkennen, indem der **Inhalt** der Dateien verglichen wird, nicht der Name
- In Archive hineinschauen und prüfen, ob deren Inhalt schon ausgepackt vorliegt
- Die vorhandene Ordnerstruktur spiegeln und die Ergebnisse ablegen

Das ist Buchhaltung, keine Intelligenz. Es läuft zuverlässig, kostet fast nichts und liefert bei gleichem Eingang immer dasselbe Ergebnis.

Echte KI kommt erst dort ins Spiel, wo etwas **verstanden** werden muss:

- Worum geht es in diesem Dokument überhaupt?
- Eine knappe Zusammenfassung formulieren
- Die Eckdaten herausziehen: Beteiligte, Daten, Beträge, Fristen
- Aus den einzelnen Dokumenten eine stimmige Übersicht des ganzen Vorgangs bilden

Diese Aufteilung ist bewusst so gewählt. Sie hält die Kosten niedrig, weil das teure Sprachmodell nur für den kleinen Teil gebraucht wird, bei dem es wirklich hilft. Und sie macht das Ergebnis überprüfbarer: Wo ein Programm gearbeitet hat, kann man die Regel nachlesen.

Auch die Texterkennung ist keine Zauberei

Die Texterkennung (OCR) ist zwar ein trainiertes Verfahren, aber ein eng spezialisiertes: Sie macht aus Bildpunkten Buchstaben und versteht dabei nichts vom Inhalt. Sie gehört in diesem Sinne zur Werkzeugkiste, nicht zur denkenden Seite.

Zwei Betriebsarten

Der Grundscan — läuft automatisch. Der Rechner schaut regelmäßig von selbst nach, ob in den Ablageordnern etwas Neues oder Geändertes liegt: tagsüber häufiger ein kurzer Blick, nachts ein vollständiger Durchgang. Er erfasst dabei **bewusst nur das Wesentliche zur Auffindbarkeit** — welche Dokumente es gibt, worum es ungefähr geht, und die Eckdaten. Er liest nicht jedes Dokument von vorne bis hinten. Für Finden und Einordnen reicht das.

Der Tiefenscan — auf Anforderung. Braucht jemand den **vollständigen Text** eines Vorgangs, fordert er einen Tiefenscan an. Der Scan-Rechner liest dann alle Seiten vollständig und legt das Ergebnis als durchsuchbare Analyse neben das Originaldokument. Der Tiefenscan **ergänzt** den Grundscan, er ersetzt ihn nicht.

Wann man ihn tatsächlich braucht. Der Grundscan genügt für den Alltag: Wo liegt der Vorgang, worum ging es, wer war beteiligt, was ist der Stand. Ein Tiefenscan lohnt sich, wenn es auf den Wortlaut ankommt — etwa wenn jemand einen sechzigseitigen Vertrag auf eine bestimmte Klausel prüfen soll, wenn ein umfangreiches Gutachten inhaltlich ausgewertet wird, oder wenn ein Vorgang übergeben wird und die Nachfolgerin ihn vollständig verstehen muss. Man fordert ihn also gezielt für einzelne Vorgänge an, statt den ganzen Bestand auf Verdacht vollständig zu erfassen.

Wie ein Eintrag konkret aussieht

Alles, was der Scan-Rechner erzeugt, ist einfacher Text. Zwei Sorten von Einträgen entstehen dabei.

Je Dokument eine Kurz-Info. Sie hält in wenigen Zeilen fest, was für ein Schriftstück vorliegt, worum es geht, welche Eckdaten darin vorkommen und wo das Original liegt. Sinngemäß so:

Art: Schreiben, 3 Seiten, eingescannt (per Texterkennung erschlossen) **Worum es geht:** Rückfrage zu einer Position aus dem Angebot vom Vormonat, mit Bitte um Klarstellung bis zum Monatsende. **Eckdaten:** Beteiligte · Datum · betroffene Position · genannte Frist **Original:** Dateiname und Ablageort **Hinweis:** liegt zusätzlich in einem zweiten Ordner (Dublette, nicht erneut erfasst)

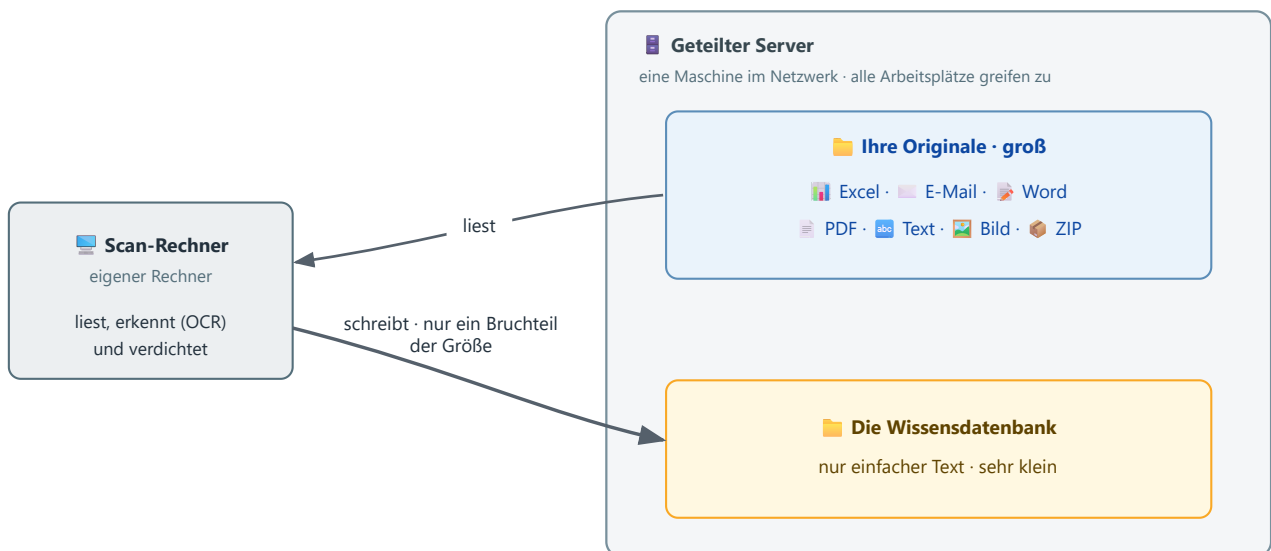
Je Vorgang eine Übersicht. Sie fasst zusammen, worum der Vorgang insgesamt geht, welche Dokumente dazugehören, wie sie zeitlich zusammenhängen und wo er gerade steht.

Beides kann ein Mensch lesen und beim Draufschaun sofort prüfen. Genau das ist der Grund, warum die Wissensdatenbank auch dann noch Wert hat, wenn man das Werkzeug dahinter wechselt.

Gleiche Struktur, ein Bruchteil der Größe

Die Wissensdatenbank **spiegelt Ihre gewohnte Ordnerstruktur eins zu eins** — dieselben Ordner, dieselben Vorgänge, in derselben Anordnung. Der Unterschied liegt im Inhalt: Wo im Original ein umfangreicher Vertragsscan oder ein ganzer Fotoordner liegt, steht daneben nur eine schlanke Textnotiz.

Beides liegt am selben Ort. Originaldokumente **und** Wissensdatenbank liegen auf einem **geteilten Server im Netzwerk**, auf den alle Arbeitsplätze zugreifen. Der Scan-Rechner ist eine davon getrennte Maschine; er liest von diesem Server die Dokumente und legt dort die Wissensdatenbank gleich daneben ab.



Aus großen Dateien wird schlanker Text

Die Wissensdatenbank ist nur ein Bruchteil der Originale groß — sie enthält keine Bilder, Scans oder Layouts, sondern nur deren Inhalt in Textform. Dadurch lässt sie sich in Sekunden durchsuchen und mühelos sichern, während die großen Originale unangetastet an ihrem Platz bleiben.

Nichts doppelt: Dubletten und Archive

Gewachsene Ablagen sind selten aufgeräumt. Dasselbe Dokument liegt oft mehrfach herum, und Archive bleiben nach dem Auspacken liegen.

Dubletten — jedes Dokument nur einmal. Der Scan-Rechner erkennt Doppelungen, indem er die Dateien an ihrem **Inhalt** vergleicht, nicht bloß am Namen, und nimmt jedes Dokument nur **einmal** auf. Die weiteren Fundstellen werden vermerkt, aber es entsteht kein doppelter Eintrag.

ZIP-Archive — hineinschauen statt doppelt zählen. Ein häufiger Fall: Jemand legt ein Archiv ab, packt es aus und lässt das Archiv liegen. Der Scan-Rechner **schaut hinein**: Ist der Inhalt bereits ausgepackt vorhanden, wird er nicht ein zweites Mal erfasst. Nur was ausschließlich im Archiv steckt, kommt zusätzlich dazu.

Bewusste Grenzen

Nicht alles ergibt als Text Sinn, und das ist Absicht:

- **Rein bildhafte Dokumente** wie großformatige Pläne, Zeichnungen oder Karten werden nicht Volltext-erfasst; sie liefern kaum verwertbaren Text. Sie werden vermerkt, Details klärt man am Original.
- **Sehr große Datentabellen** (lange Mess-, Protokoll- oder Exportdaten) werden nicht ausgeschrieben, sondern als „Massendaten“ vermerkt. Die wertet man besser direkt an der Quelle aus.

Diese Zurückhaltung ist ein Qualitätsmerkmal. Die Wissensdatenbank bleibt schlank und schnell, statt sich mit Ballast zu füllen.

Was der Scan-Rechner nicht tut

Der Scan-Rechner **liest nur**. Er verändert, verschiebt oder löscht niemals ein Originaldokument, und er kann es auch gar nicht: Auf die Originale hat er keinen schreibenden Zugriff. Sie bleiben unangetastet an ihrem

gewohnten Ort; daneben entstehen ausschließlich zusätzliche Notizen und Verweise.

Und der Scan-Vorgang selbst läuft **vollständig im Haus**, auf Ihrer eigenen Technik. Beim Aufbereiten verlassen Ihre Dokumente den Betrieb nicht.

Nicht zu verwechseln mit dem späteren Arbeiten

Das gilt für **den Scan-Vorgang**. Wenn später jemand mit dem KI-Agenten arbeitet, gehen die zur Frage passenden Ausschnitte sehr wohl nach außen an ein großes Sprachmodell. Was dabei genau hinausgeht, wohin, und wie sich das regeln lässt, steht im Hauptteil im Kapitel „Datenschutz und Vertrauen“.