

# Wissensdatenbank (RAG) aufbauen

- [Was ist Sinn und Zweck einer KI-Gestützten Wissensdatenbank](#)
- [Warum wir Word-Dateien vor dem Hochladen in die Wissensdatenbank umwandeln](#)
- [Word → Markdown: Checkliste für saubere Konvertierungen](#)

# Was ist Sinn und Zweck einer KI-Gestützten Wissensdatenbank

# Warum wir Word-Dateien vor dem Hochladen in die Wissensdatenbank umwandeln

Eine KI-gestützte Wissensdatenbank soll euch schnelle, präzise Antworten liefern – aber nur, wenn sie die Inhalte eurer Dokumente auch wirklich versteht. Genau hier liegt das Problem mit klassischen Word-Dateien.

## Das Problem: optische Formatierung ist für die KI "Rauschen"

Word-Dokumente enthalten viele Elemente, die für das menschliche Auge hilfreich sind, für eine KI aber Störfaktoren darstellen:

- **Tabellen** – besonders verschachtelte oder mit verbundenen Zellen – werden beim Einlesen oft in eine unlogische Reihenfolge zerlegt. Eine Tabelle mit Produktname, Preis und Verfügbarkeit kann so zu einer Zeile werden, in der Preis und Produktname vertauscht erscheinen.
- **Kopf- und Fußzeilen** (z. B. Seitenzahlen, Firmenname, Datum) werden teils mitten im Fließtext eingefügt und unterbrechen einen zusammenhängenden Absatz.
- **Textfelder** liegen technisch "über" dem eigentlichen Text. Viele Auswerteprogramme überspringen sie komplett – wichtige Inhalte, die in einer Infobox oder einem Kasten stehen, gehen dadurch schlicht verloren.
- **Verschachtelte Aufzählungen** (Listen mit mehreren Ebenen, Unterpunkten, Einrückungen) verlieren beim Einlesen häufig ihre Hierarchie – aus "Hauptpunkt mit drei Unterpunkten" wird eine flache Liste ohne erkennbare Zusammengehörigkeit.

- **Eingebettete Bilder und Sonderzeichen** können die Textstruktur durcheinanderbringen, etwa wenn ein Bild mitten im Satz platziert ist und der Text danach an einer unerwarteten Stelle "wieder einsetzt".

Für Menschen ist das alles selbstverständlich lesbar. Eine KI muss diese Struktur aber erst technisch entschlüsseln – und tut das nicht immer zuverlässig.

# Die Lösung: Umwandlung in ein einfacheres Format (z. B. Markdown)

Wird ein Dokument vorher in ein reduziertes Format wie Markdown umgewandelt, bleibt nur der reine Inhalt übrig: klare Überschriften, saubere Absätze und eindeutig getrennte Abschnitte. Die KI kann dadurch viel besser erkennen,

- wo ein Thema anfängt und wo es aufhört,
- was die wichtigsten Punkte sind,
- und wie einzelne Informationen inhaltlich zusammenhängen.

**Praxisbeispiel:** Eine mehrseitige Word-Datei mit Firmenlogo im Kopfbereich, einer verschachtelten Tabelle zu Ansprechpartnern und einem Textfeld mit "Wichtiger Hinweis" liefert nach der Umwandlung einen klar gegliederten Text mit Überschriften pro Abschnitt, einer einfachen Tabelle ohne Formatierungscode und dem Hinweistext direkt im Fließtext, statt in einer separaten, möglicherweise unsichtbaren Box.

# Das Ergebnis für eure Fragen an die Wissensdatenbank

Wenn später jemand eine Frage stellt, findet die KI die passende Antwort schneller und genauer, weil sie nicht erst durch Formatierungs-Ballast "hindurchsuchen" muss. Zusätzlich verhindert die Umwandlung typische Fehlerquellen:

- Tabellen, die durch die Umwandlung unübersichtlich werden,
- Textfelder oder Kopf-/Fußzeilen, die komplett ignoriert werden,
- Sonderzeichen oder eingebettete Bilder, die den Lesefluss der KI unterbrechen.

# Nicht jedes Dokument ist gleich betroffen

Ein einfaches Word-Dokument mit reinem Fließtext, wenigen Überschriften und ohne Tabellen, Textfelder oder komplexe Layouts bereitet in der Regel auch ohne Umwandlung keine größeren Probleme. Der Umwandelungsschritt lohnt sich vor allem bei Dokumenten mit:

- Tabellen,
- mehrspaltigem Layout,
- Kopf-/Fußzeilen mit relevantem Inhalt,
- Textfeldern oder Infoboxen,
- oder stark verschachtelten Aufzählungen.

## Fazit

Der zusätzliche Umwandelungsschritt kostet ein paar Minuten, sorgt aber dafür, dass eure Dokumente in der Wissensdatenbank so ankommen, wie sie gemeint sind – klar strukturiert und für die KI optimal lesbar. Der konkrete Nutzen für euch: weniger falsche oder unvollständige Antworten, weil die KI eure Inhalte von Anfang an richtig erfasst.

# Word → Markdown:

## Checkliste für saubere Konvertierungen

**Kernerkenntnis:** Fast alle Konvertierungsprobleme entstehen dadurch, dass Word visuelle Formatierung statt semantischer Struktur nutzt. Konverter wie Pandoc lesen Formatvorlagen (Styles) – nicht Optik. Was optisch wie eine Überschrift aussieht, aber nur fett und größer formatiert wurde, bleibt im Markdown ein normaler Absatz.

Vor dem Export eines Word-Dokuments nach Markdown diese Punkte prüfen:

---

### 1. Formatvorlagen statt manueller Formatierung

Überschrift 1/2/3 verwenden statt Text nur fett und größer zu setzen. Konverter erkennen ausschließlich Formatvorlagen und wandeln sie in #, ##, ### um – manuell "nachgebaute" Überschriften bleiben normaler Fließtext. **So geht's:** Start → Formatvorlagen-Galerie → „Überschrift 1“ (nicht Schriftgröße/Fett manuell setzen).

### 2. Einheitliche Überschriften-Hierarchie

Keine Ebenen überspringen (z. B. direkt von Überschrift 1 zu Überschrift 3). Markdown bildet die Hierarchie strikt ab – ein Sprung erzeugt eine falsche Gliederungstiefe im Wiki-Artikel. **So geht's:** Vor dem Export über die Navigationsleiste (Ansicht → Navigationsbereich) die Gliederungsebenen prüfen und lückenlos durchnummerieren lassen.

### 3. Änderungen nachverfolgen abschließen

Track-Changes vor dem Export annehmen oder ablehnen, sonst landen Löschungen und Einfügungen doppelt im Markdown. **So geht's:** Überprüfen → Nachverfolgung → Alle Änderungen annehmen.

## 4. Kommentare entfernen

Kommentare werden sonst als Fußnoten oder Störtext mitkonvertiert. **So geht's:** Überprüfen → Kommentare → Alle Kommentare löschen.

## 5. Keine Textfelder oder eingebetteten Objekte

Inhalte in Textfeldern, SmartArt oder OLE-Objekten werden von Konvertern oft komplett ignoriert oder nur als Bild gerendert – der Text darin geht verloren. **So geht's:** Textfelder vermeiden, Inhalt direkt in den Fließtext schreiben.

## 6. Tabellen einfach halten

Keine verschachtelten Tabellen, verbundenen Zellen oder Tabellen-in-Tabellen, da Markdown nur einfache Raster-Tabellen abbildet. **So geht's:** Tabelle einfügen → einfaches Raster, „Zellen verbinden“ nur wenn unbedingt nötig.

## 7. Bilder mit Alt-Text versehen

Ohne Alt-Text fehlt im Markdown die Bildbeschreibung (`![]()`) komplett. **So geht's:** Bild markieren → rechte Maustaste → Alternativtext bearbeiten.

## 8. Einheitliche Listenformate

Aufzählungs- und Nummerierungsvorlagen verwenden statt manueller Bindestriche oder eigener Nummerierung per Hand – sonst erkennt der Konverter keine zusammenhängende Liste. **So geht's:** Start → Aufzählungszeichen bzw. Nummerierung aus der Galerie wählen, nicht `-` oder `1.` von Hand eintippen.

## 9. Code/Monospace-Text als Zeichenformat auszeichnen

Für Befehle, Dateinamen oder Codeschnipsel das Zeichenformat „HTML Code“ bzw. eine Monospace-Formatvorlage nutzen statt nur die Schriftart manuell auf Courier zu ändern. Nur echte Formatvorlagen werden zu ``Code`` im Markdown. **So geht's:** Text markieren → Formatvorlage

„HTML Code" (oder eigene Zeichenformatvorlage) zuweisen.

## 10. Echte Hyperlinks statt reinem Text

Links als Word-Hyperlink einfügen (Strg+K), nicht nur die URL als Text schreiben. Nur echte Hyperlinks werden zu `[Text] (URL)` – reiner Text bleibt unverlinkt. **So geht's:** Text markieren → Einfügen → Link → Adresse eintragen.

## 11. Sparsam mit Sonderzeichen und typografischen Zeichen

Automatisch ersetzte typografische Anführungszeichen, Gedankenstriche oder Symbole können je nach Konverter-Einstellung uneinheitlich im Markdown landen. Wo es auf exakte Zeichen ankommt (z. B. Befehle, Pfade), einfache Anführungszeichen/Standardzeichen verwenden. **So geht's:** Datei → Optionen → Dokumentprüfung → AutoKorrektur-Optionen → „Anführungszeichen ersetzen" bei Bedarf deaktivieren.

## 12. Keine manuellen Zeilenumbrüche (Shift+Enter)

Zeilenumbrüche innerhalb eines Absatzes statt eines neuen Absatzes erzeugen im Markdown keine saubere Absatztrennung und werden je nach Konverter verschluckt oder falsch dargestellt. **So geht's:** Enter für neue Absätze verwenden, Shift+Enter vermeiden.

## 13. Keine Spalten-, Abschnittslayouts oder Fuß-/Endnoten

Mehrspaltige Layouts, Abschnittsumbrüche und Fuß-/Endnoten werden von Konvertern meist an falscher Stelle im Fließtext eingefügt oder ganz verworfen. **So geht's:** Einspaltiges Standardlayout verwenden, Anmerkungen direkt im Fließtext oder als eigener Absatz formulieren statt als Fußnote.

## 14. Ein Thema pro Dokument

Ein eigenes .docx pro Anleitung/Prozess statt Sammeldokumente – vermeidet Vermischung von Überschriftenebenen und erleichtert die spätere Aufteilung in einzelne Wiki-Artikel. **So geht's:** Für jede Anleitung eine eigene Datei anlegen statt mehrere Themen in einem Dokument zu sammeln.

---

**Kurzfassung zum Merken:** Alles, was in Word nur "gut aussieht", aber keine echte Formatvorlage, kein echtes Werkzeug (Link, Formatvorlage, Alt-Text) dahinter hat, geht bei der Konvertierung verloren oder wird falsch interpretiert.