



## Technische Informationen Sage Copilot

### Wie funktioniert Sage Copilot?

- **Verifizierte DPW-Quellen:** Es werden ausschließlich DPW-spezifische Inhalte verarbeitet, die von Sage DPW Expert\*innen ausgewählt wurden.
- **Versteht Kundenanfragen:** Sage Copilot erkennt die Bedeutung von Fragen und gleicht sie mit vorhandenen Inhalten ab.
- **Minimiert falsche Antworten:** Durch eine Technik namens RAG (Retrieval Augmented Generation) werden Fehler und „Halluzinationen“ reduziert.
- **Erzeugt menschenähnliche Antworten:** Mithilfe von großen Sprachmodellen (LLMs) werden Antworten formuliert, die wie von einer echten Person wirken.
- **Optimierte Fragestellungen (Prompt Engineering):** Je besser die Eingabe (Prompt) formuliert ist, desto besser und relevanter ist die Antwort von Sage Copilot.

### Wie werden Dokumente verarbeitet?

- Dokumente werden von Sage DPW in die Sage DPW Cloud hochgeladen, in kleinere Abschnitte zerlegt und in eine Form gebracht, welche die KI verstehen kann (sogenannte „Vektoren“).
- Diese Informationen werden in einer speziellen Datenbank gespeichert, die nach Ähnlichkeiten sucht.
- Wenn ein User eine Frage stellt, wird diese ebenfalls in einen Vektor umgewandelt. Dann sucht Sage Copilot nach den relevantesten Inhalten und erstellt daraus eine Antwort.

### Welche Technologien stecken dahinter?

- **Amazon Titan:** Modell für die Umwandlung von Text in Vektoren (Embeddings).
- **AWS (Amazon Web Services) OpenSearch:** Datenbank für die Suche nach ähnlichen Inhalten (Similarity Search).
- **Anthropic Claude Haiku:** Sage Copilot basiert auf Anthropic Claude Haiku, einem KI-Sprachmodell (Large Language Model, LLM), das Texte versteht, Fragen beantwortet und Inhalte zusammenfasst.
- **LangChain:** Framework zur Erleichterung der Integration von einzelnen Komponenten (vom Hochladen der Dokumente über die Nutzeranfragen bis hin zur passenden Antwort).
- **DynamoDB:** Speicherung von Chats und Sitzungen für Analysen und Produktverbesserungen.