



AI IN A NUTSHELL 1

Anbindung eines RAG-Subsystems an ein KI-System mit Sprachmodell

Ausgangslage

Einem KI-System implementierte Sprachmodelle nutzen bei der Reaktion auf Prompts grundsätzlich nur Daten, mit denen sie trainiert sind, sowie Informationen aus Nutzenden-prompts, Systemprompts und gegebenenfalls der Historie. Durch „Retrieval Augmented Generation“ (RAG) werden sie in die Lage versetzt, weitere (außerhalb des Sprachmodells liegende) Quellen wie Dokumentensammlungen, Datenbanken oder einen Knowledge Graph heranzuziehen (sogenanntes RAG-Subsystem).

Hierfür werden diese Datenquellen zunächst aufbereitet, beispielsweise durch Aufteilung in kürzere Absätze (Chunks), Überführung dieser Chunks in Vektoren mittels eines KI-Embedding-Modells und Speicherung der Vektoren in einer Vektordatenbank.

Wann immer Nutzende dann Prompts eingeben, wird die Anfrage im RAG-Subsystem mittels des Embedding-Modells ihrerseits in einen Anfragevektor überführt und an den sogenannten Retriever übergeben, der die Vektorbank nach denjenigen Vektoren durchsucht, welche eine möglichst geringe Distanz (das heißt große semantische Nähe) zum Anfragevektor haben.

Anschließend wird die Eingabe der oder des Nutzenden um das Suchergebnis des Retrievers, das heißt die Informationen aus den angebundenen Datenquellen, ergänzt, bevor sie vom RAG-Subsystem an ein KI-System mit Sprachmodell weitergegeben wird. Dieses soll ausschließlich (technisch derzeit noch nicht erreichbarer Idealfall) beziehungsweise jedenfalls vorrangig nur diese relevanten Inhalte als Kontext für die Beantwortung des Prompts nutzen und das KI-Sprachmodell zur reinen Textgenerierung ohne Rückgriff auf „eigenes“, im Training erlerntes Wissen verwenden.

Das hier vereinfacht dargestellte Gesamtsystem aus KI-System mit Sprachmodell und RAG-Subsystem wird als RAG-System bezeichnet.

Einsatzszenario

Die Anbindung eines RAG-Subsystems an ein KI-System mit Sprachmodell ist besonders geeignet für die schnelle und unmittelbare Zugänglichmachung externer, auch behördeninterner, und/oder sich häufig verändernder Datenquellen, da diese bei Anbindung mittels RAG-Subsystems – im Gegensatz zum Initialtraining oder Fine-Tuning – nicht in das Sprachmodell selbst integriert werden.

Die Datenquellen können dadurch nicht nur leichter angepasst oder ausgetauscht werden; bei abgeschottetem beziehungsweise strikt mandantengetrenntem Betrieb und entsprechender Konfiguration ist auch keine Preisgabe interner Informationen an weitere Nutzende desselben Sprachmodells zu befürchten.

Zentrale KI-datenschutzrechtliche Fragestellungen zur Beachtung bei Beschaffung und Nutzung von RAG-Subsystemen

Hinweis: Zur Anbindung eines RAG-Subsystems an ein KI-System mit Sprachmodell allgemein ausführlich Bayerischer Landesbeauftragter für den Datenschutz, Datenschutz bei KI-Projekten in der bayerischen Verwaltung, Orientierungshilfe, Stand 3/2026, Rn. 178 ff. m. w. N., im Internet abrufbar unter <https://www.datenschutz-bayern.de/ki> (OH KI). Zu den spezifischen Fragestellungen im Folgenden siehe die einzelnen Verweise.

[1] Personenbezogene Daten in mittels RAG angebundenen Quellen

Thematik: Enthalten die mittels RAG angebundenen Datenquellen personenbezogene Daten, findet auch auf diese das Datenschutzrecht Anwendung. Dies gilt sowohl in Bezug auf die Datenverarbeitungen im RAG-Subsystem (zum Beispiel die Erhebung der Datenquellen [vgl. OH KI, Rn. 156 ff.], die Aufbereitung der Datenquellen [vgl. OH KI, Rn. 159] und die Vektorisierung [siehe 2]) als auch hinsichtlich der Verarbeitungen im RAG-System insgesamt (beispielsweise die Verarbeitung des um die Informationen aus der Vektordatenbank ergänzten Prompts mittels des Sprachmodells [vgl. OH KI, Rn. 175 – Eingabe – und OH KI, Rn. 176 ff. – Verarbeitung innerhalb eines KI-Systems] oder die Ergebnisausgabe durch das KI-System mit Sprachmodell [OH KI, Rn. 187 f.]).

Insbesondere erforderliches Tätigwerden:

- ▶ Prüfung der anzubindenden Datenquellen auf Personenbezug; gezielte Auswahl und Kuratierung der Datenquellen, gegebenenfalls Anonymisierung;

Praxishinweis: Sowohl in Word- als auch in PDF-Dateien finden sich – unabhängig vom Inhalt – in den Metainformationen regelmäßig Autorennamen, die für sich genommen bereits personenbezogene Daten darstellen.

- ▶ Prüfung und Abgrenzung sämtlicher Verarbeitungsvorgänge sowohl innerhalb des RAG-Subsystems als auch innerhalb des RAG-Systems insgesamt; insoweit sind regelmäßig Informationen des Anbieters des jeweiligen KI-Systems/KI-Modells (hier: zumindest KI-Embedding-Modell, KI-System mit Sprachmodell) erforderlich.

[2] KI-gestütztes Embedding

Thematik: Für das KI-gestützte Embedding, das heißt die Überführung der Chunks in Vektoren, innerhalb des RAG-Subsystems kommen regelmäßig spezialisierte KI-Embedding-Modelle (oft auf Sprachmodellen basierend) zum Einsatz. Enthalten die Datenquellen personenbezogene Daten (siehe 1), bewirkt das KI-gestützte Embedding eine Verarbeitung dieser personenbezogenen Daten mittels KI (vgl. OH KI, Rn. 175 –

Eingabe – und OH KI, Rn. 176 ff. – Verarbeitung innerhalb eines KI-Systems). Die daraus entstehenden Vektoren sind regelmäßig weiterhin als personenbezogene Daten zu betrachten, da aus ihnen über semantische Ähnlichkeitssuche Rückschlüsse auf die zugrundeliegenden Inhalte gezogen werden können.

Insbesondere erforderliches Tätigwerden:

- ▶ Verhinderung möglicher Angriffe und daraus folgend (unberechtigter) Zugriffe auf die Vektordatenbank oder Rekonstruktionen aus dieser Datenbank (vgl. OH KI, Rn. 215);
- ▶ Berücksichtigung der Verwendung mehrerer KI-Modelle innerhalb des RAG-Systems in der Risikobetrachtung (vgl. OH KI, Rn. 349 ff.);
- ▶ Bewertung der Datenschutzkonformität des eingesetzten KI-Systems, sofern dieses bereits zum Zeitpunkt der Bereitstellung personenbezogene Daten enthält (vgl. OH KI, Rn. 166 ff.).

[3] Informationspflichten und Betroffenenrechte

Thematik: Da in einem RAG-System die Inhalte wesentlich über das RAG-Subsystem gesteuert werden sollen und das im KI-System implementierte Sprachmodell (vorrangig) nur zur Textgenerierung verwendet werden soll, ist es im Einzelfall auch möglich, ein RAG-Subsystem (mit gegebenenfalls personenbezogenen Datenquellen) an ein KI-System mit Sprachmodell ohne Personenbezug beziehungsweise mit anonymisiertem Sprachmodell anzubinden.

Insbesondere erforderliches Tätigwerden:

- ▶ Gewährleistung und Umsetzung der Informationspflichten (vgl. OH KI, Rn. 296ff.) und Betroffenenrechte (vgl. OH KI, Rn. 310 ff.) sowie Einhaltung der Verarbeitungsgrundsätze – insbesondere der Richtigkeit der Daten, ihrer Integrität und Vertraulichkeit sowie der Datenminimierung und Speicherbegrenzung (vgl. OH KI, Rn. 206 ff.) – können in Bezug auf die mittels RAG angeschlossenen externen Datenquellen durch den insoweit Verantwortlichen – beispielsweise die einsetzende öffentliche Stelle – gesteuert oder zumindest beeinflusst werden;
- ▶ Herausforderungen – und gegebenenfalls (vertraglich abzusichernder) Unterstützungsbedarf durch den Anbieter des jeweiligen KI-Systems/KI-Modells (hier: zumindest KI-Embedding-Modell, KI-System mit Sprachmodell) – bestehen dagegen weiterhin insbesondere hinsichtlich der Gewährleistung der Informationspflichten und Betroffenenrechte im Hinblick auf das RAG-System und die dort zum Einsatz kommenden KI-Modelle insgesamt sowie in Bezug auf die Grundsätze der Rechtmäßigkeit (OH KI, Rn. 195 f.), Transparenz (OH KI, Rn. 197 f.) und Zweckbindung (OH KI, Rn. 198 ff.).

Hinweis: Unabhängig von dem konkreten zur Anwendung kommenden Produkt und dem jeweiligen Einsatzszenario sind sämtliche Vorgaben des Datenschutzrechts und insbesondere der Datenschutz-Grundverordnung einzuhalten und nachzuweisen, sobald und soweit bayerische öffentliche Stellen personenbezogene Daten mittels Künstlicher Intelligenz verarbeiten (näher OH KI, Rn. 192 ff.).