Bot-Camp

Wissensbasierte KI-Assistenten an Hochschulen einsetzen

Einführung

Univ.-Prof. Dr. Malte Persike

persike@cls.rwth-aachen.de

+49 (170) 10 57 54 1

Zum Auftakt werden wir:

Verstehen, wann und warum wir Custom Chatbots brauchen.

Die Komponenten eine Custom Chatbot kennenlernen.

Risiken für die Qualität der Antworten demonstrieren.

KI für Heute

Das Team

CLS Center für Lehr- und Lernservices

WEST AI KI-Servicezentrum

OSKI.nrw

Ki:connect

Will Agentic AI Break Higher Education?

A new app offers to do students' work for them. Professors aren't ready.

ILLUSTRATION BY THE CHRONICLE; GETTY; WIKIMEDIA COMMONS

By Jason Gulya

March 3, 2026

Einstein first came to my attention on February 25. It was promoted as an app that logged into a student's Canvas account and completed all the work on their behalf. As advertised on their website, "Einstein is an AI with a computer. He logs into Canvas every day, watches lectures, reads essays, writes papers, participates in discussions, and submits your homework — automatically."

Quelle: https://www.chronicle.com/article/will-agentic-ai-break-higher-education

KI Agenten

auch Skills genannt

sind Custom Chatbots, die Schnittstellen mit anderen digitalen Systemen oder KI-Modellen haben.

Sie interagieren sowohl mit den Nutzenden als auch **mit anderen Softwarediensten** und können dort Aktionen auslösen.

Custom Chatbots

auch CustomGPT, Gem oder Model genannt

sind KI-Modelle, die auf bestimmte Antworttypen optimiert sind.

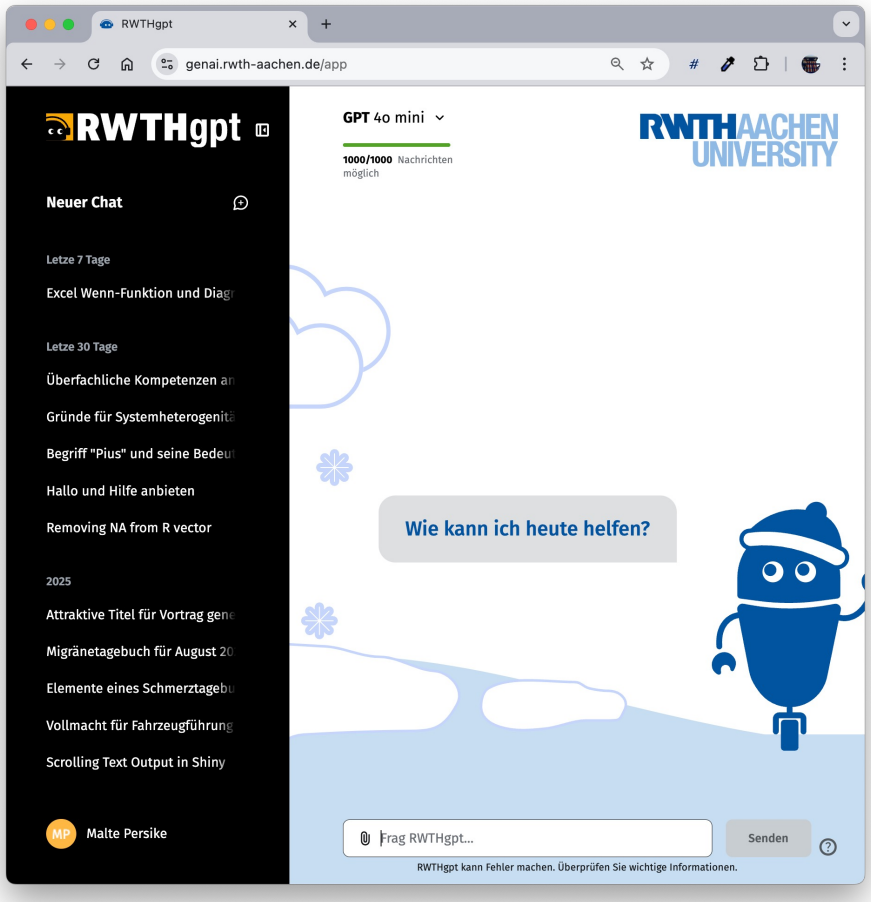
Sie interagieren nur mit den Nutzenden und können aus ihrem Chat-Interface **nicht „ausbrechen“**.

1

Warum Custom Chatbots?

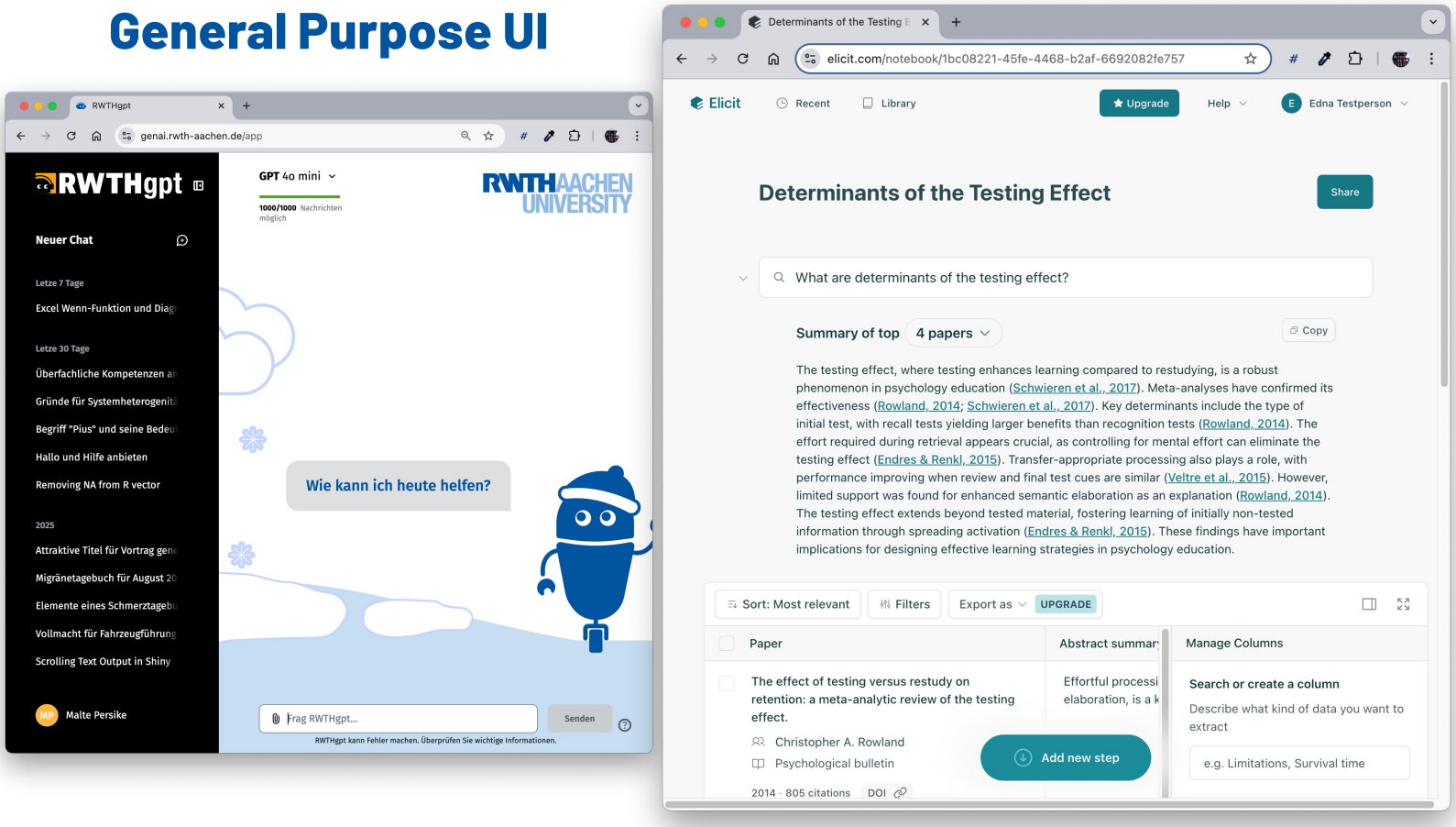
Source: https://unsplash.com/photos/Tggjbc4DD2g

Integrationsebenen von KI-Systemen



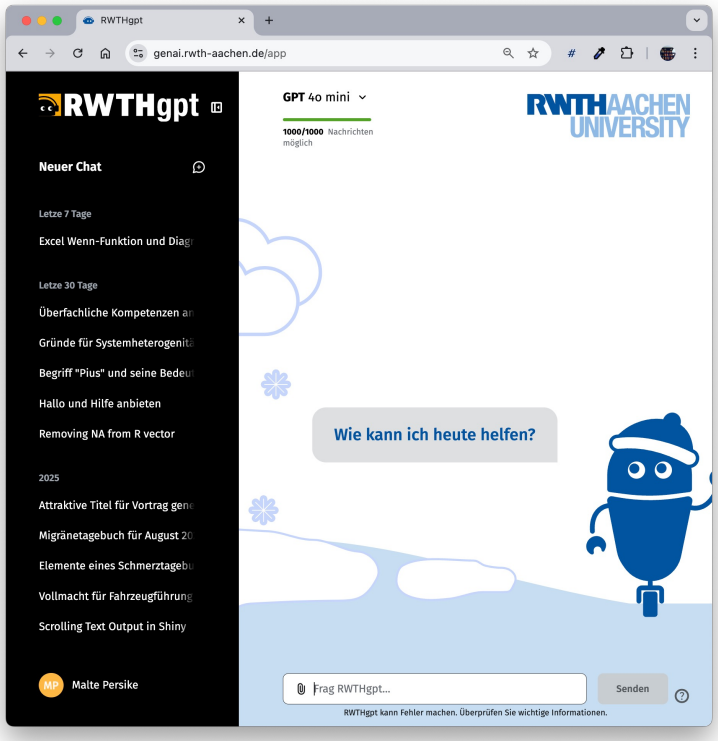
Integrationsebenen von KI-Systemen

General Purpose UI

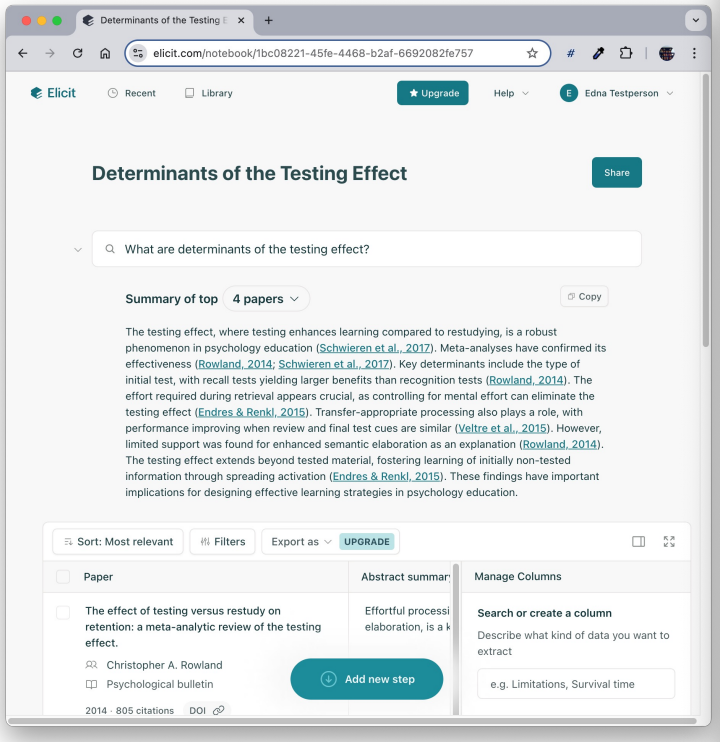


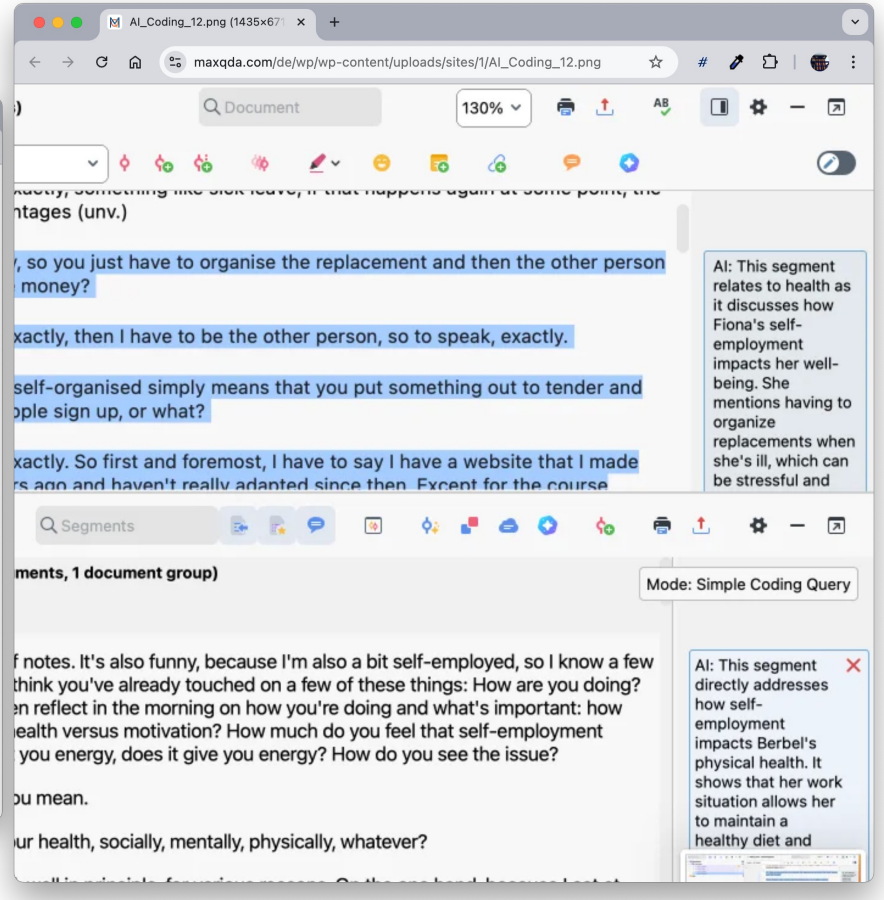
Integrationsebenen von KI-Systemen

General Purpose UI



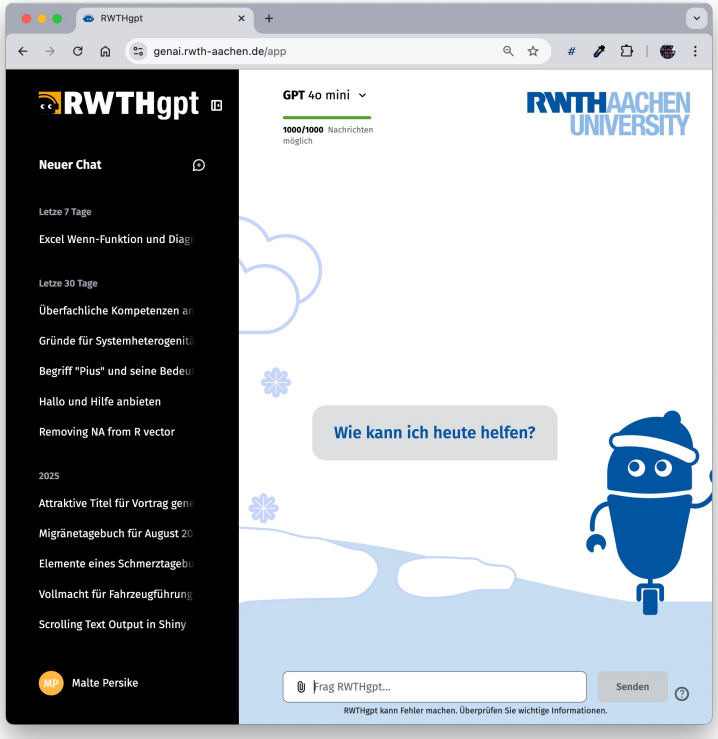
Special Purpose UI / Custom Chatbot



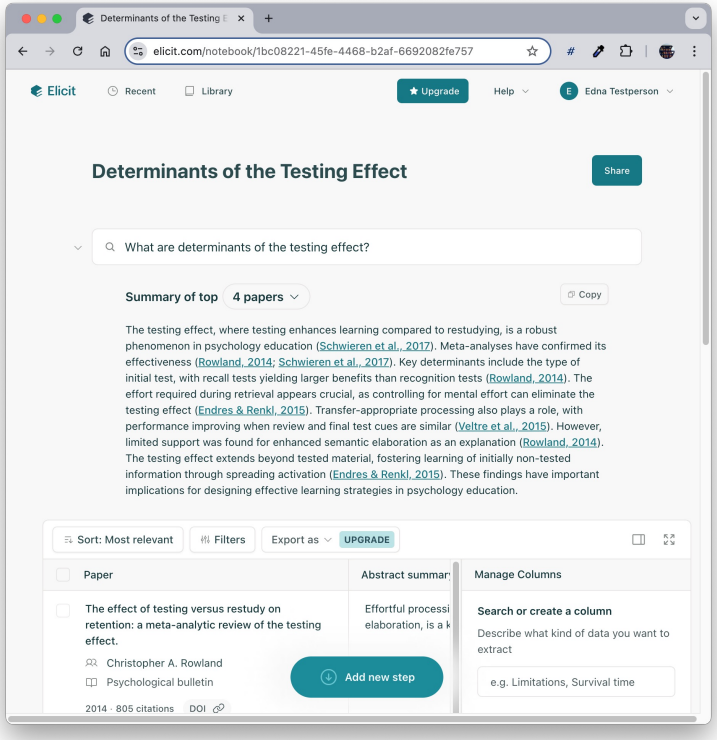


Integrationsebenen von KI-Systemen

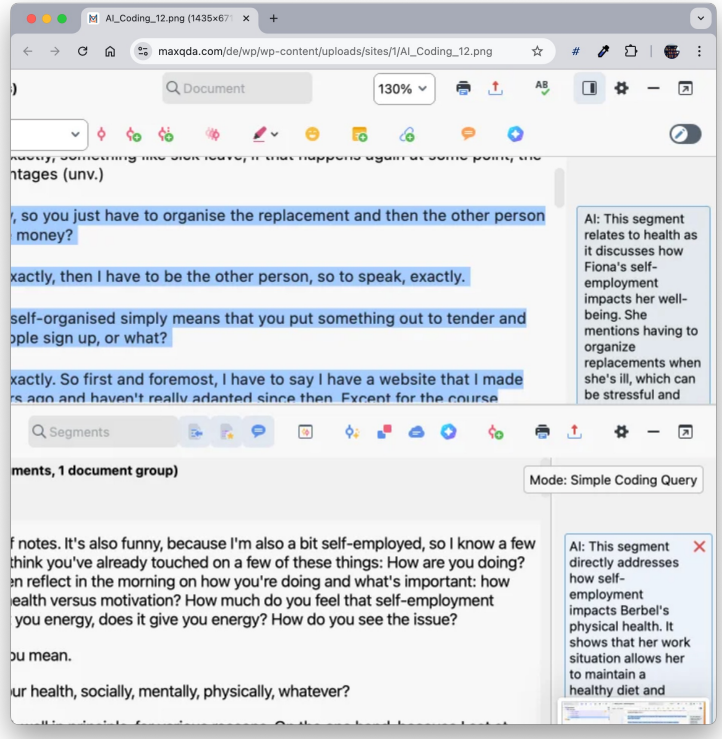
General Purpose UI



Special Purpose UI / Custom Chatbot



In-App UI



Starke Kompetenzerfordernisse

Schwache Kompetenzerfordernisse

KI-Systeme als Unterstützungswerkzeug für Alle

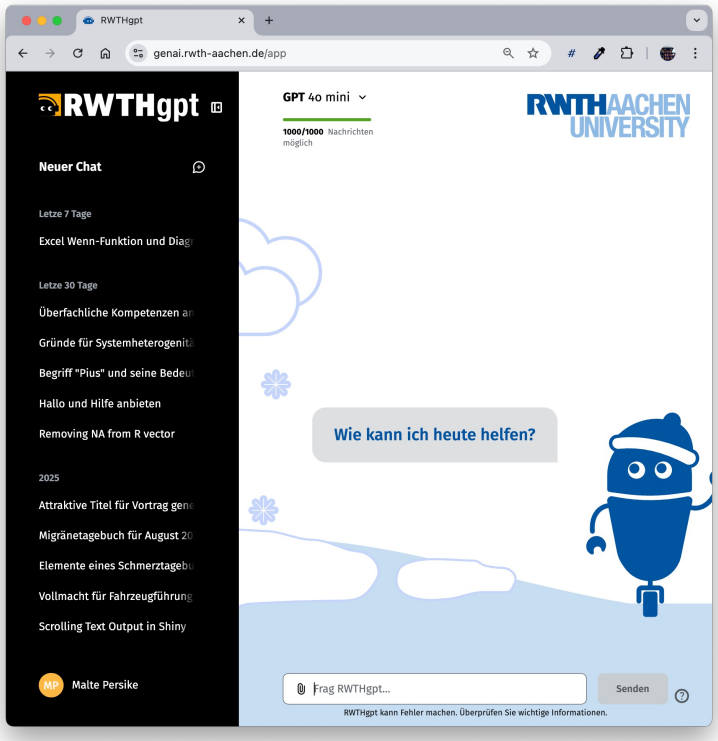
Variante 1: Selbständige Nutzung eines allgemeinen LLM wie z.B. ChatGPT

Variante 2: Nutzung eines allgemeinen LLM mit vorgefertigten Prompts und Dokumenten

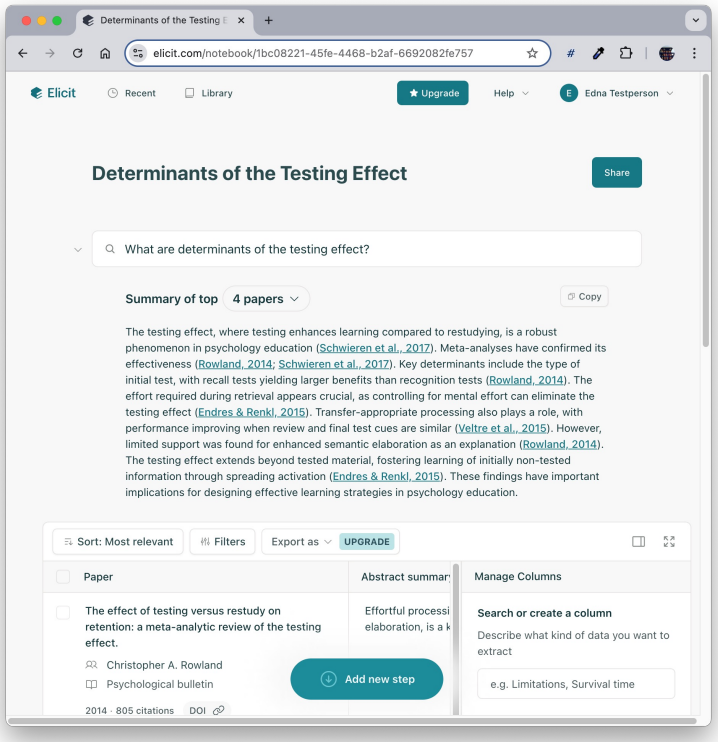
Variante 3: Nutzung von Custom Chatbots oder Spezialapplikationen mit eigens trainierten und instruierten LLMs

Integrationsebenen von KI-Systemen

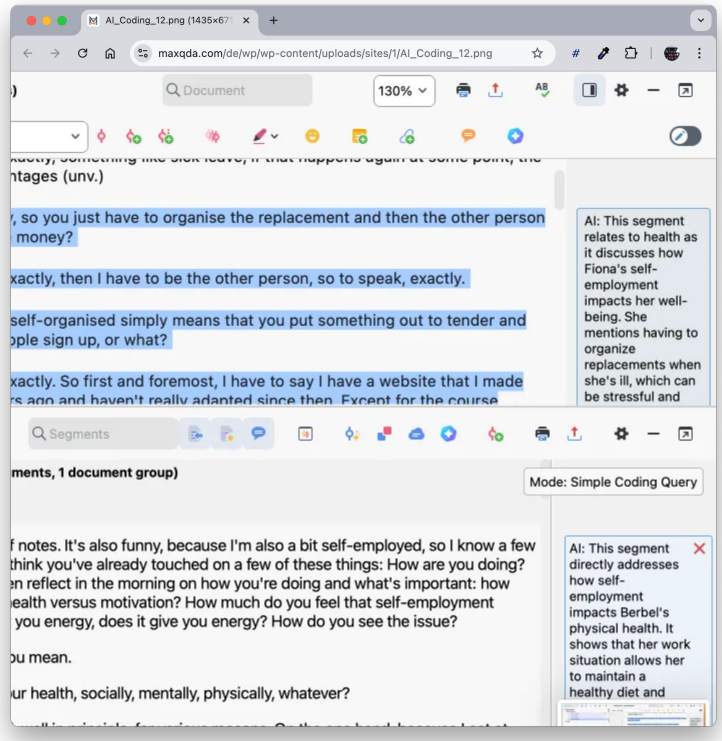
General Purpose UI



Special Purpose UI / Custom Chatbot



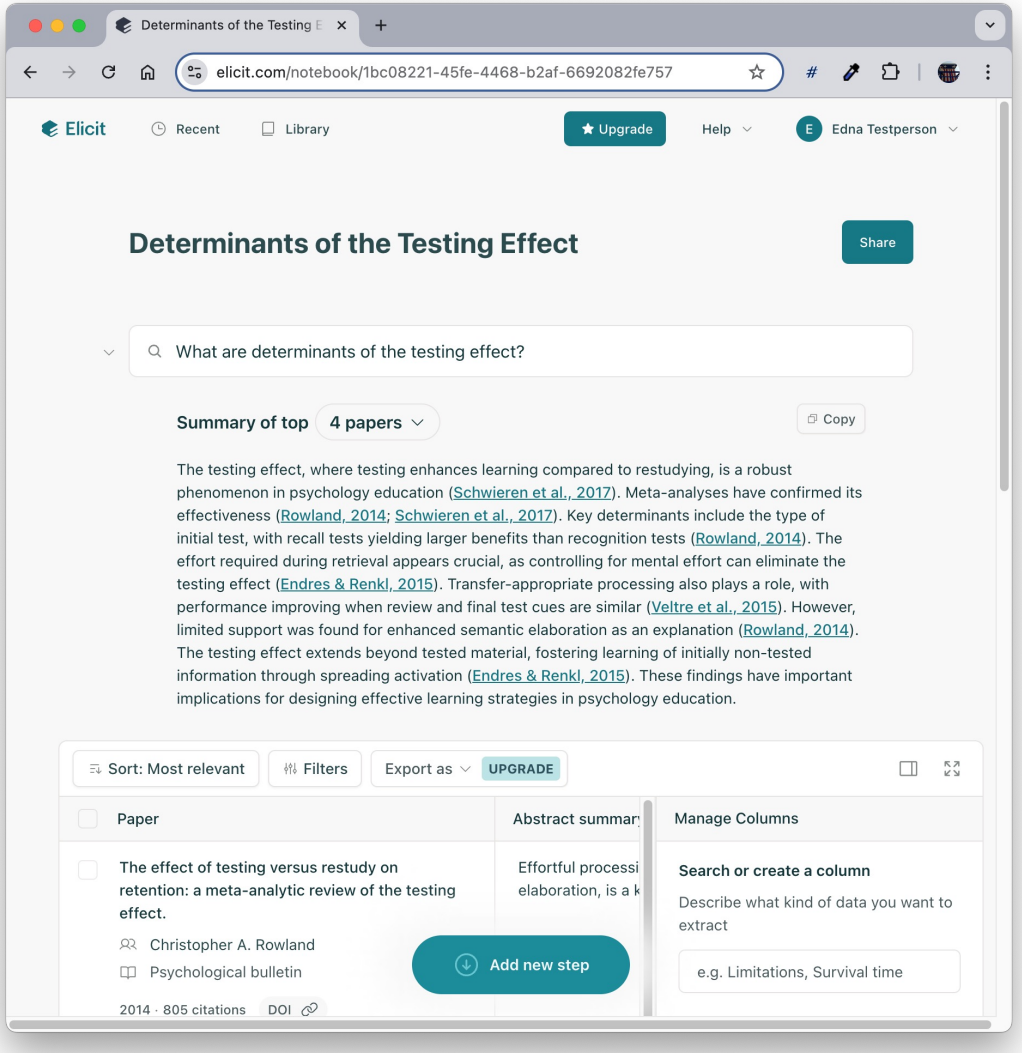
In-App UI



Spezialapplikation oder Custom Chatbot?

Gezeigt ist das Literaturrecherche-Werkzeug **Elicit** (https://elicit.com).

Im Kern ist auch eine solche Spezial-applikation kaum mehr als ein **Custom Chatbot** mit schicker Oberfläche.

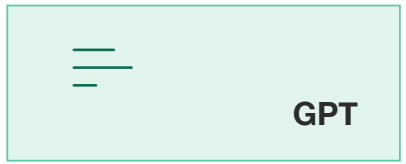


2

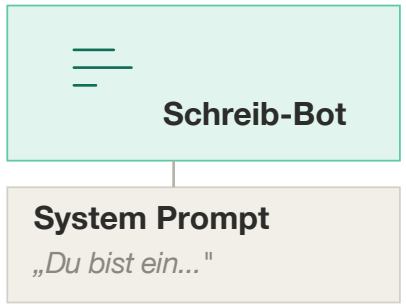
Komponenten eines Custom Chatbot

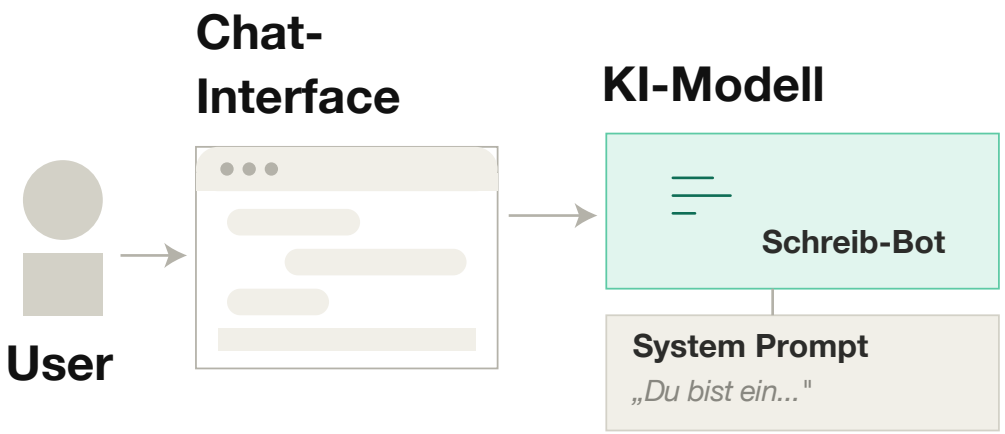
Bildquelle: https://unsplash.com/photos/1K9T5YiZ2WU

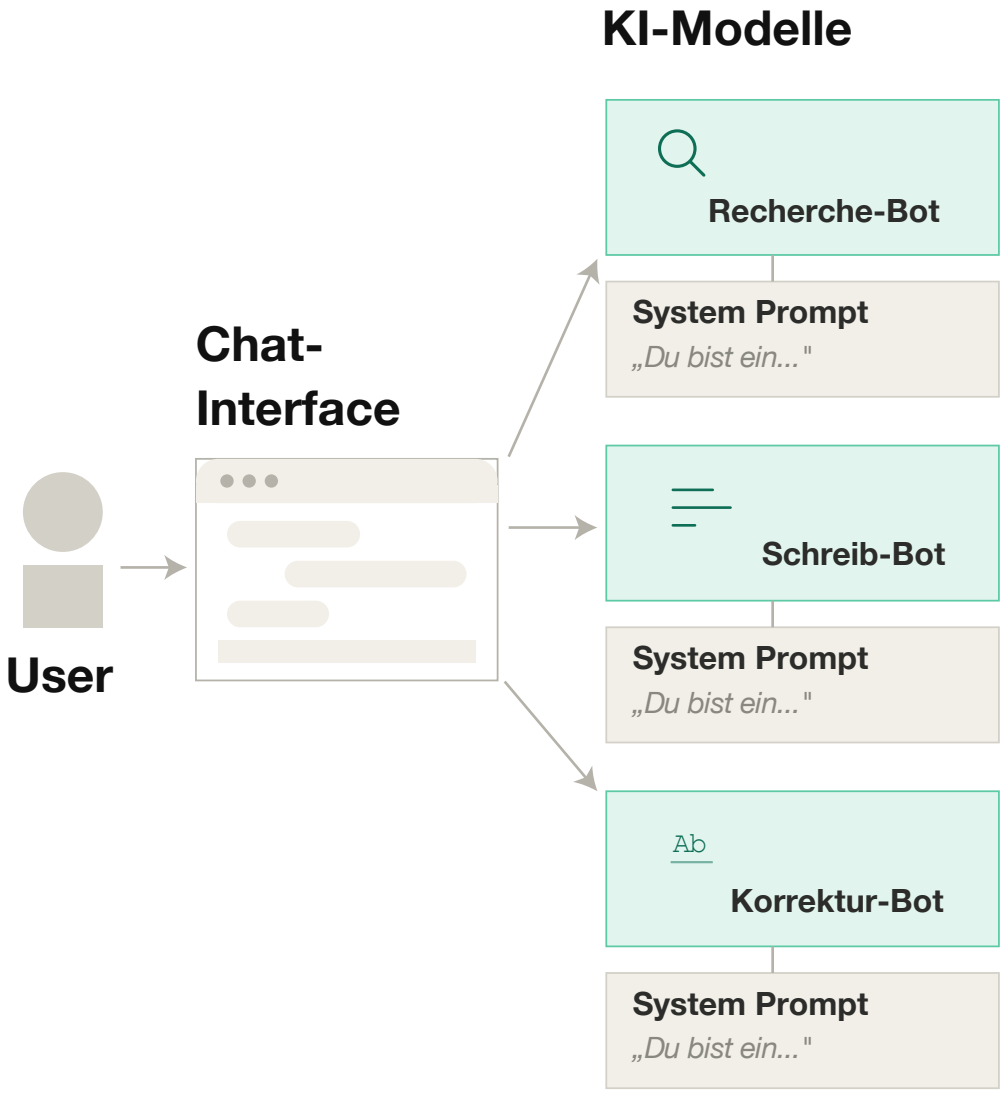
KI-Modell

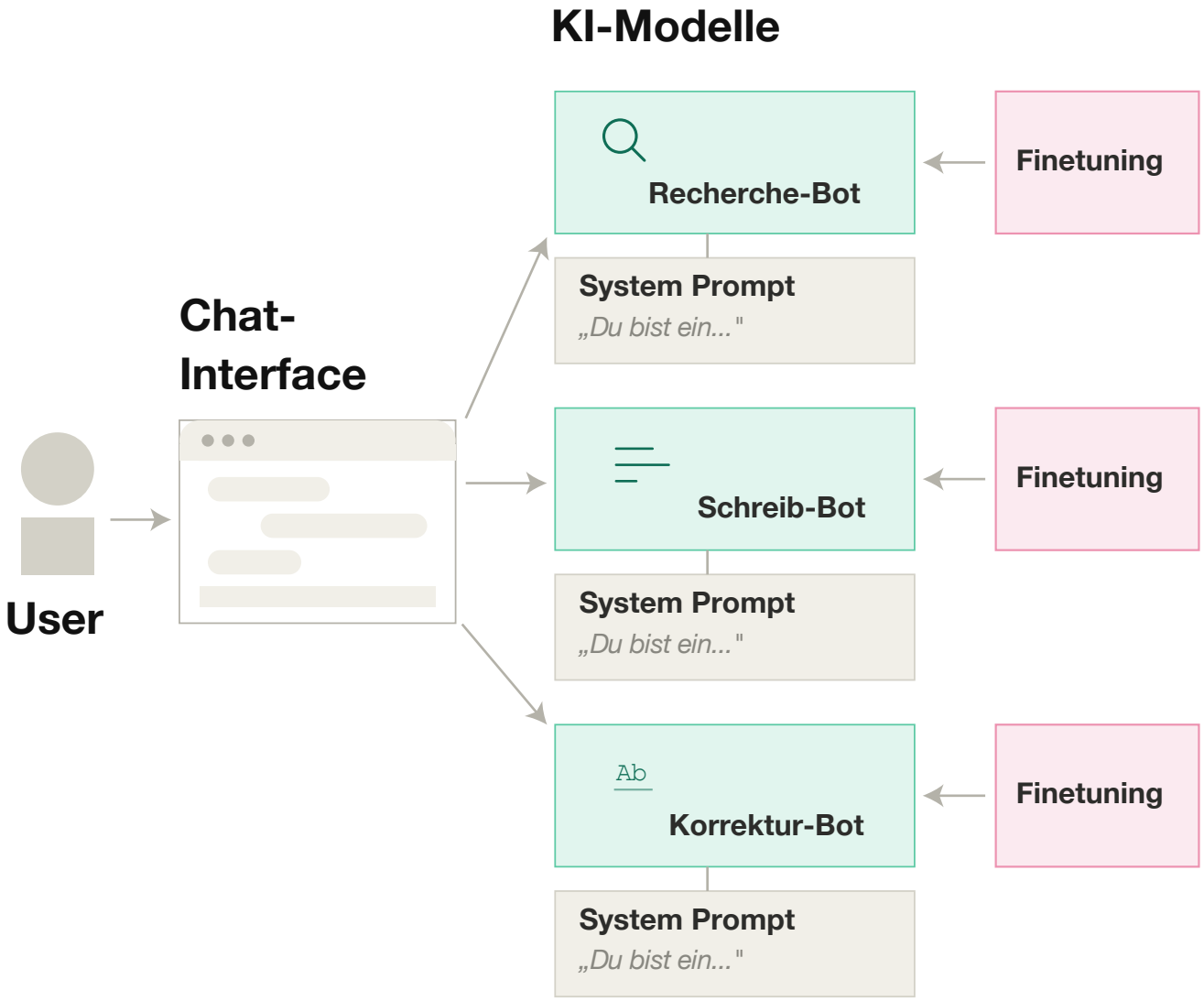


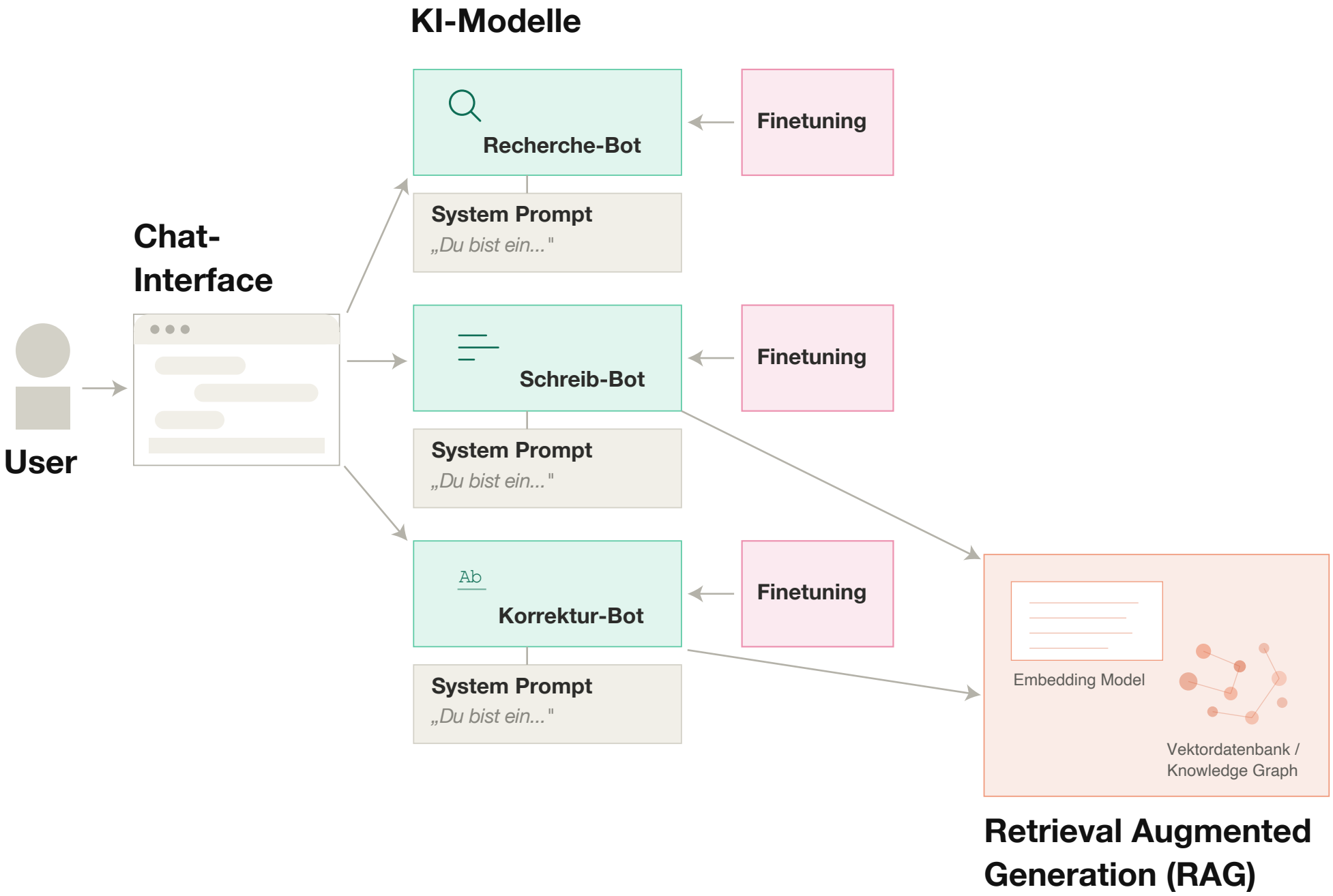
KI-Modell

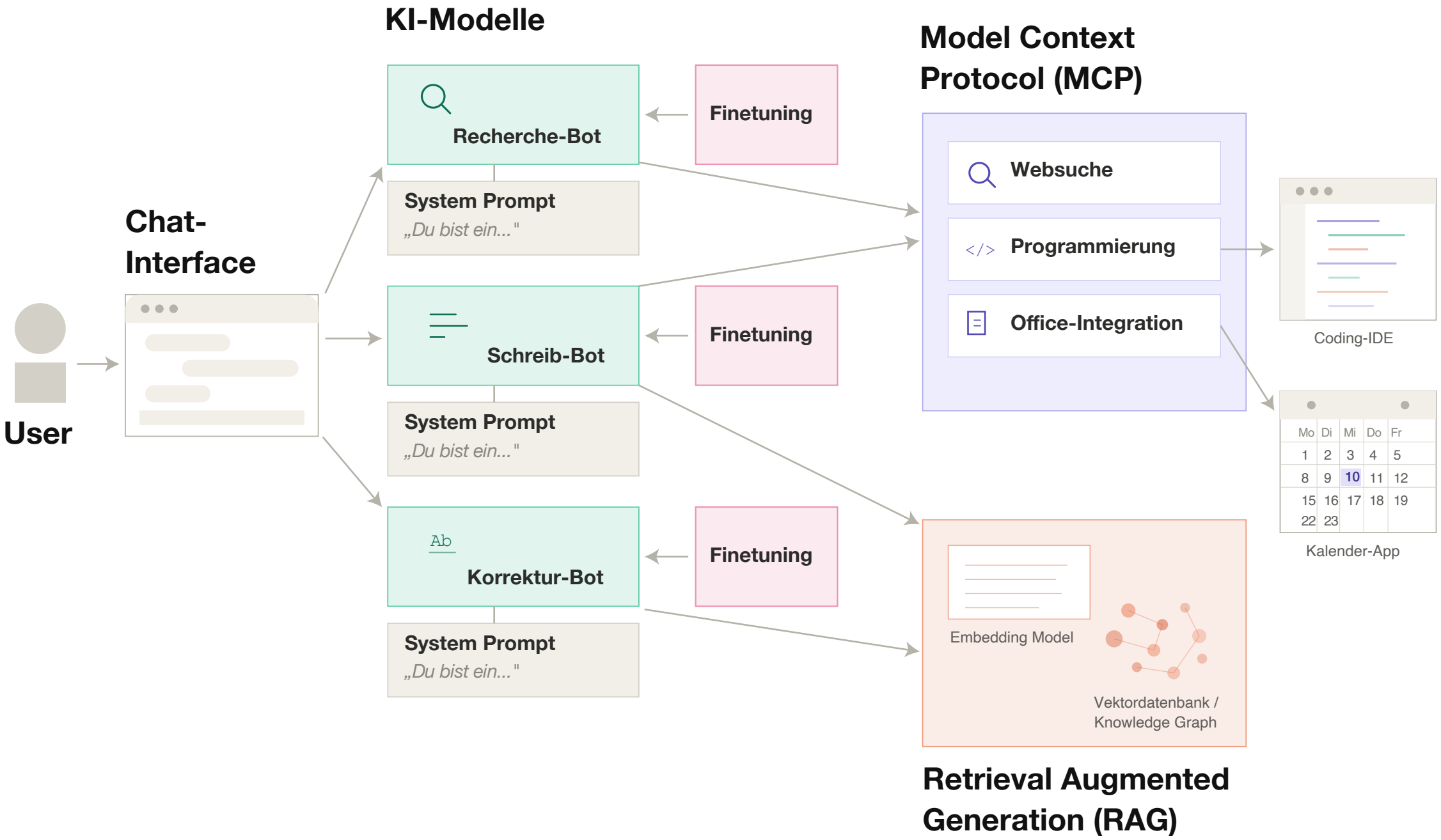


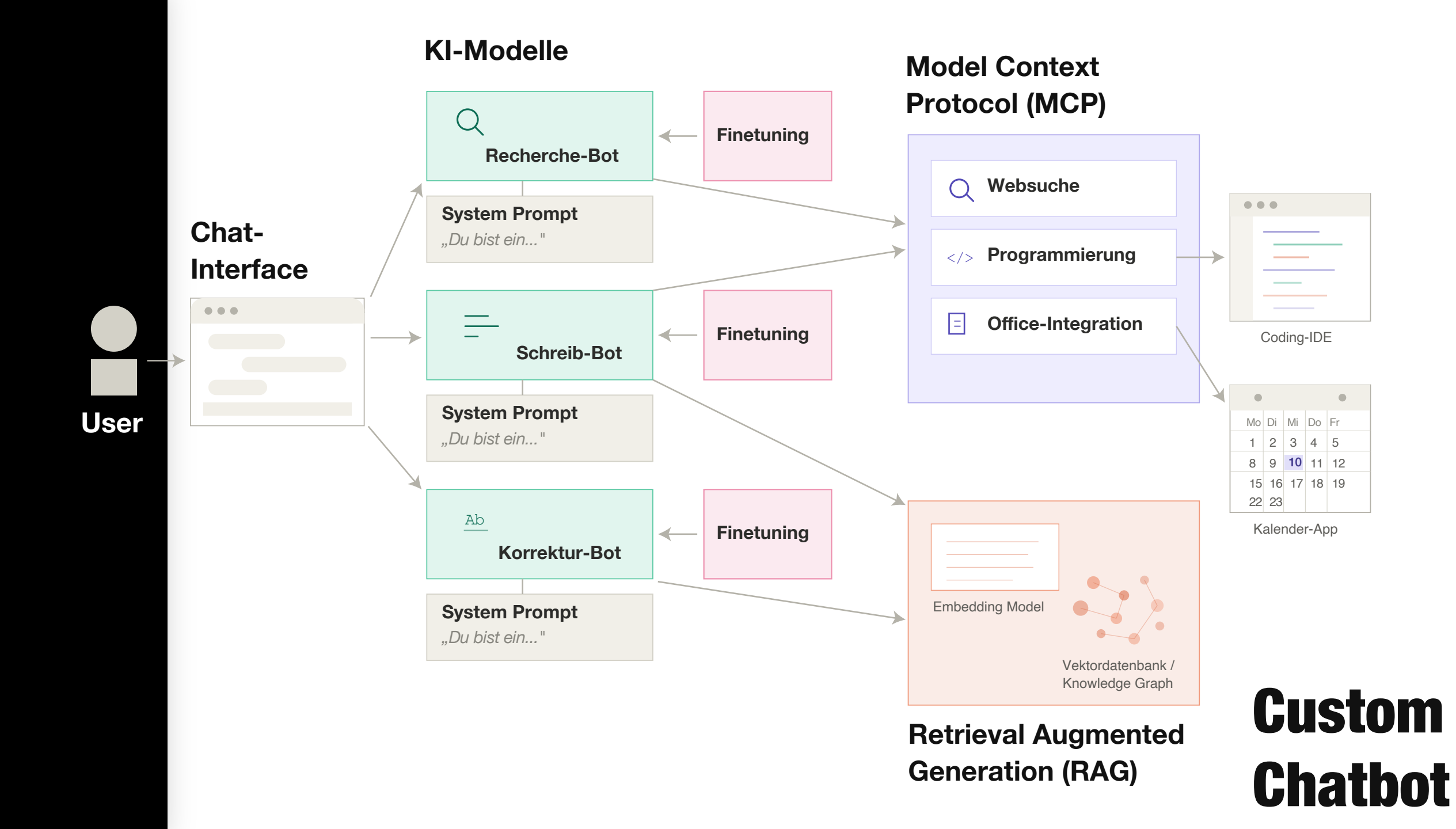

Chat-
Interface
KI-Modell
Schreib-Bot
System Prompt
„Du bist ein..."
User

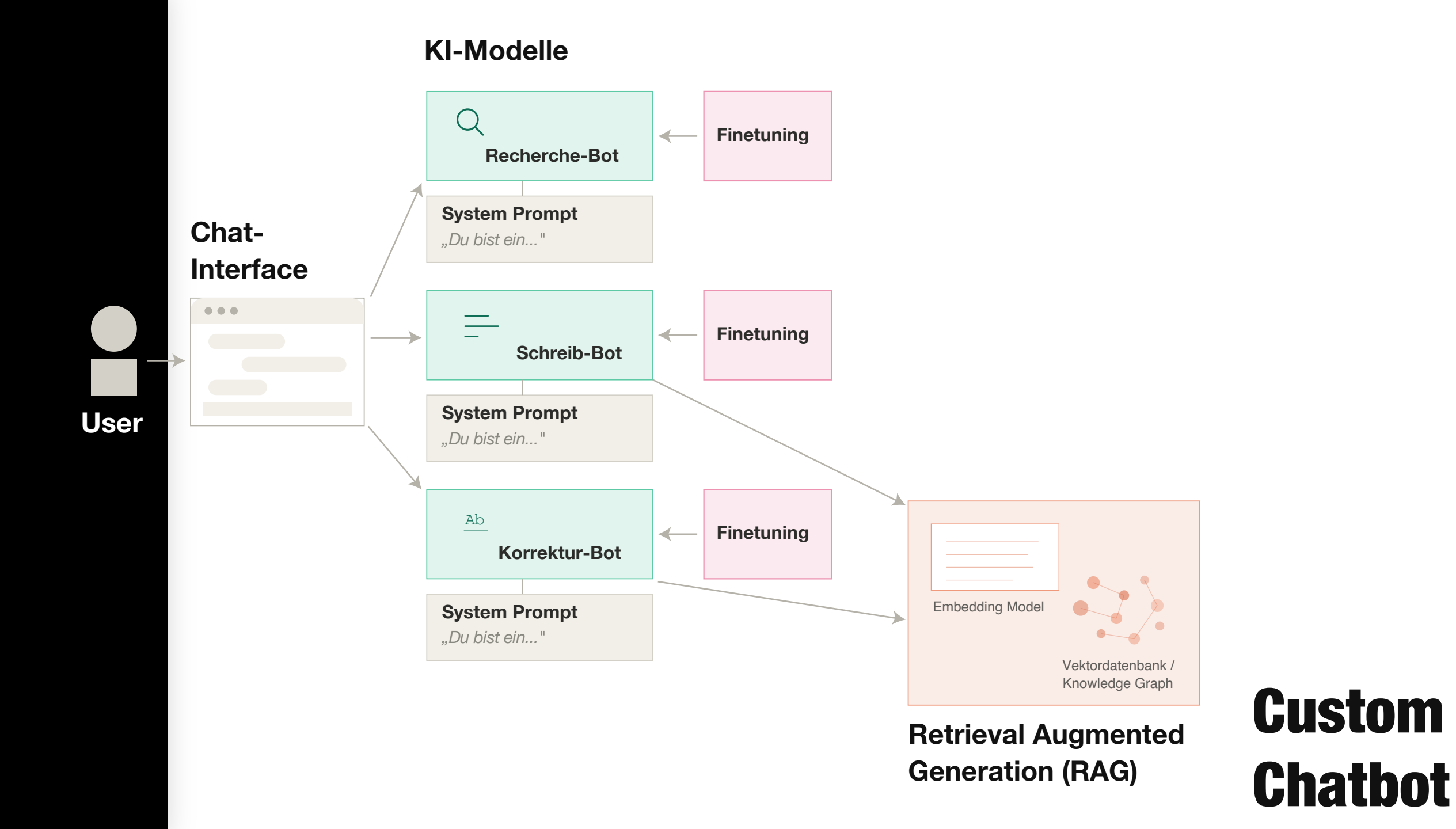

KI-Modelle
Recherche-Bot
System Prompt
„Du bist ein..."
Chat-
Interface
Schreib-Bot
System Prompt
„Du bist ein..."
User
Ab
Korrektur-Bot
System Prompt
„Du bist ein..."

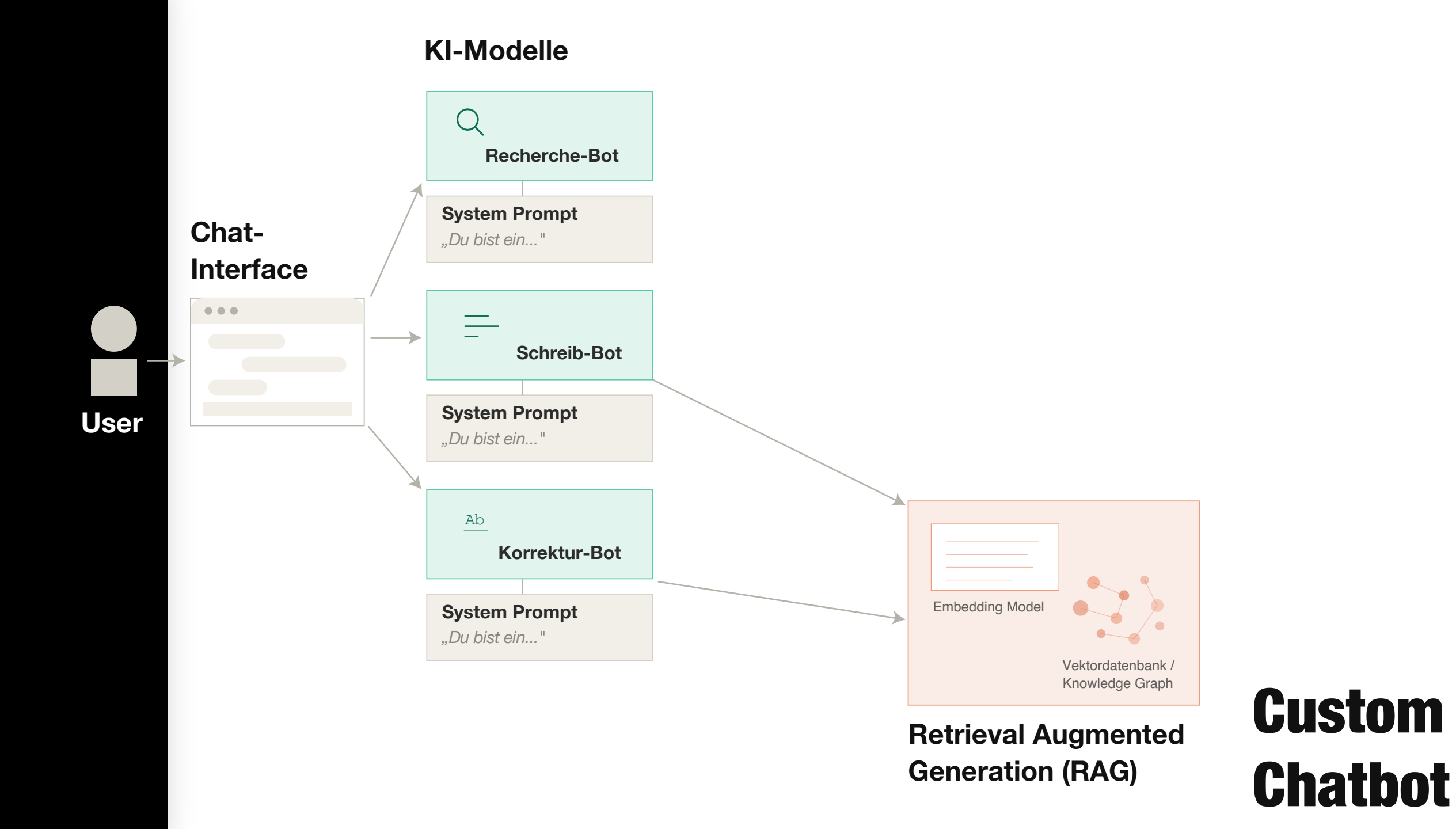

KI-Modelle
Recherche-Bot
Finetuning
System Prompt
„Du bist ein..."
Chat-
Interface
User
Schreib-Bot
Finetuning
System Prompt
„Du bist ein..."
Ab
Korrektur-Bot
Finetuning
System Prompt
„Du bist ein..."


KI-Modelle
Recherche-Bot
Finetuning
System Prompt
„Du bist ein..."
Chat-
Interface
User
Schreib-Bot
Finetuning
System Prompt
„Du bist ein..."
Ab
Korrektur-Bot
Finetuning
System Prompt
„Du bist ein..."
Embedding Model
Vektordatenbank /
Knowledge Graph
Retrieval Augmented
Generation (RAG)


KI-Modelle
Model Context Protocol (MCP)
Chat-Interface
User
Recherche-Bot
Finetuning
System Prompt
„Du bist ein..."
Schreib-Bot
Finetuning
System Prompt
„Du bist ein..."
Ab
Korrektur-Bot
Finetuning
System Prompt
„Du bist ein..."
Websuche
Programmierung
Office-Integration
Coding-IDE
Mo Di Mi Do Fr
1 2 3 4 5
8 9 10 11 12
15 16 17 18 19
22 23
Kalender-App
Embedding Model
Vektordatenbank /
Knowledge Graph
Retrieval Augmented Generation (RAG)


User
Chat-Interface
KI-Modelle
Recherche-Bot
Finetuning
System Prompt
„Du bist ein..."
Schreib-Bot
Finetuning
System Prompt
„Du bist ein..."
Ab
Korrektur-Bot
Finetuning
System Prompt
„Du bist ein..."
Model Context Protocol (MCP)
Websuche
Programmierung
Office-Integration
Coding-IDE
Mo Di Mi Do Fr
1 2 3 4 5
8 9 10 11 12
15 16 17 18 19
22 23
Kalender-App
Embedding Model
Vektordatenbank / Knowledge Graph
Retrieval Augmented Generation (RAG)
Custom Chatbot


KI-Modelle
Recherche-Bot
Finetuning
System Prompt
„Du bist ein..."
Chat-
Interface
Schreib-Bot
Finetuning
System Prompt
„Du bist ein..."
User
Ab
Korrektur-Bot
Finetuning
System Prompt
„Du bist ein..."
Embedding Model
Vektordatenbank /
Knowledge Graph
Retrieval Augmented
Generation (RAG)
Custom
Chatbot


KI-Modelle
Recherche-Bot
System Prompt
„Du bist ein..."
Chat-
Interface
Schreib-Bot
System Prompt
„Du bist ein..."
Ab
Korrektur-Bot
System Prompt
„Du bist ein..."
User
Embedding Model
Vektordatenbank /
Knowledge Graph
Retrieval Augmented
Generation (RAG)
Custom
Chatbot

User

Chat-Interface

KI-Modell

Schreib-Bot

System Prompt

„Du bist ein..."

Embedding Model

Vektordatenbank / Knowledge Graph

Retrieval Augmented Generation (RAG)

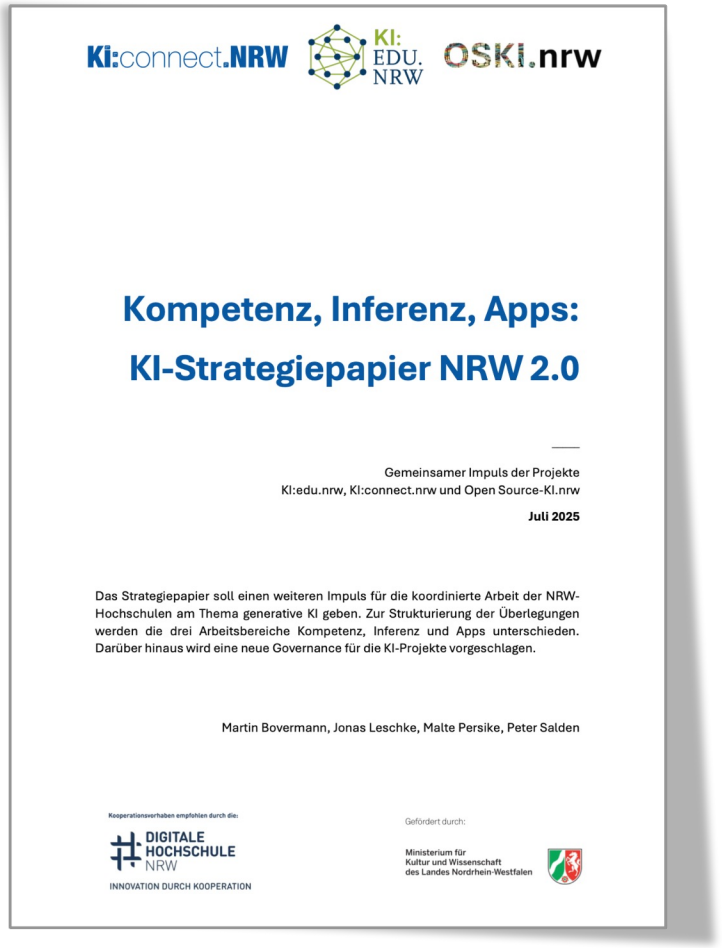
Bot-Camp

3

Risiken für die Antwortqualität

Source: https://unsplash.com/photos/hxCIfi47FgQ

1 Wir laden PDFs hoch



Ki:connect.NRW KI: EDU. NRW OSKI.nrw

Kompetenz, Inferenz, Apps:
KI-Strategiepapier NRW 2.0

Gemeinsamer Impuls der Projekte
KI:edu.nrw, KI:connect.nrw und Open Source-KI.nrw

Juli 2025

Das Strategiepapier soll einen weiteren Impuls für die koordinierte Arbeit der NRW-Hochschulen am Thema generative KI geben. Zur Strukturierung der Überlegungen werden die drei Arbeitsbereiche Kompetenz, Inferenz und Apps unterschieden. Darüber hinaus wird eine neue Governance für die KI-Projekte vorgeschlagen.

Martin Bovermann, Jonas Leschke, Malte Persike, Peter Salden

Kooperationsverhaben empfohlen durch die:
DIGITALE HOCHSCHULE NRW
INNOVATION DURCH KOOPERATION

Gefördert durch:
Ministerium für Kultur und Wissenschaft des Landes Nordrhein-Westfalen

2 Die PDFs werden in Sinneinheiten („Chunks“) gestückelt

Kompetenz, Inferenz, Apps: KI-Strategiepapier NRW 2.0

Gemeinsamer Impuls der Projekte KI:edu.nrw, KI:connect.nrw und Open Source-KI.nrw

Juli 2025

Das Strategiepapier soll einen weiteren Impuls für die koordinierte Arbeit der NRW-Hochschulen am Thema generative KI geben. Zur …

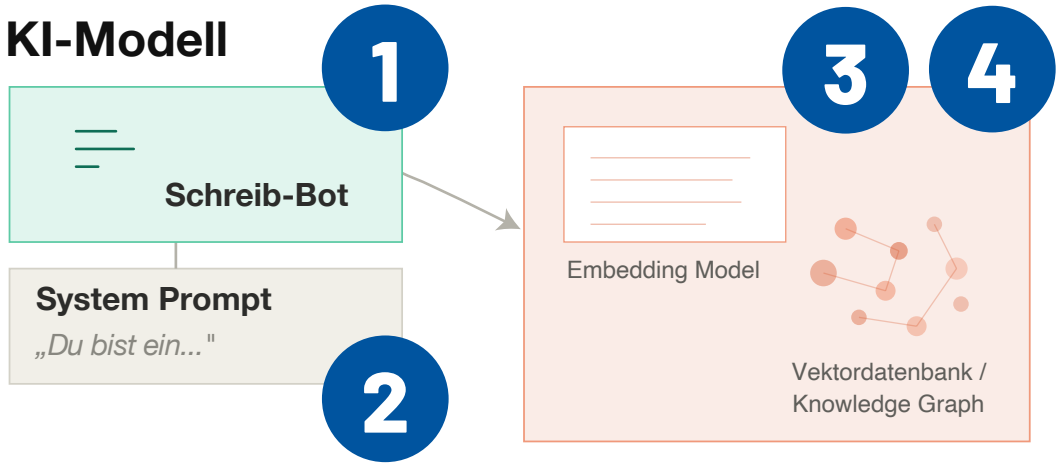
3 Die Chunks werden in einer Datenbank gespeichert

Vector Datenbank

4 Relevante Chunks werden abgerufen und in die KI-Antworten eingebaut

Im Juli 2025 wurde das Strategiepapier "Kompetenz, Inferenz, Apps: KI-Strategiepapier NRW 2.0" veröffentlicht. Es enthält einen Impuls für die koordinierte Arbeit der NRW-Hochschulen am Thema generative KI.

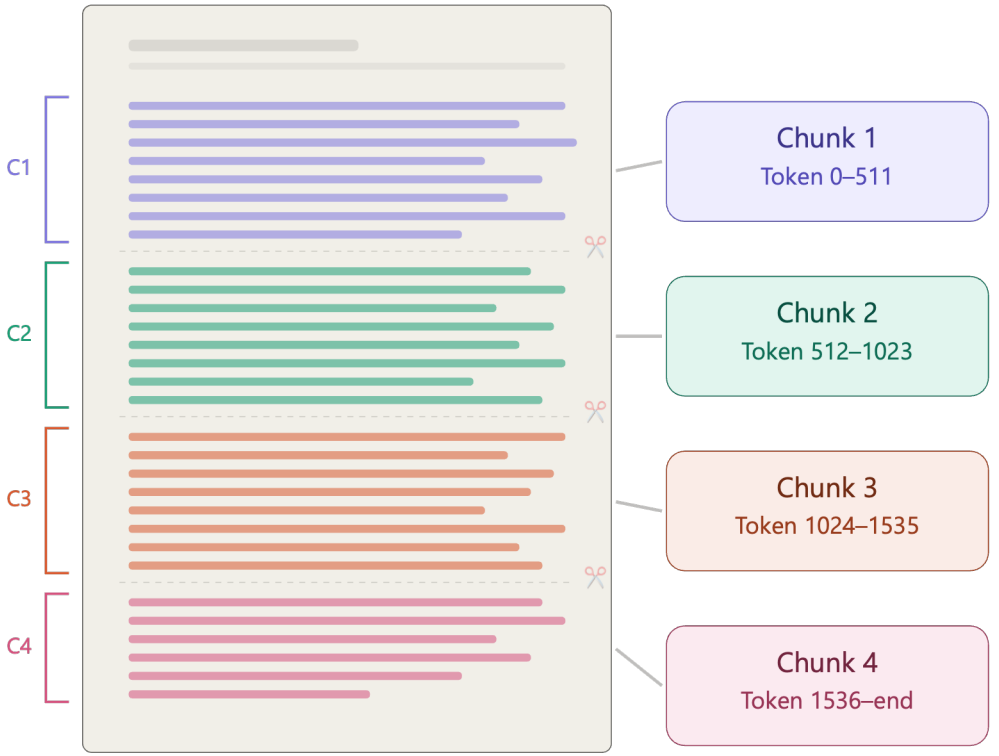
Risiken für die Antwortqualität



1. **Trivialer Fall:** KI-Halluzinationen
2. **Vermeidbar:** Defizite im System Prompt
3. **Problemfeld 1:** Chunking-Methoden beim RAG
4. **Problemfeld 2:** Retrieval-Methoden beim RAG

Einfluss der Chunking-Methode

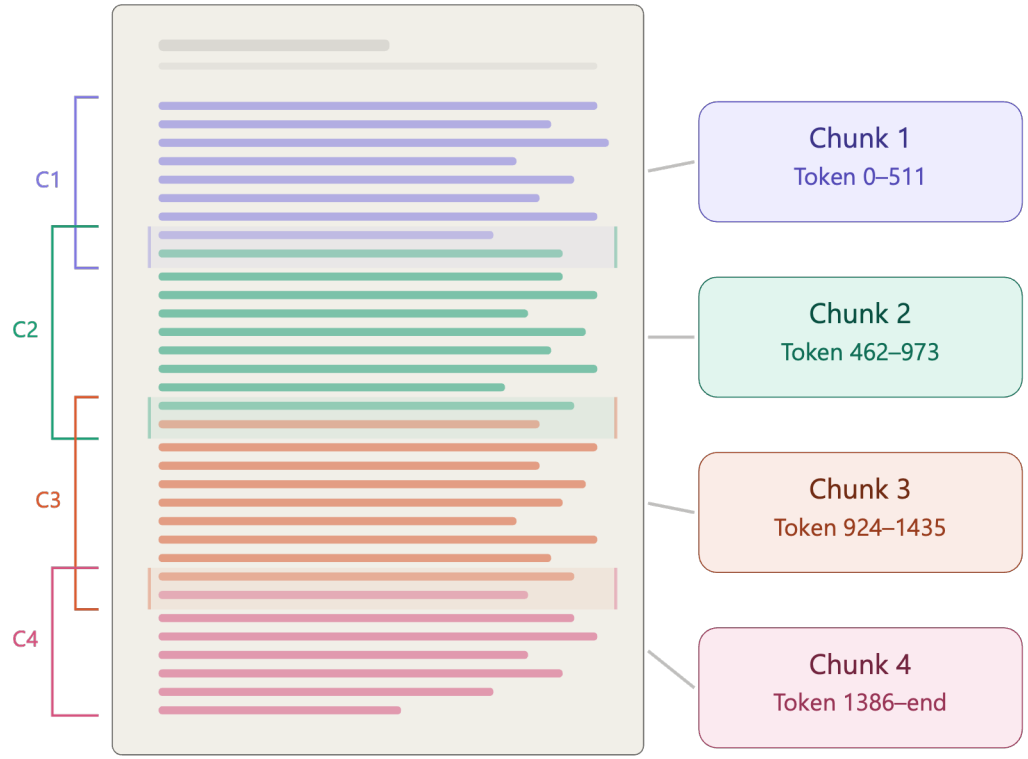
Fixed Chunking: Chunks sind immer eine feste Anzahl von Tokens.



Einfluss der Chunking-Methode

Fixed Chunking: Chunks sind immer eine feste Anzahl von Tokens.

Sliding Window: Chunks überlappen sich, um den Sinngehalt der umgebenden Chunks zu erhalten

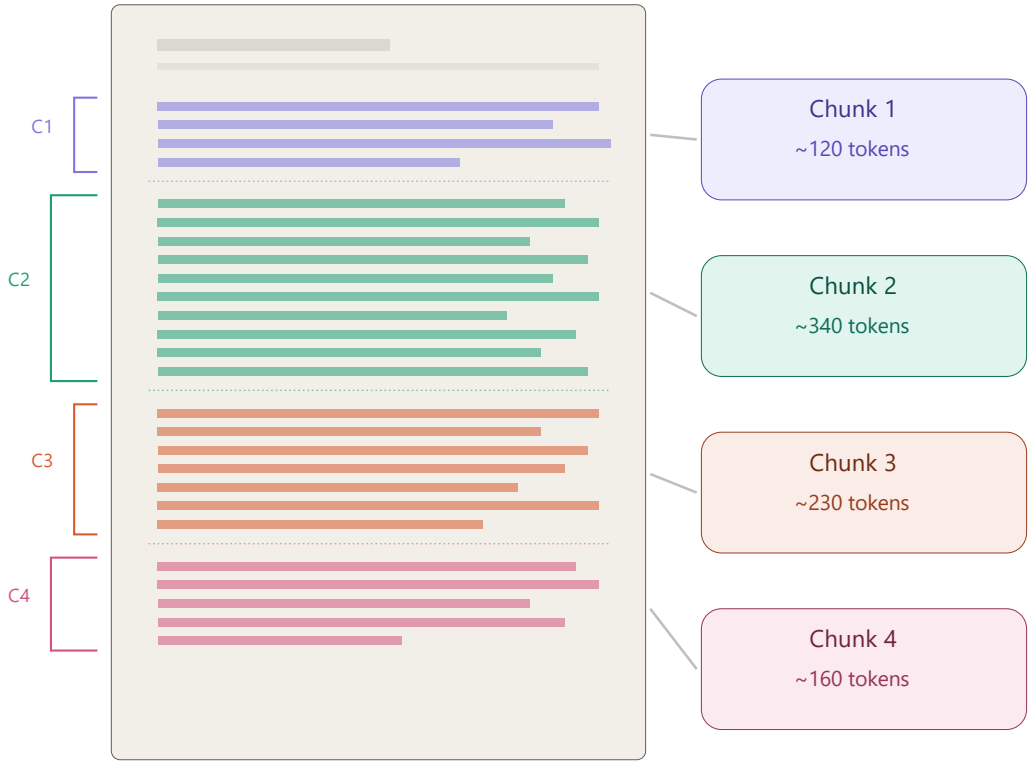


Einfluss der Chunking-Methode

Fixed Chunking: Chunks sind immer eine feste Anzahl von Tokens.

Sliding Window: Chunks überlappen sich, um den Sinngehalt der umgebenden Chunks zu erhalten

Semantic Chunking: Sinneinheiten werden gechunkt.



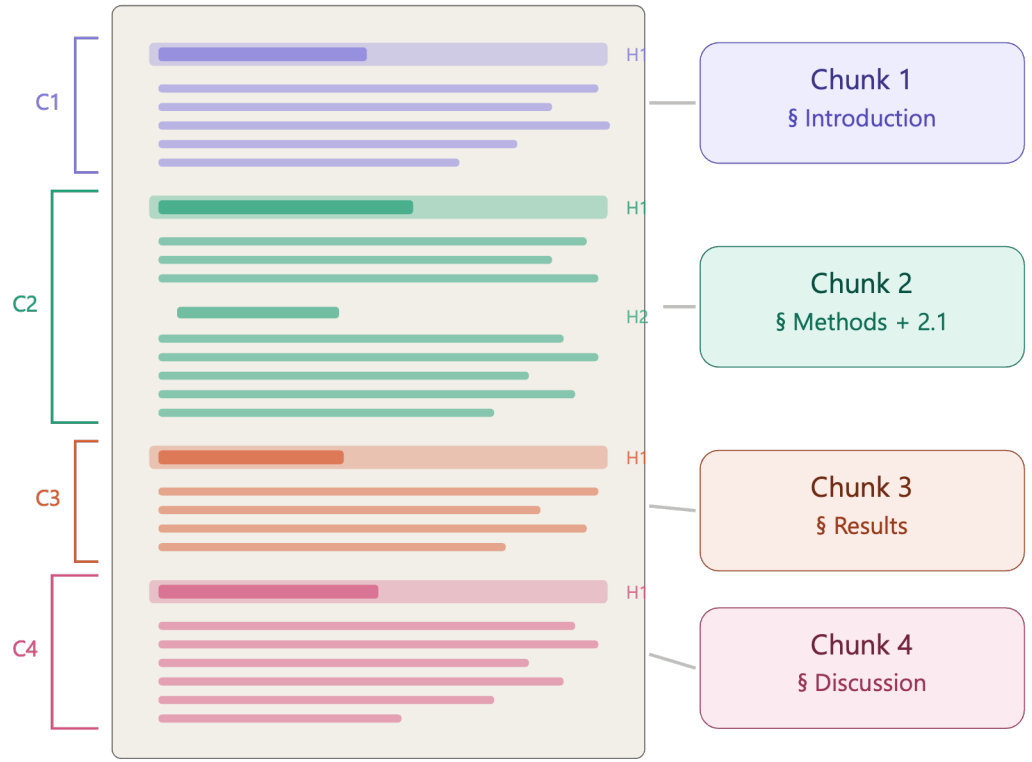
Einfluss der Chunking-Methode

Fixed Chunking: Chunks sind immer eine feste Anzahl von Tokens.

Sliding Window: Chunks überlappen sich, um den Sinngehalt der umgebenden Chunks zu erhalten

Semantic Chunking: Sinneinheiten werden gechunkt.

Document Aware Chunking: Chunks folgen Strukturelementen des Dokuments (z.B. Überschriften).



Einfluss der Chunking-Methode

Fixed Chunking: Chunks sind immer eine feste Anzahl von Tokens.

Sliding Window: Chunks überlappen sich, um den Sinngehalt der umgebenden Chunks zu erhalten

Semantic Chunking: Sinneinheiten werden gechunkt.

Document Aware Chunking: Chunks folgen Strukturelementen des Dokuments (z.B. Überschriften).

Vision Based Chunking: Bildelemente bleiben erhalten.

Einfluss der Retrieval-Methode

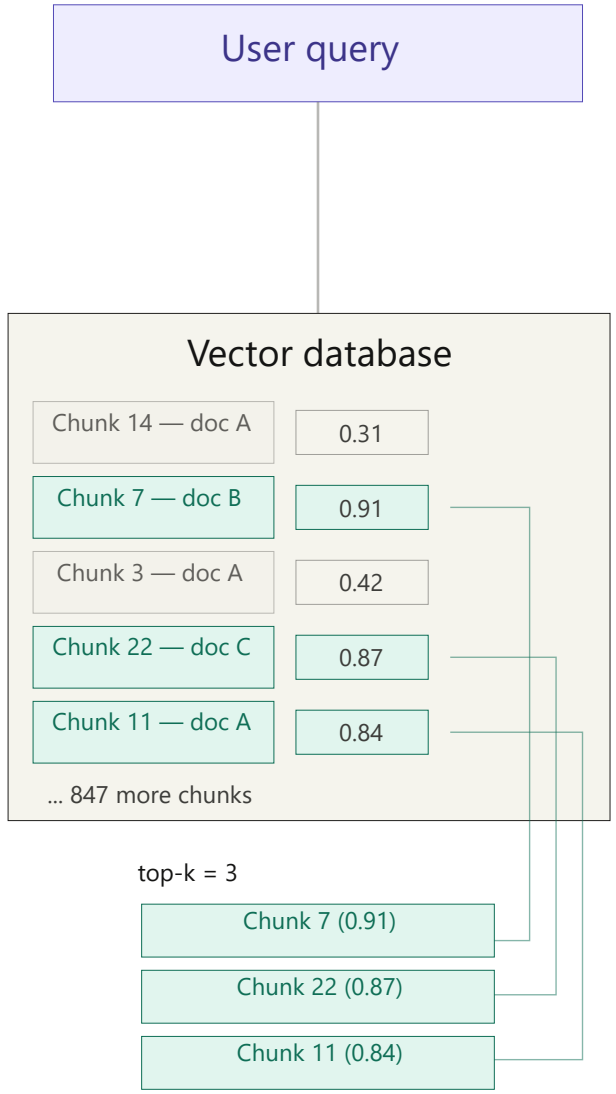
Sparse / Dense / Hybrid Retrieval
Suche mit einfachen Statistiken oder mit LLM?

Reranking
Wie werden Fundstellen nach Wichtigkeit sortiert?

Query Transformation
Wie werden Nutzendenanfragen möglichst günstig umformuliert?

Fortgeschrittene Verfahren
Multi-Hop Retrieval, Contextual Retrieval etc.

Art der Datenbank
Vektordatenbank oder Knowledge Graph Datenbank?



Was nutzen wir im Bot-Camp?

Erkenntnis: Die eine optimale Chunking-Methode für alle Dokumentarten existiert nicht.

Ebenso verhält es sich mit den Retrieval-Methoden.

Also: Wir haben Methoden ausgewählt, die ein „möglichst wenig schädlicher“ Kompromiss aus Allgemeingültigkeit und Ressourcenaufwand sind.

Folge: Beim RAG für Eure Dokumente werden mit Sicherheit Fehler auftreten.