



# Smart Student Support Chatbot

Bachelorarbeit – Universität Wien  
Betreut von Dipl.-Ing. Dr. Kalchgruber  
Muza Sofiia Shutova  
Juni 2025

# Motivation & Problemstellung



VIELE INFOQUELLEN –  
UNÜBERSICHTLICH FÜR  
STUDIERENDE



HOHE  
ANFRAGEBELASTUNG BEI  
SSC/SPL

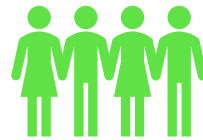


BEDARF NACH 24/7  
UNTERSTÜTZUNG

# Ziel der Arbeit



Entwicklung eines KI-  
gestützten Chatbots



Mehrsprachig, lokal  
gehostet, DSGVO-konform



Prototyp für universitäre  
Nutzung

# Technologischer Hintergrund

NLP: Verarbeitung  
natürlicher Sprache

RAG:  
Antwortgenerierung  
mit  
Dokumentenwissen

Ollama, Flask, Vue.js

# Systemarchitektur



Modular: Frontend  
(Vue.js), Backend (Flask),  
LLM (Ollama)



Docker-basierte Container  
für Portabilität



Kommunikation über REST-  
APIs

# Datenquellen & Vorbereitung

u:find XML-Daten (Lehrveranstaltungen, Kontakte)

Curriculum-PDFs → manuell in Markdown umgewandelt

Fakultäts- & SSC-Webseiten

# RAG-Architektur & Modellwahl (Retrieval-Augmented Generation)

## Embedding-Modell


Zuständig für semantische Suche  
und Kontextabruf

## Hauptmodell

Generiert die finalen Antworten  
auf Nutzeranfragen

# Embedding-Modell & Vektor-Datenbank

Die Daten werden in einer **Vektordatenbank (ChromaDB)** gespeichert  
Jeder Textabschnitt wird in einen Vektor umgewandelt  
Diese Vektoren ermöglichen kontextbasierte, sprachunabhängige Suche



## Zwei Favoriten Embedding-Modelle:

**bge-m3:** Sehr gute Ergebnisse,  
mehrsprachig, stabil

**e5-base-sts-en-de:** Besonders stark bei  
deutsch-englischem Matching





# Hauptmodell: Gemma3:27b

- Verwendet zur **Antwortgenerierung** im Chatbot
- Läuft **lokal über Ollama** → Keine Cloud notwendig
- **Effizienter als andere getestete Modelle** (z. B. LLaMA 3)

# Modellauswahl



EMBEDDING: BGE-M3  
(MEHRSPRACHIG,  
PERFORMANT)



LLM: GEMMA3:27B  
(PRÄZISE, LOKAL  
AUSFÜHRBAR)



VERGLEICH MIT ANDEREN  
MODELLEN: BESSERE  
ERGEBNISSE

# Evaluierungsfragen

1. Wie lauten die Kontaktdaten von Professor Peter Reichl?

2. Wie lauten die Öffnungszeiten und Kontaktdaten des SSC der Fakultät Informatik?

3. Welche Teilnahmevoraussetzungen gibt es für das Softwarepraktikum im Rahmen der Bachelorarbeit?

4. Wie viele ECTS-Punkte hat das Fach PR2 (Programmieren 2)?

5. Welche Aufnahmeverfahren gibt es generell an der Fakultät für Informatik?

6. What courses does Professor Peter Kalchgruber teach

7. How much is the tuition fee for non-EU students at the University of Vienna?

8. What courses are recommended in the third semester of the Bachelor in Business Informatics?

9. What is STEOP and why is it important?

10. Who is the dean of the Faculty of Computer Science and how can I contact them?

# Embedding: bge-m3

	bge-m3			
Frage	Genauigkeit	Verständlichkeit	Antwortdauer (s)	Gesamtbewertung
1	vollständig korrekt	5	4	5
2	vollständig korrekt	5	4	5
3	vollständig korrekt	5	3	5
4	vollständig korrekt	5	4	5
5	vollständig korrekt	5	5	5
6	vollständig korrekt	5	3	5
7	vollständig korrekt	5	4	5
8	FALSCH	3	6	1
9	vollständig korrekt	5	4	5
10	vollständig korrekt	5	5	5

# Embedding: e5-base-sts-en-de

		e5-base-sts-en-de		
Frage	Genauigkeit	Verständlichkeit	Antwortdauer (s)	Gesamtbewertung
1	vollständig korrekt	5	5	5
2	vollständig korrekt	5	4	5
3	vollständig korrekt	5	6	5
4	vollständig korrekt	5	5	5
5	vollständig korrekt	5	5	5
6	vollständig korrekt	5	4	5
7	FALSCH	2	6	1
8	FALSCH	2	6	1
9	vollständig korrekt	5	3	5
10	vollständig korrekt	5	5	5

# Hauptmodell: Gemma3:27b

	Gemma3:27b			
Frage	Genauigkeit	Verständlichkeit	Antwortdauer (s)	Gesamtbewertung
1	vollständig korrekt	5	4	5
2	vollständig korrekt	5	4	5
3	vollständig korrekt	5	3	5
4	vollständig korrekt	5	4	5
5	vollständig korrekt	5	5	5
6	vollständig korrekt	5	3	5
7	vollständig korrekt	5	4	5
8	FALSCH	3	6	1
9	vollständig korrekt	5	4	5
10	vollständig korrekt	5	5	5

# Hauptmodell: Llama3.1:8b

	llama3.1:8b			
Frage	Genauigkeit	Verständlichkeit	Antwortdauer (s)	Gesamtbewertung
1	vollständig korrekt	5	4	5
2	vollständig korrekt	5	5	5
3	teilweise korrekt	3	4	3
4	vollständig korrekt	5	5	5
5	vollständig korrekt	5	3	5
6	vollständig korrekt	5	4	5
7	vollständig korrekt	5	5	5
8	FALSCH	2	4	1
9	vollständig korrekt	5	5	5
10	FALSCH	2	3	1

# Hauptmodell: Gemma3:1b

	gemma3:1b			
Frage	Genauigkeit	Verständlichkeit	Antwortdauer (s)	Gesamtbewertung
1	vollständig korrekt	5	4	5
2	vollständig korrekt	5	2	5
3	FALSCH	1	1	1
4	vollständig korrekt	5	1	5
5	FALSCH	1	2	1
6	FALSCH	1	1	1
7	FALSCH	1	2	1
8	FALSCH	1	2	1
9	vollständig korrekt	3	2	4
10	FALSCH	3	2	1



# Evaluierung & Ergebnisse

Relevante Antworten  
ohne externes  
Feintuning

Hohe Genauigkeit  
bei konkreten Fragen

**Bge-m3 +  
Gemma3:27b**

Gute Performance  
auch lokal

# Herausforderungen & Ausblick

Komplexe Datenaufbereitung (PDF  
→ Markdown)



Hardwarebedarf für größere  
Modelle



Zukünftige Integration in u:space,  
Lernanalysen



# Demo & Fragen



# Abschluss & Dank



PROTOTYP FUNKTIONIERT  
ZUVERLÄSSIG, LOKAL &  
DSGVO-KONFORM



HILFREICH FÜR  
STUDIERENDE IM  
STUDIENALLTAG



DANKE AN BETREUER,  
UNTERSTÜTZENDE &  
ZUHÖRENDE!