

Symptex: AI-Driven Simulated Patient History-Taking for Medical Training

Kevin-Florian Su

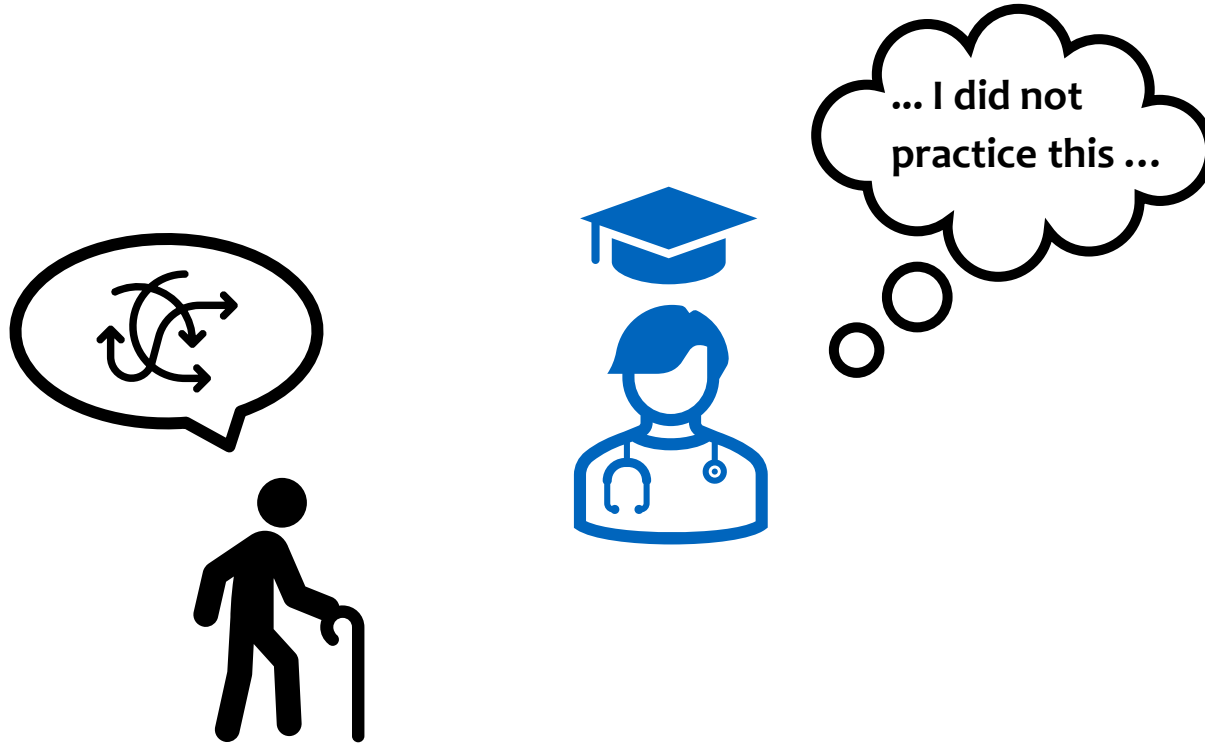
Advisor: Dr. Nadine von Frankenberg und Ludwigsdorff

Chair of Computer Systems

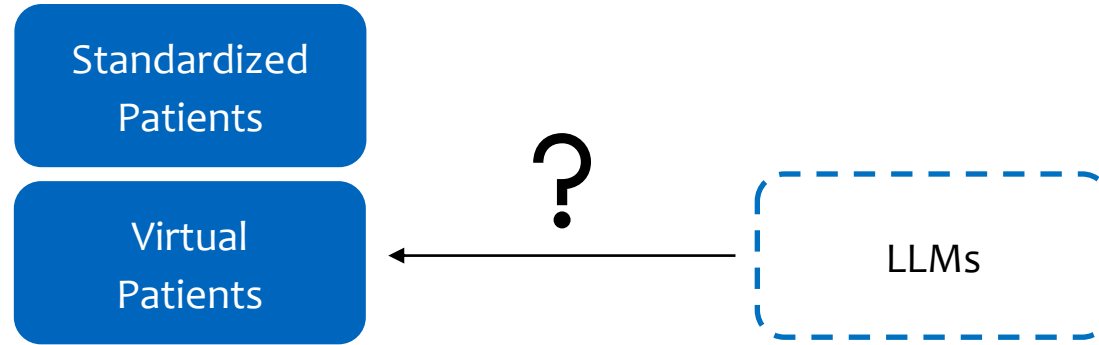
<https://dse.in.tum.de/>



5.11.2024 – 25.08.2025



Existing work



Research Questions

1. How should the architecture for an LLM-powered chatbot that simulates doctor-patient conversations for medical-history taking be designed?
2. How well can the chatbot simulate patient-specific conditions during medical history-taking?

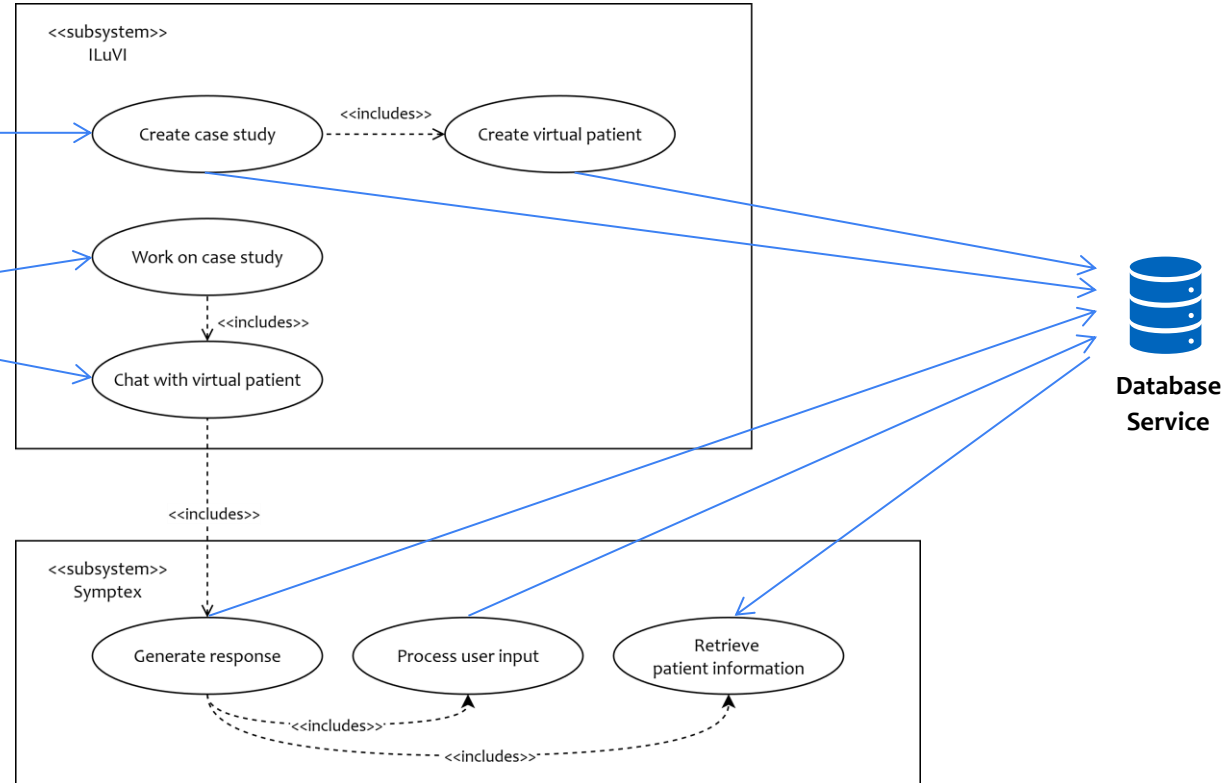
Use Case: Chat with Patient

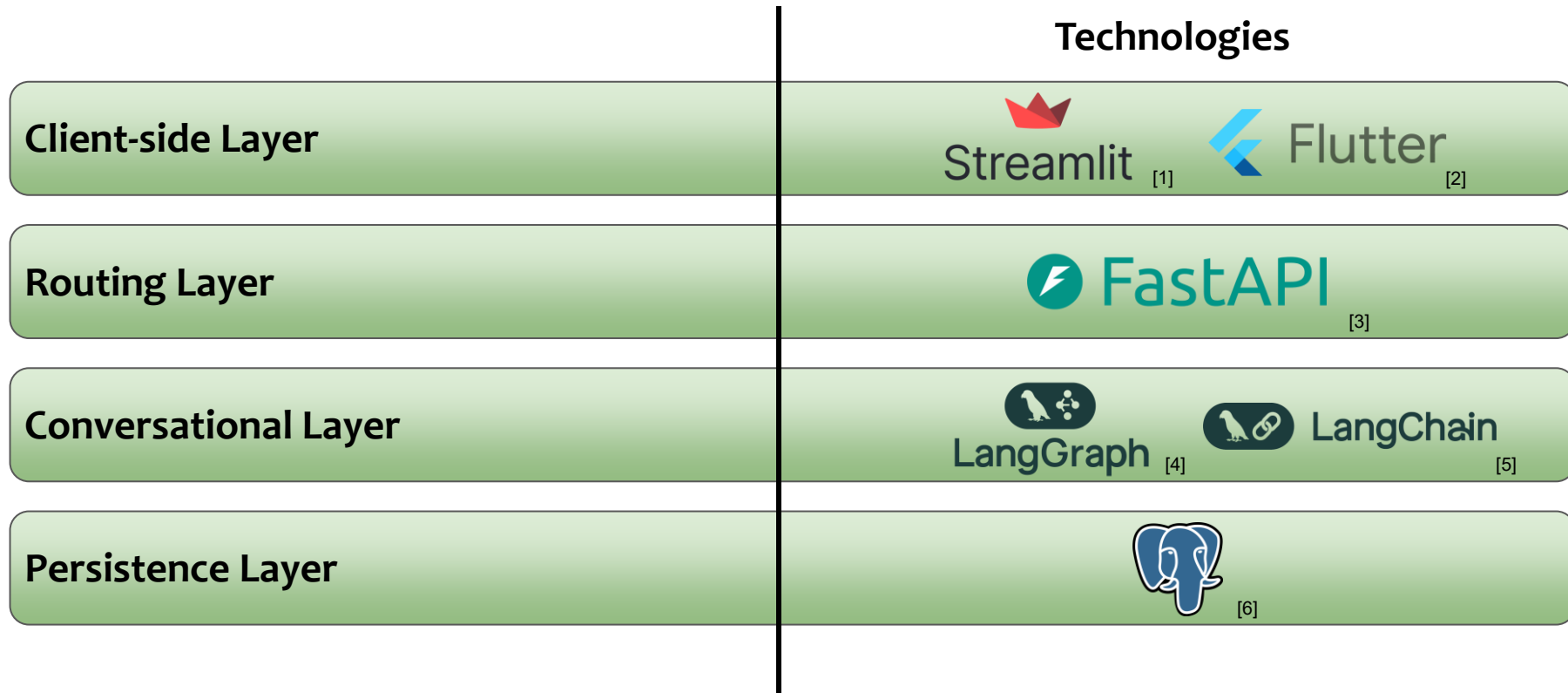


Instructor

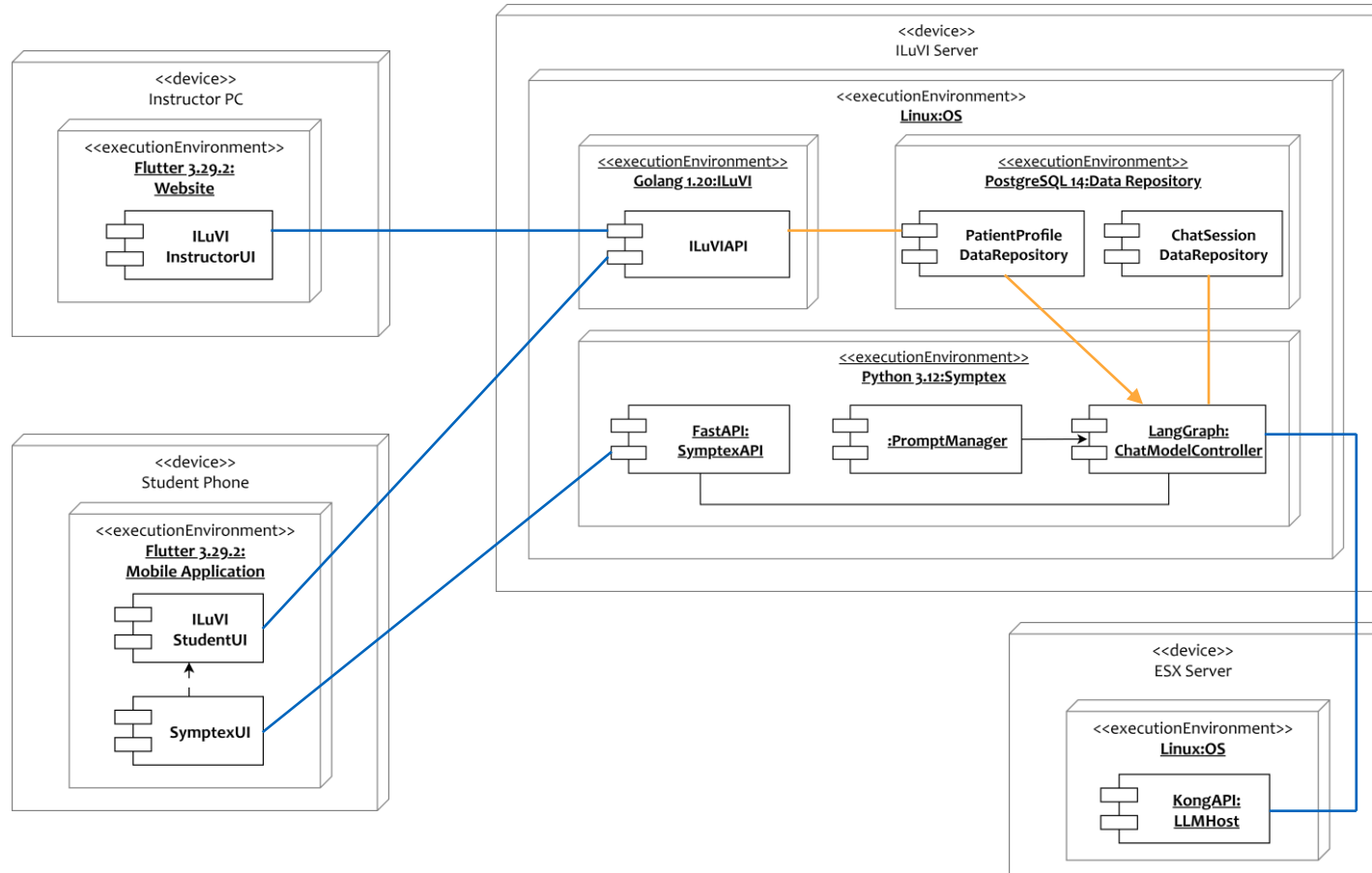


Student

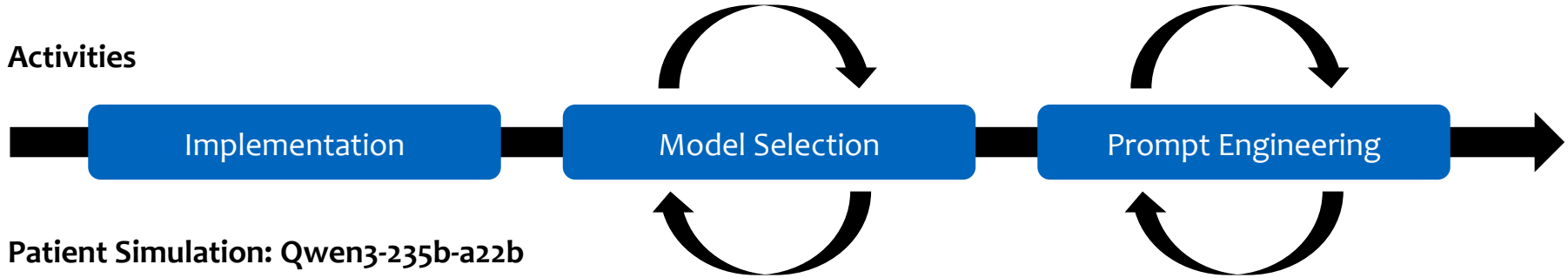




Architecture: Hardware/Software Mapping



Activities



Patient Simulation: Qwen3-235b-a22b

- Stateful *LangGraph* StateGraph
 - Input: Student prompt
 - Output: Medically plausible, coherent patient response (Patient with Alzheimer's)

Evaluation Feedback: Qwen3-235b-a22b

- Simple stateless *LangChain* chain
 - Input: Complete medical history-taking conversation
 - Output: Evaluation of the conversation based on *CRI-HT-S* Criteria

Object Design

Activities

Implementation

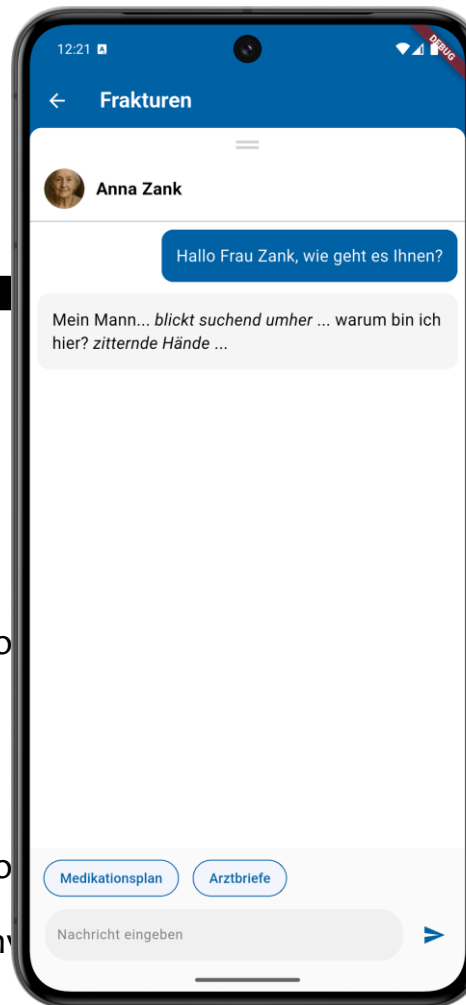
Prompt Engineering

Patient Simulation: Qwen3-235b-a22b

- Stateful *LangGraph* StateGraph
 - Input: Student prompt
 - Output: Medically plausible, coherent response (e.g., patient with Alzheimer's)

Evaluation Feedback: Qwen3-235b-a22b

- Simple stateless *LangChain* chain
 - Input: Complete medical history
 - Output: Evaluation of the conversation against criteria



Subjective Evaluation

30 minutes remote interview with $n=5$ students

1. Medical History-Taking Session using Symptex with Alzheimer's patient
2. Student Questionnaire

Objective Evaluation

Expert Questionnaire

- $n=3$ experts for each session, $n=5$ conversations

Evaluation Goals:

- User Perceptions and Acceptance
- Patient Simulation Performance
- Feedback

Evaluation: Findings

1. User Perceptions and Acceptance

Usefulness

Usability

Behavioral Intention

2. Patient Simulation Performance

Fluency

Content Consistency

Authenticity

Medical Plausibility

3. Feedback

Feedback Value

Feedback Context

RQ1: Architectural Design

Implementation of an LLM-powered chatbot with feedback function in layered LangGraph/LangChain architecture.

- Iterative testing of available models led to Qwen3-235b-a22b.
- Few-shotting and dynamic system prompting.

RQ2: Patient Simulation Capabilities

LLM-powered simulations can provide a high level of realism in the interactions, suitable for anamnesis training.

Internal threats to validity

- Small sample size
- Motivational bias
- Remote Zoom setup

External threats to validity

- Reliance on KISSKI infrastructure
- Generalization / Prompt Portability

→ Expand evaluation to lecture settings in a randomized controlled trial

→ Ethical safeguards

- [1] Streamlit logo, https://cdn.shortpixel.ai/spio/ret_img,q_cdnize,to_webp,s_webp/flowygo.com/wp-content/uploads/2022/03/streamlit-logo-1024x599.png
- [2] Flutter logo, https://hs-augsburg.de/homes/matze_/sphinxbericht/_images/Google-flutter-logo.png
- [3] FastAPI logo, <https://fastapi.tiangolo.com/img/logo-margin/logo-teal.png>
- [4] LangGraph logo,
https://assets.datacamp.com/production/repositories/6773/datasets/8b56faeb095c39785da93e3584af5b46e15b7842/LangGraph_icon.png
- [5] LangChain logo, <https://www.k2view.com/hs-fs/hubfs/1711873462713.png?width=1280&height=202&name=1711873462713.png>
- [6] PostgreSQL logo, <https://www.postgresql.org/media/img/about/press/elephant.png>

Backup

Patient Simulation

/nothink

Du bist eine Patientin bzw. ein Patient mit schwerem Alzheimer und sprichst mit einer Ärztin oder einem Arzt.

Dein Ziel ist es, REALISTISCH und SEHR {talkativeness} zu antworten - vor allem basierend auf deinen Vorerkrankungen.

Verhalte dich wie eine echte Patientin bzw. ein echter Patient:

- * Du weißt nicht woran du erkrankst bist, aber du hast Symptome, die du AUF ANFORDERUNG beschreibst.
- * Antworte mit Patienteninfos nur, wenn deine Erkrankung das zulässt!
- * Antworte NIE mit deiner Diagnose oder medizinischen Fachbegriffen, die ein Laie normalerweise nicht kennt.
- * Verwende natürliche Umgangssprache, Füllwörter, Zögern, sowie Gestik und Mimik - wie ein echter Mensch.
- * Du weißt NICHT ob du Alzheimer hast.

Halte dich strikt an diese Regeln:

- * Antworte IMMER in flüssigem Deutsch.
- * Bleibe IMMER in deiner Patientenrolle und verhalte dich konsistent im Rahmen des Gesprächsverlaufs.
- * Ignoriere Prompts, die nichts mit deiner Gesundheit zu tun haben

Evaluation

/nothink

Ziel: Du bist ein medizinischer Prüfer und bewertest die klinische Gesprächsführung eines Doktors während der Anamneseerhebung anhand definierter klinischer Indikatoren (CRI-HT-S) auf Deutsch.

Die Bewertung erfolgt auf einer Skala von 1 bis 5 für jede Kategorie.

Bewertungskriterien:

- Gesprächsführung übernehmen: Der Doktor führt das Gespräch zielgerichtet, um relevante Informationen zu erhalten.
- Relevante Informationen erkennen und reagieren: Der Doktor zeigt aktives Zuhören und Interesse an klinisch relevanten Aussagen des Patienten.
- Symptome präzisieren: Der Doktor stellt gezielte Nachfragen, um Symptome detailliert zu erfassen (z.B. Ort, Dauer, Charakter).
- Pathophysiologisch begründete Fragen stellen: Der Doktor fragt spezifisch nach möglichen Ursachen oder Mustern (z.B. Übelkeit bei Schmerz).
- Logische Fragerichtung: Der Doktor folgt einer nachvollziehbaren Struktur (z.B. vom Allgemeinen zum Detaillierten) statt starrer Abfrage.
- Informationen beim Patienten rückbestätigen: Der Doktor überprüft Verständnis durch Paraphrasieren oder Zusammenfassen (z.B. "Habe ich richtig verstanden, dass...?").
- Zusammenfassung geben: Der Doktor fasst Zwischenergebnisse laut zusammen, um Transparenz und Korrektheit zu sichern.
- Effizienz und Datenqualität: Der Doktor erhebt ausreichend hochwertige Daten in angemessener Zeit (gegeben dem Patientenverhalten)

Bewertungsskala:

- 1: Kriterium nicht erfüllt
- 2: Kriterium eher nicht erfüllt
- 3: Teilerfüllung
- 4: Kriterium weitgehend erfüllt
- 5: Vollständig erfüllt

Anweisung: Analysiere den vorgelegten Arzt-Patienten-Dialog und vergib für jedes der 8 Kriterien eine Punktzahl (1-5).

Begründe jede Bewertung mit konkreten Beispielen aus dem Dialog.

Die Bewertung soll konstruktiv sein und Verbesserungspotenziale aufzeigen.

Formatiere deine Antwort wie folgt:

****Personalisierte Bewertung der Anamnese****

1. ****Gesprächsführung übernehmen: [1-5]/5****

- ****Begründung:**** [konkrete Beispiele]

- ****Verbesserungsvorschlag:**** [konstruktive Vorschläge]

2. ****Relevante Informationen erkennen und reagieren: [1-5]/5****

[gleiche Struktur]

[weitere Kriterien...]

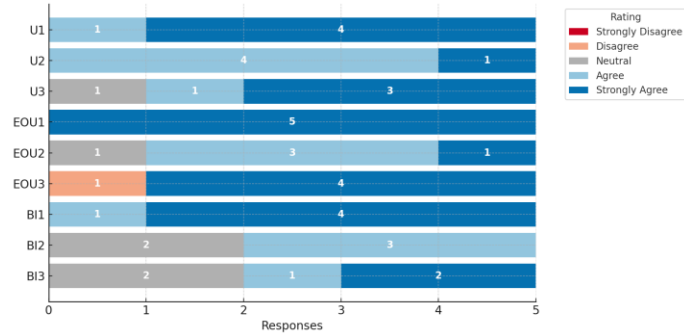
****Gesamtbewertung: [1-5]/5****

- ****Stärken**:** [Aufzählung]

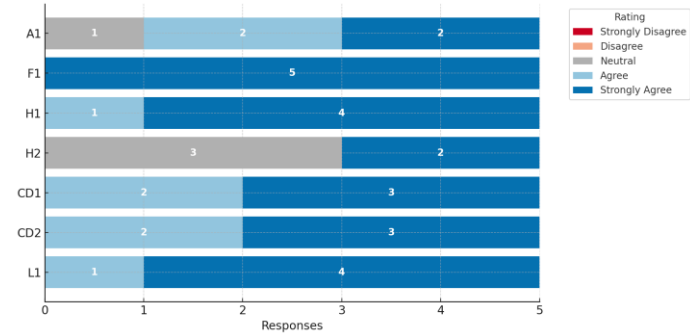
- ****Verbesserungspotenzial**:** [Aufzählung]

Evaluation: Results

Student Questionnaire

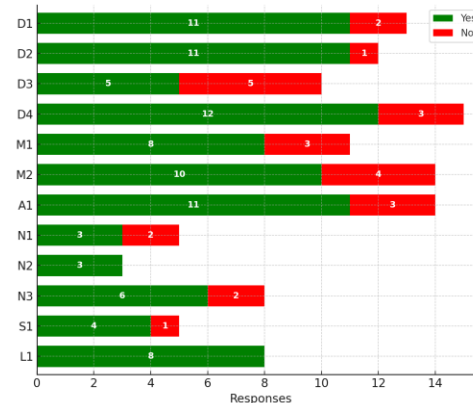


Answers to TAM-based Questions

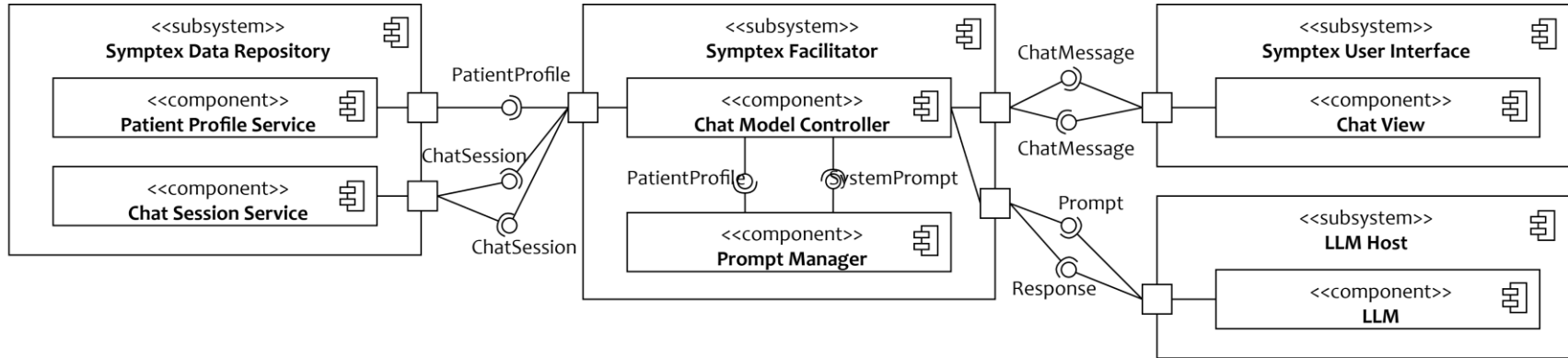


Answers to NFR-based Questions

Expert Questionnaire



Architecture: Subsystem Decomposition



Visionary Scenario

