

Einführung von KI im Biobanking Konzept zur Unterstützung im Projektmanagement

Kann KI im Biobanking bereits Routineaufgaben erleichtern oder sogar komplett übernehmen?



?



Überblick

Large Language Models (LLM)

KI-(Sprach-)Modelle mit riesigen Textmengen trainiert. Sie erkennen Muster menschlicher Sprache und generieren Texte durch das Treffen von Vorhersagen.



Tech-Giganten

Intensive Investition und starker Konkurrenzkampf führen zu schneller Einführung neuer Modelle und Versionen.



Automatisierungspotenzial Beispiele

Kundensupport



Textgenerierung und -
übersetzung



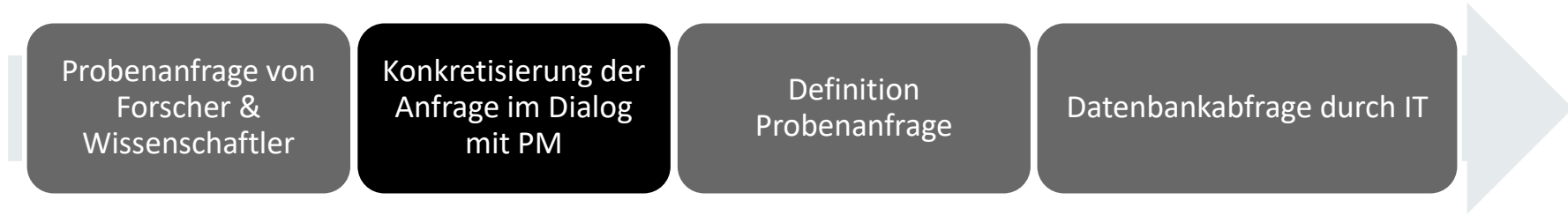
Coding



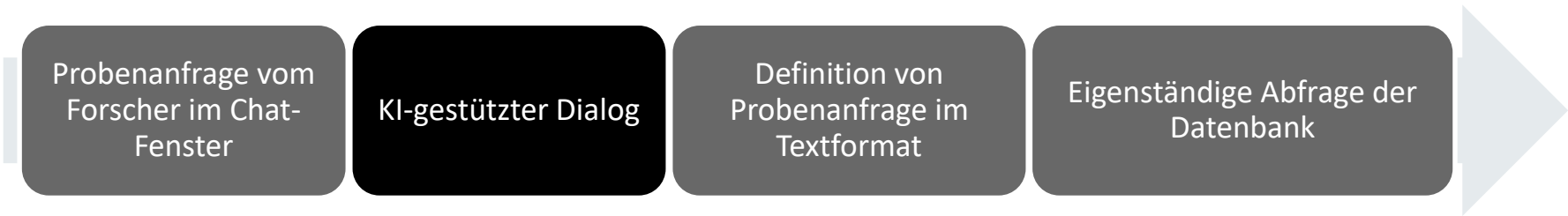
KI-Potenzial im Biobanking

Pilot-Studie: Effizienzsteigerung bei Kundenanfragen zu Probenauslagerung mithilfe von LLM

Ablauf bisher



Ablauf HUB-Chatbot Prototyp



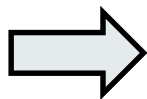
Qualitativer Vergleich und Rechendauer der sechs führenden LLMs

LLM	Domänenspez. Wissen	Qualität des Dialogs	Sprachliche Natürlichkeit	Ø Umfang der Antwort [Zeichenanzahl]	Ø Antwortzeit [s]	Berechnete Anzahl von Zeichen pro Sekunde
Falcon 180B	8	9	9	216,6	5	43,32
GPT 4 Turbo	10	10	9	2.075,5	18,7	110,99
GPT 3.5	8	8	9	1.928,8	8	241,1
Llama-2 13B	5	6	6	1.922,29	70	27,46
Gemini 1.0 Pro	8	9	9	533,49	7,5	71,13
Claude 3.5 Sonnet	9	8	9	1.149,1	8,3	138,45

10 Testfragen (Bsp.: Was ist Plasma? Wofür steht PBMC im Biobanking?)

Herausforderungen und Architektur des KI-Prototyps (Python)

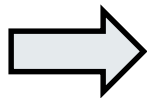
- LLMs basieren auf öffentlichen Daten
- Keine Kenntnis über Codes und Datenbankstrukturen der HUB



- Retrieval Augmented Generation (RAG): Erweiterung des LLMs mit zusätzlichen Dokumenten
- Zugriff auf spezifische Informationen bei der Antwortberechnung

Datenschutz:

- Führende LLMs operieren über Cloud-Schnittstellen
- Sensible Daten sollen die Biobank nicht verlassen



Modularer Aufbau:

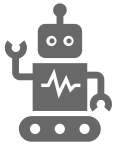
- Abkopplung der tatsächlichen Abfrage
- Lokale Ausführung der Abfrage
- Gewährleistung der Sicherheit von Patientendaten

1. GPT-4 Turbo Anbindung
Aufgabe: Anfrage konkretisieren und JSON-Datei erstellen



2. Datenbank Anbindung: Probenbestand in ElasticSearch (CentraXX)
Aufgabe: Probenabfrage mit JSON-Datei

Einblick in den Dialog



Hallo! Ich bin hier, um dir bei der Auswahl der benötigten Bioproben zu helfen. Teile mir bitte mit welche Probenart du benötigst.

Es gibt zwei Optionen für Plasma-Proben, die zur Auswahl stehen:

- Plasma, einmal zentrifugiert
- Plasma, doppelt zentrifugiert.

Welche dieser Probenarten benötigst du genau?

Für Ihre Anfrage sind aktuell 284 Proben verfügbar.

Plasma

Einmal zentrifugiert



Ergebnis und Ausblick

Evaluation

Test mit 10 Real-World Probenanfragen durchgeführt
Korrekte Probenanzahl als Ergebnis erhalten



Weiterentwicklungsmöglichkeiten:

Neue Funktionen:

- Zeitliche Bezüge
- Kombinationsanfragen (mehrere Probenarten)

LLM austauschen:

Integration von selbst gehostetem
Open-Source-Modellen

Fine-Tuning:
LLM-Training



Vielen Dank für Ihre Aufmerksamkeit!