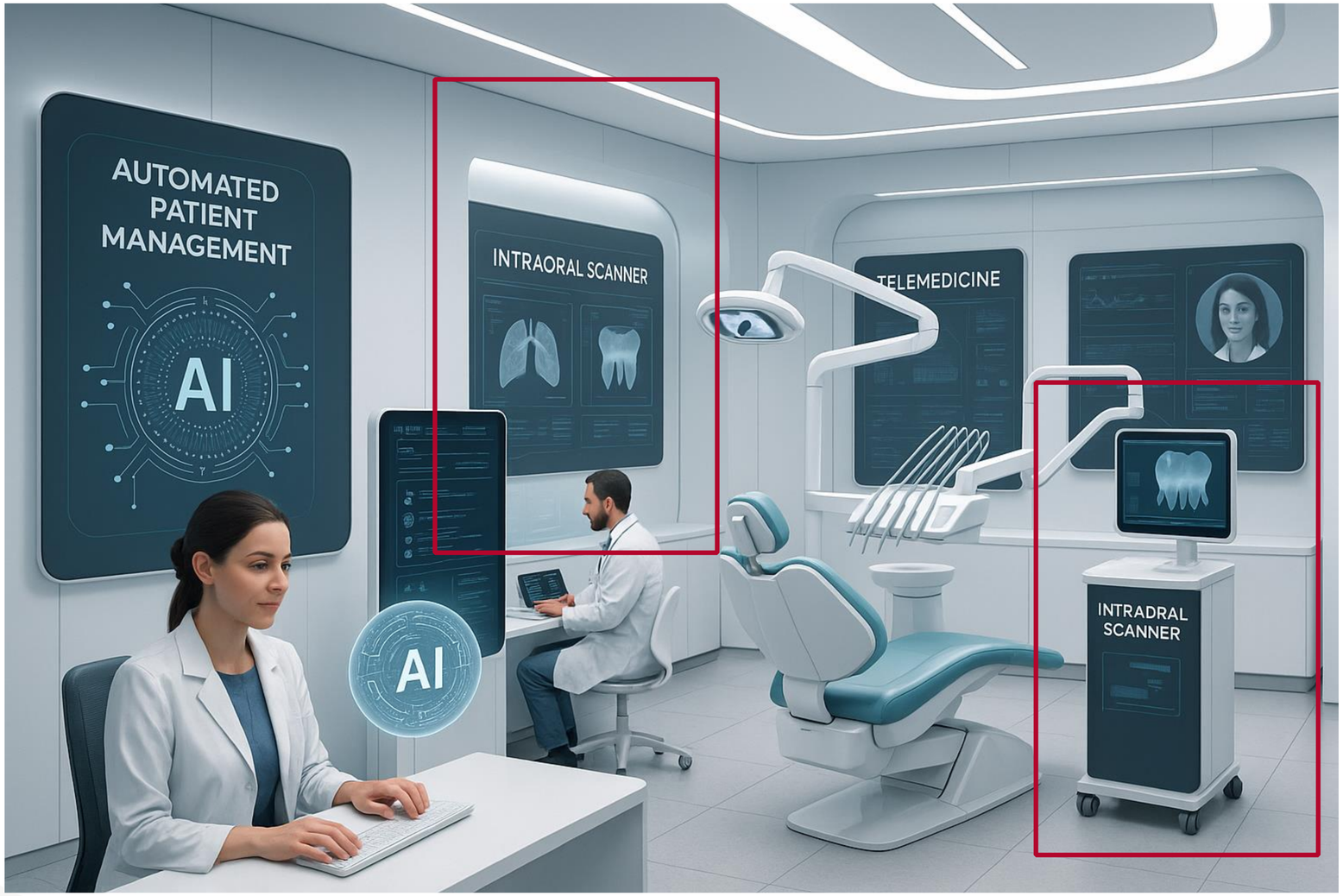
Elsevier Clinical Decision Support Symposium

MedBot vs. RealDoc: Effizienz großer Sprachmodelle in der Arzt-Patient-Kommunikation bei seltenen Krankheiten

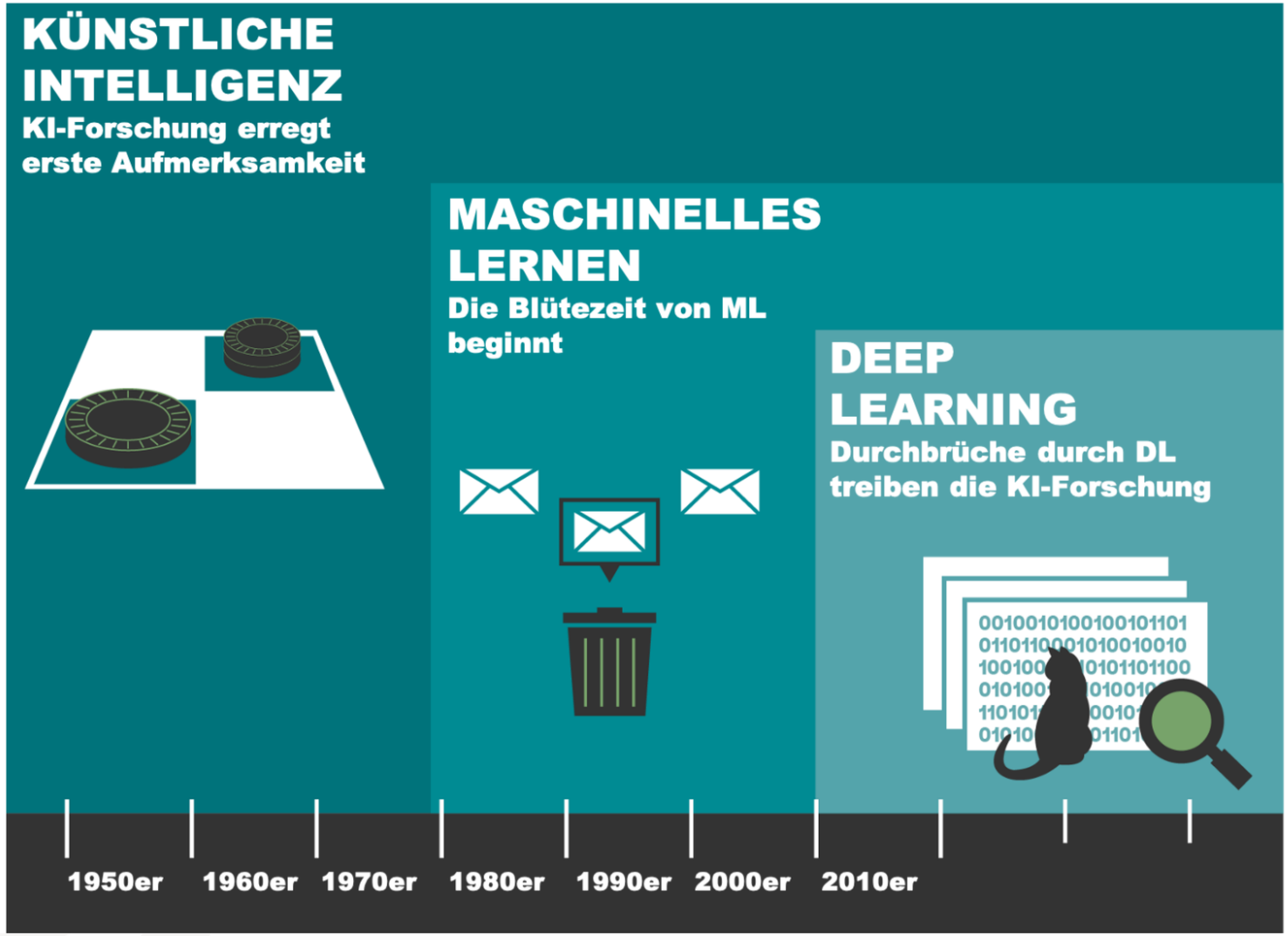
Prof. Dr. Jannik Schaaf, Institut für Medizininformatik, Goethe-Universität Frankfurt

Relevanz des Themas



[Hörzu, 14. Ausgabe, 28.03.24]

Relevanz des Themas



BUSINESS INSIDER PLUS WIRTSCHAFT KARRIERE POLITIK LEBEN NEWSLETTER PODCAST GRÜNDERSZENE Abonnement

HOME > LEBEN > CHAT GPT DIAGNOSTIZIERT KRANKHEIT – NACH DEM 17 ÄRZTE SCHEITERTEN

Chat GPT diagnostizierte die mysteriöse Krankheit eines Vierjährigen korrekt, nachdem 17 Ärzte versagt hatten

Kate Hull

Eine Mutter in den USA fragte Chat GPT nach einer Diagnose für die Symptome ihres Sohnes. (Symbolbild) ©Svetlana Iakusheva/Getty Images

COMPUTERWOCHE VOICE OF DIGITAL Alle aktuellen IDG Webcasts

Generative AI

Generative AI Hintergrund Ratgeber Bilder Video News

Generative AI in der Notaufnahme

Mit ChatGPT zu schnelleren Diagnosen

Von Manfred Bremmer (Senior Editor IoT & Mobile) FOLGEN

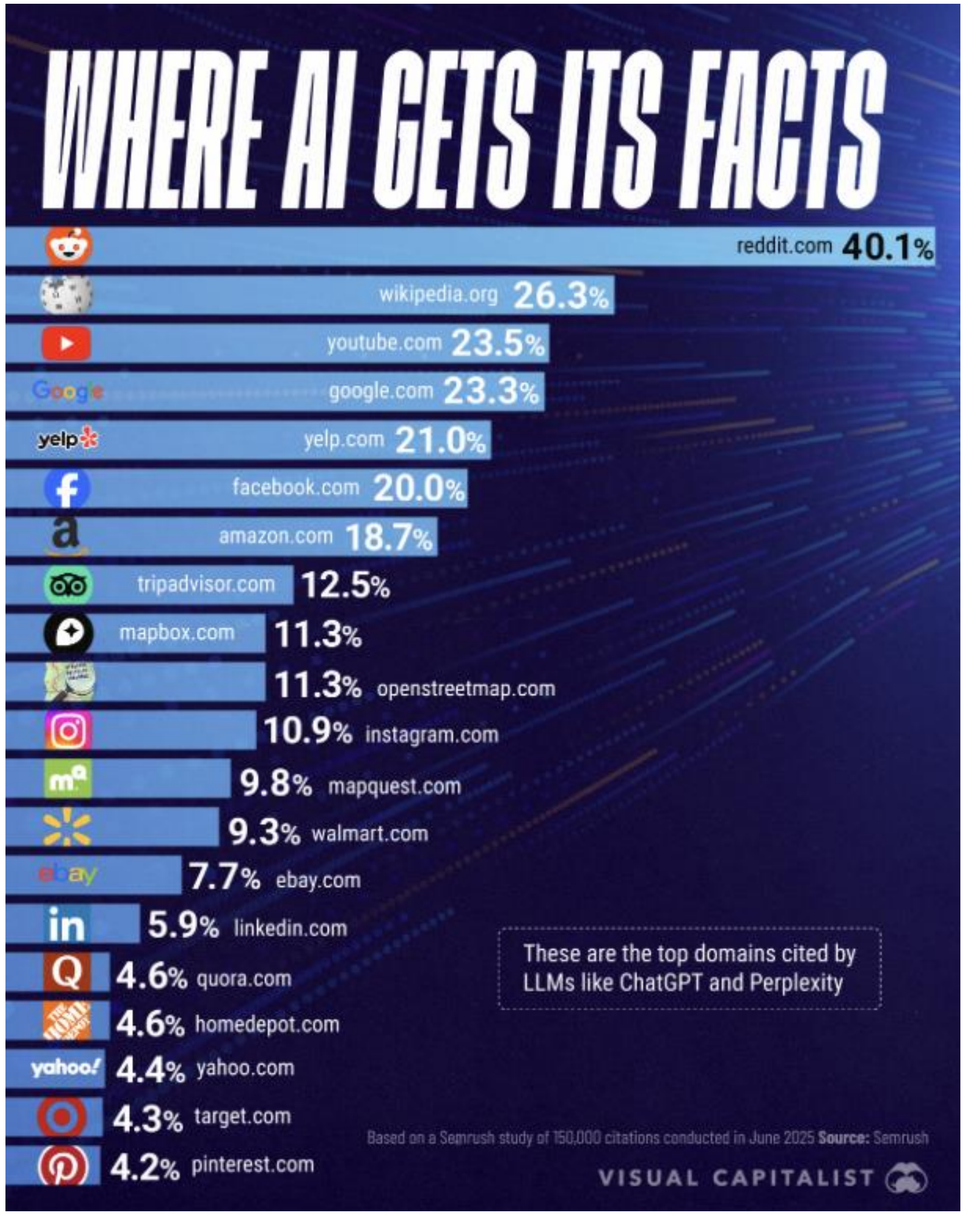
Obwohl nicht speziell dafür trainiert, ist ChatGPT laut einer aktuellen Studie offenbar gut dafür geeignet, wahrscheinliche Diagnosen für Patienten in der Notaufnahme vorzuschlagen.

Empfehlen
Drucken
PDF
URL
Xing
LinkedIn
Twitter
Facebook
Feedback

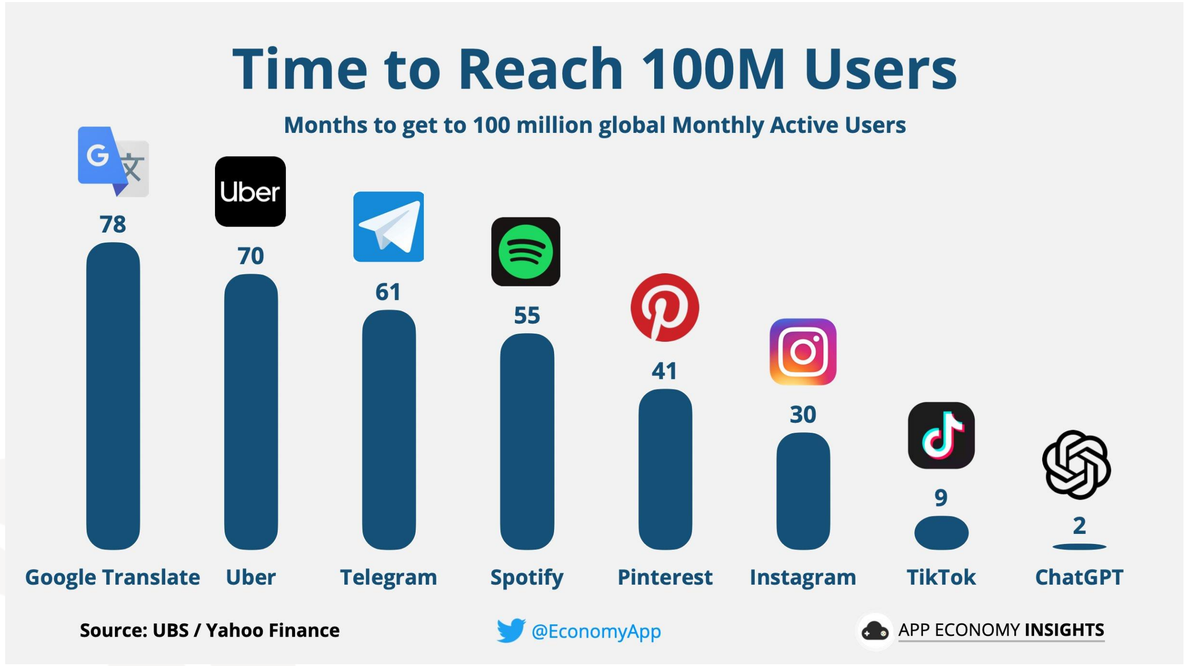
Auch in der Medizin gibt es etliche Anwendungsbereiche für (Generative) KI, beispielsweise in der Diagnostik.
Foto: chayanuphol - shutterstock.com

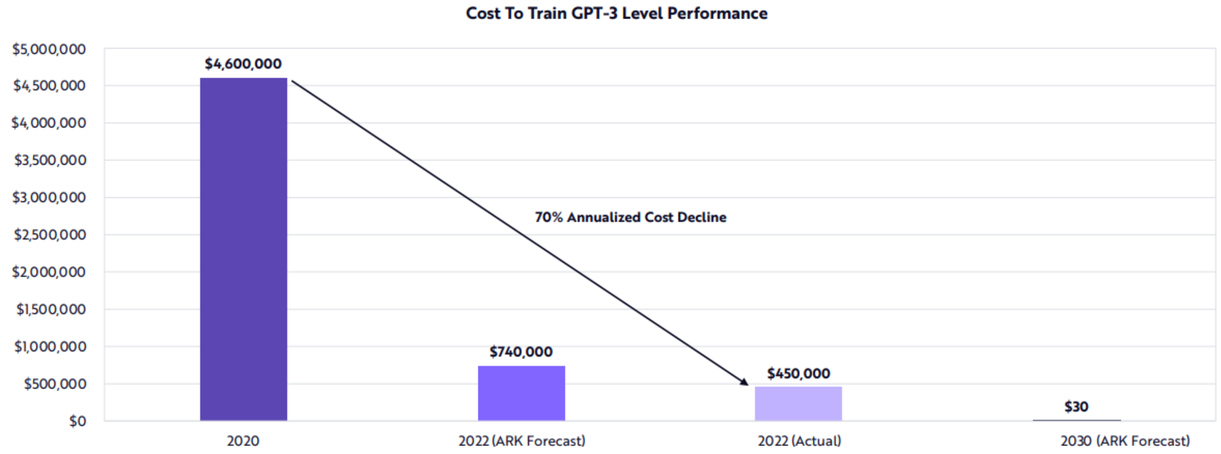
Die schnelle Diagnose von Patienten in der Notaufnahme ist entscheidend für die Verbesserung der Behandlungsergebnisse und die Verkürzung der Aufenthaltsdauer. Das Problem: Um eine genaue Diagnose zu stellen, müssen sich Ärzte auf verschiedene Informationsquellen stützen, darunter die Krankengeschichte, die körperliche Untersuchung, die Medikamenteneinnahme und weitere diagnostische Untersuchungen.

Relevanz des Themas



Künstliche Intelligenz





Künstliche Intelligenz

Meta will 600 Milliarden in KI-Infrastruktur in USA investieren

Um im Rennen um Künstliche Intelligenz mit Amazon, Microsoft und Co. mitzuhalten, will der Facebook-Mutterkonzern 600 Milliarden Dollar in US-Infrastruktur und Arbeitsplätze stecken.

07.11.2025 - 21:29 Uhr

Artikel anhören 02:04

Meta-Chef Zuckerberg: „Es ist die richtige Strategie, die Kapazitäten aggressiv vorab aufzubauen." Foto: AP

Studie MedDoc vs. Realbot

Journal of the American Medical Informatics Association, 2025, 1–9
https://doi.org/10.1093/jamia/ocaf034
Research and Applications

Research and Applications

MedBot vs RealDoc: efficacy of large language modeling in physician-patient communication for rare diseases

Magdalena T. Weber, MSc*,[1], Richard Noll, MSc*,[1], Alexandra Marchl, MSc[1], Carlo Facchinello, PhD[2], Achim Grünewaldt, PhD[3], Christian Hügel, PhD[4], Khader Musleh, PhD[3], Thomas O.F. Wagner, PhD[5], Holger Storf, PhD[1], Jannik Schaaf, PhD[1]

[1]Institute of Medical Informatics, University Medicine Frankfurt, Goethe University Frankfurt, Frankfurt 60590, Germany, [2]Santagostino Medical Center, Bologna 40138, Italy, [3]Department of Respiratory Medicine and Allergology, University Medicine Frankfurt, Goethe University Frankfurt, Frankfurt 60590, Germany, [4]HELIOS Dr Horst Schmidt Kliniken Wiesbaden, Klinik für Pneumologie, Wiesbaden 65199, Germany, [5]European Reference Network for Rare Respiratory Diseases (ERN-LUNG), University Medicine Frankfurt, Frankfurt 60590, Germany

*Corresponding authors: Magdalena T. Weber, MSc, Institute of Medical Informatics, University Medicine Frankfurt, Goethe University Frankfurt, Theodor-Stern-Kai 7, 60590 Frankfurt am Main, Germany (ma.weber@med.uni-frankfurt.de) and Richard Noll, MSc, Institute of Medical Informatics, University Medicine Frankfurt, Goethe University Frankfurt, Theodor-Stern-Kai 7, 60590 Frankfurt am Main, Germany (noll@med.uni-frankfurt.de)

Abstract

Objectives: This study assesses the abilities of 2 large language models (LLMs), GPT-4 and BioMistral 7B, in responding to patient queries, particularly concerning rare diseases, and compares their performance with that of physicians.

Materials and Methods: A total of 103 patient queries and corresponding physician answers were extracted from EXABO, a question-answering forum dedicated to rare respiratory diseases. The responses provided by physicians and generated by LLMs were ranked on a Likert scale by a panel of 4 experts based on 4 key quality criteria for health communication: correctness, comprehensibility, relevance, and empathy.

Results: The performance of generative pretrained transformer 4 (GPT-4) was significantly better than the performance of the physicians and BioMistral 7B. While the overall ranking considers GPT-4's responses to be mostly correct, comprehensive, relevant, and emphatic, the responses provided by BioMistral 7B were only partially correct and empathetic. The responses given by physicians rank in between. The experts concur that an LLM could lighten the load for physicians, rigorous validation is considered essential to guarantee dependability and efficacy.

Discussion: Open-source models such as BioMistral 7B offer the advantage of privacy by running locally in health-care settings. GPT-4, on the other hand, demonstrates proficiency in communication and knowledge depth. However, challenges persist, including the management of response variability, the balancing of comprehensibility with medical accuracy, and the assurance of consistent performance across different languages.

Conclusion: The performance of GPT-4 underscores the potential of LLMs in facilitating physician-patient communication. However, it is imperative that these systems are handled with care, as erroneous responses have the potential to cause harm without the requisite validation procedures.

Key words: artificial intelligence; medical informatics; natural language processing; health communication; rare diseases.

Background and significance

The advent of large language models (LLMs) led to a public comeback of artificial intelligence (AI), with profound implications for a variety of sectors, including health care.[1] Models such as OpenAI's generative pretrained transformer 4 (GPT-4) are capable of understanding and generating human-like text and performing various tasks, such as summarization, translation, and even coding.[2] Given their broad applicability, it seems feasible to advance automation in a variety of fields by employing LLMs, a prospect that is currently memorize prior conversations.[4] One of the principal benefits of employing LLMs in the health-care sector is their capacity to facilitate improved access to medical information. These models can provide round-the-clock support, tending to patients with immediate responses to common medical questions and helping to alleviate disparities in health-care access, particularly in underserved regions.[5] By augmenting traditional health-care services, LLMs can reduce the burden on health-care professionals.[6] In addition to convenience, LLMs can deliver personalized advice, drawing on vast databases of

UNIVERSITÄTS MEDIZIN FRANKFURT GOETHE-UNIVERSITÄT

GOETHE UNIVERSITÄT FRANKFURT AM MAIN

European Reference Network | Respiratory Diseases (ERN-LUNG)

FRZSE Frankfurter Referenzzentrum für Seltene Erkrankungen

ERN-LUNG addresses the following rare and complex pulmonary diseases and rare respiratory tumours

NON-CF BRONCHIECTASIS (NCFB)
Idiopathic Bronchiectasis (IB)
Allergic Bronchopulmonary Aspergillosis (ABPA)
Hyper igE syndrome
Rheumatoid Arthritis (RA)
Immunodeficiency associated bronchiectasis

PULMONARY HYPERTENSION (PH)
Chronic thromboembolic pulmonary hypertension (CTEPH)
Pulmonary arterial hypertension (PAH)
PH owing to lung disease and/or hypoxia
PH with unclear multifactorial mechanism
Pulmonary venoocclusive disease

MESOTHELIOMA (MSTO)

INTERSTITIAL LUNG DISEASES (ILD)
Idiopathic Pulmonary Fibrosis (IPF)
Sarcoidosis
Lymphangioleiomyomatosis (LAM)
Connective tissue diseases
Pulmonary alveolar proteinosis
Eosinophilic pneumonia
Langerhans Hystiocitosis X

PRIMARY CILIARY DYSKINESIA (PCD)
PCD including Reduced Generation of multiple Motile Cilia (RGMC)
Kartagener Syndrome

ALPHA-1-ANTITRYPSIN DEFICIENCY (AATD)
Pulmonary emphysema
Liver cirrhosis
Panniculitis

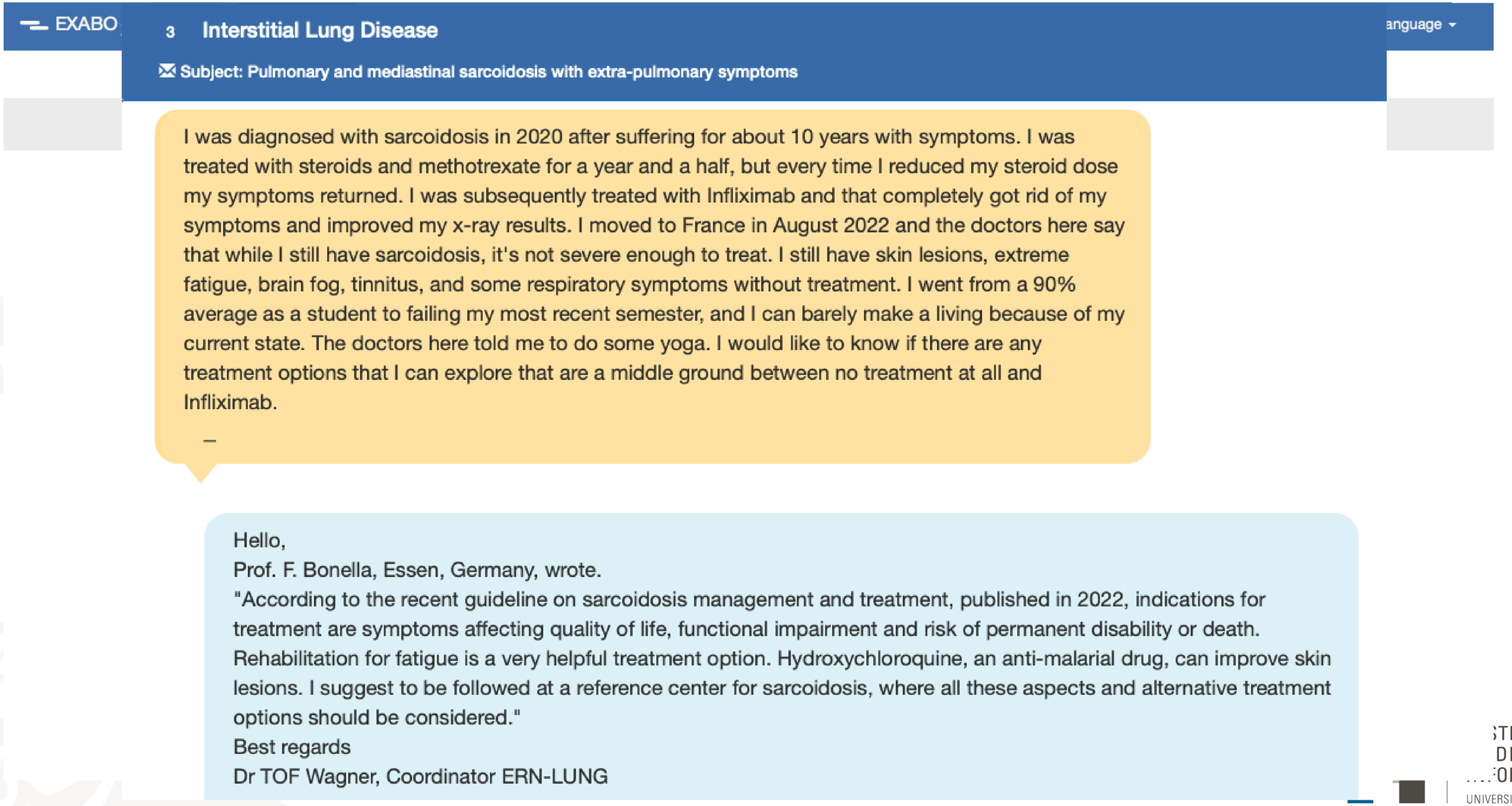
CYSTIC FIBROSIS (CF)

CHRONIC LUNG ALLOGRAFT DYSFUNCTION (CLAD)
Bronchiolitis obliterans syndrome (BOS)
Neutrophilic reversible allograft dysfunction
Restrictive allograft syndrome (RAS)

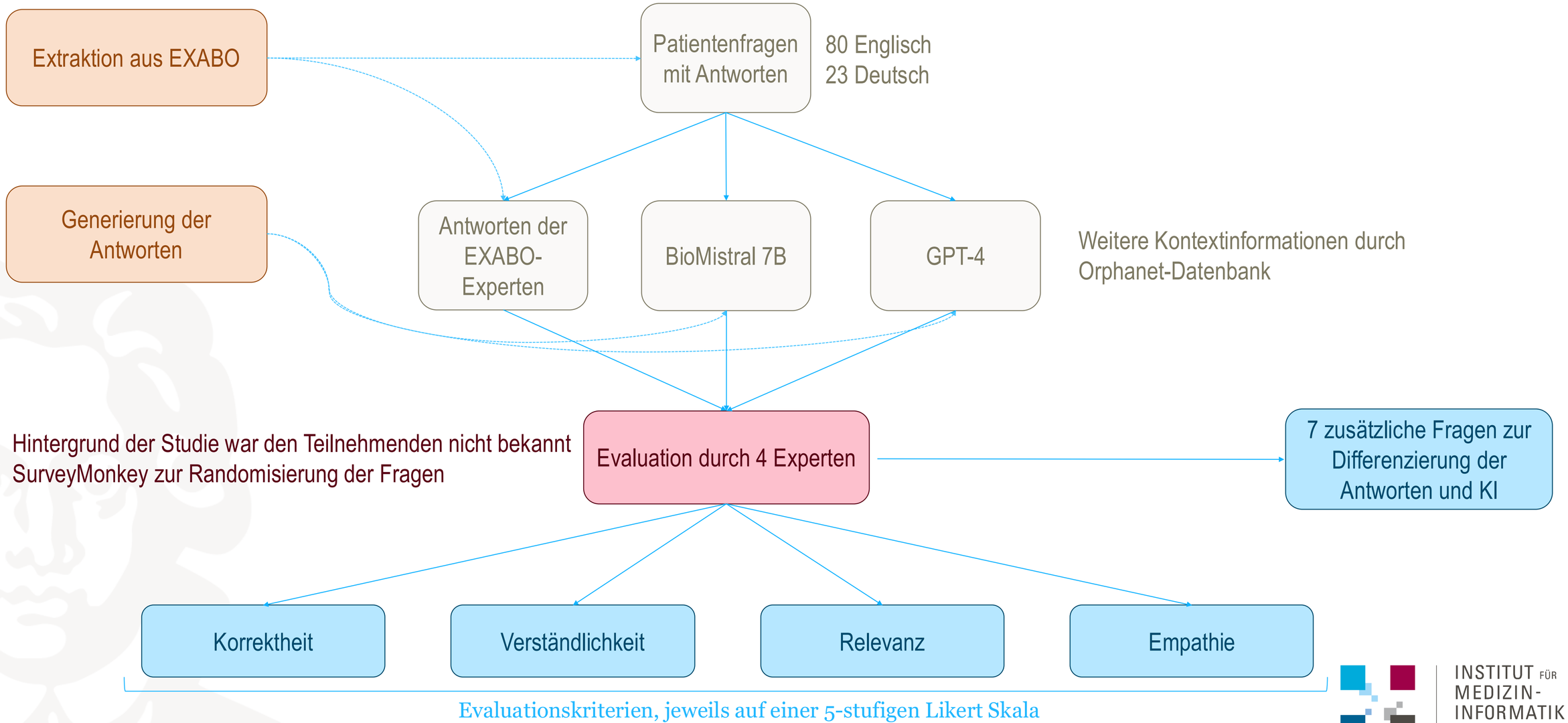
OTHER RARE LUNG DISEASES (ORLD)
Lung malformations
Disorders of respiratory drive (Ondine curse)
Acute Respiratory Distress Syndrome
Infantile Apnea

INSTITUT FÜR MEDIZIN-INFORMATIK
UNIVERSITÄTSMEDIZIN FRANKFURT

Studie MedDoc vs. Realbot

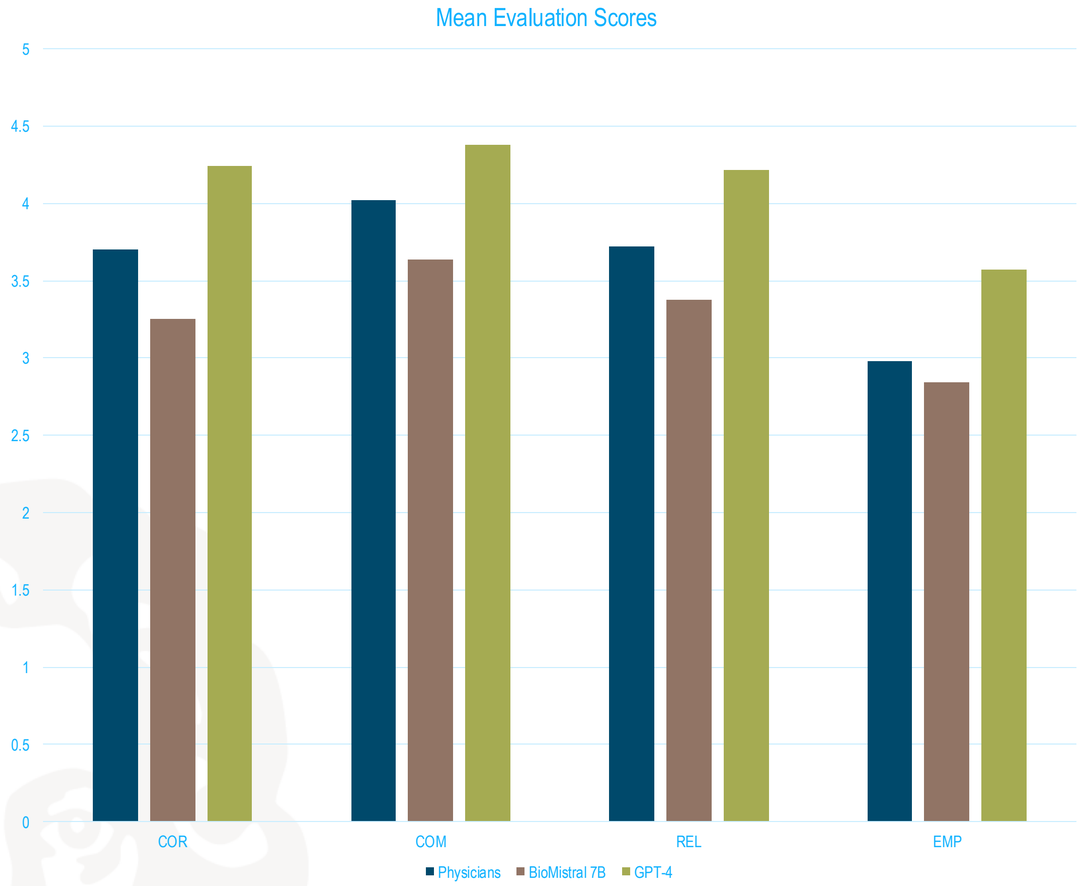


MedBot vs. RealDoc



INSTITUT FÜR MEDIZIN-INFORMATIK
UNIVERSITÄTSMEDIZIN FRANKFURT

Studie zur Large Language Modells



COR: Correctness COM: Comprehensibility
REL: Relevance EMP: Empathy

	GPT-4		BioMistral 7B		Physicians	
	Median	**Mean (SD)**	**Median**	**Mean (SD)**	**Median**	**Mean (SD)**
Correctness	4	4.24 (0.77)	3	3.25 (1.10)	4	3.70 (0.96)
Comprehensibility	4	4.38 (0.64)	4	3.64 (1.00)	4	4.02 (0.87)
Relevance	4	4.22 (0.75)	4	3.38 (1.06)	4	3.72 (0.93)
Empathy	4	3.57 (0.89)	3	2.84 (0.93)	3	2.98 (0.92)
Total		4.1 (0.76)		3.3 (1.02)		3.6 (0.92)

Intraklassenkorrelationskoeffizient (Reliabilität)

Correctness (0.728)
Comprehensibility (0.629)
Relevance (0.701)
Empathy (0.663)

Antworten im Fragebogen – Zusammenfassung

Frage	Antwort
Inwieweit könnten Chatbot-Antworten dazu beitragen, die Arbeitsbelastung von Ärzten zu verringern? Würden Sie eine solche Antwort als ersten Entwurf betrachten?	Alle Teilnehmer empfinden die vorgefertigten Antworten eines Chatbots für hilfreich und würden sie nutzen. Sie erwarten, dass er grundlegende Informationen aus der Literatur bereitstellt.
Welche potenziellen Risiken sehen Sie bei der Verwendung von Chatbots, insbesondere im Hinblick auf Fehlinformationen oder das Übersehen wichtiger Informationen?	Die Teilnehmer sind sich weitgehend einig, dass diese Risiken vermeidbar sind, wenn das Tool den Fachpublikationen und Richtlinien entspricht und einer gründlichen Validierung unterzogen wird.
Welche Überlegungen würden Sie anstellen, bevor Sie den Einsatz eines Chatbots in der Praxis in Betracht ziehen?	Die Teilnehmer würden das Risiko von Halluzinationen überprüfen, Tests durchführen und Aspekte wie Datenschutz, Benutzerfreundlichkeit und Kosten berücksichtigen.
Wie beurteilen Sie die ethischen Aspekte der Verwendung eines Chatbots, insbesondere im Hinblick auf die Vertraulichkeit von Patientendaten?	Die Teilnehmer sind sich einig, dass der Schutz der Privatsphäre der Patienten von größter Bedeutung ist. Sie gehen davon aus, dass man diesen Schutz gewährleisten kann.
Gibt es in den Antworten irgendwelche rückblickenden Merkmale oder Hinweise, die Sie vermuten lassen, dass sie von einem Arzt oder einem Chatbot stammen? Wenn ja, welche?	Die Teilnehmer geben formale Korrektheit, aber Unlogik, Unverständlichkeit, unvollständige Sätze, Allgemeinheit und Studienzitate als Hinweise für Chatbot-Antworten an.
Welche rechtlichen Herausforderungen sehen Sie bei der Implementierung eines KI-basierten Chatbots in der Arzt-Patient-Kommunikation?	Die Meinungen der Teilnehmer zu diesem Thema gehen auseinander. Einige verweisen auf Datenschutzprobleme, während andere optimistisch sind, dass der Chatbot als Unterstützungssystem umgesetzt werden kann.

Diskussionen

- GPT4 erzielte signifikant bessere Ergebnisse
- Hohe Empathie; jedoch kulturell eingeschränkt
- Biomistral 7B signifikant schlechter; mögliche Ursache geringe Parameteranzahl und Defizite im Deutschen
- Biomistral 7B kann jedoch lokal betrieben werden
- Gefahr von Halluzinationen sind groß
- LLMs in der Patientenkommunikation könnten Zeitdruck minimieren und fehlendes Wissen kompensieren
- Begrenzte Repräsentativität: Nur 4 Ärzte als Bewertende, jedoch moderate bis hohe Inter-Rater-Reliabilitäten

Vielen Dank für Ihre Aufmerksamkeit!

Prof. Dr. Jannik Schaaf

Johann Wolfgang Goethe-Universität Frankfurt
Universitätsmedizin Frankfurt
Institut für Medizininformatik – IMI

Haus 4
Theodor-Stern-Kai 7
60590 Frankfurt

schaaf@med.uni-frankfurt.de
www.imi-frankfurt.de

INSTITUT FÜR MEDIZIN-INFORMATIK
UNIVERSITÄTSMEDIZIN FRANKFURT