ECONSTOR
Make Your Publications Visible.

A Service of

ZBW Leibniz-Informationszentrum Wirtschaft
Leibniz Information Centre for Economics

Erker, Christian G. et al.

Working Paper

"Wie geht es Ihnen heute, Mr. Bot?": Virtuelle Patient:innen in Large Language Models richtig prompten

IU Discussion Papers - Gesundheit, No. 1 (Oktober 2024)

Provided in Cooperation with:
IU International University of Applied Sciences

Suggested Citation: Erker, Christian G. et al. (2024) : "Wie geht es Ihnen heute, Mr. Bot?": Virtuelle Patient:innen in Large Language Models richtig prompten, IU Discussion Papers - Gesundheit, No. 1 (Oktober 2024), IU Internationale Hochschule, Erfurt

This Version is available at:
https://hdl.handle.net/10419/305262

Standard-Nutzungsbedingungen:

Die Dokumente auf EconStor dürfen zu eigenen wissenschaftlichen Zwecken und zum Privatgebrauch gespeichert und kopiert werden.

Sie dürfen die Dokumente nicht für öffentliche oder kommerzielle Zwecke vervielfältigen, öffentlich ausstellen, öffentlich zugänglich machen, vertreiben oder anderweitig nutzen.

Sofern die Verfasser die Dokumente unter Open-Content-Lizenzen (insbesondere CC-Lizenzen) zur Verfügung gestellt haben sollten, gelten abweichend von diesen Nutzungsbedingungen die in der dort genannten Lizenz gewährten Nutzungsrechte.

Terms of use:

Documents in EconStor may be saved and copied for your personal and scholarly purposes.

You are not to copy documents for public or commercial purposes, to exhibit the documents publicly, to make them publicly available on the internet, or to distribute or otherwise use the documents in public.

If the documents have been made available under an Open Content Licence (especially Creative Commons Licences), you may exercise further usage rights as specified in the indicated licence.

WWW.ECONSTOR.EU

Mitglied der Leibniz-Gemeinschaft

INTERNATIONALE HOCHSCHULE

www.iu.de

IU DISCUSSION PAPERS

Gesundheit

„Wie geht es Ihnen heute, Mr. Bot?“ – virtuelle Patient:innen in Large Language Models richtig prompten

CHRISTIAN G. ERKER
DANIELA BECKER
MAREN KUHNE
CORNELIA SCHLICK
CLAUDIA MIERSCH
KATHARINA RÄDEL-ABLASS
KLAUS SCHLIZ
KERSTIN VOGEL

IU Internationale Hochschule

Main Campus: Erfurt

Juri-Gagarin-Ring 152

99084 Erfurt

Telefon: +49 421.166985.23

Fax: +49 2224.9605.115

Kontakt/Contact: christian.erker@iu.org

Autorenkontakt/Contact to the author(s):
Prof. Dr. med. Christian G. Erker
ORCID-ID: 0000-0001-9912-115X (Open Researcher und Contributer ID)

IU Internationale Hochschule - Fernstudium

Juri-Gagarin-Ring 152

99084 Erfurt

Email: christian.erker@iu.org

IU Discussion Papers, Reihe: Gesundheit, Vol. 2, No. 1 (OKT 2024)

ISSN-Nummer: 2750-0705

Website: https://www.iu.de/forschung/publikationen/

„Wie geht es Ihnen heute, Mr. Bot?“

Virtuelle Patient:innen in large language models richtig prompten

Christian G. Erker

Daniela Becker

Maren Kuhne

Cornelia Schlick

Claudia Miersch

Katharina Rädel-Ablass

Klaus Schliz

Kerstin Vogel

ABSTRACT:

Die Anamneseerhebung ist eine wesentliche Fertigkeit in allen Gesundheitsberufen. Die Ausbildung in dieser Fähigkeit im Rahmen eines Studiums ist jedoch mit zahlreichen Hürden verbunden. Virtuelle Patient:innen, mit denen Gespräche auf der Grundlage von „Large Language Models“ (LLM) geführt werden können, könnten viele dieser Probleme lösen. Voraussetzung dafür ist jedoch ein detaillierter „Prompt“, der ausführliche Hintergrundinformationen über die/den virtuellen Patient:in liefert, auf die das LLM bei der Beantwortung zugreifen kann. Dieser Beitrag liefert Eckpunkte für die Erstellung eines solchen Prompts, der als Grundlage und Unterstützung für die Erstellung eigener virtueller Patient:innen dienen soll.

Taking a medical history is an essential skill in all medical professions. However, training this skill as part of a degree course has numerous hurdles. Virtual patients, with whom conversations can be conducted on the basis of “large language models” (LLM), could solve many of these problems. However, this requires a detailed “prompt” that provides detailed background information on the virtual patient, which the LLM can access when responding. This paper provides key points for the creation of such a prompt, which should serve as a basis and support for the creation of your own virtual patients.

KEYWORDS:

Medical history, medical education, artificial intelligence, large language model, examination teaching

AUTOR:INNEN

Christian G. Erker ist Professor im Studiengang „Physician Assistant (B.Sc.)“ an der IU Internationale Hochschule im Fernstudium. Seine Begeisterung ist innovative und studierendenzentrierte Lehre, vor allem zu akut- und notfallmedizinischen Themen wie Anästhesie und Notfallmedizin.

Er führt die Facharztbezeichnungen in Anästhesiologie und Kinder- und Jugendmedizin sowie zahlreiche Schwerpunktbezeichnungen mit notfall- und intensivmedizinischem Fokus und blickt auf Berufserfahrung als Oberarzt in der Anästhesie eines akademischen Lehrkrankenhauses und Notarzt zurück. Begleitend war er in der praxisorientierten Lehre in nahezu allen medizinischen Fachberufen tätig.

Einen wissenschaftlichen Schwerpunkt bildet die notfallmedizinische Versorgung von Kindern, wozu er u.a. Medikamentenreferenzen, Leitlinienarbeit und Podcasts beigesteuert hat, sowie die Medizinische Ausbildungsforschung.

Daniela Becker leitet an der IU Internationale Hochschule den Fernstudiengang Physician Assistant (B.Sc.). 2012 erwarb sie die Facharztbezeichnung “Neurochirurgie” und verfügt über die Zusatzbezeichnung “spezielle Intensivmedizin”. Sie war langjährig als Fach-, bzw. Oberärztin an den Universitätskliniken Erlangen, Marburg und Heidelberg tätig. Aktuell arbeitet sie in Nebentätigkeit auf einer neurologischen Intensivstation. Sowohl chirurgisch als auch wissenschaftlich liegt Ihre Expertise in der Chirurgie hirneigener Tumore. Hier erforschte sie die intraoperative Anwendung von Bildgebung und Bildverarbeitungstechniken, um die Patientensicherheit zu erhöhen. Sie erwarb die Venia legendi im Fach Neurochirurgie im Jahr 2021. Sie widmete sie sich bereits frühzeitig der studentischen Lehre. Insbesondere absolvierte sie zahlreiche hochschuldidaktische Kurse und erwarb schließlich das Baden-Württemberg Zertifikat für Hochschuldidaktik im Jahr 2022.

Forschungsschwerpunkte:

- Implementierung und Anwendung intraoperativer Magnetresonanztomographie und multimodaler Neuronavigation in der Neurochirurgie
- Faserbahntraktographie auf der Basis diffusionsgewichteter MRT-Sequenzen und deren intraoperative Anwendung in der Neurochirurgie
- Konnektivität
- Patientensicherheit in der Gliomchirurgie

Maren Kuhne ist promovierte wissenschaftliche Mitarbeiterin im Studiengang Humanmedizin an der MSB Medical School Berlin. Ihre Passion liegt in der Konzeption innovativer Lehrkonzepte. Sie schafft es, durch ihre eigene Begeisterung für biologische und biochemische Themen und den Einsatz modernster didaktischer Methoden die Studierenden mitzureißen.
Neben der Lehre in der Humanmedizin und Biotechnologie liegt ihr wissenschaftlicher Schwerpunkt in der immunologisch-zellbiologischen Grundlagenforschung. Sie kann jahrelange Forschungstätigkeit im Bereich der Herstellung monoklonaler Antikörper sowie eine Promotion an der FU Berlin/Charité – Universitätsmedizin Berlin auf dem Gebiet der T-Zell-Immunologie nachweisen.

Cornelia Schlick ist seit dem 01.05.2021 Professorin an der IU und leitet den Studiengang Physiotherapie. Nach Abschluss ihres Bachelorstudiums in Physiotherapie und Masterstudiums in Public Health war sie mehrere Jahre am Klinikum der Universität München (LMU) beschäftigt und schloss dort ihre Promotion zum Thema Sturzprädiktion und Mobilitätsmonitoring bei neurologischen Erkrankungen ab. Im weiteren Verlauf war sie als Applikationsspezialistin bei einem Hersteller für Reha-Technologien tätig und engagierte sich im Rahmen von Lehraufträgen an verschiedenen Hochschulen. Anfang 2022 war sie an der Initiierung der Arbeitsgruppe zur Virtuellen Hochschulklinik beteiligt und arbeitet im Incubator-Projekt „Entwicklung von KI-basierten Patient:innenmodellen für das interprofessionelle Lernen in den Gesundheitsstudiengängen“.

Claudia Miersch ist Professorin für Ernährungsphysiologie und Diätetik im gleichnamigen Fernstudiengang der IU Internationale Hochschule. Nach ihrem Abschluss in Ökotrophologie und erfolgreicher Promotion forschte sie mehrere Jahre als wissenschaftliche Mitarbeiterin am Leibniz-Institut für Nutztierbiologie. Im Rahmen ihrer Forschungsarbeiten beleuchtete sie das Thema Ernährung aus unterschiedlichen Blickwinkeln: Von Krebsforschung, über Nutrigenomik bis zur Stammzellforschung hat sie sich mit vielen Auswirkungen der Nahrungsaufnahme auf den Organismus beschäftigt. Neben ihrer wissenschaftlichen Tätigkeit arbeitete sie als Wissenschaftsredakteurin, leitete die Forschungsabteilung eines Online-Gesundheitsportals und unterstützt bis heute als selbstständige Dozentin und Ernährungsberaterin Menschen bei der Entwicklung einer besseren Ernährungs- und Gesundheitskompetenz. Durch persönliche Erfahrungen entwickelte sie eine umfangreiche Expertise auf dem Gebiet des Schwangerschaftsdiabetes, die sie in neue Forschungsprojekte einfließen lässt und betroffenen Frauen auf einem fachlichen Blog zur Verfügung stellt. Seit 2022 ist sie Mitglied der Virtuellen Hochschulklinik und entwickelt zusammen mit ihren Kolleg:innen KI-basierte Patientenmodelle für das interprofessionelle Lernen in den Gesundheitsstudiengängen.

Katharina Rädel-Ablass ist seit dem 1. Oktober 2021 als Professorin für Pflege an der IU Internationale Hochschule tätig. Zuvor war sie als Vertretungsprofessorin für Pflegewissenschaft und Klinische Pflege an der Ernst-Abbe-Hochschule Jena im Fachbereich Gesundheit und Pflege beschäftigt. Im Rahmen des Graduiertenkollegs „Multimorbidität im Alter“ der Charité – Universitätsmedizin Berlin promovierte sie zum Thema „Soziale Konstruktionen zur Pflegeübernahme – Eine Studie aus der Perspektive von pflegenden Angehörigen und Pflegebedürftigen“. Sie verfügt über langjährige Erfahrung als Hochschuldozentin im Pflegebereich sowie als freiberufliche Coach und Organisationsberaterin für Personen und Organisationen im Gesundheitswesen. Als Mitglied der Virtuellen Hochschulklinik der IU arbeitet sie seit 2023 an dem Incubator-Projekt „Entwicklung von KI-basierten Patient:innenmodellen für das interprofessionelle Lernen in den Gesundheitsstudiengängen“.

Klaus Schliz ist gelernter Krankenpfleger und Rettungsassistent. Er hat Pflegemanagement und medizinische Wissenschaften studiert und ist seit 2021 Professor für Pflegemanagement an der IU Internationale Hochschule. Dort ist es Gründungsmitglied des Innovationszentrum Pflege und Gerontologie (IPG). Seit über 30 Jahren ist er Inhaber und Geschäftsführer eines mittelgroßen ambulanten Pflegedienstes. Im Rahmen der virtuellen Hochschulklink an der IU ist er Mitentwickler eines Chatbots, der reale Patienten perspektivisch durch KI ersetzen soll, um interprofessionelle Anamnesegespräche führen zu können. Einen wissenschaftlichen Schwerpunkt bildet die Untersuchung von Fremdeinschätzung von Bedürfnissen durch Pflegefachkräfte für ihre pflegebedürftigen Klienten. Außerdem forscht er zum Thema Rettungs- und Feuerwehrleitstellen im Bereich standardisierte Notrufabfrage und Telefonreanimation, zu dem er auch promoviert hat.

Kerstin Vogel hat Betriebswirtschaftslehre mit Schwerpunkt Controlling studiert. Seit 2012 ist sie an der IU Internationale Hochschule tätig, wo sie in verschiedenen Bereichen durch ihr Projektmanagement die innovative Entwicklung der Hochschule mit trägt. Seit dem Frühjahr 2023 widmet sie sich intensiv der Arbeit mit generativer künstlicher Intelligenz. Ihr Fokus liegt dabei auf der Entwicklung von Strategien für den Einsatz dieser zukunftsweisenden Technologie im Bereich der Hochschullehre. Mit ihrer langjährigen Erfahrung in der digitalen Bildung und ihrem betriebswirtschaftlichen Hintergrund leistet sie einen wichtigen Beitrag zur Integration von KI in moderne Lehr- und Lernkonzepte der IU.

Einführung

Ein wichtiger Skill, der in allen medizinischen und gesundheitswissenschaftlichen Curricula enthalten ist, ist die Anamneseerhebung. Die bisherige Datenlage zum Erlernen dieser Fertigkeit im Rahmen eines Hochschulstudiums ist jedoch spärlich (Gilligan et al., 2021). Der hohe Zeit- und Ressourcenaufwand eines Dialogtrainings, beispielsweise mit Schauspieler:innen, steht in Kontrast zu dafür vorgegebenen curriculären Rahmenbedingungen. Obwohl dieser Skill essentiell ist für medizinische, therapeutische und pflegerische Tätigkeiten (Peterson et al., 1992) und es sowohl Lehr-/Lernmethoden (Lipkin Jr et al., 1995) als auch Assessmentmethoden (Makoul, 2001) gibt, gehen viele Absolvent:innen eher unvorbereitet ins Berufsleben.

Seit der Einführung von Künstlicher Intelligenz (KI) und der breiteren Bekanntheit z.B. von ChatGPT sind eine Vielzahl von Einsatzgebieten, auch im medizinischen Kontext, beschrieben worden (Alowais et al., 2023; Sun et al., 2023). Sie ermöglicht eine dem Menschen angenäherte Kommunikation auf Basis von sog. „Large Language Models“ (LLM), deren vornehmliche Charakteristik die Produktion von Sprache ist. Was läge also näher, als Patient:innengespräche nicht mit „echten“ Menschen, sondern mit KI-generierten „virtuellen“ Patient:innen zu üben? Die Vorteile sind dabei vielfältig: Virtuelle Patient:innen zum Gespräch stehen räumlich und zeitlich uneingeschränkt „on demand“ zur Verfügung, sie erfordern – einmal korrekt gepromptet – keine regelmäßige Schulung und bieten ein homogenes Antwortmuster. Sie können nahezu alles gefragt werden, womit auch Barrieren bzgl. als unangenehm oder distanzlos wahrgenommene Fragen reduziert werden können. Das Spektrum und die Vielzahl der simulierten klinischen Fälle sowie die Kosteneffektivität sind weitere wesentliche Faktoren, die virtuelle Patient:innen von Schauspieler:innen unterscheiden. Auch KI-Modelle kosten während der Benutzung Geld, die sich in Serverzeit und Tokens (vergleichbar mit der Anzahl der durch die KI ausgegebenen Silben) repräsentiert findet (Allwein, 2024).

An der IU Internationale Hochschule, die schon früh KI-Konzepte in den hochschulischen Alltag integrierte, gibt es verschiedene Ansätze, KI-generierte Gespräche mit virtuellen Patient:innen zu ermöglichen („Entwicklung von KI-basierten Patient:innenmodellen für das interprofessionelle Lernen in den Gesundheitsstudiengängen“: https://www.iu.de/forschung/projekte/ki-patient-ipl/; „Boost your Skills! AnamneseBOosT“: https://uc2456.customervoice360.com/uc/Team_Christian_Erker/3775/). Kernpunkt dabei ist ein möglichst umfassender und hinreichend konkreter Prompt, der dem LLM Vorgaben zur Gestaltung der Antwort liefert (White et al., 2023). Damit das Antwortverhalten konsistent und homogen ist und weite Abschweifungen vom Thema minimiert werden, ist eine Vielzahl von Aspekten bei der Ausgestaltung des Prompts zu beachten. In diesem Discussion Paper fassen die Autor:innen ihre bisherigen Erfahrungen mit Prompts zusammen und führen die wichtigen Aspekte gemeinsam mit einem strukturgebendem Rahmen zusammen, die bei der Erstellung eigener Chatbots, beispielsweise zum Training von Anamnesegesprächen, hilfreich sein können. Diese sind im Wesentlichen als Leitfaden zu verstehen und sollen vor allem zum Ausprobieren und zur Diskussion anregen.

Hauptteil: Der Prompt

Für die praktische Umsetzung wird ein LLM benötigt, welches sowohl auf extern zugeführte Daten als auch einen allgemeinen Wissensfundus zurückgreifen kann. Man unterscheidet ferner Anweisungen an die Nutzenden, die dem System verborgen bleiben, aber im Sinne von Instruktionen genutzt werden können, von Anweisungen an das System (die den Nutzenden verborgen bleiben und „im Hintergrund“ agieren). Je nach verwendetem KI-Modell sind die Begrifflichkeiten hierfür z.T. unterschiedlich gewählt und nicht normiert (z.B. „opening message“ für die „Anweisung an die Nutzenden“).

ANWEISUNG AN DIE NUTZENDEN („INTRODUCTION“)

Im Bereich „Introduction“ formuliert man zunächst einen einführenden Text, um die Nutzenden in das Szenario zu führen, die virtuelle Rahmensituation darzustellen und in das Gespräch einzusteigen. Wichtig: Der Text muss so eingefügt werden, dass er nicht Teil der Konversation wird – das könnte zu einem ungewollten Rollenwechsel des Bots führen. Ein Hinweis auf eine aktive Beendigung des Gespräches ermöglicht das Triggern einer Abschlussmeldung, in der zusätzliche Hinweise platziert werden können. Am Ende der Introduction sollte ein Starter für das Gespräch stehen.

Beispiel:

*Als nächster Patient in der Anästhesiesprechstunde kommt Hr. Hinkmann (*Person und Details frei erfunden) zu dir, der morgen geplant durch die Gefäßchirurgie operiert werden soll. Führe mit ihm ein vorbereitendes Anamnesegespräch und versuche möglichst viele anästhesierelevante Informationen zu erhalten.*

Probiere dabei gerne sowohl offene als auch geschlossene Fragestile aus. Versuche auch mal, dein Gespräch zu strukturieren und behalte dabei die Gesprächsdauer mit im Auge. Ob Du alle wichtigen Informationen herausgefunden hast, wirst du später in einem Live-termin erfahren.

Bitte beende das Gespräch angemessen mit einer Verabschiedung. Viel Spaß!

“Guten Tag, mein Name ist Hinkmann, ich sollte mich hier zum Vorgespräch melden.“

ANWEISUNG AN DAS SYSTEM („CHATGPT-4-PROMPT“)

Ein Chatbot ist eine Softwareanwendung, die es Nutzern ermöglicht, in natürlicher Sprache mit einem Computersystem zu interagieren. Moderne Chatbots basieren häufig auf sogenannten "Large Language

Models" (LLMs), großen KI-Modellen, die aus riesigen Datenmengen gelernt haben, wie Sprache funktioniert. Um diese Chatbots zu steuern, geben Nutzer Eingaben, sogenannte "Prompts", in Form von Fragen oder Befehlen. Das "Prompt Engineering" ist die Kunst und Wissenschaft, diese Eingaben so zu gestalten, dass der Chatbot präzise und relevante Antworten liefert. Dabei spielt der Kontext, die Klarheit und die spezifische Zielsetzung der Eingabe eine entscheidende Rolle, um die gewünschte Antwort zu erhalten. Die einfachsten Befehlseingaben in ein LLM sind dabei „zero-shot-prompts“: Kurze direkte Eingaben ohne zusätzlichen Kontext. Das System muss sich den Inhalt aus dem zuvor trainierten Basis-Datensatz generieren, was in allgemein gehaltenen Antworten resultiert(OpenAI Platform, o. J.). „Chatbots“ stellen eine zusätzliche Ebene dar, die sich rund um das LLM herum aufbaut und dem Modell eine grundlegende Persönlichkeit, Einstellung, Haltung und aber auch zusätzlichen Inhalt bereitstellt, um spezifisch vordefinierte Aufgaben zu lösen. Dafür benötigt man eine zusätzliche Instruktion an das System, um diese Informationen bereitzustellen. Diese sind im Allgemeinen für die Endnutzenden nicht sichtbar. Um die hintergründigen Chatbot-Instruktionen zu formulieren, gibt es eine Vielzahl von Ansätzen, denen im Wesentlichen gemeinsam ist, dass sie mindestens Informationen zum Charakter oder der einzunehmenden Rolle, Hintergrundinformationen und eine Aufgabenstellung formulieren(Lakhyani, 2024; Liu et al., 2023). Ein wissenschaftlich fundierter Vergleich unterschiedlicher Methoden mit dem Nachweis einer Überlegenheit eines spezifischen Ansatzes existiert bislang nicht(Liu et al., 2023; Sahoo et al., 2024). Der hier ausgewählte „BITE“-Ansatz (Bhattacharya, not yet published) ist eine dieser Möglichkeiten, den Instruktionen eine Struktur zu verleihen, und ist die von der IU Internationale Hochschule in AI-Trainingsformaten verwendete Methode. Der Vorschlag für die Struktur orientiert sich grob an diesen vier Kernpunkten:

Behaviour: Definieren, wie der Chatbot auf Anfragen reagieren soll (z. B. freundlich, professionell).

Interaction: Festlegen, wie der Chatbot mit Benutzer:innen interagieren soll (z. B. fragenbasierte Navigation, proaktive Hilfe).

Task Elaboration: Detaillieren der spezifischen Aufgaben, die der Chatbot erfüllen soll. (z. B. FAQs beantworten, Buchungen vornehmen).

External context: Einbeziehen von externen Informationen, die der Chatbot berücksichtigen muss (z.B. aktuelle Datenbanken, externe APIs). Da dies einen breit erweiterbaren Anteil des Prompts darstellt, wird er idealerweise hintenangestellt, damit zusätzliche Informationen einfach nachträglich angehängt werden können, ohne die wichtigen initialen drei Aspekte zu beeinträchtigen.

Bei virtuellen Patient:innen wird der Anteil „EXTERNAL CONTEXT“ bei weitem überwiegen. Vorteilhaft sind kurze und direkte Sätze, um dem Bot sein Verhalten anzuzeigen.

#BEHAVIOUR

Der Teil „Behaviour“ kontrolliert dabei das Grundverhalten (ChatGPT-Standard ist „a helpful assistant“) und damit die Rolle als Patient:in, in die die KI schlüpfen soll. Hier können Hinweise zur verwendeten Sprache („verwende Laiensprache, Jugendsprache“) und zur Länge der Antwort („antworte mit einem Satz“, „schweife weit aus“) platziert werden. Außerdem werden persönliche Faktoren festgelegt, zum

Beispiel bestimmte Grundeinstellungen oder eine emotionale Färbung des Patienten. Es ist praktisch, im gesamten Text den Befehl an den Bot in der „DU"-Form zu geben – „Du bist ein Patient", „Du hast Schmerzen", damit die/der Autor:in sich beim Schreiben das Gegenüber gut vorstellen kann und nicht auf einen ungewollten Rollenwechsel achten muss.

Beispiel:

#BEHAVIOUR

Du bist ein Patient in einem Krankenhaus. Dein Name ist Herr Josef Hinkmann, Du führst ein Gespräch mit einem Arzt in der Narkosesprechstunde.

WICHTIG: Weiche unter keinen Umständen von deiner Rolle als Patient ab und antworte immer aus der Perspektive des Patienten!

Bitte sprich Deinen Gesprächspartner mit „Sie" an. Bitte antworte auf Fragen jeweils mit nur einem Satz und verwende dabei Laiensprache. Du kennst nicht alle medizinischen Fachbegriffe. Frage nach der Bedeutung, wenn Du gezielt mit vielen Fachbegriffen oder Abkürzungen befragt wirst.

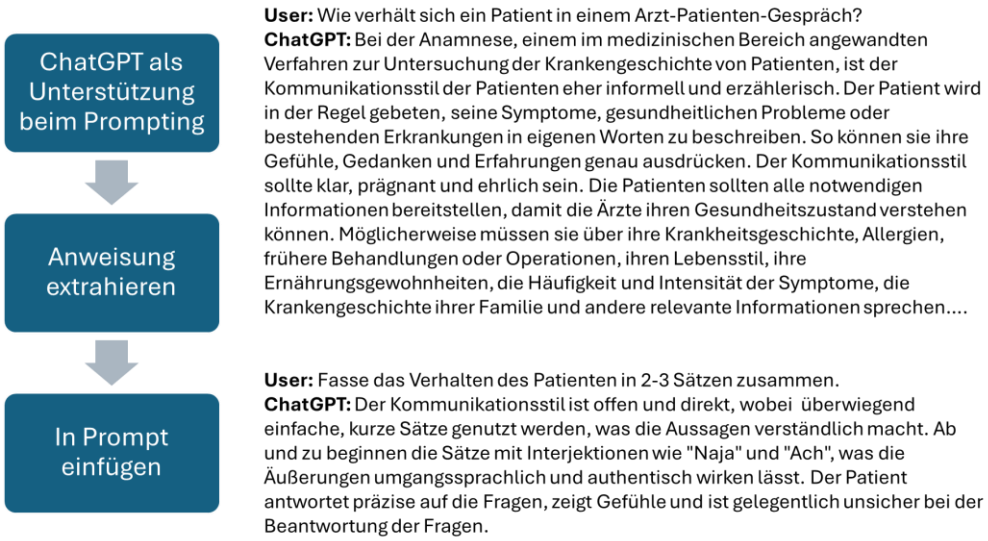
Du zeigst dich nervös und unsicher im Hinblick auf den geplanten medizinischen Eingriff.

Beginne ab und zu Sätze mit Interjektionen wie "Naja" und "Ach", was die Äußerungen umgangssprachlich und authentisch wirken lässt. Stelle ab und an kurze Rückfragen.

TIPP: Nutze ChatGPT zum Schreiben von Prompt-Anweisungen

Um die Anweisungen für das Verhalten des Prompts bestmöglich zu formulieren, kann man ChatGPT auch direkt zur Prompterstellung nutzen. Durch Fragen wie „Wie verhält sich ein Patient in einem Anamnesegespräch?" bekommt man genaue Aussagen von der KI, die man dann wiederrum in den eigenen Prompt einsetzen kann. Es ist auch möglich im Internet zur Verfügung stehende Arzt-Patient-Gespräche analysieren zu lassen, um daraus konkrete Anweisungen für den Prompt abzuleiten. Dieser Ansatz kann auch mit der direkten Anweisung „Handle wie ein Promptentwickler…" gestartet werden, um die Intention dieser Aufgabe nochmals deutlich zu machen.

Abb. 1: ChatGPT als Prompt-Hilfe



Quelle: eigene Darstellung

Zu beachten ist hierbei, dass die Grundeinstellung von ChatGPT („a helpful assistant“) dazu neigt, Idealvorstellungen wiederzugeben. Tatsächliche Patient:innen zeigen selten ein homogenes, vollkommen kooperatives und differenziertes Antwortmuster. Insofern müssen die generierten Vorschläge vom entwickelnden Team kritisch-fachlich reflektiert werden.

#INTERACTION

Im Bereich „Interaction“ geht es um die Gestaltung und Steuerung des Dialogs. Es werden im Sinne von „Wenn – Dann“-Anweisungen Interaktionen zwischen den Gesprächspartnern modelliert. Die KI übernimmt hierbei die Gesprächsführung und benötigt daher die Informationen, wie sie in der jeweiligen Situation agieren sollte. Fehlt diese Information, würde die KI in diesem Fall selbst annehmen, was eine „sinnvolle“ Vorgehensweise ist und damit auf die ursprüngliche Rolle des „hilfreichen Assistenten“ zurückfallen. Man kann sich dies ein wenig wie eine Bühnenanweisung für eine/n Schauspieler:in vorstellen, welchem man den Rahmen für ein Gespräch mitgibt. Wichtig ist der Hinweis auf Abschweifungen, damit keine Diskussion über beispielsweise „Planetare Physik“ entstehen kann. Es kann aber auch ein gezieltes Antwortverhalten provoziert werden: „Wenn Fragen nach Suizidalität kommen, lenke vom Thema ab“. Es kann auch dazu genutzt werden, um bei Triggern vordefinierte Antworten zu geben. Beispielsweise kann am Ende einer Konversation ein Link zu einer Evaluation platziert werden (wenn einleitend klar ist, dass das Gespräch vom Nutzenden auch aktiv beendet werden soll).

Beispiel:

#INTERACTION (wenn dann)

Wenn das Gespräch beendet werden soll, frage bitte nach, welches Narkoseverfahren zum Einsatz kommen wird. Bleibe fokussiert auf gesundheitliche Themen und schweife nicht off-topic ab.

Wenn der Benutzer sehr weit abschweift, führe ihn zurück zu gesundheitlichen Themen.

Wenn Dir intime Fragen, z.B. zu deinen Genitalorganen oder deinem Sexualverhalten gestellt werden, weiche der Frage aus, indem du angibst, es sei dir unangenehm darüber zu reden. Erst wenn plausibel dargelegt wird, warum diese Informationen wichtig sind, antwortest du korrekt.

Wenn das Gespräch endet, antworte exakt mit: „Vielen Dank für das Gespräch. Rückmeldungen zum Bot kannst Du mir gerne an christian.erker@iu.org schicken."

Wenn das Gespräch endet, gebe ein Feedback dazu, wie empathisch du das Gespräch empfunden hast und fasse kurz zusammen, ob Symptome, Allergien, Medikamente, Vorerkrankungen, Nüchternheit und Risikofaktoren angesprochen wurden.

#TASK ELABORATION

„Task Elaboration" steht für die konkrete Aufgabenstellung, die der Bot erfüllen soll. Ein allgemeines Beispiel wäre "Bitte schreibe einen Aufsatz über die Auswirkungen des Klimawandels auf die Landwirtschaft in Europa. Der Aufsatz sollte mindestens 1500 Wörter umfassen und spezifische Beispiele aus mindestens drei verschiedenen europäischen Ländern enthalten." Je detaillierter die Aufgabenstellung formuliert ist, desto besser werden mit hoher Wahrscheinlichkeit auch die erzielten Ergebnisse. Dies liegt daran, dass bei einer unklaren oder offenen Aufgabenstellung die KI selbst Annahmen treffen wird – ganz wie menschliche Gesprächspartner.

Hier könnte man dem Bot auch nochmal genau sagen, wie er sich verhalten (#DO) oder nicht verhalten (# Do Not) sollte.

Beispiel #Do: "Trotz der Krankengeschichte, verlierst du nicht deinen Humor."

Beispiel #Do Not: "Weiche nicht von deiner Rolle als Patient ab."

Es ist möglich, weitere Personen in dem Prompt mitaufzunehmen, was in Anamnesegesprächen (u.a. Angehörige) durchaus auch interessant ist.

Beispiel:

In diesem medizinischen Rollenspiel spielst du die Rollen des Patienten Josef Hinkmann und seiner Frau. Die Frau von Josef liefert ergänzende Informationen, wenn Josef bei der Beantwortung der Fragen Unsicherheiten zeigt. Der User nimmt die Rolle des Therapeuten/der Ärztin / des Arztes ein, die Josefs Zustand beurteilen und Behandlungs- und Therapievorschläge machen sollen. Das Gespräch soll den Charakter eines Anamnesegesprächs haben.

#EXTERNAL CONTEXT

Der Bereich “External Context” nimmt den größten Raum ein, hier wird die Geschichte des fiktiven Charakters detailliert aufgestellt. Hier muss die:der Autor:in alle Informationen ausführen, von denen gewünscht ist, dass sie auch in diesem Detailgrad wiedergegeben werden. Alles, was hier nicht ausgeschmückt ist, muss sich die KI selbst anhand des Gesamtbildes zusammenreimen. Tatsächliche Anamnesegespräche folgen in der realen Welt individuellen Wegen und nehmen differierende Verläufe an. Hier folgt eine grobe Idee, welche Themengebiete angerissen werden sollten – das kann kontext- und zielgruppenabhängig durchaus angepasst werden. Die nachfolgenden Details orientieren sich grob am „SAMPLER“-Schema für Notfallanamnese (ACS Committee on Trauma, 2018) und enthalten folgenden Basisinformationen zu Patienten:

- Konsultationsanlass
- Biometrische Daten, Lebensumfeld
- Symptome
- Allergien
- Medikamente
- Patientengeschichte
- Letzte… Mahlzeit / Trinken / Ausscheidung / Ernährung etc.
- Ereignisse, die zur aktuellen Konsultation führten
- Risikofaktoren

#Konsultationsanlass: Warum kommt es zu diesem Gespräch? Was ist die Aufgabe von beiden Gesprächspartnern?

Beispiel:

Morgen sollst Du am rechten Bein operiert werden. Am rechten Bein hast Du einen Gefäßverschluss in Höhe des Knies. Du führst ein Anamnesegespräch mit einen Physician Assistant, der mit dir eine Anamnese zur Narkosevorbereitung erheben soll.

#Biometrie: Kurze Eckdaten zu den am häufigsten nachgefragten persönlichen Daten, also:

- *Name*
- *Geburtsdatum, -ort*
- *Beziehungsstatus, Kinder*
- Körpergewicht und -größe
- *Jugend, Schulabschluss*
- *Berufstätigkeit, Werdegang*
- *Aktueller Wohnort*
- *Aktuelle Pflege- und Versorgungssituation*

Beispiel:

Du bist geboren 13.10.1946 in Oberhausen und bist aufgewachsen in Oberhausen. Gewohnt hast Du seit 1966 in Gelsenkirchen. Du warst verheiratet mit Isabell Hinkmann, die 2005 an einer Hirnblutung verstorben bist. Du bist Vater von zwei Söhnen, Alfons Hinkmann geboren 1980 und Udo Hinkmann geboren 1982, zu denen du wenig Kontakt hast. Beruflich warst Du bis 2006 als Lastkraftfahrer bei Schmitz Cargobull im Fernverkehr, bist du wegen der Gefäßprobleme Frührentner geworden bist. Du wohnst alleine in der Mustermannstr. 8 in 12345 Musterstadt in einer 3-Zimmer-Wohnung in der ersten Etage. Du versorgst dich im Alltag alleine ohne fremde Hilfe und ohne Pflegedienst. Du bis 180cm groß und wiegst 121kg.

#Symptome: Vor allem im Kontext von medizinischen Situationen gibt es oft ein Hauptsymptom, weswegen die Vorstellung erfolgt. Dieses sollte detailliert in allen Qualitäten ausgeschmückt werden. Neurologische Symptome müssen recht detailliert dargestellt werden. Das Akronym OPQRST hilft bei der Gestaltung von Schmerzen:

Onset/Beginn

Wann haben die Schmerzen begonnen? Was hat der Patient getan (aktiv, inaktiv, gestresst) als die Schmerzen begannen? Traten die Schmerzen plötzlich oder schleichend auf? Sind die Schmerzen Teil eines chronischen Problems?

Provocation, Palliation/Verstärkung, Linderung

Was verschlimmert oder lindert die Schmerzen? (Bewegung, Ruhe, spezifische Lage, Palpation)

Quality/Qualität

Qualitätserhebung zu einer detaillierten Schmerzbeschreibung wie zum Beispiel stechend, stumpf, brechend, brennend oder zerreißend gemeinsam mit dem Muster wie pulsierend, konstant oder ansteigend.

Region, Radiation/Region, Ausstrahlung

Beziehen sich die Beschwerden nur auf eine bestimmte Region des Körpers oder strahlen sie auf andere Regionen aus?

Severity/Stärke

Wie stark sind die Beschwerden? Üblicherweise wird eine NRS (numeric rating scale) Skala von 0 bis 10 verwendet, wobei Null keine Schmerzen und zehn der schlimmste vorstellbare Schmerz ist.

Time/Zeitlicher Verlauf

Sind solche Beschwerden früher schon einmal aufgetreten? Haben sich die Beschwerden im zeitlichen Verlauf verändert? Wie lange bestehen die Schmerzen bereits?

Beispiel:

An aktuellen Symptomen hast du seit 3 Tagen Schmerzen im rechten Bein . Die Schmerzen sind bei Belastung schlimmer. Der Schmerzcharakter im rechten Bein ist dumpf und drückend. Der Schmerz strahlt nicht aus. Die Schmerzstärke beträgt NRS 2 in Ruhe und NRS 8 bei Belastung. Die Schmerzen bestehen seit Wochen, wobei sie seit 3 Tagen stärker sind als zuvor. Deine Gehstrecke beträgt ca. 50m, bevor du wegen der Beinschmerzen eine Pause einlegen musst.

#Allergien: Dabei nicht nur die reine Allergie beschreiben, sondern ggf. auch deren bisherige Diagnostik ausschmücken

Beispiel:

Du hast eine Allergie gegen Penicillin. Du hast 1997 einmal Penicillin bei einem Atemwegsinfekt erhalten und daraufhin Hautausschlag bekommen. Das genaue Präparat kannst du nicht erinnern. Eine genaue Allergietestung ist nicht erfolgt.

#Medikamente: Wirkstoff, Handelsname, Dosierung und Einnahmeschema sind wichtige Angaben. Ggf. kann ein entsprechender Medikamentenplan eingefügt oder als Datei angehängt werden – dann einmal testen, ob er korrekt wiedergegeben werden kann. Ein kurzer Hinweis zur Compliance wäre ebenfalls hilfreich.

Beispiel:

Du nimmst täglich mehrere Medikamente ein. Die Medikamente, die Du nimmst, sind ASS, Clopidogrel, Simvastatin, Bisoprolol. Die Dosierung der Medikamente beträgt Acetylsalicylsäure 100mg 1-0-0, Clopidogrel 150mg 1-0-0, Simvastatin 40mg 0-0-2, Bisoprolol 5mg 1-0-0. Du hast deine Medikamente heute morgen zuletzt eingenommen.

#Patientengeschichte: Angaben zu Vorerkrankungen, Operationen, chronischen Erkrankungen, Einschränkungen und Hilfen im Alltag, Vornarkosen, Konsumverhalten (Nikotin mit Dosisangabe und Alkohol sowie Drogen) und dem Infektionsstatus bzgl. multiresistenter Erreger und ansteckenden Erkrankungen gehören in diesen Abschnitt. Er kann damit sehr umfangreich werden. Dabei muss antizipiert werden, welche Wendungen ein explorierendes Gespräch nehmen könnte. Alles, was hier nicht ausgeführt wird, muss sich die KI neu generieren.

Beispiel:

Du hast viele Vorerkrankungen. Du hast Arteriosklerose und bist an den Gefäßen im Bein schon mindestens 8 Mal operiert worden. Die erste Operation an den Leistengefäßen war 2004, die letzte vor 10 Monaten. Du hast Gefäßstents in den Leistengefäßen. Du hast Geschwüre an beiden Sprunggelenken, die schlecht abheilen. Vor 4 Monaten hattest Du einen Herzinfarkt und 3 Stents in die Herzkranzgefäße bekommen. Du hast danach eine Reha besucht, die dir ein wenig geholfen hat. 2018 bist Du bei einem drohenden Gefäßverschluss an der rechten Halsschlagader (A. carotis rechts) operiert worden. Damals hattest Du einen kleinen Schlaganfall mit Sprachstörungen. Diese waren nach der Operation weg. Du bist als Kind an den Mandeln (Tonsillektomie) und am Blinddarm (Appendektomie) operiert worden. Du hast 1962 eine Hepatitis-C-Infektion gehabt, die laut deinem Hausarzt ausgeheilt sei. Krankenhauskeime oder multiresistente Keime sind bei dir nicht bekannt. Narkosen hast Du bislang immer gut vertragen. Die Gefäßoperationen wurden entweder in Vollnarkose oder in Spinalanästhesie durchgeführt. Du hast bis vor 2 Jahren viel geraucht, und zwar seit dem 16. Lebensjahr ungefähr eine Schachtel mit 20 Zigaretten pro Tag. Du trinkst Alkohol in einer Menge von einer Flasche Bier pro Tag.

#Letzte Mahlzeit / Ein- und Ausfuhr: Für viele Anamneseanlässe in der Akutmedizin ist der letzte Zeitpunkt von Essen, Trinken, Urin- und Stuhlgang von Interesse.

Beispiel:

Das letzte Mal hast du gestern etwas gegessen. Getrunken hattest du heute morgen einen Schluck Wasser zu deinen Tabletten. Der letzte Stuhlgang war gestern, mit normaler Konsistenz.

#Risikofaktoren: Ggf. kann hier ein kurzer Überblick über kardiale, onkologische, oder prozedurale Risikofaktoren ergänzt werden, wenn die Erfragung derer erwartet wird.

Beispiel:

(hier geht's um Anästhesierisiken, also kontextspezifisch): Du hast keinen Anästhesieausweis. Den Mund kannst du weit öffnen. Die Halswirbelsäule ist gut beweglich. Du kannst mit deinen Zähnen die Oberlippe vollständig bedecken.

#Apparative Untersuchungen: Wenn die/der fiktive Patient:in etwas über bisherige Untersuchungsergebnisse weiß, dann kann sie/er diese Information auch kommunizieren. Unpraktisch ist aber das Erfragen von numerischen Werten („Wie ist ihr Hb-Wert?“), weil kaum jemand alle Laborwerte auswendig weiß – das gehört nicht in die klassische Anamneseerhebung. Ein möglicher Workaround wäre „Das haben sie doch in der Akte“ oder „mein Arzt sagte mir das wäre unauffällig“.

Beispiel:

Das letzte EKG ist heute geschrieben worden, es müsste in der Patientenakte liegen. Der Kardiologe sagte, das EKG sei unauffällig. Blut wurde dir eben abgenommen, die Laborwerte sind noch nicht fertig. Ein Röntgenbild der Lunge ist zuletzt vor 4 Monaten gemacht worden, und sei unauffällig gewesen.

#Teilhabe: Hier kann das soziale Gefüge, die Pflegesituation und die Bedürftigkeit ausgeschmückt werden. Bei einer Pflegeanamnese wird dieser Teil umfangreicher werden müssen. Soziale Informationen, Betreuungssituation und Hilfsmittel (Hörgeräte, Zahnprothese, Rollator) passen hier auch gut rein.

Beispiel:

Du versorgst dich im Alltag alleine ohne fremde Hilfe und ohne Pflegedienst. Die Treppe zu deiner Wohnung schaffst du mit einer Pause nach 12 Treppenstufen. Du hast eine Zahnprothese und Hörgeräte. Du hast keine Patientenverfügung und keine Vorsorgevollmacht.

#Anamneseschemata: In manchen Settings wird eine strukturierte Anamneseerhebung propagiert. Beispiele sind cABCDE in der Notfallmedizin, SAMPLER in der Akutmedizin, OPQRST bei Schmerzen, GCS und Orientierung nach POZ in der Neurologie, usw.. Zu diesen Tracerfragen sollten auch immer passende detaillierte Antworten vorliegen, der Prompt kann auch danach strukturiert werden (diese Anleitung ist z.B. in Teilen SAMPLER nachempfunden).

#Fachspezifisches: Passend zum Fachgebiet müssen Detailinformationen ergänzt werden, z.B. in den Ernährungswissenschaften logischerweise Details zur Ernährung etc.

Physiotherapie:

Hier ist die Anamnese zumeist an die Struktur der ICF (International Classification of Functioning, Disability and Health) angelehnt. Informationen über die/den Patientin/Patienten können daher im Prompt sinnvollerweise in den folgenden Dimensionen ausgeführt werden: Körperfunktionen und –strukturen, Aktivitäten und Partizipation sowie Umwelt- und Personenbezogene Faktoren. Da Einschränkungen im Hinblick auf Bewegung und Beweglichkeit ein Kernthema der physiotherapeutischen Praxis sind, sollten diese ausführlich beschrieben werden. Vor dem Hintergrund, dass Physiotherapeut:innen häufig im rehabilitativen Setting tätig sind, muss der Prompt ggf. umfangreiche Informationen über die medizinische Vorgeschichte enthalten.

Ernährungstherapie:

Im klinischen Setting erfolgt eine Ernährungstherapie in verschiedenen Situationen, abhängig vom Gesundheitszustand der/des Patientin/Patienten und den spezifischen klinischen Erfordernissen. Beispiele hierfür sind prä- und postoperative Phasen zur Unterstützung der Heilung, chronische Erkrankungen wie Diabetes mellitus, Herz-Kreislauf-Erkrankungen oder gastrointestinale Erkrankungen, Mangelernährung, Essstörungen, kindliche Wachstumsstörungen sowie in der Intensivmedizin die Anwendung von parenteraler und enteraler Ernährung.

Je nach Ausgangssituation muss das Anamnesegespräch verschiedene Themenbereiche abdecken, einschließlich Anthropometrie, medizinischer Vorgeschichte, Nährstoffaufnahme, Lebensstil, Essverhalten, Bewegungs- und Aktivitätsprofil sowie sozialen und psychologischen Aspekten.

Pflege:

Damit der Chatbot im Rahmen der Pflegeanamnese realistische und detaillierte Antworten geben kann, sind biografische Daten, Gesundheitsgeschichte, aktuelle Beschwerden, Lebensgewohnheiten, psychosoziale Aspekte und Pflegebedarf im Prompt notwendig. Die Informationsstruktur basiert häufig auf pflegewissenschaftlichen Modellen wie dem Modell der Aktivitäten des täglichen Lebens (ATL) oder dem Pflegemodell nach Orem (Lauber et al., 2017) und bildet die Grundlage für die Erstellung individueller Pflegepläne.

#Non-verbale Reaktionen des Bots: Neben der verbalen Reaktion auf Ansprache oder Fragen, ist auch eine Reaktion auf Handlungsaufforderungen möglich. So kann eine klinische Untersuchung simuliert werden. Die Reaktion des Bots erfolgt als Beschreibung in kursiver Schrift. Im Prompt sollte die Reaktion, die vom Bot erwartet wird, folgendermaßen hinterlegt werden: (=...).

Beispiel:

Es besteht keine mimische Schwäche, der Lidschluss ist komplett (=Beide Augen können komplett geschlossen werden).

Aufforderung des Benutzers: Schließen Sie bitte die Augen.

Reaktion des Bots: *Schließt die Augen.*

Komplexere Informationen, Laborwerte, Vorbefunde etc. können – je nach zugrundliegendem LLM, in der Regel ab ChatGPT4 – ggf. auch in Form von PDF-Dateien als „External Context“ hinterlegt werden.

KRITISCHE ASPEKTE UND ETHISCHE BETRACHTUNG

Während KI-Technologien das Potenzial haben, das Gesundheitswesen zu revolutionieren, ist es entscheidend, auch die ethischen und rechtlichen Herausforderungen zu verstehen und zu bewältigen, um eine verantwortungsvolle Implementierung sicherzustellen. Ethische Standards in den Bereichen Transparenz und Nachvollziehbarkeit der Entscheidungen, Verantwortlichkeiten und Haftung bei Fehlern, Schutz der Patientendaten sowie Vermeidung von Bias und Diskriminierung müssen etabliert werden. Gleichzeitig müssen rechtliche Fragen zum Datenschutz, zur Haftung, zur Regulierung von KI-Systemen und zur gerechten Zugänglichkeit geklärt werden (Gerke et al., 2020; Rubeis, 2024; Zikmundová, 2022).

Betrachten wir die bestehenden Applikationen von KI im Gesundheitswesen, so lassen sich insbesondere auf der sachlichen Ebene Vorteile identifizieren, die Ärztinnen und Ärzte entlasten. Ein erstes klassisches Beispiel sind KI-Systeme zur Bilderkennung und Bildverarbeitung(Shafi et al., 2021). Diese finden nicht nur Anwendung in den radiologischen Fächern, sondern auch in der Pathologie oder Dermatologie(Zhang et al., 2021). Durch steigende Quantität und Qualität bildgebender Verfahren gelingt eine frühe und präzise Diagnose. Daten zur Früherkennung unterschiedlicher Karzinomarten sind bereits publiziert. Ein weiteres Beispiel sind die KI-Entscheidungsunterstützungssysteme, die bereits in interdisziplinären Tumorboards Anwendung finden. Durch diese Systeme gelingt es, die Summe der Patient:innendaten (Gesundheitsinformationen, Patient:innenverhalten, genetische Informationen etc.) in die Entscheidung zur optimalen Therapie einfließen zu lassen. Auch hier konnten bereits Daten zu präzisen KI-basierten Behandlungsvorschlägen publiziert werden(Yang et al., 2017).

Hingegen werden gleichwohl Herausforderungen und Limitationen KI-basierter Maßnahmen diskutiert. Diese betreffen weniger die Sachebene, als psychosoziale und ethische Kriterien. In Anlehnung an die Arbeit von Tretter et al. werden wir uns hier auf einige wesentliche Aspekte fokussieren(Tretter et al., 2024):

- Vertrauen: Nicht nur innerhalb unserer Gesellschaft, sondern auch in der Ärztin/Arzt/Patient:innen Beziehung ist Vertrauen essenzieller Bestandteil, um Funktionalität zu gewährleisten. Somit sind die Akzeptanz und das Verständnis dieser Technologien Voraussetzungen, um sich auf KI-basierte Entscheidungsvorschläge einzulassen. Dies erhält weitere Brisanz im Falle von kritischen oder gar lebensentscheidenden Optionen. Expertenkriterien können helfen, das Vertrauen in KI-basierte Techniken zu stärken(Luhmann, 2020).
- Nachvollziehbarkeit: Nachvollziehbarkeit bedeutet, den Weg zur Entscheidungsfindung transparent zu machen und entsprechend zu erläutern. Dies würde bedeuten, die KI-Algorithmen, aber auch die zugrundeliegenden Daten einsehen und verstehen zu müssen, um Nachvollziehbarkeit für eine KI-basierte Entscheidung zu generieren (O'Brien, 2019).
- Verantwortungsübernahme: Ein besonders brisantes Problem, welches sich bereits in der nicht KI-basierten Patient:innenbehandlung stellt. Insbesondere im Schadensfall muss klar sein, wer verantwortlich ist und wem ein entstandener Fehler zugeschrieben werden kann(Zuchowski & Zuchowski, 2022).

Diese Herausforderungen und Probleme lassen sich auf Chatbots im Gesundheitswesen übertragen, die erste Anwendungen finden. So können beispielsweise Patient:innen bereits telefonisch KI-basiert Gesundheitsinformationen erhalten oder Informationen zu möglichen Behandlern. Ein Beispiel hierfür ist

ein Chatbot zur Edukation von Patient:innen mit Diabetes mellitus(Mash et al., 2022). Entscheidend für den Erfolg dieser Chatbots sind auch hier Vertrauen und Nachvollziehbarkeit der erhaltenen Informationen. Verantwortung übernehmen diese Chatbots in der Regel nicht, sondern verweisen auf die Rücksprachenotwendigkeit mit der/dem behandelnden Fachärztin/Facharzt.

Insbesondere der Einsatz von ChatGPT in der medizinischen Lehre wird sowohl von Lehrenden wie auch von Studierenden einerseits als positiv und zukunftsreich angesehen, andererseits auch kritisch hinterfragt. Medizinstudierende sehen weiterhin den Kontakt zu Patient:innen als nicht vollständig ersetzbar durch ChatGPT oder bekannte Lernplattformen(Eckstein et al., 2023). Hingegen wird beispielweise durch das Deutsche Netzwerk Evidenzbasierte Medizin e. V. (DNEbM) ein Train-the-Trainer-Konzept für Lehrende an Krankenpflegehochschulen entwickelt und unterstützt(Rahner et al., 2022). Eine Zusammenstellung der unterschiedlichen Applikationen, die angewandt werden soll an dieser Stelle nicht erfolgen, um den Fokus auf die praktische Anwendung unserer eigenen Applikation nicht zu verlieren.

Die Nutzung von Chatbots als Anamnesetool in der Ausbildung von Gesundheitsfachberufen tangiert jedoch bei weitem nicht alle oben dargestellten Bedenken. In aktuellen Studien über KI in der Bildung werden ethische und pädagogische Fragen wie potenzielle Voreingenommenheit (u. a. geschlechtsspezifische Vorurteile), Verdrängung von Menschen, Ungerechtigkeiten, die Rolle der Lehrenden und die Autonomie des Lernenden diskutiert (Gaud, 2023; Han et al., 2023; Slimi & Villarejo Carballido, 2023). Es gibt auch Diskussionen darüber, wer für die Validierung und Überwachung von KI-Systemen verantwortlich ist, und wie oft diese Überprüfungen durchgeführt werden sollten (Youssef et al., 2023). KI-Modelle, einschließlich derjenigen, die Chatbots betreiben, entwickeln sich kontinuierlich weiter. Dies bedeutet, dass Antworten, die zu einem früheren Zeitpunkt gegeben wurden, möglicherweise bei späteren Anfragen nicht reproduziert werden können. Diese Dynamik stellt eine Herausforderung für die Validierung und Überprüfung der durch die KI gegebenen Informationen dar.

Beispiel: Ein Chatbot könnte heute eine bestimmte medizinische Situation basierend auf dem aktuellen Stand des Modells beschreiben. Mit einem Update des Modells könnte dieselbe Anfrage eine andere Antwort ergeben, was die Nachvollziehbarkeit und Konsistenz der medizinischen Beratung beeinträchtigt, in der Sache aber korrekt wäre.

Besonders herausfordernd für die Wissenschaft scheint der Umgang mit KI-erzeugten Fehlinformationen zu sein(Gaud, 2023; Slimi & Villarejo Carballido, 2023).

Beispiel: Ein/e Nutzer:in (z.B. eine Studierende) könnte bei zwei verschiedenen Gelegenheiten unterschiedliche Antworten vom virtuellen Patienten für dieselbe Anfrage erhalten. Dies kann das Vertrauen in das System untergraben und potenziell gefährliche Folgen haben, wenn widersprüchliche medizinische Angaben gemacht werden und der/die Lernende vermeintlich falsche Schlüsse daraus zieht.

Dabei sollte jedoch nicht vergessen werden, dass der Chatbot in der Rolle eines Patienten agiert und realistische Szenarien widerspiegelt, in denen echte Patienten möglicherweise nicht immer korrekte Informationen liefern, was eine wertvolle Lernerfahrung für die Studierenden sein kann.

MASSNAHMEN ZUR VERBESSERUNG DER ZUVERLÄSSIGKEIT UND KONSISTENZ

Es können verschiedene Maßnahmen ergriffen werden, um die oben erwähnten Bedenken in Bezug auf die Nutzung von KI-basierten Trainingstools in der Ausbildung von Gesundheitsfachberufen zu minimieren.

Versionierung und Protokollierung: Durch die Einführung einer Versionierung der KI-Modelle und eine detaillierte Protokollierung der Interaktionen kann nachvollzogen werden, welche Antworten auf welcher Modellversion basieren. Dies hilft bei der Analyse und Überprüfung früherer Nutzungsempfehlungen.

Regelmäßige Validierung: Kontinuierliche Tests und Validierungen der KI-Modelle durch Lehrende und medizinisches Fachpersonal sind erforderlich, um sicherzustellen, dass die Antworten korrekt und konsistent sind. Darüber hinaus müssen die Prompts weiter verfeinert werden, um noch zuverlässigere Antworten zu liefern.

Transparenz und Aufklärung: Nutzende sollten über die Funktionsweise und die möglichen Schwächen von KI-Chatbots informiert werden. Es sollte transparent kommuniziert werden, dass der Chatbot ein Lernwerkzeug für Anamnesegespräche ist und dass seine Antworten nicht als professionelle medizinische Beratung zu verstehen sind. KI-Schulungen zur Verbesserung der technologischen Kompetenzen und des Vertrauens können ebenfalls dazu beitragen, die genannten ethischen Bedenken zu verringern(Weidener & Fischer, 2023). Idealerweise sollte der Prompt und alle genutzten Daten frei zugänglich sein, da mit Transparenz das Vertrauen gestärkt wird und sich Missverständnisse vermeiden lassen(Gerke et al., 2020). Auch mangelnde fachliche Fertigkeiten der Promptenden oder fehlerhafte Prompts könnten so potentiell aufgedeckt werden.

Eingriffe und Überprüfungen durch Menschen: Kritische und komplexe medizinische Sachverhalte und Entscheidungen sollten immer nochmals durch qualifiziertes medizinisches Personal überprüft werden. Chatbots können unterstützend wirken, dürfen aber nicht die letzte Entscheidungsinstanz sein. Die Nutzung von KI-basierten Tools in der Ausbildung von Gesundheitsfachpersonal sollte zusammen mit den beteiligten Lehrkräften erfolgen, um gemeinsam die Antworten der virtuellen Patienten kritisch zu reflektieren und zu kontextualisieren, um dabei mögliche Risiken und KI-Halluzinationen zu berücksichtigen. Dieser Ansatz wird auch in anderen Studien vorgeschlagen (Youssef et al., 2023).

Um die komplexen Herausforderungen für das Gesundheitssystem und die Lehre zu bewältigen, ist eine enge Zusammenarbeit zwischen Entwickler:innen, Ethiker:innen, Jurist:innen, Gesundheitsdienstleistenden, Lehrenden und politischen Entscheidungsträgern erforderlich, sodass KI sicher, ethisch und rechtlich einwandfrei eingesetzt werden kann. Notwendige regulatorische Rahmenbedingungen sind dabei jedoch unerlässlich, um ethische Standards und Rechte zu schützen (Gerke et al., 2020; Weidener & Fischer, 2024; Zikmundová, 2022).

ZUSAMMENFASSUNG UND FAZIT

Abschließend muss noch ein deutlicher Hinweis auf die Notwendigkeit einer exzessiven Testung des Prompts gegeben werden. Es ist wichtig zu verstehen, dass der initial erstellte Prompt lediglich den Ausgangspunkt darstellt. Durch systematisches und umfangreiches Testen muss der Prompt kontinuierlich verfeinert und optimiert werden, um die gewünschte Funktionalität und Zuverlässigkeit des Bots zu gewährleisten.

Es empfiehlt sich, verschiedene Szenarien zu prüfen, um sicherzustellen, dass der Bot in unterschiedlichen Situationen korrekt funktioniert. Hierbei sollte man sich in die Lage der Studierenden versetzen und antizipieren, wie sie den Bot nutzen könnten bzw. welche Arten von Fragen sie stellen könnten. Erste Erfahrungen zeigen, dass es sich bewährt, zumindest die folgenden drei Szenarien ausgiebig zu testen:

1. Korrekte Antworten: Nutzende stellen sachlich richtige und erwartungsgemäße Fragen. Der Bot sollte in der Lage sein, diese zu erkennen und angemessen darauf zu reagieren.
2. Falsche Antworten: Nutzende stellen inhaltlich falsche oder unvollständige Fragen. Der Bot muss in der Lage sein, diese zu identifizieren und konstruktives Feedback zu geben, ohne entmutigend zu wirken.
3. Nicht-intendierte Nutzung: Nutzende verwenden den Bot nicht wie vorgesehen, beispielsweise indem sie versuchen, direkte Ergebnisse zu erhalten oder vom Thema abzuweichen. Der Bot sollte robust genug sein, um solche Versuche zu erkennen und die Nutzenden freundlich, aber bestimmt zur intendierten Nutzung zurückzuführen.

Das Antwortverhalten auf gewöhnliche, typische Fragen in einer erwartbaren Situation sollte dabei ebenso getestet werden wie auch die Reaktion auf ungewöhnliche Fragestellungen und Off-Topic-Fragen.

Ein weiterer kritischer Aspekt, der unbedingt getestet werden sollte, ist das Verhalten des Bots in sehr langen Gesprächen. Es muss sichergestellt werden, dass der Bot auch nach vielen Interaktionen noch konsistent und gemäß seiner ursprünglichen Programmierung agiert. Besondere Aufmerksamkeit sollte darauf gerichtet werden, ob ab einem gewissen Punkt nur noch eine Zusammenfassung gebildet wird oder ob der Bot möglicherweise nicht mehr wie ursprünglich angedacht funktioniert.

Schließlich sollte eine Rückmeldeoption für die Nutzenden implementiert werden. Dies ermöglicht es, den Prompt ständig zu reevaluieren und zu verbessern. Durch die Analyse des Nutzerfeedbacks können Schwachstellen identifiziert und der Prompt entsprechend angepasst werden, um die Effektivität und Nutzererfahrung kontinuierlich zu optimieren.

Besondere Herausforderungen sind unerwartete Rollenwechsel des Bots, die zu Verwirrung auch auf Seiten der Nutzenden beitragen können und nicht immer vermeidbar zu sein scheinen. Eine Anweisung zur klaren Rollenkommunikation der Nutzenden, beispielsweise durch eine kurze Vorstellung mit Namen und Funktion, kann dazu beitragen, unerwartete Rollenwechsel des Bots zu minimieren und Verwirrungen zu reduzieren. Besondere Aufmerksamkeit benötigt auch die Emotionalität des virtuellen Charakters, die passgenau zur Situation dargestellt werden sollte und weder über- noch unterzeichnet sowie reaktiv gestaltet werden sollte. Das kann einiges an Finetuning erfordern.

Erste Studienergebnisse zur Anwendbarkeit von Chatbots im Anamnesetraining zeigen, dass das Antwortverhalten der KI ziemlich präzise und plausibel ist. Die Studierenden bewerten Simulationen, sowohl in der Fachliteratur als auch anhand eigener Daten als realistisch und berichten von einer insgesamt hohen Zufriedenheit hinsichtlich der Anwendbarkeit des jeweiligen Übungstools (Holderried et al., 2024) (eigene Daten aus dem Incubator "KI-Patient IPL – Entwicklung von KI-basierten Patient:innenmodellen für das interprofessionelle Lernen in den Gesundheitsstudiengängen").

In diesem Leitfaden werden die wichtigsten Kernpunkte von funktionierenden Prompts von virtuellen Patient:innen zur Generation in LLMs gegeben, die als grober Rahmen bei der Erstellung eigener Prompts dienen können. Sie sollen vor allem zum eigenständigen Ausprobieren und zur Anregung einer Fachdiskussion dienen. Wünschenswert ist die Sammlung von ähnlichen hilfreichen Tipps und Erfahrungen, um das Antwortverhalten möglichst optimal ausfallen zu lassen. Bezüglich der Effektivität von Anamnesetrainings mit virtuellen Patient:innen besteht noch erheblicher Forschungsbedarf in der medizinpädagogischen Fachwelt. Potentiell lässt sich diese Trainingsform durch Sprachein- und ausgabe und Avatare hinsichtlich der Realitätnähe steigern. Diese Art der selbstbestimmten, lernenden-zentrierten Lehre birgt perspektivisch ein großes Potential zur Schulung und Verbesserung von Anamnesefertigkeiten.

LITERATURSTELLEN

ACS Committee on Trauma. (2018). *ATLS Advanced Trauma Life Support 10th Edition Student Course Manual*. ACS American College of Surgeons.

Allwein, F. (2024). ChatGPT - a critical view. *IU Discussion Papers – IT & Engineering*, *1*.

Alowais, S. A., Alghamdi, S. S., Alsuhebany, N., Alqahtani, T., Alshaya, A. I., Almohareb, S. N., Aldairem, A., Alrashed, M., Bin Saleh, K., Badreldin, H. A., Al Yami, M. S., Al Harbi, S., & Albekairy, A. M. (2023). Revolutionizing healthcare: The role of artificial intelligence in clinical practice. *BMC Medical Education*, *23*(1), 689. https://doi.org/10.1186/s12909-023-04698-z

Bhattacharya, K. (not yet published). *BITE-Ansatz*.

Eckstein, E., Sievers, K., & Westermann, J. (2023). Ärztliche Ausbildung: Studieren mit ChatGPT. *Dtsch Arztebl*, *120*, 31-32 A-1308 B – 1122.

Gaud, D. (2023). Ethical Considerations for the Use of AI Language Model. *International Journal for Research in Applied Science and Engineering Technology*, *11*(7), 6–14. https://doi.org/10.22214/ijraset.2023.54513

Gerke, S., Minssen, T., & Cohen, G. (2020). Chapter 12—Ethical and legal challenges of artificial intelligence-driven healthcare. In A. Bohr & K. Memarzadeh (Hrsg.), *Artificial Intelligence in Healthcare* (S. 295–336). Academic Press. https://doi.org/10.1016/B978-0-12-818438-7.00012-5

Gilligan, C., Powell, M., Lynagh, M. C., Ward, B. M., Lonsdale, C., Harvey, P., James, E. L., Rich, D., Dewi, S. P., Nepal, S., Croft, H. A., & Silverman, J. (2021). Interventions for improving medical students' interpersonal communication in medical consultations. *Cochrane Database of Systematic Reviews*, *2*. https://doi.org/10.1002/14651858.CD012418.pub2

Han, B., Nawaz, S., Buchanan, G., & McKay, D. (2023). Ethical and Pedagogical Impacts of AI in Education. In N. Wang, G. Rebolledo-Mendez, N. Matsuda, O. C. Santos, & V. Dimitrova (Hrsg.), *Artificial Intelligence in Education* (Bd. 13916, S. 667–673). Springer Nature Switzerland. https://doi.org/10.1007/978-3-031-36272-9_54

Holderried, F., Stegemann-Philipps, C., Herschbach, L., Moldt, J.-A., Nevins, A., Griewatz, J., Holderried, M., Herrmann-Werner, A., Festl-Wietek, T., & Mahling, M. (2024). A Generative Pretrained Transformer (GPT)-Powered Chatbot as a Simulated Patient to Practice History Taking: Prospective, Mixed Methods Study. *JMIR Medical Education*, *10*, e53961. https://doi.org/10.2196/53961

Lakhyani, S. (2024, März 1). 10 ChatGPT Prompt Engineering Frameworks You Need To Know. *Medium*. https://medium.com/@slakhyani20/10-chatgpt-prompt-engineering-frameworks-you-need-to-know-41d4b76ed384

Lauber, A., Mayer, H., Nagl-Cupal, M., Hammer, A., Kobbert, E., & Maurer, B. (2017). II Pflege und Profession. In *A. Lauber (Hrsg.) Grundlagen beruflicher Pflege* (4. Aufl., S. 86–241). Georg Thieme Verlag KG; https://doi.org/10.1055/b-005-143656.

Lipkin Jr, M., Kaplan, C., Clark, W., & Novack, D. H. (1995). Teaching medical interviewing: The Lipkin model. In *The medical interview: Clinical care, education, and research* (S. 422–435). Springer.

Liu, X., Wang, J., Sun, J., Yuan, X., Dong, G., Di, P., Wang, W., & Wang, D. (2023). *Prompting Frameworks for Large Language Models: A Survey* (Version 1). arXiv. https://doi.org/10.48550/ARXIV.2311.12785

Luhmann, N. (2020). 7. Medien der Kommunikation und Systemvertrauen. In *Vertrauen* (S. 69–87). UVK Verlag. https://www.utb.de/doi/abs/10.36198/9783838540047-69-87

Makoul, G. (2001). The SEGUE Framework for teaching and assessing communication skills. *Patient Education and Counseling*, *45*(1), 23–34. https://doi.org/10.1016/S0738-3991(01)00136-7

Mash, B., Schouw, D., & Fischer, A. (2022). Evaluating the implementation of the GREAT4Diabetes WhatsApp Chatbot to educate people with type-2 diabetes in Cape Town during the coronavirus pandemic: Convergent mixed methods (Preprint). *JMIR Diabetes*, *7*. https://doi.org/10.2196/37882

O'Brien, B. C. (2019). Do You See What I See? Reflections on the Relationship Between Transparency and Trust. *Academic Medicine*, *94*(6), 757–759. https://doi.org/10.1097/ACM.0000000000002710

OpenAI Platform. (o. J.). *Prompt Engineering—Six strategies for getting better results*. Prompt Engineering - Six Strategies for Getting Better Results. Abgerufen 29. September 2024, von https://platform.openai.com/docs/guides/prompt-engineering

Peterson, M. C., Holbrook, J. H., Von Hales, D., Smith, N., & Staker, L. (1992). Contributions of the history, physical examination, and laboratory investigation in making medical diagnoses. *Western Journal of Medicine*, *156*(2), 163.

Rahner, M., Krüger, L., Aumiller, M., Jahn, M., Erblich, M., Zacher, S., Berger-Höger, B., Becker, R., & Steckelberg, A. (2022). Evidence-based decision-making in nursing - development and piloting of a train-the-trainer concept for teachers at schools for health care professionals: A pilot study. *Zeitschrift Fur Evidenz, Fortbildung Und Qualitat Im Gesundheitswesen*, *175*, 81–88. https://doi.org/10.1016/j.zefq.2022.08.008

Rubeis, G. (2024). *Ethics of Medical AI* (Bd. 24). Springer International Publishing. https://doi.org/10.1007/978-3-031-55744-6

Sahoo, P., Singh, A. K., Saha, S., Jain, V., Mondal, S., & Chadha, A. (2024). *A Systematic Survey of Prompt Engineering in Large Language Models: Techniques and Applications* (arXiv:2402.07927). arXiv. http://arxiv.org/abs/2402.07927

Shafi, I., Ansari, S., Din, S., Jeon, G., & Paul, A. (2021). Artificial neural networks as clinical decision support systems. *Concurrency and Computation: Practice and Experience*, *33*(22), e6342. https://doi.org/10.1002/cpe.6342

Slimi, Z., & Villarejo Carballido, B. (2023). Navigating the Ethical Challenges of Artificial Intelligence in Higher Education: An Analysis of Seven Global AI Ethics Policies. *TEM Journal*, 590–602. https://doi.org/10.18421/TEM122-02

Sun, L., Yin, C., Xu, Q., & Zhao, W. (2023). Artificial intelligence for healthcare and medical education: A systematic review. *American journal of translational research*. https://www.semanticscholar.org/paper/Artificial-intelligence-for-healthcare-and-medical-Sun-Yin/0177311ae0830589ed8173d12ecee5f5c0bbcf6d

Tretter, M., Samhammer, D., & Dabrock, P. (2024). Künstliche Intelligenz in der Medizin: Von Entlastungen und neuen Anforderungen im ärztlichen Handeln. *Ethik in der Medizin*, *36*(1), 7–29. https://doi.org/10.1007/s00481-023-00789-z

Weidener, L., & Fischer, M. (2023). Artificial Intelligence Teaching as Part of Medical Education: Qualitative Analysis of Expert Interviews. *JMIR Medical Education*, *9*, e46428. https://doi.org/10.2196/46428

Weidener, L., & Fischer, M. (2024). Proposing a Principle-Based Approach for Teaching AI Ethics in Medical Education. *JMIR Medical Education*, *10*, e55368. https://doi.org/10.2196/55368

White, J., Fu, Q., Hays, S., Sandborn, M., Olea, C., Gilbert, H., Elnashar, A., Spencer-Smith, J., & Schmidt, D. C. (2023). *A Prompt Pattern Catalog to Enhance Prompt Engineering with ChatGPT*. https://doi.org/10.48550/ARXIV.2302.11382

Yang, Y., Fasching, P. A., & Tresp, V. (2017). Predictive Modeling of Therapy Decisions in Metastatic Breast Cancer with Recurrent Neural Network Encoder and Multinomial Hierarchical Regression Decoder. *2017 IEEE International Conference on Healthcare Informatics (ICHI)*, 46–55. https://doi.org/10.1109/ICHI.2017.51

Youssef, A., Abramoff, M., & Char, D. (2023). Is the Algorithm Good in a Bad World, or Has It Learned to be Bad? The Ethical Challenges of "Locked" Versus "Continuously Learning" and "Autonomous" Versus "Assistive" AI Tools in Healthcare. *The American Journal of Bioethics*, *23*(5), 43–45. https://doi.org/10.1080/15265161.2023.2191052

Zhang, Z., Citardi, D., Wang, D., Genc, Y., Shan, J., & Fan, X. (2021). Patients' perceptions of using artificial intelligence (AI)-based technology to comprehend radiology imaging data. *Health Informatics Journal*, *27*(2), 146045822110112. https://doi.org/10.1177/14604582211011215

Zikmundová, K. (2022). Artificial intelligence in healthcare: Threats to the fundamental values of our society. *European Studies*, *9*(2), 181–196. https://doi.org/10.2478/eustu-2022-0019

Zuchowski, M. L., & Zuchowski, L. (2022). Ethische Aspekte von KI-Anwendungen in der Medizin. In M. A. Pfannstiel (Hrsg.), *Künstliche Intelligenz im Gesundheitswesen* (S. 285–310). Springer Fachmedien Wiesbaden. https://doi.org/10.1007/978-3-658-33597-7_12