

# Generative KI im Gesundheitswesen – Chancen und Risiken

von Mirja Mittermaier

## Abstract

**Generative künstliche Intelligenz (KI) erzeugt neue Inhalte**, zum Beispiel Texte, Bilder, Filme oder Musik. Die generativen KI-Modelle werden mithilfe einer großen Menge von Daten trainiert. Nach Eingabe von Anweisungen durch den Nutzer werden neue Inhalte auf Basis der zuvor gelernten Muster und Beziehungen generiert. Für das Gesundheitswesen ist das Potenzial von generativer KI enorm. Es reicht von Abrechnungscodierung über die Erstellung von Arztbriefen bis hin zur Arzneimittelforschung. Trotz dieser Möglichkeiten steht die Nutzung von generativer KI im klinischen Alltag noch ganz am Anfang, da zahlreiche Risiken und Implementierungshürden bestehen. Dazu zählen eine ungeeignete IT-Infrastruktur, das Fehlen von klinischen Studien, die die Genauigkeit und Zuverlässigkeit der erzeugten Ergebnisse validieren, sowie offene Fragen zu Datenschutz, Haftung und Regulierung. Der nachfolgende Artikel beleuchtet sowohl die Potenziale als auch die Herausforderungen von generativer KI im Gesundheitswesen.

**Schlüsselwörter:** generative künstliche Intelligenz, generative KI im Gesundheitswesen, Large Language Models, Datensicherheit und Ethik beim Einsatz von generativer KI

**Generative artificial intelligence (AI) generates new content**, such as texts, images, films, or music. Generative AI models are trained on a large amount of training data. Upon receiving prompts (i.e. instructions) from the user, new content is generated based on the patterns and relationships learned from the training dataset. The potential of generative AI in healthcare is enormous and ranges from billing coding to the creation of medical letters and drug research. Despite these possibilities, the use of generative AI in clinical practice is still in its early stages, as numerous risks and implementation barriers exist. These include issues such as unsuitable IT infrastructure, the lack of clinical studies validating the accuracy and reliability of the generated results, and open questions about data protection, liability, and regulation. The following article explores both the potentials and challenges of generative AI in healthcare.

**Keywords:** generative artificial intelligence, generative AI in the healthcare system, large language models, data security and ethics in connection with generative AI

# 1 Einleitung

Im November 2022 hat die US-amerikanische Firma OpenAI der weltweiten Öffentlichkeit ChatGPT präsentiert und damit sowohl in der Fachwelt als auch in der breiten Öffentlichkeit große Aufmerksamkeit erregt. Chat steht dabei für „Unterhaltung“, GPT für *generative pre-trained transformer* (generativer, vortrainierter Transformator). ChatGPT ist ein sogenanntes großes Sprachmodell oder Large Language Model (LLM). Generative KI, zu der LLMs gehören, basiert auf den Prinzipien des maschinellen Lernens und wird mithilfe einer großen Anzahl an Trainingsdaten trainiert. Die im Training gelernten Muster und Zusammenhänge werden anschließend genutzt, um neue Daten, die auf den Eigenschaften der Trainingsdaten basieren, zu generieren. Dadurch ergeben sich bis dahin nicht mögliche Anwendungen. So kann generative KI aus Text Bilder oder sogar Videos generieren, virtuelle Assistenten schaffen, als Chatbot mit Menschen kommunizieren oder Deepfake-Videos erzeugen. Auch im Gesundheitswesen gibt es eine Vielzahl von potenziellen Anwendungsmöglichkeiten. Die Implementation in die klinische Versorgung steht jedoch noch ganz am Anfang und ist auf vielen Ebenen herausfordernd. In den nachfolgenden Abschnitten sollen sowohl die Möglichkeiten als auch die Risiken und Herausforderungen im Gesundheitswesen diskutiert werden.

## 1.1 Unterschied zu bisherigen Modellen

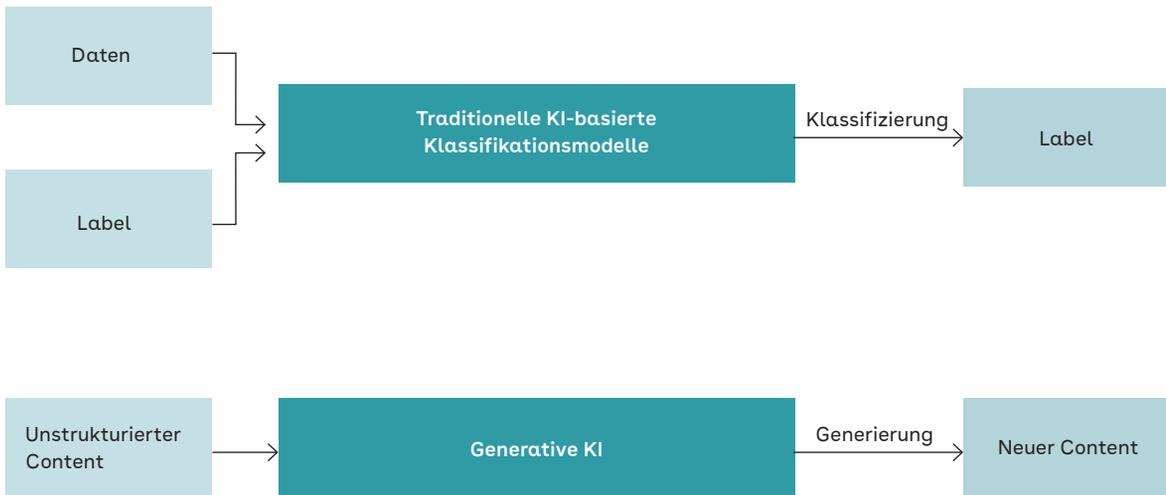
Der Hauptunterschied zwischen generativer künstlicher Intelligenz (KI) und bisherigen KI-Modellen liegt in ihrer Funktionsweise und Zielsetzung. Generative KI-Modelle sind darauf ausgelegt, neue Daten zu erzeugen, indem sie Muster und Strukturen, die sie im Trainingsdatensatz gelernt haben, für die Generierung neuer, ähnlicher Daten anwenden (Epstein et al. 2023; Takefuji 2023). Der Nutzer (*user*) gibt einen sogenannten *prompt* ein. Dies kann eine Frage, eine Anweisung oder im Fall von Bildgeneratoren die Beschreibung, wie ein zu generierendes Bild aussehen soll, sein. Anschließend erzeugt die generative KI die entsprechende Antwort. Bisherige KI-Modelle sind hingegen darauf ausgelegt, Muster in Daten zu erkennen und aufgrund dieser Muster Entscheidungen und Vorhersagen zu treffen (Abbildung 1). Dazu gehören Klassifikationsmodelle, die zum Beispiel benigne von malignen Hautveränderungen unterscheiden und

in Kategorien einteilen können (Esteva et al. 2017) und Prädiktionsmodelle, die zum Beispiel eine Sepsis frühzeitig vorhersagen können.

## 1.2 Modellarchitektur

Generative KI ist ein Oberbegriff für viele verschiedene KI-Modelle, die – abhängig von dem Ziel des Modells – unterschiedliche KI-Modell-Architekturen nutzen. Ein Beispiel für generative Algorithmen sind LLMs und große multimodale Modelle (*large multimodal models*, LMMs). Der Hauptunterschied zwischen LLMs und LMMs besteht in der Art der Daten, die sie verarbeiten und generieren können. LLMs wie OpenAIs GPT-4 oder Googles Bard sind hauptsächlich darauf trainiert, Muster in Texten zu erkennen, Fragen zu beantworten, Texte zusammenzufassen oder zu übersetzen – und das fast auf dem Niveau menschlicher Qualität. Im Unterschied dazu können LMMs gleichzeitig verschiedene Arten von Daten, wie Text, Bild, Audio und Video, verarbeiten. Dies befähigt LMMs, Aufgaben zu bewältigen, die ein Verständnis mehrerer Informationsarten erfordern, wie das Erstellen von Bildern aus Textbeschreibungen, das Verstehen von Videos anhand gesprochener oder textbasierter Prompts und ähnliche komplexe Aufgaben. Beispiele für Algorithmen, die Bilder aus Textbeschreibungen erzeugen, sind DALL-E von OpenAI oder Googles Imagen. CLIP (das Akronym steht für *contrastive language-image pre-training*) wurde von OpenAI entwickelt und mithilfe einer großen Anzahl an Bild-Text-Paaren trainiert. Es kann dadurch Bilder anhand von Textbeschreibungen erkennen und klassifizieren sowie die Bildunterschrift für ein bestimmtes Bild vorhersagen. LLMs und LMMs nutzen in der Regel eine Transformer-Architektur, die besonders gut Abhängigkeiten beziehungsweise Zusammenhänge in großen Datenmengen erfassen kann. Neben der Transformer-Architektur werden Generative Adversarial Networks (GANs), Variational Autoencoders (VAEs), Auto-Regressive-Modelle oder Diffusionsmodelle genutzt. Letztere sind sehr leistungsfähig in der Erzeugung von realistischen Bildern. GANs können ebenso neue Bilder und Videos erzeugen, die von echten nicht zu unterscheiden sind und zum Beispiel einen bestimmten Stil imitieren (Goodfellow et al. 2014). Generative KI-Modelle werden zudem als eine Untergruppe der sogenannten Foundation-Modelle (Basis-Modelle) angesehen, die sich dadurch auszeichnen, dass sie mit einer riesigen Datenmenge trainiert werden.

Abbildung 1 Traditionelle und generative KI



Quelle: theaidiscovery.com; Grafik: G+G Wissenschaft 2024

Herkömmliche KI-basierte Klassifikationsmodelle nutzen Daten und Label, um neue Daten zu klassifizieren. Generative KI hingegen erkennt auch Muster in unstrukturierten Daten und kann daraus neue Daten erzeugen.

### 1.3 Für Medizin optimierte generative KI-Modelle

Spezifisch für medizinische Fragestellungen wurde beispielsweise von Google das LLM MedPalm entwickelt, mit dem Ziel „qualitativ hochwertige Antworten auf medizinische Fragen“ zu geben ([sites.research.google/med-palm/](https://sites.research.google/med-palm/); Singhal et al. 2023). Zusätzlich entwickelte Google das LMM MedPalm M, welches multimodale Daten wie Bilder, Texte und Laborwerte verarbeiten kann, und Med-Gemini, welches unstrukturierte Daten aus umfangreichen Datensätzen wie Patientenakten besonders gut analysieren kann. Meditron wurde von Meta entwickelt und basiert auf dem Open-Source-Modell Llama 3. Es soll insbesondere in Ländern mit geringeren medizinischen Ressourcen eingesetzt werden (Meta 2024). Bei der Beantwortung medizinischer Fragen der United States Medical Licensing Examination (USMLE) erreicht MedPalm eine Genauigkeit von 67,2 Prozent, Med-PaLM 2 86,5 Prozent, GPT-4 90 Prozent, und MedGemini 91,1 Prozent – das verdeutlicht die rasante Verbesserung der Modelle innerhalb kurzer Zeit (Brin et al. 2023; Singhal et al. 2023). Meditron wird in einer Online-Validierung – genannt „Meditron MOOVE“ (*massive online open validation and evalua-*

*tion*) – weltweit von Ärzten getestet (Meta 2024). Trotz dieser Validierung ist kritisch zu betrachten, dass auch die für die Medizin entwickelten Modelle weder in großen klinischen Studien geprüft noch beispielsweise durch die US Food and Drug Administration (FDA) lizenziert sind. Auch Google betont, dass weitere umfangreiche Forschung und Validierung erfolgen müsse, bevor es zur Anwendung bei sicherheitskritischen medizinischen Aufgaben kommen kann (Google Research).

## 2 Anwendungspotenzial

Im nachfolgenden Absatz wird ein Überblick über die Potenziale (und Grenzen) von generativer KI in klinischer Praxis, Verwaltung, medizinischer Forschung und medizinischer Ausbildung gegeben.

### 2.1 Dokumentation und Verwaltung

Generative KI kann für die medizinische Dokumentation zum Beispiel zum Zusammenfassen von Gesprächen oder von Patientendaten genutzt werden. Me-

dizinische Daten wie Anamnese, Medikamente, Diagnosen, Behandlungen oder Laborergebnisse können so aufgearbeitet und prägnant zusammengefasst werden (Zhang und Boulos 2023). Zudem können Patientenakten analysiert und Diskrepanzen oder Lücken identifiziert werden und wichtige Informationen wie empfohlene Behandlungen oder ein Verlauf über die Zeit analysiert und hervorgehoben werden. Hierdurch kann die Dokumentationslast deutlich verringert und Lücken in der Behandlung der Medikamenteninteraktionen oder Kontraindikationen können besser identifiziert werden. Eine große Herausforderung stellt derzeit noch die Evaluierung dar, da es weder einheitliche noch für alle Fälle geeignete Evaluationsmetriken gibt. Zudem sind die Trainingsdaten häufig sowohl in Bezug auf die Menge als auch die Qualität nicht ausreichend, um das Modell für die medizinspezifischen Anforderungen ausreichend zu trainieren. Dies kann dazu führen, dass die Modelle nicht ausreichend generalisieren, also nicht auf eine Vielzahl von unterschiedlichen Daten übertragbar sind.

Generative KI kann zudem für die Erstellung von Abrechnungen oder zur Terminvereinbarung genutzt werden. Solche risikoärmeren Anwendungen befinden sich zum Teil schon in Gebrauch, etwa in Form des AI-Powered Medical Coding. Dennoch konnte eine kürzlich veröffentlichte Studie zeigen, dass die LLMs GPT-3.5, GPT-4, Gemini Pro, und Llama2-70b nur schlecht medizinische Abrechnungs-codes erstellt haben, weswegen die Autoren der Studie die getesteten LLMs momentan als nicht geeignet für diese Aufgabe ansehen und weitere Forschung und Anpassung der Modelle auf diese spezifische Fragestellung für notwendig erachten (Soroush et al. 2024).

## 2.2 Forschung

Bisher war die Entwicklung von neuen Medikamenten langwierig und kostspielig. Mit dem Einsatz von generativen KI-Modellen in der Arzneimittelforschung ist es nun möglich, chemische Strukturen und deren Variationen zu simulieren, um so viel schneller neue Proteine zu entwickeln, als dies mit bisherigen Methoden möglich war (Vert 2023). Darüber hinaus kann generative KI durch die Analyse großer Datenmengen potenzielle Probleme in Bezug auf Wirksamkeit, Sicherheit und Nebenwirkungen identifizieren und so zu einer Beschleunigung in einem entscheidenden Schritt der Arzneimittelentwicklung führen (Reddy

2024; Vert 2023). Zusätzlich können die Modelle für die Entwicklung personalisierter Therapien, zum Beispiel bei Krebspatienten, genutzt werden. Insgesamt bietet die generative KI enormes Potenzial, mit ihr schneller und besser wirksame und bei Bedarf personalisierte Medikamente entwickeln zu können, dennoch steht auch die Pharmabranche noch ganz am Anfang der Implementierung und Skalierung (Viswa et al. 2024).

## 2.3 Medizinische Ausbildung

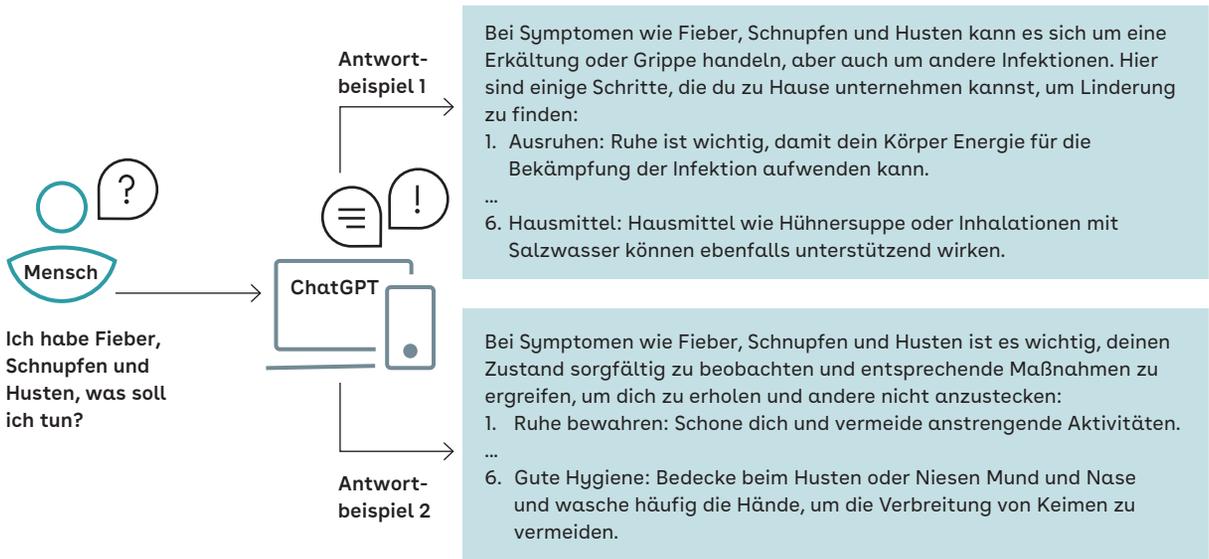
Auch für die medizinische Ausbildung, also sowohl während des Studiums als auch während der Facharzt-ausbildung, gibt es vielfältige potenzielle Anwendungen. Dazu gehören realistische Simulationsszenarien, personalisiertes Lernmaterial samt Feedback und die Bereitstellung von zusammengefassten wissenschaftlichen Artikeln. Aufgrund von Befürchtungen hinsichtlich wissenschaftlicher Integrität, Veränderungen im Lernstil oder Ungenauigkeiten in dem durch generative KI aufbereiteten Lernmaterial wurden von vielen Universitäten Regeln zur Verwendung von generativer KI erlassen. Auch wird angenommen, dass die Anwendung von (generativer) KI zu einem Wissens- und Fähigkeitsverlust in bestimmten Bereichen führen wird. Mit dem fortschreitenden Einzug von (generativer) KI in den klinischen Alltag wird es deswegen umso wichtiger, dass Studierende sowie Ärztinnen und Ärzte lernen, Algorithmen kritisch zu bewerten und ihre Limitationen zu verstehen.

## 2.4 Chatbots für Patienten

Bereits heute können LLM Chatbots wie ChatGPT und Bard von Patienten genutzt werden, um sich über Symptome zu informieren oder medizinischen Rat zu holen. Chatbots haben das Potenzial, mit Patienten gezielt zu kommunizieren. Aktuell besteht die Herausforderung, dass die LLMs nicht immer korrekt antworten. Zum Beispiel weist GPT-4 darauf hin, dass es nicht „zur Diagnose von Krankheiten oder anderen Zuständen oder zur Heilung, Linderung, Behandlung oder Prävention von Krankheiten verwendet werden soll“, und auch, dass „GPT-4 [...] Fehler machen [kann]. Es ist ratsam, wichtige Informationen zu überprüfen“ (OpenAI o. J.). Zudem wurde bisher keines der LLMs durch eine europäische Behörde oder durch die FDA überprüft. Neben der notwendigen Anpassung auf medizinische Fragestellungen und der Durchführung von klinischen Studien (Thiru-

## Abbildung 2 ChatGPT: gleiche Frage, stets unterschiedliche Antworten

Quelle: eigene Darstellung; Grafik: G+G Wissenschaft 2024



ChatGPT variiert seine Antworten auf identische Fragen. Dargestellt sind Ausschnitte aus Original-Antworten.

*navukarasu et al. 2023*) sind auch Haftungsfragen noch nicht geklärt. In **Abbildung 2** ist dargestellt, wie dieselbe medizinische Frage GPT-4 zweimal gestellt wird. Es ist auffällig, dass sich die Antworten unterscheiden. Bisherige KI-Modelle haben, einmal trainiert, in der Regel dieselbe Antwort auf eine Aufgabe gegeben. Im Fall der generativen KI wird jedes Mal etwas Neues erzeugt, weswegen sich die gegebenen Antworten unterscheiden. Dies verdeutlicht auch die Komplexität, die bei der Überprüfung der Antworten auf Korrektheit besteht. Zukünftig stellt sich die Frage, inwieweit speziell für Patienten konzipierte Chatbots entwickelt und kontinuierlich auf dem neuesten Wissensstand gehalten und gleichzeitig rigoros auf das korrekte Geben von Antworten überprüft werden können.

### 3 Chancen und Risiken bei der Implementierung

Trotz der vielversprechenden Möglichkeiten, die generative KI im Gesundheitswesen bietet, steht die Implementierung in den klinischen Alltag noch ganz am Anfang, da es zahlreiche Risiken und Herausforderungen gibt. Einige davon wurden bereits im vorherigen

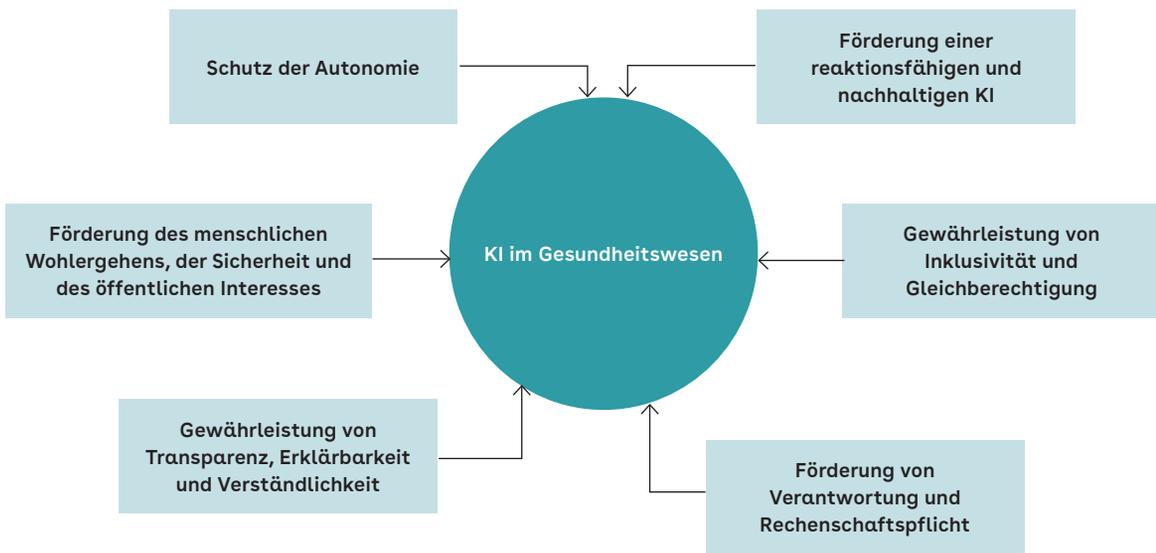
Kapitel angeschnitten, es sollen jedoch im nachfolgenden Abschnitt einige der Herausforderungen und auch erste Regelungen eingehender diskutiert werden. Die Risiken erstrecken sich von der Evaluierung der Modelle über ethische Fragestellungen, Datenschutz und Qualität der generierten Daten bis hin zu Haftungsfragen, Transparenz, Regulierung und Erklärbarkeit.

### 3.1 Halluzination, Bias und Transparenz

Generative KI und insbesondere LLMs zeichnen sich dadurch aus, dass sie kohärente, menschenähnliche Antworten geben, die zudem häufig durch Beispiele untermauert werden und so auch dann plausibel erscheinen, wenn die Antworten nicht korrekt sind. Deswegen kommt es gelegentlich vor, dass falsche, erfundene Antworten gegeben werden, die dem Leser aus den eben genannten Gründen dennoch glaubwürdig erscheinen. Dieses Phänomen nennt man Halluzination. Halluzinationen können nicht nur bei LLMs, also Texten, sondern auch bei generiertem Audio-, Bild- oder Videomaterial vorkommen. Bekannte Beispiele von Chat-GPT sind das Erfinden/Halluzinieren von

Abbildung 3 Ethikprinzipien der WHO für KI im Gesundheitssektor

Quelle: WHO 2024; Grafik: G+G Wissenschaft 2024



Aufgrund der großen Potenziale und hohen Risiken hat die WHO im Januar 2024 einen Leitfaden zum Umgang mit KI im Gesundheitswesen veröffentlicht. Er enthält sechs Prinzipien, die die WHO im Leitfaden näher ausführt.

Quellen, die es gar nicht gibt (McGowan et al. 2023). In einigen Fällen sind Halluzinationen offensichtlich falsch, viele sind jedoch nicht auf den ersten Blick erkennbar, was für die Anwendung im Gesundheitssektor ein großes Risiko darstellt.

Bias ist eine weitere Herausforderung bei KI-Modellen. Ein Bias ist eine systematische Verzerrung von Ergebnissen, wodurch vorhandene Vorurteile oder Annahmen aus den Trainingsdaten in den Antworten des KI-Modells erneut aufgezeigt oder sogar verstärkt werden können. Bias kann bei allen KI-Modellen vorkommen, nicht nur bei generativer KI. Bei ihr besteht jedoch die vermehrte Möglichkeit, dass die neu generierten falschen Aussagen als wahr angenommen werden und möglicherweise als Grundlage für weitere Trainings dienen und sich so auch selbst verstärken können. Bias kann sich somit negativ auf einzelne Gruppen oder Minderheiten auswirken und im Gesundheitsbereich beispielsweise die Chancengleichheit vermindern (Zhou et al. 2024).

Eine weitere Herausforderung bei der Beurteilung von KI-Modellen ist die sogenannte Blackbox-Problematik, die beschreibt, dass die Entscheidungen von KI-Modellen häufig für den Anwender nicht transparent oder nachvollziehbar sind. Im Gesundheitssektor

ist es jedoch von besonderer Wichtigkeit, zu verstehen, auf Basis welcher Annahmen eine Entscheidung von dem KI-Modell getroffen wurde. Obwohl schon relevante methodische Fortschritte bei der Lösung der Blackbox-Problematik erzielt wurden, bleibt sie eine Herausforderung und erfordert weitere Forschung (o. V. 2023).

### 3.2 Klinische Evaluation und Haftung

Die Evaluation von generativer KI nicht nur in Studien, sondern auch kontinuierlich im klinischen Alltag ist eine besondere Herausforderung. Die rasante Entwicklung generativer KI erschwert ihre klinische Evaluation, Regulierung und Zertifizierung. Generative KI-Modelle, die beispielsweise bei der Diagnosestellung unterstützen sollen, müssen als Medizinprodukt zertifiziert sein. Damit die generativen KI-Modelle gut funktionieren, sollten sie fortlaufend und verlässlich trainiert sowie aktualisiert werden können. Momentan ist es aber in Deutschland so, dass KI-Modelle nur „fertig trainiert“ als Medizinprodukt zertifiziert werden können und kontinuierlich weiter lernende Modelle nicht, da momentan nicht fortlaufend die Qualität dieser Modelle durch die Prüfstellen beurteilt werden kann.

## 4. Governance zur Regulierung

Im Zusammenhang mit Regelungsmechanismen für (generative) KI sind besonders Dokumente auf Welt- und Europaebene von Wichtigkeit. Sie werden im Folgenden skizziert.

### 4.1. Die WHO-Leitlinie

Aufgrund der großen Potenziale bei gleichzeitigen vielen Risiken von LLMs für die weltweite Gesundheit hat die Weltgesundheitsorganisation (WHO) im Januar 2024 den „Leitfaden Ethik und Verwaltung künstlicher Intelligenz im Gesundheitswesen: Leitlinien für große multimodale Modelle“ veröffentlicht (WHO 2024a). Ziel ist, dass durch angemessenen Einsatz von LLMs die Gesundheit der Bevölkerung gefördert, aber gleichzeitig auch der Schutz der Gesundheit gewährleistet wird. Jeremy Farrar, Chef-Wissenschaftler der WHO sagt (wörtliche Übersetzung aus dem Englischen): „Generative KI-Technologien haben das Potenzial, die Gesundheitsversorgung zu verbessern, aber nur, wenn diejenigen, die diese Technologien entwickeln, regulieren und nutzen, die damit verbundenen Risiken identifizieren und vollständig berücksichtigen. Wir benötigen transparente Informationen und Richtlinien zur Steuerung des Designs, der Entwicklung und der Nutzung von LLMs, um bessere Gesundheitsergebnisse zu erzielen und anhaltende Gesundheitsungleichheiten zu überwinden“ (WHO 2024b). Um diese Ziele zu erreichen, werden im Leitfaden mehr als 40 Empfehlungen an Regulierungsbehörden, Gesundheitsdienstleister, Wissenschaftler und (Technologie-) Unternehmen ausgesprochen. Die **Abbildung 3** präsentiert einige der ethischen Prinzipien, die beim Einsatz von KI im Gesundheitswesen von der WHO als Konsens publiziert wurden (WHO 2024a).

### 4.2 Das KI-Gesetz der EU

Die Europäische Union (EU) ist mit der Verabschiedung des EU-KI-Gesetzes (Artificial Intelligence Act) im März 2024 (EP 2024a) weltweit Vorreiter. Laut dem Gesetz werden KI-Anwendungen nach Risiko eingestuft, von minimalem Risiko (hier reicht ein *code of conduct*) über begrenztes (Transparenzkriterien), hohes (Konformitätsbewertungen) und unannehmbares Risiko (verboten) (EP 2024a). Abhängig von der Risikoklasse müssen dann unterschiedliche Auflagen erfüllt

werden. Ein ChatBot in einem Terminbuchungsportal würde als „begrenztes Risiko“ eingestuft werden und müsste demnach bestimmte Transparenzkriterien erfüllen. Eine KI, die zum Beispiel in einer Insulinpumpe verbaut ist und die Insulinabgabemenge berechnet, ist als „hohes Risiko“ einzustufen und müsste somit weitere Auflagen erfüllen. KI-Modelle wie *social scoring*, die Priorisierung von Patienten aufgrund politischer oder religiöser Ansichten oder eines mit dem Ziel der Echtzeit-Biometrie-Erkennung werden als „unannehmbares Risiko“ klassifiziert und sind demnach verboten (EP 2024b). Generative KI-Modelle fallen in den allermeisten Fällen unter die Regelung für „KI-Modelle für allgemeine Zwecke“ (*general purpose models*) und müssen besondere Auflagen, wie zum Beispiel „die technische Dokumentation des Modells einschließlich des Ausbildungs- und Prüfungsverfahrens und der Ergebnisse seiner Bewertung“, erfüllen (EP 2024c).

## 5 Fazit

Seit der Veröffentlichung von ChatGPT im November 2022 hat generative KI einen regelrechten Hype ausgelöst. Auch für das Gesundheitswesen werden das Potenzial und die Anwendungsmöglichkeiten von generativer KI als enorm eingeschätzt. Die Implementierung in den klinischen Alltag ist jedoch sehr herausfordernd und steht erst ganz am Anfang. Themen reichen von Datenschutz über Haftungsfragen, Evaluation von klinischen Studien, kontinuierlichem Monitoring von generativer KI und der Frage, wie Ärzte in Zukunft aus- oder weitergebildet werden müssen, um (generative) KI sicher im klinischen Alltag anwenden zu können. Auch die Gesetzgebung und Regulatorik steht vor neuen Herausforderungen. Mit dem Artificial Intelligence Act des EU-Parlaments wurde ein erster wichtiger Schritt gegangen, und auch die WHO hat mit ihrem Leitfaden zu LLMs zur Zusammenarbeit von Programmierern, Wissenschaftlern, Anwendern und Regulatoren aufgerufen. Dennoch sind noch viele Fragen offen und weitere gesetzliche Regelungen notwendig. Obwohl technisch schon sehr viel möglich ist, kann eine sichere Anwendung in vielen Bereichen im Gesundheitswesen noch nicht erfolgen. In Deutschland benötigen wir zudem innovative Ansätze, um relevante Akteure zusammenzubringen, damit Lösungen gefunden werden, wie generative KI dabei unterstützen kann, das Gesundheitssystem besser, qualitativ hochwertiger und effizienter (insbesondere vor dem Hintergrund des Fachkräftemangels) gestalten zu können.

## Literatur

- Brin D et al. (2023): Comparing ChatGPT and GPT-4 Performance in USMLE Soft Skill Assessments. Scientific Reports, Vol. 13, No. 1, Article number 16492
- EP (Europäisches Parlament) (2024a): Artificial Intelligence Act – European Parliament Legislative Resolution of 13 March 2024 on the Proposal for a Regulation of the European Parliament and of the Council on Laying down Harmonised Rules on Artificial Intelligence (Artificial Intelligence Act) and Amending Certain Union Legislative Acts (COM(2021)0206 – C9-0146/2021 – 2021/0106(COD)); europa.eu/doceo/document/TA-9-2024-0138\_EN.pdf
- EP (Europäisches Parlament) (2024b): KI-Gesetz: erste Regulierung der künstlichen Intelligenz; europa.eu/topics/de → Digitales → Künstliche Intelligenz
- EP (Europäisches Parlament) (2024c): Artificial Intelligence Act – Artikel 53: Verpflichtungen für Anbieter von KI-Modellen für allgemeine Zwecke; artificialintelligenceact.eu/de/article/53/
- Epstein Z, Hertzmann A and the Investigators of Human Creativity (2023): Art and the Science of Generative AI. Understanding Shifts in Creative Work Will Help Guide AI's Impact on the Media Ecosystem. Science, Vol. 380, No. 6650, 1110–1111
- Esteva A et al. (2017): Dermatologist-Level Classification of Skin Cancer with Deep Neural Networks. Nature, Vol. 542, No. 7639, 115–118; nature.com → Explore content → Browse issues → Vol. 542, No. 7639
- Goodfellow, IJ et al. (2014): Generative Adversarial Nets. arXiv, 10 Jun 2014; arxiv.org/pdf/1406.2661
- Google Research (o. J.): Med-PaLM: A Large Language Model from Google Research, Designed for the Medical Domain; sites.research.google/med-palm/
- McGowan A et al. (2023): ChatGPT and Bard Exhibit Spontaneous Citation Fabrication during Psychiatry Literature Search. Psychiatry Research, Vol. 326, Article No. 115334
- Meta (2024): Meditron: An LLM Suite Especially Suited for Low-Resource Medical Settings Leveraging Meta Llama. April 25, 2024; ai.meta.com/blog → More from AI at Meta
- o. V. (ohne Verfasser) (2023): ChatGPT is a Black Box: How AI Research Can Break it Open. Nature, Vol. 619, No. 7971, 671–672; nature.com → Explore content → Browse issues → Vol. 619, No. 7971
- OpenAI (o. J.): ChatGPT 4; chatgpt.com
- Reddy S (2024): Generative AI in Healthcare: An Implementation Science Informed Translational Path on Application, Integration and Governance. Implementation Science, Vol. 19, No. 1, 27
- Singhal K et al. (2023): Large Language Models Encode Clinical Knowledge. Nature, Vol. 620, No. 7972, 172–180; nature.com → Explore content → Browse issues → Vol. 620, No. 7972
- Soroush et al. (2024): Large Language Models Are Poor Medical Coders – Benchmarking of Medical Code Querying. NEJM AI, Vol. 1, No. 5; ai.nejm.org → current issue → browse all issues → Vol. 1, No. 5
- Takefuji Y (2023): A Brief Tutorial on Generative AI. British Dental Journal, Vol. 234, No. 12, 845
- Thirunavukarasu AJ et al. (2023): Large Language Models in Medicine. Nature Medicine, Vol. 29, No. 8, 1930–1940
- Vert JP (2023): How Will Generative AI Disrupt Data Science in Drug Discovery? Nature Biotechnology, Vol. 41, No. 6, 750–751
- Viswa CA et al. (2024): Generative AI in the Pharmaceutical Industry: Moving from Hype to Reality; mckinsey.com → Suche: Generative AI in the pharmaceutical industry
- WHO (World Health Organization) (2024a): Ethics and Governance of Artificial Intelligence for Health: Guidance on Large Multi-Modal Models; iris.who.int → Search: Large Multi-Modal Models
- WHO (World Health Organization) (2024b): WHO Releases AI Ethics and Governance Guidance for Large Multi-Modal Models; who.int → Search: Large Multi-Modal Models
- Zhang P, Boulos MNK (2023): Generative AI in Medicine and Healthcare: Promises, Opportunities and Challenges. Future Internet, Vol. 15, No. 9, 286; mdpi.com → Search Title: Generative AI in Medicine and Healthcare; Author: Zhang
- Zhou M et al. (2024): Bias in Generative AI. arXiv, 5 Mar 2024; arxiv.org/abs/2403.02726

(letzter Zugriff auf alle Internetquellen: 3. Juni 2024)

## Die Autorin



**Dr. med. Mirja Mittermaier** hat Humanmedizin an der Medizinischen Hochschule Hannover studiert und am dortigen Institut für Virologie promoviert. Auslandsaufenthalte führten sie an die University of Cambridge, UK, und an die Tongji University, China. Zudem hat sie einen Master of Biomedical Informatics an der Harvard Medical School, USA, erworben. Seit 2016 arbeitet sie als Ärztin an der Charité – Universitätsmedizin Berlin. Parallel ist sie wissenschaftlich tätig.

### Kontakt

Dr. med. Mirja Mittermaier, Fächerverbund für Infektiologie, Pneumologie und Intensivmedizin, Charité – Universitätsmedizin Berlin, Charitéplatz 1, 10117, Berlin, Telefon: 030 450653334, E-Mail: mirja.mittermaier@charite.de