

Künstliche Intelligenz und Apps in der Anästhesiologie

Artificial intelligence and apps used in anaesthesiology

R. Rümmler¹

► **Zitierweise:** Rümmler R: Künstliche Intelligenz und Apps in der Anästhesiologie. Anästh Intensivmed 2025;66:284–291. DOI: 10.19224/ai2025.284

Zusammenfassung

Das Manuskript behandelt die zunehmende Integration von künstlicher Intelligenz (KI) und mobilen Anwendungen in die Anästhesiologie und Intensivmedizin. Es beschreibt die Grundlagen der KI, einschließlich verschiedener Lernmethoden (Supervised, Unsupervised, Reinforcement Learning) und KI-Typen (Generalisierte, Generative, Klassische KI). Anwendungsgebiete wie präoperative Risikoeinschätzung, intraoperative Überwachung und postoperative Intensivmedizin werden detailliert dargestellt. Zudem werden Limitationen wie unzureichende Trainingsdaten, Modellkomplexität und fehlende Transparenz diskutiert. Der Einsatz von Apps, ihre Kategorien und Herausforderungen im klinischen Kontext sowie Datenschutzproblematiken bei Messenger-Diensten und die Nutzung von Large Language Models werden ebenfalls thematisiert. Abschließend wird die Bedeutung einer kritischen Auseinandersetzung und gezielter Implementierung dieser Technologien hervorgehoben.

Summary

The manuscript addresses the increasing integration of artificial intelligence (AI) and mobile applications in anaesthesiology and intensive care medicine. It outlines the fundamentals of AI, including various learning methods (Supervised, Unsupervised, Reinforcement Learning) and types of AI (Generalized, Generative, Classical AI). Application areas such as preoperative risk assessment, intra-

operative monitoring, and postoperative intensive care are detailed. Moreover, limitations such as insufficient training data, model complexity, and lack of transparency are discussed. The use of apps, their categories and challenges in the clinical context, as well as privacy issues occurring in connection with messaging services and the utilization of Large Language Models are also addressed. Finally, the importance of a critical discussion and targeted implementation of these technologies is emphasized.

Einleitung

Rasante Fortschritte auf dem Gebiet der Implementierung von **künstlicher Intelligenz (KI)** und von **maschinellern Lernen (ML)** nehmen zunehmend auch Einzug in den medizinischen Alltag. Besonders in Anästhesiologie und Intensivmedizin haben sich hier aufgrund vieler gut standardisierter sowie algorithmenbasierter Behandlungspfade zahlreiche Ansätze, KI-gestützte Systeme zur Entscheidungsfindung zu nutzen, identifizieren lassen. Die immer größer werdende Menge an Daten, die während der Behandlung von Patient*Innen in modernen Gesundheitssystemen generiert wird, scheint für einzelne Behandelnde nicht mehr allein sinnvoll interpretierbar, und auch hier könnten maschinell gestützte Analysen helfen, modellbasiert systematische Symptomkonstellationen zu erkennen, lange bevor dies selbst erfahrenen Kliniker*Innen gelingt. Zusätzlich drängen immer mehr Hersteller

Zertifizierte Fortbildung

CME online

BDA- und DGAI-Mitglieder müssen sich mit ihren Zugangsdaten aus dem geschlossenen Bereich der BDA- und DGAI-Webseite unter der Domain www.cme-anesthesiologie.de anmelden, um auf das Kursangebot zugreifen zu können.

1 Klinik für Anästhesiologie und Intensivmedizin, Universitätsmedizin der Johannes Gutenberg-Universität Mainz (Direktor: Prof. Dr. C. Werner)

Interessenkonflikt

Der Autor erhielt Reiseunterstützungen zu Symposien von Stryker und Medtronic.

Schlüsselwörter

Künstliche Intelligenz – Maschinelles Lernen – Mobile Anwendungen – Datenschutz

Keywords

Artificial Intelligence – Machine Learning – Mobile Applications – Data Protection

von mobilen Anwendungen (Apps) sowohl auf Patient*Innenseite als auch auf Seiten der Behandelnden auf den Markt und zwingen alle Beteiligten, sich ernsthaft mit sich zügig ändernden Prozessen zu befassen und zu arrangieren. Eine exponentiell größer werdende Menge an Publikationen zu den Themen KI und App-basierte Behandlungen (Abb. 1) unterstreicht die Notwendigkeit nach einem Grundverständnis der genutzten Begriffe und einer akkuraten Kommunikation, um Missverständnisse zu vermeiden [1].

Getrieben von einer im Rahmen der COVID-19-Pandemie notgedrungen stark beschleunigten Digitalisierung vieler Kommunikationswege im Medizinsektor und einem regelrechten Hype um die Entwicklung generativer Sprachmodelle (i. e. ChatGPT, DeepSeek u. ä.) wurden Umwälzungsprozesse in Gang gesetzt, die auch unser Fachgebiet nachhaltig verändern werden. Der folgende Artikel soll als Handreichung gedacht sein, um in dem komplexen Geflecht aus Fachbegriffen, multimedialer Dauerexposition, unübersichtlicher Evidenzlage und teils sehr individuellen Datenschutzüberzeugungen auf Basis eines grundlegenden Verständnisses für die Thematik argumentieren und ggf. Entscheidungen treffen zu können.

Für jegliche konkret genannten Anwendungen bestehen dabei keinerlei Interessenkonflikte seitens des Autors. Aufgrund der Vielschichtigkeit der angesprochenen Themen besteht kein Anspruch auf Vollständigkeit. Spezielle Einzelaspekte wurden bereits in anderen Arbeiten in dieser Zeitschrift ausführlich diskutiert und werden hier nicht dezidiert besprochen [2,3].

Grundlagen der Künstlichen Intelligenz in der Anästhesie

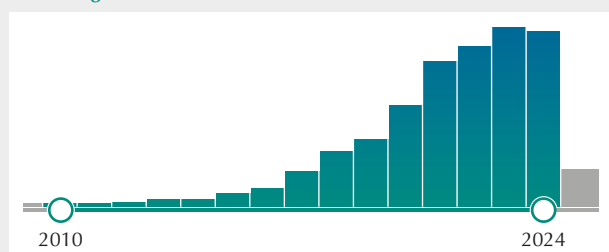
KI umfasst eine Vielzahl von Algorithmen, die aus Daten lernen und Muster erkennen können. Insbesondere ML ist in der Medizin vielversprechend. Das Verständnis und die Implementierung dieser Technologien erfordern jedoch ein hohes Maß an Fachwissen und die Bereitschaft, kontinuierlich zu lernen und sich anzupassen. Verschiedene Ansätze im Bereich des ML bieten unterschiedliche Vorteile und Herausforderungen, je nachdem, wie die Algorithmen trainiert und angewendet werden. Hierbei lassen sich verschiedene Ansätze unterscheiden:

- **Supervised Learning:** Algorithmen lernen aus gekennzeichneten Daten und können Vorhersagen treffen. Die Fragestellung ist in der Regel sehr eng gefasst und der Lösungsweg vorgezeichnet, jedoch häufig sehr **rechenintensiv**.
- **Unsupervised Learning:** Daten werden ohne vorherige Etikettierung analysiert, um Cluster oder Zusammenhänge zu finden. Vor allem zur Mustererkennung in großen Datenmengen oder unübersichtlichen Datenkonstellationen genutzt, häufig als erster Schritt zu einer sinnvollen Hypothesengeneration.
- **Reinforcement Learning:** Ein System verbessert sich durch Feedback, ähnlich einem Lernprozess durch Versuch und Irrtum. Gestartet wird oft mit einer konkreten Aufgabe, der Lösungsweg wird über eine meist sehr hohe Zahl an Iterationen erreicht und im Verlauf als optimaler Lösungsweg gespeichert.

Abgeleitet von den oben genannten Lernprotokollen können in hochmodernen KI-Programmen mittels weiterer Analyseschritte und interner Feedbackschleifen (sogenannte „Blackbox“) komplexe neuronale Netzwerke simuliert werden und eine theoretisch komplett autarke Generierung von Analyseergebnissen und Interpretationen ermöglicht werden. Je nach Reifegrad dieser Systeme und der genutzten Lernprotokolle lassen sich die Systeme weiter aufteilen:

- **Generalisierte oder allgemeine KI:** Ein völlig autarkes System, das unabhängig von Input und Kontext selbstständig arbeitet, lernt und grundsätzlich (abhängig vom Zugang zu Lernmaterial) auch zur Entwicklung „eigener Ideen“ befähigt ist. Während unsaubere Berichterstattung und popkulturelle (Fehl-) Interpretationen dies häufig als den Archetyp der KI definieren (siehe „Skynet“), existieren zum jetzigen Zeitpunkt noch keine Systeme, die ernsthaft in diese Klassifikation zu fassen wären.
- **Generative KI:** Ein Algorithmus, der auf der Basis teils immenser Datenmengen und Lerniterationen probabilistische Entscheidungen auf der Basis einer konkreten Anfrage („Prompt“) ausgibt. Der mögliche Output kann hier immer nur so gut sein wie die genutzte (möglichst kuratierte) Datenbasis, weshalb hier zumeist (teil-)supervidierte Lernmodelle genutzt werden. Beispiele sind Large-Language-Model-Generatoren wie ChatGPT oder DeepSeek sowie komplexere Bildanalysesysteme wie DALL-E oder Canva.
- **Klassische oder „einfache“ KI:** Programme mit ausschließlich supervidiertem Lernalgorithmus und stark begrenztem oder singulärem Anwendungszweck. Meistens werden konkrete Datenströme auf bestimmte Fragestellungen analysiert und eine Auswahl an dezidierten Reaktionen/Empfehlungen erwartet. Einfache KI-Anwendungen finden sich mittlerweile in fast allen modernen Systemen, die mit Messtechnik jedweder Art arbeiten. Sie stellen

Abbildung 1



Ergebnis der Suchanfrage „artificial intelligence + anesthesia“ über Pubmed.org, Ergebnisse von 2010 bis 2024, Abfragedatum 12.02.2025. Vergleichbare Ergebnisse werden erzielt zum Suchbegriff „app-based anesthesia“.

bislang auch die einzigen evidenzbasiert angewendeten Modelle in der Medizin dar (z. B. BIS, NOL Index, Hypotension Prediction Index).

Eine adäquate Nomenklatur relevanter KI-Modelle ist in der Kommunikation essenziell, um Unklarheiten und Verunsicherungen zu vermeiden.

Anwendungsgebiete von KI in der Anästhesie

Überblick

Wie bereits erwähnt, eignet sich besonders ein auf Messwerte ausgerichtetes Fach wie die Anästhesiologie zumindest theoretisch sehr gut zu einer Anwendung datengetriebener KI-Programme. Häufig wird zur besseren Übersichtlichkeit dabei auf eine Einteilung auf die einzelnen Anästhesiesäulen (AINSP) zurückgegriffen. Potenziell von ML profitierende Prozesse lassen sich jedoch vor allem im prä-, peri- und postoperativen Setting finden, weshalb diese Aufteilung hier gewählt werden soll. Notfallmedizinische KI-Projekte, die häufig im Rahmen telemedizinischer Kommunikationsinitiativen verortet sind, werden in dieser Arbeit nicht behandelt.

Präoperative Planung und Risikoeinschätzung

Während der präoperativen Vorbereitung besteht seitens der anästhesiologischen Betreuung häufig nur ein sehr kurzer Kontakt zu Patient*Innen und der Informationsfluss- und -zugang kann je nach spezifisch vor Ort herrschenden Prozessabläufen deutlich eingeschränkt sein. Umso wichtiger ist es, aus den zugänglichen Untersuchungen ein mögliches Maximum an **prädiktiven Outcome-relevanten Datenpunkten** zu extrahieren und zu **optimalen Behandlungsentscheidungen** zu integrieren. KI-Modelle können hier besonders in der **Risikoprädiktion** gute Dienste leisten und konnten hier auch schon in verschiedenen Kontexten evaluiert werden.

Tavolara et al. zeigten beispielsweise anhand eines trainierten Gesichtserkennungsalgorithmus durch die Nutzung einer einzigen frontalen Fotoaufnahme eine mindestens genauso gute – und teilweise sogar bessere – Prädiktion schwieriger Intubationsbedingungen als konventionelle Scoring-Systeme [4].

Auch vor dem Hintergrund der Patient*Innensicherheit und der Vermeidung von Patient*Innenverwechslungen können KI-Applikationen nützliche und objektivierbar zuverlässigere Unterstützung bieten. So nutzten chinesische Forscher*Innen biometrische Daten zur sicheren präoperativen Patient*Innenidentifikation vor der Einschleusung in den OP-Bereich [5]. Inwiefern solche Verfahren in datenschutzrechtlich rigider aufgestellten Systemen realistisch wären, soll an späterer Stelle diskutiert werden.

Intraoperative Überwachung und Entscheidungsunterstützung

Im intraoperativen Setting ist die schnelle und präzise Beurteilung verschiedenster Patient*Innenparameter Kernaufgabe anästhesiologischer Versorgung. Gerade bei komplexen Eingriffen kann es hier zur Überfrachtung kommen und es droht eine unterbewusste Selektion vermeintlich unwichtiger Vitalzeichen. In Echtzeit durchgeführte Datenanalysen können hier helfen, komplexe Parametergruppen zu überwachen, etwaige risikobehaftete Konstellationen zu identifizieren und diese im besten Fall den Behandelnden in aufbereiteter Form zu präsentieren.

Eines der am meisten im Alltag genutzten Produkte stellt der **Bispektralindex** zur Überwachung der Narkosetiefe dar, aber auch **Nozizeptionsmessungen** greifen auf eine indizierte Vereinfachung einer multiparametrischen Messerhebung zurück [6,7]. Als einer der am häufigsten untersuchten Vertreter „neuer“ KI-basierter Algorithmen wird seit mehreren Jahren der **Hypotension Prediction Index (HPI)** der Firma Edwards beschrieben. Auch hier wird aus mehreren ohnehin erhobenen Vitalparametern während einer anästhesiologischen Behandlung ein Score ermittelt, mit dem das Risiko eines hypotensiven Ereignisses in den nächsten Minuten vorhergesagt wer-

den soll. Ob das Produkt trotz bereits erfolgter Kommerzialisierung allerdings tatsächlich einen relevanten Unterschied zur Evaluation eines anästhesiologischen Kollegen vor Ort machen würde, ist nicht zuletzt auch durch ein möglicherweise unvorteilhaft gewähltes Trainingskollektiv weiter Teil lebhafter Debatten und beispielhaft für die Probleme bei der Implementierung neuer Technologien in den Narkosealltag [2,3,8,9].

Neben der klassischen Datenaggregation in dimensionslose Scores zur besseren Visualisierung existieren aber auch andere Ansätze zur maschinellen Unterstützung der Narkoseführung. So wurden Patient*Innen im Rahmen einer Analgosedierung mit Mikrofonen überwacht und Änderungen in der Schwingungsfrequenz der oberen Atemwege analysiert. Frequenzänderungen wurden mit einer drohenden Atemwegsverlegung korreliert und eine entsprechende Warnung wurde an die Behandelnden übermittelt, um gegebenenfalls die Narkosetiefe zu reevaluieren oder andere Maßnahmen zu ergreifen, um den Atemweg zu schützen [10].

Postoperative Überwachung

Den bislang größten Einfluss zeigen ML-basierte Ansätze aktuell in der postoperativen und hier speziell **intensivmedizinischen Überwachung**. Die im Vergleich zur perioperativen Versorgung weitaus größere Menge erhobener Datenpunkte, ergänzt durch bildgebende Verfahren, mikrobiologische Proben, Laborwerte etc. sowie die Fokussierung auf mehrere Patient*Innen gleichzeitig scheint geradezu prädestiniert dazu, zuverlässig objektive KI-Programme in die klinische Einschätzung einzubinden. Extensive Datenanalysen konnten hier nicht nur zur Etablierung prädiktiver Modelle zur **frühzeitigen Sepsiserkennung** beitragen [11], die unsupervidierte Kontrolle großer Datenkollektive durch ML-Prozesse konnte auch dabei helfen, **neue Sepsis-Phänotypen** zu identifizieren, um so möglicherweise individuellere Therapieansätze und präzisere Stratifizierungen zu ermöglichen [12]. Ähnliche Untersuchungen waren in der Lage, differenzierte Aussagen über die

Sedierungstiefe und neuronale Aktivität zu liefern [7,13] oder spezifische **Endotypen für hypotensive Entgleisungen** zu definieren [14].

Die potenziellen Nutzungsmöglichkeiten erstrecken sich aber auch auf die tägliche pragmatische Therapie. Computergestützte Feedbacksysteme für die **adäquate Anordnung von Medikamenten** sowie die KI-gestützte **Diagnostik radiologischer Bilddaten** sind nur einige Beispiele der momentan beforschten und teils bereits implementierten Therapiebereiche [15]. Zusätzlich sind ML-Module dazu in der Lage, klinische Assessments zu standardisieren. So konnte eine auf die visuelle Analogskala zur Schmerzwahrnehmung trainierte KI anhand von Bildern mit guter Übereinstimmung auch bei unmöglicher Kommunikation ein Schmerzniveau identifizieren [16]. Eben diese Kommunikation wird andererseits durch ubiquitär verfügbare multilinguale Chatbots mittlerweile beinahe barrierefrei möglich. Die Nutzung derselben Sprachgeneratoren zur effektiven und zeitnahen Sicherstellung interdisziplinärer Kommunikation mittels automatisierter Anordnungen und Arztbriefherstellungen könnte zur mehr als nötigen Entlastung der Kolleg*Innen dienen und auch hier drängen immer mehr Anbieter auf den Markt.

Limitationen

Bei aller Euphorie existieren nicht ohne Grund Zweifel an der Fülle neu generierter oder bewerteter Datenmengen, vermeintlich vollkommen innovativer Therapieansätze und revolutionärer Messgeräte. Gondode et al. beschreiben in ihrer Zusammenfassung die drei größten Risikofaktoren für KI-basierte „Halluzinationen“, auf deren Basis keine klinischen Entscheidungen getroffen werden sollten, und beschreiben, wie diese zu vermeiden sind [17].

- **Insuffiziente Trainingsdaten:** Wie mehrfach angeklungen, sind faktisch alle aktuell genutzten ML-Systeme klassisch konditionierte Programme. Das bedeutet auch, dass aus diesen Systemen generierte Anwendungen und Therapieentscheidungen nur so gut sein können wie der zum

Training verwendete Datensatz. Unzureichende Daten, Stichprobenverzerrung und weitere Störfaktoren könnten sonst zu Schäden bei Patient*Innen außerhalb des Trainingskollektivs führen.

- **Hohe Modellkomplexität:** Je komplexer ein zu analysierendes Modell gewählt wird und je mehr Daten in seine Berechnungen eingehen, desto fehleranfälliger ist es auch. Speziell bei online arbeitenden, perioperativen Scores zeigen sich häufig bereits massive Abweichungen von der klinischen Realität, wenn nur ein kleiner Teil der zuführenden Sensorkanäle artefaktbedingten Schwankungen unterlegen ist. Ein blindes Vertrauen in die gelieferten Scores kann hier zu einem falschen Sicherheitsgefühl oder inadäquaten Medikamentendosierungen führen und somit Patient*Innen einem unnötigen Risiko aussetzen.
- **Fehlende Transparenz:** Je multifaktorieller KI-Systeme arbeiten müssen und je mehr Analyseschritte innerhalb der programmeigenen „Blackbox“ stattfinden, desto schwerer wird es für die Behandelnden, die resultierenden Therapieempfehlungen nachzuvollziehen. Besonders in der medikolegalen Verantwortungsinterpretation droht dadurch entweder ein blindes Befolgen der ausgegebenen Empfehlungen oder ein grundsätzliches Misstrauen gegenüber allen KI-basierten Empfehlungen. Die Entwicklung verständlicher ML-Modelle und intensive Kommunikation zwischen Herstellern und Anwendern könnte hier Abhilfe schaffen.

Eine ausführlichere Zusammenfassung möglicher **KI-Halluzinationen** bietet Tabelle 1.

Eine kritische Auseinandersetzung mit automatisiert generierten Behandlungsempfehlungen ist essenziell, um auf Messartefakten basierende Fehlentscheidungen zu vermeiden.

Apps in der Anästhesiologie

Überblick

Abseits der großen und aufwendigen KI-getriebenen klinischen Systeme gibt es schon seit Jahren eine Vielzahl an kleineren Anwendungen (engl. Applications oder kurz „Apps“), die häufig direkt auf dem Smartphone abgelegt werden können oder via Internet-Browser erreichbar sind. Die allermeisten dieser Programme sind **nicht wissenschaftlich validiert**, ihre sofortige Zugänglichkeit und ihre einfache Nutzung führen aber dazu, dass sie sich im klinischen Alltag nicht selten etabliert haben. Dies gilt nicht nur für die ärztliche Nutzung, sondern vor allem für immer mehr an Patient*Innen gerichtete Angebote, sowohl als kostenlose Freeware als auch von den Krankenkassen als digitale Gesundheitsanwendung (DiGa) vergütete „App auf Rezept“.

Eine komplette Übersicht oder Bewertung all dieser vorhandenen Optionen führt weit über die Grenzen dieses Artikels hinaus, es soll im Folgenden aber auf einige grundlegende Aspekte hingewiesen werden, die dabei helfen können, besser zu unterscheiden, welche Anwendungen existieren und welche potenziellen Probleme bei der Anwendung im klinischen Kontext entstehen können. Grundsätzlich fallen die meisten Apps in eine der folgenden Kategorien:

- **(Dosis-)Rechner-Apps:** Diese Apps stellen die wahrscheinlich am häufigsten genutzte Gruppe von Anwendungen dar. Zumeist werden vorgefertigte Formelsammlungen thematisch gebündelt angeboten, um die entsprechend gesuchten Werte zügig berechnen zu lassen. Dies reicht von einfachen BMI-Berechnungen in der Sprechstunde bis zu komplexeren Risiko-Scores oder Analysen anhand von erweitertem kardiovaskulärem Monitoring. Dosierungshilfen für Medikamente, beispielsweise bei der Versorgung von Kindern oder zur Optimierung der Gerinnung nach thrombelastometrischen Messungen, fallen ebenfalls in diese Kategorie und

Tabelle 1
KI-Halluzinationen in der Anästhesiologie (nach [17]).

Anästhesiebereich	Einfluss von KI auf die Anästhesiologie	Bedenken hinsichtlich KI-Halluzinationen
Präoperative Risikobeurteilung	Eine KI optimiert die präoperative Beurteilung durch Analyse von Patient*Innendaten, Identifizierung von Risikofaktoren und bietet maßgeschneiderte Empfehlungen.	Das Auftreten von halluzinierten Daten kann Ungenauigkeiten einführen, die potenziell Anästhesiepläne und Patient*Innenmanagementstrategien stören.
Präoperative Phase	KI erleichtert die präoperative Planung, Prognose von Ergebnissen und die interdisziplinäre Kommunikation innerhalb von Behandlungsteams.	Das Risiko der Generierung falscher Prognosen oder Empfehlungen beeinflusst chirurgische Strategien und Patient*Innenergebnisse.
PACU	KI verbessert die postanästhesiologische Versorgung durch kontinuierliche Überwachung, Vorhersage von Komplikationen und erleichtert rechtzeitige Interventionen.	Halluzinierte Daten können fehlerhafte Warnungen oder Vorhersagen auslösen, was Risiken für ein effektives Patient*Innenmanagement und die postoperative Versorgung darstellt.
Intraoperatives Management	KI stärkt die intraoperative Sicherheit durch verbesserte Überwachung, Anpassung der Anästhesie und chirurgische Koordination.	Die Generierung irreführender Überwachungsdaten oder Warnungen gefährdet Medikamentendosierungen und chirurgische Versorgung.
Patient*Innenkommunikation	KI-gesteuerte Tools optimieren die Patient*Innenaufklärung und die Kommunikation mit Gesundheitsdienstleistern und gewährleisten personalisierte Interaktionen.	KI-Halluzinationen in Kommunikationsplattformen können ungenaue oder verwirrende Informationen an Patient*Innen weitergeben, was die Wirksamkeit der Gesundheitsversorgung beeinträchtigt.
Intensivmedizin	KI ermöglicht die frühzeitige Erkennung von Zustandsverschlechterungen, Prognose von Ergebnissen und Optimierung von Behandlungsmodalitäten.	KI-Halluzinationen in Überwachungssystemen der Intensivmedizin können falsche Alarmer oder fehlerhafte klinische Vorhersagen auslösen, was Risiken für das Patient*Innenwohl darstellt.
Schmerzmedizin	KI unterstützt bei der Schmerzeinschätzung, der Erstellung individueller Behandlungspläne und der Vorhersage von Therapieerfolgen.	Ungenauigkeiten bei der Schmerzeinschätzung oder Therapievorhersagen können die Patient*Innenergebnisse negativ beeinflussen.
Forschung	KI beschleunigt Forschungsbemühungen durch die Analyse von Datensätzen und fördert die Entwicklung neuer therapeutischer Interventionen.	Kann Verzerrungen oder Ungenauigkeiten einführen, die die Integrität wissenschaftlicher Untersuchungen untergraben.
Ausbildung	KI-gestützte Simulationen verbessern die Ausbildung von Assistenzärzt*Innen durch immersive Erfahrungen und Verfeinerung klinischer Fähigkeiten.	KI-Halluzinationen in Simulationsszenarien können unrealistische oder gefährliche Trainingsumgebungen schaffen, die die Ausbildung und Fähigkeiten der Assistenzärzt*Innen beeinträchtigen.
Bildung	KI-gesteuerte Bildungsplattformen bieten Lernenden personalisierte Lernerfahrungen und Echtzeit-Feedback, fördern den Kompetenzerwerb.	KI-Halluzinationen in Bildungsmaterialien können irreführende Informationen oder Feedback an Lernende vermitteln, was die Effektivität von Bildungsmaßnahmen beeinträchtigt.

PACU: postanaesthesia care unit.

- können helfen, Dosierungsfehler zu vermeiden [18].

 - **Checklisten und visuelle Entscheidungshilfen:** Hierzu zählen Leitlinienzusammenstellungen, Sammlungen von Checklisten und visuelle Entscheidungshilfen in verschiedenster Ausprägung, die häufig mit Unterstützung der jeweiligen Fachgesellschaften entworfen werden und die bei regelmäßiger Aktualisierung nachgewiesenermaßen die Versorgung im Notfall verbessern [19].
 - **Selbstmanagement-Apps:** Diese Apps sind besonders im Rahmen der postoperativen Evaluation, Schmerztherapie und Verhaltensprävention nützlich und im Grunde die einzigen
- relevanten Vertreter unter den zugelassenen DiGa. Häufig werden in einer optimierten Bedienungsmaske Möglichkeiten zur Symptombdokumentation angeboten und auf Basis der angegebenen Beschwerden Behandlungsinformationen oder Empfehlungen zur Verfügung gestellt [20,21]. Als Variation werden ähnliche Programme auch genutzt, um pädiatrischen Patient*Innen spielerisch die Angst vor Eingriffen und Untersuchungen zu nehmen oder sie während dieser abzulenken [22,23].

 - **Edukative Apps:** Hierunter werden visuell ansprechend aufbereitete, häufig in ein umfassenderes Lernumfeld eingebettete Unterstützungspro-
- gramme subsummiert. Diese können von reinen Zusammenfassungen der zugehörigen Bücher hin zu eigenständigen Fort- und Weiterbildungsplattformen mit eigener Lernfortschrittsverfolgung reichen. Die Einbettung von unterstützendem Bild- und Videomaterial (zum Beispiel zum Erlernen regional-anästhesiologischer Verfahren oder spezifischen Ultraschalluntersuchungen) ist genauso möglich wie ein umfassendes Angebot an Wissensabfragen [24]. Auch im Rahmen der Weiterbildungsdokumentation im e-Logbuch oder der Planung und Durchführung von Entrustable Professional Abilities (EPAs) sind diese Apps interessant.

Die „ideale“ App

Unter Anästhesist*Innen gehören Überlegungen zu den Eigenschaften des idealen Narkosegases zu einer der Standardfragestellungen im Rahmen der Weiterbildung. Analog hierzu stellt Tabelle 2 eine Zusammenstellung von Attributen vor, die – ähnlich utopische – Forderungen an ideale Applikationen stellen. Während die Auseinandersetzung mit den unterschiedlichen Aspekten der von einem selbst im Alltag genutzten Apps zunächst eher spielerischen Charakter haben mag, ist besonders im interdisziplinären Kontext und ab dem Zeitpunkt, an dem Patient*Innendaten eine konkrete Rolle spielen, Vorsicht in der unkritischen Nutzung geboten.

Exkurs: Messenger-Dienste und die Datenschutz-Grundverordnung (DSGVO)

Auf einer gemeinsamen Fachtagung der Pflege und Ärzteschaft im Mai 2024 ließ sich der Präsident der Landesärztekammer Rheinland-Pfalz mit den Worten zitieren: „Wenn wir weitermachen wie bisher, ist der Datenschutz in einigen Jahren die häufigste Todesursache in Deutschland“ [25]. Die – keinesfalls als Einzelmeinung aufzufassende – Polemisierung der Datenschutzproblematik ist Ausdruck einer Frustration, die an vielen Stellen in der medizinischen Versorgung und Forschung Fuß gefasst hat. Die effektive Kommunikation immer größer werdender Datenmengen zwischen verschiedenen Behandlungspartnern, sowohl innerklinisch als auch von Klinik zu weiterbehandelnden niedergelassenen Kolleg*Innen, stellt eine immer größere Herausforderung dar. Instant **Messaging Apps** sind hierbei aus dem persönlichen und professionellen Alltag kaum noch wegzudenken und werden nahezu ubiquitär verwendet. Sowohl das fehlende Verständnis für die tatsächlichen Vorgaben der DSGVO als auch die Übersicht über die verschiedenen Messenger-Angebote mit unterschiedlichen Formen von Datenschutzversprechungen führt an vielen Stellen zu Resignation und nicht selten im Verlauf

Tabelle 2
Attribute zur Beurteilung von Apps.

„Ideale“ App-Attribute
zuverlässig
übersichtlich
intuitiv
aktuell
fachlich adäquat kuratiert
unabhängig
offline nutzbar
günstig/kostenlos
belegbarer Nutzen/disruptives Potenzial
ansprechendes Design
adäquater Datenschutz

zu einer eher unkritischen Nutzung von zweifelhaften Anbietern. Dabei sind neben den möglichen berufsrechtlichen Konsequenzen bei unsicher übermittelten Patient*Innendaten oder gar ganzen Befundvideos durch Verstöße gegen das Arztgeheimnis auch strafrechtliche Konsequenzen gegenüber dem jeweiligen Arbeitgeber zu bedenken. Verstöße gegen die DSGVO können im Extremfall mit bis zu 4 % des Gesamtumsatzes eines Betriebes bußgeldlich sanktioniert werden.

Schon 2019 veröffentlichte eine Datenschutzkonferenz auf Bundesebene ein sogenanntes **Whitepaper zur Vorgabe für Messengernutzungen im Medizinsektor** [26]. Die Liste der Vorgaben ist berechtigterweise lang und als Fazit muss konstatiert werden, dass fast kein kommerziell erhältlicher Messenger alle Vorgaben erfüllt. Besonders die adäquate Verschlüsselung der Daten und die genaue Kontrolle über die Datenströme stellen ein Problem dar, weshalb eigentlich nur sogenannte **Silo-Lösungen**, also Nachrichtendienste, deren Datenströme die eigenen Klinikserver nicht verlassen, wirklich sicher scheinen. Zur Lösung des Problems wurde 2021 via Bundesbeschluss der Auftrag an die Nationale Agentur für Digitale Medizin (gematik) erteilt, eine sichere, deutsche Messengerlösung zu entwickeln, über die alle Ärzt*Innen miteinander vernetzt

werden und Patient*Innendaten sicher ausgetauscht werden könnten. Verschiedenste Schnittstellenproblematiken und Entwicklungsverzögerungen sorgen aber bis heute dafür, dass der sogenannte TI-Messenger nicht außerhalb von Modellregionen verfügbar ist [27].

Die Nutzung kommerzieller Messenger für die Kommunikation von Patient*Innendaten oder Befunden ist in den meisten Fällen nicht DSGVO-konform und stellt ein Risiko dar.

Exkurs: Large-Language-Modelle (LLM)

Während automatisierte Chatprogramme wie IBM Watson auch im medizinischen Kontext bereits seit 2011 mit der Fähigkeit, natürliche Sprache zu imitieren und Fachwissen zu implementieren, ausgestattet wurden [28], revolutionierten die Arbeiten einer von Google finanzierten Forschergruppe 2017 die Art, wie KI-gesteuerte Sprachgeneratoren programmiert werden können [28, 29]. **Selbstaufmerksamkeits-Algorithmen** erlaubten neuen Sprachgeneratoren, deutlich präziser eingegebene Texte zu analysieren, relevante Aspekte zu gewichten und mit hoher Genauigkeit sinnvolle Ergebnisse zu produzieren. Die Qualität der ausgegebenen Texte hängt dabei maßgeblich von den zuvor genutzten Trainingsinhalten ab, deren Daten später dann in komplexen Deep-Learning-Schritten iterativ zu neuen Antworten verarbeitet werden [30,31]. Diese neue Art der Sprachgeneratoren, sogenannte **Generative Pre-trained Transformer** (GPTs), stellen die grundsätzliche Basis aller seit 2022 kommerziell und kostenfrei geradezu explosionsartig verbreiteten KI-Modelle dar. Unterschiede in der Leistungsfähigkeit ergeben sich hierbei vor allem in der Menge, Qualität und Aktualität der Trainingsdaten, der zur Verfügung stehenden Rechenkapazität der jeweiligen KI und der von Anbieter zu Anbieter unterschiedlich gestalteten teilsupervi-

dierten Kontrolle der Textausgaben. So lassen sich zwar exzellent allgemeine Zusammenhänge mit sehr hoher Sprachgenauigkeit generieren, sobald jedoch auf restringierte Inhalte wie wissenschaftliche Artikel hinter einer Bezahlbarriere zugegriffen werden soll, stoßen die Transformer an ihre Grenzen. Häufig entstehen in diesen Situationen ungenaue Zitationen (Halluzinationen) oder es wird nur sehr oberflächlich aus den zugänglichen Abstracts zitiert. Dies muss zu jeder Zeit bedacht werden, wenn sich auf künstlich generierte Texte verlassen werden soll.

Tabelle 3 zeigt beispielhaft die Ergebnisse eines Selbstversuchs mit dem Vergleich eines eher einfachen sowie eines komplexeren Prompts (= Anfrage/Aufgabe an die KI-Plattform). Neben Unterschieden in der Ausgabezeit und Wörteranzahl generierten alle KIs medizinisch akkurate Antworten, teilweise mit dezidierten Dosierungsempfehlungen für einzelne Medikamente. Besonders, wenn nach wissenschaftlichen Zitaten gefragt wird, berufen sich beinahe alle Programme jedoch nur auf allgemeine Zusammenfassungen, Open-Access-Reviews mit teilweise fragwürdiger Relevanz oder Buchkapitelauszüge. Soll der

Fokus einer Anfrage also auf fachlicher Genauigkeit liegen, stellen die Ausgaben entgegen der stilistisch und sprachlich sehr hohen Qualität weiterhin bestfalls den Ausgangspunkt für genauere Recherchen dar. Zusätzlich ist zu beachten, dass die Genauigkeit der Anfrage maßgeblich zum Ergebnis der Ausgabe beiträgt. Mehrfache Iterationen mit detaillierteren Prompts sollten daher in Betracht gezogen werden. Vor allem für wissenschaftliche Publikationen muss die KI-Nutzung abschließend weiter kritisch bewertet werden, während ihr Vorteil bei Übersetzungen und Editierung langer Textpassagen nicht von der Hand zu weisen ist.

Fazit

KI und mobile Applikationen bieten ein enormes Potenzial für die Anästhesiologie, um Patient*Innenversorgung und Sicherheitsstandards zu verbessern. Die Integration dieser Technologien erfordert jedoch eine sorgfältige Validierung, Datenschutzmaßnahmen und gezielte Schulungsprogramme. In den kommenden Jahren wird die Kombination aus Big Data, KI und mobilen Anwendungen auch die Art und Weise, wie Anästhesist*Innen arbeiten, er-

heblich verändern. Eine frühzeitige und differenzierte Auseinandersetzung mit den teilweise sehr komplexen Themenfeldern ist hierbei essenziell, um auch zukünftig sowohl im Bereich der Patient*Innenversorgung als auch der wissenschaftlichen Errungenschaften national und international wettbewerbsfähig bleiben zu können.

Literatur

1. Xie BH, Li TT, Ma FT, Li QJ, Xiao QX, Xiong LL, et al: Artificial intelligence in anesthesiology: a bibliometric analysis. *Perioper Med* 2024;13:121
2. Luckscheiter, Zink W, Thiel M, Schneider-Lindner V: Maschinelles Lernen in der Anästhesiologie – Anwendungen, Entwicklungsprozess und Ausblick. *Anästh Intensivmed* 2024;65:466–478
3. Shimada K, Inokuchi R, Ohigashi T, Iwagami M, Tanaka M, Goshō M, et al: Artificial intelligence-assisted interventions for perioperative anesthetic management: a systematic review and meta-analysis. *BMC Anesthesiol* 2024;24:306
4. Tavolara TE, Gurcan MN, Segal S, Niazi MKK: Identification of difficult to intubate patients from frontal face images using an ensemble of deep learning models. *Comput Biol Med* 2021;136:104737

Tabelle 3
Textausgaben, Quellen und Rechenzeit verschiedener frei zugänglicher Large-Language-Modelle (LLM) für einfache und komplexe Anfragen. Da nicht alle Modelle für deutsche Prompts ausgelegt sind, wurden die Anfragen („Erläutere das Konzept der Spinalanästhesie und verwende wissenschaftliche Referenzen“ bzw. „Plane eine Anästhesieeinleitung bei einem bekannten schwierigen Atemweg; gebe wissenschaftliche Referenzen für die Entscheidungen an“) auf Englisch formuliert und allen GPTs gleich zur Verfügung gestellt. Die resultierenden Halluzinationen sind zum größten Teil fehlverlinkte oder nicht existente Artikel, die vermittelten Informationen der Ausgaben waren jedoch in allen Fällen ohne grobe inhaltliche Fehler. Allgemeine Bezüge zu Abstracts realer Artikel wurden nicht als Halluzination gewertet, stellen aber einen erheblichen Anteil der angegebenen Quellen dar.

Prompt	„Explain the concept of spinal anesthesia to me and use scientific references.“				„Plan an anesthetic induction for a patient with a known difficult airway. Provide scientific references for your choices.“			
LLM	Wörter	Quellen	Halluzinationen	Zeit	Wörter	Quellen	Halluzinationen	Zeit
DeepSeek-R1	277	3	1	45 s	675	4	2	50 s
GPT-4	311	5	0	22 s	421	4	0	25 s
GPT-4o	304	15	4	16 s	360	7	0	78 s
GPT-4.5	281	5	0	43 s	288	5	0	44 s
GPT-4.5 Deep	1.404	4	0	442 s	4.458	23	0	2.225 s
Scholar GPT	265	9	0	117 s	490	8	0	80 s
Gemini Flash 2.0	364	0	0	2 s	582	2	0	3 s
Qwen 2.5	517	5	4	4 s	752	11	3	2 s
Grok 3	297	5	0	14 s	539	7	0	19 s

Medical Education

Review Articles

5. Jeon B, Jeong B, Jee S, Huang Y, Kim Y, Park GH, et al: A Facial Recognition Mobile App for Patient Safety and Biometric Identification: Design, Development, and Validation. *JMIR Mhealth Uhealth*. 2019;7:e11472
6. Aagaard N, Aasvang EK, Meyhoff CS: Discrepancies between Promised and Actual AI Capabilities in the Continuous Vital Sign Monitoring of In-Hospital Patients: A Review of the Current Evidence. *Sensors* 2024;24:6497
7. Chen M, He Y, Yang Z: A Deep Learning Framework for Anesthesia Depth Prediction from Drug Infusion History. *Sensors* 2023;23:8994
8. Rellum SR, Kho E, Schenk J, van der Ster BJP, Vlaar APJ, Veelo DP: A comparison between invasive and noninvasive measurement of the Hypotension Prediction Index: A post hoc analysis of a prospective cohort study. *Eur J Anaesth* 2025;42:131–139
9. Massari D, de Keijzer IN, Vos JJ: Comparing the Hypotension Prediction Index to Mean Arterial Pressure and Linear Extrapolated Mean Arterial Pressure for the Prediction of Intraoperative Hypotension: A Secondary Analysis. *Anesthesiology* 2024;141:1200–1202
10. Shimizu Y, Saeki N, Ohshimo S, Doi M, Oue K, Yoshida M, et al: Usefulness of new acoustic respiratory sound monitoring with artificial intelligence for upper airway assessment in obese patients during monitored anesthesia care. *J Med Invest* 2023;70:430–435
11. Wang D, Li J, Sun Y, Ding X, Zhang X, Liu S, et al: A Machine Learning Model for Accurate Prediction of Sepsis in ICU Patients. *Front Public Health* 2021;9:754348
12. Seymour CW, Kennedy JN, Wang S, Chang CH, Elliott CF, Xu Z, et al: Derivation, Validation, and Potential Treatment Implications of Novel Clinical Phenotypes for Sepsis. *JAMA* 2019;321:2003–2017
13. Schmierer T, Li T, Li Y: Harnessing machine learning for EEG signal analysis: Innovations in depth of anaesthesia assessment. *Artif Intell Med* 2024;151:102869
14. Jian Z, Liu X, Kouz K, Settels JJ, Davies S, Scheeren TWL, et al: Deep learning model to identify and validate hypotension endotypes in surgical and critically ill patients. *Br J Anaesth* 2025;134:308–316
15. Chiumello D, Coppola S, Catozzi G, Danzo F, Santus P, Radovanovic D: Lung Imaging and Artificial Intelligence in ARDS. *J Clin Med* 2024;13:305
16. Cascella M, Shariff MN, Lo Bianco G, Monaco F, Gargano F, Simonini A, et al: Employing the Artificial Intelligence Object Detection Tool YOLOv8 for Real-Time Pain Detection: A Feasibility Study. *J Pain Res* 2024;17:3681–3696
17. Gondode P, Duggal S, Mahor V: Artificial intelligence hallucinations in anaesthesia: Causes, consequences and countermeasures. *Indian J Anaesth* 2024;68:658–661
18. Fubini PE, Savoldelli GL, Beckmann TS, Samer CF, Suppan M: Impact of a Mobile App (LoAD Calc) on the Calculation of Maximum Safe Doses of Local Anesthetics: Protocol for a Randomized Controlled Trial. *JMIR Res Protoc* 2024;13:e53679
19. Rückert F, Truxa V, Dussmann P, Schmidt T, Seyfried T: Behandlungsqualität mit und ohne elektronische Gedächtnis- und Entscheidungshilfe (eGENA) – Die DANGER-Pilotstudie Teil 2 bei randomisierten kontrollierten „In-situ“-Notfallsimulationen. *Anesthesiologie* 2025;74:15–23
20. Nordstoga AL, Aasdahl L, Sandal LF, Dalager T, Kongsvold A, Mork PJ, et al: The Role of Pain Duration and Pain Intensity on the Effectiveness of App-Delivered Self-Management for Low Back Pain (selfBACK): Secondary Analysis of a Randomized Controlled Trial. *JMIR Mhealth Uhealth* 2023;11:e40422
21. Lang EV, Jackson W, Senn P, Aroni DKK, Finkelman MD, Corino TA, et al: Efficacy of a Self-Hypnotic Relaxation App on Pain and Anxiety in a Randomized Clinical Trial: Results and Considerations on the Design of Active and Control Apps. *Int J Clin Exp Hypn* 2021;69:277–295
22. Dunsmuir D, Wu H, Sun T, West NC, Lauder GR, Görges M, et al: A Postoperative Pain Management Mobile App (Panda) for Children at Home After Discharge: Usability and Feasibility. *JMIR Perioper Med* 2019;2:e12305
23. Geuens S, Lemiere J, Nijs J, Treunen M, Aertsen M, Toelen J, et al: Testing a Home Solution for Preparing Young Children for an Awake MRI: A Promising Smartphone Application. *Children* 2023;10:1866
24. Jones A, Tang R, Dabo-Trubelja A, Yeoh CB, Richards L, Gottumukkala V: Optimising artificial intelligence ultrasound tools in anaesthesiology and perioperative medicine: The next frontier for advanced technology application. *Indian J Anaesth* 2024;68:1016–1021
25. Kabon M: Fachtagung: Medizinisch-pflegerische Versorgung – Innovative Lösungsansätze statt Mangelverwaltung. *Ärzteblatt RLP* 12/2024;32–35
26. Bundesbeauftragte für den Datenschutz und die Informationsfreiheit (BfDI): „Whitepaper“ der DSK zu technischen Datenschutzanforderungen an Messenger-Dienste im Krankenhausbereich 2019. https://www.bfdi.bund.de/SharedDocs/Downloads/DE/DSK/DSKBeschluessePositionsspa-piere/98DSK_Whitepaper-Messenger-Krankenhausbereich.html (Zugriffsdatum: 09.05.2025)
27. gematik: TI-Messenger Anwendungen 2025. <https://www.gematik.de/anwendungen/ti-messenger> (Zugriffsdatum: 09.05.2025)
28. Abdellatif A, Badran K, Costa DE, Shihab E: A comparison of natural language understanding platforms for chatbots in software engineering. *IEEE Trans Softw Eng* 2021;48:3087–3102
29. Vaswani A, Shazeer N, Parmar N, Uszkoreit J, Jones L, Gomez AN, et al: Attention is all you need. *Advances in neural information processing systems* 2017;30:1–13
30. Brown T, Mann B, Ryder N, Subbiah M, Kaplan JD, Dhariwal P, et al: Language models are few-shot learners. *Advances in neural information processing systems* 2020;33:1877–1901
31. Kaplan J, McCandlish S, Henighan T, Brown TB, Chess B, Child R, et al: Scaling laws for neural language models 2020;arXiv preprint arXiv:2001.08361.

Korrespondenz-
adresse

**Priv.-Doz. Dr.
med. habil.
Robert Rümmler,
DESAIC, MBA**



Klinik für Anästhesiologie und
Intensivmedizin
Universitätsmedizin der Johannes
Gutenberg-Universität
Langenbeckstraße 1
55131 Mainz, Deutschland

Tel.: 06131-17 6755

Fax: 06131-17 5599

E-Mail: robert.ruemmler@
unimedizin-mainz.de

ORCID-ID: 0000-0002-9265-6052