



Künstliche Intelligenz in chirurgischen Disziplinen: Einsatz, Nutzen und Potenzial – ein Delphi-Expertenkonsensus

G. Duwe¹ · K. Moench¹ · V. Kauth¹ · M. Angeloni² · J. Eckhoff^{3,15} · M. Görtz^{4,16} · S. Hoefert⁵ · T. D. Kocar⁶ · L. Kollitsch⁷ · S. Mehravand⁸ · D. Mercier⁹ · J. Rudolph¹⁰ · J. Rueckel¹¹ · R. Schönhof⁶ · M. Sondermann⁸ · CAJ von Klot¹² · A. Zamzow¹³ · J. P. Struck^{14,17} · H. Borgmann^{14,17}

¹ Klinik und Poliklinik für Urologie und Kinderurologie, Universitätsmedizin der Johannes Gutenberg-Universität Mainz, Mainz, Deutschland; ² Pathologisches Institut, Universitätsklinikum der Friedrich-Alexander-Universität Erlangen, Erlangen, Deutschland; ³ Klinik für Allgemein-, Viszeral-, Thorax- und Transplantationschirurgie, Universitätsklinikum Köln, Köln, Deutschland; ⁴ Nachwuchs-Klinische Kooperationsseinheit „Multiparametrische Methoden zur Früherkennung von Prostatakrebs“, Deutsches Krebsforschungszentrum (DKFZ), Heidelberg, Deutschland; ⁵ Klinik und Poliklinik für Mund-, Kiefer- und Gesichtschirurgie, Universitätsklinikum Tübingen, Tübingen, Deutschland; ⁶ Institut für Geriatrie Forschung, Universitätsklinikum Ulm, Ulm, Deutschland; ⁷ Abteilung für Urologie und Andrologie, Klinik Donaustadt, Wien, Österreich; ⁸ Klinik und Poliklinik für Urologie, Universitätsklinikum Carl Gustav Carus Dresden, Dresden, Deutschland; ⁹ Deutsches Forschungszentrum für Künstliche Intelligenz (DFKI), Kaiserslautern, Deutschland; ¹⁰ Klinik und Poliklinik für Radiologie, LMU Klinikum, LMU München, München, Deutschland; ¹¹ Institut für Diagnostische und Interventionelle Neuroradiologie, LMU Klinikum, LMU München, München, Deutschland; ¹² Klinik für Urologie und Urologische Onkologie, Medizinische Hochschule Hannover, Hannover, Deutschland; ¹³ Institut für Medizinische Lehre und Ausbildungsforschung, Julius-Maximilians-Universität Würzburg, Würzburg, Deutschland; ¹⁴ Klinik für Urologie und Kinderurologie, Universitätsklinikum Brandenburg an der Havel, Brandenburg an der Havel, Deutschland; ¹⁵ Surgical Artificial Intelligence & Innovation Laboratory (SAIIL), Massachusetts General Hospital, Boston, USA; ¹⁶ Urologische Klinik, Universitätsklinikum Heidelberg, Heidelberg, Deutschland; ¹⁷ Fakultät für Gesundheitswissenschaften Brandenburg, Medizinische Hochschule Brandenburg Theodor Fontane, Brandenburg, Deutschland

Gregor Duwe (Konsensusgruppenleiter 1) und Kerstin Moench (Konsensusgruppenleiterin 2) haben gleichermaßen als Erstautoren beigetragen. Julian Struck und Hendrik Borgmann haben gleichermaßen als Letztautoren beigetragen.

Anmerkung: Zur besseren Lesbarkeit wird in dieser Publikation das generische Maskulinum verwendet. Die in dieser Publikation verwendeten Personenbezeichnungen beziehen sich auf beide biologische Geschlechter.



QR-Code scannen & Beitrag online lesen

Künstliche Intelligenz (KI) ist aus allen Bereichen des chirurgischen Alltags nicht mehr wegzudenken. Um unsere Patienten bestmöglich zu versorgen und als deutsches Gesundheitswesen international wettbewerbsfähig zu bleiben, gilt es, die Möglichkeiten und Herausforderungen zu benennen und in interdisziplinären Lösungen umzusetzen. 16 Experten beschreiben die aktuelle Situation, Zukunftswünsche sowie Ideen für den Weg in die digitale Transformation.

Hintergrund

Die technischen Entwicklungen im Bereich der KI zeigen das Potenzial für einen Paradigmenwechsel in der chirurgischen Medizin auf allen Ebenen der ärztlichen Tätig-

keiten [1]. Im Zuge einer chirurgischen Behandlung könnten KI-Anwendungen theoretisch alle Behandlungsschritte unterstützen: beginnend mit der radiologischen und pathologischen Diagnostik, Patientenaufklärung und edukativer Begleitung während der Behandlung, über eine intraoperative Unterstützung, z. B. durch autonome roboterassistierte Arbeitsschritte bis hin zur postoperativen Überwachung und Vorhersage von Behandlungsverläufen. Ziel aller Anwendungsbereiche ist es, die Behandlungsqualität und damit die Patientensicherheit zu verbessern.

Aufgrund der rasanten Entwicklung im Bereich der KI ist eine frühzeitige, strukturierte Zusammenarbeit zwischen den medizinischen, ärztlichen Anwendern, der Politik für die Verantwortung der rechtlichen Rahmenbedingungen und gesellschaftli-

chen Zielsetzungen sowie der freien Wirtschaft als Innovations- und Transformatorenstreiber notwendig. Hierfür haben wir als interdisziplinäre Expertengruppe für KI-Anwendungen in der Chirurgie im Rahmen des 2. Digital Health Summit am 30.08.2024 in Brandenburg an der Havel eine Konsenskonferenz nach Delphi-Maßgaben umgesetzt, um einen Status quo zur Realität sowie die notwendigen wissenschaftlichen und technischen Transformationsschritte, die eine skizzierte Utopie ermöglichen könnten, zu definieren [2]. Hierzu wurden neben den medizinischen Fragestellungen und Anforderungen insbesondere auch Aspekte zu der notwendigen „KI-Infrastruktur“ im Bereich Datenschutz, IT-Sicherheit und klinischer IT-Infrastruktur im Sinne der Digitalisierung diskutiert.

Mit den Ergebnissen dieser Konsensuskonferenz möchten wir allen beteiligten Akteuren in chirurgischen Disziplinen (und verwandten medizinischen Bereichen), Politik und Wirtschaft einen interdisziplinären, expertenvalidierten Überblick über den aktuellen Stand sowie die notwendigen Transformationsprozesse geben, um KI zukünftig in der chirurgischen Behandlung auf allen Ebenen erfolgreich und sinnvoll zu etablieren. Hierbei werden auch insbesondere berufsrechtliche und ethische Limitationen aufgezeigt, um aus ärztlicher Perspektive zielführende wissenschaftliche, technische und finanzielle Investitionen vorzuschlagen.

Methodik

Wir führten ein modifiziertes Delphi-Verfahren mit 16 interdisziplinären Forschern zum Themenkomplex ‚Künstliche Intelligenz in chirurgischen Disziplinen‘ durch. Die Zusammenstellung des Expertenpanels erfolgte durch E-Mail-Einladungen an Erst-/Letztautoren aus Deutschland und Österreich von Forschungsarbeiten auf PubMed zum Themenfeld. Die interdisziplinäre Zusammensetzung des Delphi-Panels erfolgte nach E-Mail-Einladungen an Autorinnen und Autoren relevanter Forschungspublikationen der vergangenen 5 Jahre zum Themenfeld KI in der Medizin, die zuvor durch die Leitung des 2. Digital Health Summit identifiziert worden waren. Das Panel umfasste Experten aus klinisch-

Hintergrund: Künstliche Intelligenz (KI) in der chirurgischen Medizin hat das Potenzial, alle Behandlungsphasen im Sinne einer verbesserten Behandlungsqualität und Patientensicherheit zu unterstützen. Eine Expertengruppe hat im Rahmen einer Konsenskonferenz beim 2. Digital Health Summit in Brandenburg im August 2024 den aktuellen Stand und die erforderlichen Schritte zur erfolgreichen Integration von KI in chirurgische Fächer diskutiert.

Methodik: Es wurde ein modifiziertes Delphi-Verfahren mit 16 interdisziplinären Forschern zum Themenkomplex KI in chirurgischen Disziplinen durchgeführt. In zwei Online-Meetings mit anschließenden Delphi-Umfragerunden (LimeSurvey) und einem abschließenden Hybrid-Meeting wurden von allen 16 Teilnehmern individuelle Stellungnahmen erstellt, gemeinsam diskutiert und in Anlehnung an das AWMF-Regelwerk Leitlinien konsentiert.

Ergebnisse: Aus 103 eingereichten Stellungnahmen wurden 36 Statements zu Realität ($n=12$), Utopie ($n=13$) und Möglichkeiten der digitalen Transformation ($n=11$) nach Diskussion und Modifikation konsentiert. Wir erzielten einen Konsens von mindestens 75 % in allen hier aufgeführten Stellungnahmen, 6 der Stellungnahmen erreichten mit 100 % Zustimmung starken Konsens.

Schlussfolgerung: Die konsentierten Stellungnahmen zeigen das große Potenzial von KI für die Verbesserung in allen Bereichen der Patientenversorgung in der chirurgischen Medizin. Herausforderungen wie fehlende Digitalisierungsstrukturen und rechtliche Rahmenbedingungen wurden identifiziert und praxisnahe Vorschläge zur Umsetzung entwickelt. Die Notwendigkeit einer interdisziplinären Zusammenarbeit zwischen Medizin, Politik und Wirtschaft wird betont, um das deutsche Gesundheitswesen national wie international wettbewerbsfähig und zukunftssicher zu gestalten.

Schlüsselwörter

Künstliche Intelligenz · Chirurgie · Delphi-Konsensuskonferenz · Rechtliche Rahmenbedingungen · Digitalisierung

therapeutischen, diagnostischen und forschenden Bereichen – einschließlich Lehr- und Ausbildungsforschung – verschiedener deutscher Universitätskliniken, eines österreichischen Schwerpunktkrankenhauses sowie des Deutschen Forschungszentrums für Künstliche Intelligenz. Mit Ausnahme der Letztautoren sind sämtliche Mitglieder des Delphi-Panels zugleich Autorinnen und Autoren dieser Publikation. Die wissenschaftliche Methodik wurde in Zusammenarbeit mit dem Zentrum für Wissenstransfer „UroEvidence“ der Deutschen Gesellschaft für Urologie e.V. konzipiert. Die Delphi-Methode wurde in den 1950er-Jahren entwickelt und ist ein strukturiertes, mehrstufiges Befragungsverfahren zur Konsensbildung. Sie basiert auf der moderierten Sammlung anonymisierter Einschätzungen eines Expertenpanels [3].

Die Methodik des modifizierten Delphi-Verfahrens wurde im Voraus durch ein schriftliches Protokoll festgelegt. Alle eingeladenen Experten gaben ihre Interessenkonflikte offen an. Den Experten wurde ein Projektexposé mit allen relevanten Hintergrundinformationen bereitgestellt. Das

Protokoll umfasste außerdem eine umfassende narrative Literaturrecherche zum Einsatz, Nutzen und Potenzial von KI in chirurgischen Disziplinen.

Es fanden insgesamt 2 Online-Meetings sowie 1 abschließendes Hybrid-Meeting (Präsenz + Online) mit den eingeladenen Experten statt. Im ersten Online-Meeting (via Microsoft Teams) erfolgte die Projektvorstellung sowie die Sammlung von Kernfragen und Argumenten. Nachfolgend entwickelten die Konsensusgruppenteilnehmer nach Literaturrecherche passende Konsensusstatements. Diese wurden durch die Konsensusgruppenleitung gebündelt und in die Bereiche Realität, Utopie und Transformation kategorisiert sowie thematisch gruppiert. Nachfolgend wurde die Konsensusumfrage auf der Plattform LimeSurvey an die Experten zur Abstimmung versandt. Der Grad der Zustimmung/Ablehnung zum Konsensusstatement erfolgte auf einer 5-Punkte-Likert-Skala („stimme voll und ganz zu; stimme eher zu, stimme weder zu noch nicht zu, stimme eher nicht zu, stimme gar nicht zu“). Im zweiten Online-Meeting erfolgte die Vorstellung und Dis-

kussion der bisherigen Konsensusergebnisse. Die Konsensusstatements wurden hier abhängig vom Zustimmungsgrad und der Diskussion belassen, geschärft, verändert oder verworfen. Nachfolgend wurde eine zweite Konsensusumfrage auf der Plattform LimeSurvey an die Experten zur Abstimmung versandt. Beim abschließenden Hybrid-Meeting wurden die Konsensusstatements dann final diskutiert, ggfs. adjustiert und final abgestimmt.

Der Grad der Konsensstärke in Abhängigkeit der Zustimmung der Experten zu den Konsensusstatements wurde in Anlehnung an das AWMF-Regelwerk Leitlinien wie folgt definiert:

- **starker Konsens:** Zustimmung von > 95 % der Teilnehmenden,
- **Konsens:** Zustimmung von > 75–95 % der Teilnehmenden,
- **mehrheitliche Zustimmung:** Zustimmung von > 50–75 % der Teilnehmenden,
- **keine mehrheitliche Zustimmung:** Zustimmung von ≤ 50 % der Teilnehmenden.

„Starker Konsens“ (> 95 % der Teilnehmenden) wurde aufgrund der Zusammensetzung der Gruppe aus 16 Experten nur dann erreicht, wenn alle 16 (100 %) eine Zustimmung angegeben haben. 15 Zustimmungen entsprechen hier 94 % und damit formal einem „Konsens“.

Ergebnisse

Die erarbeiteten und konsentierten Statements beschäftigen sich zunächst mit der Realität (Status quo) der aktuellen KI-Nutzung in chirurgischen und angeschlossenen Disziplinen, definieren dann eine Utopie (maximale Potenzilausschöpfung) und beschreiben schließlich die aktuellen Möglichkeiten zur digitalen Transformation (vom Status quo zur maximalen Potenzilausschöpfung). Viele relevante Bereiche der chirurgischen Tätigkeiten werden thematisiert, jedoch besteht kein Anspruch auf Vollständigkeit.

Realität (Status quo)

Trotz rasanter wissenschaftlicher Weiterentwicklung von KI-Modellen und Erweiterung der Anwendungsgebiete mit

Fokus auf chirurgische Sicherheit fehlen aktuell in Deutschland und europaweit essentielle strukturelle Voraussetzungen für die klinische Translation (technische, organisatorische, wirtschaftliche und regulatorische Voraussetzungen, sowie Akzeptanz- und Ausbildungsbarrieren). (KONSENS)

Der begrenzte Zugang zu multizentrischen, umfassenden Datensätzen und die fehlende Interoperabilität zwischen chirurgischen Systemen erschweren den Datenaustausch und die Integration von KI-basierten Anwendungen, z.B. im Operationsaal. KI-Modelle müssen daher häufig an Hand von monozentrischen, spezifischen und weniger repräsentativen Datensätzen trainiert werden, was ihre Fähigkeit einschränkt, die klinische und demografische Vielfalt abzubilden [4, 5]. Die fehlende regulatorische Klarheit und unzureichende technische Infrastruktur zur nahtlosen Datenerfassung und -verarbeitung verhindern den Einsatz von KI, insbesondere bei zeitkritischen Entscheidungen im intraoperativen Setting.

Für die nationale klinische Anwendung von KI-basierten Assistenzsystemen stellen die rechtlichen Verwaltungsstrukturen des Föderalismus für die flächendeckende Umsetzung von Vorgaben des Datenschutzes (mit einzelnen Datenschutzbeauftragten), der Informationssicherheit sowie der noch sehr heterogenen Digitalisierungsstandards eine aktuell komplexe politische, strukturelle Herausforderung dar. (STARKER KONSENS)

Der aktuelle Einsatz von KI-Systemen im deutschen Gesundheitswesen erfolgt aktuell v. a. anhand einzelner Projekte in Kliniken bzw. Klinikverbunden. Somit muss jedes Krankenhaus eigenverantwortlich mit dem IT-Abteilungen sowie Datenschutzbeauftragten gesetzeskonforme und umsetzbare digitale Prozesse erarbeiten, die anschließend KI-Applikationen anwenden können. Für einen zukünftigen, effizienten, klinik- und länderübergreifenden Einsatz und Austausch von Daten für erfolgreiche KI-Systeme sollten in der Politik bundeseinheitliche Maßnahmen und Vorgaben erarbeitet werden, um diese bestehenden gesetzlichen Limitation zu verbessern [6–8].

Die klinische Einführung von KI-Tools im medizinischen Bereich wird v. a. durch

folgende Faktoren beeinflusst und aktuell verzögert: das Fehlen einer prospektiven klinischen Validierung, die Notwendigkeit einer behördlichen Genehmigung und das Fehlen einer angemessenen institutionellen Infrastruktur für die systematische Datenerfassung und -verarbeitung. (KONSENS)

a) Verwendungszweck sowie analytische und klinische Validierung, b) Datenqualität und c) Privatsphäre und Datenschutz gehören zu den sechs zentralen Themenbereichen, die von der WHO für eine sichere und schnelle Nutzung von KI-Systemen im Gesundheitswesen angesprochen werden [9]. Vor allem die Erhebung vielfältiger und für die Zielpopulation repräsentativer Daten ist für die Gewährleistung der „algorithmischen Sicherheit“ unerlässlich [10]. Außerdem sollte das Paradigma des „federated learning“ in Betracht gezogen werden, um den Datenschutz und Fragen der gemeinsamen Nutzung von Daten zu wahren und gleichzeitig die Entwicklung robuster, institutionenübergreifender Modelle zu gewährleisten [11].

Der demografische Wandel führt durch die alternde Bevölkerung zu einem steigenden Bedarf an Gesundheitsleistungen, den das Gesundheitssystem ohne den Einsatz von KI kaum bewältigen kann. (KONSENS)

Der Einsatz von KI im Gesundheitswesen birgt vielversprechendes Potenzial, wirft aber auch ethische Bedenken und Herausforderungen auf. Die Integration von KI-Lösungen in die Gesundheitsversorgung älterer Menschen erfordert daher eine sorgfältige Planung und die Berücksichtigung von Aspekten wie externer Validierung, Transparenz und Schutz vor Fehlern. Um den individuellen Bedürfnissen älterer Menschen angemessen Rechnung zu tragen, ist ein personenzentrierter Ansatz unerlässlich [12].

Die technischen Operationspunkte von Algorithmen zur Detektion pathologischer Befunde in radiologischen Bildern balancieren eine häufig priorisierte hohe Sensitivität mit der im Rahmen dessen noch erreichbaren Spezifitäten und klinisch-prädiktiven Werten [13–15]. Zwar eignen sich hochsensitive Algorithmen als „second reader“, um die Anzahl übersehener Befunde zu reduzieren

[16–18]. Jedoch bergen hochsensitive Algorithmen mit resultierend auch falsch-positiven Detektionen in Kombination mit einem möglichen „confirmation bias“ klinischer Anwender die Gefahr, dass Patienten unnötigerweise einer erweiterten Diagnostik und Therapie zugeführt werden [14, 19, 20]. (KONSENS)

Bereits in den 1980er-/1990er-Jahren führten „Computer-aided-design“ (CAD)-Systeme in der Mammographie zu kritischen Diskussionen über unnötige Abklärungen [19, 20] durch (zu) hohe Sensitivitäten und limitierte prädiktive Werte [13, 14, 16–18]. Ähnliche Herausforderungen bringen auch Anwendungen modernerer KI-Algorithmen mit sich [15].

Der Einsatz von KI-basierten Chatbots zur Beantwortung von Patientenfragen, z. B. bei der präoperativen Schulung oder in der postoperativen Nachsorge, beschränkt sich auf wenige Studien. Ebenso ist die KI-gestützte Überwachung postoperativer Fortschritte sowie die frühzeitige Erkennung von Komplikationen nicht Teil der klinischen Routine. (STAR-KER KONSENS)

Studienergebnisse, wie z. B. bei der Wissensvermittlung in der Prostatakrebsdiagnostik, zeigen, dass Chatbots von Patienten genutzt und erfolgreich klinisch eingesetzt werden können [21]. Obwohl vielversprechende Studienergebnisse zu beobachten sind, erfordert die Integration in klinische Arbeitsabläufe weitere Forschung, um ethische, technische und regulatorische Herausforderungen zu bewältigen. KI-basierte Chatbots können Ärzte und Patienten unterstützen, ersetzen jedoch nicht die zwischenmenschliche Arzt-Patienten-Beziehung.

Die KI zur intraoperativen Qualitätssicherung und Verbesserung der Patientensicherheit beschränkt sich derzeit auf wissenschaftliche retrospektive Analysen von operativen Videos, zur Segmentierung von anatomischen Strukturen, Phasenerkennung und Bewertung chirurgischer Fähigkeiten. (KONSENS)

Obwohl KI-Anwendungen zunehmend in der präoperativen Diagnostik, Therapieplanung, Risikostratifizierung und postoperativen Verlaufskontrolle eingesetzt werden, bleibt ihr Potenzial für intraoperative Qualitätssicherung und Patientensicherheit bislang weitgehend

ungenutzt. Aktuelle, überwiegend wissenschaftlich-experimentelle Ansätze zielen primär auf ein automatisiertes Verständnis intraoperativer Abläufe sowie auf die retrospektive Erkennung und Klassifikation räumlicher und zeitlicher Aspekte im endoskopischen Video ab (z. B. Erkennung operativer Phasen, Vorhersage nächster Schritte oder Einordnung von Instrument-Gewebe-Interaktionen) sowie die Bewertung chirurgischer Fähigkeiten [22–27].

Das Training von KI-System in der Medizin erfordert aktuell große Mengen an Gesundheitsdaten. Gesundheitsdaten werden unter Einsatz einer doppelten Pseudonymisierung durch eine Vertrauensstelle von potenziell identifizierbaren Informationen getrennt. (KONSENS)

Mit dem am 26.03.2024 in Kraft getretenen Gesundheitsdatennutzungsgesetz (GDNG) besteht der rechtliche Rahmen, medizinische Daten in breiterer Anwendung nutzbar zu machen. Zum aktuellen Zeitpunkt fehlt aber die Infrastruktur, diese Gesundheitsdaten anzufordern, da sich das neue Forschungsdatenzentrum Gesundheit der gesetzlichen Ombudsbehörde, dem Bundesinstitut für Arzneimittel und Medizinprodukte, noch im Aufbau befindet. Anträge werden voraussichtlich ab Anfang 2025 bearbeitet werden können [28–30].

Zum aktuellen Zeitpunkt bestehen keinerlei gesetzliche Haftungs- und Urheberrechte für die Verwendung von medizinischen, KI-generierten Inhalten. Damit Inhalte urheberrechtlich geschützt sind, müssen sie gemäß § 2 Abs. 2 UrhG als „Werk“ gelten, was eine „persönliche geistige Schöpfung“ voraussetzt. Diese Schöpfung muss auf menschlicher kreativer Tätigkeit beruhen, um als „persönlich“ zu gelten. KI-erzeugte Inhalte gelten in der Regel nicht als urheberrechtlich geschützt, da keine „persönliche geistige Schöpfung“ vorliegt. (KONSENS)

Um Verantwortung übernehmen zu können, muss ein Subjekt im traditionellen Sinne handlungsfähig, also in der Lage sein, Entscheidungen zu treffen. KI erfüllen diese Voraussetzungen nicht, weshalb Verantwortung für deren Verhalten nicht auf die KI selbst übertragen werden kann [31].

Zu jedem Zeitpunkt unterliegt die Verantwortung der Diagnostik, Diagnosestellung und Therapie, auch bei Verwendung von KI-gestützten Systemen, den behandelnden Ärzten. (KONSENS)

Für Schnittstellen zwischen Menschen und KI besteht das Risiko einer Verantwortungsdiffusion. Bei der Zuweisung von Verantwortung tritt an die Stelle des Arzt-Patienten-Verhältnisses ein Zusammenspiel von Herstellern, Programmierern, Regulierungsbehörden, medizinischen Anwendenden und Patienten. KI kann selbst nicht für Fehler zur Verantwortung gezogen werden [31]. Akteure sind in der Pflicht, die Verantwortung eindeutig zuzuweisen [32].

Die aktuellen medizinischen Curricula sind nicht unter Berücksichtigung der hohen Wahrscheinlichkeit, dass in den kommenden Jahrzehnten vermehrt KI-basierte Assistenzsysteme Einzug in die klinische Anwendung halten werden, verfasst worden. (KONSENS)

Die Inhalte des Medizinstudiums orientieren sich einerseits an den Erfahrungen der bisherigen Lehre, sollen aber gleichzeitig bereits frühzeitig sich abzeichnende Trends integrieren [33, 34]. Die aktuell für 2027 geplante Novelle der Approbationsordnung stellt eine günstige Gelegenheit dar, es den Studierenden zu ermöglichen, bereits im Rahmen ihrer Ausbildung zu lernen, die Entscheidungsprozesse von KI-Modellen in ihren Grundzügen nachzuvollziehen und dadurch ein Verständnis für Potenzial und Begrenzungen bestimmter Methoden zu entwickeln.

Aktuell gibt es keine speziell für medizinische Anwendungen trainierte KI-Systeme, die effektiv und zuverlässig für den Wissenserwerb in der medizinischen Ausbildung empfohlen werden können. Die Entwicklung solcher Systeme steht noch am Anfang und bedarf weiterer Forschung und Validierung, bevor sie in der medizinischen Lehre flächendeckend eingesetzt werden können. (KONSENS)

Obwohl „large language models“ beeindruckende Ergebnisse bei der Beantwortung medizinischer Fragestellungen erbringen, mangelt es aktuell an medizinspezifischem Training, Präzision und Zuverlässigkeit, um sie als vertrauenswürdige Werkzeuge für den Wissenserwerb zu nutzen [35–42]. Sorgfältige Planung, rigorose Testungen, fortlaufende Updates,

aktive Wartungen, ethische Überlegungen und Datenschutz sind von entscheidender Bedeutung für die Entwicklung solcher Systeme.

Utopie (maximale Potenzialerschöpfung)

Der Datenzugriff, die Datenverarbeitung, der Daten-Output sowie die Speicherung des KI-generierten Inhalts unterliegt nicht der Verantwortung der Ärzte. Die Nutzung und Anwendung dessen für Klinik, Forschung und Lehre jedoch unterliegt der Verantwortung der Ärzte. (KONSENS)

Hersteller müssen sicherstellen, dass KI-Systeme transparent und verantwortungsvoll genutzt werden können. Patienten sollten Entscheidungsunterstützungssysteme in Grundzügen verstehen können, um einen „informed consent“ zu ermöglichen. Ärzte sollen in der Lage sein, eine Risikobewertung bei der Nutzung eines KI-Systems vorzunehmen. Die moralische und rechtliche Letztverantwortung bei KI-gestützten Anwendungen verbleibt beim behandelnden Fachpersonal. Das Vertrauen in die Arzt-Patienten-Beziehung bleibt somit zentral für den Behandlungserfolg. Das Ziel des KI-Einsatzes in der Medizin soll nicht darin bestehen, medizinisches Fachpersonal zu ersetzen, sondern den Behandlungsprozess zu unterstützen [32].

Die Ergebnisse der im medizinischen Bereich eingesetzten KI-Tools sollten für Kliniker leicht interpretierbar sein, sie sollten plausible, logische Begründungen für das Ergebnis liefern (auch bei Unsicherheiten) und transparent sein. (KONSENS)

Die KI-Systeme sollten transparente Entscheidungsprozesse zeigen, um Vertrauen zu schaffen. Vor allem sollten KI-Tools nicht nur Vorhersagen, sondern auch Begründungen liefern, die mit dem medizinischen Wissen übereinstimmen. Plausibilität, gepaart mit Interpretierbarkeit, ist für Kliniker entscheidend, damit sie KI-Ergebnissen vertrauen und sie überprüfen können, insbesondere in kritischen Bereichen wie Diagnostik und Behandlungsplanung. Techniken wie Entscheidungsbäume, „attention maps“ in neuronalen Netzen und „feature importance scores“ können die Interpretierbarkeit verbessern. Die Verankerung von

Kompetenzen im Bereich der digitalen Gesundheit in der Aus- und Weiterbildung von Ärzten [43] schafft die Voraussetzungen für eine sachkundige Nutzung und Sensibilisierung für diese Technologien.

Die KI-Anwendungen können zu einer verbesserten Erkennung von Patienten führen, die von einer chirurgischen Therapie profitieren können. Gleichzeitig könnte durch eine Reduktion von unnötiger invasiver Diagnostik Schaden vom Patienten abgewendet werden. (KONSENS)

Die Früherkennung von Erkrankungen in chirurgischen Disziplinen kann Behandlungsergebnisse signifikant verbessern [44]. Allerdings ist die diagnostische Genauigkeit vieler Screening-Methoden limitiert. Moderne KI, insbesondere das „Deep Learning“, kann für spezifische Fragestellungen die menschliche Leistungsfähigkeit erreichen [45] und potenziell zu einer verbesserten Erkennung von Patienten führen, die von einer frühzeitigen Therapie profitieren können. Gleichzeitig könnte durch Reduktion von invasiver Diagnostik Schaden vom Patienten abgewendet werden.

Die KI-Systeme könnten zukünftig personalisierte und präzise Therapieempfehlungen liefern, die die Behandlungsergebnisse erheblich verbessern könnten. (KONSENS)

Die KI-Systeme haben das Potenzial, durch die Analyse großer Datenmengen präzise Therapieempfehlungen zu generieren, die individualisierte Behandlungspläne ermöglichen. Diese Systeme könnten klinische Daten, genetische Informationen und Patientenpräferenzen verarbeiten, um maßgeschneiderte Therapievor schläge zu bieten, die die Behandlungsergebnisse optimieren. Beispielsweise könnten sie Ärzten dabei helfen, spezifische Medikationen oder Therapien für einzelne Patienten basierend auf deren individuellen Risikoprofilen auszuwählen und so die Sicherheit der Behandlung zu erhöhen [46]. Studien zeigen, dass die Anwendung solcher Systeme in der Krebsbehandlung bereits zu maßgeschneiderten, multidrug-basierten Therapien geführt hat, die Überlebensraten verbessern konnten [47].

Die KI-gestützten Therapieunterstützungssysteme können klinische Abläufe zeitlich optimieren und einen Beitrag

zur optimierten Patientenversorgung leisten, da individuelle Erfahrungsstände des Personals ausgeglichen werden können, einrichtungsübergreifend Erfahrung genutzt werden kann und komplexe Zusammenhänge kompakter dargestellt werden können. (KONSENS)

Die KI-Methoden sind aufgrund ihrer Rechenleistung in der Lage, Entscheidungen basierend auf einer riesigen Anzahl an Faktoren zu treffen, wodurch sie Empfehlungen optimieren könnten [48]. Durch eine potenziell enorme Menge an medizinischen Daten sind sie in der Lage, personalisierte Vorhersagen zu treffen [49].

Klinische KI-Software soll barrierefrei mit synthetischen, frei zugänglichen und keine personenbezogenen Informationen enthaltenden Daten trainiert werden, die reale Daten in gleicher Weise abbilden. (KONSENS)

Die KI ist in der Lage, authentische synthetische Gesundheitsdaten zu erzeugen, die keine personenbezogenen Informationen enthalten [50–53]. Synthetische Daten können ohne datenschutzrechtliche Einschränkungen genutzt werden und bestehende Datensätze erweitern oder ersetzen. Synthetische Daten müssen wiederum initial von KI mithilfe von echten Daten („Big Data“) erzeugt werden [54]. Damit kann u. a. die Genauigkeit KI-basierter Diagnosesysteme verbessert werden und sie sind barrierefrei in der Klinik, Forschung und Lehre etc. einsetzbar, ohne auf Bundesdatenbanken angewiesen zu sein [28, 29].

Das partizipative ärztliche Aufklärungsgespräch kann durch KI-Systeme nicht ersetzt, sondern nur unterstützt werden. (KONSENS)

Die KI-Systeme sollten das partizipative ärztliche Aufklärungsgespräch nicht vollständigen ersetzen, obwohl die Systeme das Gespräch durch Informationsvermittlung zu Vorgehensweisen, prozeduralen Abläufen, Risikoeinschätzungen und Formalitäten unterstützen können [55]. Es fehlt die Fähigkeit individuelle Bedürfnisse, Fragen und Emotionen einzuschätzen und deren Konsequenzen in den Entscheidungsprozess einfließen zu lassen [56]. Somit bleibt das Aufklärungsgespräch eine ärztliche Kernkompetenz mit zwischenmenschlicher Interaktion und Empathie, welche nicht durch Technologie ersetzt werden sollte [57].

Zukünftig sollte KI, unter Verwendung aller in der robotischen Chirurgie verfügbaren Datenmodalitäten (z. B. Video und Kinematik), intraoperative Warnsysteme und automatisierte präventive Intervention ermöglichen. (KONSENS)

Die Vision zielt darauf ab, eine kontinuierliche, datengestützte Überwachung der chirurgischen Umgebung zu ermöglichen, wobei kritische Strukturen hervorgehoben und Bewegungsmuster analysiert werden, um chirurgische Fähigkeiten objektiv zu bewerten und intraoperative Entscheidungen zu verbessern. Insbesondere die robotische Chirurgie bietet hierfür hervorragende Grundvoraussetzungen, da nicht nur visuelle Informationen aufgezeichnet werden, sondern auch kinematische und haptische Daten analysiert werden können [58, 59] und damit intraoperative Warnsysteme und automatisierte präventive Interventionen ermöglicht werden [60, 61]. Zusammengefasst zeigt sich im erzielten Konsensus, dass sowohl aktuelle als auch zukünftige KI-Technologien das Potenzial haben, die robotische Chirurgie sicherer und effizienter zu gestalten.

Zukünftige KI-Anwendungen sollen intraoperative Echtzeitüberwachung und Bildverarbeitung zur automatischen Fehlererkennung und -vermeidung sowie detaillierte, effiziente Dokumentation ermöglichen. (STARKER KONSENS)

Der ideale Einsatz von KI für die intraoperative Qualitätssicherung und Verbesserung der Patientensicherheit wird in der automatisierten Erkennung von operativen Risiken und zeitgerechten Warnung bzw. Intervention zur rechtzeitigen Vermeidung von operativen Komplikationen gesehen. Solche Systeme könnten beispielsweise kritische anatomische Strukturen in Echtzeit markieren oder bei fehlerhaften Instrumentenbewegungen Warnungen ausgeben [22, 62]. Darüber hinaus würde eine KI-gestützte Dokumentation den operativen Ablauf automatisch aufzeichnen und analysieren, wodurch nicht nur der administrative Aufwand reduziert, sondern auch wertvolle Daten für zukünftige Operationen generiert werden könnten. Der Weg zu diesem Idealzustand erfordert jedoch erhebliche Fortschritte im Datenmanagement, vorhandener Technologie und regulatorischen Rahmenbedingungen.

Die verschiedenen Schweregrade unerwünschter Arzneimittelwirkungen (UAW) werden in klinischen Studien hinsichtlich ihrer Häufigkeit berichtet, jedoch oft ohne individualisierbare oder für eine KI-prädiktive Analyse verwertbare statistische Zusammenhänge. Perspektivisch soll die KI-gestützte Vorhersage von UAW durch eine Verbesserung der Datenverfügbarkeit für die wissenschaftliche Öffentlichkeit ermöglicht werden, um eine Optimierung der Therapie zu erreichen. (KONSENS)

Primäre Endpunkte der Zulassungsstudien konzentrieren sich auf onkologische Endpunkte wie Gesamtüberleben, progressionsfreies Überleben und die Zeit bis zur Folgetherapie. Die Abhängigkeit dieser Endpunkte von klinischen Parametern wird gewöhnlich im Kontext von Subgruppenanalysen erörtert. Toxizitäten werden hinsichtlich der Häufigkeit und der unterschiedlichen Schweregrade berichtet, jedoch in der Regel ohne die Durchführung von Subgruppenanalysen oder die Etablierung statistisch signifikanter Zusammenhänge zu patientenbezogenen Parametern. Für die zukünftige prädiktive Modellierung von UAW unter Verwendung maschineller Lernverfahren sollten die benötigten Rohdaten sowie relevante Informationen für potenzielles „Feature Engineering“ – analog zu den Analysen zur onkologischen Wirksamkeit – im Rahmen einer Subgruppenanalyse bereitgestellt werden.

Zur zielgerichteten Wissensvermittlung für Patienten könnten sorgfältig validierte KI-Systeme entwickelt und unter der Aufsicht eines Arztes eingesetzt werden. Langfristig könnten diese eine maßgeschneiderte Schulung für Patienten anbieten, die auf die individuellen Bedürfnisse und Gesundheitszustände der Patienten zugeschnitten sind. (KONSENS)

Die KI kann dabei helfen, die Gesundheitskompetenz der Patienten zu verbessern und individuell zugeschnittene Informationen bereitzustellen. Eine entscheidende Rolle spielt die Supervision durch Ärzte, um sicherzustellen, dass die Technologie korrekt angewendet wird und keine Risiken für die Patienten entstehen. Langfristig könnte KI eine zentrale Rolle in der Unterstützung der personalisierten Patien-

tenbetreuung spielen, wenn Ethik und Datenschutz bei ihrer Implementierung berücksichtigt werden [63].

Die KI-Systeme können zukünftig die Wissensbeschaffung wesentlich erleichtern und optimieren, wodurch sich Auszubildende vermehrt auf den Wissenserwerb konzentrieren können. (KONSENS)

Das „Deep-Learning-Konzept“ ermöglicht KI-Modellen, sehr große Datenmengen effizient zu verarbeiten, zu analysieren und aufzubereiten [64, 65]. KI-gestützte Lernsysteme könnten den Wissenszugang revolutionieren, in dem sie das Navigieren durch umfangreiche medizinische Literatur erleichtern. Sie können adaptive Anweisungen sowie multimodale Inhalte bereitstellen und sich an individuelle Lernbedürfnisse anpassen. Durch kognitive Modellierung und effektives Feedback optimieren sie den Lernprozess, während interaktive Szenarien praxisnahe Erfahrungen ermöglichen. Diese Ansätze verbessern nicht nur die Wissensaneignung, sondern auch Selbstreflexion und Lernstrategien, wodurch der Lernerfolg langfristig gesteigert werden könnte [66].

Durch den Einsatz von KI-basierten Lehrmethoden soll die medizinische Lehre skalierbar weiter personalisiert und letztendlich der individuelle Lernerfolg gesteigert werden. (KONSENS)

Eine möglichst personalisierte Lehre erhöht gerade im Medizinstudium den Lernerfolg signifikant [67–69]. Gleichzeitig ist das Studium der Humanmedizin einer der teuersten Studiengänge in Deutschland [70] und die medizinischen Fakultäten operieren bereits am oberen Rand ihrer personellen und finanziellen Kapazitäten. Der Einsatz von KI-basierten Tools wie z. B. Sprachmodellen stellt eine sehr valide Möglichkeit dar, um die Qualität der Lehre und damit die Wissensretention und den Fähigkeitserwerb zu erhöhen, ohne dabei die Kosten des Studiums per capita wesentlich zu steigern [71–73].

Digitale Transformation (vom Status quo zur maximalen Potenzilausschöpfung)

Es ist eine interdisziplinäre Bündelung aller Ressourcen erforderlich, die es ermöglicht, durch Förderung von Kooperationsnetzwerken zwischen Medizinern

verschiedener Fachdisziplinen, Informatikern und anderen Experten auf dem Gebiet der digitalen Medizin, valide KI-Modelle zu entwickeln. (STARKER KONSENS)

Die Entwicklung von KI-Anwendungen für die Medizin erfordert die Kooperation zwischen Spezialisten aus sehr unterschiedlichen Fachrichtungen. Dies erfordert interdisziplinäres Arbeiten, das direktes Feedback ermöglicht, aber auch Fort- und Weiterbildungsarbeit, um entsprechendes fachfremdes Wissen grundlegend zu beherrschen und ein Verständnis für die speziellen Anforderungen zu entwickeln [74].

Hierbei sollte die technische Entwicklung mit der praktischen in Einklang gebracht werden. Dieses sollte bereits in der Planungsphase gewährleistet sein, um praktisch relevante Modelle für den klinischen Nutzen zu entwickeln. Auf diese Weise wird es möglich, dass sich KI-Tools mit der klinischen Praxis weiterentwickeln und die diagnostische Genauigkeit und Effizienz verbessern.

Durch eine idealerweise einrichtungsübergreifende Standardisierung der klinischen Prozesse und Dokumentationen könnte die Entwicklung KI-basierter Unterstützungssysteme inklusive Sprachmodellen begünstigt werden, da Engpässe bezüglich der Datenverfügbarkeit, des Datenaustausches und der Datenstandardisierung besser adressiert wären. (KONSENS)

Daten werden durch Standardisierung interoperabel und erst dadurch analysierbar [75, 76]. Inkonsistente Daten oder Daten mit fehlenden Werten erschweren das Training von KI-Methoden. Ebenso wird in beiden Publikationen darauf hingewiesen, dass eine stärkere Zusammenarbeit zwischen Medizinerinnen und Informatikern dazu beitragen kann, eine Repräsentation der Daten zu finden, welche eine effiziente Verarbeitung mittels KI-Methoden und einen Datenaustausch ermöglicht.

Es sollten spezielle Sprachmodelle trainiert werden, die objektive Gesundheitsdaten trotz individueller Textformulierungen sicher verarbeiten können. (STARKER KONSENS)

Die Komplexität und Menge von Patientendaten, die abteilungsspezifisch und damit heterogen dokumentiert werden,

erfordert eine sichere und vollständige Erkennung der relevanten Informationen [77]. Das aktuell hohe Risiko für Halluzinationen einerseits und Mängeln bei der Befunderkennung durch KI andererseits begünstigt Fehlinterpretationen [78], die mit steigender Qualität der Modelle reduziert werden.

Der Überführung diagnostischer KI-Algorithmen zur Analyse radiologischer Bildgebung in die klinische Routine sollen externe Validierungsstudien vorausgehen [79–82]. Diese Validierungsstudien sollten spezifisch den klinischen Anwendungsfall an repräsentativen Patientenkohorten [80, 83, 84] simulieren, die Algorithmus-Performance an einem robusten Referenzstandard (z. B. Konsensus-Reading mehrerer im jeweiligen Anwendungsfeld erfahrener radiologischer Fachärzte [80, 85–87]) messen und auch auf mögliche „Confounder“ [88–92] in den zugrunde liegenden Trainingsdaten testen, um systematische Fehler zu vermeiden und die Robustheit der Algorithmen zu gewährleisten [92]. (KONSENS)

Diese Validierungsstudien müssen robuste Referenzstandards [79–82, 85–87] nutzen, auf „Confounder“ [88–92] prüfen und an realen Patientenkohorten [80, 83, 84] durchgeführt werden, um systematische Fehler zu vermeiden [92].

Bei der Implementierung und Nutzung diagnostischer Algorithmen zur Analyse radiologischer Bildgebung sollten auch potenziell negativen Auswirkungen sorgfältig berücksichtigt werden und Maßnahmen (ggf. auch instituts-/klinikspezifisch) ergriffen werden, um diese Risiken zu minimieren. (KONSENS)

Um klinische Risiken durch den Einsatz von KI zu minimieren, sollten prospektive Studien durchgeführt werden, die auch klinische Endpunkte einbeziehen. Diese Studien bieten die Möglichkeit, die tatsächlichen Auswirkungen auf Patienten und mögliche negative Effekte wie Fehldiagnosen besser zu verstehen [19, 93].

Die Ausgabe eines KI-Algorithmus sollte dem Arzt verständlich erklärt werden sowie dessen Unsicherheit mitgeteilt werden, anstelle einer simplifizierenden Ja-oder-Nein-Antwort, damit die Ausgabe ärztlich auditiert, und an den Patienten kommuniziert werden kann. (KONSENS)

Ein KI-Algorithmus sollte dem Anwender die individuelle Ausgabe erklären können, um so Transparenz und Verständnis zu schaffen („explainable AI“). Dies kann erreicht werden durch die Verwendung intrinsisch erklärbarer Modelle wie linearer Modelle oder Entscheidungsbäume. Alternativ können auch Surrogatmodelle wie LIME („local interpretable model-agnostic explanations“) oder Methoden aus der kooperativen Spieltheorie wie SHAP („shapley additive explanations“) eingesetzt werden, um die Entscheidungsfindung des Modells nachvollziehen zu können.

Ein KI-Algorithmus sollte zudem eine Einschätzung seiner Unsicherheit bezüglich der Ausgabe liefern. Dies könnte durch eine probabilistische Kalibrierung an die Zielpopulation erfolgen, beispielsweise in Form von Konfidenzintervallen für Regressionsaufgaben oder Wahrscheinlichkeiten für Klassifizierungsaufgaben. Ein geeignetes Framework hierfür bietet „Conformal Prediction“ [94, 95].

Modelle des maschinellen Lernens (ML) sollen implementiert werden, um UAW frühzeitig zu identifizieren oder vorherzusagen. Diese Modelle sind kontinuierlich zu aktualisieren und zu validieren. (STARKER KONSENS)

Die Implementierung dieser Modelle in der Medizin wird zunehmend als transformative Methode anerkannt. ML kann z. B. erfolgreich den Behandlungserfolg bei Psoriasispatienten vorhersagen, wobei eine Genauigkeit von >80% bei der Vorhersage des Abbruchs von Medikamenten aufgrund unerwünschter Ereignisse oder Verlust der Wirksamkeit erreicht wurde [96]. Die kontinuierliche Aktualisierung und Validierung dieser Modelle ist entscheidend, um ihre Relevanz und Genauigkeit in einem sich schnell wandelnden Gesundheitsumfeld aufrechtzuerhalten. Ross et al. betonten die Bedeutung von Echtzeitprognoseanalysen aus elektronischen Patientenakten, was eine fortlaufende Verfeinerung der Modelle erfordert, um sich an neue Daten und aufkommende Muster anzupassen [97].

Um Patienten zielgerichteter mit Informationen zu versorgen, soll die Entwicklung und Implementierung von KI-gestützten Chatbots vorangetrieben

werden. Dabei soll die Integration in bestehende Versorgungsstrukturen gewährleistet sowie die Datensicherheit und der Datenschutz sichergestellt werden. (KONSENS)

Die Entwicklung von KI-gestützten Chatbots bietet großes Potenzial, um Patienten schnell, effizient und personalisiert mit Gesundheitsinformationen zu versorgen. Eine erfolgreiche Implementierung setzt voraus, dass die Chatbots für die Patienten benutzerfreundlich sind und sich nahtlos in bestehende Versorgungsstrukturen integrieren lassen, um die Arbeitsabläufe von Ärzten und Pflegepersonal zu unterstützen [98]. Entscheidend ist auch, dass die Datensicherheit und der Schutz sensibler Gesundheitsdaten durch sorgfältige Regulierung und Verschlüsselungstechnologien gewährleistet werden.

Um KI effizient und zuverlässig zur Verbesserung der Patientensicherheit und chirurgischen Qualität einzusetzen, müssen krankenhauspolitische und rechtliche Voraussetzungen geschaffen werden, welche die weitreichende Nutzung von multizentrischen klinischen sowie Video- und Bilddatensätzen für das Training und die Validierung von KI-Modellen unterstützen. (KONSENS)

Neben einer umfassenden technischen und regulatorischen Infrastruktur bedarf die klinische Implementierung von KI in intraoperative chirurgische Prozesse v. a. umfassende klinische Studien zur Validierung der Sicherheit, Zuverlässigkeit und Relevanz der Modelle. KI-Methoden sollten im Kontext realer Entscheidungsszenarien und klinischer Outcomes evaluiert werden, um die Praxistauglichkeit sicherzustellen.

Der klinische Einsatz von KI zur Qualitätssicherung und Verbesserung der Patientensicherheit in der Chirurgie erfordert die zunehmende Einbeziehung aller Interessensgruppen. Dies schließt Vertreter von Datenschutz, Informationstechnik und Industrie sowie aktuell unterrepräsentierte Gruppen wie Pflegekräften, Verwaltung, Medizinstudenten, Datenschutzvertretern und Patienten mit ein. (KONSENS)

Neben chirurgischen Experten, die die klinische Relevanz von KI bewerten und deren zielgerichtete Entwicklung mitbeeinflussen sowie Computerwissenschaftlern,

die adäquate technische Lösungen entwickeln, bedarf es diverser anderer Interessensgruppen, um eine klinisch fundierte KI zu entwickeln. Neben administrativen und juristischen Fachkräften, um datenschutzrechtliche und regulatorische Anforderungen zu adressieren [99, 100], bedarf es v. a. auch der Einbindung von Patienten und Pflegekräften, die die Nutzung von KI im klinischen Rahmen bewerten. Ethische, moralische und kulturelle Aspekte sollten berücksichtigt werden, um den Weg für die klinische Implementierung zu ebnen. Mit der zunehmenden Komplexität und der fortschreitenden Integration von KI in die klinische Praxis wächst die Liste der Interessensgruppen stetig. Dies sichert nicht nur die Entwicklung robuster KI-Modelle, sondern trägt auch zur Akzeptanz und Sicherheit von KI-gestützten Anwendungen in der Chirurgie bei.

Die Unterstützung durch Politik und öffentliche Krankenhäuser ist entscheidend für die Entwicklung und Implementierung von KI-Modellen im Gesundheitswesen. (KONSENS)

Politische Unterstützung ist ausschlaggebend, um die klinische Anwendung von KI voranzutreiben und deren Potenzial zur Verbesserung der Gesundheitsversorgung auszuschöpfen. Die Schaffung regulatorischer Rahmenbedingungen, Bereitstellung notwendiger Ressourcen, Innovationsförderung, Stärkung der Zusammenarbeit von „Stakeholdern“, Etablierung von Datenschutzrichtlinien und ethischen Standards sollten gefördert werden [101, 102]. Öffentliche Krankenhäuser bieten eine ideale Plattform für die Entwicklung und Implementierung von KI-Technologien, da sie über große und diverse Patientendaten verfügen und in der Praxis validierte Modelle benötigen. Ein ganzheitlicher Ansatz ist notwendig, um langfristige Verbesserungen zu erzielen [103].

Diskussion

Der erarbeitete Delphi-Konsensusprozess für KI-Anwendungen in der Chirurgie zeigt eindrücklich, welche großen Chancen KI für eine verbesserte Patientenversorgung darstellt. Die Anwendungen umfassen nahezu alle Bereiche der chirurgischen Behandlungsabläufe, von der Prävention über die

Diagnostik, Aufklärung bis hin zur Therapie und Nachsorge der Patienten. Um dieses enorme Potenzial zu verwirklichen, wurden im Zuge des systematischen Delphi-Konsensusprozesses bestehende Limitationen identifiziert, die insbesondere auf die Verbesserung einheitlicher Digitalisierungsstrukturen sowie die rechtlichen Rahmenbedingungen fokussieren. Die interdisziplinäre Expertengruppe hat hieraus konkrete und praxisnahe Vorschläge für eine erfolgreiche Transformation erstellt, die alle verantwortlichen Akteure in der Medizin, Politik und Wirtschaft adressiert. Limitation dieses Delphi-Konsensus ist die relativ geringe Anzahl von 16 Akteuren, die jedoch durch die breite interdisziplinäre Zusammensetzung sowie die intensiven und konstruktiven Diskussionen in diesem kleinen Rahmen ausbalanciert wurde. Wir appellieren daran, die entwickelten Lösungsansätze zukünftig ebenso interdisziplinär zu erarbeiten und zu beraten, um im internationalen Vergleich und Wettbewerb das deutsche Gesundheitswesen auch weiterhin zukunftsfähig in die digitale Transformation zu führen.

Fazit für die Praxis

- Künstliche Intelligenz (KI) hat bedeutendes Potenzial, die Patientenversorgung in allen Bereichen chirurgischer und angeschlossener Fachrichtungen zu verbessern.
- KI-basierte Gesundheitslösungen sollten von Grund auf multidisziplinär konzipiert und entwickelt werden. Dazu gehören u. a. medizinische Leistungserbringer, Medizintechnik und Pharmaindustrie, digitale Medienpräsenz, Patientenvertreter und Experten aus den Bereichen Ethik und Recht.
- Für die nationale klinische Anwendung von KI-basierten Assistenzsystemen stellen die rechtlichen Verwaltungsstrukturen des Föderalismus eine komplexe politisch-strukturelle Herausforderung dar. Dies betrifft insbesondere die Umsetzung von Datenschutzvorgaben, Informationssicherheit sowie bestehende heterogene Digitalisierungsstandards.
- Es besteht die medizinische Notwendigkeit, dass die Verantwortung für Diagnostik, Diagnosestellung und Therapie, auch bei der Verwendung KI-gestützter Systeme, zu jeder Zeit den behandelnden Ärzten unterliegt.

Korrespondenzadresse



Dr. med. G. Duwe

Klinik und Poliklinik für Urologie und Kinderurologie, Universitätsmedizin der Johannes Gutenberg-Universität Mainz
Langenbeckstr. 1, 55131 Mainz, Deutschland
gregor.duwe@unimedizin-mainz.de

Förderung. Keine

Funding. Open Access funding enabled and organized by Projekt DEAL.

Datenverfügbarkeit. Detaillierte Daten zur Konsistenzstärke der einzelnen Statements können auf Anfrage eingesehen werden.

Einhaltung ethischer Richtlinien

Interessenkonflikt. G. Duwe, K. Moench, V. Kauth, M. Angeloni, J. Eckhoff, M. Götz, S. Hoefert, T.D. Kocar, L. Kollitsch, S. Mehralivand, D. Mercier, J. Rudolph, J. Rueckel, R. Schönhof, M. Sondermann, C. von Klot, A. Zamzow, J.P. Struck und H. Borgmann geben an, dass kein Interessenkonflikt besteht.

Für diesen Beitrag wurden von den Autor/-innen keine Studien an Menschen oder Tieren durchgeführt. Für die aufgeführten Studien gelten die jeweils dort angegebenen ethischen Richtlinien.

Open Access. Dieser Artikel wird unter der Creative Commons Namensnennung 4.0 International Lizenz veröffentlicht, welche die Nutzung, Vervielfältigung, Bearbeitung, Verbreitung und Wiedergabe in jeglichem Medium und Format erlaubt, sofern Sie den/die ursprünglichen Autor(en) und die Quelle ordnungsgemäß nennen, einen Link zur Creative Commons Lizenz beifügen und angeben, ob Änderungen vorgenommen wurden. Die in diesem Artikel enthaltenen Bilder und sonstiges Drittmaterial unterliegen ebenfalls der genannten Creative Commons Lizenz, sofern sich aus der Abbildungslegende nichts anderes ergibt. Sofern das betreffende Material nicht unter der genannten Creative Commons Lizenz steht und die betreffende Handlung nicht nach gesetzlichen Vorschriften erlaubt ist, ist für die oben aufgeführten Weiterverwendungen des Materials die Einwilligung des jeweiligen Rechteinhabers einzuholen. Weitere Details zur Lizenz entnehmen Sie bitte der Lizenzinformation auf <http://creativecommons.org/licenses/by/4.0/deed.de>.

Literatur

1. Hashimoto DA, Rosman G, Rus D, Meireles OR (2018) Artificial Intelligence in Surgery: Promises and Perils. *Ann Surg* 268(1):70–76. <https://doi.org/10.1097/SLA.0000000000002693>
2. <https://www.aerzteblatt.de/nachrichten/153917/Expertengruppe-legt-Zehn-Punkte-Plan-zur-Digitalisierung-vor>. Letzter Aufruf 20. November 2024.
3. Niederberger M, Deckert S (2022) The Delphi technique: Methodology, variants and usage examples. *Z Evid Fortbild Qual Gesundhwes* 174:11–19. <https://doi.org/10.1016/j.zefq.2022.08.007>
4. Eckhoff JA, Rosman G, Altieri MS, Speidel S, Stoyanov D, Anvari M et al (2023) SAGES consensus recommendations on surgical video data use, structure, and exploration (for research in artificial intelligence, clinical quality improvement, and surgical education). *Surg Endosc* 37(11):8690–8707. <https://doi.org/10.1007/s00464-023-10288-3>
5. Maier-Hein L, Eisenmann M, Sarikaya D, Marz K, Collins T, Malpani A et al (2022) Surgical data science – from concepts toward clinical translation. *Med Image Anal* 76:102306. <https://doi.org/10.1016/j.media.2021.102306>
6. Hiltawsky K, al. e. KI für bessere Abläufe in Medizin und Pflege. Anwendungen und Potenziale in organisatorischen Prozessen. Whitepaper aus der Plattform Lernende Systeme, München.: Bundesministerium für Bildung und Forschung (BMBF); 2024.
7. <https://www.aerzteblatt.de/nachrichten/151383/KI-in-der-Medizin-Aerzteschaft-fuer-verbindlichen-Rechtsrahmen>. Letzter Aufruf 20. November 2024.
8. <https://www.aerzteblatt.de/nachrichten/sw/K%FCnstliche%20Intelligenz?s=&p=1&n=1&nid=155037>. Letzter Aufruf 20. November 2024.
9. WHO. Regulatory considerations on artificial intelligence for health. Geneva: World Health Organization; 2023. Licence: CCBY-NC-SA 3.0 IGO.
10. Arora A, Alderman JE, Palmer J et al (2023) The value of standards for health datasets in artificial intelligence-based applications. *Nat Med* 29(11):2929–2938. <https://doi.org/10.1038/s41591-023-02608-w>
11. Rieke N, Hancox J, Li W et al (2020) The future of digital health with federated learning. *NPJ Digit Med* 14(3):119–119. <https://doi.org/10.1038/s41746-020-00323-1>
12. Shiwani T, Relton S, Evans R, Kale A, Heaven A, Clegg A et al (2023) New Horizons in artificial intelligence in the healthcare of older people. *Age Ageing*. <https://doi.org/10.1093/ageing/afad219>
13. Oren O, Gersh BJ, Bhatt DL (2020) Artificial intelligence in medical imaging: switching from radiographic pathological data to clinically meaningful endpoints. *Lancet Digit Health* 2(9):e486–e488. [https://doi.org/10.1016/S2589-7500\(20\)30160-6](https://doi.org/10.1016/S2589-7500(20)30160-6)
14. Yoon JH, Han K, Suh HJ, Youk JH, Lee SE, Kim EK (2023) Artificial intelligence-based computer-assisted detection/diagnosis (AI-CAD) for screening mammography: Outcomes of AI-CAD in the mammographic interpretation workflow. *Eur J Radiol Open* 11:100509. <https://doi.org/10.1016/j.ejro.2023.100509>
15. Kuwabara M, Ikawa F, Sakamoto S, Okazaki T, Ishii D, Hosogai M et al (2023) Effectiveness of tuning an artificial intelligence algorithm for cerebral aneurysm diagnosis: a study of 10,000 consecutive cases. *Sci Rep* 13(1):16202. <https://doi.org/10.1038/s41598-023-43418-x>
16. Rueckel J, Sperl JJ, Kaestle S, Hoppe BF, Fink N, Rudolph J et al (2021) Reduction of missed thoracic findings in emergency whole-body computed tomography using artificial intelligence assistance. *Quant Imaging Med Surg* 11(6):2486–2498. <https://doi.org/10.21037/qims-20-1037>
17. Peters AA, Wiescholek N, Muller M, Klaus J, Strodka F, Macek A et al (2024) Impact of artificial intelligence assistance on pulmonary nodule detection and localization in chest CT: a comparative study among radiologists of varying experience levels. *Sci Rep* 14(1):22447. <https://doi.org/10.1038/s41598-024-73435-3>
18. Topff L, Steltenpool S, Ranschaert ER, Ramanauskas N, Menezes R, Visser JJ et al (2024) Artificial intelligence-assisted double reading of chest radiographs to detect clinically relevant missed findings: a two-centre evaluation. *Eur Radiol* 34(9):5876–5885. <https://doi.org/10.1007/s00330-024-10676-w>
19. Bernstein MH, Atalay MK, Dibble EH, Maxwell AWP, Karam AR, Agarwal S et al (2023) Can incorrect artificial intelligence (AI) results impact radiologists, and if so, what can we do about it? A multi-reader pilot study of lung cancer detection with chest radiography. *Eur Radiol* 33(11):8263–8269. <https://doi.org/10.1007/s00330-023-09747-1>
20. Le EPV, Wang Y, Huang Y, Hickman S, Gilbert FJ (2019) Artificial intelligence in breast imaging. *Clin Radiol* 74(5):357–366. <https://doi.org/10.1016/j.crad.2019.02.006>
21. Gortz M, Baumgartner K, Schmid T, Muschko M, Woessner P, Gerlach A et al (2023) An artificial intelligence-based chatbot for prostate cancer education: Design and patient evaluation study. *Digit Health* 9:20552076231173304. <https://doi.org/10.1177/20552076231173304>
22. Ban Y, Rosman G, Eckhoff JA, Ward TM, Hashimoto DA, Kondo T, et al. SUPR-GAN: SURgical PRediction GAN for Event Anticipation in Laparoscopic and Robotic Surgery. *IEEE Robotics and Automation Letters*. 2022:5741–8.
23. Demir KC, Schieber H, Roth D, Maier A, Yang SH (2022) Surgical Phase Recognition: A Review and Evaluation of Current Approaches. *TechRxiv*. <https://doi.org/10.36227/techrxiv.19665717.v1>
24. Nwoye CI et al (2020) Recognition of Instrument-Tissue Interactions in Endoscopic Videos via Action Triplets. *arXiv [eess.IV]*.
25. Ayobi N (2024) Pixel-Wise Recognition for Holistic Surgical Scene Understanding. <https://arxiv.org/abs/2401.00000>
26. Kirtac K (2022) Surgical Phase Recognition: From Public Datasets to Real-World Data. *NATO Adv. Sci. Inst. Ser. E Appl. Sci* 12:8746
27. Boal MWE, Anastasiou D, Tesfai F, Ghamrawi W, Mazomenos E, Curtis N et al (2024) Evaluation of objective tools and artificial intelligence in robotic surgery technical skills assessment: a systematic review. *Br J Surg*. <https://doi.org/10.1093/bjs/znad331>
28. Gesetz zur verbesserten Nutzung von Gesundheitsdaten, (25.03.2024).
29. Europäischer Raum für Gesundheitsdaten (EHDS) der Europäischen Union, (24.04.2024).
30. Das FDZ Gesundheit 2024; <https://www.forschungsdatenzentrum-gesundheit.de/forschung-digitalisierung>. Letzter Aufruf 20. November 2024.
31. Künstliche Intelligenz und Urheberrecht - Fragen und Antworten -, Bundesministerium der Justiz, Editor. März 2024.

32. Gerst T. KI in der Medizin – Ärztliche Letztverantwortung bleibt bestehen. *Rheinisches Ärzteblatt*. 2024;(5):25.
33. https://www.bundesärztekammer.de/fileadmin/user_upload/BAEK/Politik/Stellungnahmen/Aenderung_Approbnationsordnung_ueberarbeiteter_Referentenentwurf_SN_BAEK_10082023.pdf. Letzter Aufruf 20. November 2024.
34. https://www.bundesgesundheitsministerium.de/fileadmin/Dateien/3_Downloads/Gesetze_und_Verordnungen/GuV/A/Referentenentwurf_AEApprO.pdf. Letzter Aufruf 20. November 2024.
35. Gilson A, Safranek CW, Huang T, Socrates V, Chi L, Taylor RA et al (2023) How Does ChatGPT Perform on the United States Medical Licensing Examination (USMLE)? The Implications of Large Language Models for Medical Education and Knowledge Assessment. *JMIR Med Educ* 9:e45312. <https://doi.org/10.2196/45312>
36. Jung LB, Gudera JA, Wiegand TLT, Allmendinger S, Dimitriadis K, Koerte IK (2023) ChatGPT Passes German State Examination in Medicine With Picture Questions Omitted. *Dtsch Arztebl Int* 120(21):373–374. <https://doi.org/10.3238/arztebl.m2023.0113>
37. Kollitsch L, Eredics K, Marszalek M, Rauchenwald M, Brookman-May SD, Burger M et al (2024) How does artificial intelligence master urological board examinations? A comparative analysis of different Large Language Models' accuracy and reliability in the 2022 In-Service Assessment of the European Board of Urology. *World J Urol* 42(1):20. <https://doi.org/10.1007/s00345-023-04749-6>
38. Kung TH, Cheatham M, Medenilla A, Sillos C, De Leon L, Elepano C et al (2023) Performance of ChatGPT on USMLE: Potential for AI-assisted medical education using large language models. *PLOS Digit Health* 2(2):e0000198. <https://doi.org/10.1371/journal.pdig.0000198>
39. May M, Korner-Riffard K, Kollitsch L, Burger M, Brookman-May SD, Rauchenwald M et al (2024) Evaluating the Efficacy of AI Chatbots as Tutors in Urology: A Comparative Analysis of Responses to the 2022 In-Service Assessment of the European Board of Urology. *Urol Int* 108(4):359–366. <https://doi.org/10.1159/000537854>
40. May M, Korner-Riffard K, Marszalek M, Eredics K (2024) Would Uro-Chat, a Newly Developed Generative Artificial Intelligence Large Language Model, Have Successfully Passed the In-Service Assessment Questions of the European Board of Urology in 2022? *Eur Urol Oncol* 7(1):155–156. <https://doi.org/10.1016/j.euo.2023.08.013>
41. Singhal K, Tu T, Gottweis J, al. e. Towards expert-level medical question answering with large language models. *arXiv* <https://doi.org/10.48550/arXiv.2305.09617>.
42. Takagi S, Watari T, Erabi A, Sakaguchi K (2023) Performance of GPT-3.5 and GPT-4 on the Japanese Medical Licensing Examination: Comparison Study. *JMIR Med Educ* 9:e48002. <https://doi.org/10.2196/48002>
43. Scott IA, Shaw T, Slade C, Wan TT, Coorey C, Johnson SLJ et al (2023) Digital health competencies for the next generation of physicians. *Intern Med J* 53(6):1042–1049. <https://doi.org/10.1111/imj.16122>
44. Hugosson J, Roobol MJ, Månsson M, Tammela TLJ, Zappa M, Nelen V et al (2019) A 16-yr Follow-up of the European Randomized study of Screening for Prostate Cancer. *Eur Urol Oncol* 76(1):43–51. <https://doi.org/10.1016/j.eururo.2019.02.009>
45. Lotter W, Diab AR, Haslam B, Kim JG, Grisot G, Wu E et al (2021) Robust breast cancer detection in mammography and digital breast tomosynthesis using an annotation-efficient deep learning approach. *Nat Med* 27(2):244–249. <https://doi.org/10.1038/s41591-020-01174-9>
46. Topol E (2019) Deep Medicine: How Artificial Intelligence Can Make Healthcare Human Again, 1 edn. Basic Books
47. Sicklick JK, Kato S, Okamura R, Schwaederle M, Hahn ME, Williams CB et al (2019) Molecular profiling of cancer patients enables personalized combination therapy: the I-PREDICT study. *Nat Med* 25(5):744–750. <https://doi.org/10.1038/s41591-019-0407-5>
48. Giordano C, Brennan M, Mohamed B, Rashidi P, Modave F, Tighe P (2021) Accessing Artificial Intelligence for Clinical Decision-Making. *Front Digit Health* 3:645232. <https://doi.org/10.3389/fdgh.2021.645232>
49. Arowoogun JO, Babawarun O, Chidi R, Adeniyi AO, Okolo CA (2024) A comprehensive review of data analytics in healthcare management: Leveraging big data for decision-making. *World Journal of Advanced Research and Reviews* 21(2):1810–1821
50. Arora A, Arora A (2022) Generative adversarial networks and synthetic patient data: current challenges and future perspectives. *Future Healthc J* 9(2):190–193. <https://doi.org/10.7861/fhj.2022-0013>
51. Schoenhof R, Schoenhof R, Blumenstock G, Lethaus B, Hoefert S (2024) Synthetic, non-person related panoramic radiographs created by generative adversarial networks in research, clinical, and teaching applications. *J Dent* 146:105042. <https://doi.org/10.1016/j.jdent.2024.105042>
52. Sun C, van Soest J, Dumontier M (2023) Generating synthetic personal health data using conditional generative adversarial networks combining with differential privacy. *AMIA Annu Symp Proc* 143:104404. <https://doi.org/10.1016/j.jbi.2023.104404>
53. Giuffrè M, Shung DL (2023) Harnessing the power of synthetic data in healthcare: innovation, application, and privacy. *NPJ Digit Med* 6(1):186. <https://doi.org/10.1038/s41746-023-00927-3>
54. Kaulartz M, Braegelmann T, Ammann T (2020) *Rechtshandbuch Artificial Intelligence und Machine Learning*. C.H.Beck
55. Beck S, Faber M, Gerndt S (2023) *Rechtliche Aspekte des Einsatzes von KI und Robotik in Medizin und Pflege*. Ethik Med 35:247–263
56. Misselhorn C (2021) *Künstliche Intelligenz und Empathie. Vom Leben mit Emotionserkennung, Sexrobotern & Co.* Reclam-Verlag
57. § 630e BGB (Bürgerliches Gesetzbuch): Aufklärungspflichten.
58. Chen J, Cheng N, Cacciamani G, Oh P, Lin-Brandt M, Remulla D et al (2019) Objective Assessment of Robotic Surgical Technical Skill: A Systematic Review. *J Urol* 201(3):461–469. <https://doi.org/10.1016/j.juro.2018.06.078>
59. Knudsen JE, Ghaffar U, Ma R, Hung AJ (2024) Clinical applications of artificial intelligence in robotic surgery. *J Robot Surg* 18(1):102–102. <https://doi.org/10.1007/s11701-024-01867-0>
60. Kruger CM, Ruckbeil O, Sebestyen U, Schlick T, Kurbis J, Riediger H (2021) DeRAS I-German situation of robotic-assisted surgery-an online survey. *Chirurgie* 92(12):1107–1113. <https://doi.org/10.1007/s00104-021-01404-x>
61. Alip SL, Kim J, Rha KH, Han WK (2022) Future Platforms of Robotic Surgery. *Urol Clin North Am* 49(1):23–38. <https://doi.org/10.1016/j.ucl.2021.07.008>
62. Fei N, Lu Z, Gao Y, Yang G, Huo Y, Wen J et al (2022) Towards artificial general intelligence via a multimodal foundation model. *Nat Commun* 13(1):3094. <https://doi.org/10.1038/s41467-022-30761-2>
63. WHO. Guidance on Ethics & Governance of Artificial Intelligence for Health: World Health Organization (WHO): www.who.int; 2019.
64. Borchers M. Künstliche Intelligenz in der Medizin. *Uro-News* [Internet]. 2023; 10.
65. Ray P (2023) ChatGPT: A comprehensive review on background, applications, key challenges, bias, ethics, limitations and future scope. *Internet of Things and Cyber-Physical Systems* 3:121–154. <https://doi.org/10.1016/j.iotcps.2023.04.003>
66. As'ad A (2024) Intelligent Tutoring Systems, Generative Artificial Intelligence (AI), and Healthcare Agents: A Proof of Concept and Dual-Layer Approach. *Cureus* 16(9):e69710. <https://doi.org/10.7759/cureus.69710>
67. Chen F, Lui AM, Martinelli SM (2017) A systematic review of the effectiveness of flipped classrooms in medical education. *Med Educ* 51(6):585–597. <https://doi.org/10.1111/medu.13272>
68. Trullas JC, Blay C, Sarri E, Pujol R (2022) Effectiveness of problem-based learning methodology in undergraduate medical education: a scoping review. *BMC Med Educ* 22(1):104. <https://doi.org/10.1186/s12909-022-03154-8>
69. Fasino PS, Wilborn TW (2024) Pharmacology education in the medical curriculum: Challenges and opportunities for improvement. *Pharmacol Res Perspect* 12(1):e1178. <https://doi.org/10.1002/prp2.1178>
70. <https://www-genesis.destatis.de/genesis/online?sequenz=tabelleErgebnis&selectionname=21371-0001>. Letzter Aufruf 20. November 2024.
71. Igboke I (2023) Application of artificial intelligence (AI) in educational management. *International Journal of Scientific and Research Publications* 13(3):300–307
72. Kamalov F, Calonge S, Gurrib I (2023) New Era of Artificial Intelligence in Education: Towards a Sustainable Multifaceted Revolution. *Sustainability* 15(16):12451. <https://doi.org/10.3390/su151612451>
73. George B, Wooden O (2023) Managing the Strategic Transformation of Higher Education through Artificial Intelligence. *Adm Sci* 13(9):196–196. <https://doi.org/10.3390/admsci13090196>
74. Park JI, Lee HY, Kim H, Lee J, Shinn J, Kim HS (2021) Lack of Acceptance of Digital Healthcare in the Medical Market: Addressing Old Problems Raised by Various Clinical Professionals and Developing Possible Solutions. *J Korean Med Sci* 36(37):e253. <https://doi.org/10.3346/jkms.2021.36.e253>
75. Isgut M, Gloster L, Choi K, Venugopalan J, Wang MD (2023) Systematic Review of Advanced AI Methods for Improving Healthcare Data Quality in Post COVID-19 Era. *IEEE Rev Biomed Eng* 16:53–69. <https://doi.org/10.1109/RBME.2022.3216531>
76. Norori N, Hu Q, Aellen FM, Faraci FD, Tzovara A (2021) Addressing bias in big data and AI for health care: A call for open science. *Patterns* 2(10):100347. <https://doi.org/10.1016/j.patter.2021.100347>
77. Duwe G, Mercier D, Wiesmann C, Kauth V, Moench K, Junker M et al (2024) Challenges and perspectives in use of artificial intelligence to support treatment recommendations in clinical oncology. *Cancer Med* 13(12):e7398. <https://doi.org/10.1002/cam4.7398>
78. Freyer O, Wiest IC, Kather JN, Gilbert S (2024) A future role for health applications of large language models depends on regulators enforcing safety

- standards. *Lancet Digit Health* 6(9):e662–e672. [https://doi.org/10.1016/S2589-7500\(24\)00124-9](https://doi.org/10.1016/S2589-7500(24)00124-9)
79. Duron L, Ducarouge A, Gillibert A, Laine J, Allouche C, Cherel N et al (2021) Assessment of an AI Aid in Detection of Adult Appendicular Skeletal Fractures by Emergency Physicians and Radiologists: A Multicenter Cross-sectional Diagnostic Study. *Radiol Imaging Cancer* 300(1):120–129. <https://doi.org/10.1148/radiol.2021203886>
 80. Dembrower K, Crippa A, Colon E, Eklund M, Strand F, ScreenTrust CADTC (2023) Artificial intelligence for breast cancer detection in screening mammography in Sweden: a prospective, population-based, paired-reader, non-inferiority study. *Lancet Digit Health* 5(10):e703–e711. [https://doi.org/10.1016/S2589-7500\(23\)00153-X](https://doi.org/10.1016/S2589-7500(23)00153-X)
 81. Grafe D, Beeskow AB, Pfaffle R, Rosolowski M, Chung TS, DiFranco MD (2024) Automated bone age assessment in a German pediatric cohort: agreement between an artificial intelligence software and the manual Greulich and Pyle method. *Eur Radiol* 34(7):4407–4413. <https://doi.org/10.1007/s00330-023-10543-0>
 82. Park SH, Han K (2018) Methodologic Guide for Evaluating Clinical Performance and Effect of Artificial Intelligence Technology for Medical Diagnosis and Prediction. *Radiol Imaging Cancer* 286(3):800–809. <https://doi.org/10.1148/radiol.2017171920>
 83. Rudolph J, Schachtner B, Fink N, Koliogiannis V, Schwarze V, Goller S et al (2022) Clinically focused multi-cohort benchmarking as a tool for external validation of artificial intelligence algorithm performance in basic chest radiography analysis. *Sci Rep* 12(1):12764. <https://doi.org/10.1038/s41598-022-16514-7>
 84. Savage CH, Tanwar M, Elkassem AA, Sturdivant A, Hamki O, Sotoudeh H et al (2024) Prospective Evaluation of Artificial Intelligence Triage of Intracranial Hemorrhage on Noncontrast Head CT Examinations. *AJR Am J Roentgenol*. <https://doi.org/10.2214/AJR.24.31639>
 85. Rudolph J, Huemmer C, Preuhs A, Buizza G, Hoppe BF, Dinkel J et al (2024) Nonradiology Health Care Professionals Significantly Benefit From AI Assistance in Emergency-Related Chest Radiography Interpretation. *Chest* 166(1):157–170. <https://doi.org/10.1016/j.chest.2024.01.039>
 86. Rudolph J, Huemmer C, Ghesu FC, Mansoor A, Preuhs A, Fieselmann A et al (2022) Artificial Intelligence in Chest Radiography Reporting Accuracy: Added Clinical Value in the Emergency Unit Setting Without 24/7 Radiology Coverage. *Invest Radiol* 57(2):90–98. <https://doi.org/10.1097/RLI.0000000000000813>
 87. Rudolph J, Rueckel J, Döpfert J, Wen X, Opalka J, Brem C et al (2024) Artificial intelligence-based rapid brain volumetry substantially improves differential diagnosis in dementia. *Alzheimers Dement* 70037. <https://doi.org/10.1002/dad2.70037>
 88. Rueckel J, Trappmann L, Schachtner B, Wesp P, Hoppe BF, Fink N et al (2020) Impact of Confounding Thoracic Tubes and Pleural Dehiscence Extent on Artificial Intelligence Pneumothorax Detection in Chest Radiographs. *Invest Radiol* 55(12):792–798. <https://doi.org/10.1097/RLI.0000000000000707>
 89. Rueckel J, Huemmer C, Fieselmann A, Ghesu FC, Mansoor A, Schachtner B et al (2021) Pneumothorax detection in chest radiographs: optimizing artificial intelligence system for accuracy and confounding bias reduction using image annotations in algorithm training. *Eur Ra-*

Artificial intelligence in surgical disciplines: Clinical application, advantages, and potential—a Delphi expert consensus

Background: Artificial intelligence (AI) in surgical disciplines has the potential to support all areas of patient care, with the goal of improving treatment quality and patient safety. A group of multidisciplinary experts discussed the current situation as well as steps required to successfully integrate AI into surgical disciplines in the context of a consensus conference at the second Digital Health Summit (Brandenburg an der Havel, Germany) in August 2024.

Methods: A modified Delphi procedure was performed with 16 multidisciplinary physicians and scientists on the topic of AI in surgical disciplines and beyond. In two online meetings with subsequent Delphi survey rounds (LimeSurvey) and a final hybrid meeting, individual statements were contributed, discussed, and consented by all 16 participants based on current national clinical guidelines.

Results: From a total of 103 submitted statements, 36 statements on reality ($n = 12$), utopia ($n = 13$), and opportunities for digital transformation ($n = 11$) were consented after discussion and modification. We achieved a consensus of at least 75% for all the statements presented, with six of the statements achieving a strong consensus of 100% agreement.

Conclusion: The consensus statements show the great potential of AI for improving patient care in surgical disciplines. Challenges such as the lack of digitalization structures and legal frameworks were identified, and practice-oriented proposals for implementation were developed. The need for multidisciplinary cooperation between medical professionals, politics, and industry was emphasized in order to facilitate the German healthcare system remaining competitive for the future, both nationally and internationally.

Keywords

Artificial intelligence · Surgery · Delphi consensus conference · Legal framework · Digitalization

- diol 31(10):7888–7900. <https://doi.org/10.1007/s00330-021-07833-w>
90. Zhao Q, Adeli E, Pohl KM (2020) Training confounder-free deep learning models for medical applications. *Nat Commun* 11(1):6010. <https://doi.org/10.1038/s41467-020-19784-9>
91. Rao A, Monteiro JM, Mourao-Miranda J, Alzheimer's Disease I (2017) Predictive modelling using neuroimaging data in the presence of confounds. *Neuroimage* 150:23–49. <https://doi.org/10.1016/j.neuroimage.2017.01.066>
92. Kocak B, Ponsiglione A, Stanzione A, Bluethgen C, Santinha J, Ugga L et al (2024) Bias in artificial intelligence for medical imaging: fundamentals, detection, avoidance, mitigation, challenges, ethics, and prospects. *Diagn Interv Radiol*. <https://doi.org/10.4274/dir.2024.242854>
93. Reyes M, Meier R, Pereira S, Silva CA, Dahlweid FM, von Tengg-Koblick H et al (2020) On the Interpretability of Artificial Intelligence in Radiology: Challenges and Opportunities. *Radiol Imaging Cancer* 2(3):e190043. <https://doi.org/10.1148/ryai.2020190043>
94. Molnar C (2022) Interpretable Machine Learning: A Guide For Making Black Box Models Explainable. Independently published
95. Manokhin V (2023) Practical Guide to Applied Conformal Prediction in Python: Learn and apply the best uncertainty frameworks to your industry applications, 1 edn. Packt Publishing
96. Emam S, Du AX, Surmanowicz P, Thomsen SF, Greiner R, Gniadecki R (2020) Predicting the long-term outcomes of biologics in patients with psoriasis using machine learning. *Br J Dermatol* 182(5):1305–1307. <https://doi.org/10.1111/bjd.18741>
97. Ross EG, Shah NH, Dalman RL, Nead KT, Cooke JP, Leeper NJ (2016) The use of machine learning for the identification of peripheral artery disease and future mortality risk. *J Vasc Surg* 64(5):1515–1522. <https://doi.org/10.1016/j.jvs.2016.04.026>
98. Nadarzynski T, Miles O, Cowie A, Ridge D (2019) Acceptability of artificial intelligence (AI)-led chatbot services in healthcare: A mixed-methods study. *Digit Health* 5:2055207619871808. <https://doi.org/10.1177/2055207619871808>
99. Tranter-Entwistle I, Wang H, Daly K, Maxwell S, Connor S (2021) The Challenges of Implementing Artificial Intelligence into Surgical Practice. *World J Surg* 45(2):420–428. <https://doi.org/10.1007/s00268-020-05820-8>
100. Cobianchi L, Verde JM, Loftus TJ, Piccolo D, Dal Mas F, Mascagni P et al (2022) Artificial Intelligence and Surgery: Ethical Dilemmas and Open Issues. *J Am Coll Surg* 235(2):268–275. <https://doi.org/10.1097/XCS.0000000000000242>
101. Shipton L, Vitale L (2024) Artificial intelligence and the politics of avoidance in global health. *SSM Health Syst* 359:117274. <https://doi.org/10.1016/j.socscimed.2024.117274>
102. Eppler M, Ganjavi C, Ramacciotti LS, Piazza P, Rodler S, Checucci E et al (2024) Awareness and Use of ChatGPT and Large Language Models: A Prospective Cross-sectional Global Survey in Urology. *Eur Urol Oncol* 85(2):146–153. <https://doi.org/10.1016/j.eururo.2023.10.014>
103. Senthil R, Anand T, Somala CS, Saravanan KM (2024) Bibliometric analysis of artificial intelligence in

healthcare research: Trends and future directions.
Future Healthc J 11(3):100182. <https://doi.org/10.1016/j.fhj.2024.100182>

Hinweis des Verlags. Der Verlag bleibt in Hinblick auf geografische Zuordnungen und Gebietsbezeichnungen in veröffentlichten Karten und Institutsadressen neutral.