

TOP Ic Gesundheits-, Sozial- und ärztliche Berufspolitik: Aussprache zur Rede des Präsidenten und zum Leitantrag - Aktuelle Fragen der ärztlichen Berufsausübung

Titel: Qualitätsstandards für Studien zu KI-Anwendungen in der Medizin

Beschlussantrag

Von: Vorstand der Bundesärztekammer

Die Ärzteschaft fordert den Gesetzgeber, den Gemeinsamen Bundesausschuss (G-BA) und das Bundesinstitut für Arzneimittel und Medizinprodukte (BfArM), aber auch Fachgesellschaften, Forschungsinstitute und Hersteller von KI-Anwendungen auf, sich dafür einzusetzen, dass prospektive und methodisch hochwertige klinische Studien für KI-Anwendungen durchgeführt werden, wenn diese diagnostische oder therapeutische Entscheidungen unterstützen, beeinflussen oder vorbereiten. Insbesondere bei patientenrelevanten Auswirkungen müssen regelhaft auch randomisierte kontrollierte Studien durchgeführt werden.

Dabei ist insbesondere voranzusetzen, dass die verwendeten Daten die Zielpopulationen und Versorgungskontexte angemessen abbilden, nicht einseitig oder unvollständig sind und relevante Subgruppen einschließen. Verzerrungen, Einschränkungen der Übertragbarkeit sowie potenzielle Benachteiligungen einzelner Patientengruppen sollen systematisch geprüft, offengelegt und in der Bewertung berücksichtigt werden.

Für diese Studien sollen klare Anforderungen gelten:

- Vor ihrem Einsatz im medizinischen Bereich sollen KI-Systeme anhand unabhängiger Datensätze und möglichst auch in unterschiedlichen Versorgungskontexten geprüft und extern validiert werden. Aussagekräftig ist insbesondere eine Validierung im Rahmen prospektiver, kontrollierter klinischer Studien in Bezug auf den Behandlungserfolg.
- Die Funktionsweise der Künstlichen Intelligenz (KI) muss nachvollziehbar beschrieben werden. Dazu zählen insbesondere Angaben zu Art und Zweck des Modells, Einsatzkontext, Herkunft, Zusammensetzung und Qualität der Trainings-, Validierungs- und Testdaten, Version des Systems, Schnittstellen zur ärztlichen Anwendung sowie Art, Zeitpunkt und Häufigkeit von Updates.
- In den Studien soll untersucht und dargelegt werden, wie die KI in klinische Entscheidungsprozesse eingebunden ist, welchen Einfluss sie auf ärztliche Entscheidungen hat und welche Risiken durch Fehlinterpretation, Übervertrauen oder Automationsbias bestehen.

ANGENOMMEN

Angenommen: ☒ Abgelehnt: ☐ Vorstandsüberweisung: ☐ Entfallen: ☐ Zurückgezogen: ☐ Nichtbefassung: ☐

Stimmen Ja: 195

Stimmen Nein: 2

Enthaltungen: 8

- Etablierte Leitlinien zur Studienplanung und -berichterstattung (wie CONSORT-AI und SPIRIT-AI) sollen regelhaft angewendet und bei der Bewertung der Studien verbindlich berücksichtigt werden.
- Studien müssen patientenrelevante und klinisch aussagekräftige Endpunkte untersuchen, insbesondere Behandlungserfolg, Sicherheit für Patientinnen und Patienten, Kosteneffektivität sowie Auswirkungen (u. a. organisatorisch, zeitlich) auf Versorgungsabläufe in Praxis und Klinik. Die Studiendaten sollen so dokumentiert werden, dass sie unabhängig geprüft und mit anderen Studien verglichen werden können.

Zusätzlich sollen praxisnahe Handlungs- und Bewertungsleitfäden sowie standardisierte Prüfkriterien für Ärztinnen und Ärzte entwickelt werden, die helfen, Studien zu KI-Systemen besser zu verstehen und einzuordnen.

Begründung:

KI-Anwendungen finden mit großer Dynamik Eingang in Diagnostik, Therapieplanung und Versorgungssteuerung im Gesundheitswesen.

Eine im Fachjournal Nature Communications veröffentlichte Übersichtsarbeit zu Studien über KI-Anwendungen in der Medizin zeigt, dass viele dieser Studien wichtige Qualitätskriterien nur unvollständig erfüllen. Dazu gehören unter anderem:

- eine unklare oder unzureichende Beschreibung der eingesetzten KI,
- fehlende oder unvollständige Angaben zu den verwendeten Trainingsdaten und zur Aktualisierung des Systems,
- unklare oder wenig aussagekräftige Endpunkte sowie
- unvollständige Angaben zu Nebenwirkungen, Risiken und möglichen Fehlanwendungen.

Die Autorinnen und Autoren der Übersichtsarbeit schließen daraus, dass eine unzureichende Berichtsqualität die Übertragbarkeit der Studienergebnisse in die Versorgung erheblich einschränkt und einen belastbaren Vergleich verschiedener KI-Systeme kaum zulässt.

Es besteht das Risiko, dass bestimmte KI-gestützte Verfahren ohne hinreichend gesicherte Evidenz faktisch zum Versorgungsstandard werden, etwa weil sie stark beworben oder früh in Leitlinien aufgenommen werden. Eine solche Fehlentwicklung mit Auswirkungen auf das Patientenwohl wäre im Nachhinein nur schwer zu korrigieren.

Mit dieser Forderung trägt der 130. Deutsche Ärztetag 2026 dazu bei, die Patientensicherheit zu stärken, insbesondere vulnerable Patientengruppen zu schützen,

ANGENOMMEN

Fehlanwendungen von KI-Systemen zu verhindern und gleichzeitig solche KI-Innovationen zu fördern, die im Sinne des Patientenwohls auf einer soliden, transparenten, unabhängigen und methodisch sauberen Evidenzbasis stehen.

ANGENOMMEN