

Leitfaden für sichere KI-Medizinprodukte: Dresdner Forschende rücken menschliche Faktoren in den Fokus

KI-gestützte Medizinprodukte versprechen eine bessere medizinische Versorgung und Unterstützung von Gesundheitsfachkräften. Sicherheit und Leistungsfähigkeit solcher Systeme hängen jedoch nicht allein vom passenden Algorithmus oder den technischen Eigenschaften ab. Ebenso entscheidend ist, wie Menschen diese Geräte und Anwendungen nutzen. In einer aktuellen Veröffentlichung in der Fachzeitschrift NEJM AI analysiert ein Forschungsteam um Prof. Stephen Gilbert vom Else Kröner Fresenius Zentrum (EKFZ) für Digitale Gesundheit an der TU Dresden (TUD) systematisch Risiken, die bei der Interaktion zwischen Mensch und KI entstehen können und formuliert Empfehlungen für Hersteller und Prüfstellen.

Die Autorinnen und Autoren zeigen, dass bestehende regulatorische Vorgaben bei der Zulassung viele dieser durch Menschen verursachte Risiken bislang nur teilweise berücksichtigen. Dadurch können Lücken entstehen, die sich auf die Sicherheit und Qualität der Versorgung auswirken. Um diese zu schließen, identifizieren die Forschenden sieben zentrale Risiken und entwickeln praktische Handlungsempfehlungen, die sich in bestehende regulatorische und dokumentarische Prozesse integrieren lassen.

Risiken im Umgang mit KI-Systemen

KI-basierte Medizinprodukte können in verschiedenen klinischen Bereichen eingesetzt werden. In der Radiologie unterstützen sie beispielsweise die Erkennung von Krebserkrankungen. Darüber hinaus helfen klinische Entscheidungsunterstützungssysteme dabei, personalisierte Therapien für Patient:innen auszuwählen. KI kann zudem Echtzeitüberwachungs- und Frühwarnsysteme sowie Chatbots für Anwendungen wie die Patientenkommunikation und Software unterstützen, die automatisch medizinische Berichte erstellt oder Befunde zusammenfasst. Im Mittelpunkt der Analyse stehen Risiken, die im praktischen Umgang mit solchen KI-Systemen auftreten können. Dazu gehört beispielsweise eine erhöhte Wahrscheinlichkeit, dass Ergebnisse aufgrund der teilweise intransparenten Natur von KI-Systemen missverstanden oder fehlinterpretiert werden. Probleme können auch entstehen, wenn das Vertrauen in die Anwendung nicht angemessen ist: Menschen verlassen sich entweder zu stark auf die KI-Unterstützung oder ignorieren relevante Hinweise. Die Forschenden weisen zudem auf das Risiko eines Automatisierungsbias hin: die Tendenz, Empfehlungen unkritisch zu übernehmen, wodurch Fehler übersehen oder eigenständige Urteile vernachlässigt werden können. Weitere mögliche Risiken umfassen einen Kompetenzverlust (Deskilling), Technostress bei Nutzer:innen, eine ungeprüfte Ausweitung der Einsatzgebiete über den ursprünglich vorgesehenen Bereich hinaus sowie Fehler, die durch Änderungen an Systemen und unterschiedliche Betriebsmodi entstehen. Solche Faktoren können im klinischen Alltag zusätzliche Belastungen oder unerwartete Fehlerquellen schaffen – selbst dann, wenn die technische Leistung eines Systems an sich gut ist.

Ein praxisnaher Leitfaden für Hersteller und Prüfstellen

Für ihre Analyse wertete das Forschungsteam bestehende Standards zur Benutzerfreundlichkeit und Sicherheit, regulatorische Leitlinien sowie wissenschaftliche Literatur zu KI im Gesundheitswesen

aus. Ergänzend flossen Expertendiskussionen aus den Bereichen klinische Anwendung, Regulierung und menschlicher Faktoren ein. Das Ergebnis ist ein praxisorientierter Leitfaden, der mit sieben Empfehlungen eine Lücke in den aktuellen Standards schließt. Diese sollen Hersteller und Prüfstellen sowohl vor als auch nach der Markteinführung eines Produkts unterstützen, damit KI-spezifische Risiken im Zusammenspiel mit menschlichen Nutzer:innen frühzeitig erkannt und systematisch adressiert werden können.

Der Leitfaden empfiehlt, KI-Medizinprodukte so zu entwickeln und einzusetzen, dass klar definiert ist, wer sie nutzt, in welchem Bereich sie eingesetzt werden und welche Aufgaben Mensch und System jeweils übernehmen. Zudem sollen Ergebnisse verständlich dargestellt, in bestehende klinische Abläufe integriert und durch erforderliche Schulungen sowie sichere Ausweichlösungen bei Systemausfällen ergänzt werden. Darüber hinaus betonen die Autor:innen die Bedeutung kontinuierlicher Prüfung nach der Markteinführung. Nutzungsmuster, mögliche Fehlanwendungen oder Übervertrauen in KI-Systeme sollten systematisch beobachtet und bei Bedarf korrigiert werden. Auch Änderungen an den Systemen müssten transparent kommuniziert werden, damit Arbeitsprozesse entsprechend angepasst werden können.

Die Empfehlungen sind bewusst allgemein, aber an den regulatorischen Vorschriften orientiert formuliert, damit sie für unterschiedliche KI-gestützte Medizinprodukte und Anwendungsszenarien genutzt werden können. In einem nächsten Schritt möchten die Forschenden ihre Empfehlungen anhand konkreter Pilotanwendungen mit KI-gestützten Medizinprodukten überprüfen und weiterentwickeln. Langfristig sollen menschliche Einflussfaktoren systematisch in Regulierung und Bewertung KI-basierter Gesundheitstechnologien miteinbezogen werden – mit dem Ziel, vermeidbare Risiken zu reduzieren und sichere Innovationen in der Medizin zu fördern.

Die Veröffentlichung wurde von Forschenden der TUD (EKFZ für Digitale Gesundheit, Professur für Technisches Design und Fakultät Wirtschaftswissenschaften) in Zusammenarbeit mit Expert:innen der University of Oxford (Vereinigtes Königreich) und des Geneva University Hospital (Schweiz) verfasst.

Publikation

Rebecca Mathias, Anne Schmitt, Mateo Campos, Baptiste Vasey, Sebastian Lorenz, Peter McCulloch, Stephen Gilbert: Evaluation of Human Factors-Related Risks in AI-Enabled Medical Devices: A Practical Guide, NEJM AI, 2026. Link: <https://ai.nejm.org/doi/full/10.1056/AIpc2501297>

Else Kröner Fresenius Zentrum (EKFZ) für Digitale Gesundheit

Das EKFZ für Digitale Gesundheit an der Medizinischen Fakultät der Technischen Universität Dresden (TUD) und dem Universitätsklinikum Carl Gustav Carus Dresden wurde im September 2019 gegründet. Es wird mit einer Fördersumme von 40 Millionen Euro für eine Laufzeit von zehn Jahren von der Else Kröner-Fresenius-Stiftung gefördert. Das Zentrum konzentriert seine Forschungsaktivitäten auf innovative, medizinische und digitale Technologien an der direkten Schnittstelle zu den Patientinnen und Patienten. Das Ziel ist dabei, das Potenzial der Digitalisierung in der Medizin voll auszuschöpfen, um die Gesundheitsversorgung, die medizinische Forschung und die klinische Praxis nachhaltig zu verbessern.

Originalpublikation:

<https://ai.nejm.org/doi/full/10.1056/AIpc2501297>