KI 01.10.2026

Künstliche Intelligenz in der medizinischen Dokumentation: Potenziale, Risiken und ökologische Implikationen



Die klinische Dokumentation hat sich in den vergangenen Jahrzehnten zu einer der zeitintensivsten Tätigkeiten im ärztlichen Berufsalltag entwickelt. Beobachtungsstudien im ambulanten Setting zeigen, dass ein erheblicher Anteil der Arbeitszeit nicht auf die direkte Patientenversorgung, sondern auf administrative Tätigkeiten entfällt. Als Konsequenz berichten viele Ärztinnen und Ärzte von reduziertem Blickkontakt mit Patientinnen und Patienten, vermehrten Copy-Paste-Fehlern sowie einem erhöhten Burnout-Risiko – ein Phänomen, das auch als „Pajama Time“ beschrieben wird, da die Nachbearbeitung von Dokumentation häufig erst nach Dienstende stattfindet.

Vor diesem Hintergrund rücken KI-gestützte Dokumentationssysteme zunehmend in den Fokus. Diese versprechen, durch automatische Spracherkennung, Strukturierung medizinischer Inhalte und Codierungsunterstützung eine signifikante Entlastung herbeizuführen. Der vorliegende Beitrag analysiert kritisch den aktuellen Evidenzstand, diskutiert systeminhärente Risiken und lenkt den Blick auf die ökologischen Kosten des KI-Einsatzes im Gesundheitswesen.

Funktionsweise KI-gestützter Dokumentationssysteme

Moderne KI-Dokumentationslösungen lassen sich in vier funktionale Kernbereiche unterteilen:

Sprache → Text

Diktat oder das gesamte Arzt-Patienten-Gespräch wird in ein Rohtranskript umgewandelt. Fortgeschrittene Systeme leisten dabei Sprecher:innen-Trennung und Zeitstempelung.

Text → Struktur

Aus dem Freitext werden standardisierte Dokumentformate generiert (z. B. SOAP-Noten, OP-Berichte, Arztbriefe), inklusive Problem- und Medikamentenlisten sowie klinisch relevanter Negativa.

Struktur → Codes

Das System schlägt automatisch ICD-, OPS- und DRG-Codes vor, führt Plausibilitätsprüfungen durch und generiert sogenannte ‚Coding Hints' zur Optimierung der Vergütung.

Qualität und Compliance

Checklisten (z. B. Sepsis-Bundle, OP-Site-Verification), Warnhinweise auf Allergien und Wechselwirkungen sowie stilistische Optimierungen für klarere und lesbarere Texte runden das Funktionsspektrum ab.

Evidenzlage: Was sagen die Studien?

Die klinische Evidenz zu AI Scribes ist noch jung, aber wächst rasch. Ein randomisiertes kontrolliertes Design (RCT, Laufzeit November 2024 bis Januar 2025) verglich zwei kommerzielle AI-Scribe-Systeme (Microsoft DAX und Nabla) mit üblicher Dokumentation (Usual Care). In einer großen Real-World-Kohorte (238 Ärztinnen und Ärzte, ca. 72.000 Begegnungen) zeigte Nabla eine Reduktion der Dokumentationszeit um circa 10 % gegenüber der Kontrollgruppe, Burnout- und Stresswerte verbesserten sich in Teilen.

Besonders eindrücklich sind Daten aus dem Bereich der Notaufnahme: Die Schreibzeit für Entlassbriefe sank im Median von 69,5 Sekunden auf 32,0 Sekunden unter LLM(Large Language Models)-Assistenz. In einem hochfrequenten Setting wie der Notaufnahme stellt eine Halbierung der Dokumentationszeit eine klinisch relevante Entlastung dar [4].

Zusammenfassend gilt: KI spart nachweislich Zeit – unter der Bedingung, dass Workflow-Integration, Datenschutzkonformität und die notwendige Nachbearbeitung sichergestellt sind.

Klinische Risiken

Halluzinationen: KI-Systeme können plausibel klingende, aber faktisch falsche Inhalte generieren. Dies stellt im medizinischen Kontext ein erhebliches Patientensicherheitsrisiko dar.

Omission Bias: Relevante klinische Informationen können in der Dokumentation fehlen, wenn sie im Gespräch nicht explizit geäußert wurden.

Bias bei sprachlicher oder ethnischer Diversität: Die Qualität der generierten Dokumentation kann je nach Sprache, Dialekt und Patientengruppe signifikant variieren, was zu Ungleichheiten in der Versorgungsqualität führen kann.

Rechtliche und datenschutzrechtliche Risiken

Audiodaten, Freitexte und Diagnosen zählen zu den sensibelsten Datenkategorien im Sinne der DSGVO. Ein lückenloser Audit-Trail ist unverzichtbar, um nachvollziehen zu können, wer welche Inhalte wann erstellt oder verändert hat. Der Grundsatz ‚KI hat es geschrieben' stellt keinen Haftungsausschluss dar – die rechtliche und fachliche Verantwortung für die Dokumentation verbleibt bei der unterzeichnenden Ärztin bzw. dem unterzeichnenden Arzt.

Pragmatische Leitlinie: KI erstellt den Entwurf – die Verantwortung liegt beim Menschen. Im Zweifelsfall gilt: 30 Sekunden zusätzliche Prüfung sind besser als 30 Tage Erklärungsbedarf.

Der ökologische Fußabdruck von KI

Ein oft vernachlässigter Aspekt des KI-Einsatzes ist dessen ökologische Dimension. KI-Systeme sind keine immateriellen Werkzeuge: Hinter jeder Anfrage stehen Rechenzentren, Netzinfrastruktur, Kühlsysteme und Halbleiterproduktion.

Energieverbrauch

Laut einer Studie des Öko-Instituts im Auftrag von Greenpeace (2025) wird der Strombedarf von KI-Rechenzentren bis 2030 auf rund das 11-Fache des Werts von 2023 ansteigen [2]. Die Internationale Energieagentur (IEA) prognostiziert für Rechenzentren weltweit einen Strombedarf von ca. 945 TWh bis 2030 – in fortgeschrittenen Volkswirtschaften könnten Rechenzentren für mehr als 20 % des gesamten Strombedarfszuwachses verantwortlich sein [3].

CO_2-Emissionen

Die globalen CO2-Äquivalent-Emissionen aus Rechenzentren stiegen laut derselben Studie von 213 Megatonnen (2023) auf projizierte 356 Megatonnen (2030). Der KI-spezifische Anteil erhöht sich voraussichtlich von 29 auf 166 Megatonnen – nahezu

eine Versechsfachung. Selbst bei einem vollständig CO2-neutralen Strommix bis 2040 wird erwartet, dass Emissionen zunehmen und fossile Kraftwerke länger in Betrieb bleiben.

Wasserverbrauch

Der globale Wasserverbrauch von Rechenzentren soll sich von 175 Milliarden Litern (2023) auf 664 Milliarden Liter (2030) nahezu vervierfachen. Treiber sind erhöhter Kühlbedarf durch stetig steigende Rechenleistung sowie indirekter Wasserverbrauch durch Stromerzeugung und Chipproduktion. Forschungsarbeiten zum Water Footprint von KI-Modellen schätzen, dass allein das Training von GPT-3 direkt bis zu 700.000 Liter Frischwasser verdunsten ließ [5].

Empfehlungen für einen verantwortungsvollen KI-Einsatz

Auf Basis der vorliegenden Evidenz und der dargelegten Risiken werden folgende fünf Handlungsempfehlungen formuliert:

1. Nutzenprüfung vor Implementierung:
   KI sollte ausschließlich dort eingesetzt werden, wo ein messbarer Nutzen für Patientensicherheit oder Arbeitslastreduzierung nachgewiesen ist.
2. Präferenz für ressourceneffiziente Modelle:
   Kleinere, destillierte Modelle sowie Retrieval-Augmented-Generation-Ansätze und konsequentes Caching können den ökologischen Fußabdruck erheblich reduzieren.
3. Obligatorisches Human-in-the-Loop-
   Prinzip:
   Jede KI-generierte Dokumentation muss von einer qualifizierten Person geprüft, verantwortet und signiert werden. Versionierung und Audit-Trail sind Pflicht.
4. Umweltkriterien bei der Beschaffung:
   In Vergabeverfahren sollten Transparenzanforderungen zu Strommix, Power Usage Effectiveness (PUE), Water Usage Effectiveness (WUE), Standort und Abwärmenutzung verankert werden.
5. Kontinuierliches Monitoring und Reporting:
   Einrichtungen sollten KPIs (Key Performance Indicators) zu Dokumentationszeit, Fehlerquoten sowie Energie- und Wasserverbräuchen erheben und berichten.

Deepfake Medical Imaging: Eine neue klinische Bedrohung

Eine besonders beunruhigende Entwicklung stellt der Einsatz generativer KI zur Erstellung gefälschter medizinischer Bildgebung dar. Eine kürzlich im Fachjournal Radiology veröffentlichte Studie [1] untersuchte systematisch, ob Radiologen und aktuelle KI-Modelle in der Lage sind, KI-generierte Röntgenbilder von echten Aufnahmen zu unterscheiden.

In einem zweiarmigen Studiendesign wurden 77 durch ChatGPT-4o synthetisch erzeugte Radiographien mit authentischen Röntgenbildern verglichen (Datensatz 1). Danach wurden 55 durch RoentGen, (organ-spezifisches Diffusionsmodell für Thoraxröntgenbilder) erzeugte Radiographien mit 55 authentischen Röntgenbildern verglichen (Datensatz 2).

17 Radiologinnen und Radiologen erkannten bei der Betrachtung von Datensatz 1 lediglich mit einer Wahrscheinlichkeit von 41 % die Medical Deepfakes. Die gleichen 17 Radiologinnen und Radiologen sowie vier Large Language Models (GPT-5, GPT-4o, Gemini 2.5 Pro, Llama 4 Maverick) sollten dann bei Vorlage des Datensatzes 2 echte von generierten Aufnahmen unterscheiden. Die diagnostische Trennschärfe aller Prüfer inklusive der KIs lag deutlich über dem Zufallsniveau (75 %), wies jedoch eine bemerkenswerte Variabilität auf.

Als diagnostischer Hinweis auf KI-generierte Traumabilder identifizierten die Autoren unnatürlich glatte, konsistente Frakturlinien sowie unvollständige kortikale Destruktionsmuster Modelle – Merkmale, die realen Frakturen nicht entsprechen. Die Studie unterstreicht den dringenden Bedarf an robusten Verifikationsverfahren für medizinisches Bildmaterial und einer kritischen Haltung gegenüber digital übermittelten diagnostischen Befunden.

Fazit

KI-gestützte Dokumentationssysteme bieten ein erhebliches Potenzial zur Entlastung klinisch tätiger Ärztinnen und Ärzte. Die Evidenz zu AI Scribes ist vielversprechend, aber noch begrenzt. Klinische Risiken – insbesondere Halluzinationen und Omission Bias – sowie rechtliche Verantwortlichkeiten erfordern ein konsequentes Human-in-the-Loop-Prinzip (Der Mensch bleibt bei der Verwendung von KI alleinig in der Verantwortung).

Der ökologische Fußabdruck von KI-Systemen wird in der medizinischen Fachdiskussion bislang unzureichend berücksichtigt. Angesichts prognostizierter Steigerungen des Strom- und Wasserverbrauchs ist eine strategische Auseinandersetzung mit Effizienz und Ressourcenverantwortung im Gesundheitswesen geboten. Das aufkommende Phänomen der Deepfake Medical Imaging verlangt nach neuen Verifikationsstandards und erhöhter kritischer Kompetenz aller klinisch Tätigen.

Die größte Gefahr von KI ist nicht, dass sie Fehler macht – sondern dass wir ihr glauben. Verantwortung BLEIBT menschlich.

Literatur

[1] *Tordjman M et al. The rise of deepfake medical imaging. Radiology. 2026 Mar.*

[2] *Öko-Institut im Auftrag von Greenpeace. Treibhausgasemissionen und Wasserverbrauch durch KI-Rechenzentren. 2025.*

[3] *International Energy Agency (IEA). Energy and AI. 2024/2025.*

[4] *Lukac P et al. Ambient AI Scribes in Clinical Practice: A Randomized Trial. DOI: 10.1056/aioa2501000. 2025 Dec*

[5] *Water Footprint of Large Language Models: Freshwater consumption during training and inference.*

Goller R, Mautner S, Lorenz R: Künstliche Intelligenz in der medizinischen Dokumentation: Potenziale, Risiken und ökologische Implikationen. Passion Chirurgie. 2026 Oktober; 16(10): Artikel 09.

Autor:innen des Artikels



Dr. med. Rudolf Goller, FEBS-AWS

Katharina Klinik Zerbst/Anhalt
Zentrum für Chirurgie

Kontakt





Dr. med. Ralph Lorenz

1. Vorsitzender des BDC LV|Berlin
Stellv. Sprecher der BDC-Landesverbandsvorsitzenden
Universitätsklinikum Brandenburg an der Havel
Hernienzentrum 3+CHIRURGEN
Berlin

Kontakt