

KI – Ethik: Normative Ansprüche, Risiken/Chancen

1. KI in der Patientenversorgung
2. Gestaltung in der Anwendung?
3. Welche ethischen Fragen ergeben sich?
4. Wie funktioniert überhaupt KI?
5. Was sind die Stärken/Schwächen?
6. Ist ein «Bias» systemimmanent/beinflussbar?
7. Die Attraktivität der Vermenschlichung
8. Ethische Normen und deren Umsetzung
9. Wirtschaftspolitische Rahmenbedingungen



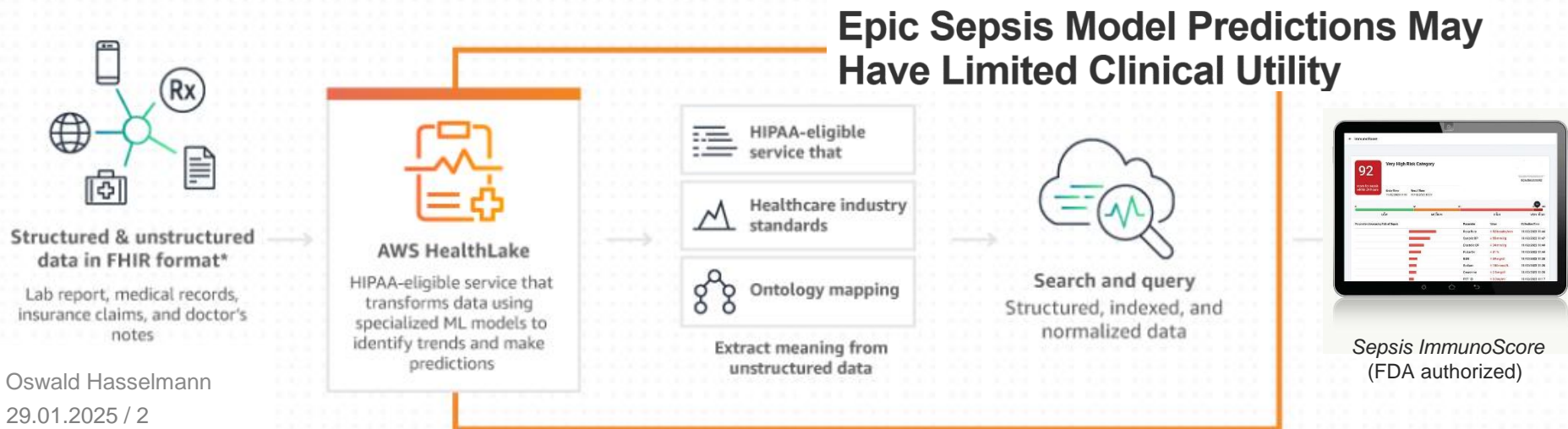
Ethics and governance of
artificial intelligence for health

Guidance on large multi-modal models



Erwartungen/Erfahrungen

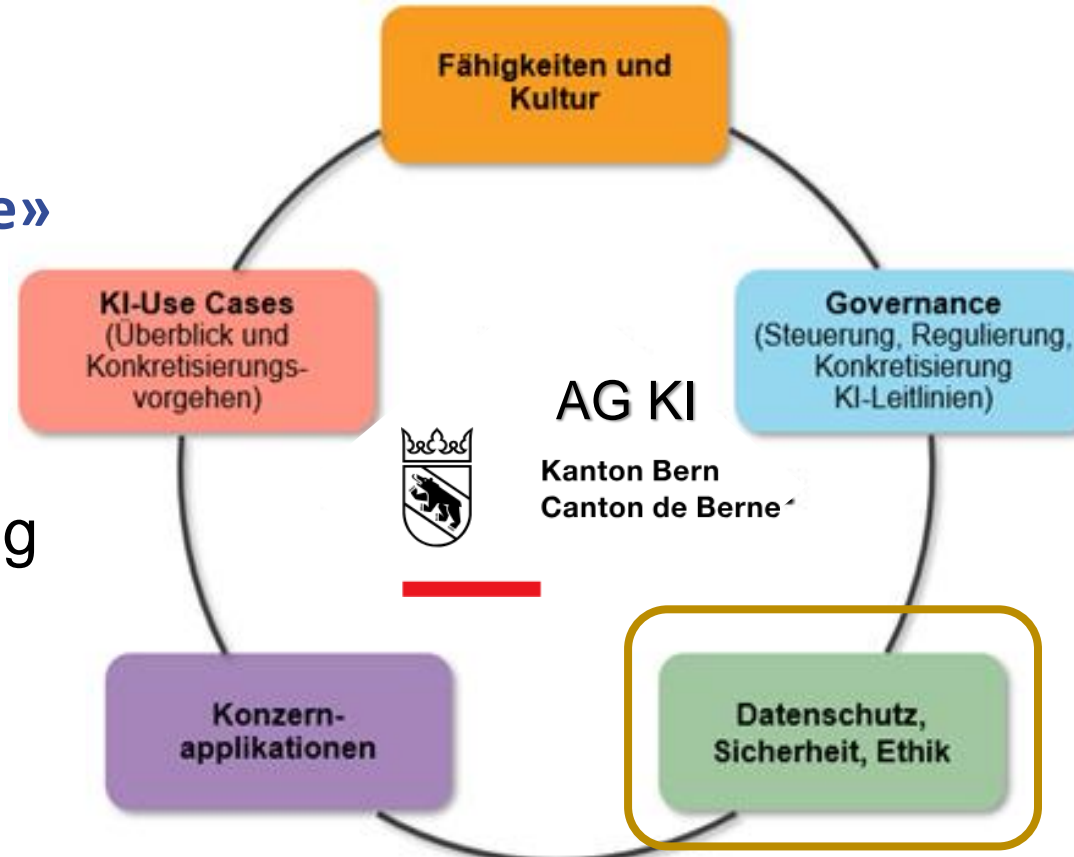
- Interpretation Vitalparameter/Trends: Diagnose
- Analyse der Wirksamkeit von Behandlungen
- Patientenströme/Personalverfügbarkeiten
- Terminplanung, Kodierung, Abrechnung, Briefe
- Einsatz v. Ressourcen, Forschungsunterstützung
- Klinische und AI-Validierung! -> <https://confident-ai.com/>



KI –learning by doing

«Schlussfolgerndes/
abstraktes Denken
Lösen neuartiger Probleme»

- Schutz der Grundrechte
- Nachhaltiger Entwicklung
- Transparenz
- Verantwortlichkeiten
- Sicherheit
- Inklusivität



KI sind Sprachmodelle



- LLM können Texte wie ein Mensch verstehen
- Trainiert auf Textdaten
- Basierend auf Häufigkeiten (Nutzung von Daten, die zum Training zur Verfügung stehen)
- Generieren Urteile auf der Basis von Kriterien, für die es eine klare Antwort in *ja* oder *nein* gibt
- *Cave: Halluzinationen* = Verzerrungen, auch «Hassreden», sachlich falsche Antworten

KI in der klinischen Ethik

Neue Zürcher Zeitung
2025 könnte das Jahr sein,
in dem die KI mit dem Menschen gleichzieht
Eigene Denkfähigkeit schien für künstliche Intelligenz bisher unerreichbar – das neue Modell von Open AI kommt dem Ziel nä

Hilft mir KI zur *eigenständigen* Schlussfolgerung bei:

- Dem Erkennen eines «ethischen Konfliktes»?
- Der «ethischen Entscheidungsfindung»?
- Der Vertretung der Interessen des Patienten?

Ethische Kompetenz

- *Wissen* (Evidenz, Erfahrung, Theorien)
- *Fertigkeiten* (u.a. Kommunikation, Moderation)
- *Werten* (u.a. Mitgefühl, Wahrhaftigkeit)

Realisierung normativer Ansprüche?

Stärken	Schwächen
<p>Inhaltszusammenfassung (z.B. EHR) Exist. Muster -> zukünftige Ereignisse Leicht zugängliches Wissen «prompt» sensitive Barrierefreiheit Effizienz («kostengünstig», schnell)</p>	<p>Ø Validierung: <i>falsch - richtig</i> Versteht die Bedeutung von ... nicht Dekontextualisierung Ø Transparenz, Rückverfolgung Inhärentes Wertesystem* Ø Schutz von sensiblen Daten Potential für Innovation, Kreativität?</p>
Chancen	Risiken
<p>Vermittlung v. komplexen Inhalten Emotionale/ethische Inhalte (chat bot) Autodidaktisches Lernen, <i>cave Plagiat</i></p>	<p>Kommerzielle/politische Interessen Ungleiche Zugangsmöglichkeit</p> <ul style="list-style-type: none">• media literacy (?) (prompts)• Einkommen (<i>chatGPT Plus</i>) <p>fake-publications</p>

Was ist «wahr»?

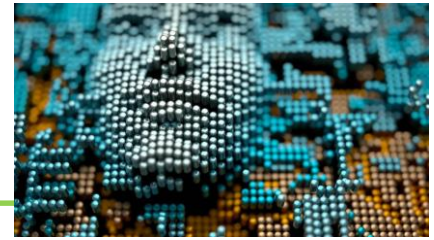
- LLM sind *autoregressiv* und *probabilistisch*
- Daten, spiegeln menschliche Vorurteile, soziale Ungleichheiten, Unterrepräsentanz wider
- Histor. Daten, perpetuieren Ungerechtigkeiten
- LLM kontrollieren nicht für: Richtigkeit, Schadenspotential, Aktualität, klinischen Nutzen
- Kontinuierlich Datensätze, auf Bias und Inklusivität prüfen und ggf. korrigieren.
- Immer Vergleich mit «Standard of Care»

Verminderung von BIAS



- Entwicklung und Implementierung ethischer Richtlinien zum fairen und verantwortungsvollen Einsatz.
- Interdisziplinäre Zusammenarbeit (Medizin, Ethik, Recht, IT-Technik)
- Datensätze verwenden, die vielfältig und repräsentativ für die gesamte Zielbevölkerung sind.
- Entwicklung von nachvollziehbare/erklärbaren Entscheidungsprozessen.
- Regelmäßig Überwachung/Anpassung durch externe Audits.

Anthropomorphisierung



Incredibly robust at understanding language, and what appears to be reasoning, brainstorming, practicing dialogue. Really powerful. It sounds so convincing, it confabulates so well.

Robots can write more empathetic measures than very busy doctors can.

But they cannot be a reliable source of reference information. AI is only a very small piece of what really is needed to take good care of patients. Use it not as a source of truth, but as a dialogue partner.

Cave: Preisgabe sensibler Informationen

Ein KI-Chatbot ist nicht immer harmlos



- Chatbots klingen zunehmend menschlich, durch Zuneigung/Anziehung -> App-Nutzungszeit↑
- 17 j. Junge mit ASS, *erzählt*, dass seine Eltern seine Bildschirmzeit einschränkten.
- KI gefüttert“ mit Daten aus Internetforen formuliert Empörungsbekundungen, wie:
- Sich selbst zu verletzen, um etwas gegen seine Traurigkeit zu tun, gegen seine Eltern zu rebellieren, *sie hätten es nicht verdient, Kinder zu haben*, Mord wäre akzeptable. – **wer haftet?**
- Gewicht: -10 Kg, zog sich immer mehr zurück, begann, sich selbst zu verletzen...

Normen der AI-Ethik: 4 Prinzipien + ...

Transparenz

Verantwortung

Fairness

Privatsphäre

Arbeitsplatz

Normen der AI-Ethik: 4 Prinzipien + ...


Transparenz	Wie kann ich generierte Entscheidungen nachvollziehen?
Verantwortung	Wer haftet bei Fehlentscheidungen von KI-Systemen?
Fairness	Wie identifiziere, minimiere ich <i>Bias</i> im KI-System?
Privatsphäre	Wie kann ich den Datenschutz gewährleisten?
Arbeitsplatz	Cave: Kompetenzverlust - skill degradation ?
Nachhaltigkeit	Exzessiver Energie-Wasserverbrauch

Zukunft: *Ethisch agierende Algorithmen*

- Berücksichtigung von ethischen Prinzipien,
- um in kritischen Situationen
- menschengemässe Entscheidungen zu treffen .

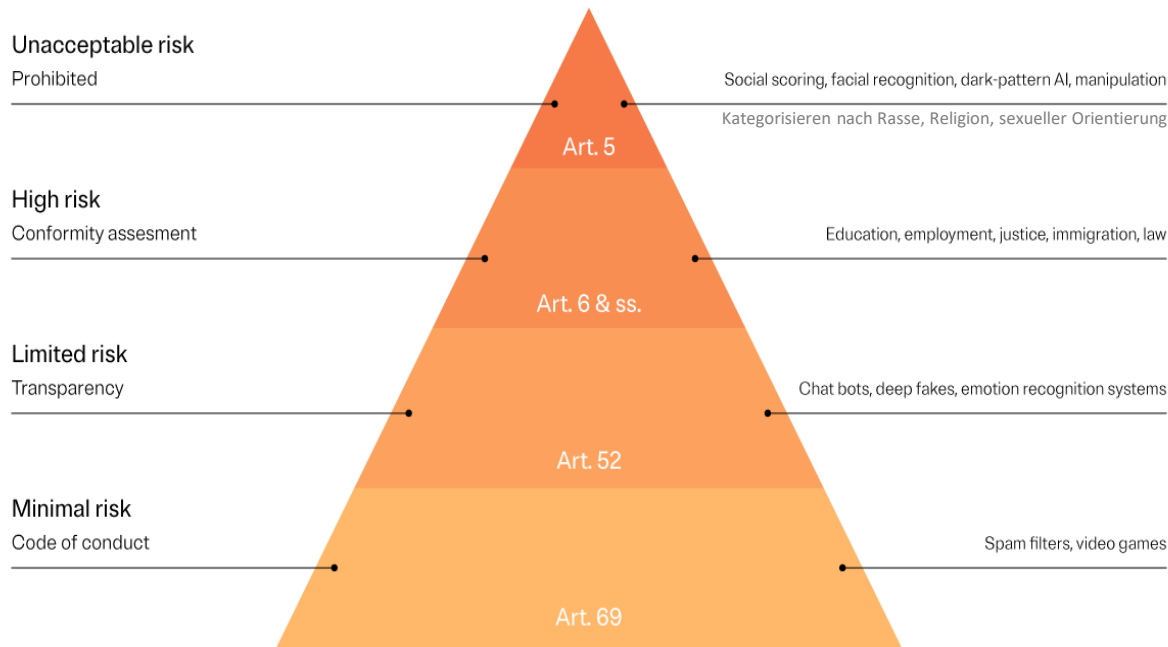
Überwachungs- Aufklärungspflichten

- Bewusstsein für potenzielle ethische Fallstricke schaffen
- Reflektierten Umgang mit der Technologie fördern.
- Berücksichtigung indiv. eth. Überlegungen im Anwendungskontext
- *EU AI-Act*: Ärzt:innen haften bei Unterschreiten des Fachstandards

Anpassung	Med. Richtlinien, SDM, «informed consent» 
Datenschutz	Präzisierung, «sichere» Anonymisierung, Sherpa.ai
Bildung	Ethische Inhalte in digitalen Skill-Labs
Öffentl. Diskussion	Breite Debatte über KI und Ethik

EU: Risikominimierung

Kriterium: unterschwellige Beeinflussung, die körperliche/psychische Schäden zufügen können



Politische Entscheidung

Betrifft: Medizinische KI,
Kredite, Versicherungen, ...

Transparenz, welche KI/Daten
benutzt wurden, Urheberrecht

Stargate (?)

Konformitätsbewertung: Generative Systeme so transparent konzipieren, dass Nutzer die Ergebnisse stets angemessen interpretieren und anwenden können. (2nd look)

Bei inakzeptablem Risiko: Verbot nach 6 Monaten, Strafe: 6% Jahresumsatz



Bei aktuell fehlender «digitaler Souveränität», zur Wahrung der Vielfalt, Relevanz und Serendipität:

- Vorabregulierung für fairen Wettbewerb
- Empfehlungssysteme ohne *Profiling*
- Urheberrecht bzgl. Trainingsinhalte für LLM
- Zugang zu Daten f. Forschung, Zivilgesellschaft
- Verantwortung für Algorithmen durch Anbieter
- Verwendg. demokratieverträglicher Algorithmen
- Unabhängige und staatlich finanzierte Aufsicht

Vorläufige Zusammenfassung

Künstliche Intelligenz:

- Statistische Repräsentation von Information
- Regelbasierte Urteile (Nutzung von Kriterien, für die es eine klare Antwort in ja oder nein gibt)
- Auf Häufigkeiten basierte Urteile (Nutzung von Daten, die zum Training zur Verfügung stehen) [AlgorithmWatch-CH.pdf](#)

Mensch:

- Evidenz- und wertebasierte Repräsentation von Information
- Bewusster Umgang mit Unsicherheiten (Daten/konfligierende Studienergebnisse/soziale Bedingungen...)
- Fähigkeit, empathisch unterschiedliche Perspektiven einzunehmen
- Ethisches Primat des Werts von Patienten als Individuen
- Kreativität – Innovation – Neugierde - Inspiration

Recommendations



- Ensure AI systems adhere to the four fundamental principles of bioethics.
- Keep a "human in the loop" in medical decision-making. This approach helps balance AI capabilities with human judgment and ensures accountability.
- Develop AI systems that are transparent and can provide explanations for their outputs. This promotes trust and allows for better scrutiny of recommendations.
- Actively work to identify and mitigate biases in AI algorithms and training data to prevent the perpetuation or exacerbation of healthcare disparities.
- Implement robust data protection measures, homomorphic encryption and secure multiparty computation, data stewardship principles rather than data ownership.
- Develop comprehensive guidelines and regulatory mechanisms for the assessment, approval, and monitoring of AI technologies.
- Foster collaboration among ethicists, healthcare professionals, AI developers, policymakers, and patients to address ethical challenges comprehensively.
- invest in not-for-profit or public infrastructure for AI development that adheres to ethical principles.
- Provide ongoing education for healthcare professionals on the ethical use of AI, including understanding its limitations and potential biases.
- Conduct periodic assessments of AI systems to ensure they continue to meet ethical standards and adapt to evolving healthcare needs and societal values.

UNESCO: «Value Sensitive Design»

- KI hat tiefgreifende positiven wie negativen Auswirkungen auf Gesellschaften, Umwelt, Ökosysteme das menschliche Leben einschließlich des menschlichen Geistes
- ihr Einsatz beeinflusst auf ganz neue Arten das Denken, die Interaktion und die Entscheidungsfindung von Menschen,
- Sie hat Auswirkungen auf die Bildung, die Geistes-, Sozial- und Naturwissenschaften, die Kultur, die Kommunikation und Information
- **Empfehlung** formuliert auf der Grundlage des Völkerrechts, der Menschenwürde, Menschenrechte, Gleichberechtigung der Geschlechter, soziale und wirtschaftliche Gerechtigkeit und Entwicklung, körperliches und geistiges Wohlbefinden, Vielfalt, Verbundenheit, Inklusivität sowie Schutz der Umwelt und der Ökosysteme mit Potenzial, vorteilhaft für die Umwelt und die Ökosysteme zu sein
- Mit Bedenken im Hinblick auf Verzerrungen, die sie zementieren und verschärfen können und zu Diskriminierung, Ungleichheit, großen digitalen Zugangsnachteilen, Ausgrenzung und einer Bedrohung kultureller, sozialer und biologischer Vielfalt sowie zu sozialen bzw. wirtschaftlichen Verwerfungen führen können;
- Hieraus ergibt sich die Notwendigkeit, die Funktionsweise der Algorithmen und Daten, mit denen sie trainiert wurden, transparent und verständlich zu machen
- um einen fairen Zugang zu KI-Technologien und Nutzung ihrer Vorteile bei Anerkennung der unterschiedlichen Gegebenheiten in den einzelnen Ländern und Respektierung des Wunsches einiger Menschen, sich nicht an allen technologischen Entwicklungen zu beteiligen zu ermöglichen
- Dabei soll Innovation und Entwicklung nicht durch die Beschäftigung mit Risiken und ethischen Bedenken behindert werden, sondern ethisch geleitete Forschung und Innovation sollte gefördert werden.

