Ethische Herausforderungen von künstlicher Intelligenz und digitalen Tools in der Psychotherapie

Ethik-Tag UPK
19. Juni 2025

Dr. Jan Schürmann MA, Abteilung Klinische Ethik
Universitätsspital Basel
Universitäre Psychiatrische Kliniken Basel
Universitäts-Kinderspital beider Basel
Universitäre Altersmedizin FELIX PLATTER

Gliederung

1 | Kurze Einführung zur Künstlichen Intelligenz (KI) in der Medizin

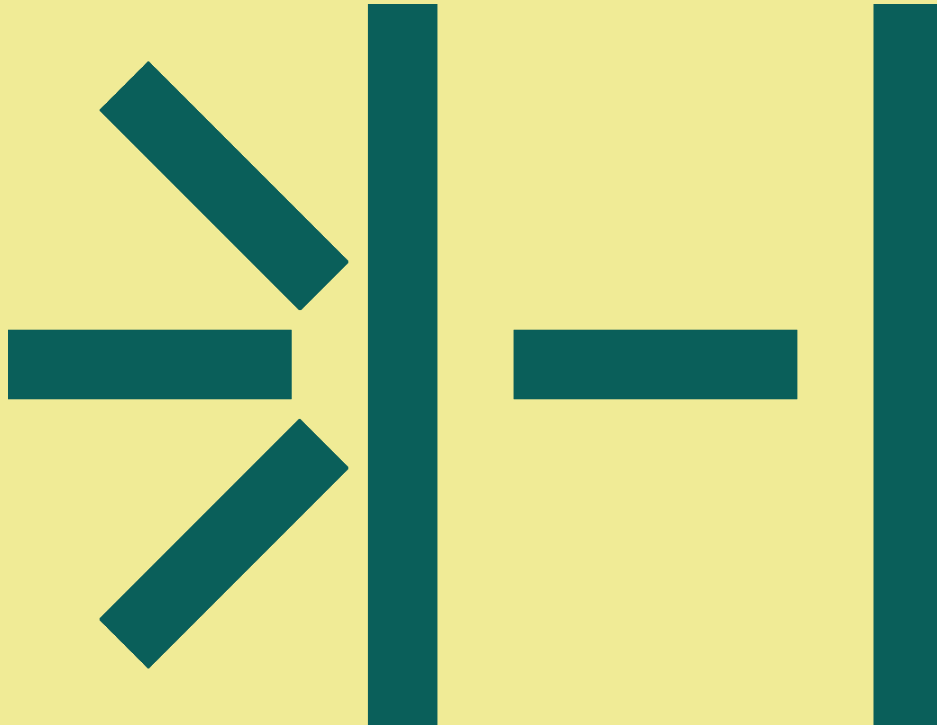
2 | Anwendungsbereiche von KI für die psychosoziale Gesundheit

3 | Ethische Herausforderungen von KI in der Medizin

4 | Lösungsansätze: Klug mit KI umgehen

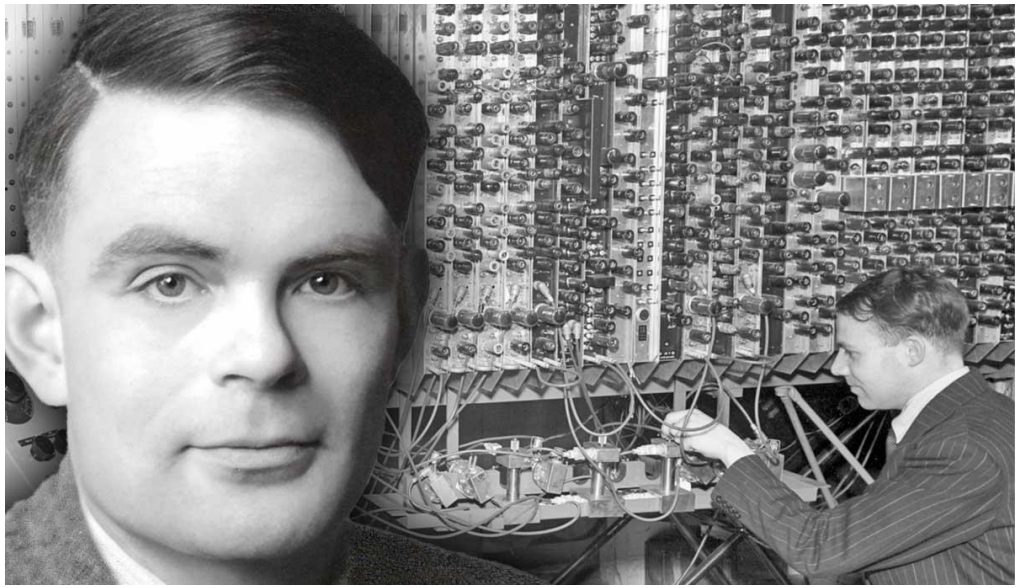
01

Kurze Einführung zur Künstlichen Intelligenz (KI)

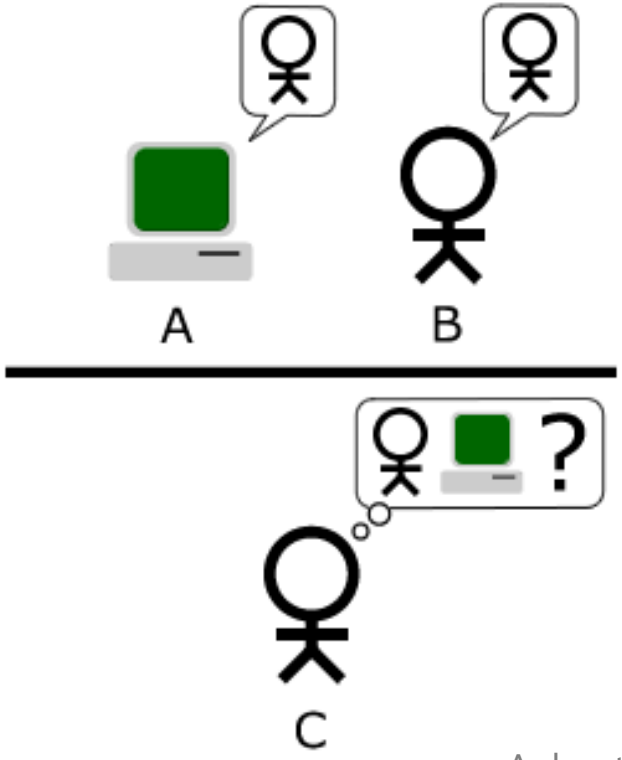


Definitionen

Künstliche Intelligenz (KI) ist die Fähigkeit eines Computers, kognitive Leistungen zu erbringen, die denen des **Menschen ähnlich** sind, insbesondere die Umwelt **wahrzunehmen**, Ereignisse **vorherzusagen** und aus Erfahrungen **zu lernen**.



Alan Turing (1912-1954)



Adaptiert nach Rubeis 2024

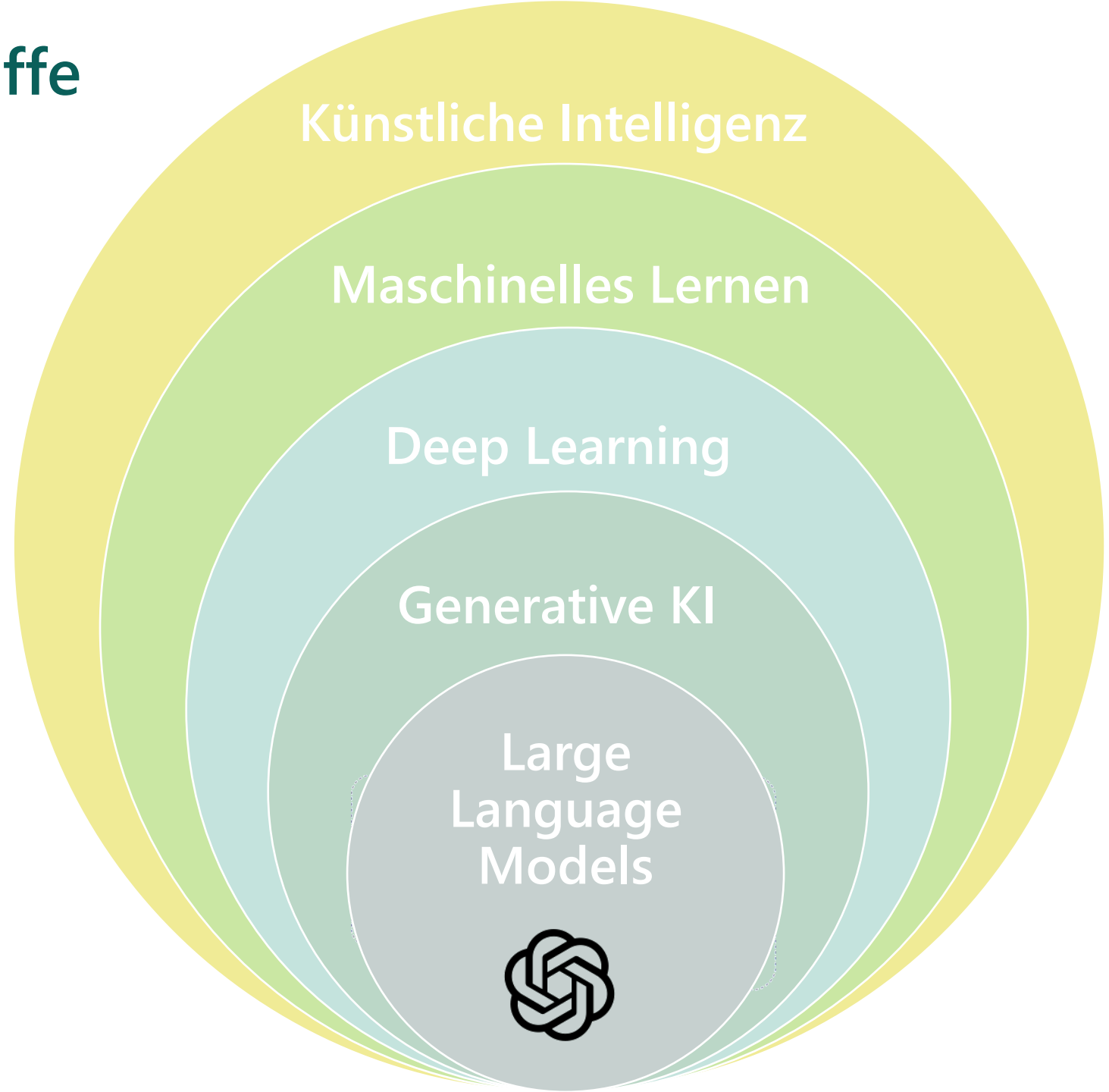
Der erste Psychotherapie-Chatbot

```
 Welcome to
                EEEEEE  LL      IIII  ZZZZZZZ   AAAAA
                EE      LL       II       ZZ   AA   AA
                EEEEE   LL       II     ZZZ    AAAAAAA
                EE      LL       II    ZZ      AA   AA
                EEEEEE  LLLLLL  IIII  ZZZZZZZ  AA   AA

 Eliza is a mock Rogerian psychotherapist.
 The original program was described by Joseph Weizenbaum in 1966.
 This implementation by Norbert Landsteiner 2005.


ELIZA: Is something troubling you ?
YOU:   Men are all alike.
ELIZA: What is the connection, do you suppose ?
YOU:   They're always bugging us about something or other.
ELIZA: Can you think of a specific example ?
YOU:   Well, my boyfriend made me come here.
ELIZA: Is it important to you that your boyfriend made you come here ?
YOU:   He says I'm depressed much of the time.
ELIZA: I am sorry to hear that you are depressed.
YOU:   It's true. I am unhappy.
ELIZA: Can you explain what made you unhappy ?
YOU:
```

Joseph Weizenbaum (1923-2008)

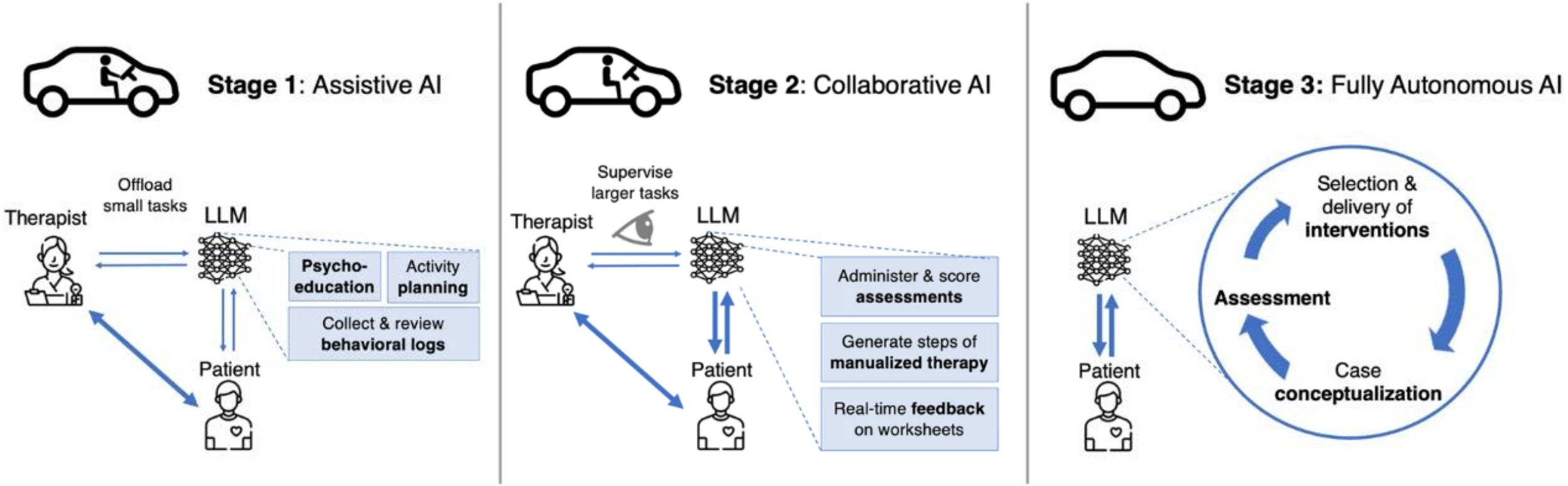
https://web.njit.edu/~ronkowit/eliza.html

Grundbegriffe



Schwache, Starke und Super KI

Grade der Autonomie



«Machine in the Loop»

«Human in the Loop»

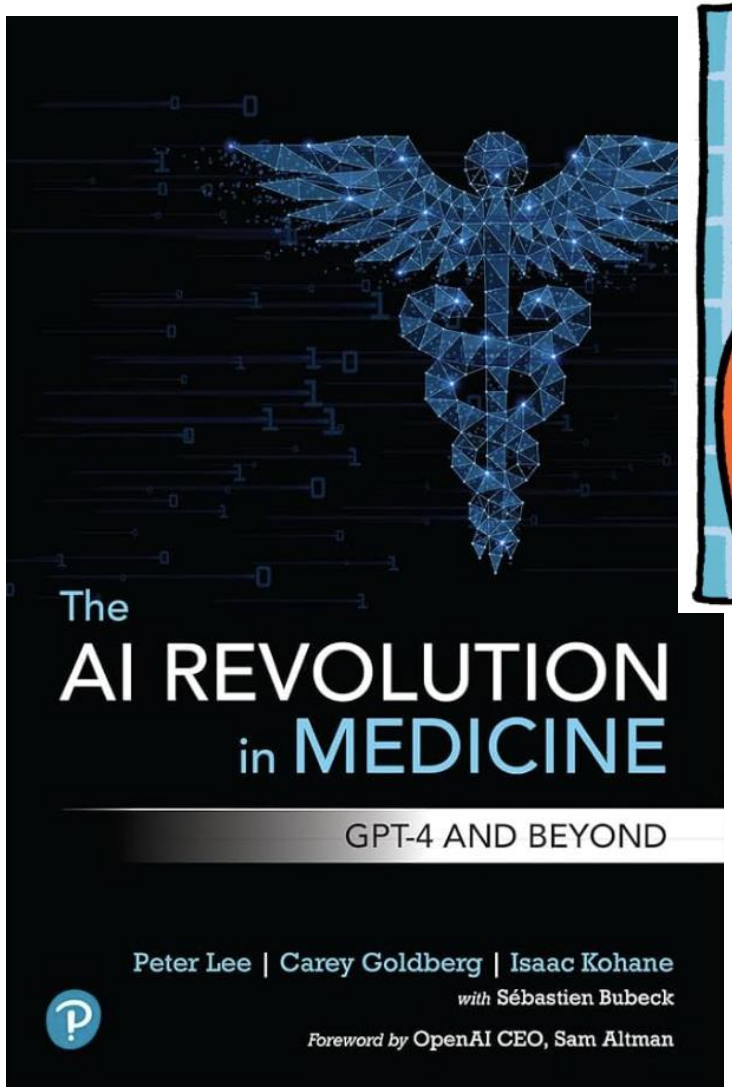
«Human out of the Loop»

Stade et al. 2025

Hoffnungen…

Five ways AI could improve the world: 'We can cure all diseases, stabilise our climate, halt poverty'

The Guardian



NZZ FORMAT

Sind Algorithmen bald die besseren Ärzte? Eine Dokumentation, die zeigt, wie KI die Medizin revolutioniert

Algorithmen erkennen Krebsmuster, erstellen bei Schlaganfällen Prognosen und unterstützen bei Operationen. Der Alltag von Ärzten wird sich durch KI verändern. Wie gut sind die KI-Systeme wirklich? Oft sind sie wie eine Blackbox, und die Mediziner verstehen nicht, wie es zu den Diagnosen kommt.

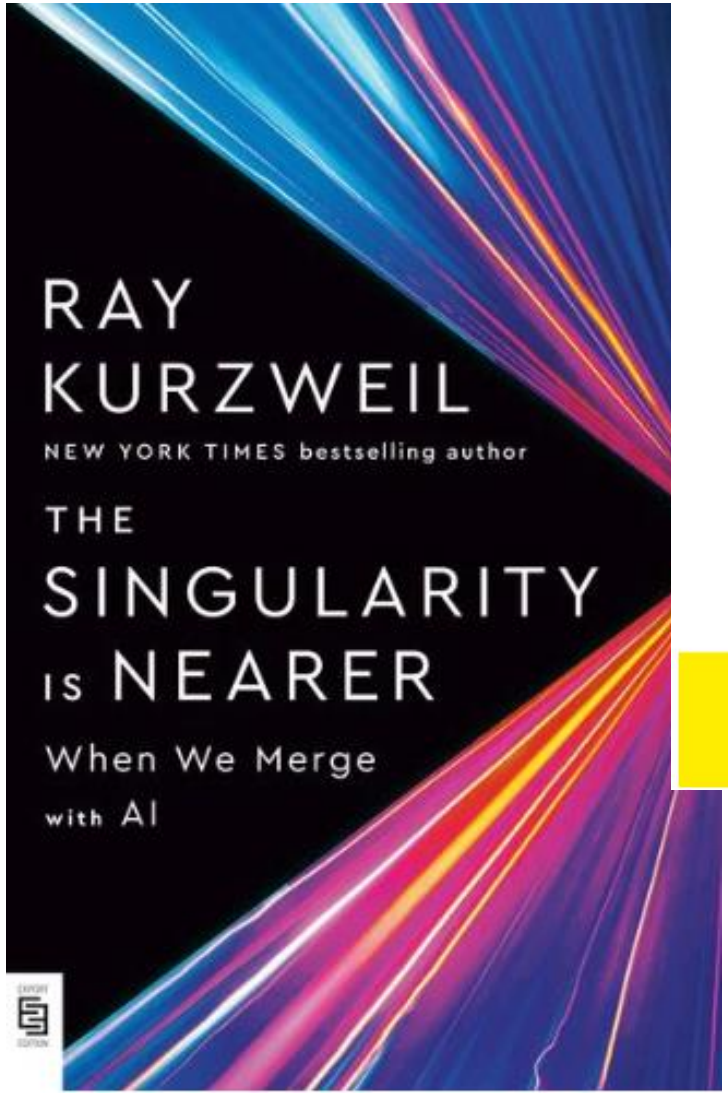
Andrea Hauner
17.11.2023, 06.30 Uhr

Merken Drucken Teilen

...und Ängste

Five ways AI might destroy the world: 'Everyone on Earth could fall over dead in the same second'

The Guardian



B DAS MAGAZIN

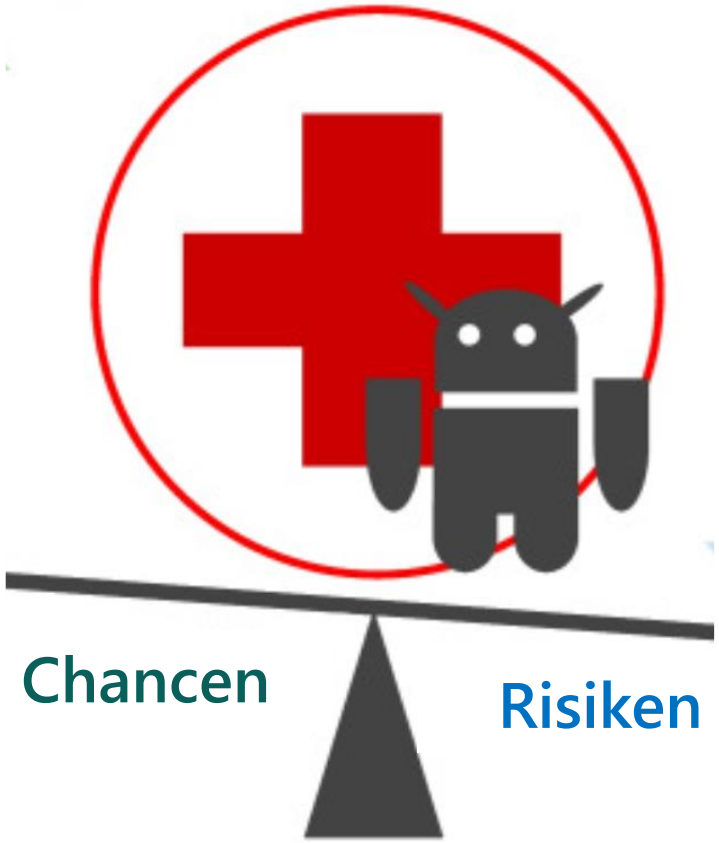
Abo Präsident der ETH Lausanne über KI

«Haben Sie Chat-GPT ausprobiert?» – «Natürlich. Ich bin erschrocken, weil das Programm so schlecht ist»

Was ist künstliche Intelligenz überhaupt? Und warum macht sie so viele Fehler? Martin Vetterli, Präsident der ETH Lausanne, erklärt, warum das Wettrüsten um KI so gefährlich ist.

Hannes Grassegger (Das Magazin)
Publiziert: 12.05.2023, 14:07
Aktualisiert: 12.05.2023, 16:16

Risiko-Nutzen-Abwägung

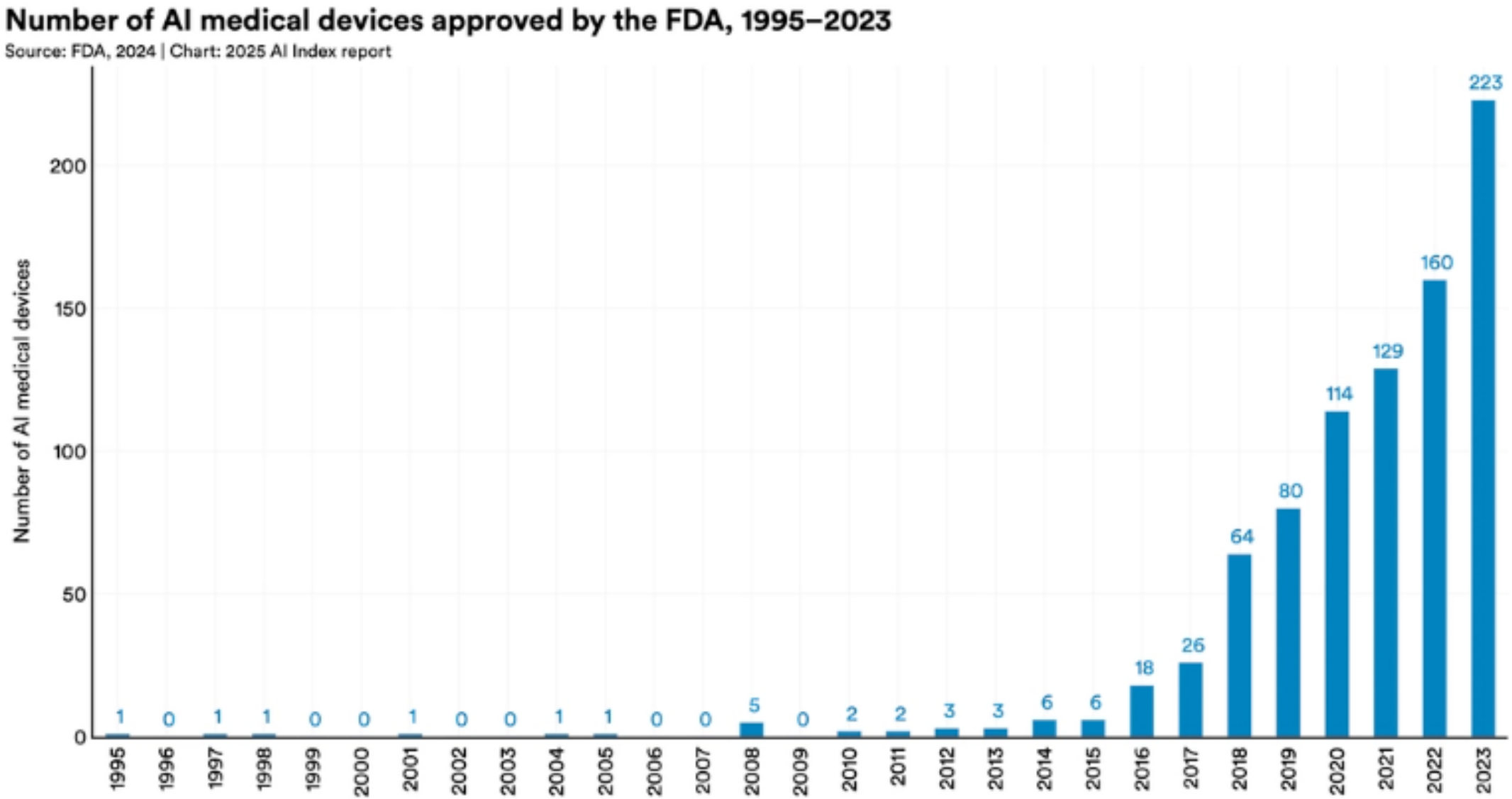


Pagallo et al. 2024

KI im Gesundheitswesen

WHO 2023

KI im Gesundheitswesen



HAI 2025

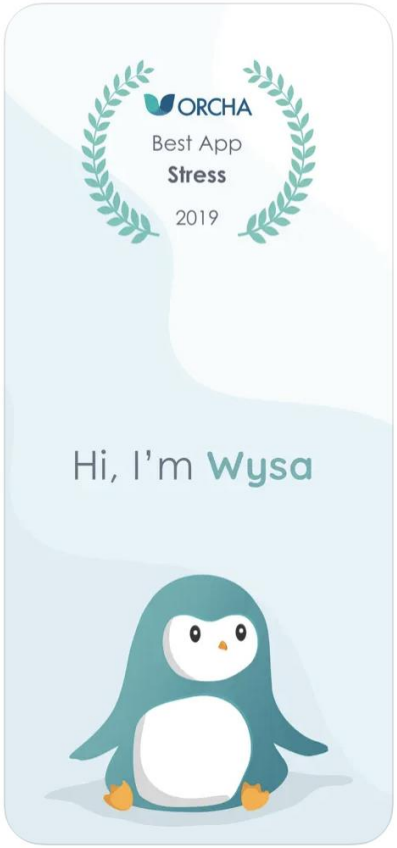
02

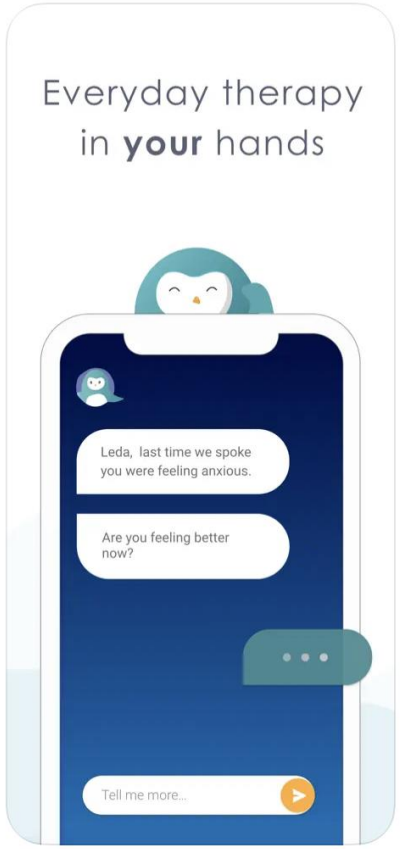
Anwendungsbereiche von KI für die psychosoziale Gesundheit

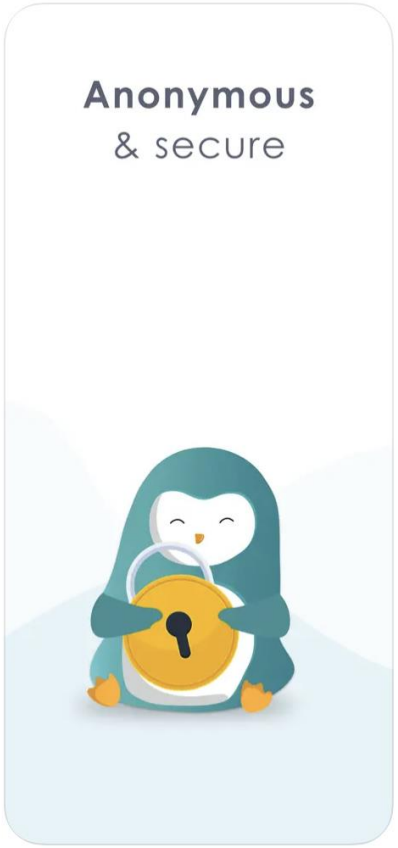
Anwendungsbereiche für die mentale Gesundheit

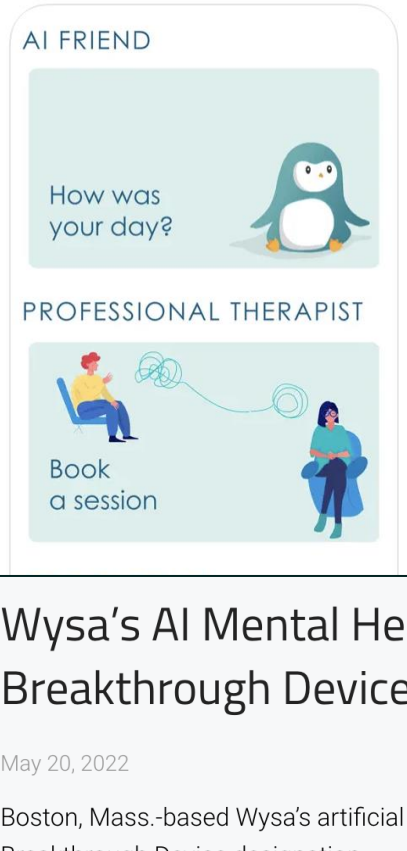
- Digitale Therapien (z.B. Chatbots, VR-Therapien)
- Professionelle Entscheidungshilfen (z.B. zur personalisierten Behandlungsplanung)
- Patientenüberwachung (z.B. Überwachung von Risiko-verhalten)
- Support bei administrativen Tätigkeiten
- Kommunikationstools (z.B. Übersetzungen oder einfache Sprache)
- Förderung der Gesundheitskompetenz

Psychotherapie-Chatbots









© https://www.wysa.io/for-individuals

Wysa's AI Mental Health Software Designated an FDA Breakthrough Device

May 20, 2022

Boston, Mass.-based Wysa's artificial intelligence (AI) mental health software has been granted an FDA Breakthrough Device designation.

The Wysa software-as-a-medical-device delivers cognitive behavioral therapy (CBT) via a smartphone-based "conversational agent" to reduce the symptoms of depression and anxiety, reduce pain interference and improve physical function.

Therapeutische Allianz

Working Alliance Inventory-Short Revised (WAI-SR, Skala 1-5)

	Number of participants	Bond subscore, *M* (*SD*)	Goal subscore, *M* (*SD*)	Task subscore, *M* (*SD*)	Total score, *M* (*SD*)
Assessment 1	1,205	3.9 (0.94)	3.62 (0.93)	3.33 (0.84)	3.64 (0.81)
Assessment 2 (3 days later)	226	4.05 (0.91)	3.70 (0.86)	3.50 (0.77)	3.75 (0.80)

Study	Modality	*M* (*SD*)
Wysa (this study)	Digital (*N* = 1,205)	3.98 (0.94)
Munder et al. (47)	In-person outpatient (*N* = 88)	4.00 (0.78)
Jasper et al. (48)	Group CBT (*N* = 26)	3.80 (0.80)

Beatty et al. 2022

THE MOST TRUSTED AI IN MENTAL HEALTHCARE

Clinical AI that makes mental healthcare more human

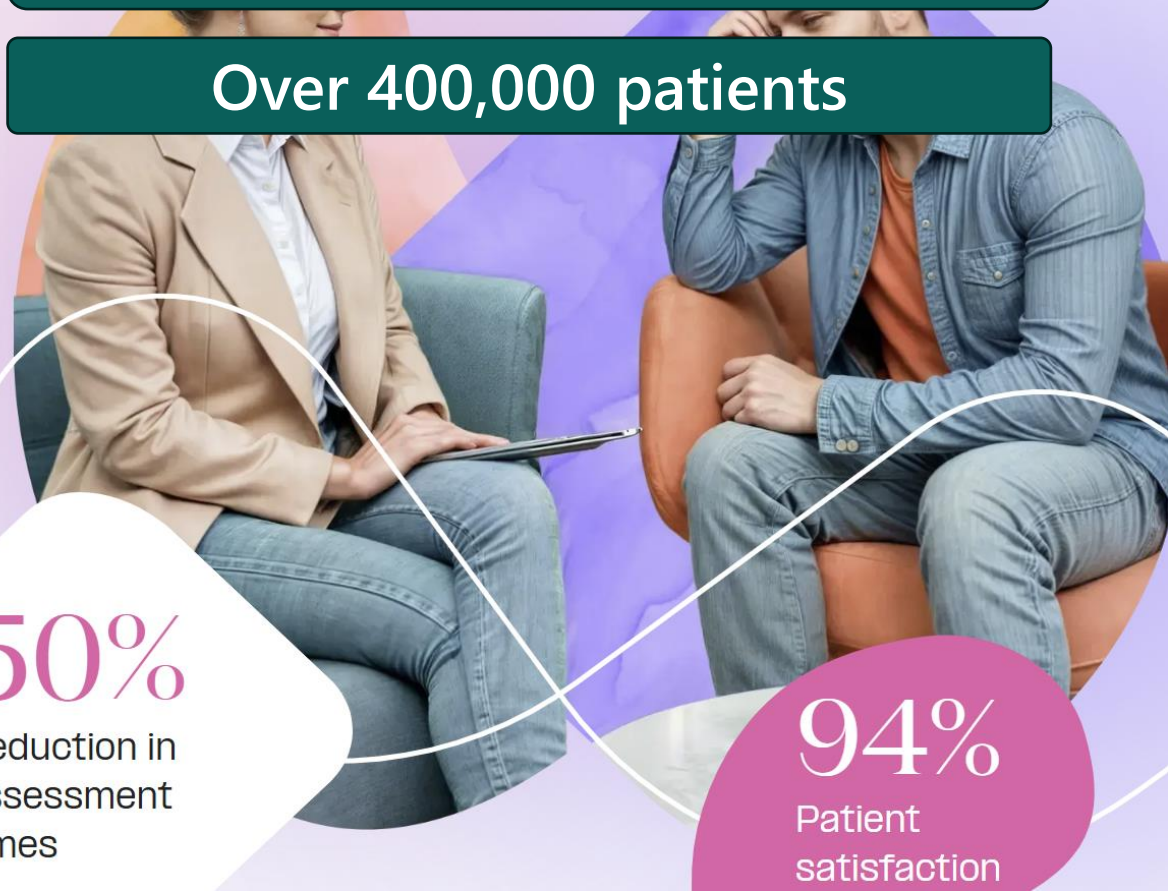
Reduce staff burnout, improve community engagement, and drive better patient outcomes. That's the power of clinical AI.

Learn more

Class IIa UK medical device

40% of NHS Talking Therapies use Limbic as routine care

Over 400,000 patients

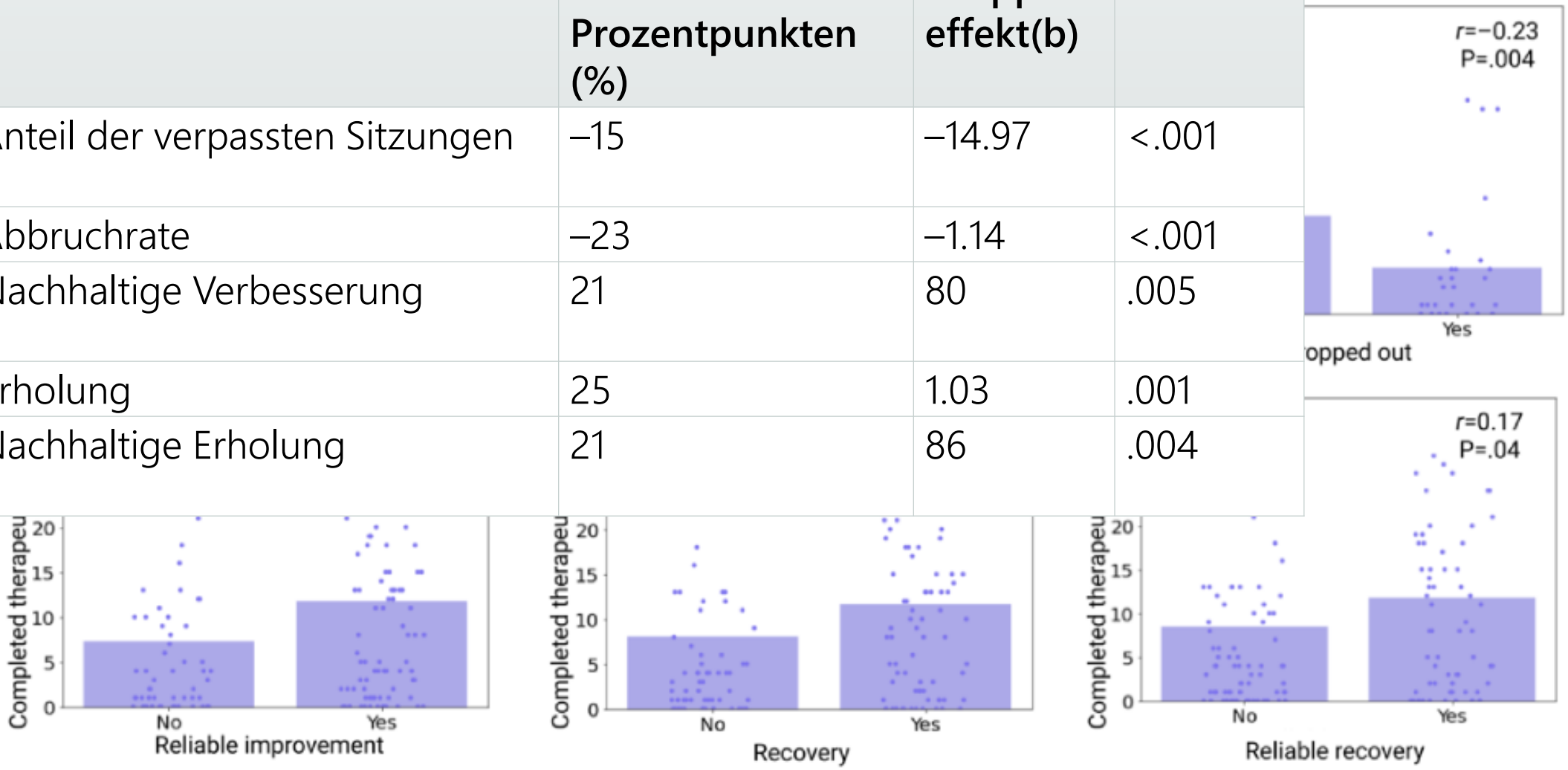


50%
Reduction in assessment times

94%
Patient satisfaction

Wirksamkeit

Outcome	Differenz in Prozentpunkten (%)	Gruppen-effekt(b)	P-Wert
Anteil der verpassten Sitzungen	–15	–14.97	<.001
Abbruchrate	–23	–1.14	<.001
Nachhaltige Verbesserung	21	80	.005
Erholung	25	1.03	.001
Nachhaltige Erholung	21	86	.004



Habicht et al. 2025

Mental health needs have multiplied. Support hasn't.

Our mission is to make mental health support radically accessible by building the future of chat-based AI wellness tools.

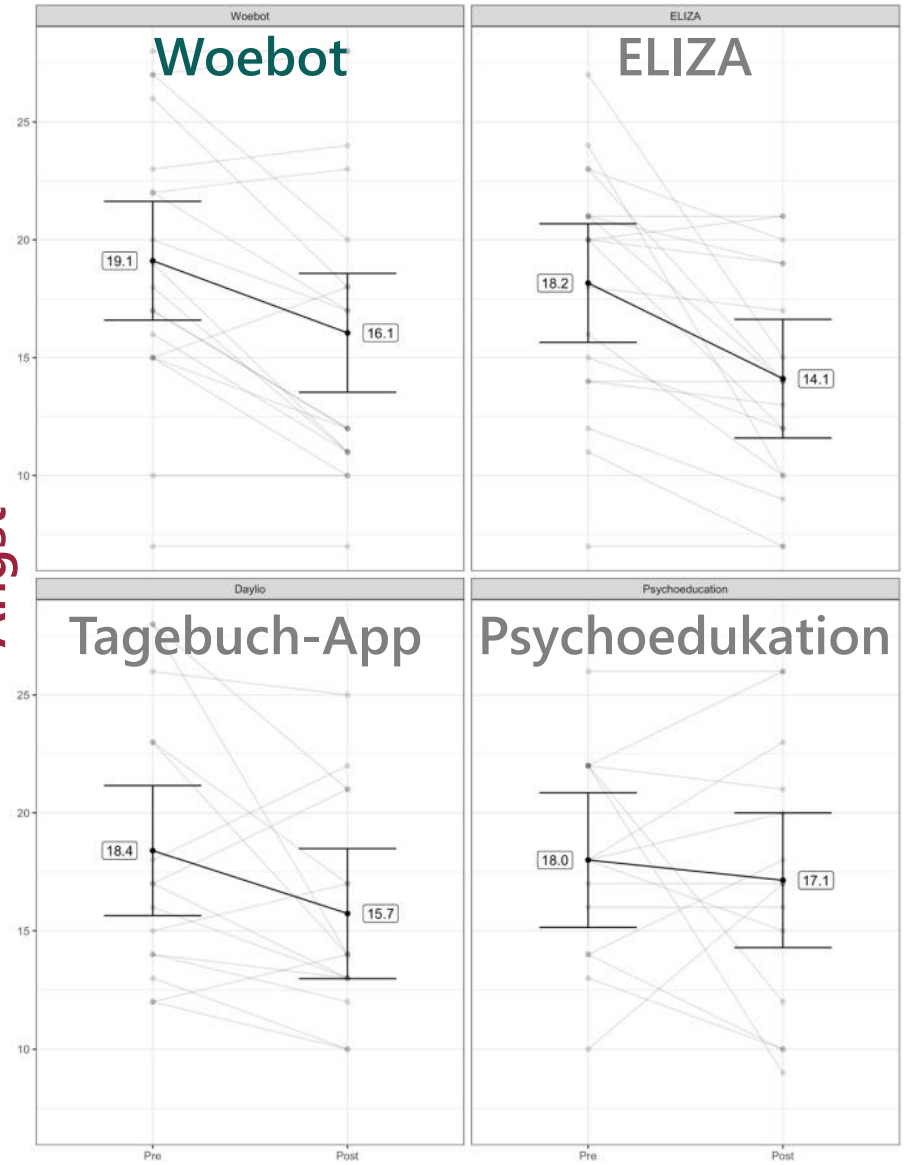
We were honored to be included in Newsweek's inaugural list of the World's Best Digital Health Companies, which recognizes stand-out companies with truly beneficial products

Too often, people don't feel supported—

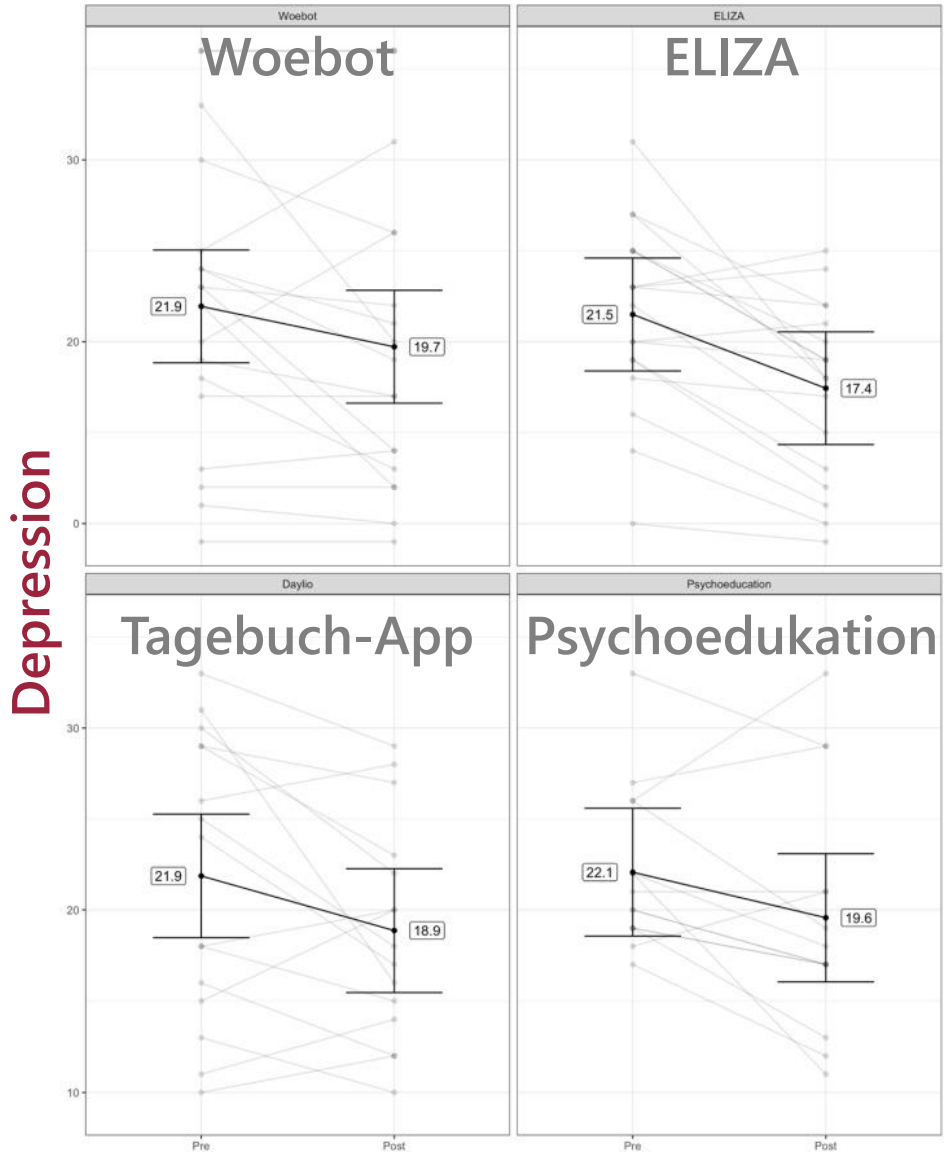
2–3 Months

Wirksamkeit

Angst



Depression



Eltahawy et al. 2024

TalktoThera - Conversational Therapy.

Therabot

Blog About Services Contact Support Us CHAT NOW

Good news! Our WhatsApp chatbot is back online! Access it now at this link.

New Find a community

Your Compassionate Digital Partner

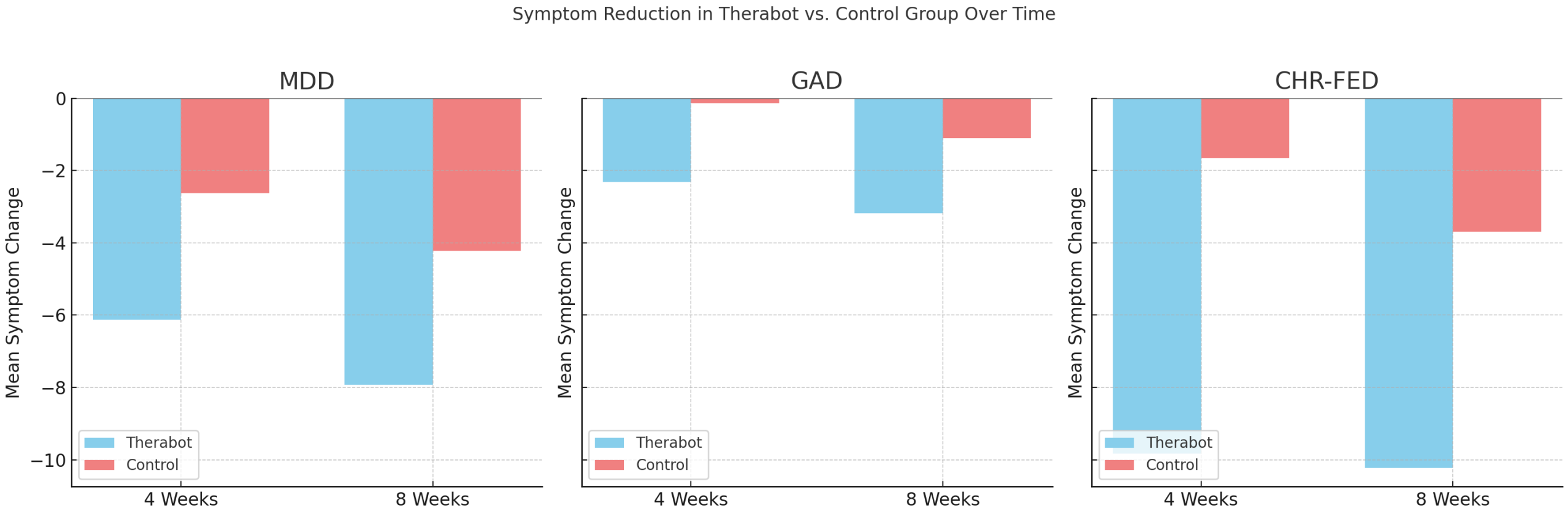
Life's journey can be daunting but remember that you aren't alone. Therabot is here to support you, offering encouragement, resilience, and hope. Together, we will navigate the path ahead.

Chat NOW | Group Therapy | Join a Community

https://www.trytherabot.com/

Wirksamkeit

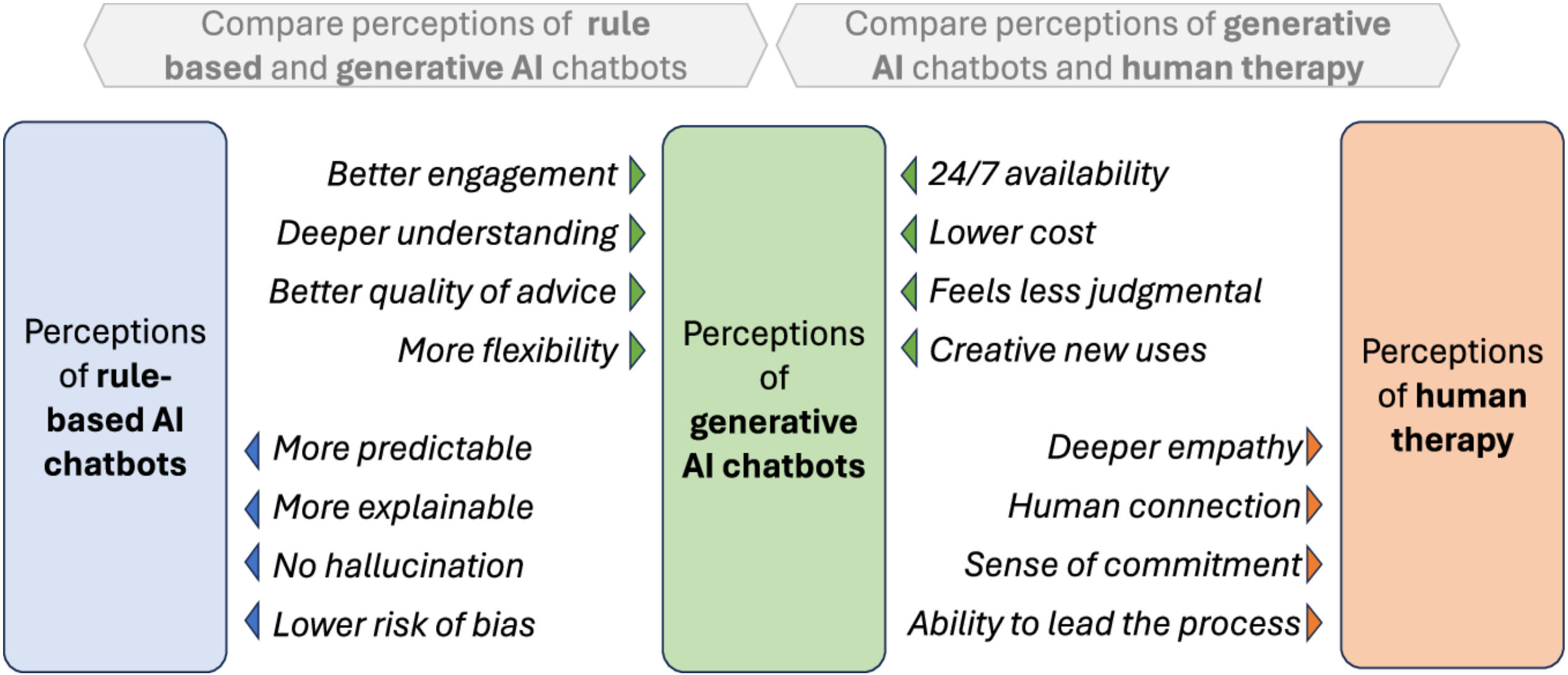
Therabot



Grafik erstellt mit Hilfe von ChatGPT

Heinz et al. 2025

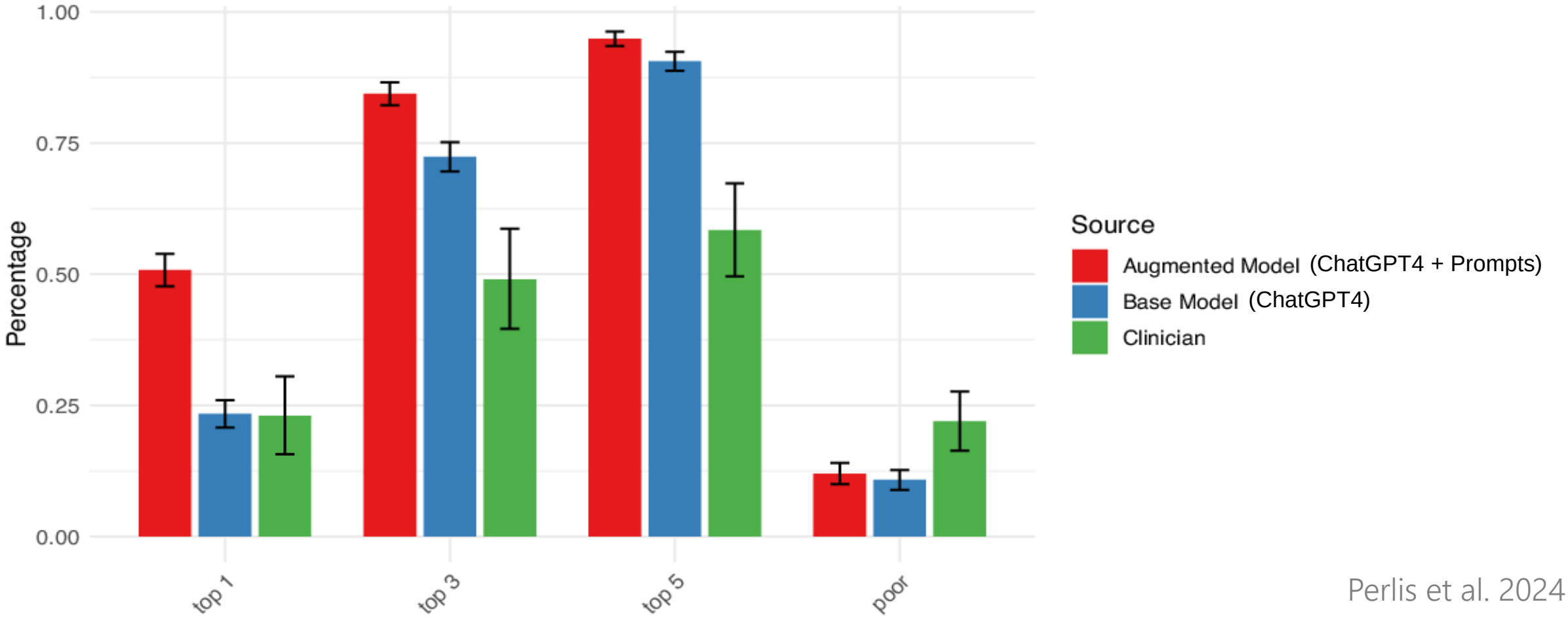
Einstellungen der Patient*innen



Siddals et al. 2024

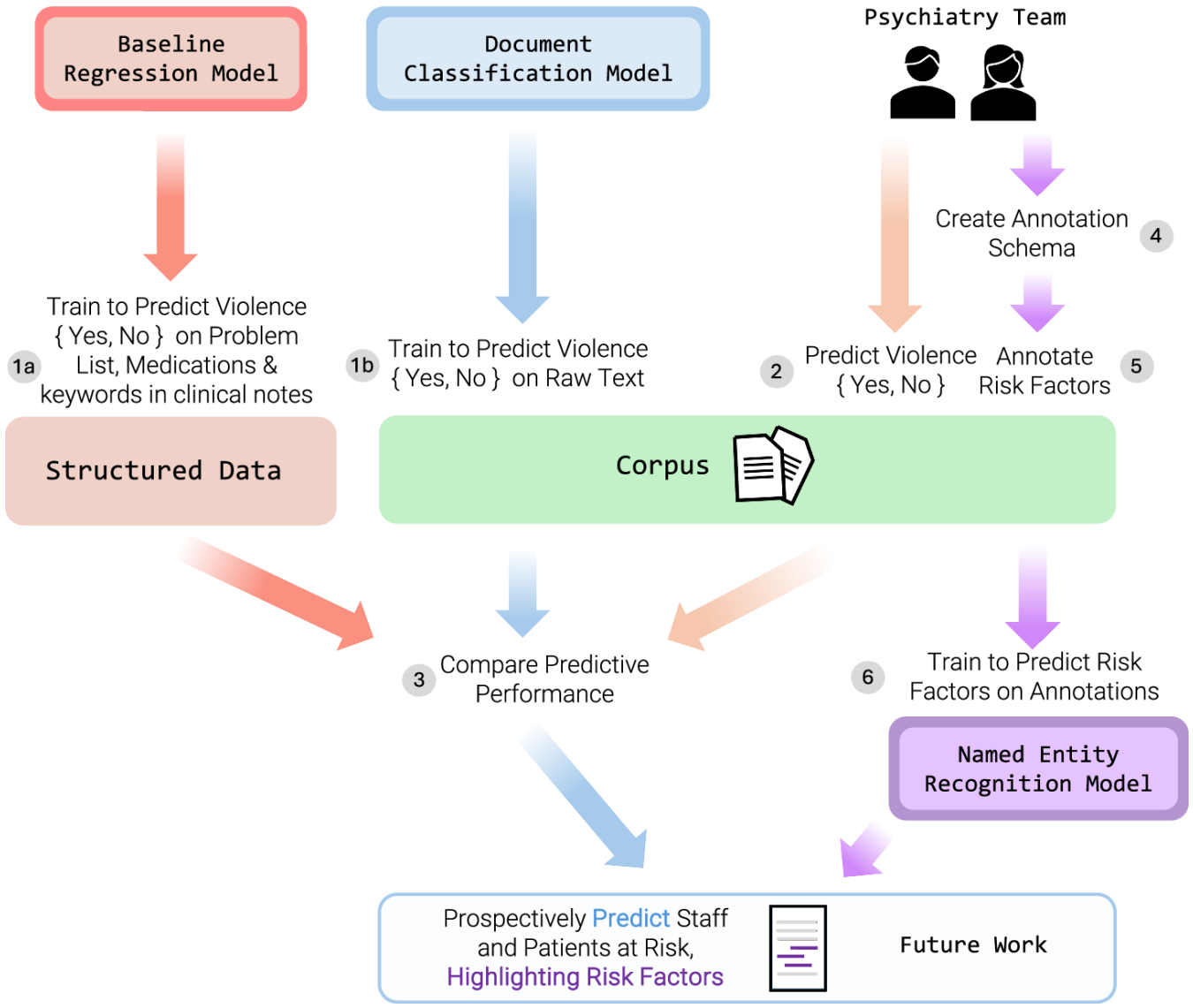
KI für die Behandlungsplanung

«ChatGPT, gib mir für folgende Fallbeschreibungen zu Personen mit bipolarer Depression die besten Medikamente (Top 1, Top 3, Top 5) für die Behandlung an.»



Perlis et al. 2024

Patientenüberwachung: Vorhersage von Gewalt



Modell	Prädiktive Leistung (F1, 0-1)
Psychiatrie-Team	0.5
Baseline Regression Model	0.72
Document Classification Model	**0.75**

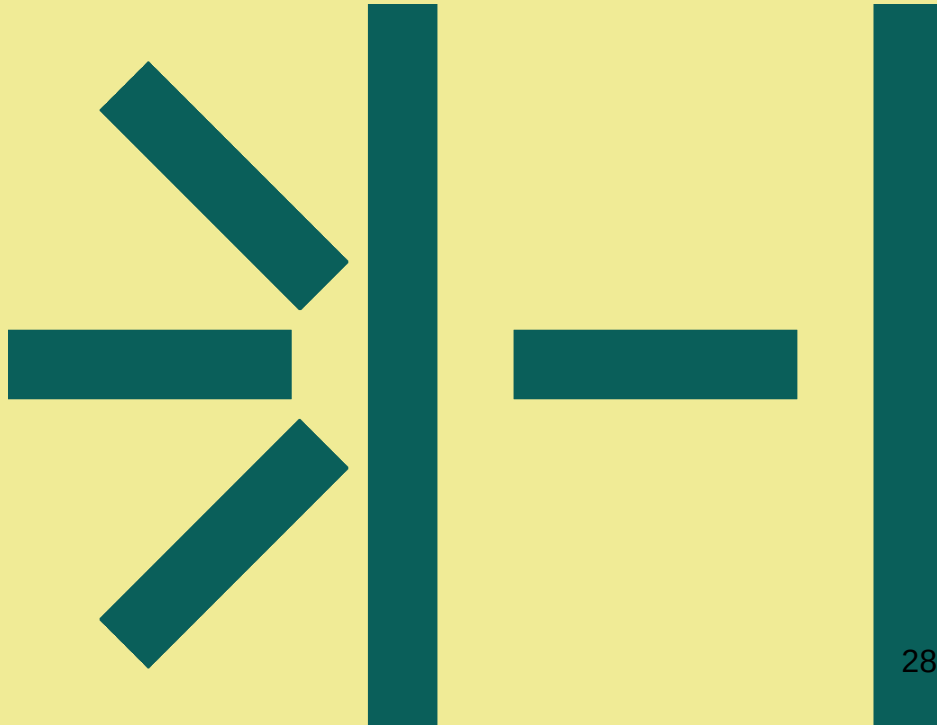
Dobbins et al. 2024

Potenzieller Nutzen von KI im Bereich mentale Gesundheit

- Verbesserter Zugang zu psychologischer Unterstützung (Online-Plattformen, Mehrsprachigkeit)
- Potenzial zur Senkung der Kosten und von Personalengpässen
- Personalisierte Behandlung (durch Integration verschiedener Daten)
- Frühzeitige Unterstützung (z.B. durch Vorhersagen von Krisen)
- Verbesserte diagnostische Genauigkeit
- Entwicklung neuer therapeutischer Ansätze (z.B. Therapien in der virtuellen Realität)
- Unterstützung der Forschung und Wissensgenerierung

03

Ethische Herausforderungen von KI in der Medizin



Ethische Herausforderungen im Überblick

- Autonomie
- Schaden und Haftung
- Transparenz und Erklärbarkeit
- Voreingenommenheit und Fairness
- Datenschutz und Datensicherheit
- Nachhaltigkeit

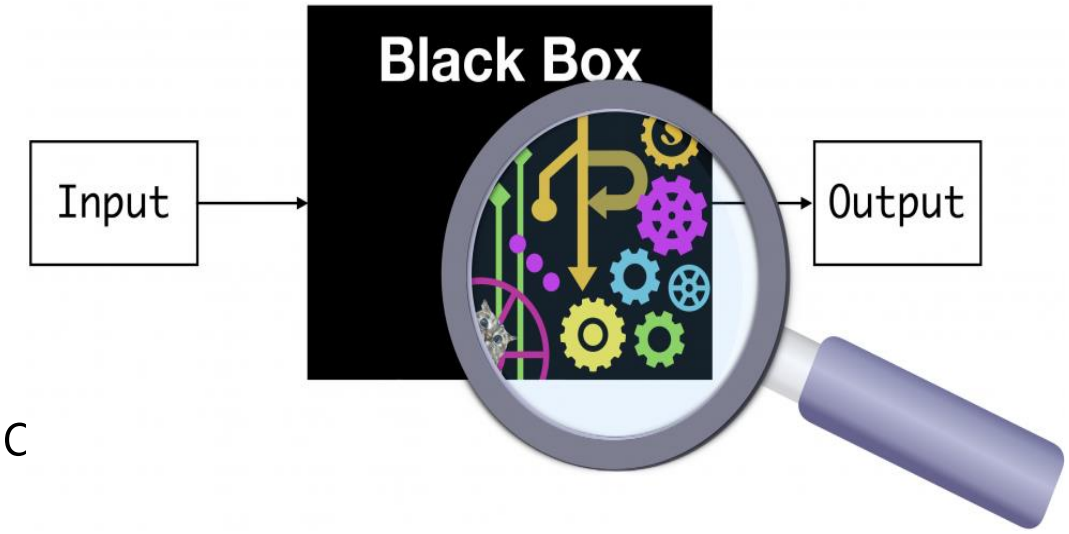
1 | Autonomie

- KI-Systeme können die Autonomie von Patient*innen und Fachpersonen einschränken
  - Kein standardisiertes **Verfahren zur Einwilligung** in die Nutzung von KI-Tools
  - Gefahr einer **paternalistischen Arzt-KI-Patienten-Beziehung**
  - **Wertfrage**: Welche Werte werden von KI-Systemen wie gewichtet?
  - **Darstellung** von Daten beeinflusst Entscheidungs-träger*innen (Ordnungseffekte)
  - KI-Systeme sind keine **moralischen Akteure**
  - **Depersonalisierung** der Gesundheitsversorgung

McDougall 2019, Debrabander & Mertes 2022; Lorenzini et al. 2022, 2023; Sauerbrei et al. 2023

2 | Transparenz und Erklärbarkeit

- Transparenz und Erklärbarkeit von KI-Systemen können eingeschränkt sein
  - KI-Systeme sind eine **Black-Box**
  - Angemessene Information als Voraussetzung einer **autonomen Entscheidung** und von **Vertrauen** in der therapeutischen Beziehung
  - Sind Ärzt*innen und Patient*innen hinreichend **kompetent** im Umgang mit KI-Systemen?



  - Erfüllen die KI-Systeme die **Transparenz-Kriterien** für eine vertrauenswürdige KI?

Alami et al. 2020; Fehr et al. 2024; Holm 2023

3 | Schaden und Haftung

- KI-Systeme können Patient*innen aufgrund von Fehlern schaden
  - Wie hoch ist der Grad der **Genauigkeit** von KI-Systemen?
  - Fehlerhafte Systeme können zu **Fehlentscheidungen** und **schwerwiegenden gesundheitlichen Schaden** führen
  - Wer ist für **Fehler** von KI-Systemen haftbar? Wer ist verantwortlich für ihre Überwachung,
    Kontrolle und Wartung?
  - **Overreliance** und **De-Skilling** bei Gesundheitsfachpersonen
  - Patient*innen können im Umgang mit KI-Systemen **überfordert** sein

Cheetham et al. 2022; Mirbabaie et al. 2022; Stroud et al. 2024

4 | Voreingenommenheit und Fairness

- KI-Systeme können Ergebnisse verzerren und gesundheitliche Ungleichheiten verschärfen
  - Repräsentiert der Trainingsdatensatz die **Ausgangspopulation** und die **Zielpopulation**?
  - KI-Systeme könnten **menschliche Vorurteile** übernehmen
  - KI-Systeme können bestehende soziale **Ungleichheiten** aufrechterhalten oder verstärken
  - **Ethik-Dumping** bei Datengenerierung
  - Gefahr einer **digitalen Kluft** («digital divide»)

nature

NEWS | 24 October 2019 | Update 26 October 2019

Millions of black people affected by racial bias in health-care algorithms

Study reveals rampant racism in decision-making software used by US hospitals – and highlights ways to correct it.

Heidi Ledford

Vayena et al. 2018; Obermeyer et al. 2019; Kuhn et al. 2021; Scott et al. 2021

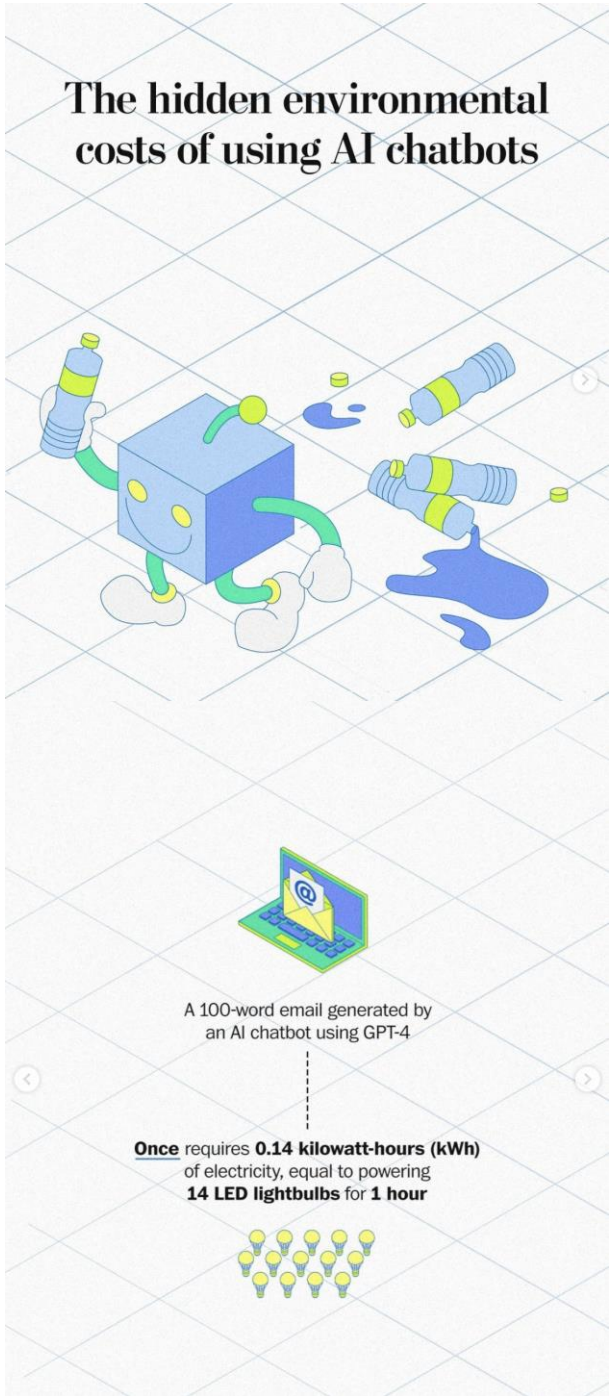
5 | Datenschutz und Datensicherheit

- Datenschutz und Datensicherheit können durch KI-Systeme gefährdet sein
  - Wie werden die Daten **erhoben und gespeichert**, und wer kann sie unter welchen Bedingungen nutzen oder wiederverwenden?
  - Wem **gehören** die Daten, wer ist für sie **verantwortlich**?
  - **Interessenkonflikte** beim Datenschutz
  - **Cybersicherheit** von KI-Systemen (KI-spezifische Angriffe)
  - Hohes Potential für **Missbrauch** von Dat oder KI-Modellen

Alami et al. 2020, Murphy et al. 2021

6 | Nachhaltigkeit

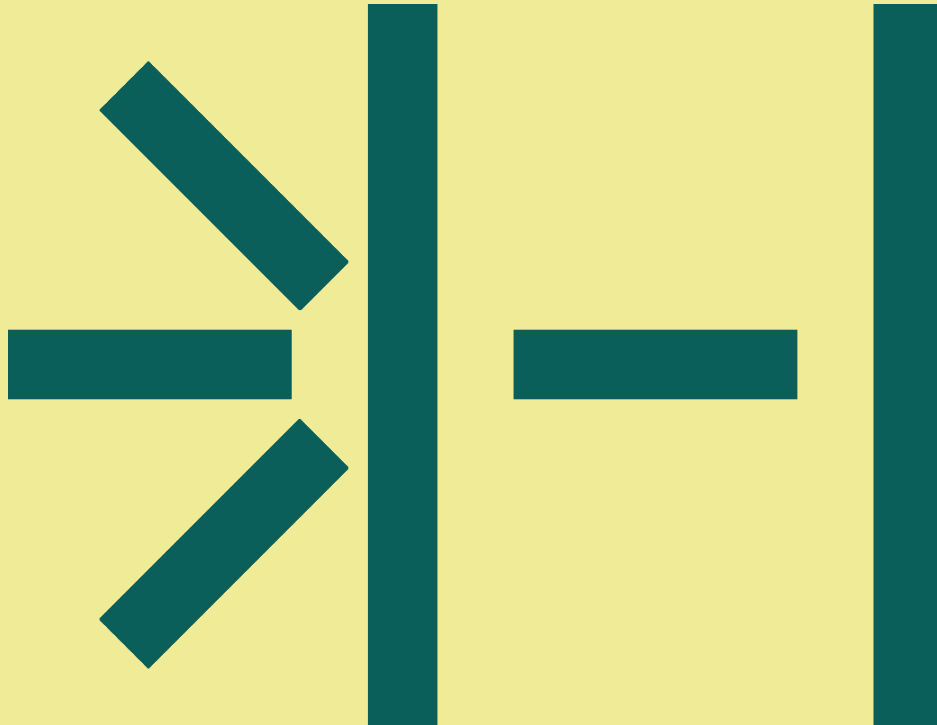
- KI-Systeme verbrauchen viele Ressourcen
  - Grosser **ökologischer Fussabdruck** von KI-Systemen
  - In die KI-Forschung, -Entwicklung und -Implementierung werden riesige **finanzielle Mittel** investiert, die an anderer Stelle fehlen (2024: 109 Mrd. $ in den USA)
  - **Finanzielle Interessen** von KI-Firmen in Spannung von Nachhaltigkeit & Sicherheit
  - **Nutzen-Lücke** zwischen theoretischen Versprechungen und Alltagsanwendungen
  - Verschärfen **globaler Ungleichheiten** im Gesundheitswesen



Verma & Tan 2024; Morely et al. 2025

03

Lösungsansätze:
Klug mit KI umgehen



Lösungsansätze: Klug mit KI umgehen

Klugheit (phrónesis) ist die Fähigkeit zu angemessenem Handeln im konkreten Einzelfall unter Berücksichtigung aller für die Situation relevanten Faktoren, Handlungsziele und Einsichten.

- Ethische Prinzipien
- Medizinisch-ethische Richtlinien
- Minimalstandards der Aufklärung – KI-Patientenrechte
- Gesellschaftliche Auseinandersetzung mit KI
- Rechtliche Regulierung von KI

Ethische Prinzipien

- **WHO-Prinzipien für den Einsatz von KI im Gesundheitswesen**

Protect autonomy

(incl. data protection)

Promote human well-being, human safety and the public interest

Ensure transparency, explainability and intelligibility

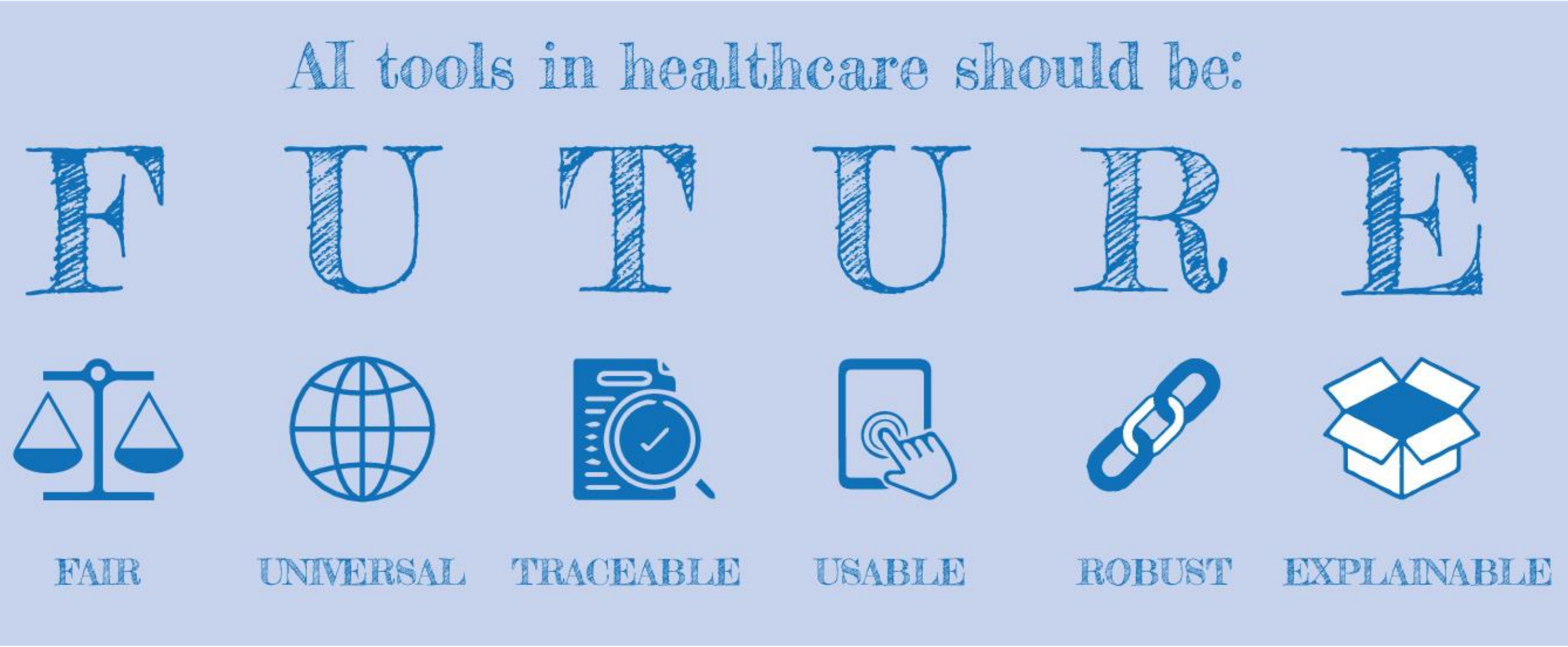
Foster responsibility and accountability

Ensure inclusiveness and equity

Promote AI that is responsive and sustainable

WHO 2021, 2024

FUTURE Richtlinie



Lekadir et al. 2025

Kriterien für die Psychotherapie

- **KI in der Psychotherapie muss…**
  - Schadensrisiken erkennen (insb. Suizidalität / Mordgedanken, Missbrauch, Gewalt)
  - unerwünschte Verhaltensweisen vermeiden
  - in der psychodiagnostischen Beurteilung helfen
  - in der Beratung reaktionsschnell und flexibel sein
  - aufhören, wenn sie nicht bzw. wahrscheinlich nicht helfen
  - fair und inklusiv sein sowie Vorteilen vermeiden
  - empathisch sein
  - offen als KI erkennbar sein

Stade et al. 2025

FMH-Forderungen zur KI in der Medizin

- **Ärztinnen und Ärzte müssen digitale Technologien verantwortungsvoll und ethisch einsetzen können**

1. Neue digitale Kompetenzen in interdisziplinäre Lehrpläne der Weiter- und Fortbildung aufnehmen
2. Bedeutung der semantischen Interoperabilität medizinischer Informationen vermitteln
3. Rechtliche und ethische Rahmenbedingungen kennen
4. Umgang mit KI fördern und die Sensibilität hinsichtlich deren Risiken erhöhen

FMH 2024

Minimalstandard Informierte Aufklärung

- **Bisher keine explizite Regelung** im CH-Recht
- **Informationspflicht der Fachperson über KI-Systeme, wenn:**
  - damit ein **relevantes Risiko** für Patient*innen verbunden ist oder aus anderen Gründen anzunehmen ist, dass die Information **für die Entscheidung des Patienten relevant** ist
  - die KI-Systeme **unabhängige Entscheidungen** treffen oder **ausserhalb medizinischer Standards** verwendet werden
- **Auskunftsrecht der Patientin:**
  - Information, ob bei der Diagnose oder Behandlung ein **KI-System verwendet** wird
  - Information, wie das KI-System in **Grundzügen funktioniert**

Universität Basel

Touvenin et al. 2024, Shaw et al 2025

KI-Patientenrechte

- **Minimal:**
  - Das Recht zu erfahren, dass KI-Systeme in die klinische Behandlung involviert sind und auf welche Weise
  - Transparenz in Bezug auf die vom KI-System verwendeten Daten

- **Erweitert:**
  - Recht auf eine Zweitmeinung
  - Recht auf Ablehnung von KI-Systemen
  - Recht auf menschlichen Kontakt
  - Recht auf mentale Privatheit
  - Recht auf digitale Gesundheitskompetenz
  - Recht auf Zugang zu KI-Systemen

Ploug & Holm 2020; 2023; Van Kolfschooten 2024

Gesellschaftliche Auseinandersetzung

- Kooperation von Fachgesellschaften
- Datenrelevanz
- Reflexives Design
- Datenschutz und Privatsphäre
- Zugangsgerechtigkeit
- Integration in Aus- und Weiterbildung
- Anwendungskompetenz und Deskilling
- Transparente Aufklärung und Kommunikation
- Erhalt menschlicher Verantwortung

Regulierung: EU Artificial Intelligence Act

Unannehmbares Risiko
- Soziale Scoring-Systeme für Gesundheitsleistungen

Verboten

Hohes Risiko
- KI-basierte Medizinprodukte (Risikoklasse IIa oder höher)
- KI-basierte klinische Entscheidungssysteme
- KI-Systeme zur Triage von Gesundheitsleistungen
- KI-Systeme zum Risiko-Assessment bei Krankenversicherungen

Risikomanagement
Datenmanagement
Qualitätsmanagement
Menschliche Aufsicht
CE-Registrierung

Begrenztes Risiko
- KI-Chatbots mit Gesundheitsempfehlungen
- Medizinische Deepfakes (z.B. Tumor von Bildern hinzufügen / entfernen)
- KI basierte Bewegungs- oder Ernährungssensoren in Altersheimen

AI-Kompetenz
Transparenz

Minimales Risiko
- KI in der pharmazeutischen Forschung und Entwicklung
- KI-Systeme in der Administration im Gesundheitswesen

Keine Anforderungen

KI-Modelle mit allgemeinem Verwendungszweck
- LLM-Systeme zur Zusammenfassung von Patientendaten
- LLM-Systeme zur Erstellung von Berichten

Dokumentation
Transparenz & Information
Modell-Evaluation
Risikomanagement

Van Kolfschooten et al. 2024

Vielen Dank für Ihre Aufmerksamkeit!

jan.schuermann@usb.ch