



Departement Medizin

## KI im Ärztealltag: Facts and Figures statt Hypes und Likes

eHealth Summit Aarau

Topic Session: 2024 PERMIT: GPT dEfinierte Referatstitel aM  
ehealthsummit

Prof. Dr. med. Balthasar L. Hug  
12. September 2024

# Inhalt

- Hintergrund
- Anwendungen im ärztlichen Alltag
- Ausblick

## PERSÖNLICHE VORSTELLUNG

1998	Internist FMH
2003	Ex. MBA, Universität St. Gallen
2011	Master of Public Health, Harvard School of Public Health
2017	Co-Chefarzt Allgemeine Innere Medizin, LUKS
2018 -	Chefarzt AIM, LUKS
2017-2020	Physician Lead Epic Implementation
2021 -	Professur für Community Medicine, Universität Luzern



Mitglied Qualitätskommission der SGAIM

Keine persönlichen Interessenskonflikte mit dem besprochenen Thema

# Hintergrund

- **Thema AI:** sehr wichtig für die medizinischen Wissenschaften. Grosse Themenkreise, welche uns beschäftigen: gen. KI, Bedeutung für Diagnostik (Bildanalyse, Pathologie, Algorithmen für Temperatur, EKG-Analyse etc.)
- **«Facts and Figures»:**
  - Facts: objektiv, wie die Anwendungen heute verwendet werden (auszugsweise)
  - Figures: heute keine Zahlen. Noch zu früh, zu neues Feld
- **«Hypes and Likes»:**
  - Hypes: subjektiv, kritiklose Annahme neuer Technologien, «alle unsere Probleme lösen», «Herdentrieb», z.B. Kurs Nvidia
  - Likes: persönliche Vorlieben, von sich auf alle schliessen
- **Challenges, Herausforderungen:** Veränderung des zwischenmenschlichen Kontaktes, s. LRB-Artikel über SF, junge Frauen, Depressionen, ChatBots, Massenbeeinflussungen durch Algorithmen, s. Politik, falsches Training der Algorithmen, Einsatz für Krieg (Drohnenschwarm im Ukraine-Krieg, Roboterhunde etc.) -> **Im Vortrag nicht oder nur ansatzweise abgedeckt**

# Hintergrund

Marktbericht > Nvidia

104.75 USD

+104.71 (261'775.00%) ↑ immer

Geschlossen: 12. Aug., 08:20 GMT-4 • Haftungsausschluss  
Vorbörse 105.80 +1.05 (1.00%)

1 T. | 5 T. | 1 M. | 6 M. | YTD | 1 J. | 5 J. | Max.



Eröffnung	105.64	Marktkap.	2.58 Bio.	CDP-Rating	B
Hoch	106.60	KGV	61.30	52-Wo-Hoch	140.76
Tief	103.43	Rendite	0.038%	52-Wo-Tief	39.23

## London Review OF BOOKS

Vol. 46 No. 3 · 8 February 2024

In the Shadow of Silicon Valley  
Rebecca Solnit

# Hintergrund

- Erster Boom im Computing: mathematische Intelligenz\* -> Rechenleistung
- Zweiter Boom im Computing: sprachliche Intelligenz\* -> sprachliche Leistung
- Dritter Boom im Computing: Quantum Computing -> Computer-Leistung per se (Geschwindigkeit, Speicher)  
-> neue Verschlüsselungstechniken notwendig
- Weitere Booms? Z.B. Emotionen, s. Studien zu «Freundlichkeit» der LLM-Modelle etc.

\*Intelligenz: lat. Verb «intellegere» = verstehen, auseinanderhalten («auseinander pflücken»)

# Hintergrund

- Definitionen (Auszug)
- Alignment: Ausrichtung, Abgleich

Glossary
<b>Alignment:</b> The degree to which the behaviors and actions of an artificial intelligence (AI) system are congruent with human values.
<b>Generative AI:</b> A form of AI designed to produce new and original data outputs, including those that resemble human-made content, with a range of output types that span text, code, images, audio, and video.
<b>Human Values:</b> A broad term for the principles, standards, and preferences that reflect human goals and guide human behaviors.
<b>Large Language Model (LLM):</b> A type of AI model that interprets and generates text. LLMs are often pre-trained with large text corpora and then fine-tuned through supervised fine-tuning and reinforcement learning from human feedback.
<b>Model Card:</b> A document that includes a comprehensive overview and performance characteristics of a machine-learning model, for example, training and evaluation data and training procedure; existing evaluations, for example, observed safety or bias challenges and existing remediation strategies; intended use cases; and model performance across populations, for example, key demographic or clinical groups). A model card is similar to the "System Card" for GPT-4. <sup>37</sup>
<b>Reinforcement Learning from Human Feedback:</b> A method of fine-tuning LLMs where humans rank responses to prompts; reinforcement learning is then used to adjust the output to align with human preferences.
<b>Supervised Fine-Tuning:</b> A method of fine-tuning LLMs that uses human-written responses to example prompts.
<b>Utility:</b> The quantitative measure used in decision analysis to assess the value of a health state or outcome. Utilities may be elicited directly from individual patients or groups, or they can be learned from data. Utilities may be applied to persons, groups, or populations.

## Medical Artificial Intelligence and Human Values

Kun-Hsing Yu, M.D., Ph.D., Elizabeth Healey, S.B., Tze-Yun Leong, Ph.D.,  
Isaac S. Kohane, M.D., Ph.D., and Arjun K. Manrai, Ph.D.

# Hintergrund

The NEW ENGLAND JOURNAL of MEDICINE

REVIEW ARTICLE

AI IN MEDICINE  
Jeffrey M. Drazen, M.D., Editor

## Medical Artificial Intelligence and Human Values

Kun-Hsing Yu, M.D., Ph.D., Elizabeth Healey, S.B., Tze-Yun Leong, Ph.D.,  
Isaac S. Kohane, M.D., Ph.D., and Arjun K. Manrai, Ph.D.

KEY POINTS

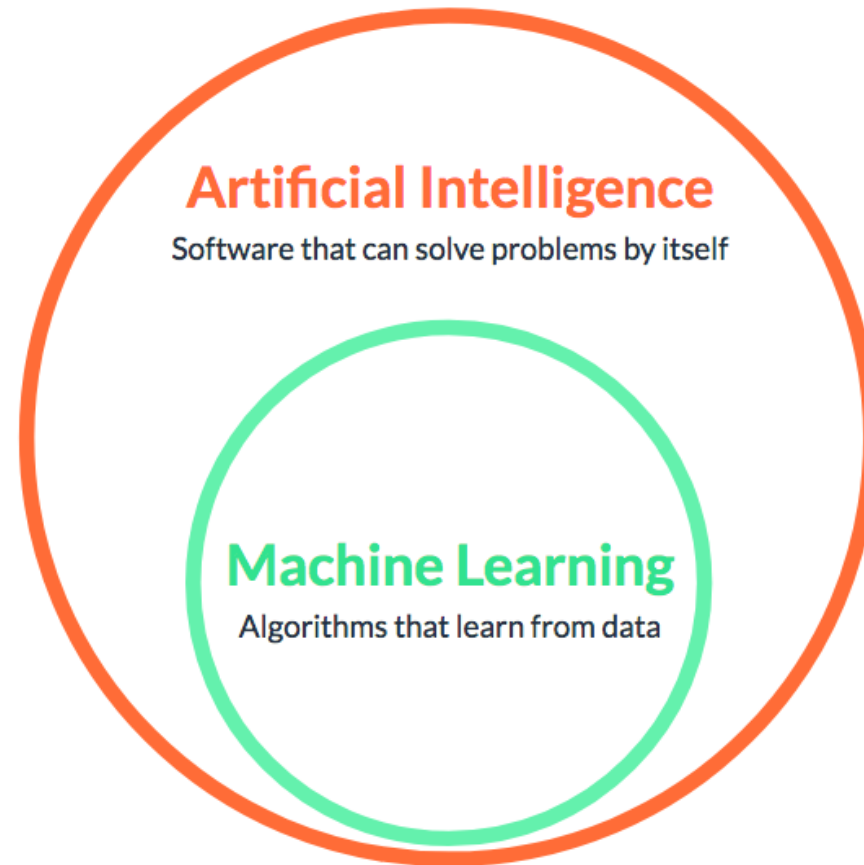
**MEDICAL ARTIFICIAL INTELLIGENCE AND HUMAN VALUES**

- As large language models and other artificial intelligence models are used more in medicine, **ethical dilemmas** can arise depending on how the model was trained. **A user must understand how human decisions and values can shape model outputs.** Medical decision analysis offers lessons on measuring human values.
- **A large language model will respond differently depending on the exact way a query is worded and how the model was directed by its makers and users.** Caution is advised when considering the use of model output in decision making.



# Hintergrund

- Definitionen



# Anwendungen im ärztlichen Alltag

- Anwendungen im ärztlichen Alltag: Kurzfortbildungen für Assistenzärzt\*innen in den vergangenen zwei Jahren (2023-24) -> «Facts and figures».
- Hintergrund: AI-Anwendungen für den ärztlichen Alltag spielerisch kennenlernen und sich damit auseinandersetzen, um zukünftige Herausforderungen meistern zu können.



# Anwendungen im ärztlichen Alltag: AI-Flash I

- Perplexity.ai : \*\*\*Suchen mit Quellenverzeichnis, programmieren von Statistik etc



1 What is the code for linear regression in R?

2 Sources



R Linear Regression Tutorial: lm Function i... - datacamp.com

A linear regression can be calculated in R with the command `lm()`. In the next example, we use this command to calculate estimate height based on...



View 7 more

## Linear Regression in R!!!

```
Call:
lm(formula = size ~ weight, data = mouse.data)
```

Residuals:

Min 10 Median 30 Max

-1.9462 -0.8037 0.1180 0.8180 1.8852

Coefficients:

Estimate Std. Error t value Pr(>|t|)

(Intercept) 0.5813 0.9567 0.603 0.5558

weight 0.7778 0.2334 3.332 0.0024 \*

So Easy!!!!

Watch

3 Answer

To perform linear regression in R, you can use the `lm()` function. Below is a step-by-step guide with code examples for both simple and multiple linear regression.

### Simple Linear Regression

#### 1. Load the Data:

First, load your dataset into R. For example, if your data is in a CSV file:

4

```
r
data <- read.csv("data.csv")
```

#### 2. Fit the Linear Model:

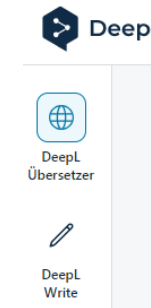
Use the `lm()` function to create a linear regression model. Here, we are predicting `y` based on `x`.

```
r
model <- lm(y ~ x, data = data)
```



# Anwendungen im ärztlichen Alltag: AI-Flash II

- DeepL, Quillbot (<https://quillbot.com/>): Zusammenfassen, \*\*\*paraphrasieren, Grammatik, Ideen, Vorschläge, \*\*\*übersetzen



- Paraphraser
- Grammar Checker
- AI Detector
- Plagiarism Checker
- Summarizer
- Translator
- Citation Generator
- QuillBot Flow
- QuillBot Premium
- QuillBot for Chrome
- Apps & Extensions

- ChatPDF (<https://www.chatpdf.com/>): \*\*\*Interaktive PDF-Analyse





# Anwendungen im ärztlichen Alltag: AI-Flash II

Anokye 2024.pdf

1 / 12

Chat



## ARTICLE IN PRESS

American Journal of  
Preventive Medicine

## REVIEW ARTICLE

## Impact of Cardiovascular Imaging Results on Medication Use and Adherence: A Systematic Review and Meta-Analysis

Reindolf Anokye, MSc, PhD,<sup>1</sup> Jack Dalla Via, PhD,<sup>2</sup> James Dimmock, PhD,<sup>3</sup> Ben Jackson, PhD,<sup>4,5</sup>  
Carl Schultz, MD, PhD,<sup>6,7</sup> Mie Schaeffer, BSc Med,<sup>8</sup> Joanne M. Dickson, PhD,<sup>2,9,10</sup>  
Lauren C. Blekkenhorst, PhD,<sup>2,7</sup> Mandy Stanley, PhD,<sup>1</sup> Jonathan M. Hodgson, PhD,<sup>2,7</sup>  
Joshua R. Lewis, PhD<sup>2,7,11</sup>

**Introduction:** Cardiovascular imaging results offer valuable information that can guide health decisions, but their impact on medication use and adherence is unclear. This systematic review and meta-analysis aimed to determine the downstream impact of cardiovascular imaging results on medication use and adherence.

**Methods:** Searches were conducted across databases, including MEDLINE, PsychINFO, EMBASE, and relevant references up to 2024. Data were extracted from studies comparing outcomes for individuals with diseased versus normal arteries and trials comparing outcomes for individuals who were provided imaging results versus those with no access to imaging results and analysed in 2023 and 2024. Pooled odds ratios (ORs) for outcomes were calculated.

**Results:** The analysis included 29 studies with 24 contributing data points. Initiation (OR:2.77;95% CI:1.82–4.20) and continuation (OR:2.06;95% CI:1.28–3.30) of lipid-lowering medications (LLMs), antihypertensives (OR:2.02;95% CI:1.76–2.33), and antiplatelets (OR:2.47;95% CI:1.68–3.64) were significantly higher in individuals with diseased arteries. The proportion of individuals on LLM increased by 2.7-fold in those with diseased arteries and 1.5-fold in those with normal arteries post-screening. The proportion on LLM increased by 4.2 times in the imaging group and 2.2 times in the “no imaging group” post-screening. There was a significant increase in LLM initiation (OR:2.37;95% CI: 1.17–4.79) in the imaging group, but medication continuation did not significantly differ between the imaging and “no imaging group”.

**Discussion:** Cardiovascular imaging results can prompt initiation of medications, particularly lipid-lowering medications, reflecting a proactive response to identified risk factors. However, evidence regarding medication continuation is mixed, and further research is required.

*Am J Prev Med* 2024;000(000):1–12. © 2024 The Author(s). Published by Elsevier Inc. This is an open access article under the CC BY license (<http://creativecommons.org/licenses/by/4.0/>).

From the <sup>1</sup>School of Medical and Health Sciences, Edith Cowan University, Perth, Western Australia, Australia; <sup>2</sup>Nutrition & Health Innovation Research Institute, School of Medical and Health Sciences, Edith Cowan University, Royal Perth Hospital Research Foundation, Perth, Western Australia, Australia; <sup>3</sup>Department of Psychology, College of Healthcare Sciences, James Cook University, Townsville, Queensland, Australia; <sup>4</sup>Telethon Kids Institute, Perth, Western Australia, Australia; <sup>5</sup>School of Human Sciences (Exercise and Sport Science), The University of Western Australia, Perth, Western Australia, Australia; <sup>6</sup>Cardiology Department, Royal Perth Hospital, Perth, Western Australia, Australia; <sup>7</sup>Medical School, The University of Western Australia, Perth, Western Australia, Australia; <sup>8</sup>Department of Cardiology, Odense University Hospital,

Denmark; <sup>9</sup>School of Arts and Humanities (Psychology Discipline), Edith Cowan University, Joondalup, Western Australia, Australia; <sup>10</sup>Department of Psychological Science, University of Liverpool, UK; and <sup>11</sup>Centre for Kidney Research, Children's Hospital at Westmead, School of Public Health, Sydney Medical School, The University of Sydney, Sydney, New South Wales, Australia  
Address reprint requests to Reindolf Anokye, School of Medical and Health Sciences, Edith Cowan University, 270 Joondalup Drive, Joondalup, WA, 6027, Australia. E-mail: [reindolfanokye@yahoo.com](mailto:reindolfanokye@yahoo.com).  
0749-3797/\$36.00  
<https://doi.org/10.1016/j.amepre.2024.06.008>

© 2024 The Author(s). Published by Elsevier Inc.

*Am J Prev Med* 2024;000(000):1–12 1

This is an open access article under the CC BY license (<http://creativecommons.org/licenses/by/4.0/>).

Hello and welcome! We're excited to share insights from our systematic review and meta-analysis on the impact of cardiovascular imaging results on medication use and adherence. This comprehensive study highlights how imaging can significantly influence the initiation and continuation of essential medications, such as lipid-lowering medications, antihypertensives, and antiplatelets. We hope you find this information valuable in understanding the role of cardiovascular imaging in enhancing patient care!

Here are three example questions you can ask about the file:

- ▶ What were the main findings regarding medication adherence in patients with cardiovascular imaging results?
- ▶ How many studies were included in the analysis, and what types of medications were examined?
- ▶ What databases were searched for relevant studies in this systematic review?

What were the main findings regarding medication adherence in patients with cardiovascular imaging results?

The main findings regarding medication adherence in patients with cardiovascular imaging results indicate that imaging significantly influences both the initiation and continuation of medications. Specifically:

- Medication Initiation:** There was a notable increase in the initiation of lipid-lowering medications (LLMs), antihypertensives, and antiplatelets among individuals with diseased arteries compared to those with normal arteries. The odds ratios (ORs) for initiation were as follows: LLMs (OR: 2.77), antihypertensives (OR: 2.02), and antiplatelets (OR: 2.47) <sup>1</sup>.
- Proportion of Individuals on Medications:** The proportion of individuals on LLMs increased significantly post-screening, with a 2.7-fold increase in those with diseased arteries and a 4.2-fold increase in the imaging group <sup>1</sup>.
- Continuation of Medications:** While initiation rates were significantly higher in the imaging group, the evidence regarding the continuation of medications was mixed. No significant difference in medication continuation rates was observed between the imaging and no-imaging groups, suggesting that while imaging prompts initial treatment, it may not necessarily lead to sustained adherence <sup>10, 11</sup>.

Overall, the findings suggest that cardiovascular imaging results can effectively prompt the initiation of medications, particularly in response to identified risk factors, but further research is needed to understand the factors influencing medication continuation <sup>1</sup>.

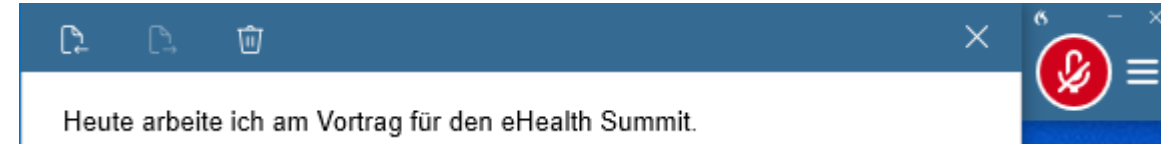
Ask any question...





# Anwendungen im ärztlichen Alltag: AI-Flash II

- \*\*\*Diktat mit Spracherkennung. Z.B. Dragon Medical



- \*\*Leitlinien zu AI in der Akademie

- UniLU Webseite: «Künstliche Intelligenz in der Lehre» <https://www.unilu.ch/universitaet/dienste/lehre/zentrum-lehre/kuenstliche-intelligenz-in-der-lehre/#tab=c154806>
- Unlocking the Power of Generative AI Models and Systems such as GPT-4 and ChatGPT for Higher Education A Guide for Students and Lecturers (Universität Hohenheim) [Unlocking the Power of Generative AI Models and Systems such as GPT-4 and ChatGPT for Higher Education: A Guide for Students and Lecturers \(uni-hohenheim.de\)](https://www.uni-hohenheim.de/unlocking-the-power-of-generative-ai-models-and-systems-such-as-gpt-4-and-chatgpt-for-higher-education-a-guide-for-students-and-lecturers)

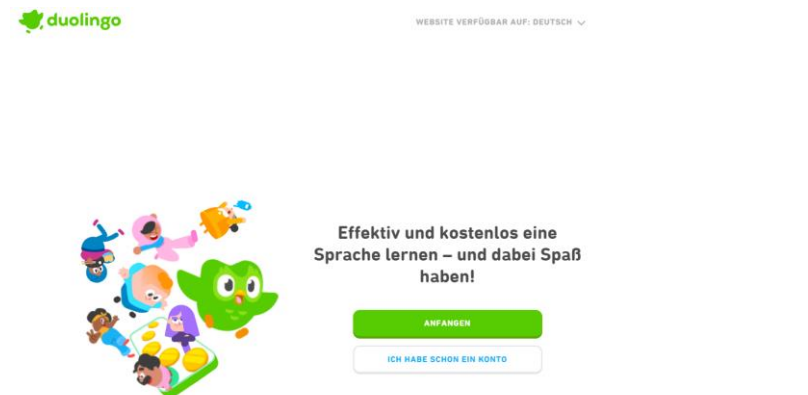
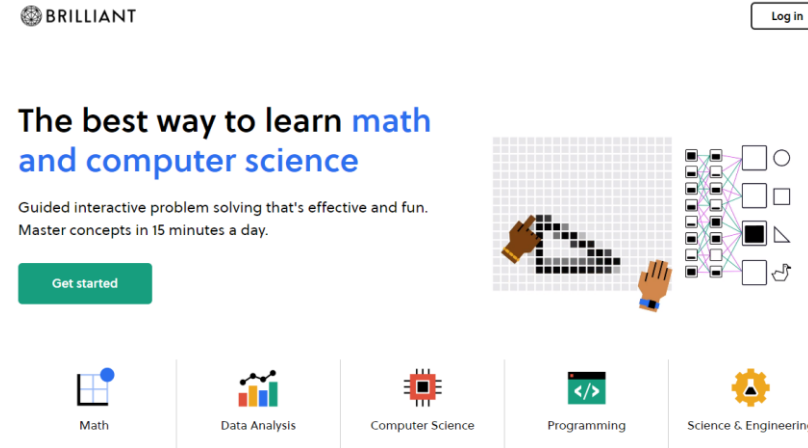
Merke:  
Paraphrasieren OK,  
durch AI schreiben  
lassen nicht

# Anwendungen im ärztlichen Alltag: AI-Flash III



## ■ \*\*\*Lernen

- Trick: nur 5 - 10'/Tag
- Kennt Ihre persönlichen Fehler
- Fehler können separat geübt werden
- Sie können immer und überall lernen



# Anwendungen im ärztlichen Alltag: AI-Flash III



medRxiv  
THE PREPRINT SERVER FOR HEALTH SCIENCES



BMJ Yale

## ■ \*->\*\*\*Entscheiden in der Medizin

- 100 Pat. mit Stroke Symptomen
  - Behandlungsentscheid
  - 90-Tage Mortalität
- 
- Cave: noch nicht klinikreif!

Keine  
klinischen  
Daten in  
ChatBots!

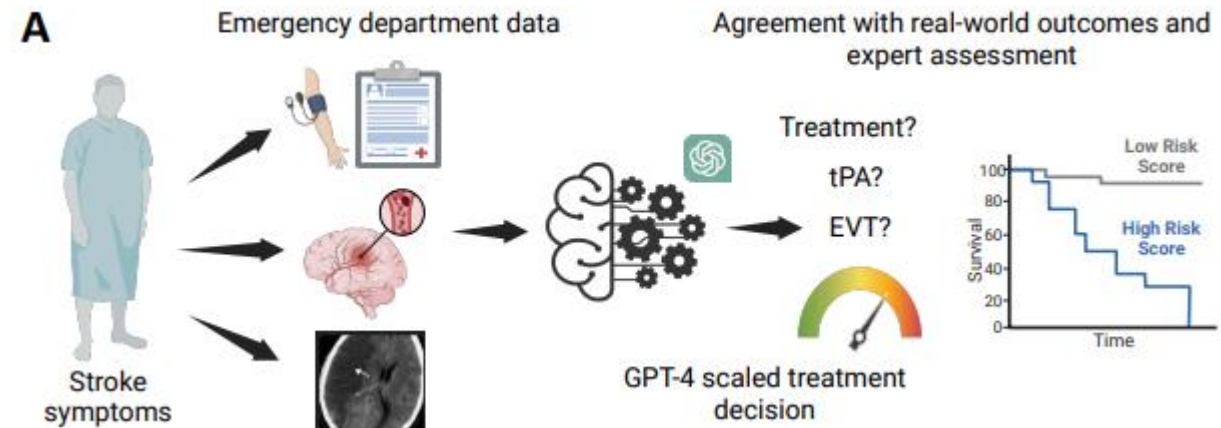
🔔 Follow this preprint

### Evaluating GPT-4 as a Clinical Decision Support Tool in Ischemic Stroke Management

Amit Haim, Mark Katson, Michal Cohen-Shelly, Shlomi Peretz, Dvir Aran, Shahar Shelly

doi: <https://doi.org/10.1101/2024.01.18.24301409>

**This article is a preprint and has not been peer-reviewed [what does this mean?]. It reports new medical research that has yet to be evaluated and so should not be used to guide clinical practice.**





# Anwendungen im ärztlichen Alltag: AI-Flash III



medRxiv

THE PREPRINT SERVER FOR HEALTH SCIENCES



Cold  
Spring  
Harbor  
Laboratory

BMJ Yale

Follow this preprint

## Evaluating GPT-4 as a Clinical Decision Support Tool in Ischemic Stroke Management

Amit Haim, Mark Katson,  Michal Cohen-Shelly, Shlomi Peretz,  Dvir Aran,  Shahar Shelly

doi: <https://doi.org/10.1101/2024.01.18.24301409>

This article is a preprint and has not been peer-reviewed [what does this mean?]. It reports new medical research that has yet to be evaluated and so should *not* be used to guide clinical practice.

decision. The agreement of GPT-4's recommendations with the expert opinion yielded an Area Under the Curve (AUC) of 0.85 [95% CI: 0.77-0.93], and with real-world treatment decisions, an AUC of 0.80 [0.69-0.91]. In terms of mortality prediction, out of 13 patients who died within 90 days, GPT-4 accurately identified 10 within its top 25 high-risk predictions (AUC = 0.89 [95% CI: 0.8077-0.9739]; HR: 6.98 [95% CI: 2.88-16.9]), surpassing supervised machine-learning models. This study demonstrates the potential of GPT-4 as a viable clinical decision support tool



# Anwendungen im ärztlichen Alltag: AI-Flash III

- \*\_>\*\*\*? Entwicklung von Arzneimitteln



Integrated & Experimentally-Validated

**ARTIFICIAL INTELLIGENCE  
FOR EVERY STEP  
OF PHARMACEUTICAL  
RESEARCH AND DEVELOPMENT**



417 0 Kommentare

**Menarini Group und Insilico Medicine schließen exklusiven globalen Lizenzvertrag für neuen KAT6-Inhibitor für mögliche Behandlungen von Brustkrebs und andere onkologische Indikationen ab**

Verfasst von [PR Newswire \(dt.\)](#)

Letzte Änderung 06.01.2024, 13:43

Quelle: Wallstreet Journal online



# Anwendungen im ärztlichen Alltag: AI-Flash VI

- Herausforderung: gem. KPMG alle 70 Tage Verdoppelung des Wissens
- \*\*\*Literatursuche & -Übersichten
- <https://consensus.app/search>
- <https://www.semanticscholar.org>
- Einzelne Papers können gesucht werden
- Zusammenfassungen einzelner und Gruppen von Papers (Co-Pilot Funktion)
- Empfehlenswert für: Übersicht über ein Thema, vertiefte Suchen, Vorbereitung wissenschaftlicher Diskussionen



## Manage Your Papers All in One Place

Store and organize all the papers that interest you in your online library and view it anywhere, anytime when you sign in to Semantic Scholar.



# Anwendungen im ärztlichen Alltag: AI-Flash VI

- \*\*\*Literatursuche & -Übersichten
- <https://consensus.app/search>
- <https://www.semanticscholar.org>
- Übungsfrage: Was ist Malakoplakie?
  1. Consensus: Literatur Zusammenstellung
  2. Co-Pilot: Erklärende Zusammenfassung
  3. Semantic Scholar: Ausdehnung der Literatursuche
  4. ChatPDF: Analyse einzelnes Paper



# Anwendungen im ärztlichen Alltag: AI-Flash VI

- Herausforderung: gem. KPMG alle 70 Tage Verdoppelung des Wissens



- \*\*\*Literatursuche & -Übersichten

- Für Systematic Reviews:
- <https://www.covidence.org>
- Erklärungsvideo: <https://youtu.be/S6OddRoqctY>

The World's #1  
Systematic Review Tool

Reviewers

Organizations

# Ausblick



- Motto: «Die Zukunft ist jetzt»
- Bildanalyse Bsp. Radiologie, Histologie, s. Pathologie LUKS
- Erfassung, Kondensierung und Gewichtung von Real World Data & Integration in EHR (z.B. Vitalzeichen T, BD, P, EKG), Zusammenfassungen Behandlungen durch AI
- Textanalyse, -zusammenfassung Bsp: Bewältigung Med. Literatur (s. Analyse Deloitte 2019,
- Beschleunigung von Planungsprozessen (Dienstplanung LUKS mit AI)
- Patientenkontakt: ChatBots; vgl. auch Banken, Versicherungen etc.
- Ethik Richtlinien: Copy rights, Anwendungen wo, vgl. Genschere CRISPR-CAS9 und chin. Forschung über HIV, Verminderung von Gewalt und Zerstörung, s. Ukrainekrieg, politische Beeinflussung s. US-Wahlen, Cyberkrieg, persönliche Beeinflussung -> z.B. Yu (2024), Daneshvar (2024)
- \*\*\*Artificial Intelligence and Human Values, s. NEJM Yu et al (2024)

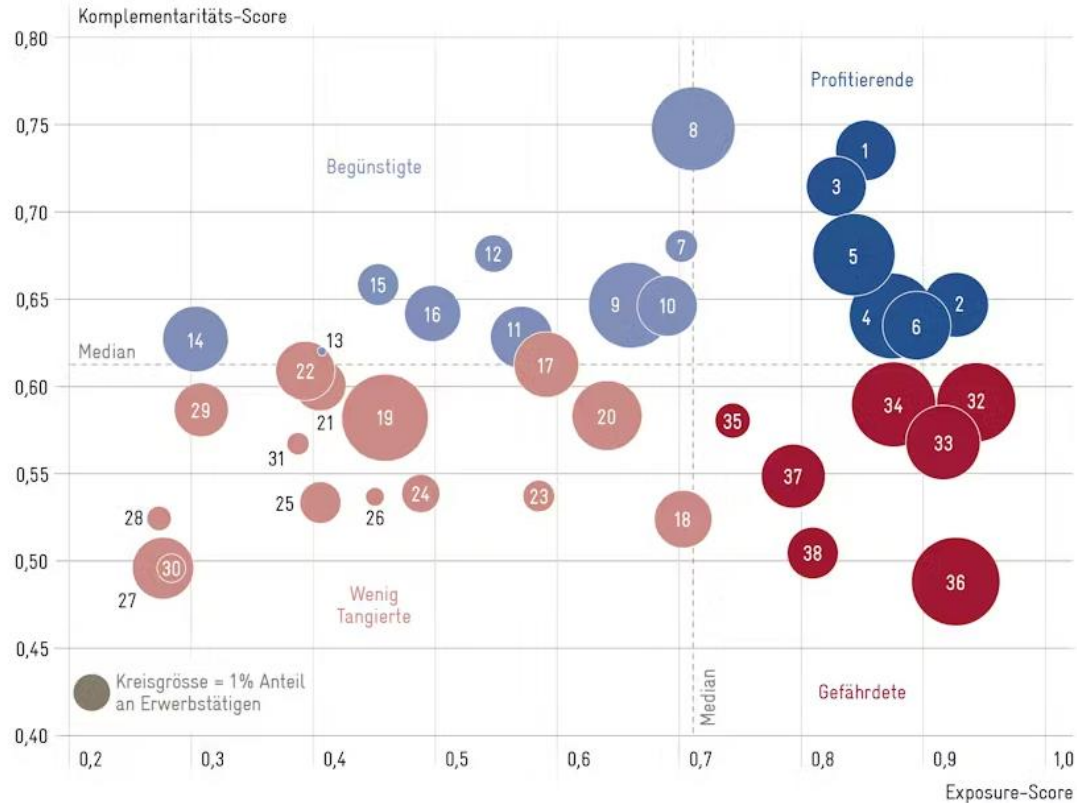
# Ausblick

- Beeinflussung Berufsbild durch KI

## Zukunftssichere Berufe? Wie KI den Schweizer Arbeitsmarkt beeinflusst

### KI wirkt an vielen (Arbeits-)Stellen

Von den Anwendungsmöglichkeiten der KI sind diverse Berufsgruppen betroffen. Vor allem Geschäftsführer, Lehrkräfte sowie Ärzte und verwandtes Gesundheitspersonal dürften aus den Technologien einen produktivitätssteigernden Nutzen ziehen. Für Bürokräfte könnten die neuen Technologien hingegen als Wettbewerber in Erscheinung treten.



**Profitierende:** 1. Geschäftsführer 2. Kaufm. Führungskräfte 3. Führungskräfte Produktion 4. Naturwissenschaftler, Ingenieure 5. Lehrkräfte 6. Sozialwissenschaftler, Juristen  
**Begünstigte:** 7. Führungskräfte Hotel und Gastronomie 8. Ärzte und verwandte Gesundheitsberufe (akad.) 9. Ingenieurtechnische Fachkräfte 10. Juristische, sozialpflegerische Berufe (nicht-akad.) 11. Betreuungskräfte 12. Sicherheitskräfte 13. Forstwirte, Fischer 14. Bau-Fachkräfte 15. Elektriker 16. Fahrer, Bediener mobiler Anlagen  
**Wenig Tangierte:** 17. Assistenzberufe Gesundheit 18. Bürokräfte mit Kundenkontakt 19. Personenbezogene DL 20. Verkaufskräfte 21. Fachkräfte Landwirtschaft 22. Metallarbeiter, Mechaniker 23. Präzisionshandwerker 24. Nahrungsmittelverarbeiter 25. Bediener von Maschinen 26. Montageberufe 27. Reinigungspersonal 28. Hilfsarbeiter Landwirtschaft 29. Hilfsarbeiter Bau 30. Hilfsarbeiter Nahrungsmittelverarbeitung 31. Abfallentsorgung  
**Gefährdete:** 32. Betriebswirte (akad.) 33. IKT-Techniker (akad.) 34. Betriebswirte (nicht-akad.) 35. IKT-Techniker 36. Allg. Bürokräfte 37. Bürokräfte Finanzen 38. Spezialisierte Bürokräfte

Quelle: BIS (SAKE, 2022), eigene Berechnungen

Grafik: Avenir Suisse



# Ausblick

## Annals of Internal Medicine

## POSITION PAPER

### Artificial Intelligence in the Provision of Health Care: An American College of Physicians Policy Position Paper

Nadia Daneshvar, JD, MPH; Deepti Pandita, MD; Shari Erickson, MPH; Lois Snyder Sulmasy, JD; and Matthew DeCamp, MD, PhD; for the ACP Medical Informatics Committee and the Ethics, Professionalism and Human Rights Committee\*

- \*\*\*Daneshvar, N., et al. (2024):

- Bsp:
  - R1: “ACP firmly believes that AI-enabled technologies should complement and not supplant the logic and decision making of physicians and other clinicians.”
  - R4: “ACP reaffirms that AI developers, implementers, and researchers should prioritize the privacy and confidentiality of patient and clinician data collected and used for AI model development and deployment.”
  - R7: “ACP recommends that developers of AI must be accountable for the performance of their models (...).”

Keine  
klinischen  
Daten in AI-  
Programme!



# Ausblick

- \*\*\*Freyer, O., et al. (2024), Lancet Digital Health:

---

**A future role for health applications of large language models depends on regulators enforcing safety standards**

*Oscar Freyer, Isabella Catharina Wiest, Jakob Nikolas Kather, Stephen Gilbert*



# Ausblick

- Freyer, O., et al. (2024)
- ChatPDF:

Freyer 2024.pdf

## Viewpoint

### A future role for health applications of large language models depends on regulators enforcing safety standards

Oscar Freyer, Isabella Catharina Wiest, Jakob Nikolas Kather, Stephen Gilbert

Among the rapid integration of artificial intelligence in clinical settings, large language models (LLMs), such as Generative Pre-trained Transformer-4, have emerged as multifaceted tools that have potential for health-care delivery, diagnosis, and patient care. However, deployment of LLMs raises substantial regulatory and safety concerns. Due to their high output variability, poor inherent explainability, and the risk of so-called AI hallucinations, LLM-based health-care applications that serve a medical purpose face regulatory challenges for approval as medical devices under US and EU laws, including the recently passed EU Artificial Intelligence Act. Despite unaddressed risks for patients, including misdiagnosis and unverified medical advice, such applications are available on the market. The regulatory ambiguity surrounding these tools creates an urgent need for frameworks that accommodate their unique capabilities and limitations. Alongside the development of these frameworks, existing regulations should be enforced. If regulators fear enforcing the regulations in a market dominated by supply or development by large technology companies, the consequences of layperson harm will force belated action, damaging the potentiality of LLM-based applications for layperson medical advice.

#### Introduction

The use of artificial intelligence (AI) in clinical settings has increased since 2015, with various applications being used for specific purposes, such as medical imaging, patient care, or emergency medicine.<sup>1</sup> Since 2022, technological advancements have led to the development of multipurpose algorithms, such as generative AI, with large language models (LLMs) as one of its most noteworthy subsets.<sup>1,2</sup> The introduction of Chat Generative Pre-trained Transformer (GPT) in November 2022,<sup>3</sup> marked a momentous shift, with LLMs rapidly gaining traction across various domains, including health-care research.<sup>2</sup>

Since the introduction of LLMs, various models have been introduced and continuously improved by different developers such as Google, Meta, OpenAI, and Microsoft,<sup>4,5</sup> offering increasingly better performance<sup>6</sup> in, for example, medical applications.<sup>7</sup> Experts expect that LLMs will transform multiple areas of health care—eg, cancer diagnosis and treatment, virtual nursing, predictive medicine, clinical notetaking, and health-care professional-patient interaction.<sup>2,8</sup>

Extensive research efforts are currently dedicated to assessing the use of LLMs within medical settings, particularly as Clinical Decision Support Systems (CDSS). LLM-based applications that are not limited to health care, such as ChatGPT, have shown great potential in timesaving for radiologists in the classification of fractures,<sup>9</sup> as a diagnostic tool evaluated with artificial case vignettes in which the majority of responses aligned with established medical guidelines,<sup>10</sup> and as a CDSS for health-care professionals, effectively generating diagnostic insights from presented symptom patterns.<sup>11</sup> LLM-based applications that are focused for health care, such as Med-Gemini, perform well in multiple medical tests (eg, the MedQA benchmark, which is based on medical board exams such as the US Medical Licensing Examination), medical text summarisation, medical reasoning, and

multimodal analysis, including radiology images.<sup>12,13</sup> The performance of individual LLMs and LLM-based systems, such as GPT-3.5, GPT-4, ChatGPT, Articulate Medical Intelligence Explorer, and Llama 2, was also measured in application scenarios for layperson end users, showing growing potential and superiority over a simple internet search,<sup>14</sup> or in text-based consultations when compared with primary care physicians.<sup>5</sup> However, validating and translating these results into useful and safe clinical applications remains challenging. We are in an era where conflicting reports from many supporters and critics are being published about the accuracy and suitability of these approaches to medicine. Several publications show the disappointing performance of LLM-based applications in a series of medical tasks, including poor performance in generating medical billing codes,<sup>15</sup> high error rates in oncology questions,<sup>16</sup> potentially harmful<sup>18</sup> and prolonged patient communication,<sup>19</sup> and misleading risk stratifications of patients with chest pain<sup>18</sup> or breast cancer.<sup>20</sup> Additionally, researchers have identified open ethical and legal questions regarding data privacy and intellectual property,<sup>21</sup> and the perpetuation of racial and gender biases.<sup>22</sup>

In summary, studies showing highly encouraging, positive performance of LLMs have often been limited by their evaluation of real-world medical scenarios, whereas studies showing disappointing performance have generally been based on unoptimised approaches with generic LLM-based applications and do not use technologies deemed important for LLMs in medicine, such as retrieval-augmented generation.<sup>2,24</sup> These studies highlight the importance of rigorous and independent evaluation frameworks, such as the Developmental and Exploratory Clinical Investigations of Decision support systems driven by Artificial Intelligence,<sup>25</sup> and the need for development of independent test platforms—an early example of which is the UK AI Safety Institute's open-source framework.<sup>26</sup>



Lancet Digit Health 2024;  
6: e662-72  
Eike Kröner Fresenius Center for  
Digital Health, TUD Dresden  
University of Technology,  
Dresden, Germany (O Freyer,  
IC West Dr med,  
Prof JN Kather Dr med,  
Prof S Gilbert PhD), Department  
of Medicine, Medical Faculty  
Mannheim, Heidelberg  
University, Mannheim,  
Germany (C Wiest),  
Department of Medicine,  
University Hospital Dresden,  
Dresden, Germany  
(Prof JN Kather), Medical  
Oncology, National Center for  
Tumor Diseases, University  
Hospital Heidelberg,  
Heidelberg, Germany  
(Prof JN Kather)  
Correspondence to:  
Prof Stephen Gilbert, The Kröner  
Fresenius Center for Digital  
Health, TUD Dresden University  
of Technology, Dresden 01302,  
Germany  
stephen.gilbert@tu-dresden.de

## Chat

### Hello and Welcome! 🌟

We're excited to share insights from the latest viewpoint published in *The Lancet Digital Health*. This article discusses the promising role of large language models (LLMs) in healthcare, highlighting their potential to enhance clinical delivery, diagnosis, and patient care. However, it also emphasizes the importance of regulatory oversight to address safety concerns, variability in outputs, and the challenges posed by AI hallucinations. Join us in exploring the future of AI in health applications!

Here are three example questions you can ask about the file:

- ▶ What are the main benefits of using large language models in healthcare?
- ▶ What regulatory concerns are associated with the deployment of LLMs in medical applications?
- ▶ How do AI hallucinations impact the reliability of LLMs in clinical settings?

### What are the main benefits of using large language models in healthcare?

The main benefits of using large language models (LLMs) in healthcare include:

- Enhanced Clinical Decision Support:** LLMs can assist healthcare professionals by providing evidence-based recommendations, improving diagnostic accuracy, and offering insights from vast medical literature, which can lead to better patient outcomes 1.
- Improved Patient Interaction:** These models can facilitate more natural and effective communication between patients and healthcare providers, enabling better understanding and engagement in their care 1.
- Automation of Administrative Tasks:** LLMs can help automate routine administrative tasks, such as documentation and data entry, allowing healthcare professionals to focus more on patient care rather than paperwork 1.
- Personalized Medicine:** By analyzing patient data and preferences, LLMs can contribute to personalized treatment plans, enhancing the relevance and effectiveness of medical interventions 1.
- Access to Information:** LLMs can provide patients and healthcare providers with quick access to medical information, guidelines, and resources, which can be particularly beneficial in urgent care situations 1.

These benefits highlight the potential of LLMs to transform various aspects of healthcare delivery, making it more efficient and patient-centered.

Ask any question...



# Ausblick: Weiterführende Literatur

- Institute for Health Care Improvement

- [www.ihl.org](http://www.ihl.org)

## Training

### The Impact of AI on the Safety of Patients

August 19, 2024

## Publications

### Patient Safety and Artificial Intelligence: Opportunities and Challenges for Care Delivery

## News

### IHI Lucian Leape Institute Releases Artificial Intelligence (AI) Recommendations for Patient Safety and Health System Leaders

May 15, 2024

## Publications

### Artificial Intelligence in Health Care: Implications for Patient and Workforce Safety

# Literatur

- Daneshvar, N., et al. (2024). "Artificial Intelligence in the Provision of Health Care: An American College of Physicians Policy Position Paper." Ann Intern Med.
- Haim, A. et al. (2024). "Evaluating GPT-4 as a Clinical Decision Support Tool in Ischemic Stroke Management." medRxiv.
- Freyer, O. e. a. (2024). "A future role for health applications of large language models depends on regulators enforcing safety standards." Lancet Digit Health.
- Yu, K. H., et al. (2024). "Medical Artificial Intelligence and Human Values." N Engl J Med 390(20): 1895-1904.

# Zusammenfassung

- Die Zukunft ist jetzt; AI-Anwendungen im sicheren Rahmen benützen
- Persönliche Einstellung:
  - **Nüchtern bleiben** (facts and figures),
  - **Chancen sehen und ausloten** (spielerisches Kennenlernen),
  - **Risiken bemerken und angehen** (Plausibilitätschecks, Verständnis Training der Modelle u.a.)
  - Kritiklose Euphorie (hypes and likes) und kritiklose Akzeptanz sind gefährlich
- Die Entwicklung von AI-Anwendungen hat erst begonnen

# Fragen

- Kontakt: [balthasar.hug@luks.ch](mailto:balthasar.hug@luks.ch) oder [balthasar.hug@unilu.ch](mailto:balthasar.hug@unilu.ch)

