

# KI im Gesundheitswesen

## Zusammenarbeit von Mensch und Maschine

Fabian Lechner M.Sc.

Institut für Künstliche Intelligenz in der Medizin  
Philipps Universität Marburg

Philipps



Universität  
Marburg



# Conflict of Interests

## Forschung:

- UKGM Institut für KI in der Medizin
- UKGM Institut für digitale Medizin
- Universität Bonn

## Vorträge:

- Siemens, Lilly, Novonordisk
- DDG, DKG, DEKG



# Referent

- Fabian Lechner M.Sc.
- B.Sc. in int. Business Management an der FH-Aachen
- M.Sc. in Wirtschaftsinformatik an der Philipps Uni Marburg
- Aktuell als Forscher und Doktorand am Institut für KI in der Medizin am UKGM
  
- Forschungsgebiete: NLP and LLM sowie multimodal RAG
  
- Kontakt: [fabian.lechner@uni-marburg.de](mailto:fabian.lechner@uni-marburg.de)

# Künstliche Intelligenz

Künstliche Intelligenz (KI) ist ein Bereich der Informatik, der mithilfe von Algorithmen kognitive Fähigkeiten in Computersystemen nachbildet, mit dem Ziel, diese Systeme dazu zu befähigen, Aufgaben selbstständig zu bearbeiten und sich an veränderte Bedingungen anzupassen, ohne dass jeder Schritt vorgegeben wird. Dabei bezeichnet KI auch Systeme, die Verhalten zeigen, welches typischerweise menschliche Intelligenz erfordert.

# Künstliche Intelligenz

Künstliche Intelligenz (KI) ist ein **Bereich der Informatik**, der mithilfe von Algorithmen **kognitive Fähigkeiten** in Computersystemen **nachbildet**, mit dem Ziel, diese Systeme dazu zu befähigen, **Aufgaben selbstständig zu bearbeiten** und sich an **veränderte Bedingungen anzupassen**, ohne dass jeder Schritt vorgegeben wird. Dabei bezeichnet KI auch Systeme, die Verhalten zeigen, welches **typischerweise menschliche Intelligenz** erfordert.

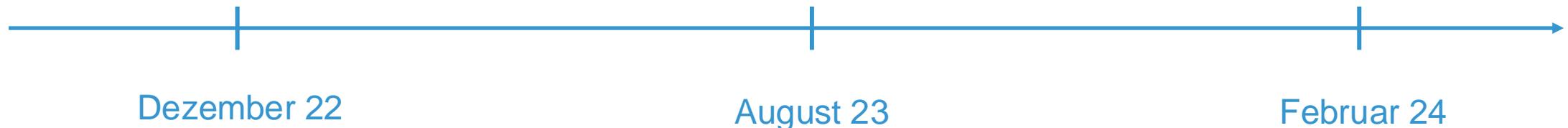
## Generative KI (Generative AI)

# Generative KI (Generative AI)

**You**  
Stell mir komplizierte Fragen zu Diabetes typ 2, die ein diabetologe anhand von Fachwissen gut beantworten könnte

**ChatGPT**  
Das Verständnis und die Behandlung von Typ-2-Diabetes erfordern umfassendes Fachwissen. Hier sind einige komplizierte Fragen, die das tiefe Verständnis eines Diabetologen oder eines Spezialisten im Bereich Endokrinologie herausfordern könnten:

1. **Integration von GLP-1-Rezeptoragonisten und SGLT2-Inhibitoren in die Behandlung:** Wie kann die Kombination von GLP-1-Rezeptoragonisten mit SGLT2-Inhibitoren im Vergleich zu traditionellen Therapiemethoden, wie Metformin und Sulfonylharnstoffen, die kardiovaskulären und renalen Outcomes bei Patienten mit Typ-2-Diabetes verbessern?
2. **Personalisierte Medizin und genetische Prädisposition:** Welche Rolle spielen genetische Variationen bei der Wirksamkeit und den Nebenwirkungen von antidiabetischen Medikamenten in der Behandlung von Typ-2-Diabetes, und wie kann die personalisierte Medizin genutzt werden, um individuelle Behandlungspläne zu entwickeln?
3. **Intermittierendes Fasten und Typ-2-Diabetes:** Wie beeinflusst intermittierendes Fasten den Glukosestoffwechsel und die Insulinempfindlichkeit bei Patienten mit Typ-2-Diabetes, und welche Evidenz unterstützt seine Anwendung als Teil eines umfassenden Behandlungsplans?
4. **Mikrobiota und Typ-2-Diabetes:** Welche Erkenntnisse gibt es über den Zusammenhang zwischen Darmmikrobiota und der Entwicklung sowie der Progression von Typ-2-Diabetes, und wie könnten Modifikationen der Darmflora als therapeutische Strategie genutzt werden?
5. **Advanced Glycation End Products (AGEs) und ihre Rolle bei Typ-2-Diabetes:** Inwiefern beeinflussen AGEs und deren Rezeptoren (RAGE) die Insulinresistenz und die Entstehung diabetischer Komplikationen, und welche Strategien existieren zur Reduktion der AGE-Belastung?



ChatGPT 

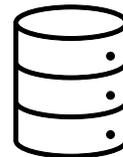
# Generative

Das nächste Wort wird  
probabilistisch berechnet



# Pretrained

Die KI ist auf riesigen  
Datenmengen trainiert



# Transformer

Die Architektur des  
neuronalen Netzes hinter  
GPT

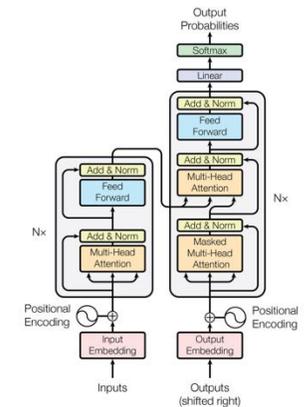


Figure 1: The Transformer - model architecture.

# Generieren

*“Insulin is not a cure for diabetes, it’s a \_\_\_\_\_”*

# Generieren

[ Insulin is not a cure for diabetes, it's a ]

Input



Neuronales Netz  
(LLM)

treatment



necessity



lifeline



requirement



medication



support



solution



Output

# Generieren

*“Insulin is not a cure for diabetes, it's a treatment”*

Frederik Banting

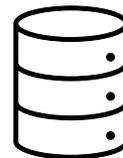
# Generative

Das nächste Wort wird  
probabilistisch berechnet



# Pretrained

Die KI ist auf riesigen  
Datenmengen trainiert



# Transformer

Die Architektur des  
neuronalen Netzes hinter  
GPT

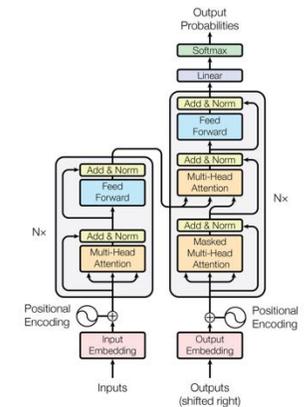


Figure 1: The Transformer - model architecture.

# Pretraining - Vortraining

 **arthurmensch** [Mistral AI\\_org](#) Oct 12, 2023  

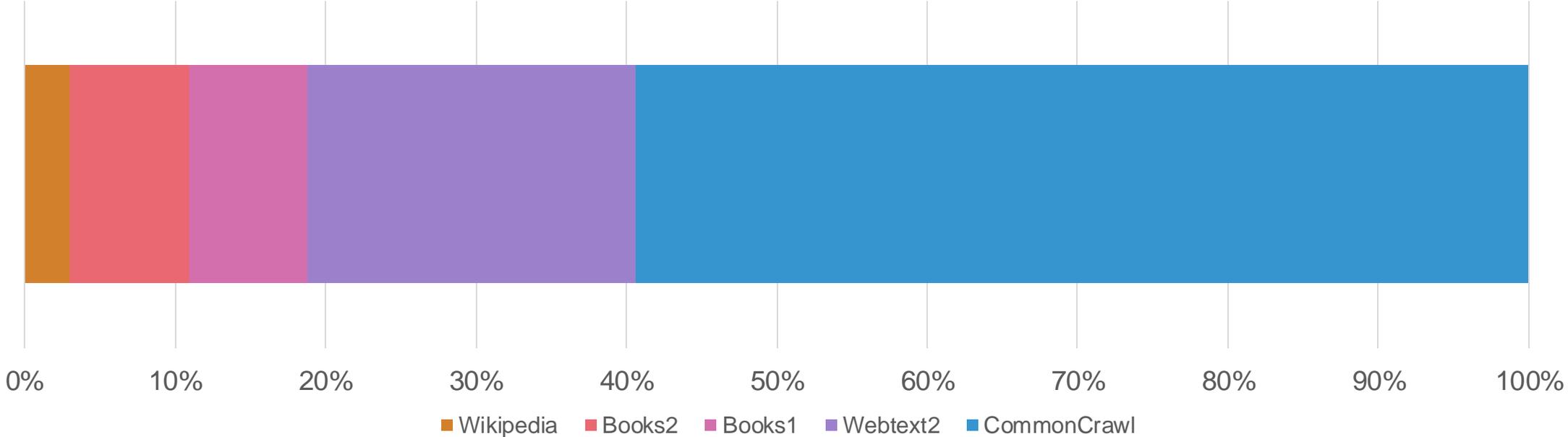
Hello, thanks for your interest and kind words! Unfortunately we're unable to share details about the training and the datasets (extracted from the open Web) due to the highly competitive nature of the field. We appreciate your understanding!

[See translation](#)

 13  4  4  1 

Trainingsdaten sind Geschäftsgeheimnis!

# Pretraining - Vortraining



500.000.000.000

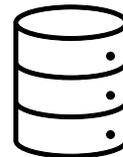
# Generative

Das nächste Wort wird  
probabilistisch berechnet



# Pretrained

Die KI ist auf riesigen  
Datenmengen trainiert



# Transformer

Die Architektur des  
neuronalen Netzes hinter  
GPT

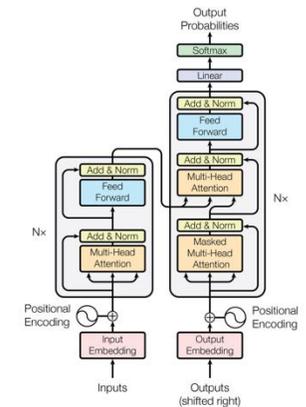
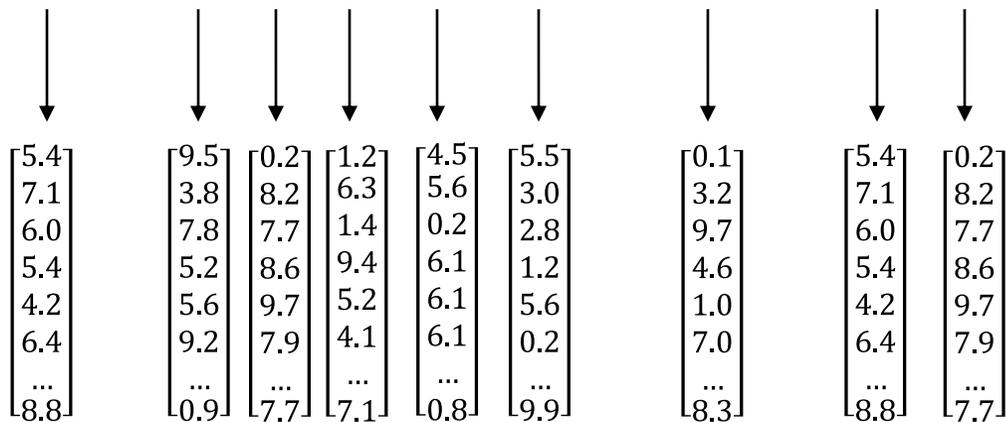


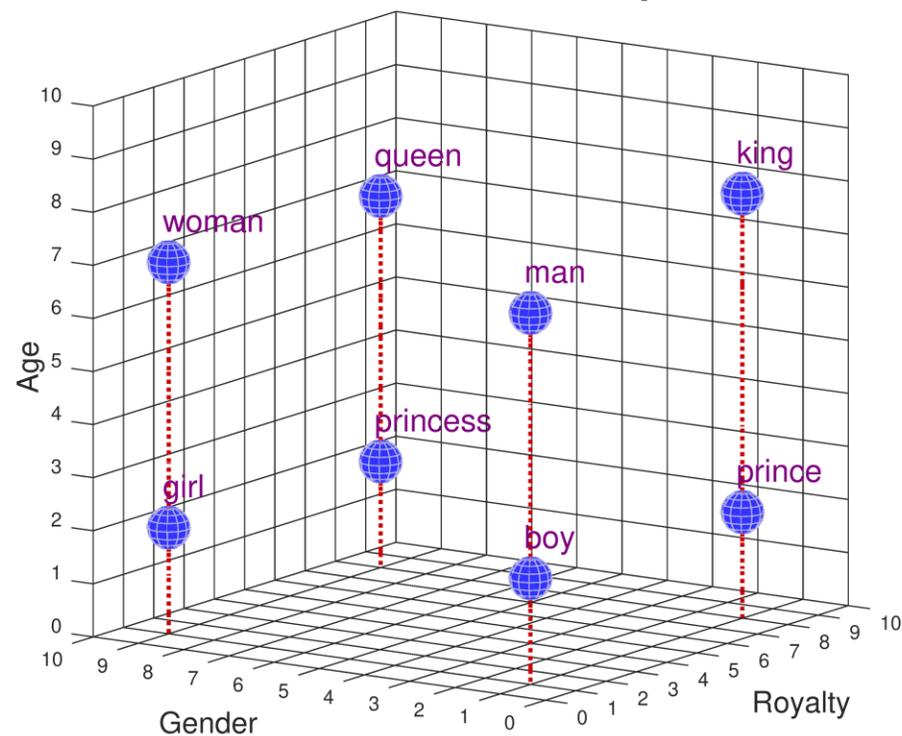
Figure 1: The Transformer - model architecture.

# Transformer - Embeddings

[Insulin is not a cure for diabetes, it's a \_\_\_\_]



3D Semantic Feature Space



# Transformer - Attention

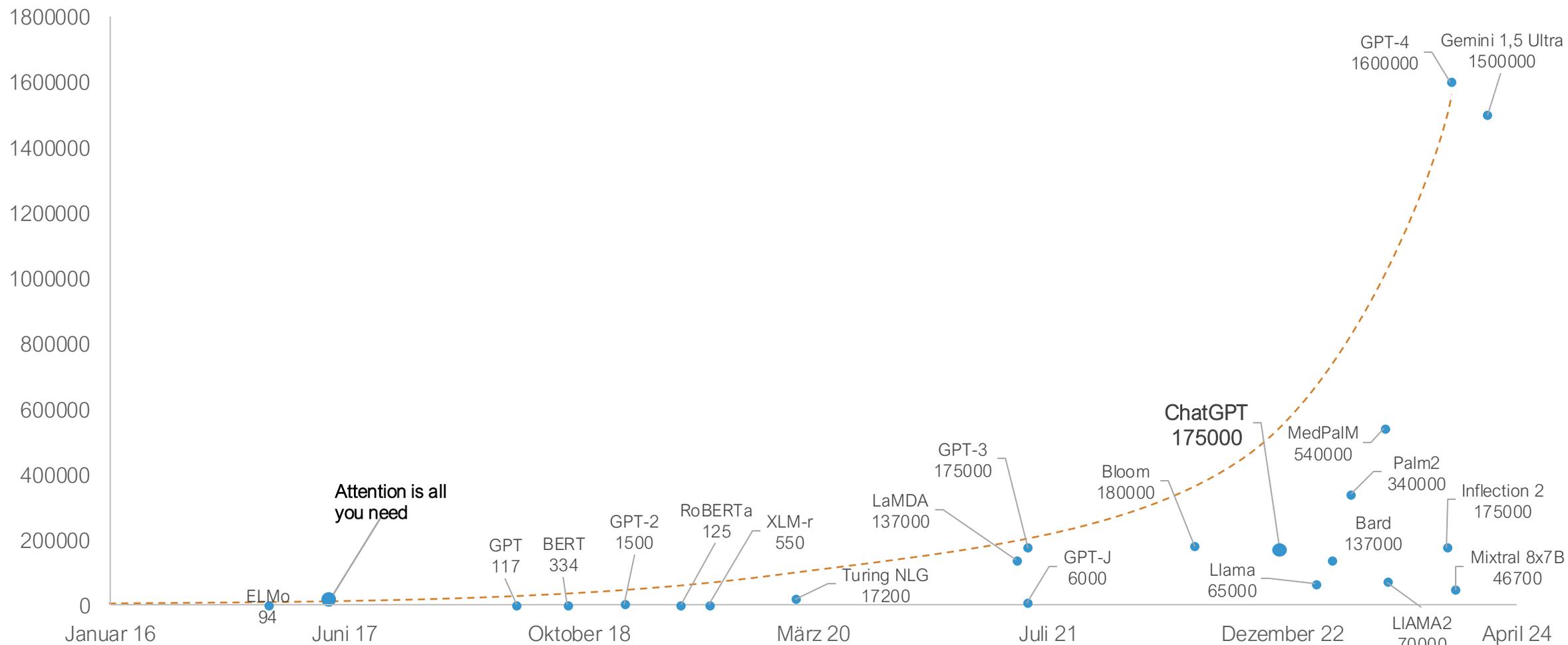


# Transformer - Attention

[**Insulin** is **not** a **cure** for **diabetes**, it's a \_\_\_\_]

# Model Größen

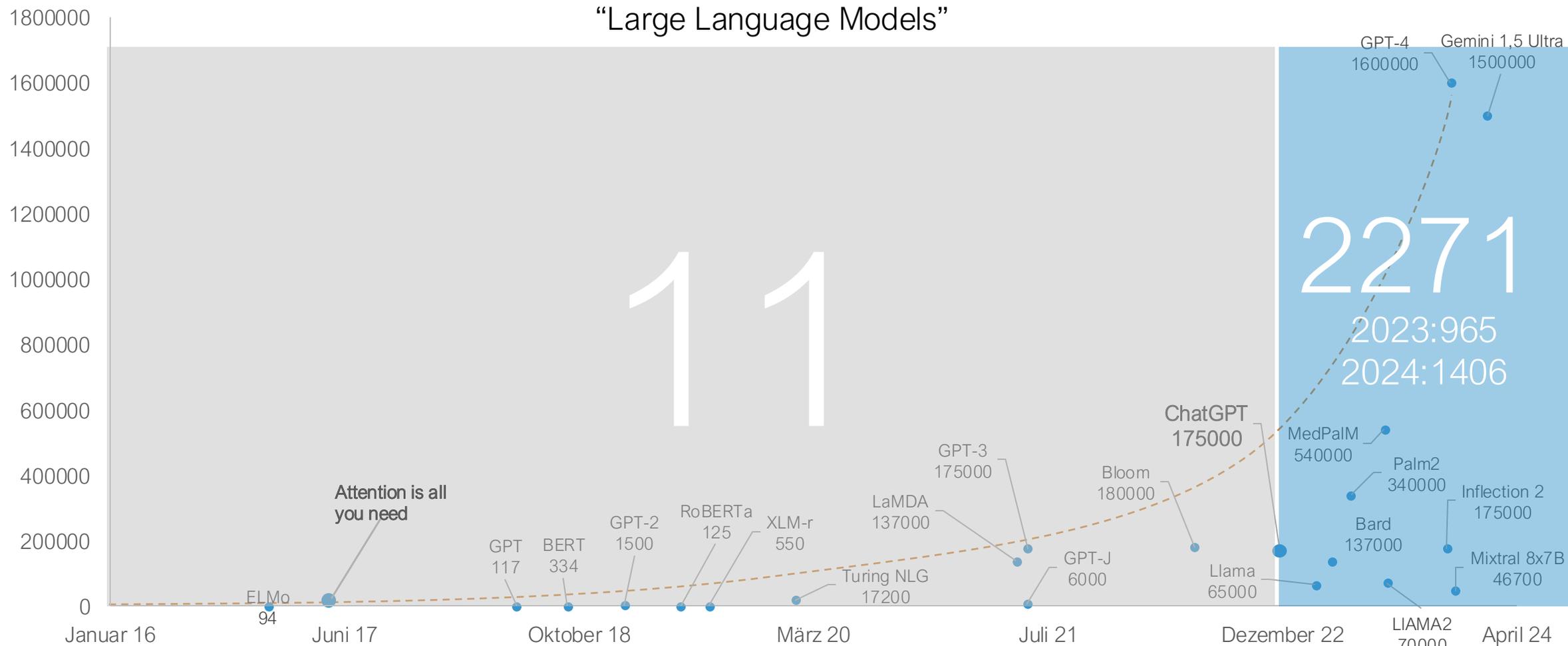
Modelgröße in Millionen Parametern



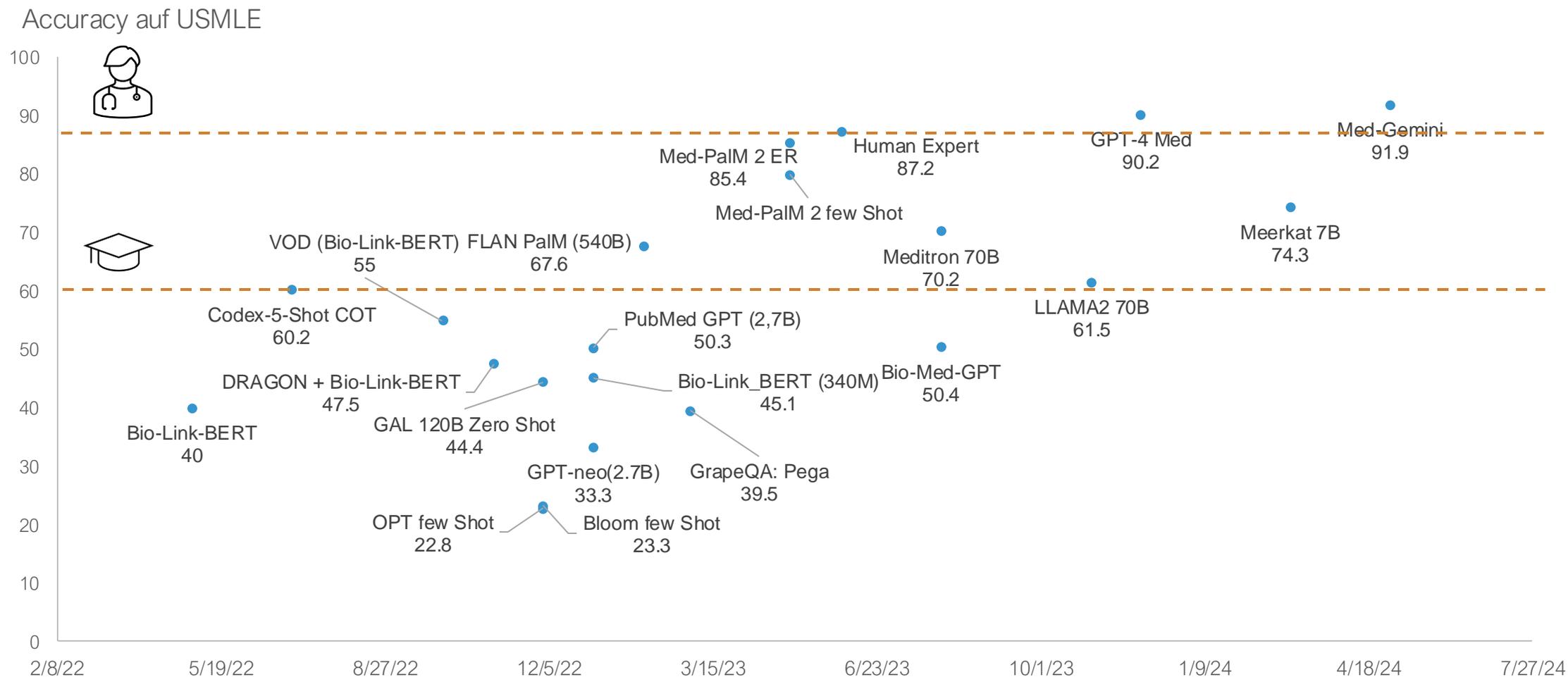
# Model Größen

Modelgröße in Millionen Parametern

“Large Language Models”



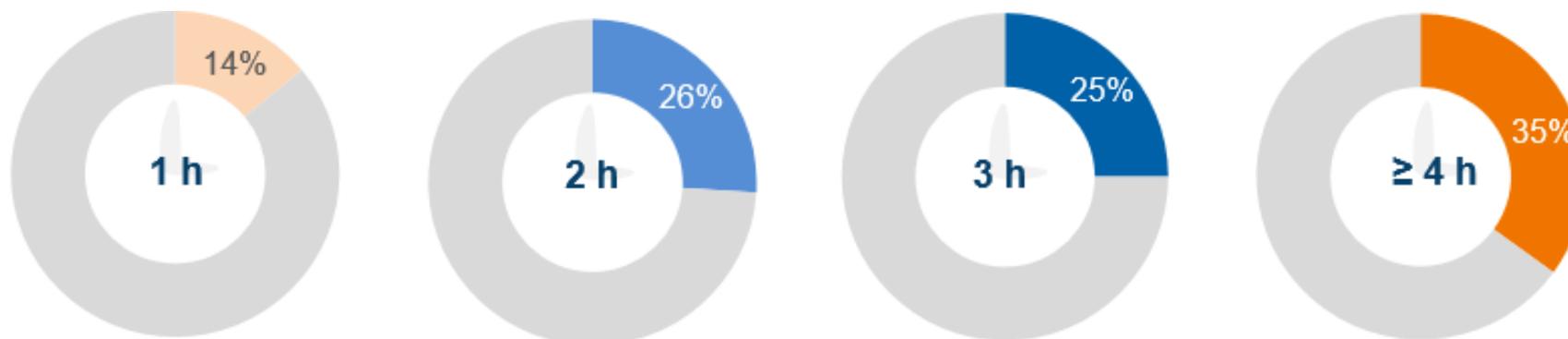
# LLMs auf medizinischen Daten



In der Forschung für die Klinik!

# Dokumentation – Status Quo

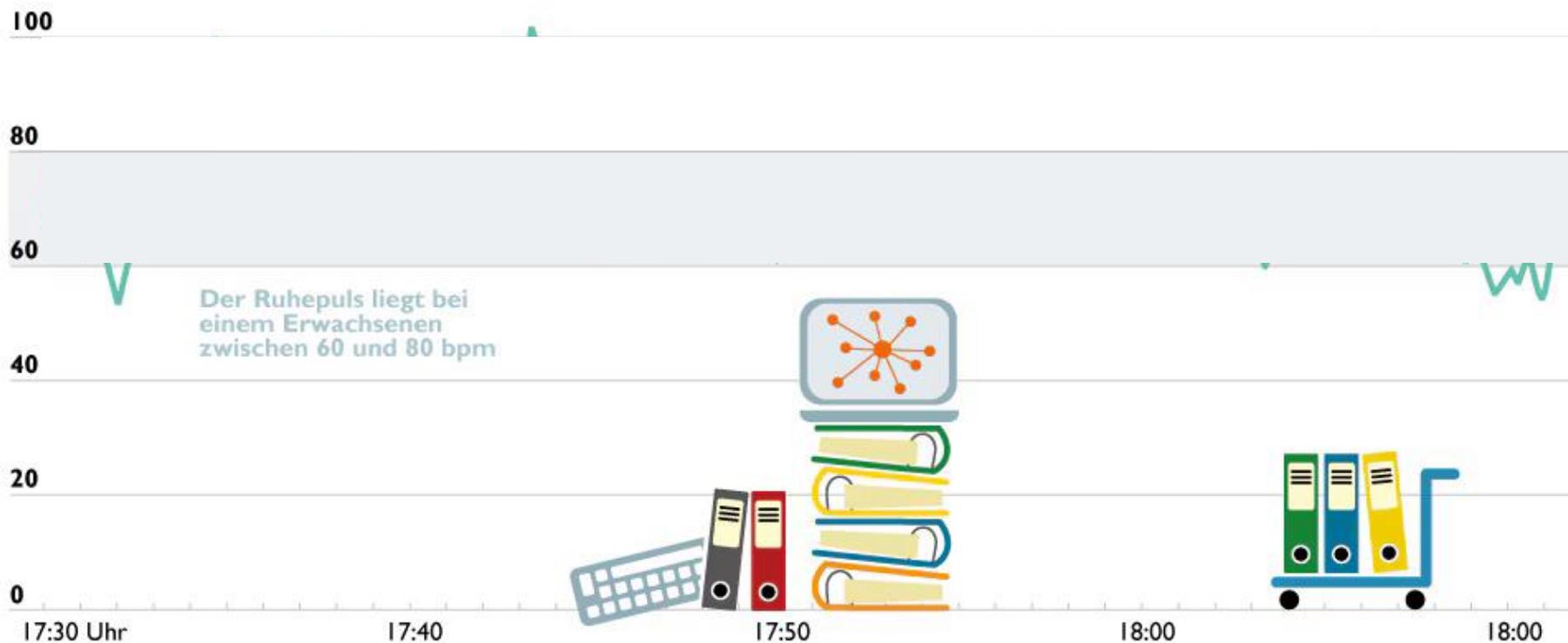
Wie hoch ist ihr täglicher Zeitaufwand für Verwaltungstätigkeiten und Organisation, die über die ärztliche Tätigkeit hinausgehen? (z.B. Datenerfassung und Dokumentation, OP-Voranmeldung)



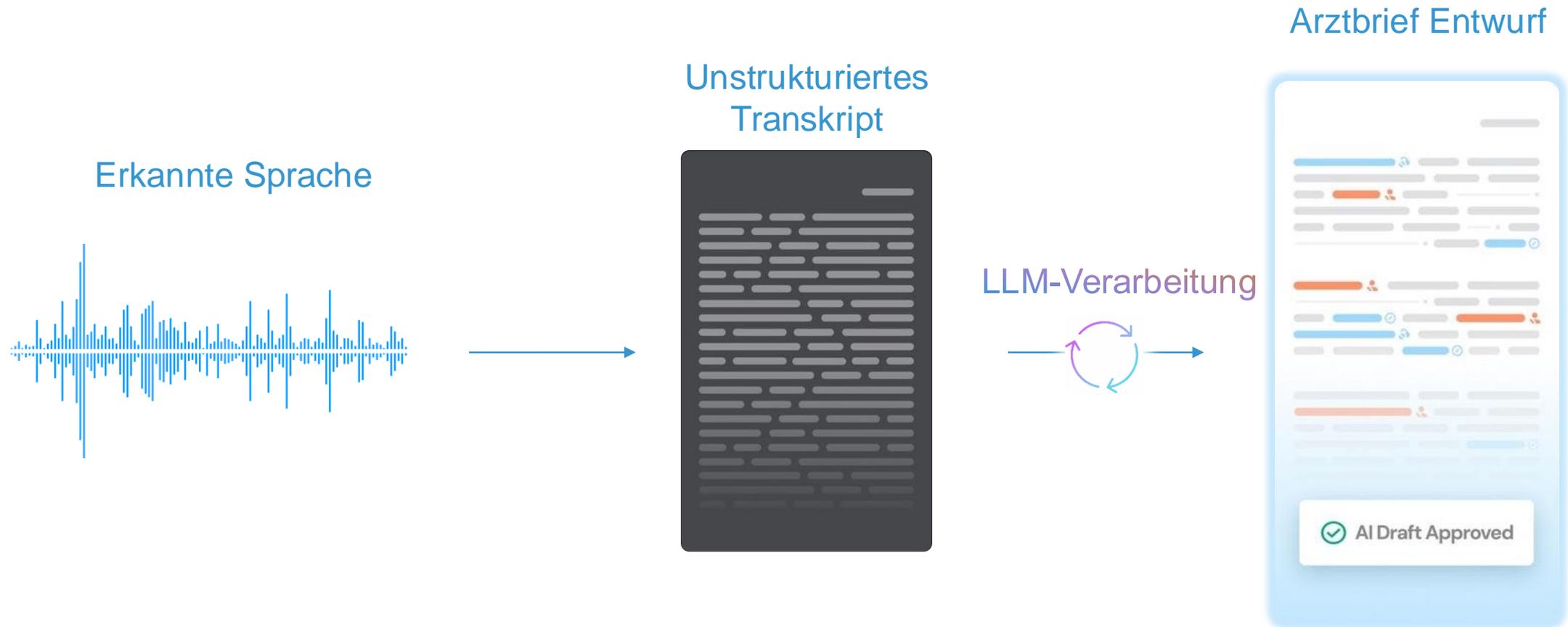
# Dokumentation – Status Quo

## Pulsfrequenz des Arztes Schade in einer Dokumentationssituation (PC-Arbeit)

120 Herzschläge pro Minute (bpm)



# Dokumentation - Ausblick



# Dokumentation – Ausblick



3. If symptoms worsen or persist after a week, consider prescription antibiotics.
4. Encourage to quit smoking and offer resources for smoking cessation, given the chronic health conditions.
5. Follow-up in 2 weeks for blood pressure and diabetes management and to ensure resolution of current symptoms.
6. Counsel on the importance of regular exercise and maintaining a healthy diet for the management of his diabetes and high blood pressure.

*Okay.*

*Before we start, I just wanted to confirm your details.*

*So it's Paul Collins and you're 46 years old and is this your address?*

*That's fine.*

*Perfect.*

*Great.*

*Okay.*

*Excellent.*

*So how can I help you today?*

*Well, I've had a cough for a few days though.*

*It's just driving me nuts really.*

*Right.*

*How long has the cough been there for?*

*I suppose about three or four days now.*

*That's it.*

*And have you got any other symptoms with that cough?*

*I've got quite sore throat.*

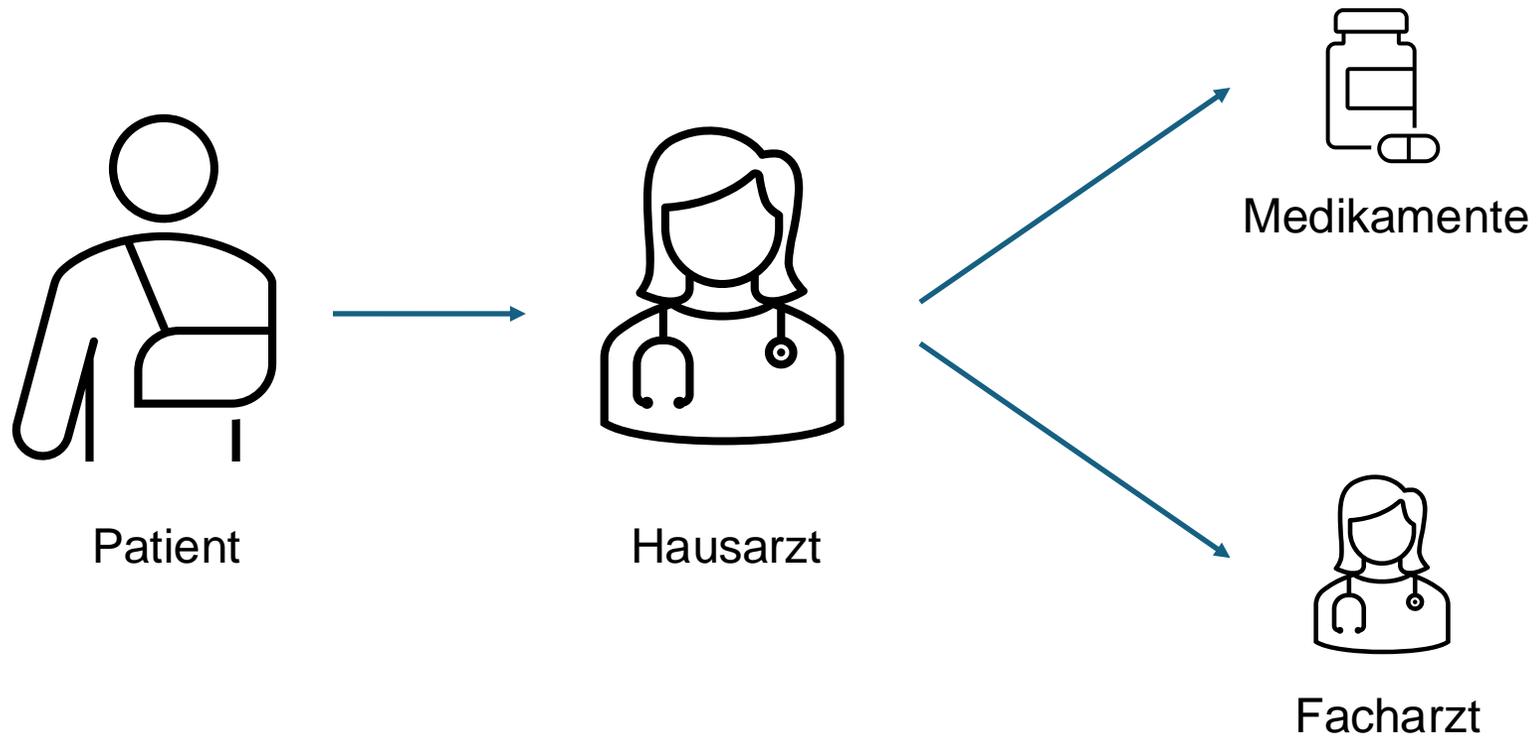
*That's quite bad as well.*

*Right.*

*And a sore throat has it been there the same amount of time?*

*Yes.*

# Diagnose – Status Quo

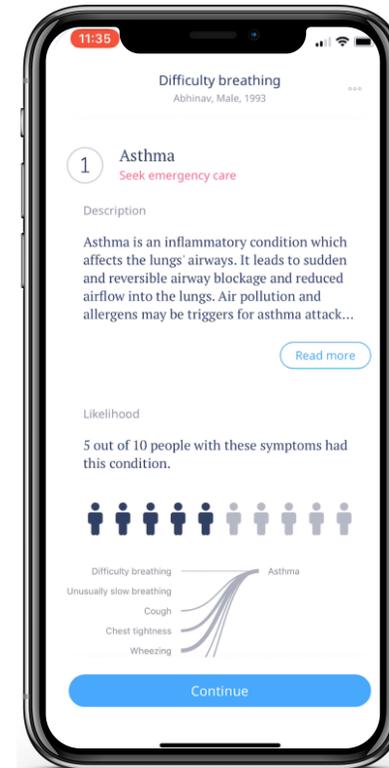
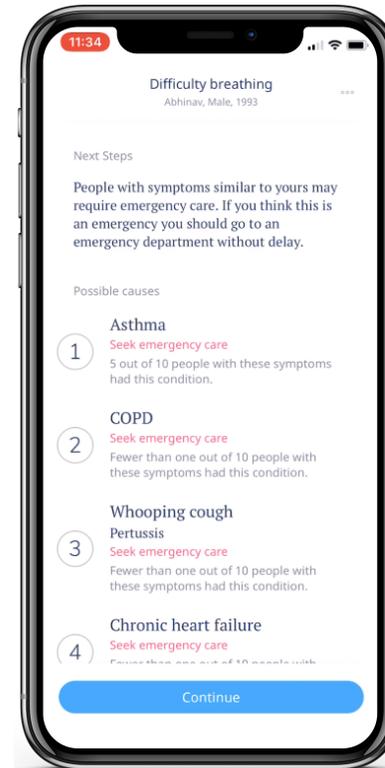
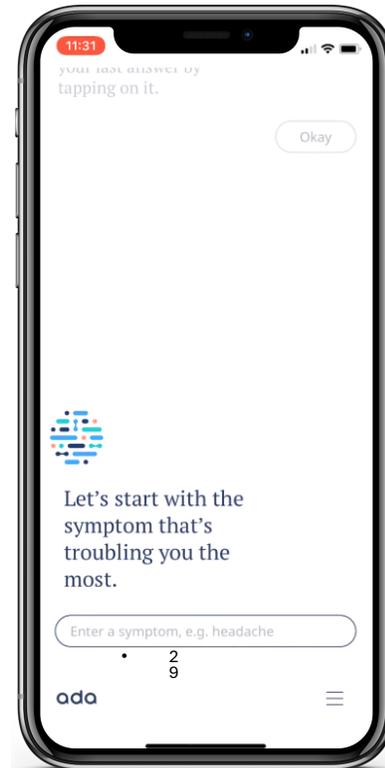
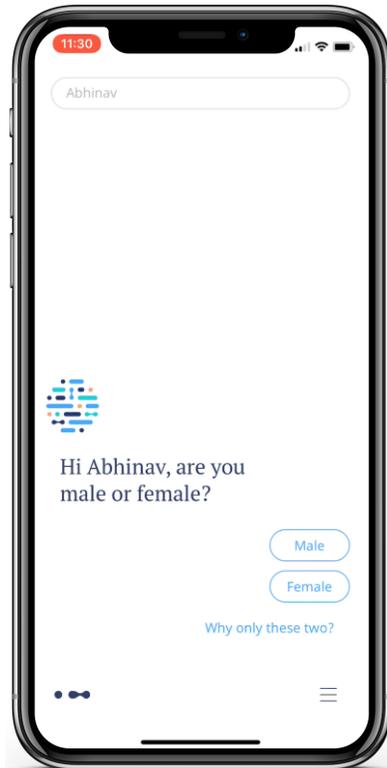


# Diagnose - ADA



Dr. Claire Novorol, Daniel Nathrath, Prof. Dr. Martin Hirsch

# Diagnose - ADA



# Diagnose - ADA

Open access

Original research

## BMJ Open How accurate are digital symptom assessment apps for suggesting conditions and urgency advice? A clinical vignettes comparison to GPs

Stephen Gilbert<sup>1</sup>, Alicia Mehl,<sup>1</sup> Adel Baluch,<sup>1</sup> Caoimhe Cawley,<sup>1</sup> Jean Challiner,<sup>1</sup> Hamish Fraser,<sup>2</sup> Elizabeth Millen,<sup>1</sup> Maryam Montazeri,<sup>1</sup> Jan Multmeier,<sup>1</sup> Fiona Pick,<sup>1</sup> Claudia Richter,<sup>1</sup> Ewelina Türk,<sup>1</sup> Shubhanan Upadhyay,<sup>1</sup> Vishaal Virani,<sup>1</sup> Nicola Vona,<sup>1</sup> Paul Wicks,<sup>1</sup> Claire Novorol<sup>1</sup>

To cite: Gilbert S, Mehl A, Baluch A, et al. How accurate are digital symptom assessment apps for suggesting conditions and urgency advice? A clinical vignettes comparison to GPs. *BMJ Open* 2020;10:e040269. doi:10.1136/bmjopen-2020-040269

Received 11 May 2020  
Revised 27 October 2020  
Accepted 16 November 2020

Check for updates

© Author(s) (or their employer(s)) 2020. Re-use permitted under CC BY-NC. No commercial re-use. See rights and permissions. Published by BMJ.

<sup>1</sup>Ada Health GmbH, Berlin, Germany  
<sup>2</sup>Brown Center for Biomedical Informatics, Brown University, Rhode Island, USA

Correspondence to: Dr Stephen Gilbert; science@ada.com

### ABSTRACT

**Objectives** To compare breadth of condition coverage, accuracy of suggested conditions and appropriateness of urgency advice of eight popular symptom assessment apps.

**Design** Vignettes study.

**Setting** 200 primary care vignettes.

**Intervention/comparator** For eight apps and seven general practitioners (GPs): breadth of coverage and condition-suggestion and urgency advice accuracy measured against the vignettes' gold-standard.

**Primary outcome measures** (1) Proportion of conditions covered by an app, that is, not excluded because the user was too young/old or pregnant, or not modelled; (2) proportion of vignettes with the correct primary diagnosis among the top 3 conditions suggested; (3) proportion of 'safe' urgency advice (ie, at gold standard level, more conservative, or no more than one level less conservative).

**Results** Condition-suggestion coverage was highly variable, with some apps not offering a suggestion for many users: in alphabetical order, Ada: 99.0%; Babylon: 51.5%; Buoy: 88.5%; K Health: 74.5%; Mediktork: 80.5%; Symptomate: 61.5%; Your.MD: 64.5%; WebMD: 93.0%. Top-3 suggestion accuracy was GPs (average): 82.1%±5.2%; Ada: 70.5%; Babylon: 32.0%; Buoy: 43.0%; K Health: 36.0%; Mediktork: 36.0%; Symptomate: 27.5%; WebMD: 35.5%; Your.MD: 23.5%. Some apps excluded certain user demographics or conditions and their performance was generally greater with the exclusion of corresponding vignettes. For safe urgency advice, tested GPs had an average of 97.0%±2.5%. For the vignettes with advice provided, only three apps had safety performance within 1 SD of the GPs—Ada: 97.0%; Babylon: 95.1%; Symptomate: 97.8%. One app had a safety performance within 2 SDs of GPs—Your.MD: 92.6%. Three apps had a safety performance outside 2 SDs of GPs—Buoy: 80.0% (p<0.001); K Health: 81.3% (p<0.001); Mediktork: 87.3% (p=1.3×10<sup>-3</sup>).

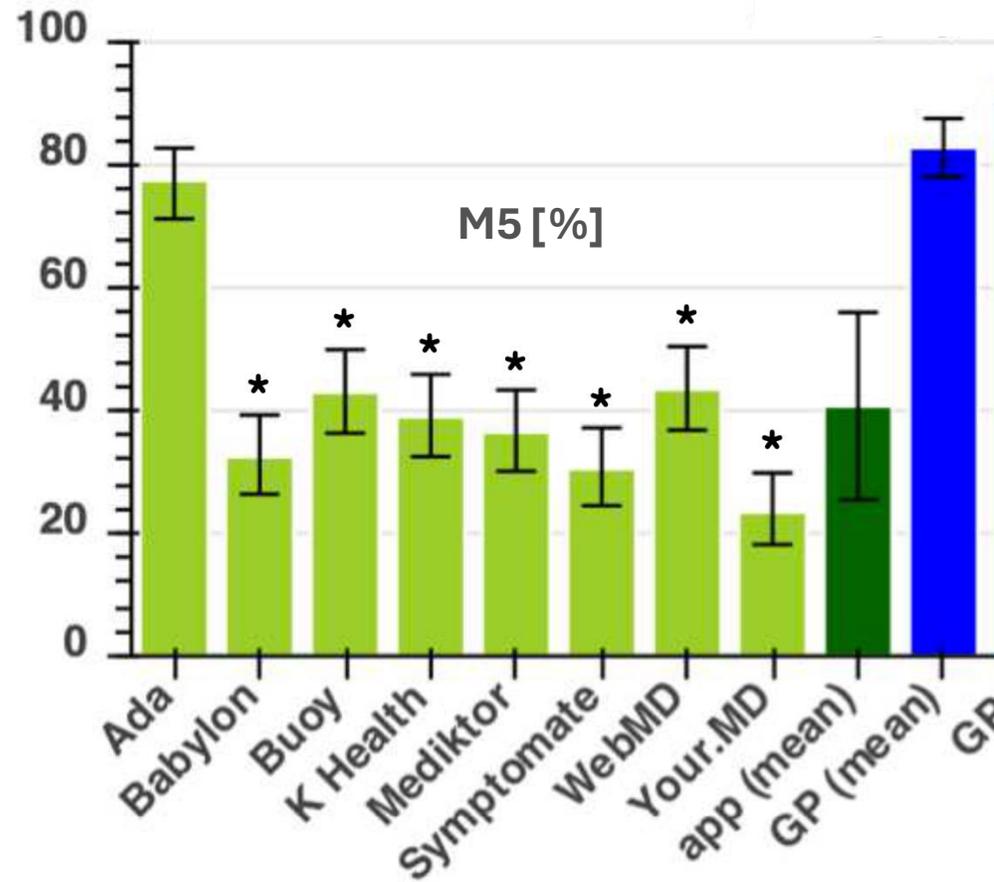
**Conclusions** The utility of digital symptom assessment apps relies on coverage, accuracy and safety. While no digital tool outperformed GPs, some came close, and the nature of iterative improvements to software offers scalable improvements to care.

### Strengths and limitations of this study

- ▶ The study included a large number of vignettes which were peer reviewed by independent and experienced primary care physicians to minimise bias.
- ▶ General practitioners and apps were tested with vignettes in a manner that simulates real clinical consultations.
- ▶ Detailed source data verification was carried out.
- ▶ Vignette entry was conducted by professionals as a recent study found that laypeople are less good at entering vignettes for symptoms that they have never experienced.
- ▶ Limitations include the lack of a rigorous and comprehensive selection process to choose the eight apps and the lack of real patient experience assessment.

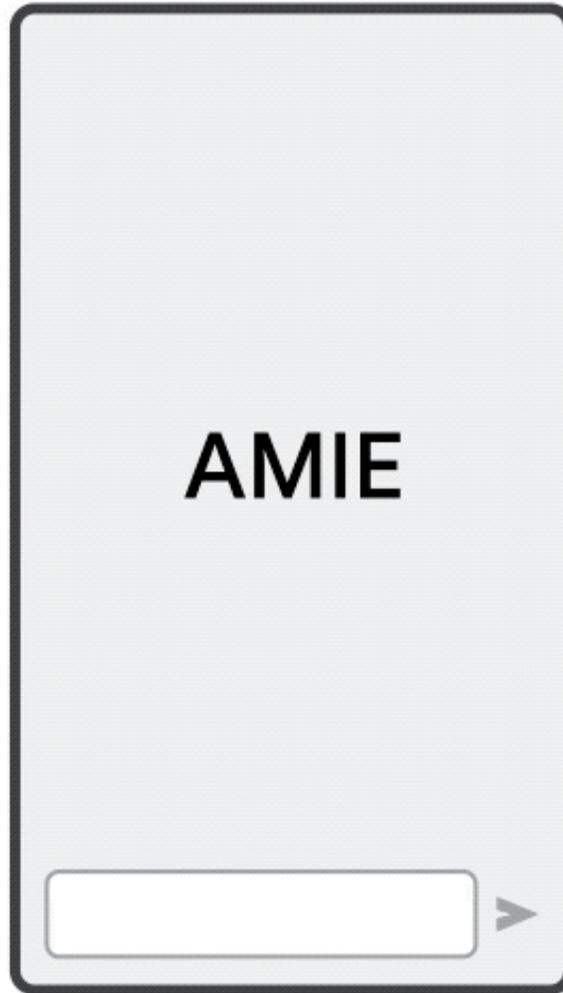
### INTRODUCTION

Against the background of an ageing population and rising pressure on medical services, the last decade has seen the internet replace general practitioners (GPs) as the first port of call for health information. A 2010 survey of over 12 000 people from 12 countries reported that 75% of respondents search for health information online,<sup>1</sup> with some two-thirds of patients in 2017 reporting that they 'google' their symptoms before going to the doctor's office.<sup>2</sup> However, online search tools like Google or Bing were not intended to provide medical advice and risk offering irrelevant or misleading information.<sup>3</sup> One potential solution is dedicated symptom assessment applications (ie, apps)<sup>4-6</sup> which use a structured interview or multiple-choice format to ask patients questions about their demographic, relevant medical history, symptoms, and presentation. In the first few screening questions, some symptom assessment apps

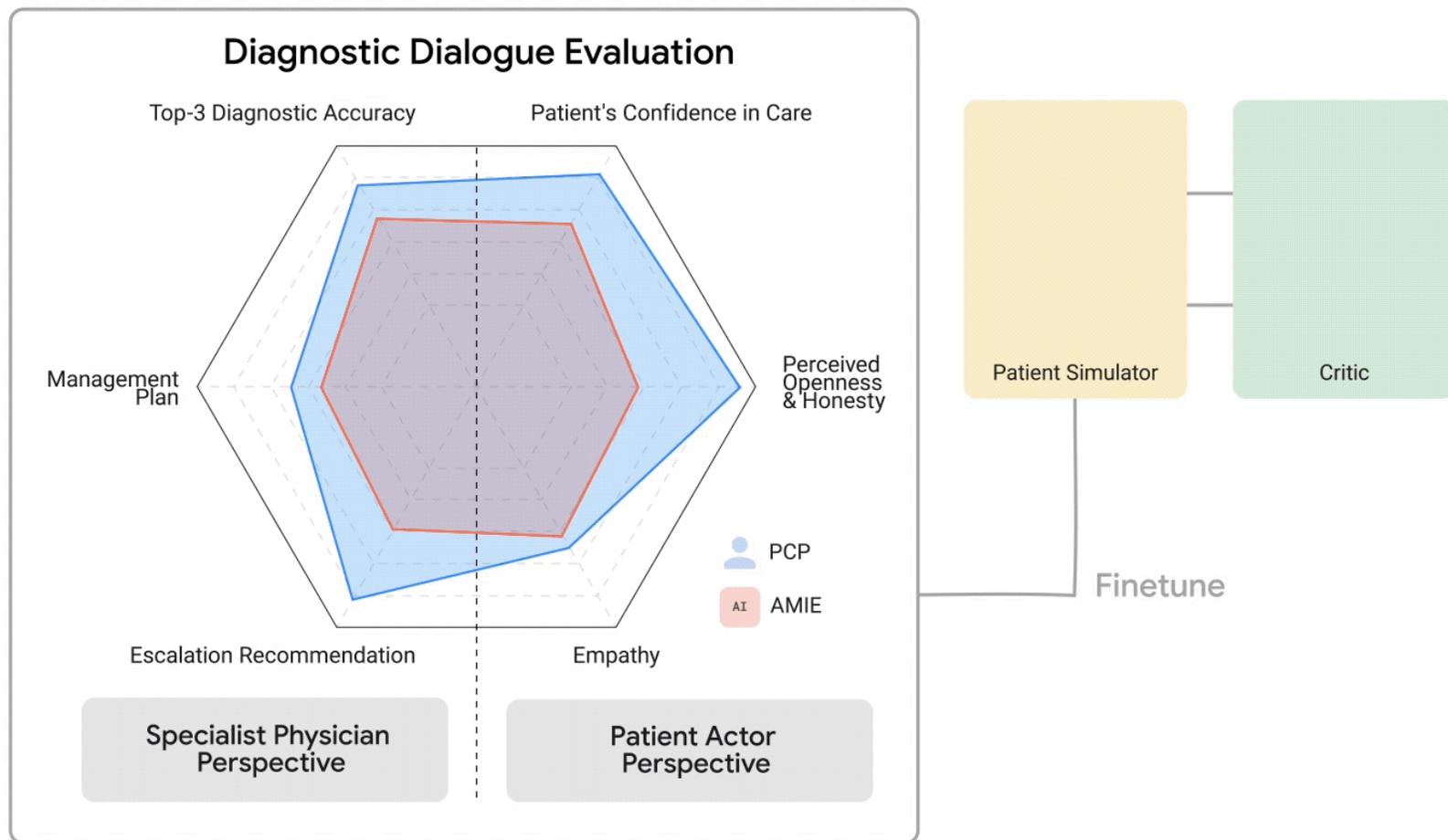


BMJ Open: first published as 10.1136/bmjopen-2020-040269 on 16 December 2020. Downloaded from http://bmjopen.bmj.com/ on January 12, 2021 by guest. Protected by copyright.

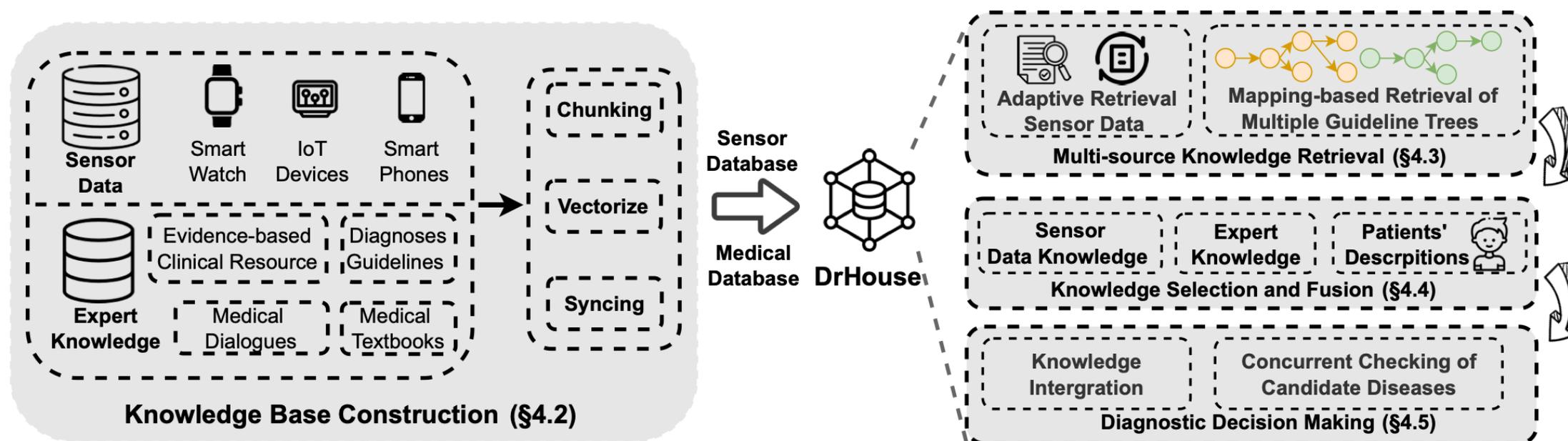
# Diagnose - AIME



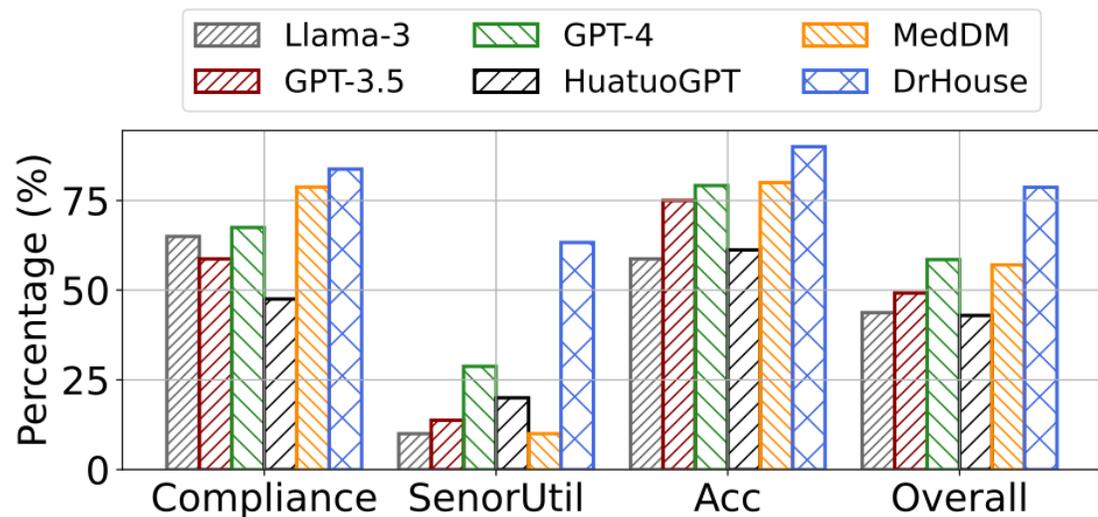
# Diagnose - AIME



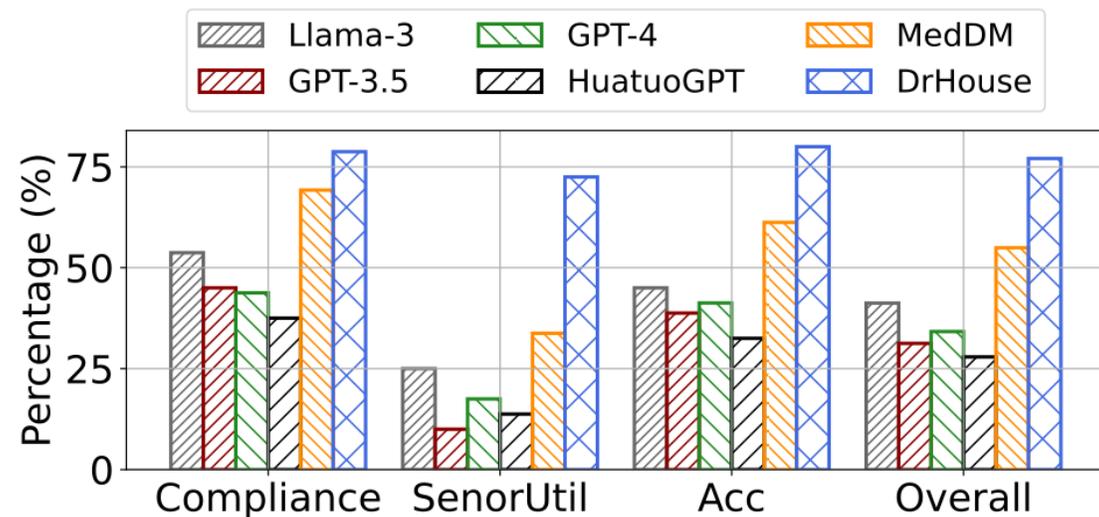
# Diagnose – DrHouse



# Diagnose – DrHouse



(a) MedDG dataset.



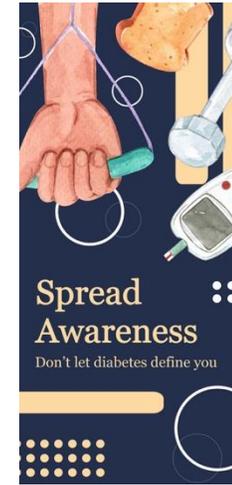
(b) KaMed dataset.

# Aufklärung – Status Quo

**Table 4.** Distribution of the disease-related knowledge levels of the patients (n = 335)

Disease Characteristics	Accurately informed		Misinformed		Does not know	
	Number	%	Number	%	Number	%
Organ from whence the diabetes originates	69	20.6	30	9.0	236	70.4
Symptoms of falling blood sugar levels	156	46.6	34	10.1	145	43.3
Organs damaged by diabetes	125	37.3	33	9.9	177	52.8
How to clip his or her nails	112	33.4	72	21.5	151	45.1
What to pay attention to while wearing shoes	123	36.7	62	18.5	150	44.8
The frequency of foot washing and checks	154	46.0	57	17.0	124	37.0
How to store the insulin pen	69	20.6	48	14.3	218	65.1
Dietary needs of a diabetic	233	69.6	48	14.3	54	16.1
Number of meals a diabetic should eat	214	63.9	61	18.2	60	17.9
Benefits of exercise	237	70.7	28	8.4	70	20.9
Knowledge that smoking should be avoided	276	82.4	14	4.2	45	13.4

# Aufklärung – Status Quo



- Persönlich
- Interaktiv
- Begrenzte Zeit
- Inhalte werden vergessen

- Immer verfügbar
- Unpersönlich
- Begrenzter Inhalt
- Nicht interaktiv

# Aufklärung - Zukunftsausblick



Gesundheitsdaten aus App



Bilder z.B. Symptome



Aufklärung-LLM

Literatur aus Leitlinien AWMF  
Arbeitsgemeinschaft der Wissenschaftlichen  
Medizinischen Fachgesellschaften e.V.

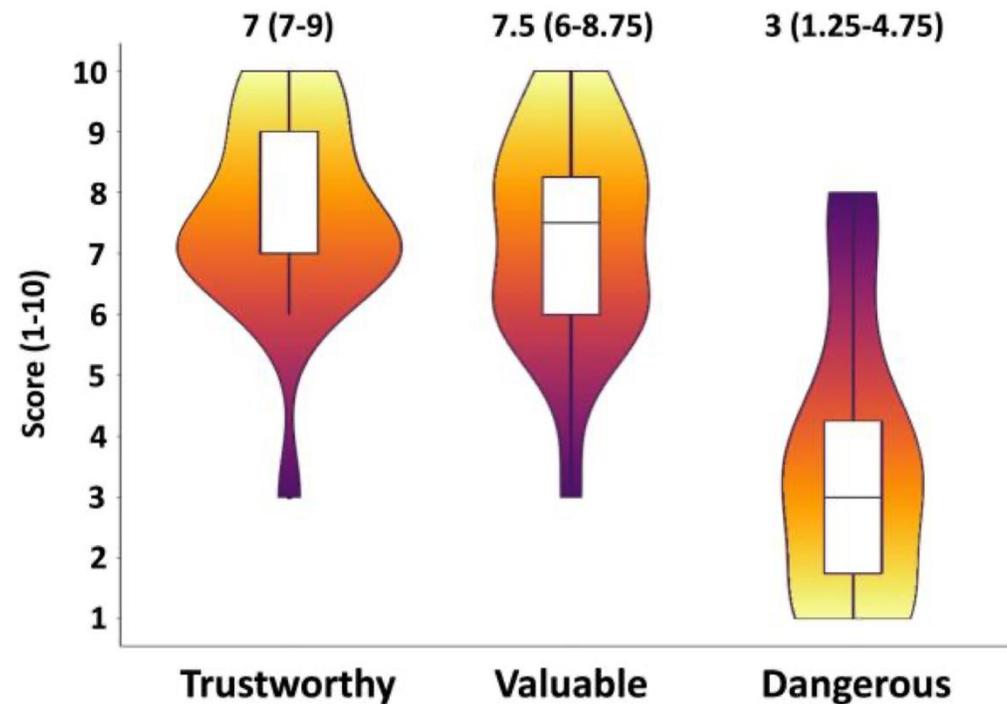
EPA aus KV App



Lokales und privates LLM

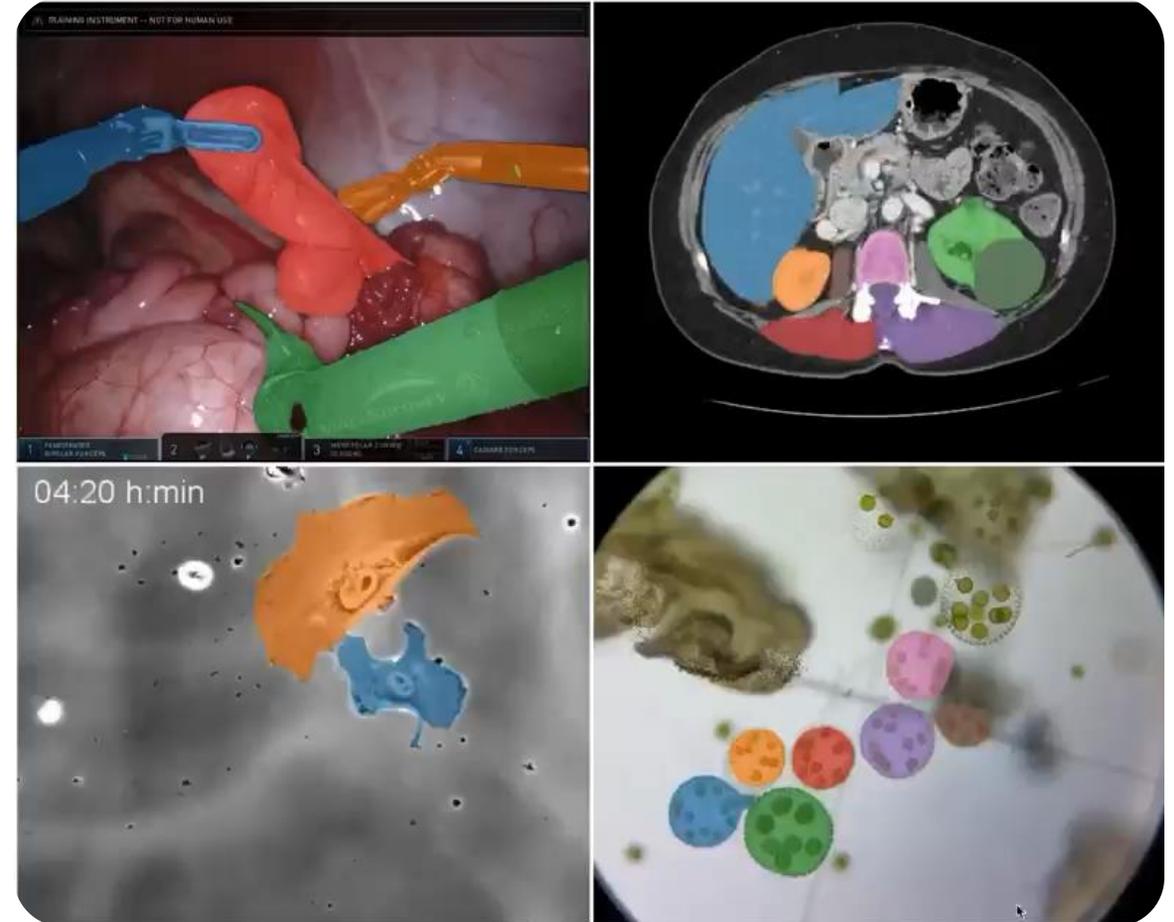


# Aufklärung - Zukunftsausblick



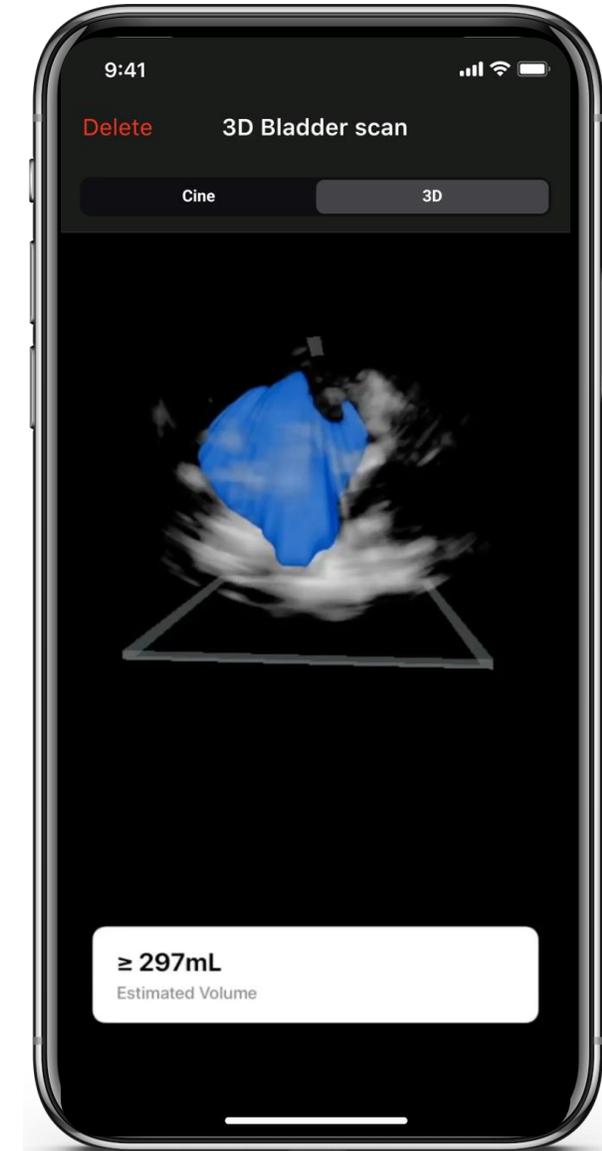
# 2D Segmentierung

- **SAM-Modell:** KI-Modell für flexible und schnelle Bildsegmentierung.
- **Datenbasis:** Trainiert auf 1 Milliarde Masken und 11 Millionen lizenzierten Bildern.
- **Medizinischer Nutzen:** Präzise Segmentierung medizinischer Bilder, vielseitig einsetzbar.
- **Zugang:** Offen verfügbar zur Förderung der medizinischen Bildverarbeitung.



# 3D Sonographie

- **Butterfly iQ:** Tragbares Ultraschallgerät
- **Funktion:** Ein-Chip-Ultraschallwandler, mit verschiedenen Bildgebungsmodi.
- **Vorteil:** Kompakt, kostengünstig, und einfach zu bedienen.
- **Anwendung:** Ermöglicht Diagnosen und Behandlung am Point-of-Care.



Viewer

Select Category

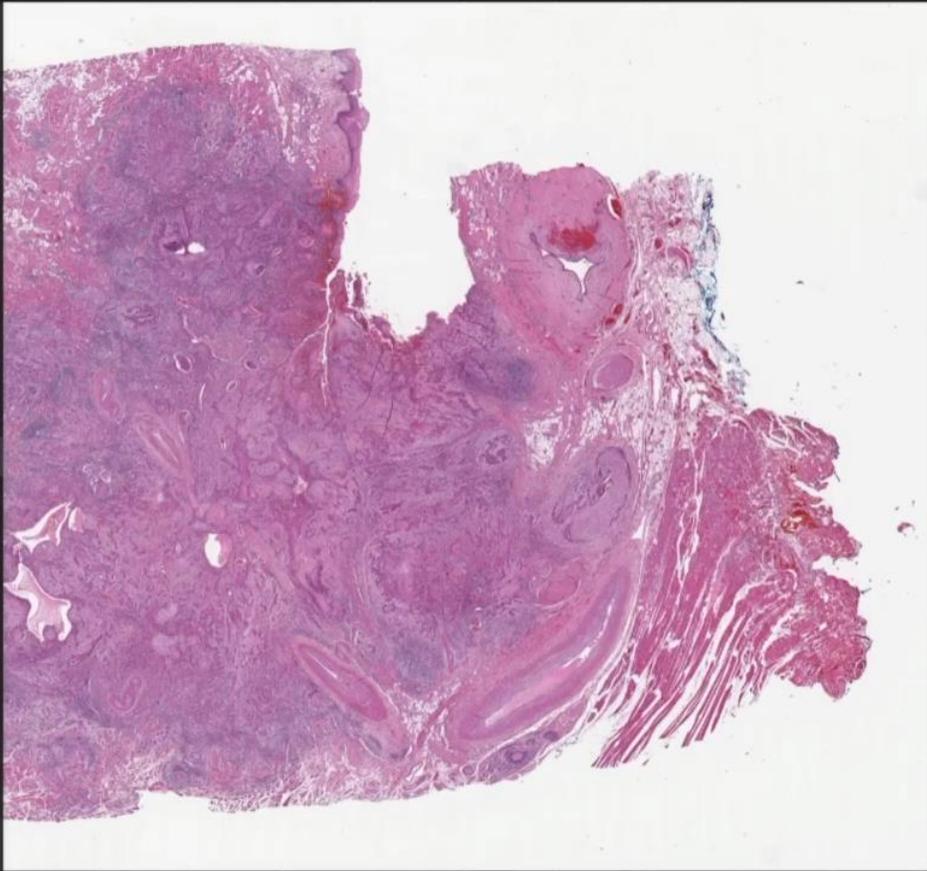
All

Select Image

TCGA-4P-AA8J-01Z-00...

Annotate

Parse Slide



Theme

Settings

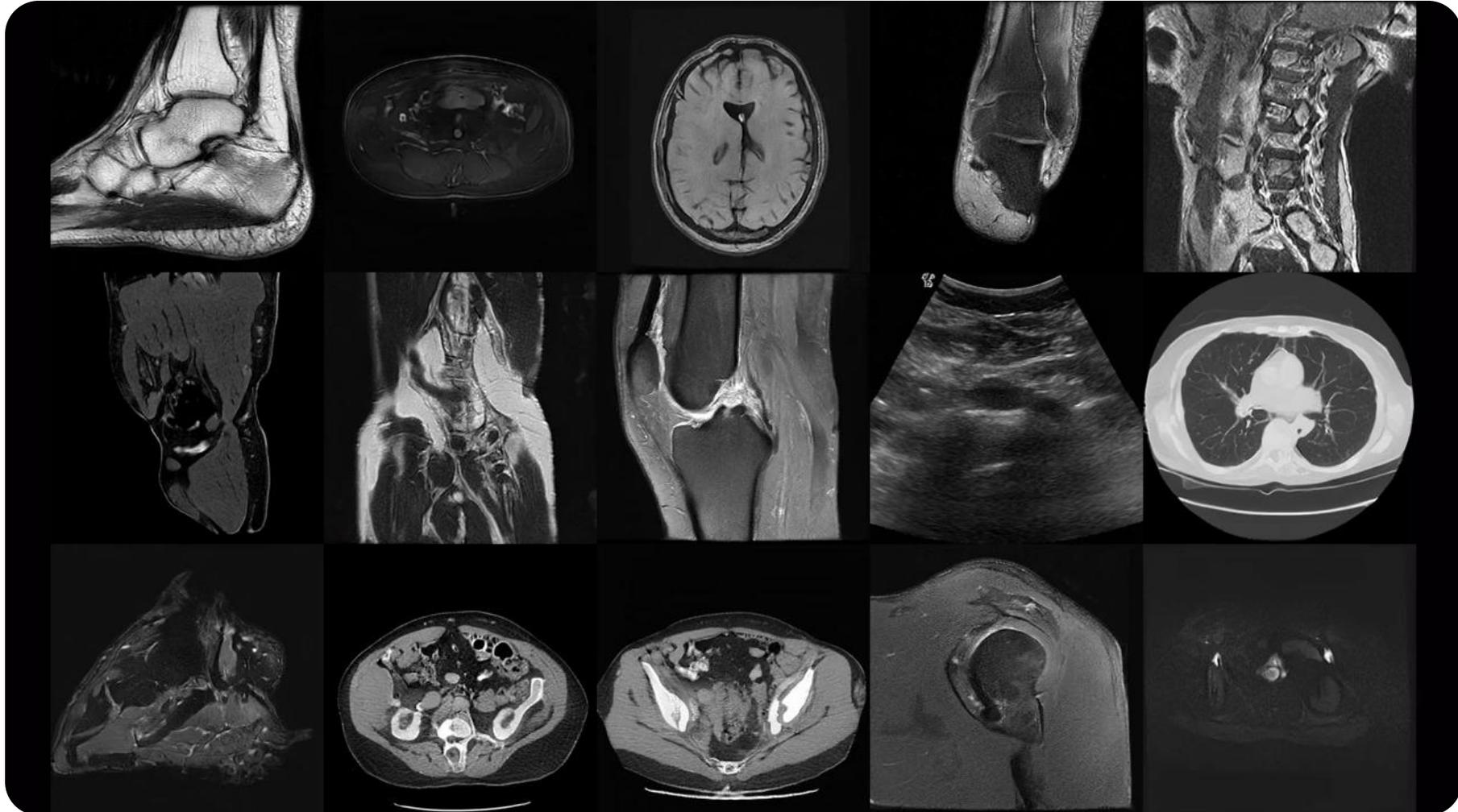
Logout

pathchat.user

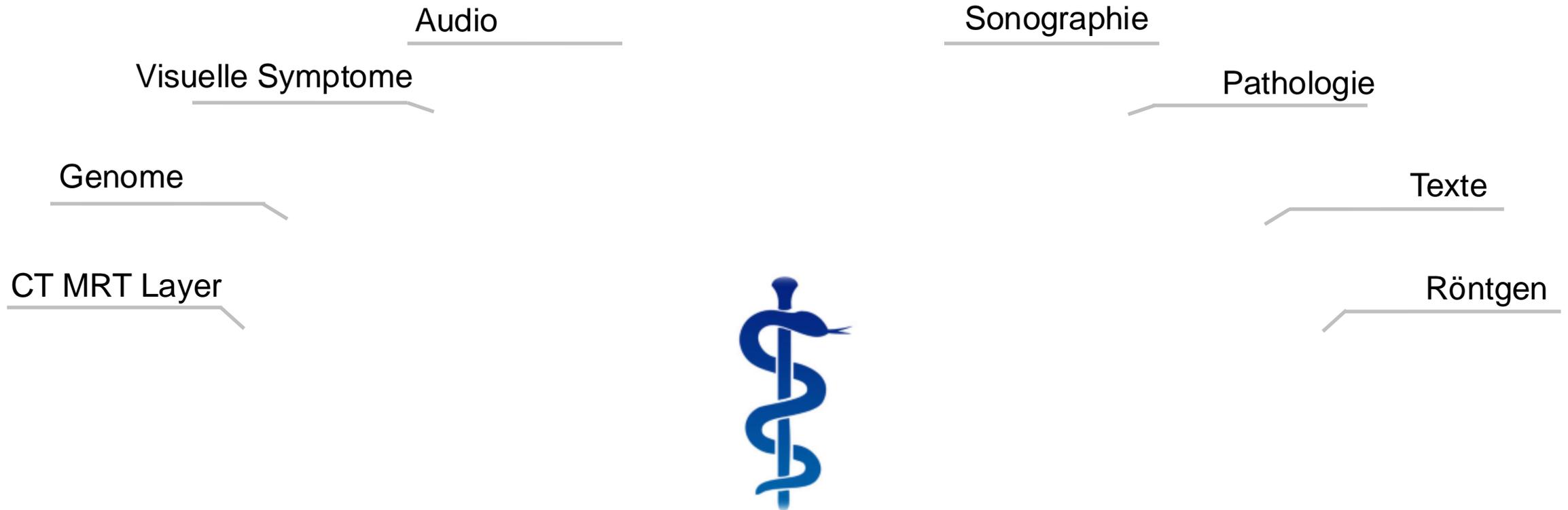
# Welcome to PathChat!

Type your message here...

# Röntgenbild Generation

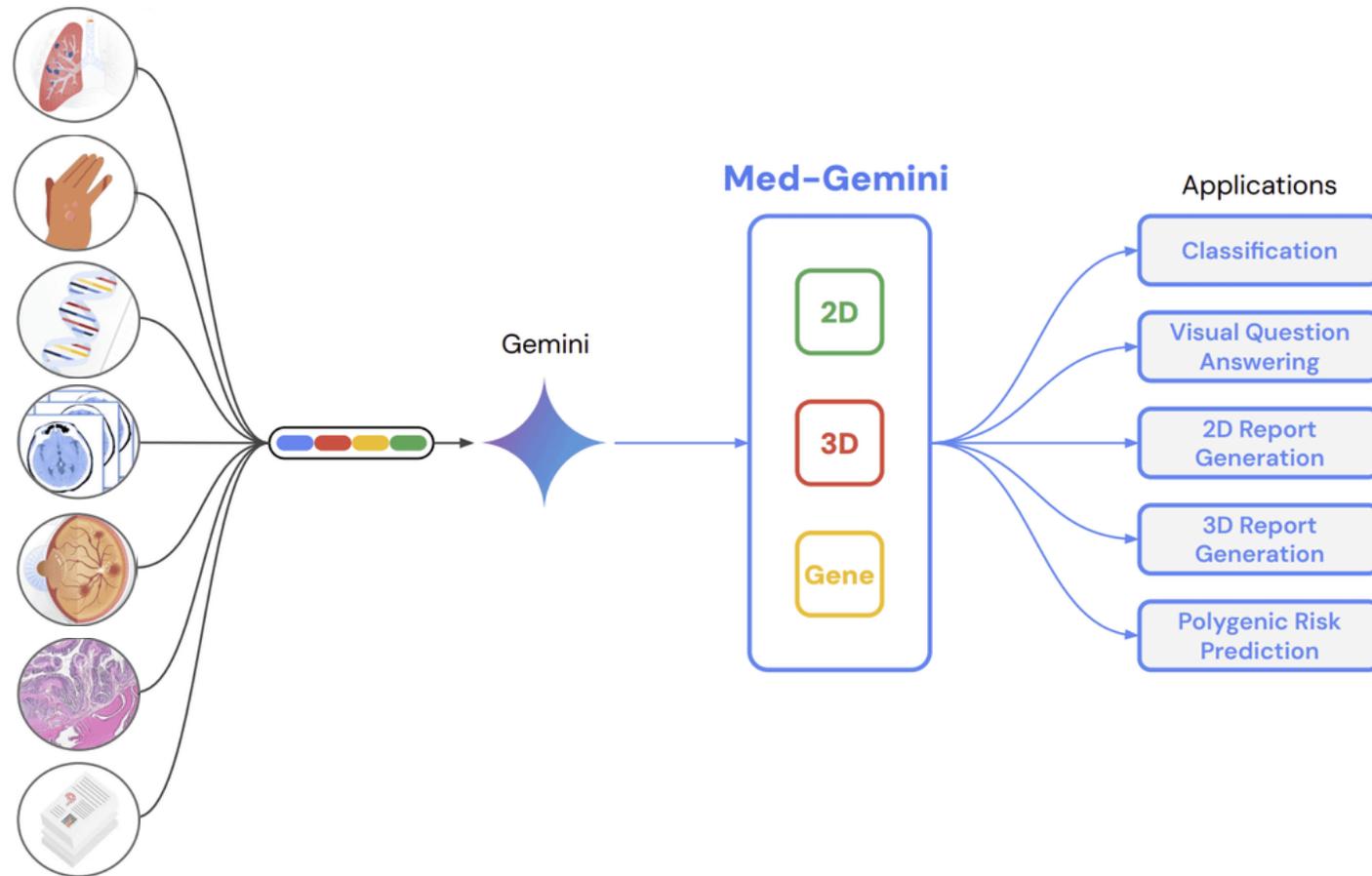


# Multimodalität



Medizin ist nicht nur Text!

# Med - Gemini 1.5



# Med - Gemini 1.5

## Prompt



...



You are a helpful medical video assistant.

**Task:** You are given a video, and a corresponding subtitle with start time and duration, followed by a question. Your task is to extract the precise video timestamps that answer the given question below.

**Instructions:** Provide one single timestamp that spans the entire length of the answer while considering the entire video. It is better to be exhaustive and providing the longest time span for the answer.

**Question:** How to relieve calf strain with foam roller massage?

## Response

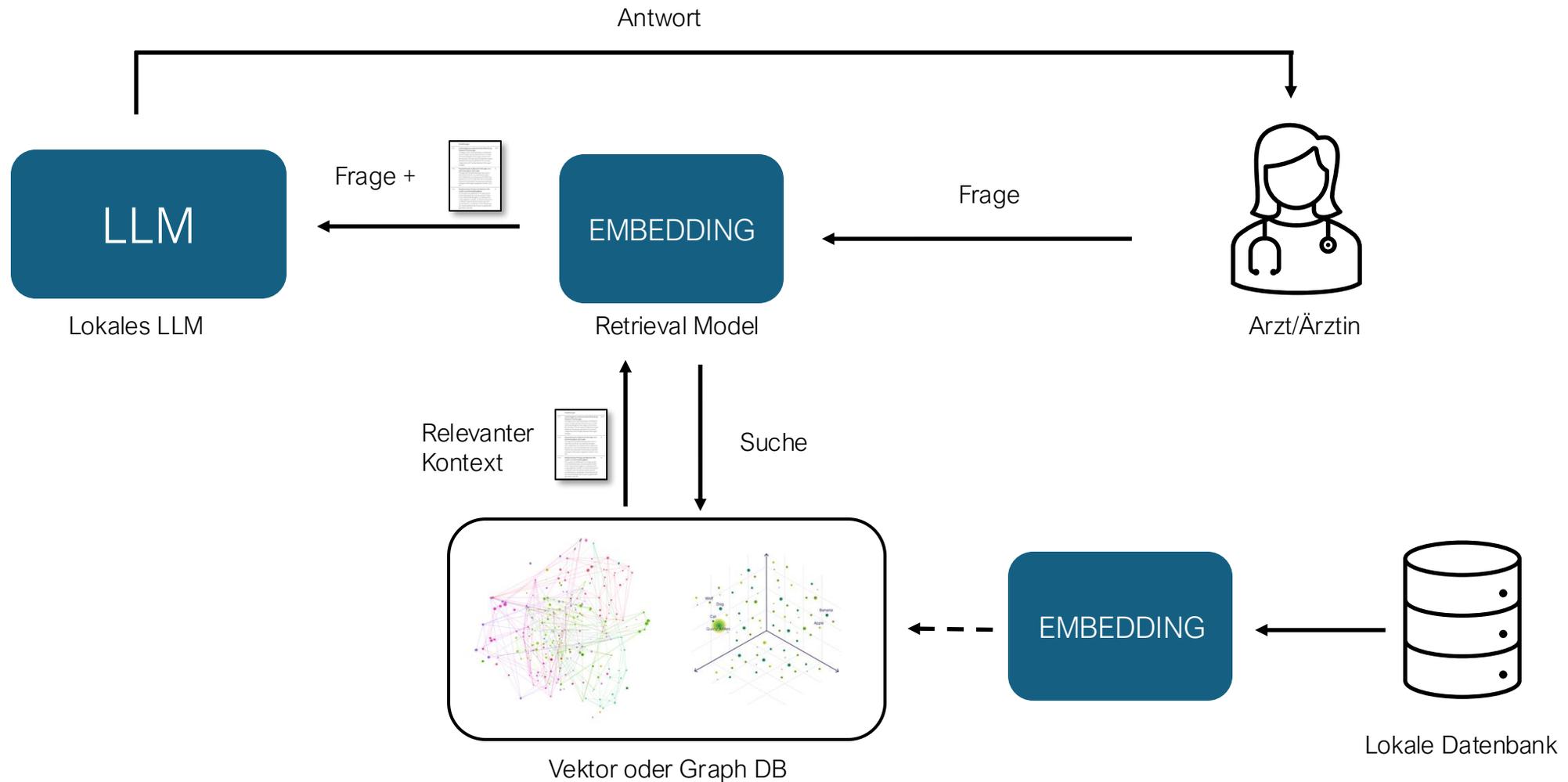
Start = 02:22

End = 02:58

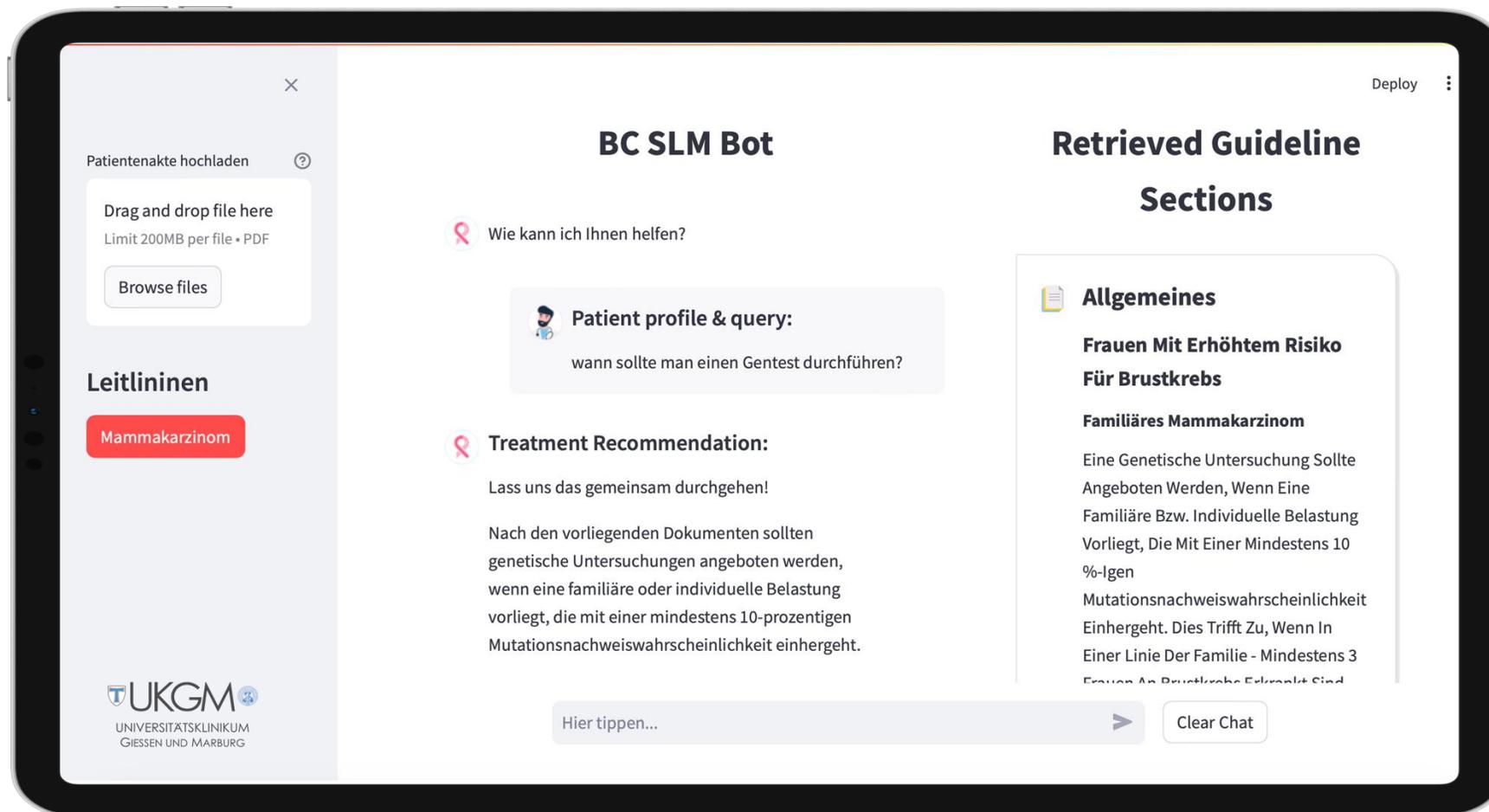
Additional details: Bob demonstrates how to foam roll your calf on a foam roller to relieve calf strain.

# Halluzinationen und keine Quellen

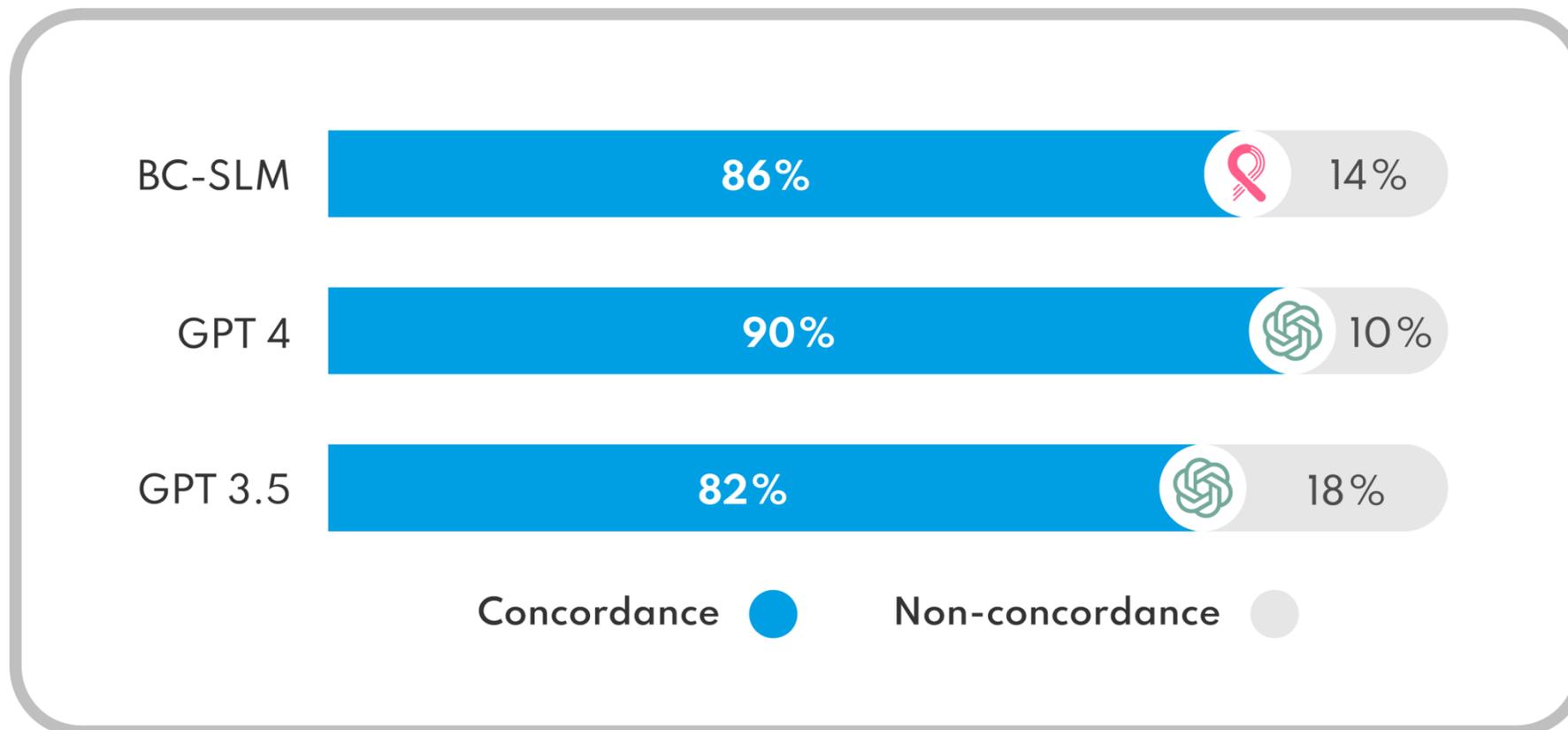
# RAG – Retrieval Augmented Generation



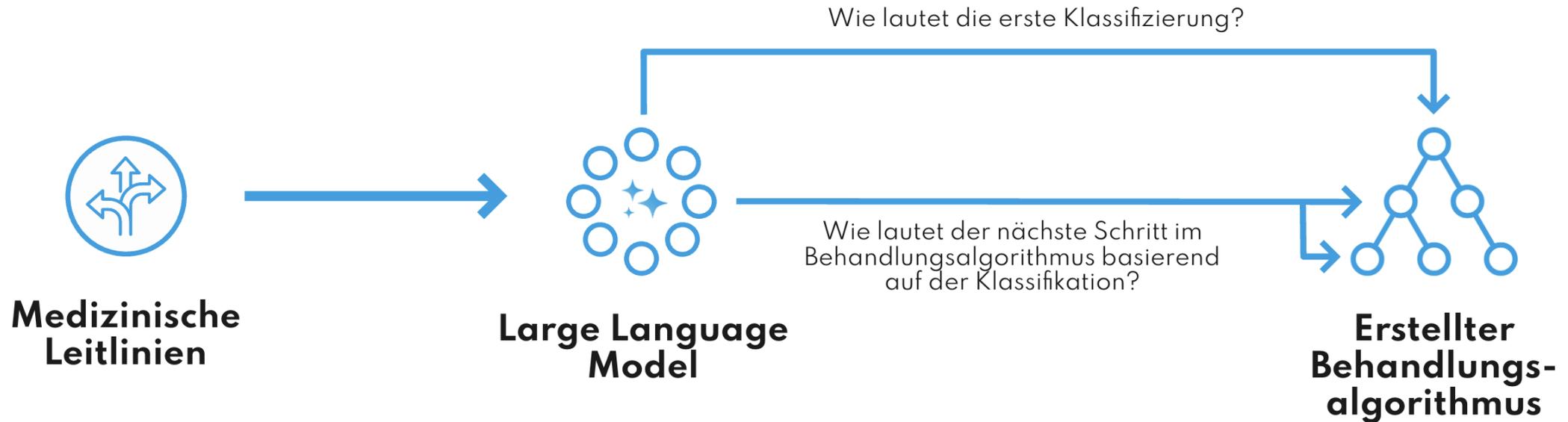
# RAG – Mammakarzinom App



# RAG – Mammakarzinom App



# RAG – Zukunft



# Umwelteinflüsse

- Ein Flug von London nach New York: ca. 500 kg/Person
- Training ChatGPT: ca. 400 Tonnen (GPT4 ist nicht bekannt)
- Die Forschung konzentriert sich nur auf Accuracy-verbesserungen und nicht auf Ressourceneffizienz
- LLMs sind überwiegend auf Englisch verfügbar
- Die globale Erderwärmung wirkt sich überwiegend auf nicht-englischsprachige Länder aus
- Eine ChatGPT Anfrage produziert so viel CO<sub>2</sub> wie 100 Google Suchen



*Medizin kann und wird eine Leitbildfunktion haben für  
die verantwortungsvolle Integration von KI in die  
Gesellschaft.*  
- Prof. Dr. Martin C. Hirsch

Vielen Dank für die Einladung



UNIVERSITÄTSKLINIKUM  
GIESSEN UND MARBURG

Contact details:

[fabian.lechner@uni-marburg.de](mailto:fabian.lechner@uni-marburg.de)

