

„Über die Grenzen zwischen menschlicher und künstlicher Intelligenz“

Dr. André Sander, ID BERLIN
Health IT Talk, 08.09.2025

Die Grenzen der KI



„Die meisten [KI-Roboter] wird es hinaustreiben, dahin, wo die meisten Ressourcen sind, also raus ins Weltall, wo man zum Beispiel im Asteroidengürtel Milliarden von selbstreplizierenden Roboterfabriken, gigantische Teleskope und alles Mögliche bauen kann, um besser zu verstehen, wie das Universum funktioniert.“

Jürgen Schmidhuber, 2016



https://de.wikipedia.org/wiki/J%C3%BCrgen_Schmidhuber

Die Grenzen der NI



„Zwei Dinge sind unendlich, das Universum und die menschliche Dummheit, aber bei dem Universum bin ich mir noch nicht ganz sicher.“

quelle Zwei Dinge sind unendlich, das Universum und die menschliche Dummheit, aber bei dem Universum bin ich mir noch nicht ganz sicher

Das Zitat "Zwei Dinge sind unendlich: das Universum und die menschliche Dummheit, aber bei dem Universum bin ich mir noch nicht ganz sicher" **wird Albert Einstein zugeschrieben, allerdings ist diese Zuschreibung fragwürdig und wahrscheinlich nicht korrekt**, da es sich um ein später veröffentlichtes, sogenanntes "Pseudo-Zitat" handelt. Die Aussage wurde erstmals um 1947 publiziert und **der Gestaltpsychologe Frederick S. Perls wird als eine Quelle genannt**, die den Satz in dieser Form von einem unbekannten Astronomen gehört haben will.

<https://www.google.com/>

quelle Zwei Dinge sind unendlich, das Universum und die menschliche Dummheit, aber bei dem Universum bin ich mir noch nicht ganz sicher

Das Zitat "Zwei Dinge sind unendlich: das Universum und die menschliche Dummheit, aber bei dem Universum bin ich mir noch nicht ganz sicher" **wird Albert Einstein zugeschrieben, allerdings ist diese Zuschreibung fragwürdig und wahrscheinlich nicht korrekt**, da es sich um ein später veröffentlichtes, sogenanntes "Pseudo-Zitat" handelt. Die Aussage wurde erstmals um 1947 publiziert und **der Gestaltpsychologe Frederick S. Perls wird als eine Quelle genannt**, die den Satz in dieser Form von einem unbekannten Astronomen gehört haben will.

Frederick S. Perls Zwei Dinge sind unendlich, das Universum und die menschliche Dummheit, aber bei dem Universum bin ich mir noch nicht ganz sicher

quelle Zwei Dinge sind unendlich, das Universum und die menschliche Dummheit, aber bei dem Universum bin ich mir noch nicht ganz sicher

Das Zitat "Zwei Dinge sind unendlich: das Universum und die menschliche Dummheit, aber bei dem Universum bin ich mir noch nicht ganz sicher" **wird Albert Einstein zugeschrieben, allerdings ist diese Zuschreibung fragwürdig und wahrscheinlich nicht korrekt**, da es sich um ein später veröffentlichtes, sogenanntes "Pseudo-Zitat" handelt. Die Aussage wurde erstmals um 1947 publiziert und **der Gestaltpsychologe Frederick S. Perls wird als eine Quelle genannt**, die den Satz in dieser Form von einem unbekannten Astronomen gehört haben will.

Frederick S. Perls Zwei Dinge sind unendlich, das Universum und die menschliche Dummheit, aber bei dem Universum bin ich mir noch nicht ganz sicher

Das bekannte Zitat "Zwei Dinge sind unendlich: das Universum und die menschliche Dummheit, aber bei dem Universum bin ich mir noch nicht ganz sicher" **stammt nicht von Frederick S. Perls, sondern von Albert Einstein**. Es wird ihm zugeschrieben und ist eines seiner bekanntesten Sprüche, um die Grenzen des Verstandes und die schiere Unermesslichkeit des Universums gegenüberzustellen.



Kann eine KI jemals einen Arzt ersetzen?*

Wie kann eine KI sinnvoll einen Arzt unterstützen, und wo sind die Grenzen?

*und möchten wir das?

André Sander



- Studium zum Dipl.-Ing. Medizininformatik an der TU Ilmenau
- Promotion Medizinwissenschaften an der Charité
- Seit 1998 bei der ID Berlin
- CTO, Mitglied der Geschäftsführung
- BVITG Arbeitsgruppe Interoperabilität und Terminologien
- Kuratorium Friedrich-Wingert-Stiftung



ID Information und Dokumentation im Gesundheitswesen GmbH & Co. KGaA



Gründung

1985



1400+

Einrichtungen in D-A-CH



Themen

- Dokumentation, Abrechnung, Medizincontrolling
- E-Medikation, AMTS
- CDSS
- Terminologien, Ontologien, NLP



120+

Mitarbeitende
(Berlin, Bielefeld, St. Gallen)



Kunden

Krankenhäuser, Universitätskliniken
Ambulante Ärztenetze
Versicherungen, Berufsgenossenschaften
Forschungseinrichtungen



Zertifikate

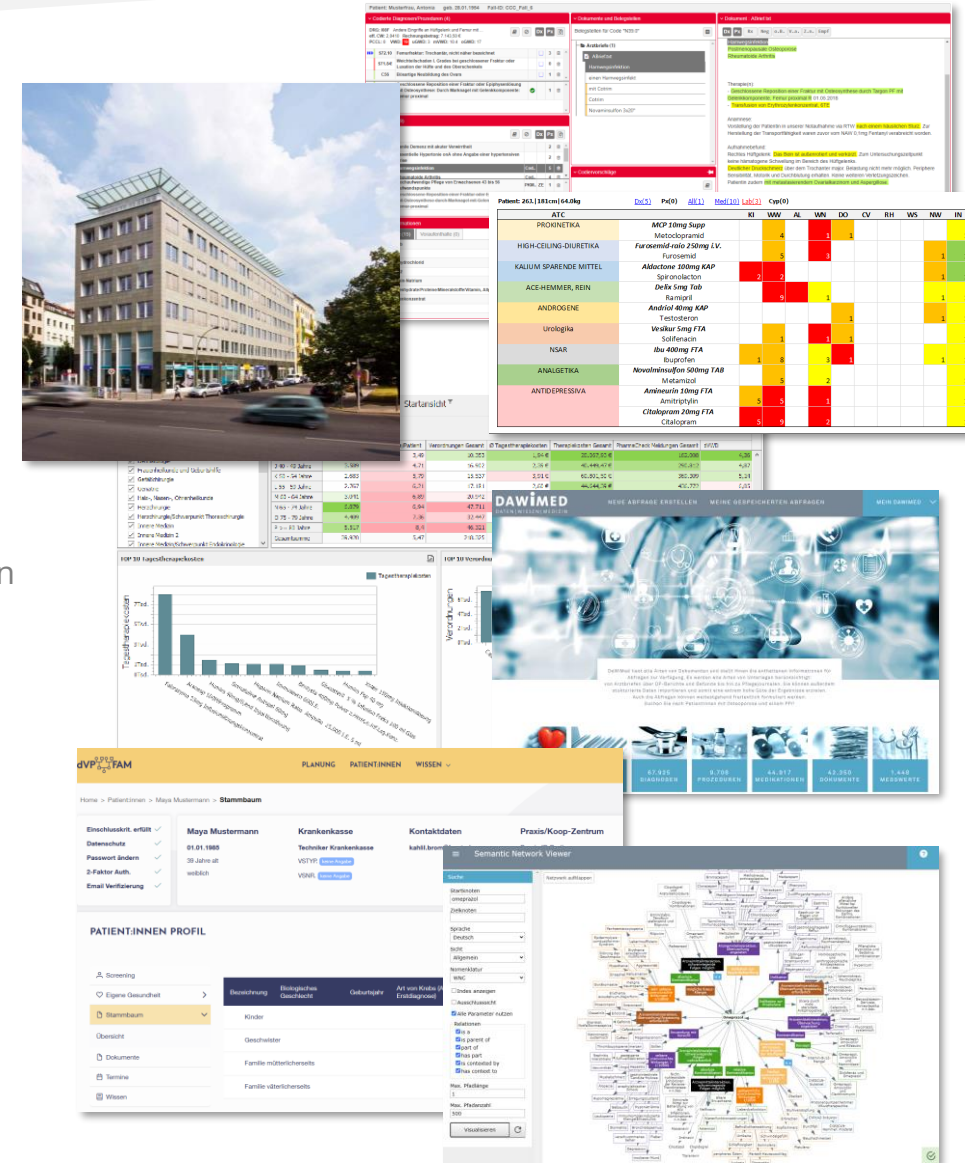
ISO 9001:2015

ISO 13485:2021

ISO 27001

KBV Zertifikat Kodierunterstützung

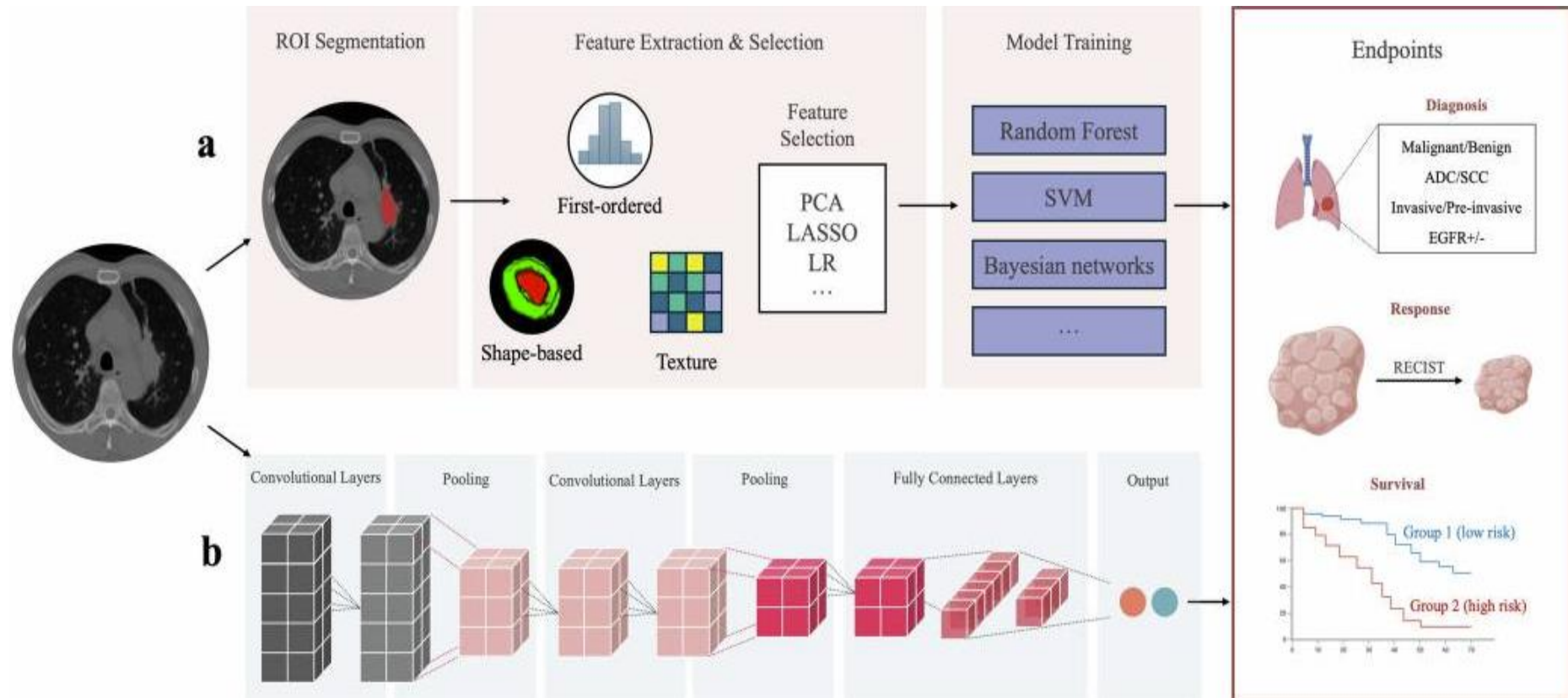
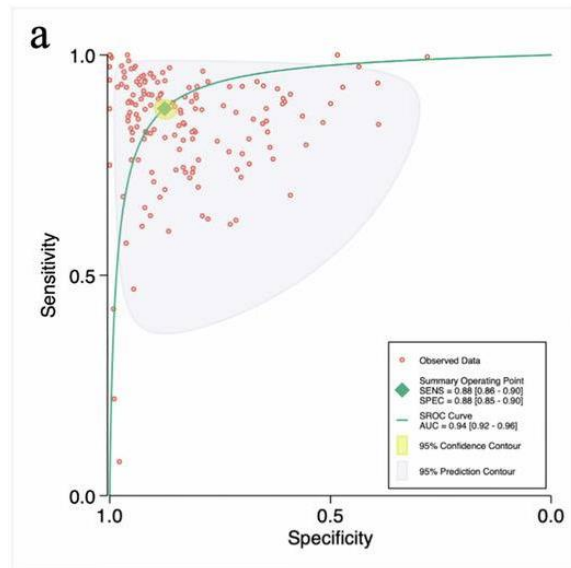
KBV Zertifikat Arzneimittelverordnung



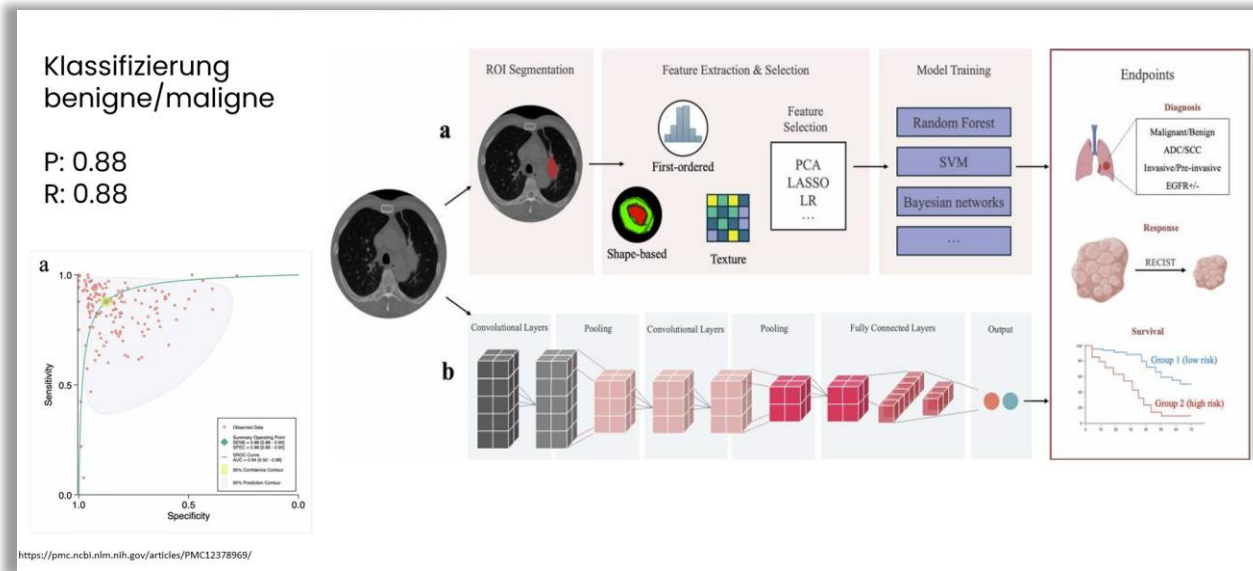
KI gestützte Erkennung von Lungenkarzinomen

Klassifizierung
benigne/maligne

P: 0.88
R: 0.88



Lungenkrebs-Screening für alle?



Fakten

- 125 Mio. Röntgenuntersuchungen in 2024
- Davon 25 Mio. Röntgen Thorax
- Prävalenz Lungenkrebs bei ca. 0,33% (10a)

Annahme

- Die „Dunkelziffer“ wird nicht berücksichtigt
- Jeder Patient erhält nur ein Röntgen
- Jeder Patient bekommt ein KI-Screening

Ergebnis

- ca. 3 Mio. Falsch-Positive!
- Überlastung des Gesundheitssystems

Fazit

- KI-Algorithmen die skaliert werden sollen, müssen eine extrem hohe Spezifität („der Prävalenz angemessen“) haben

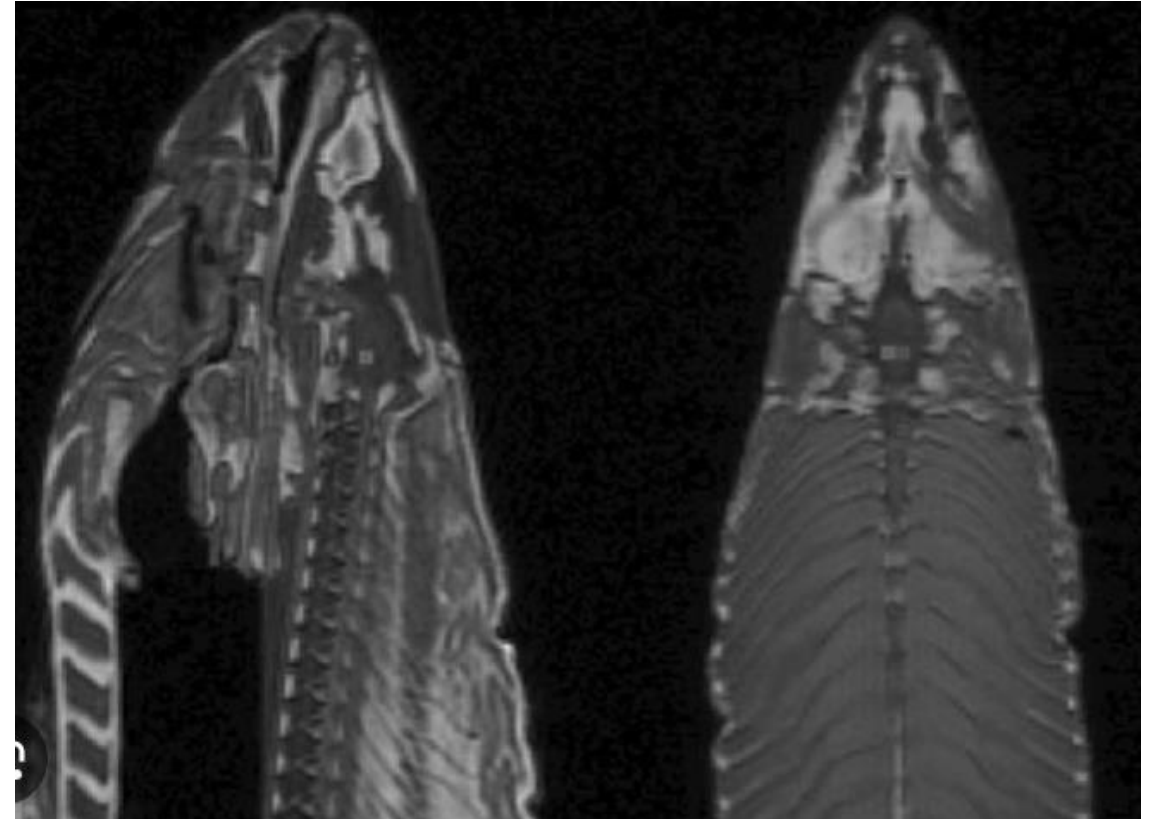
Und die Grenze?

The provided images appear to be sagittal MRI scans of the cervical and thoracic spine regions. These scans are typically used to evaluate the spinal cord, vertebrae, and surrounding soft tissues for any abnormalities or pathologies.

In the images, the spinal cord can be clearly visualized as a long, cylindrical structure running through the vertebral column. The individual vertebral bodies and intervertebral discs are also visible.

Upon careful examination, I can observe the following findings:

1. **Cervical stenosis:** There appears to be narrowing of the spinal canal in the cervical region, which could be due to degenerative changes, disc protrusions, or bony overgrowth. This narrowing can potentially compress the spinal cord and nerve roots, leading to neurological symptoms.
2. **Disc degeneration:** Several intervertebral discs in both the cervical and thoracic regions exhibit changes in signal intensity and height, suggesting varying degrees of disc degeneration and desiccation, which is a common age-related finding.
3. **Cervical lordosis:** The cervical spine appears to have an increased curvature, known as cervical lordosis, which can be a normal variant or potentially related to degenerative changes or posture.
4. **Vertebral body changes:** Some of the vertebral bodies demonstrate signal alterations, which could indicate degenerative changes, such as osteophyte formation or endplate changes.



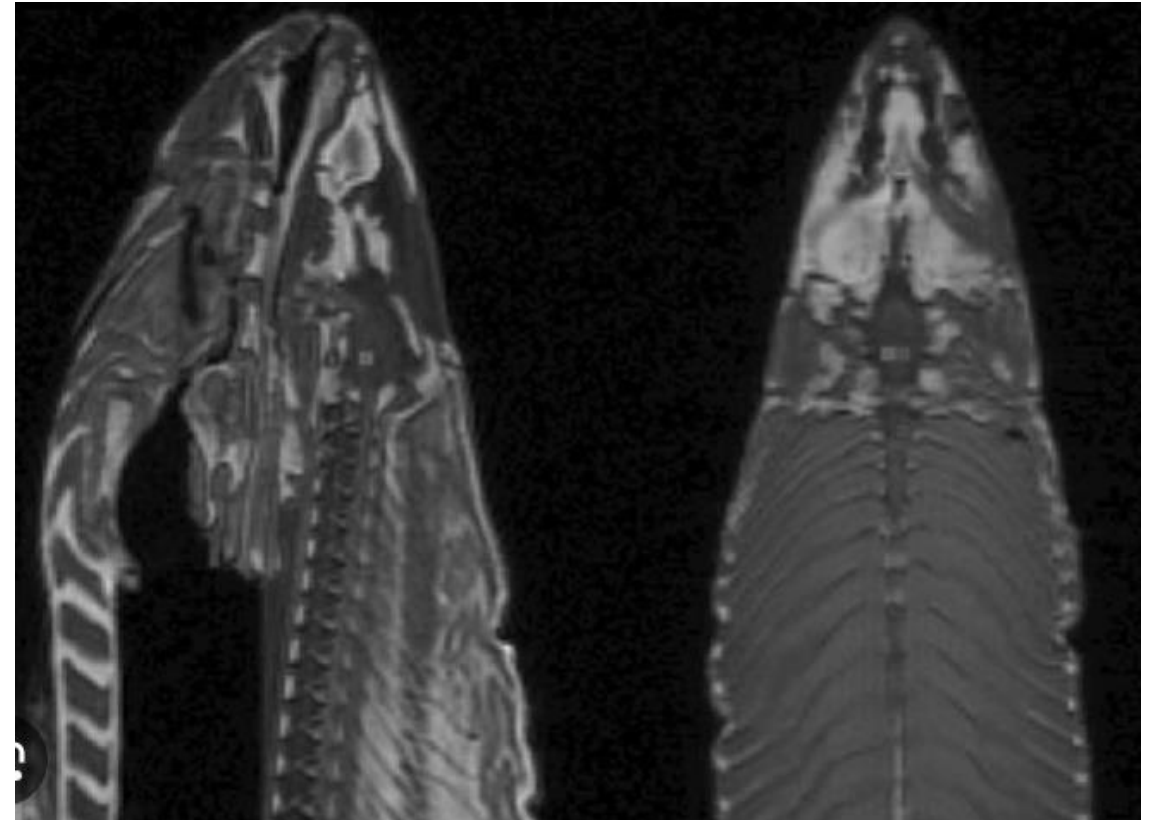
Und die Grenze?

The provided images appear to be sagittal MRI scans of the cervical and thoracic spine regions. These scans are typically used to evaluate the spinal cord, vertebrae, and surrounding soft tissues for any abnormalities or pathologies.

In the images, the spinal cord can be clearly visualized as a long, cylindrical structure running through the vertebral column. The individual vertebral bodies and intervertebral discs are also visible.

Upon careful examination, I can observe the following findings:

1. **Spinal canal narrowing:** There is a noticeable narrowing of the spinal canal in the cervical region, which could be associated with disc protrusions, or bony changes, or compression of the spinal cord or nerve roots.
2. **Disc degeneration:** Several intervertebral disc regions exhibit changes in signal intensity, suggesting varying degrees of disc degeneration and desiccation. This is a common age-related finding.
3. **Cervical lordosis:** The cervical spine appears to have an increased curvature, known as cervical lordosis, which can be a normal variant or potentially related to degenerative changes or posture.
4. **Vertebral body changes:** Some of the vertebral bodies demonstrate signal alterations, which could indicate degenerative changes, such as osteophyte formation or endplate changes.



Expect the unexpected!

Welche Technologien gibt es denn?

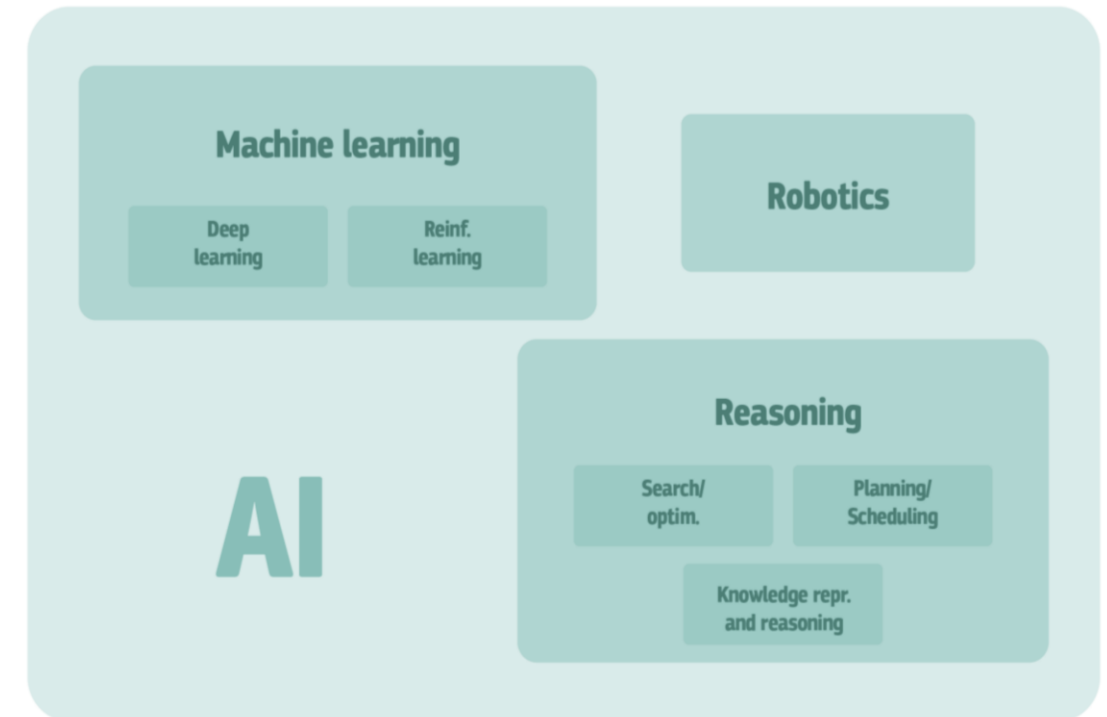
ANHANG I

TECHNIKEN UND KONZEPTE DER KÜNSTLICHEN INTELLIZENZ
gemäß Artikel 3 Absatz 1

a. Konzepte des maschinellen Lernens, mit beaufsichtigtem, unbeaufsichtigtem und bestärkendem Lernen unter Verwendung einer breiten Palette von Methoden, einschließlich des tiefen Lernens (Deep Learning);

b. Logik- und wissensgestützte Konzepte, einschließlich Wissensrepräsentation, induktiver (logischer) Programmierung, Wissensgrundlagen, Inferenz- und Deduktionsmaschinen, (symbolischer) Schlussfolgerungs- und Expertensysteme;

c. Statistische Ansätze, Bayessche Schätz-, Such- und Optimierungsmethoden.



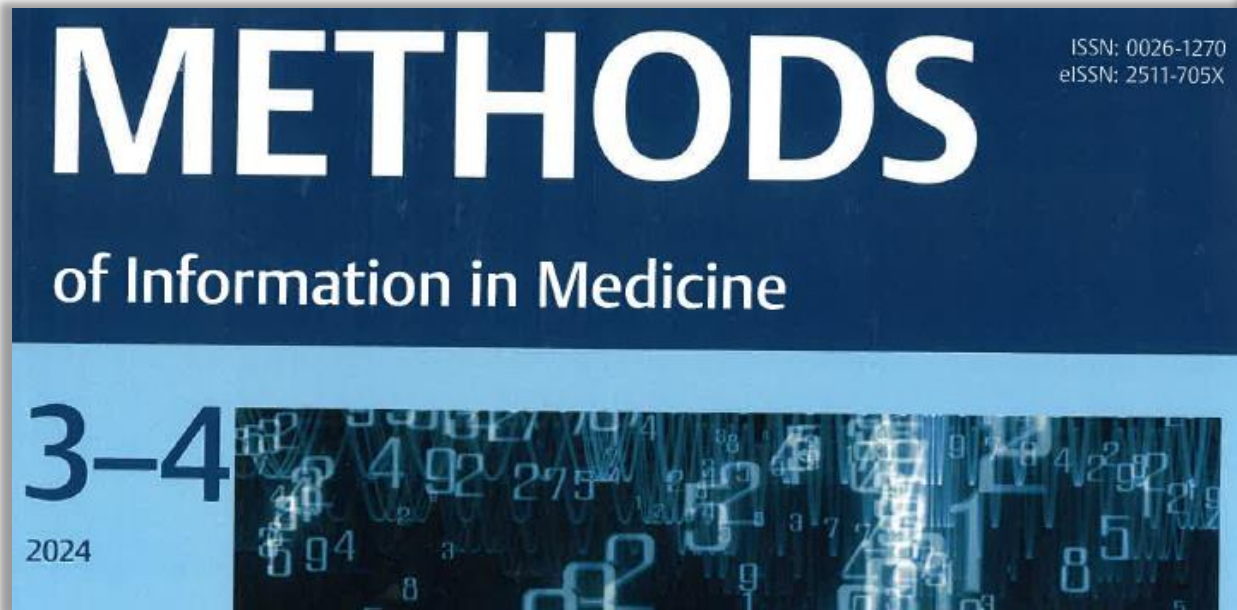
<https://eur-lex.europa.eu/legal-content/DE/TXT/HTML/?uri=CELEX:52021PC0206>

<https://digital-strategy.ec.europa.eu/en/library/definition-artificial-intelligence-main-capabilities-and-scientific-disciplines>

Kombinieren oder Ablösen?



<https://www.thieme-connect.com/products/ejournals/pdf/10.1055/a-2528-4299.pdf>



Leveraging Guideline-Based Clinical Decision Support Systems with Large Language Models: A Case Study with Breast Cancer

Solène Delourme^{1,2} Akram Redjda^{1,3} Jacques Bouaud¹ Brigitte Seroussi^{4,5}

¹Sorbonne Université, Université Sorbonne Paris Nord, INSERM, LIMICS, Paris, France

²EPITA, Paris, France

³Univ Gustave Eiffel, Aix-Marseille Univ, LBA, Marseille, France

⁴Sorbonne Université, AP-HP, Tenon Hospital, Public Health Department, INSERM, Université Sorbonne Paris Nord, Limics, Paris, France

⁵APREC, Paris, France

Address for correspondence Akram Redjda, PhD, 8 avenue de ceinture, Créteil 94000, France
(e-mail: redjalakram300@gmail.com).

Methods Inf Med 2024;63:85–96.

Abstract

Background Multidisciplinary tumor boards (MTBs) have been established in most countries to allow experts collaboratively determine the best treatment decisions for cancer patients. However, MTBs often face challenges such as case overload, which can compromise MTB decision quality. Clinical decision support systems (CDSSs) have been introduced to assist clinicians in this process. Despite their potential, CDSSs are still underutilized in routine practice. The emergence of large language models (LLMs), such as ChatGPT, offers new opportunities to improve the efficiency and usability of traditional CDSSs.

Objectives OncoDoc2 is a guideline-based CDSS developed using a documentary approach and applied to breast cancer management. This study aims to evaluate the potential of LLMs, used as question-answering (QA) systems, to improve the usability of OncoDoc2 across different prompt engineering techniques (PETs).

Methods Data extracted from breast cancer patient summaries (BCPSs), together with questions formulated by OncoDoc2, were used to create prompts for various LLMs, and several PETs were designed and tested. Using a sample of 200 randomized BCPSs, LLMs and PETs were initially compared with regard to their responses to OncoDoc2 questions using classic metrics (accuracy, precision, recall, and F1 score). Best performing LLMs and PETs were further assessed by comparing the therapeutic recommendations generated by OncoDoc2, based on LLM inputs, to those provided by MTB clinicians using OncoDoc2. Finally, the best performing method was validated using a new sample of 30 randomized BCPSs.

Results The combination of Mistral and OpenChat models under the enhanced Zero-Shot PET showed the best performance as a question-answering system. This approach gets a precision of 60.16%, a recall of 54.18%, an F1 score of 56.59%, and an accuracy of

Keywords

- clinical decision support systems
- OncoDoc2
- breast cancer
- large language models
- question-answering systems

...schwierig!

Table 1 Results of the different PETs with Mistral, Mixtral 8 × 7B, and OpenChat models on 3,142 prompts

PET	Model	Precision (%)	Recall (%)	F1 score (%)	Accuracy (%)
Zero-Shot	Mistral	54.27	49.34	47.44	61.78
	OpenChat	72.22	70.73	71.30	69.95
Enhanced Zero-Shot	Mistral	57.26	47.67	49.62	66.10
	Mixtral8 × 7B	61.69	58.06	59.45	77.08
	OpenChat	73.54	71.31	72.11	72.47
Zero-Shot CoT	Mistral	54.80	44.27	45.70	59.83
	OpenChat	55.74	49.13	51.60	69.96

Abbreviation: PET, prompt engineering technique.

...schwierig!

Table 1 Results of the different PETs with Mistral, Mixtral 8 × 7B, and OpenChat models on 3,142 prompts

PET	Model	Precision (%)	Recall (%)	F1 score (%)
Zero-Shot	Mistral	54.27	49.34	47.44
	OpenChat	72.22	70.73	71.30
Enhanced Zero-Shot	Mistral	57.26	47.67	49.62
	Mixtral8 × 7B	61.69	58.06	59.45
	OpenChat	73.54	71.31	72.11
Zero-Shot CoT	Mistral	54.80	44.27	45.70
	OpenChat	55.74	49.13	51.60

Abbreviation: PET, prompt engineering technique.

- Maximal 72% F1
- Anteil der bis 90% korrekt beantworteten Fragen liegt bei <30%
- Reicht (uns) das?

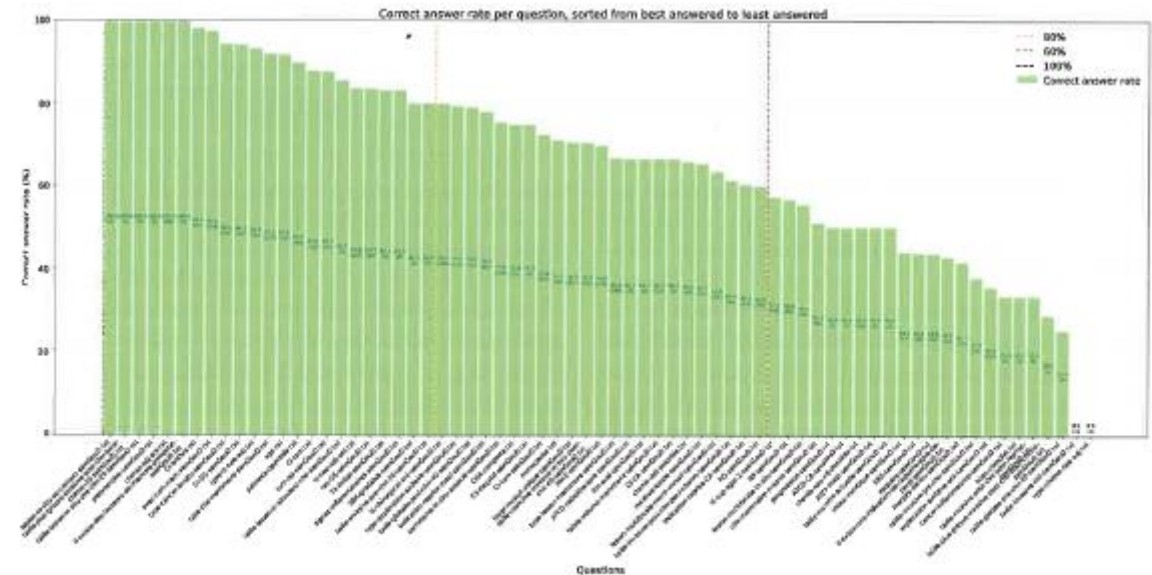
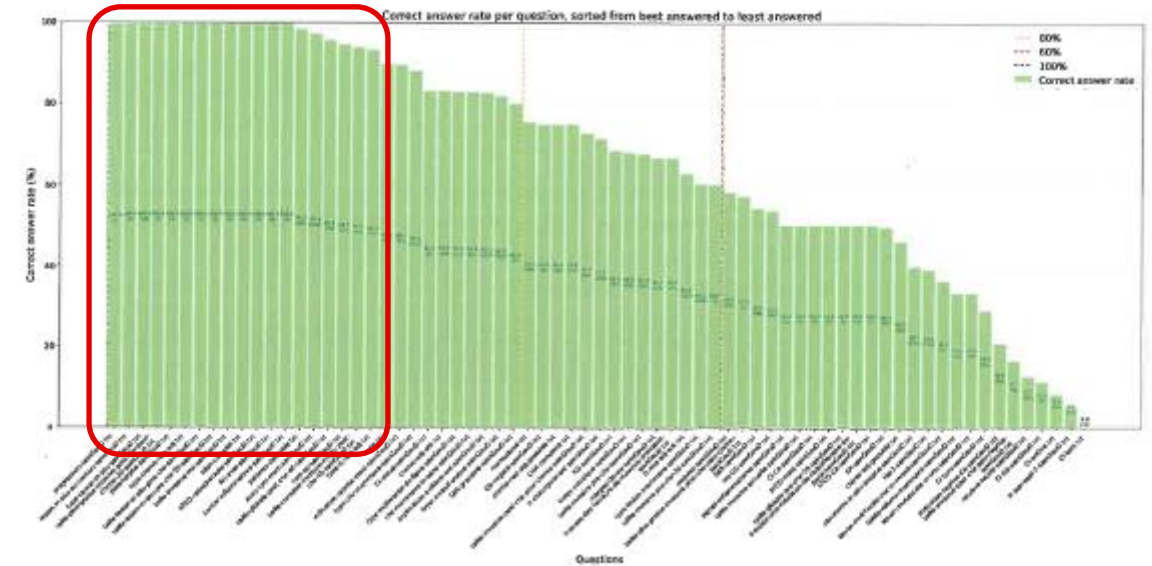


Fig. 10 Distributions of Mistral (top) and OpenChat (bottom) models' accuracy to answer OncoDoc2 questions with the enhanced Zero-Shot prompt engineering technique (PET).

Logisches Denken mit ML?

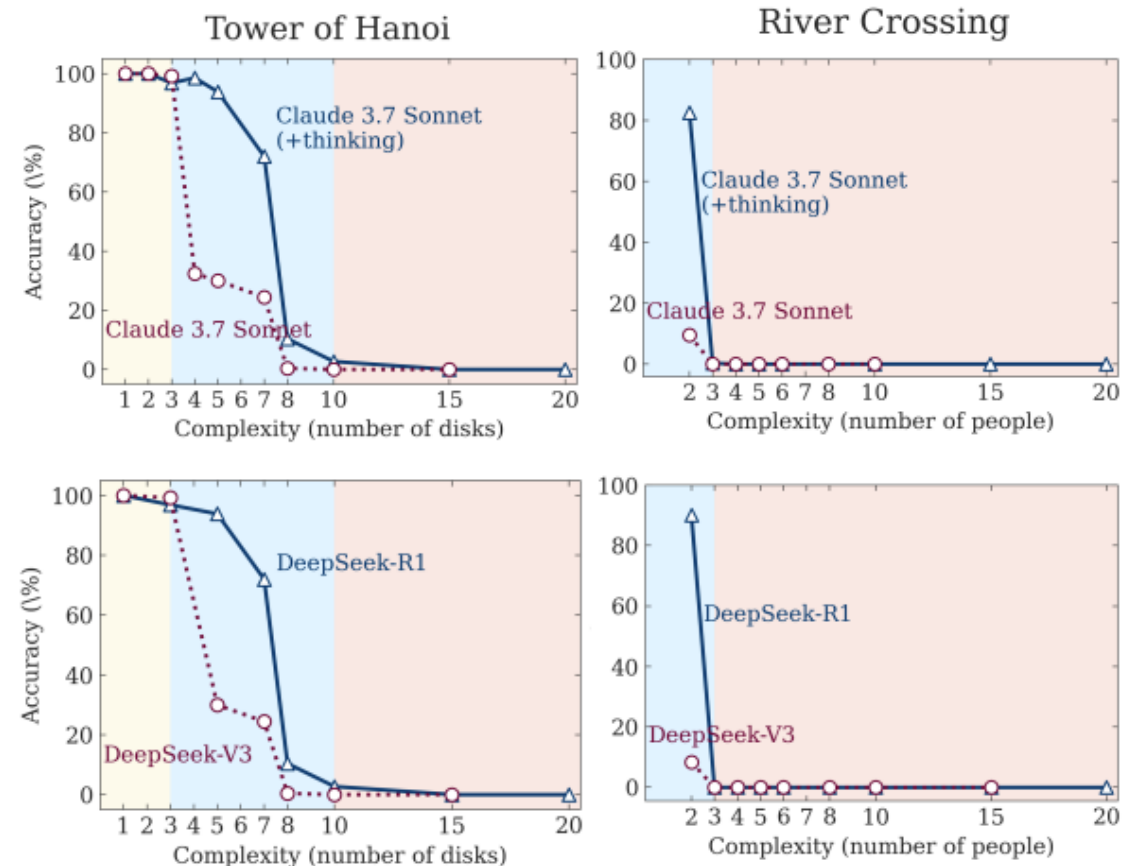
The Illusion of Thinking: Understanding the Strengths and Limitations of Reasoning Models via the Lens of Problem Complexity

Parshin Shojae*† Iman Mirzadeh* Keivan Alizadeh
Maxwell Horton Samy Bengio Mehrdad Farajtabar

Apple

Abstract

Recent generations of frontier language models have introduced Large Reasoning Models (LRMs) that generate detailed thinking processes before providing answers. While these models demonstrate improved performance on reasoning benchmarks, their fundamental capabilities, scaling properties, and limitations remain insufficiently understood. Current evaluations primarily focus on established mathematical and coding benchmarks, emphasizing final answer accuracy. However, this evaluation paradigm often suffers from data contamination and does not provide insights into the reasoning traces' structure and quality. In this work, we systematically investigate these gaps with the help of controllable puzzle environments that allow precise manipulation of compositional complexity while maintaining consistent logical structures. This setup enables the analysis of not only final answers but also the internal reasoning traces, offering insights into how LRMs "think". Through extensive experimentation across diverse puzzles, we show that frontier LRMs face a complete accuracy collapse beyond certain complexities. Moreover, they exhibit a counter-intuitive scaling limit: their reasoning effort increases with problem complexity up to a point, then declines despite having an adequate token budget. By comparing LRMs with their standard LLM counterparts under equivalent inference compute, we identify three performance regimes: (1) low-complexity tasks where standard models surprisingly outperform LRMs, (2) medium-complexity tasks where additional thinking in LRMs demonstrates advantage, and (3) high-complexity tasks where both models experience complete collapse. We found that LRMs have limitations in exact computation: they fail to use explicit algorithms and reason inconsistently across puzzles. We also investigate the reasoning traces in more depth, studying the patterns of explored solutions and analyzing the models' computational behavior, shedding light on their strengths, limitations, and ultimately raising crucial questions about their true reasoning capabilities.

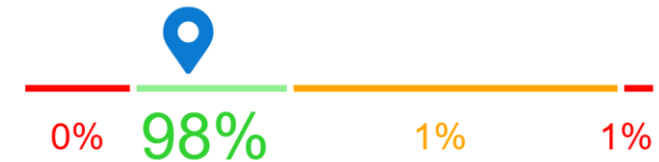


Beispiel: Deep Learning

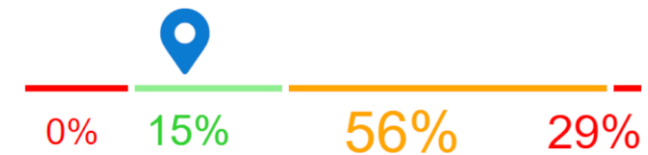


2016 <https://www.handelsblatt.com/technik/it-internet/googles-alpha-go-software-schlaegt-den-go-champion/13072718.html>

Verweildauer 3 Tage



Verweildauer 3 Tage



Versorgungsqualität
Komplikationspräktion

Wirtschaftliche Aspekte
DRG-Kalkulation

F1: ~80%

Beispiel: wissensbasierte KI

Automatisierte Prüfung der AMTS

Alter, Geschlecht, Gewicht
Schwangerschaft, Stillzeit
Diagnosen, Allergien,
Labor- und Vitalwerte
Gene

Kontraindikationen
Wechselwirkungen
Nebenwirkungen
Dosierung
Anwendungsdauer
Gabezeiten

Patientensicherheit wird
drastisch erhöht!

The screenshot displays the ID MEDICS software interface. The top section shows a medication overview for a patient named Adipositas, Alfred (born 12.09.1946, 70 years old, male). The interface includes a timeline of medication administration from 07:00 to 22:00. Below this, a list of medications is shown, including Heparin, Delix, HCT, Metformin, Nexium, Pravadatin, Cloxone, Insulin, Floxal, Duroneg, Polibexal, and Ibu. A detailed drug interaction table is also visible, listing various drugs and their interactions.

ATC	Medication	KI	WW	AL	WN	DO	CV	RH	WS	NW	IN
PROKINETIKA	MCP 10mg Supp										
	Metoclopramid		4		1	1					1
HIGH-CEILING-DIURETIKA	Furosemid-ratio 250mg i.v.		5		3					1	1
	Furosemid										
KALIUM SPARENDE MITTEL	Aldactone 100mg KAP	2	2							1	1
	Spiroolacton										
ACE-HEMMER, REIN	Delix 5mg Tab		9		1					1	1
	Ramipril										
ANDROGENE	Andriol 40mg KAP						1			1	1
	Testosteron										
Urologika	Vesikur 5mg FTA		1		1	1				1	1
	Sildenafil										
NSAR	Ibu 400mg FTA		1	8	3	1				1	1
	Ibuprofen										
ANALGETIKA	Novalmin sulfon 500mg TAB			5		2					1
	Metamizol										
ANTIDEPRESSIVA	Amineurin 10mg FTA		5	5		1					1
	Amitriptylin										
	Citalopram 20mg FTA		5	9		2					1
	Citalopram										

F1: >>99%

Und was ist mit LLMs?

2018

Brauchen wir!

- Text erzeugen (Arztbriefschreibung)
- Text verändern (Laiensprache)
- Informationen extrahieren
- Informationen sammeln (ChatBots)
- NLU (Spracheingabe)
- RAG (Dokumente erschließen)
- ...
- *Wissen vermitteln (aber nicht erzeugen)

Improving Language Understanding by Generative Pre-Training

Alec Radford OpenAI alec@openai.com	Karthik Narasimhan OpenAI karthikn@openai.com	Tim Salimans OpenAI tim@openai.com	Ilya Sutskever OpenAI ilyasu@openai.com
---	---	--	---

Abstract

Natural language understanding comprises a wide range of diverse tasks such as textual entailment, question answering, semantic similarity assessment, and document classification. Although large unlabeled text corpora are abundant, labeled data for learning these specific tasks is scarce, making it challenging for discriminatively trained models to perform adequately. We demonstrate that large gains on these tasks can be realized by *generative pre-training* of a language model on a diverse corpus of unlabeled text, followed by *discriminative fine-tuning* on each specific task. In contrast to previous approaches, we make use of task-aware input transformations during fine-tuning to achieve effective transfer while requiring minimal changes to the model architecture. We demonstrate the effectiveness of our approach on a wide range of benchmarks for natural language understanding. Our general task-agnostic model outperforms discriminatively trained models that use architectures specifically crafted for each task, significantly improving upon the state of the art in 9 out of the 12 tasks studied. For instance, we achieve absolute improvements of 8.9% on commonsense reasoning (Stories Cloze Test), 5.7% on question answering (RACE), and 1.5% on textual entailment (MultiNLI).

Eines der Trainingsziele:
„Next Sentence Prediction“

RAG im Patientenkontext

- Begründung auffälliger Laborwerte
- Abschätzung von Komplikationen
- Substitution von Medikamenten
- Anschlussbehandlung (Pflegebedürftigkeit)
- Kontrolle

Cave: Software muss Medizinprodukt sein und die Haftung übernimmt der Behandler!

* Der Screenshot zeigt eine aktuelle Studie

The screenshot displays a RAG application interface. On the left, a chat window titled 'Hallo Julian' shows a user prompt: 'Welche Ernährung kann dem Patienten empfohlen werden? Bitte antworte direkt dem Patienten mit wenigen Sätzen!'. The system response provides dietary advice for a patient with chronic kidney disease, recommending a low-protein diet (approx. 0.8g protein/kg body weight) and limiting sodium intake to 1500mg per day. On the right, a PDF document titled 'Arztbrief.pdf' is displayed. The document is a medical discharge summary for Herr Max Mustermann, dated May 22, 2023. It details his medical history, including chronic kidney disease (Stage 3), hypertension, and polyarthritis. The summary lists primary and secondary diagnoses, procedures performed (Native Computertomography of the thorax), and discharge instructions, including a recommendation for nursing care (Pflegegrad 1).

Technologien kombinieren – Werkzeuge sinnvoll einsetzen!



Wir sind für Sie da!

ID – Information und Dokumentation im Gesundheitswesen GmbH & Co. KGaA

Platz vor dem Neuen Tor 2
10115 Berlin

Tel.: +49 30 24626-0
Fax: +49 30 24626-111

Techn. Support: +49 30 24626-112
Inhaltl. Support: +49 30 24626-108

info@id-berlin.de
<http://www.id-berlin.de>

