Foundation Models – Chancen und Herausforderungen

Sven Giesselbach - Fraunhofer IAIS

Der Impact von Foundation Models

Schlagzeilen

Reuters

Microsoft co-founder Bill Gates: ChatGPT 'will change our world'

Harvard Business School

Generative AI Will Change Your Business. Here's How to Adapt.

by David C. Edelman and Mark Abraham

Financial Times

Generative AI: how will the new era of machine learning affect you?

Biztech News

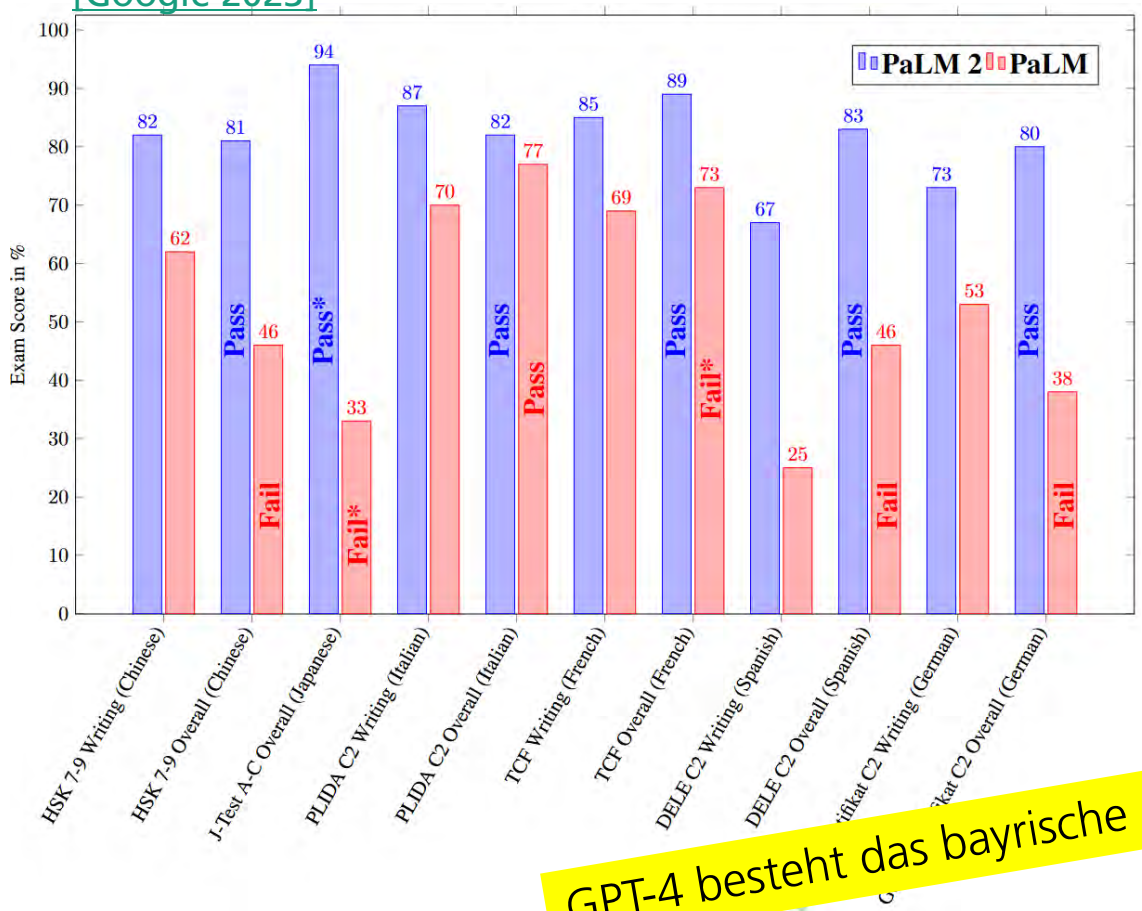
ChatGPT: AI will shape the world on a scale not seen since the iPhone revolution, says OpenAI boss

OpenAI Google ALEPH ALPHA Microsoft

Der Impact von Foundation Models

Werden Sie die (Arbeits-)Welt zu verändern? – I

- PaLM 2 Ergebnisse auf "menschlichen" Prüfungen [Google 2023]



- GPT-4 Ergebnisse auf "menschlichen" Prüfungen [OpenAI 2023]

Exam	GPT-4	GPT-4 (no vision)	GPT-3.5
Uniform Bar Exam (MBE+MEE+MPT)	298 / 400 (~90th)	298 / 400 (~90th)	213 / 400 (~10th)
LSAT	163 (~88th)	161 (~83rd)	149 (~40th)
SAT Evidence-Based Reading & Writing	710 / 800 (~93rd)	710 / 800 (~93rd)	670 / 800 (~87th)
SAT Math	700 / 800 (~89th)	690 / 800 (~89th)	590 / 800 (~70th)
Graduate Record Examination (GRE) Quantitative	163 / 170 (~80th)	157 / 170 (~62nd)	147 / 170 (~25th)
Graduate Record Examination (GRE) Verbal	169 / 170 (~99th)	165 / 170 (~96th)	154 / 170 (~63rd)
Graduate Record Examination (GRE) Writing	4 / 6 (~54th)	4 / 6 (~54th)	4 / 6 (~54th)
USABO Semifinal Exam 2020	87 / 150 (99th - 100th)	87 / 150 (99th - 100th)	43 / 150 (31st - 33rd)
USNCO Local Section Exam 2022	36 / 60	38 / 60	24 / 60
Medical Knowledge Self-Assessment Program	75 %	75 %	53 %
Codeforces Rating	392 (below 5th)	392 (below 5th)	260 (below 5th)
AP Art History	5 (86th - 100th)	5 (86th - 100th)	5 (86th - 100th)
AP Biology	5 (85th - 100th)	5 (85th - 100th)	4 (62nd - 85th)
AP Calculus BC	4 (43rd - 59th)	4 (43rd - 59th)	1 (0th - 7th)
AP Chemistry	4 (71st - 88th)	4 (71st - 88th)	2 (22nd - 46th)
AP English Language and Composition	2 (14th - 44th)	2 (14th - 44th)	2 (14th - 44th)
AP English Literature and Composition	2 (8th - 22nd)	2 (8th - 22nd)	2 (8th - 22nd)
AP Environmental Science	5 (91st - 100th)	5 (91st - 100th)	5 (91st - 100th)
AP Macroeconomics	5 (84th - 100th)	5 (84th - 100th)	2 (33rd - 48th)
AP Microeconomics	5 (82nd - 100th)	4 (60th - 82nd)	4 (60th - 82nd)
[illegible] 2	4 (66th - 84th)	4 (66th - 84th)	3 (30th - 66th)
[illegible] gy	5 (83rd - 100th)	5 (83rd - 100th)	5 (83rd - 100th)
AP Statistics	5 (85th - 100th)	5 (85th - 100th)	3 (40th - 63rd)
AP US Government	5 (88th - 100th)	5 (88th - 100th)	4 (77th - 88th)
AP US History	5 (89th - 100th)	4 (74th - 89th)	4 (74th - 89th)

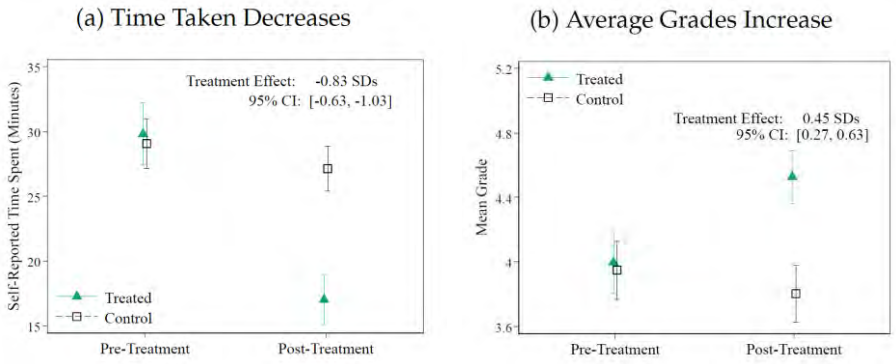
GPT-4 besteht das bayrische Abitur mit der Note 2

Der Impact von Foundation Models

Werden Sie die (Arbeits-)Welt zu verändern? – II

- Laut McKinsey können bis zu **45% der Arbeitstätigkeiten** mit Hilfe aktueller Technologien wie Foundation Models automatisiert werden[1]

- Aktuelle Studie von Noy und Zhang (2023) zeigt, dass durch Foundation Models
  - a) **Arbeitszeiten** (und somit Kosten) dramatisch **reduziert** werden können,
  - b) die **Qualität der Arbeit verbessert** werden



- Anwendungsbeispiel Chatbots:
  - Laut einer Studie von Salesforce erwarten 64 % der Kunden von Unternehmen Unterstützung in Echtzeit
  - Erreichbar durch Einsatz von Foundation Models?

Quellen:
1. *https://www.mckinsey.com/featured-insights/future-of-work/jobs-lost-jobs-gained-what-the-future-of-work-will-mean-for-jobs-skills-and-wages*
2. Noy, Shaked and Zhang, Whitney (2023). Experimental Evidence on the Productivity Effects of Generative Artificial Intelligence. *https://papers.ssrn.com/sol3/papers.cfm?abstract_id=4375283*

Anwendungsgebiete im Unternehmen

Einsatzmöglichkeiten in allen Unternehmensbereichen: Texte, Bilder, Videos & Audio

General Work

- Tools zur Effizienzsteigerung
- Unternehmens-KI
- Übersetzungen
- …

Sales & Customer Support

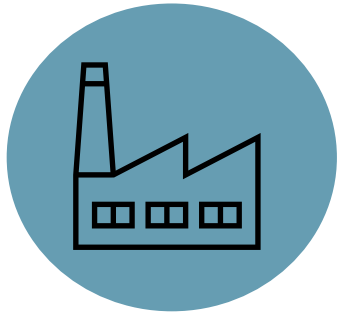
- Lead-Generierung
- Beratung & Angebote
- Beschwerdemanagement
- …

Human Ressources

- CV-Analyse
- Dokumenten-Management
- Stellenausschreibungen
- …

Marketing

- Werbetext-Entwürfe
- Inspiration für Bilder
- Personalisiertes Marketing
- …



R&D

- Protein-Strukturen
- Moleküle
- Literatur-Mining
- …

IT & Data Science

- Code erstellen
- Code erklären
- Erstellung von Daten
- …

Healthcare

- Arztbriefgenerator
- Erstberatung (Chatbots)
- Literatur-Mining
- …

Foundation Models
Die Kernbestandteile

Maschinelles Lernen

- Zumeist: Transformer Netze, selbstüberwachtes Lernen

Rechenpower

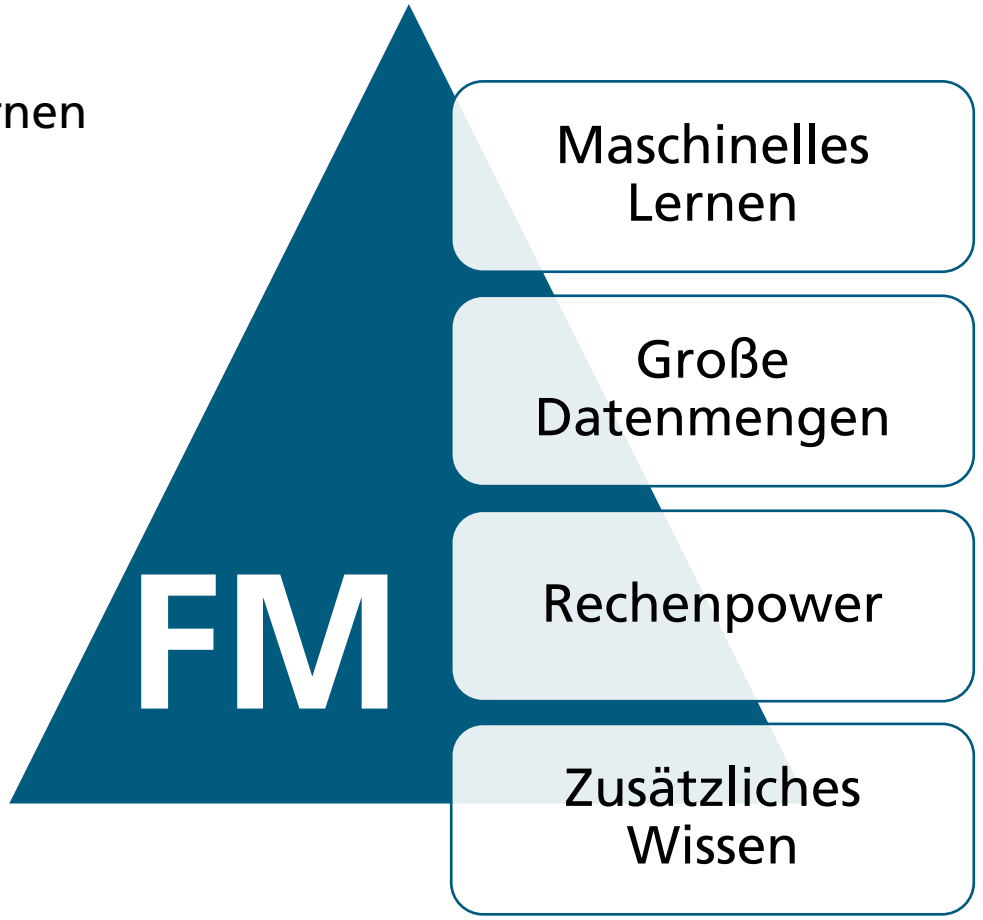
- Wachsende Rechenleistung

Große Datenmengen

- z.B. Posts aus sozialen Medien

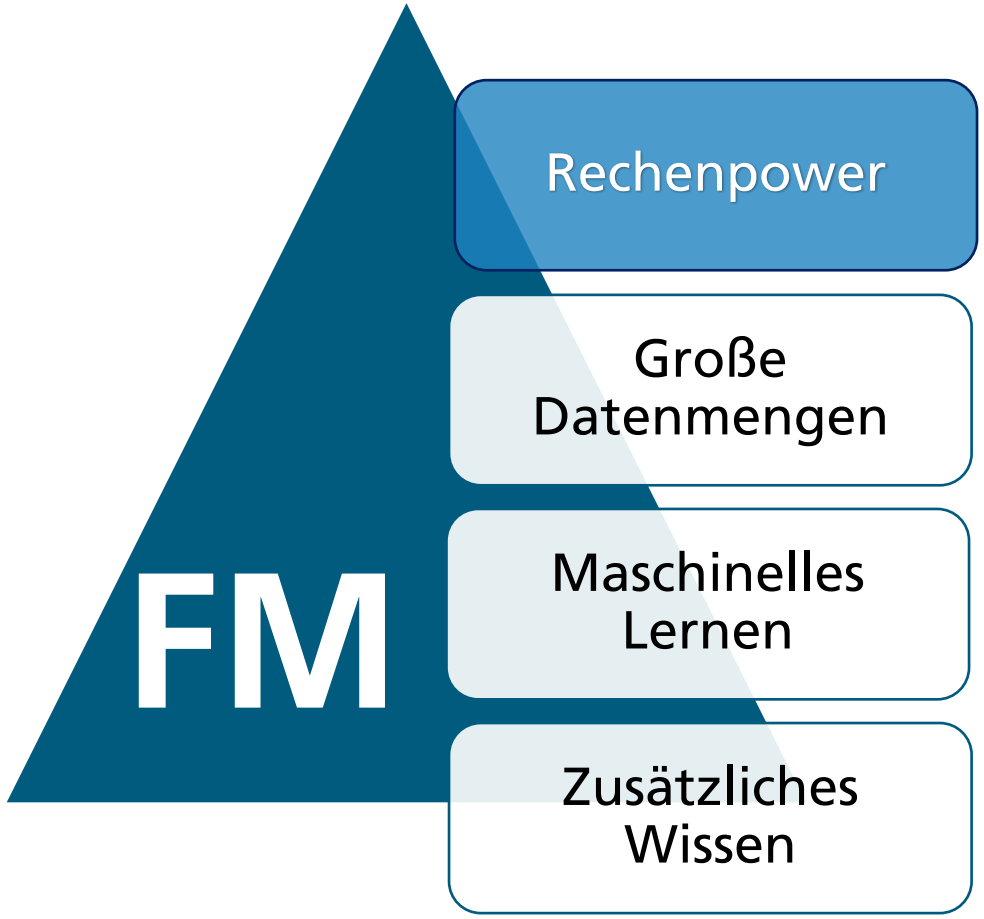
Integration von Wissen

- z.B. über Regeln, Ontologien, Probabilistic Soft Logic, etc.



Kernbestandteile von Foundation Models

Beispiel GPT-3 - Rechenpower



We are waiting for OpenAI to reveal more details about the training infrastructure and model implementation. But to put things into perspective, GPT-3 175B model **required 3.14E23 FLOPS of computing for training.** Even at theoretical 28 TFLOPS for V100 and lowest 3 year reserved cloud pricing we could find, this will take **355 GPU-years and cost $4.6M for a single training run**. Similarly, a single RTX 8000, assuming 15 TFLOPS, would take 665 years to run.

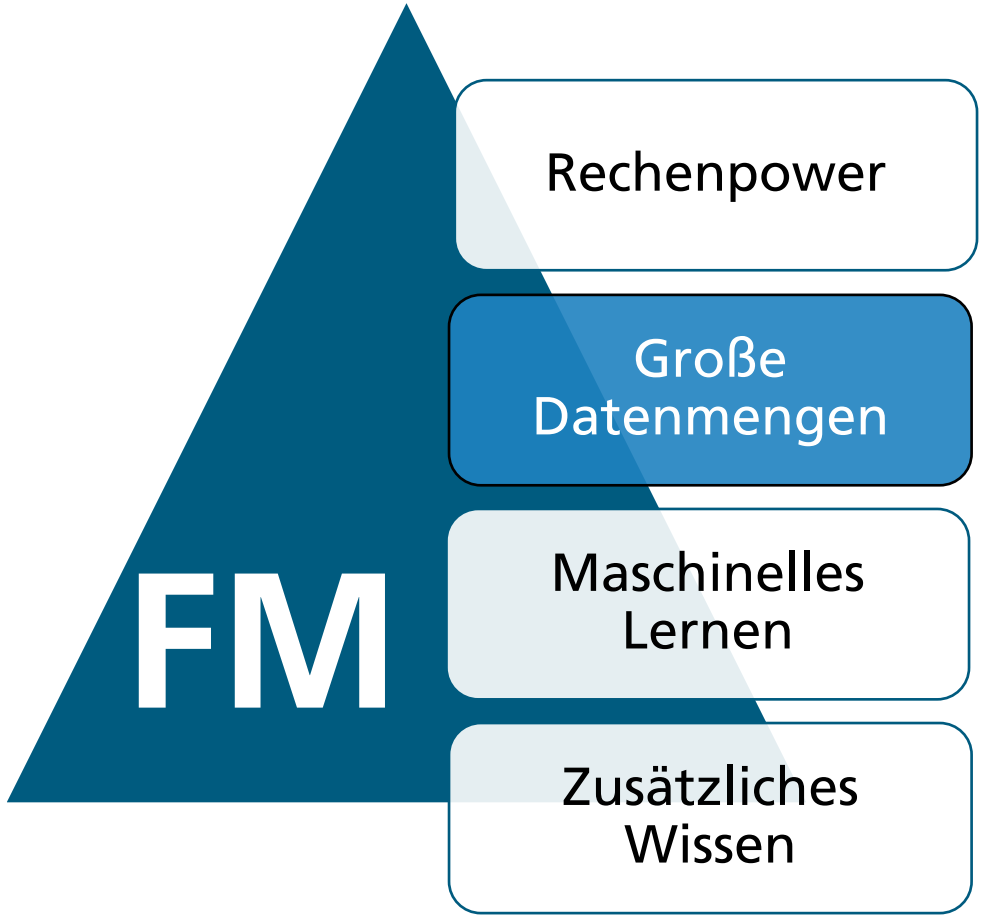
https://lambdalabs.com/blog/demystifying-gpt-3/

"The supercomputer developed for OpenAI is a single system with more than **285,000 CPU cores, 10,000 GPUs and 400 gigabits per second** of network connectivity for each GPU server," the companies stated in a blog.

https://news.developer.nvidia.com/openai-presents-gpt-3-a-175-billion-parameters-language-model/

Kernbestandteile von Foundation Models

Beispiel GPT-3 – Big Data



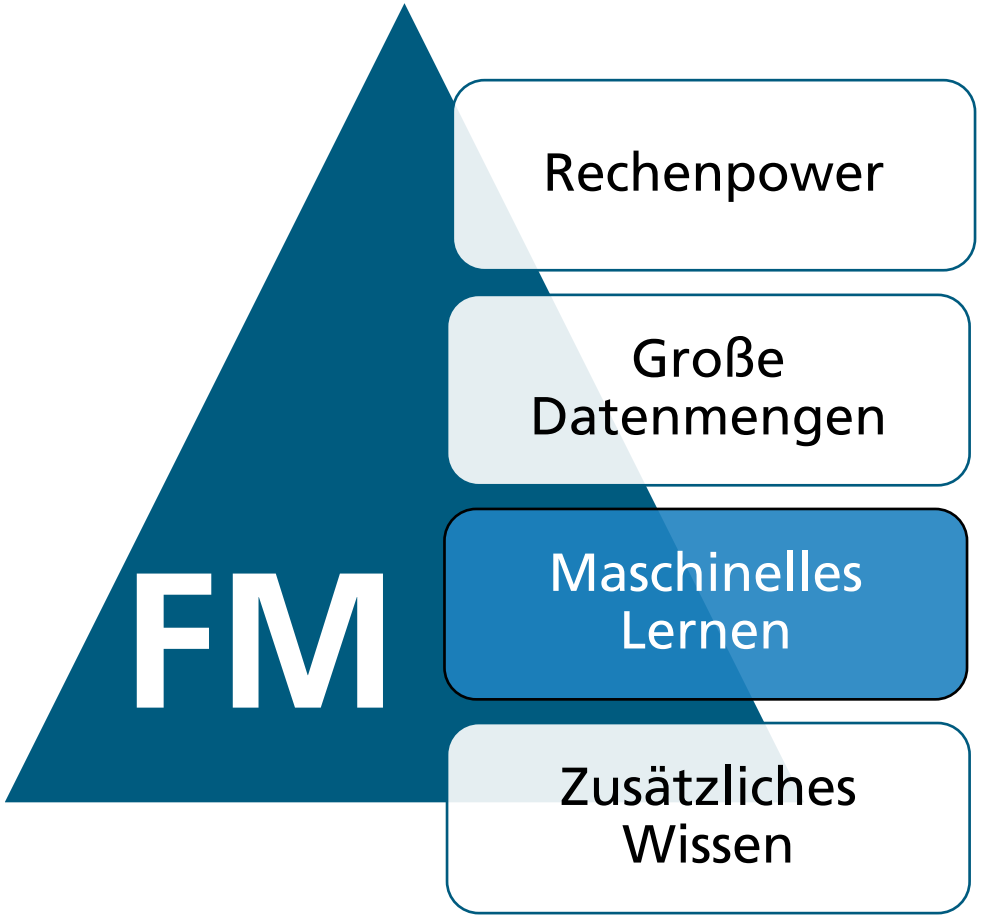
- GPT-3 wurde auf **570 GB Texten trainiert**
  - Anzahl Dokumente pro Sprache:

Englisch	235.987.420
Deutsch	3.014.597
Französisch	2.568.341

  - Paper zeigt, dass die Qualität des Modells von der **Größe des Modells, der Anzahl der Daten und der Trainingsdurchläufe** abhängt
  - Insgesamt beinhalten die Texte ca. **500 Millarden Token** (Achtung: 1 Token != 1 Wort)

Kernbestandteile von Foundation Models

Beispiel GPT-3 – Maschinelles Lernen



- GPT-3 basiert auf dem **„Decoder“** der **neuronalen „Transformer“**-Architektur
  - Die größte Variante ist ein sehr **tiefes neuronales Netz** mit **96 „Attention“ Schichten** und insgesamt **175 Milliarden Parametern**
- Das Original Transformer-Paper **„Attention is all you need“** (Vaswani et al., 2017) hat heute bereits ca. **78.500 Zitierungen**

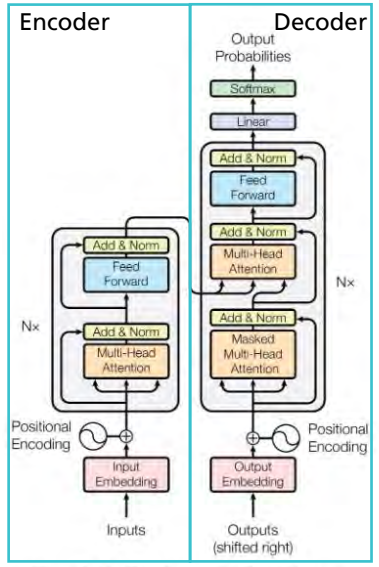
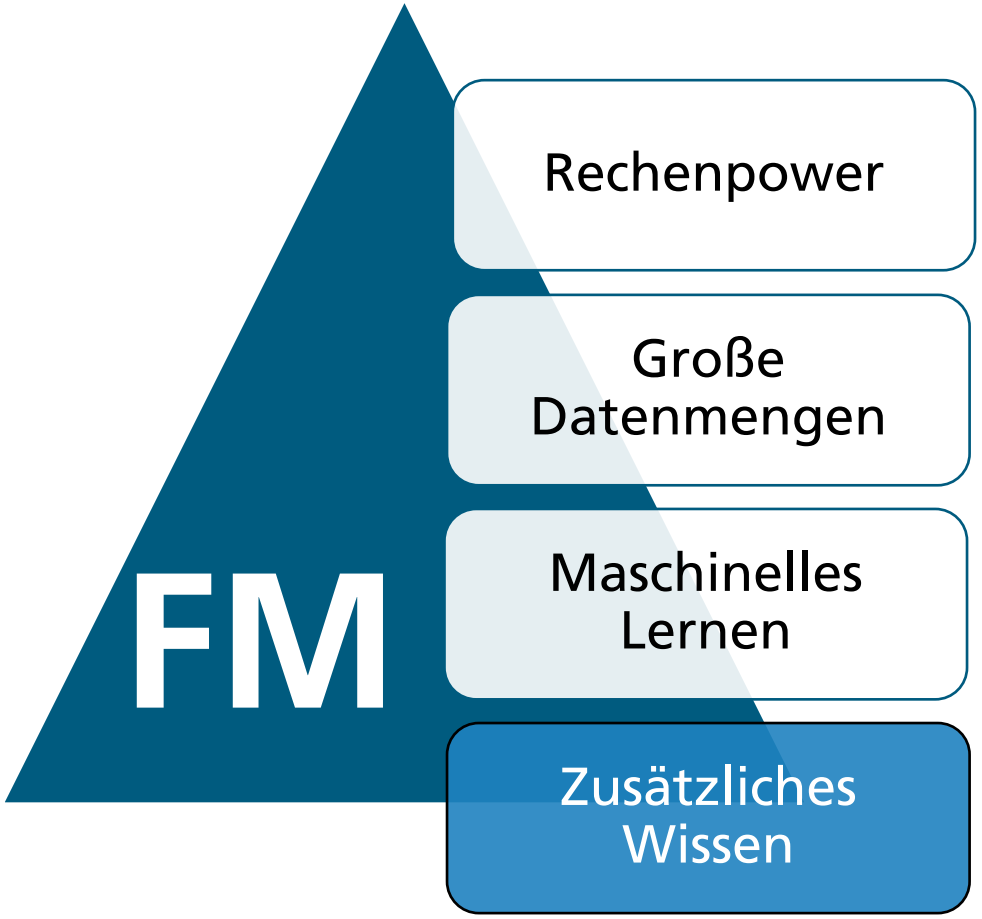


Figure 1: The Transformer - model architecture.

Vaswani et al. – Attention is all you need (2017)

Kernbestandteile von Foundation Models

Beispiel GPT-3 – Zusätzliches Wissen

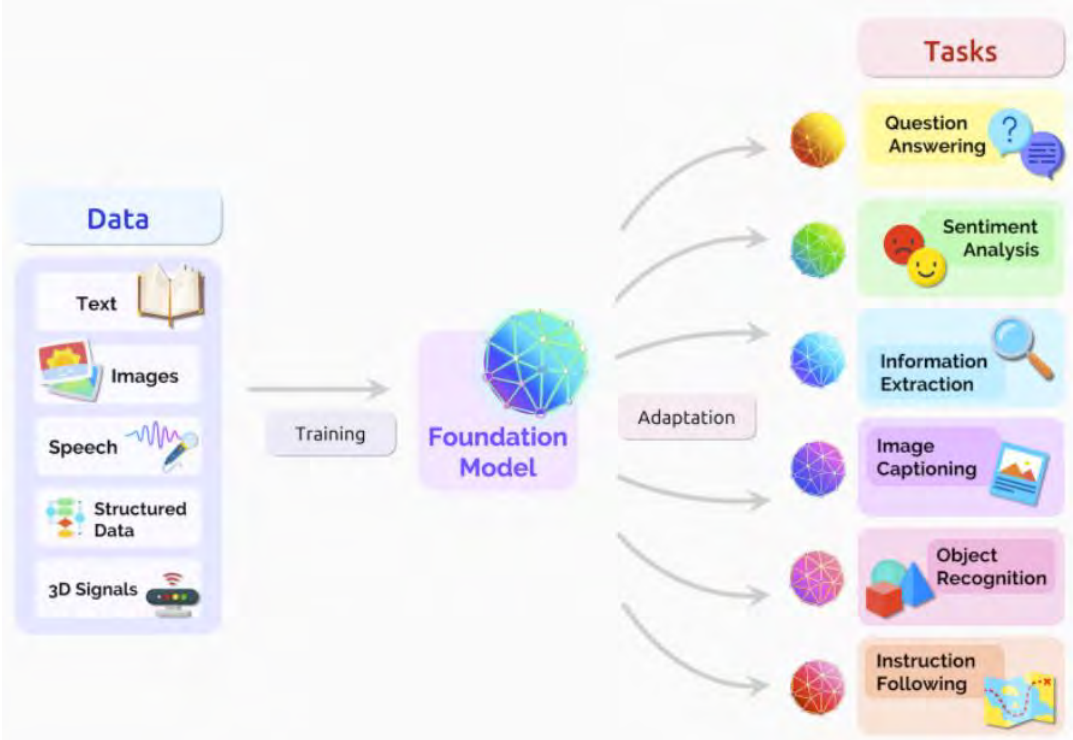


- GPT-3 nutzt **kein Wissen** außerhalb der Trainingstexte die es sieht, dies lässt **Potenzial** offen:
  - „[…] it still **sees** much **more text** […] **than a human** sees **in their lifetime** […]" – (Brown et al. 2020 – Language Models are Few Shot Learners)
  - "[…] apparently simple problems require humans to **integrate knowledge across vastly disparate sources** […] entirely **different sorts of tools** are needed, along with deep learning, if we are to **reach human-level cognitive flexibility**." – (Gary Marcus 2018 – Deep Learning – A Critical Appraisal")
  - Modelle neigen zum "Halluzinieren"

Foundation Models

Was sind ihre Vorteile gegenüber klassischen Modellen?

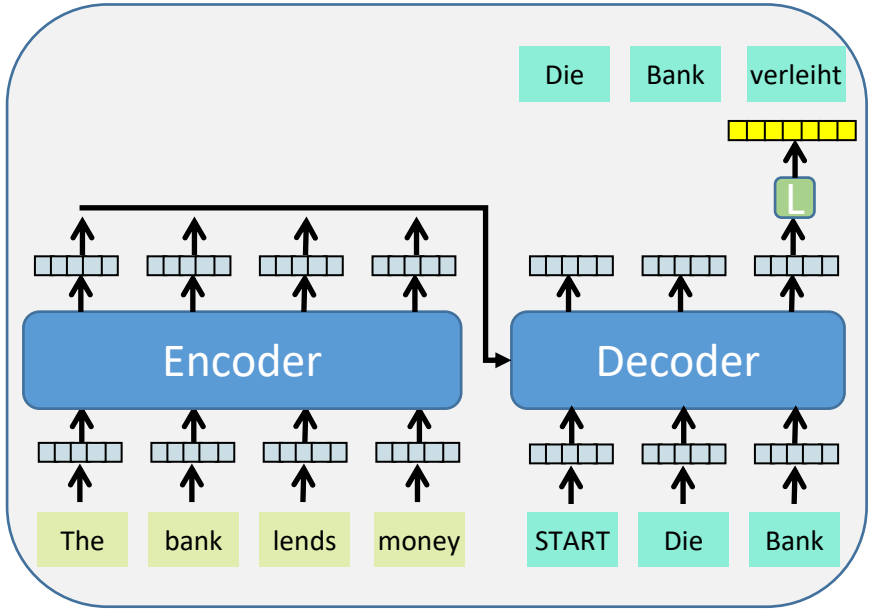
- **Adaptierbarkeit**
  - Modelle werden einmal auf vielen verschieden Daten vortrainiert (Pre-Training) und sind danach auf verschiedene Aufgaben adaptierbar
  - Adaptierung erfolgt entweder über Fine-Tuning oder Prompting
- **Emergente Fähigkeiten**
  - Große Modelle zeigen besondere Fähigkeiten z.B. bei Commonsense Reasoning
- **Bessere Performanz**
  - Zeigen bessere Ergebnisse auch bei wenigen bis keinen Trainingsbeispielen
- **Vereinheitlichte Architekturen**
  - Eine einheitliche Architektur für mehrere Modalitäten
  - Modelle können in Software einfacher und ohne größere Änderungen ersetzt werden

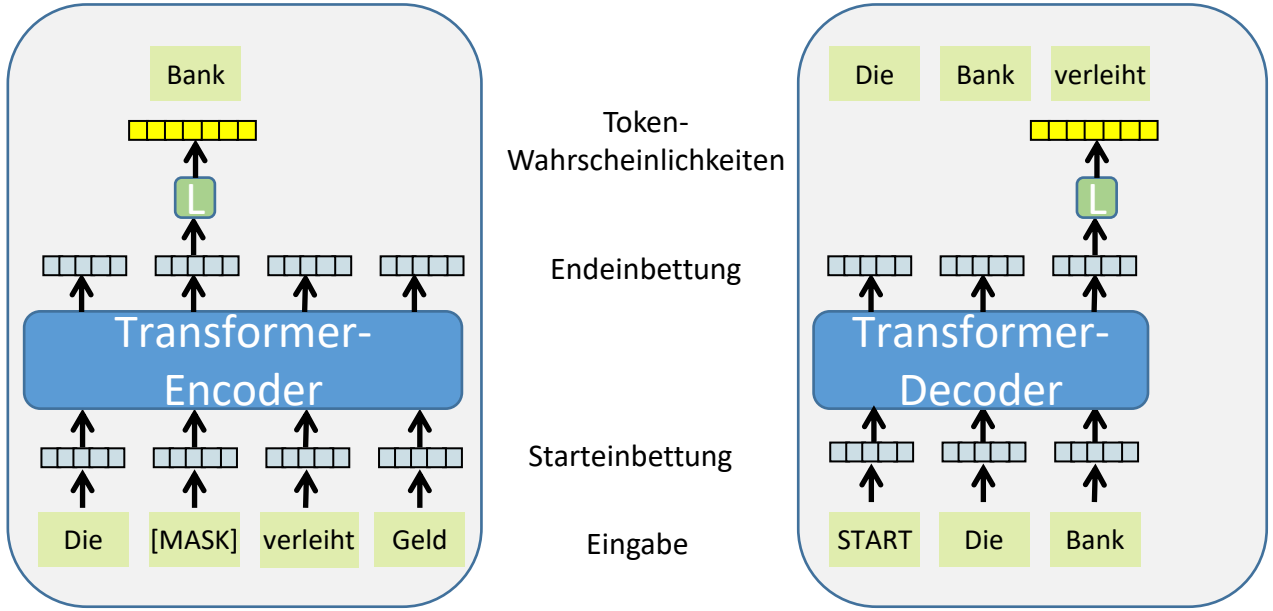


Bommasani et al. - On the Opportunities and Risks of Foundation Models - https://arxiv.org/pdf/2108.07258.pdf

Arten von textuellen Foundation Models

BERT, GPT, Transformer





Transformer: Übersetzungen eines Textes in einen anderen Text

- Kombination aus Encoder und Decoder

BERT-Modell: Pre-Training durch Vorhersage maskierter Token.

- Adaption durch Finetuning z.B. auf: Dokumentklassifikation, Informationsextraktion, etc.

GPT-Sprachmodell: Pre-Training durch Vorhersage des nächsten Tokens

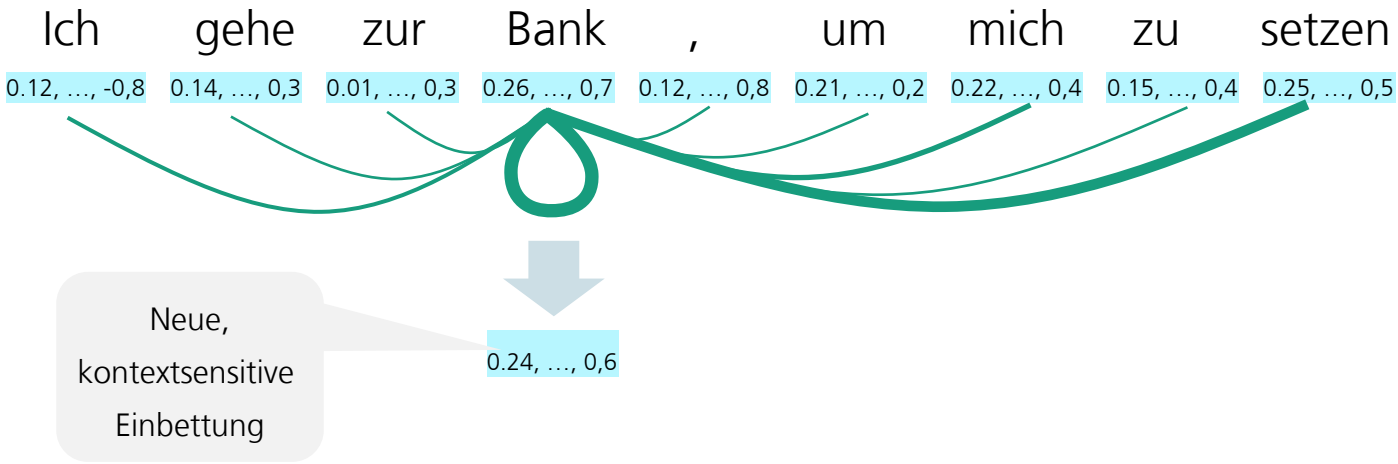
- Adaption durch Prompting, d.h. durch natürlichsprachliche Anweisungen

Erfassung von Wortbedeutungen

- Ausgangspunkt: Erfassung der **Bedeutung von Worten**
  - Ich gehe zur Bank, um Geld abzuheben. ➔ Bank ist Finanzinstitut
  - Ich gehe zur Bank, um mich zu setzen. ➔ Bank ist Sitzmöbel
  - ➔ *Bedeutung eines Wortes wird durch die anderen Worte bestimmt*
- Zu jedem Wort soll eine **Vektor (Einbettung/Embedding)** berechnet werden, die dessen Bedeutung darstellt.
  - Ziel: haben zwei Worte eine ähnliche Bedeutung, wenn deren Zahlenreihen sich wenig „unterscheiden“.

0.01, 1.2, 3.4, -1.4,…, 0,3

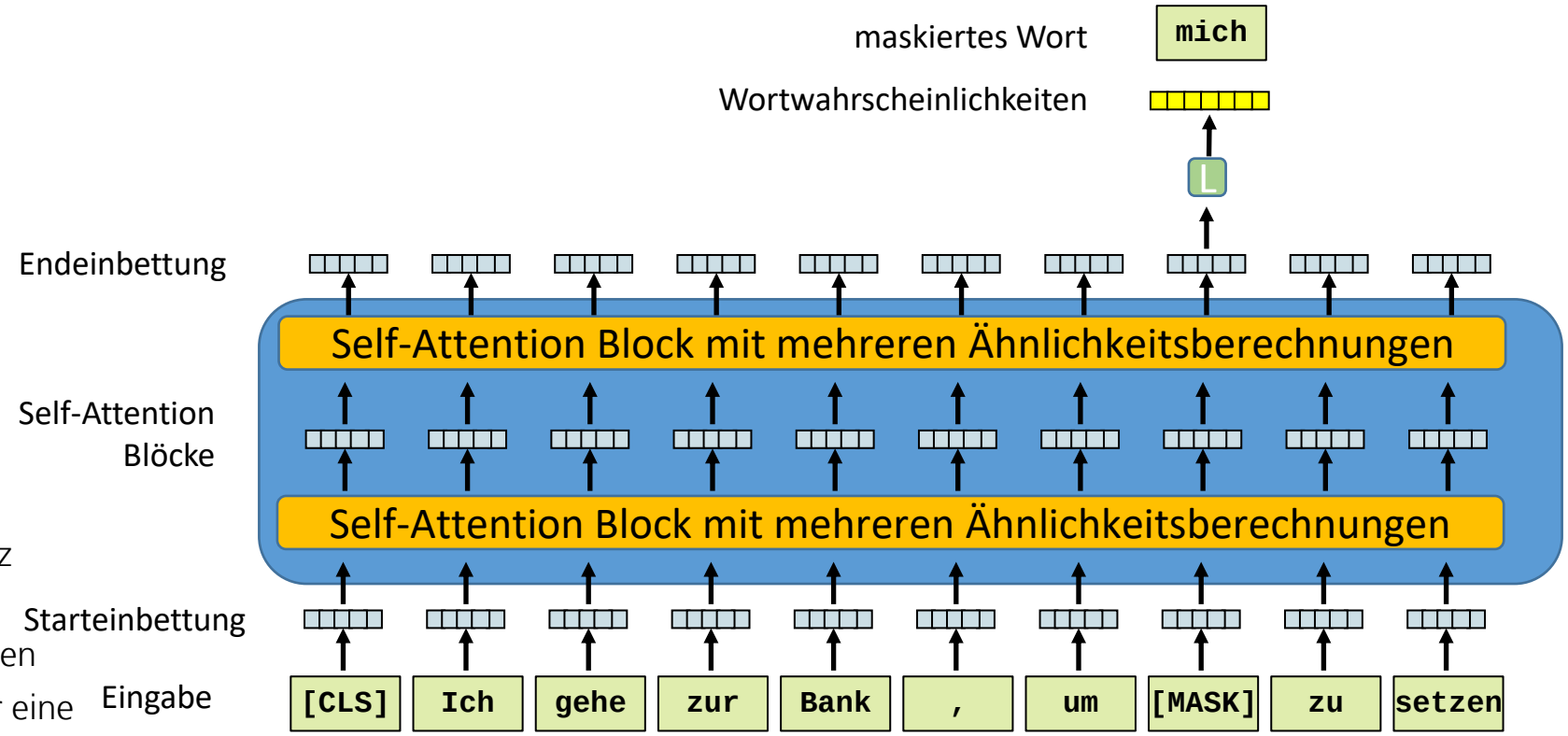
- Vorgehen
  - Starte mit einer vorgegebenen Einbettung für jedes Wort
  - Berechne Ähnlichkeiten aller Einbettungen zu der „Bank“-Einbettung durch *Self-Attention*
  - Berechne eine neue, **kontextsensitive Einbettung** für „Bank“ als einen gewichteten Mittelwert aus allen Einbettungen, wobei die ähnlichen bzw. wichtigen ein höheres Gewicht bekommen



Transferlernen

Pre-Training und Fine-Tuning I

- **Selbstüberwachtes Pre-Training:**
  - Lernen aus vielen Beispieltexten, in denen einzelne Worte **maskiert** sind
  - Berechne Wahrscheinlichkeit für das maskierte Wort
  - Ziel: ändere Netz-Parameter, so dass die maskierten Worte eine hohe Wahrscheinlichkeit erhalten



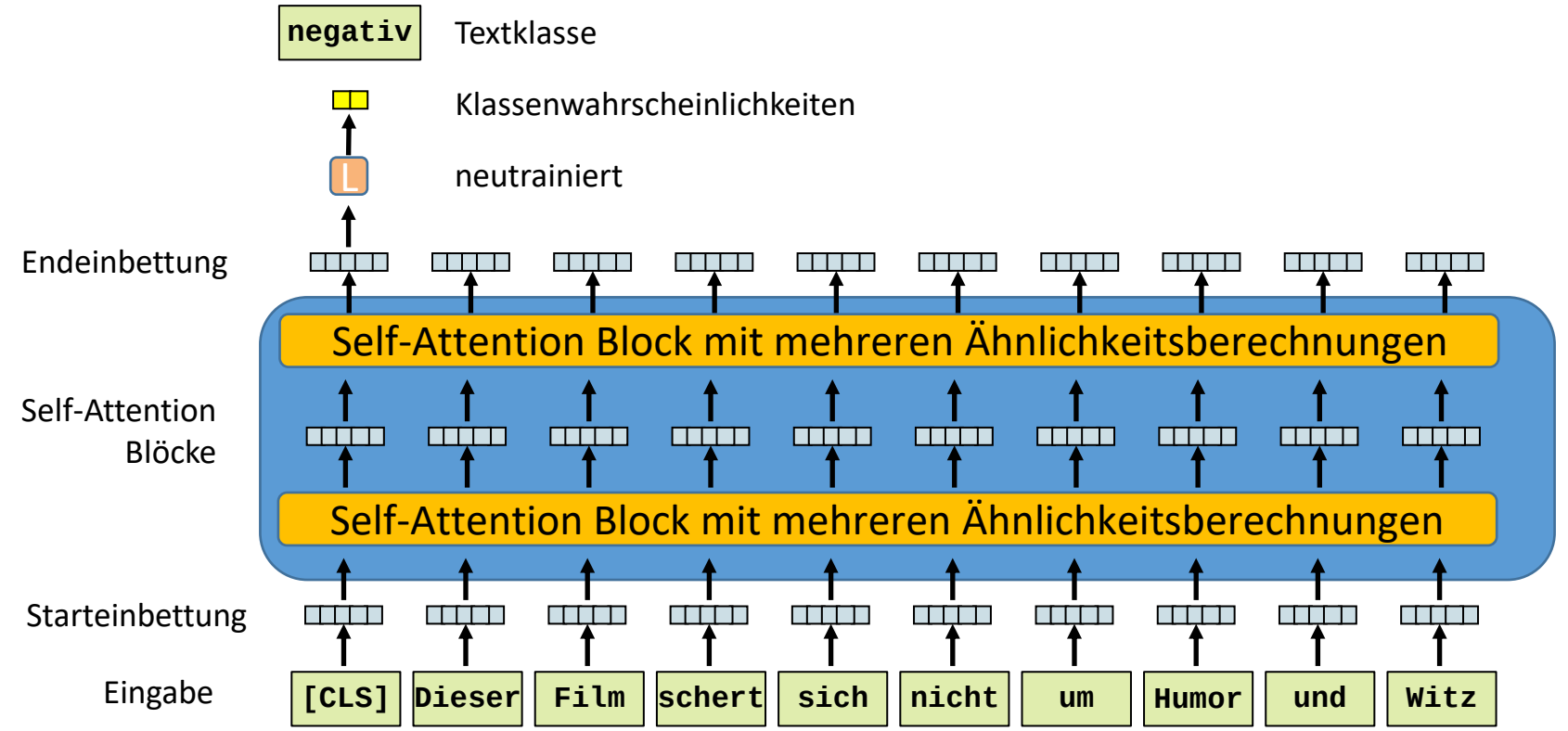
- Durch Vortraining auf einem großen Datensatz lernen die Modelle die Struktur der Sprache
  - **Syntax**: wie Wörter einen Satz bilden können
  - **Semantik**: wie man einen Sachverhalt oder eine Beziehung ausdrückt

BERT erzeugt **kontextabhängige**, sehr aussagekräftige Einbettungen

Transferlernen

Pre-Training und Fine-Tuning II

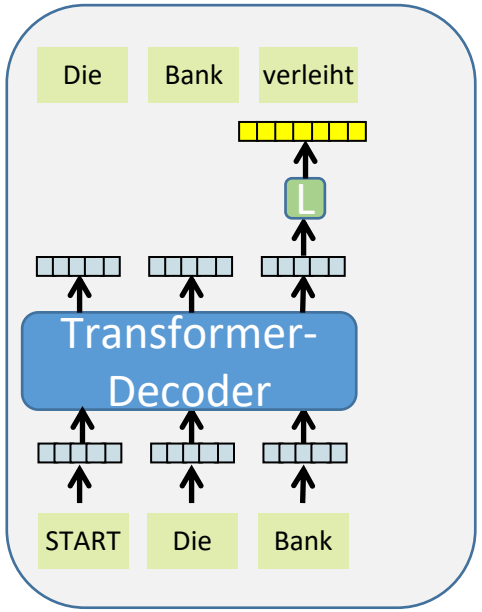
- Anpassen der Modelle durch **Finetuning** mit wenigen Trainingsdaten
  - Passe alle Parameter ein wenig an
  - Neues Modell zur Klassenprognose
  - Annotierte Beispieldaten: Filmbeschreibung, Bewertung negativ
- Viel bessere Ergebnisse als vorher mit nur wenigen annotierten Daten



Feintuning bewirkt **bessere Prognosen** mit weniger annotierten Daten

Prompting und Adaptionen

GPT-Sprachmodelle reagieren auf Anweisungen



- Ein GPT-Sprachmodell erhält einen **Starttext**
  - Es erzeugt eine Fortsetzung die syntaktisch richtig und inhaltlich plausibel ist
  - Enthält der Starttext eine **Anweisung**, so wird die entsprechende Antwort erzeugt

- Das GPT-Sprachmodell kann durch ein oder mehrere **Beispiele** instruiert werden, eine Aufgabe zu lösen:
  ➔ **Kein Finetuning** für Aufgabe erforderlich, kaum bis keine weiteren Daten notwendig

- Durch Adaptionen (z.B. Instruction-Tuning und Reinforcement Learning from Human Feedback) kann das Sprachmodell trainiert werden, …
  - **Beleidigende** Aussagen zu vermeiden
  - beliebige Anfragen **korrekt** und **zielgenau** zu erfüllen
  - auf den Dialogpartner einzugehen und den Dialog **interessant** zu gestalten

GE Hier wird einem achtjährigen erklärt, wie ein Sprachmodell funktioniert.

Hallo, lieber achtjähriger! Ich bin ein Sprachmodell, das dazu entwickelt wurde, auf Fragen zu antworten und Gespräche zu führen. Ich wurde von vielen Menschen trainiert, um zu verstehen, was sie sagen, und um darauf zu reagieren. Genau wie du lernst, wenn du Bücher liest oder wenn dir jemand etwas erzählt, lerne ich aus den Texten und Gesprächen, die ich sehe und höre.

GE Tweet: "Ich hasse es, wenn der Akku meines Handys leer ist." Stimmung: negativ. ### Tweet: "Heute lief alles super." Stimmung: positiv. ### Tweet: "Dieses neue Musikvideo ist unglaublich!". Stimmung:

positiv.

Zugang über Drittanbieter

Foundation Modelle als SaaS

Die Nutzung von Foundation Modellen lässt sich über Anbieter realisieren. Diese bieten die Anpassung auf das Einsatzfeld, den Betrieb in der Cloud und Schnittstellen für den Zugriff.

Dadurch können Angebote schnell auf den Markt gebracht werden, weil auf die Expertise und die Ressourcen den Anbieters zurückgegriffen werden kann.

TraumAInterfaces

Automatische Dokumentation im Schockraum

Idee

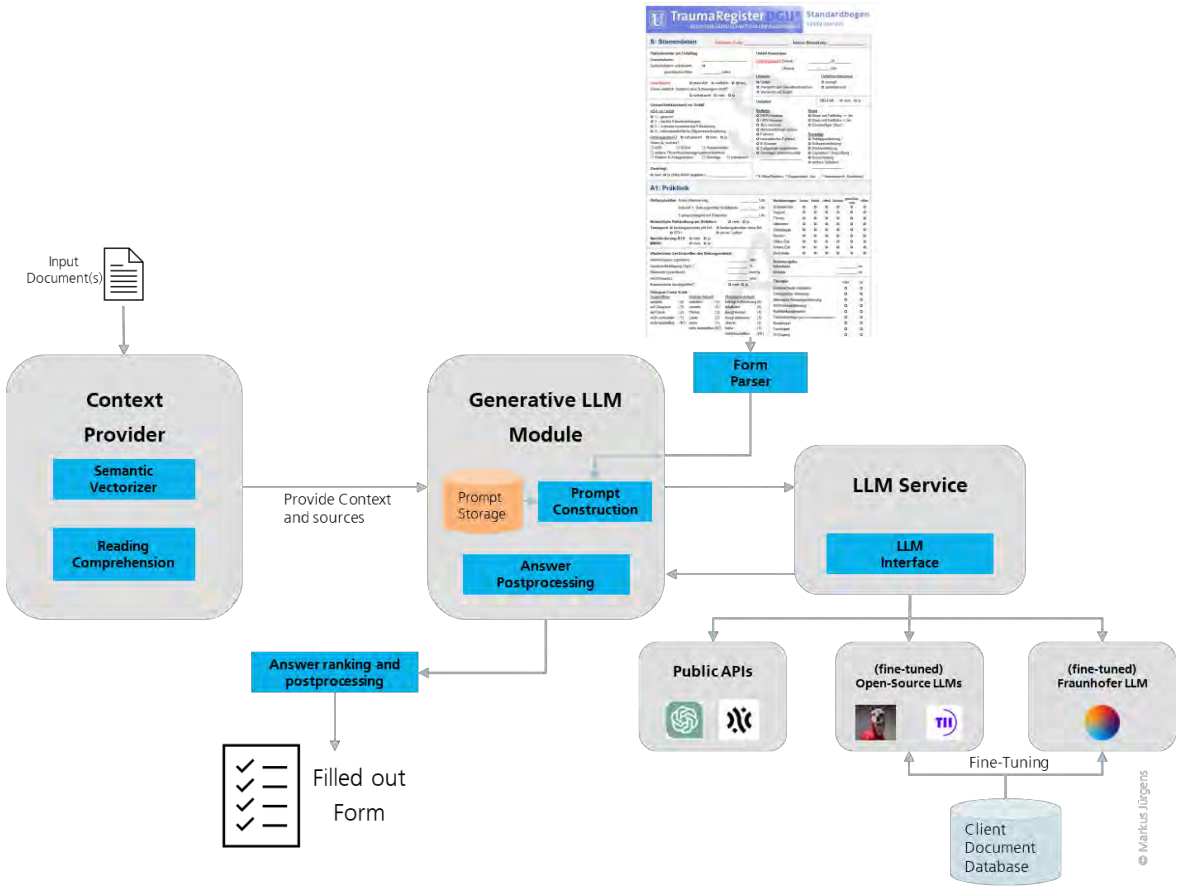
- Informationsverlust nach der Schockraumbehandlung minimieren
- Automatisiertes Befüllen des TraumaRegister Standardbogen
- Nutzen von Transkripten aus Schockraum-Simulationen zum Antrainieren

Herausforderung

- Keine Trainingsdaten vorhanden
  - Wegen Pandemie keine Möglichkeit Trainingsdaten zu erstellen: Benötigt sind die Transkripte der Behandlungssimulationen und die entsprechend ausgefüllten Bögen mit Markierung der jeweiligen Antworten die zu Einträgen geführt haben

Unsere Lösung

- Ausnutzen der Generalisierungsfähigkeit von Large Language Models
- Gegeben einem Schockraum-Transkript, werden zu jeder Frage Prompts an das Large Language Model gestellt
- Einschränkung der Antworten auf die im Formular vordefinierten Möglichkeiten
- Erfolgreiche Umsetzung mit OpenAI GPT-3.5 Modellen



Fraunhofer IAIS

Drittanbieter

Warum nicht einfach für alles GPT-4 und Co nutzen?

- **Datensicherheit**
  - Anfragen werden in der Cloud verarbeitet, d.h. Daten müssen die eigene Infrastruktur verlassen
- **Vertrauenswürdigkeit**
  - Modelle „halluzinieren“ und erfinden Teile ihrer Antworten
  - Sie geben uns kein Maß für ihre Sicherheit bei einer Antwort
  - Entscheidungen von LLMs sind nicht inhärent erklärbar, d.h. es ist nicht klar warum eine bestimmte Antwort gegeben wird
    - Referenz auf Vortrag: Lorenz Lehmhaus (Aleph Alpha GmbH)
  - Die Modelle nennen keine Quelle für ihr Wissen
  - Sie können „toxische“ Antworten produzieren
- **Souveränität**
  - Keine Kontrolle über die Modelle
  - Abhängigkeit von Drittanbietern
- **Ressourcenverbrauch**
  - Modelle brauchen riesige Ressourcen und Energie
  - Teilweise extrem lange Antwortzeiten


MATT BURGESS SECURITY 04.04.2023 05:00 PM

ChatGPT Has a Big Privacy Problem

Italy's recent ban of Open AI's generative text tool may just be the beginning of ChatGPT's regulatory woes.

https://www.wired.co.uk/article/italy-ban-chatgpt-privacy-gdpr


ChatGPT: US lawyer admits using AI for case research

27 May

REUTERS

ChatGPT can answer questions using natural, human-like language and mimic other writing styles

By Kathryn Armstrong
BBC News

A New York lawyer is facing a court hearing of his own after his firm used AI tool ChatGPT for legal research.

A judge said the court was faced with an "unprecedented circumstance" after a filing was found to reference example legal cases that did not exist.

what is heavier: kilo of water or kilo of air?

A kilogram of water is heavier than a kilogram of air. Water is a dense, liquid substance that has a mass of about 1 kilogram per liter at standard temperature and pressure. Air, on the other hand, is a gas that is much less dense than water and has a lower mass. A volume of air that weighs 1 kilogram would occupy a much larger space than a liter of water, which weighs the same amount.

https://babich.biz/4-biggest-issues-with-ai-tools/

Whoops, Samsung workers accidentally leaked trade secrets via ChatGPT

ChatGPT doesn't keep secrets.

https://sea.mashable.com/tech/23047/whoops-samsung-workers-accidentally-leaked-trade-secrets-via-chatgpt

https://www.bbc.com/news/world-us-canada-65735769

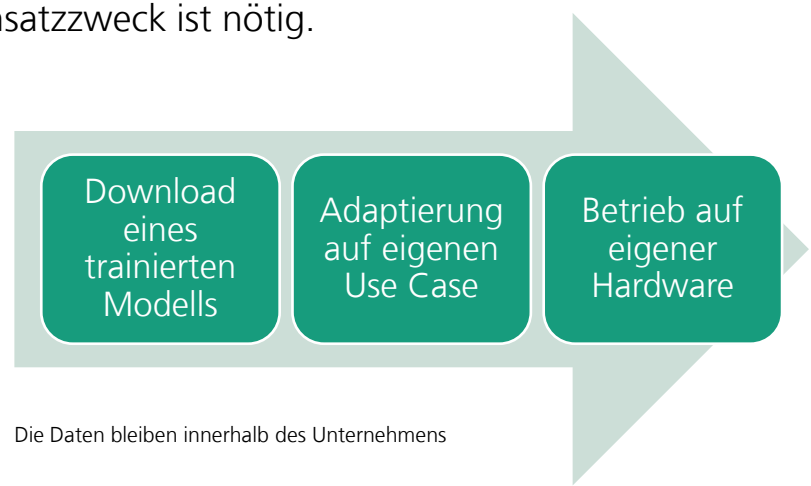
Adaptieren eigener Modelle

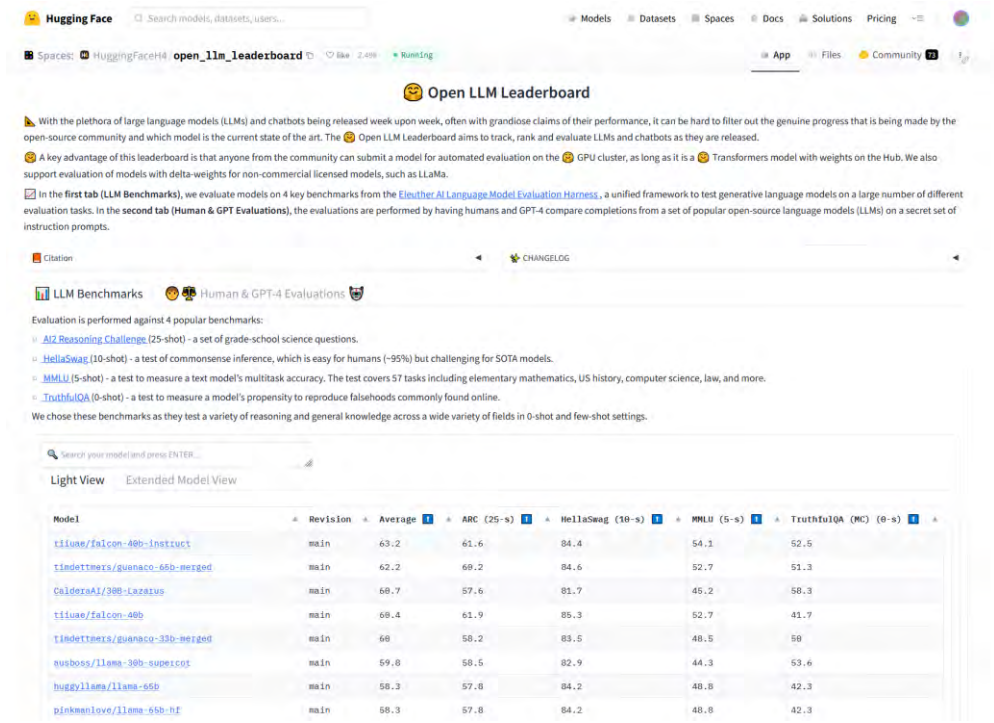
Open-Source Foundation Modelle

Viele Forschungseinrichtungen und einige Unternehmen stellen die von ihnen trainierten Modelle Open Source zur Verfügung. Diese können für den jeweiligen Use Case adaptiert werden.

Auch kleinere Modelle zeigen gute Performance, wenn Sie ausreichend lange auf ausreichend vielen Daten trainiert warden.

Durch Wiederverwendung dieser Modelle ist es nicht mehr notwendig, das umfangreiche Pre-Training durchzuführen, lediglich eine Adapation an den Einsatzzweck ist nötig.





Quelle: https://huggingface.co/spaces/HuggingFaceH4/open_llm_leaderboard

Offene Alternative zu GPT-4: OpenGPT-X:

Large Language Models made in Europe

Erstellung von LLMs mit Fokus auf Datensicherheit und europäische Sprachen

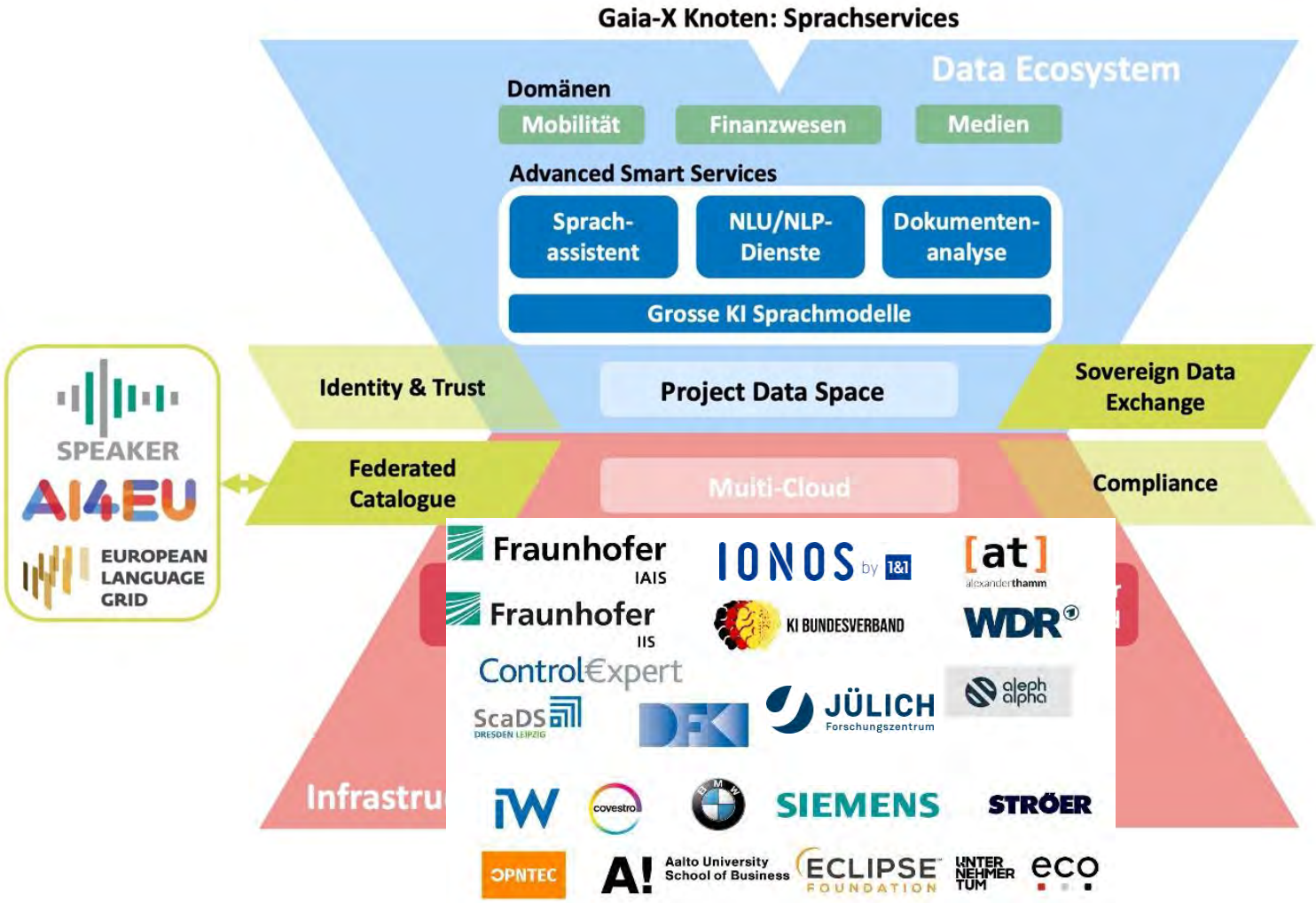
Offen verfügbare Modelle

Sicherstellung digitaler Souveränität

Kooperation von Partnern aus Industrie und Forschung

Erste 3 und 13 Mrd. Parameter Modelle wurden auf circa 600 GPUs trainiert

Nächster Schritt: **70 Mrd. Parameter Modell**

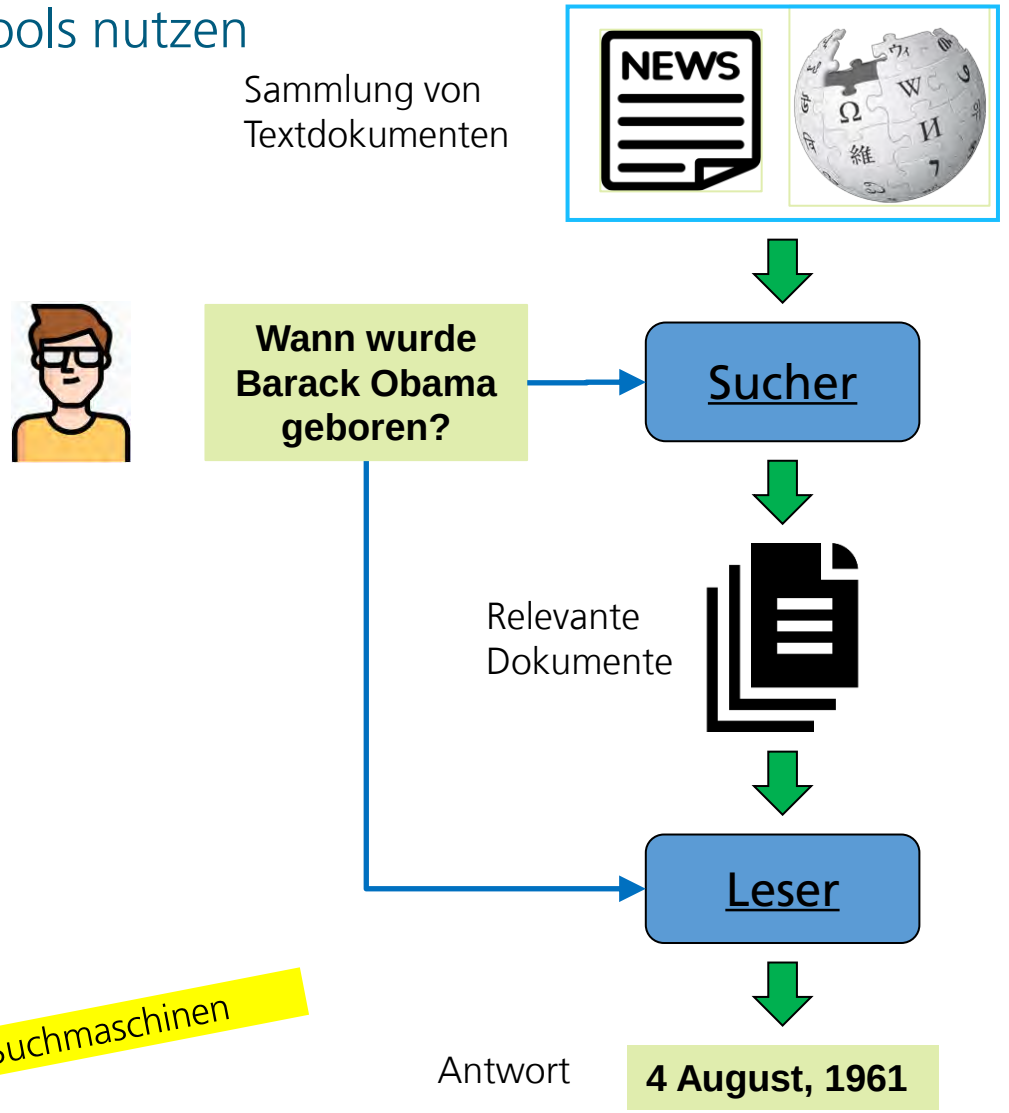


Fraunhofer

Retrieval und Tools

Foundation Models können zusätzliche Informationen und Tools nutzen

- Sprachmodelle sollten in Anwendungen eingebettet werden, die ihre Vertrauenswürdigkeit sicherstellen
  - Referenz auf Vortrag: Jochen Papenbrock (NVIDIA Corp.)
- Beispiel: Sie können mit anderen Tools kombiniert werden und sogar gezielt darauf trainiert werden Tools wie Suchmaschinen, Taschenrechner und Co zu nutzen
- Nutzung zusätzlicher Dokumente durch **Suchmaschine**
  - Nutzer stellt Anfrage
  - „Sucher“ durchsucht Dokumentensammlung und findet relevante Dokumente
  - „Leser“ Sprachmodell erhält Anfrage und relevante Dokumente
    ➔ Erzeugt daraus eine Antwort
  - Sprachmodell kann **riesiges Hintergrundwissen** verwenden
    - Wissen ist aktuell
    - Kann **alternative Formulierungen** berücksichtigen
    - Gefundene Dokumente können als Begründung / Referenzen genutzt werden
    - **Datenbanken** können analog verwendet werden

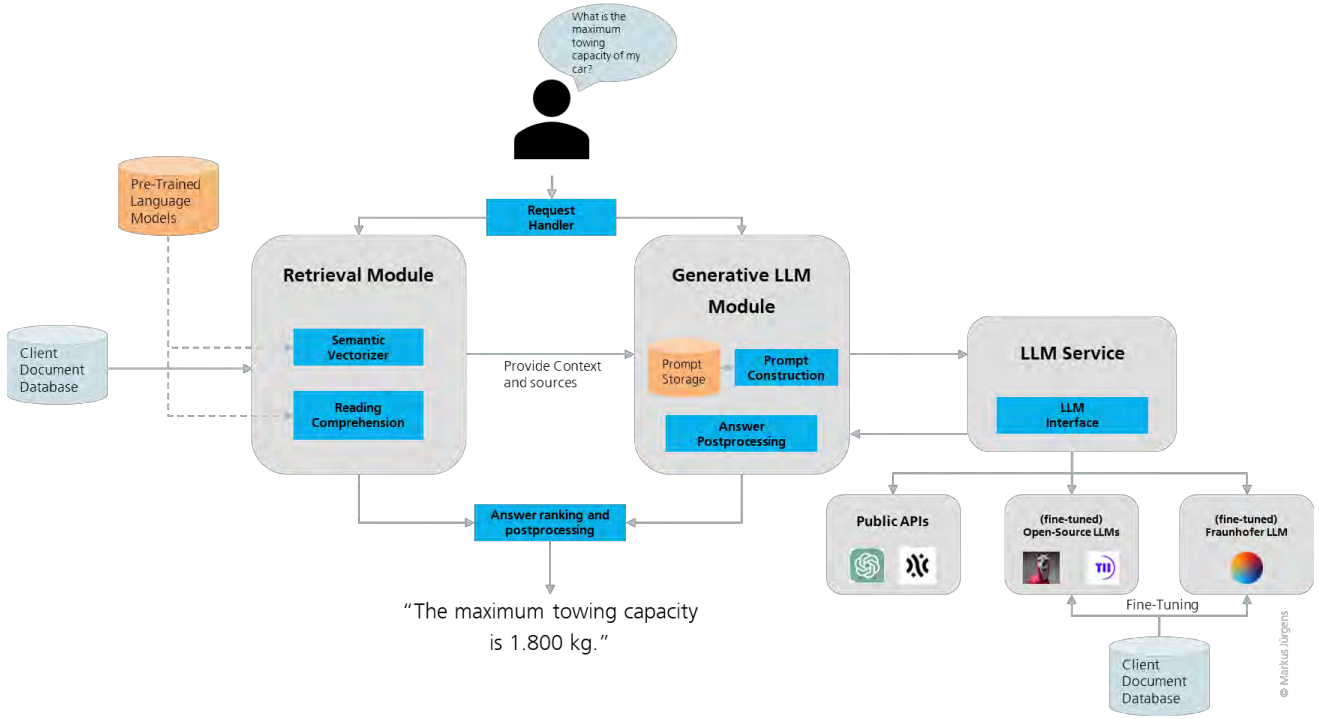


Präzisere Antworten durch Suchmaschinen

Conversational AI

Mit Large Language Models und Retrieval

- Projekt mit einem DAX-Kunden
- Transfer zum Produktionsteam des Kunden bis Ende 2023
- System kann Nutzer*innenanfragen zu Produkten des Kunden beantworten
- **Fokus:** Balance aus Faktentreue, Produktsicherheit und natürlich klingenden Antworten
  - Retrieval aus einer Datenbank mit Kundendokumenten
  - Referenzierung der Quelle der Antwort
  - Zusätzliche Möglichkeit LLM APIs anzusprechen
  - Die beste Antwort wird aus Retrieval und LLM Antwort ausgesucht, je nachdem was besser auf die Anfrage des Kunden passt.



Fraunhofer IAIS

Encoder Modelle

Die leichtgewichtige Alternative

- **Ressourcenbedarf**
  - Die meisten Encoder-Modelle sind wesentlich kleiner als ihre generativen Gegenstücke
  - Sie können auf normaler Hardware trainiert werden (z.B. eine einzelne GPU)
  - Fine-Tuning dauert meistens nur wenige Minuten bis Stunden (benötigt aber wesentlich mehr Daten als Propting)
    - Für bestimmte Aufgaben funktioniert prompt-based fine-tuning mit Encodern
- **Datensicherheit**
  - Encoder machen es nahezu unmöglich Trainingsdaten aus ihnen wiederherzustellen*
  - Modelle können einfach auf eigener Infrastruktur betrieben werden, so dass Daten die eigenen Server nicht verlassen müssen
- **Robustheit und Transparenz**
  - Modelle können so kalibriert werden, dass Sie Indikationen über ihre eigene Unsicherheit geben können
  - Modelle können einfach mit vorhandenen Erklärbarkeitsmethoden kombiniert werden
  - Antworten können über zusätzliche Ausgabeschichten einfach auf vordefinierte Werte eingeschränkt werden
- **Sehr Gute Performance**
  - Auf Aufgaben wie Informationsextraktion, Klassifikation, etc. sind die Modelle teilweise besser als generative Modelle**

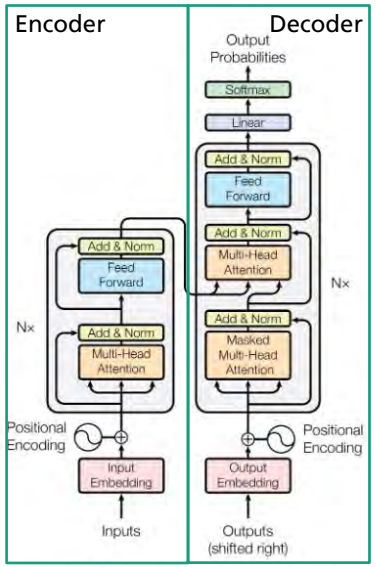


Figure 1: The Transformer - model architecture.

Vaswani et al. – Attention is all you need (2017)

*Lehman et al., 2021, "Does BERT Pretrained on Clinical Notes Reveal Sensitive Data?", Proceedings of the 2021 Conference of the NAACL

**Chen et al., 2023, "Large Language Models in Biomedical Natural Language Processing: Benchmarks, Baselines, and Recommendations"

Right-COding

Kodierung von Patient*innenakten

Idee

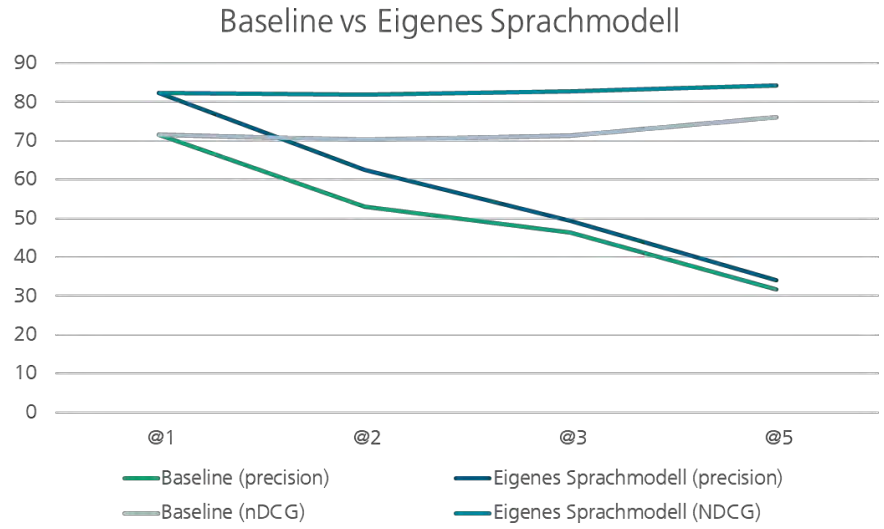
- Abrechnung von Patient*innenakten teil-automatisieren

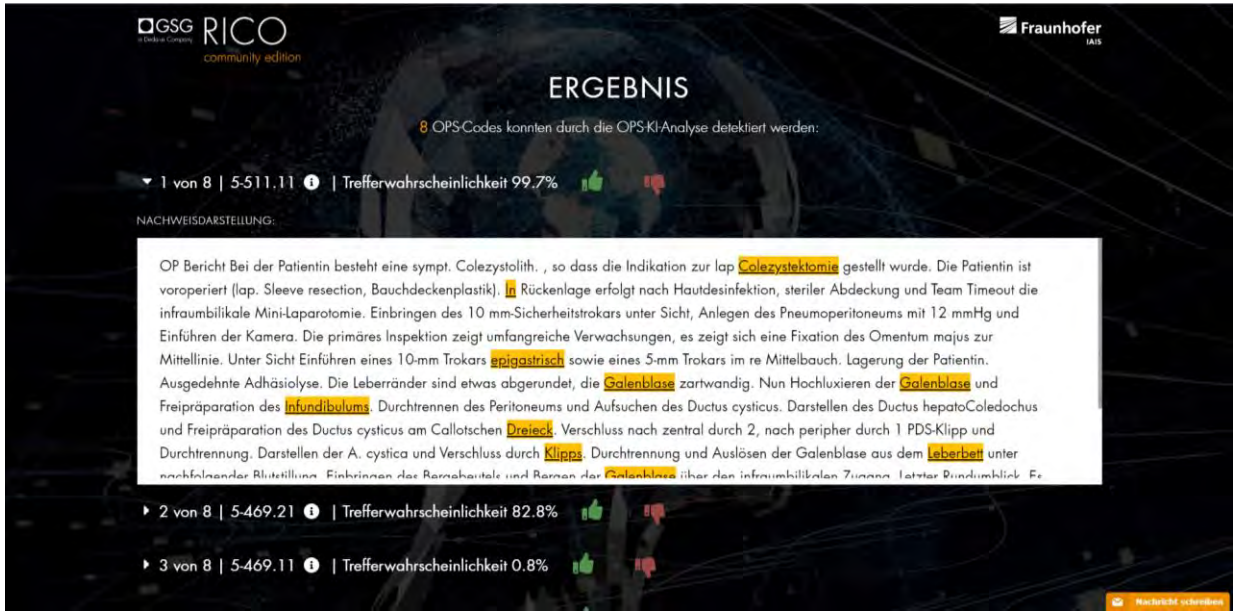
Herausforderung

- Daten dürfen die Krankenhäuser nicht verlassen
- Eingeschränkte Hardware, d.h. jeweils 1 GPU pro Krankenhaus(-verbund)
- Nachvollziehbarkeit zwingend gefordert durch Gesetzgeber
- 10.000 verschiedene Diagnose- und Prozedurcodes müssen unterschieden werden

Unsere Lösung

- Eigens vortrainiertes Encoder-Modell (auf Daten von Krankenhausverbund)
- Fine-tuning auf allen teilnehmenden Krankenhäusern durch eigenes Framework zum verteilten Lernen (Daten bleiben bei den Häusern)
- Gleiches Modell wurde auf die Vorhersage von Codes und Negationen adaptiert
- Erklärbarkeit durch Integrated Gradients Verfahren und Integration von Ontologie
- Modell gibt eine Indikation darüber wie sicher vorhersagen sind





Foundation Models

Wir unterstützen Sie gerne

Schulungen

- Wir halten Schulungen zu Foundation Models
  - Theoretischer Hintergrund
  - Praktische Beispiele

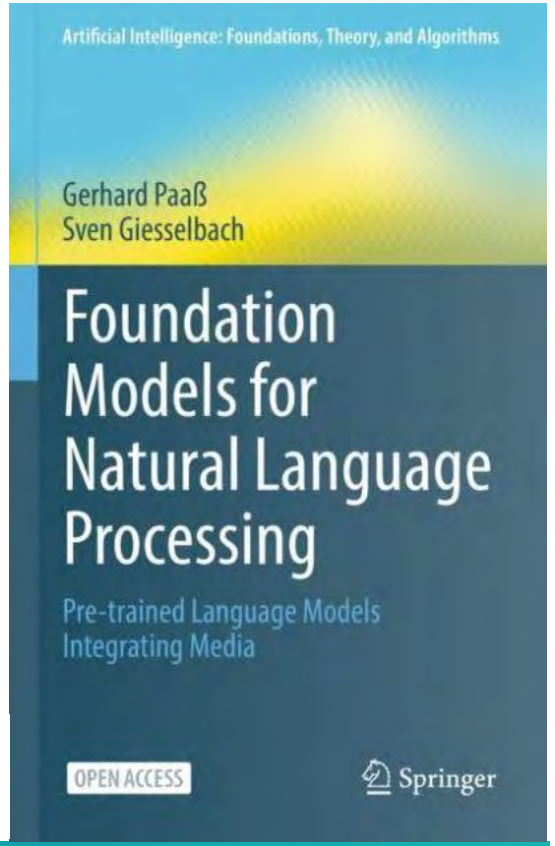
Innovation Briefing

- Wir bieten einen 2-3-stündigen Kurzeinstieg in das Thema Foundation Models
- Wir beantworten die Frage wie Businesses und Geschäftsbereiche durch Foundation Models beeinflusst werden
- Wir helfen Zwischen Hype und Anwendbarkeit zu unterscheiden

Foundation Model Assessment

- Wir identifizieren gemeinsam mit Ihnen Use Cases
- Wir unterstützen Sie bei der Implementierung

openGPT-X



https://link.springer.com/book/10.1007/978-3-031-23190-2

Fraunhofer can kickstart your Foundation Model journey

Danke für Ihre Aufmerksamkeit

Melden Sie sich gerne bei mir

Fraunhofer IAIS
Schloss Birlinghoven

53757 Sankt Augustin

www.iais.fraunhofer.de
www.iais.fraunhofer.de/nlu
https://machinelearning-blog.de/

Kontakt
Sven Giesselbach
Team Lead Natural Language Understanding
Fraunhofer IAIS
Telefon +49 2241 14-2249
E-Mail: sven.giesselbach@iais.fraunhofer.de

