

Seminararbeit

aus dem Fach Physik

Thema: Wie schlau sind ChatGPT & Co. wirklich?

Verfasser: Alex Eschbaumer
W-Seminar: Wissenschaftsjournalismus
Seminarleiter: Sebastian Bauer
Abgabetermin: 07. November 2024

Erzielte Note: in Worten:
Erzielte Punkte: in Worten:
(einfache Wertung)

Abgabe beim Oberstufenbetreuer am:

Unterschrift des Seminarleiters

Inhaltsverzeichnis

1	Einleitung	1
1.1	Forschungsfrage und Zielsetzung	1
1.2	Methodik und Aufbau	2
2	Chatbots in der Gesellschaft	3
2.1	Chatbots mit Gefühlen	3
2.2	Mediale Berichterstattung	4
2.3	Chatbots und IQ-Tests	5
3	Funktionsweise von Chatbots	6
3.1	Begriffe und Grundlagen	6
3.2	Transformer und Chatbots	8
3.3	Warum ist das noch nicht AGI?	11
3.4	Was folgt daraus?	11
3.4.1	Immer noch nur Algorithmen	11
3.4.2	Das falsche Bild in den Medien	11
3.4.3	IQ-Tests mit wenig Aussagekraft	11
4	Fazit: KI kann leicht missverstanden werden	11
	Literaturverzeichnis	12
	Abbildungsverzeichnis	14
	Tabellenverzeichnis	15
	Anhang	16

1 Einleitung

1.1 Forschungsfrage und Zielsetzung

Seit dem 30. November 2023 ist der Chatbot ChatGPT veröffentlicht und frei nutzbar. Nur nach weniger als zwei Monaten hatte ChatGPT bereits 100 Millionen Nutzer [1, S. 15]. Wegen des großen Erfolges begannen schon nach kurzer Zeit weitere Unternehmen wie Google ihre eigenen Chatbots zu entwickeln. Mittlerweile gibt es immer umfangreichere sowie leistungsfähigere KI und deren Entwicklung schreitet unglaublich schnell voran.

Diese Technologie hat einen immer größer werdenden Einfluss auf unsere moderne Gesellschaft. Chatbots sind für viele Menschen bereits Teil ihres Alltags und helfen in der Schule oder im Studium, aber auch im Beruf, beispielsweise dann, wenn es darum geht längere Texte zu verfassen [2, S. 175, S. 185]. Auf der andere Seite gibt es sogar Chatbots die keinen praktischen Nutzen haben und dazu gedacht sind, dass man sich mit ihnen die Zeit vertreibt, indem man z. B. emotionale Gespräche mit ihnen führt. Diese Art der Chatbots sollen als eine Art KI-Freund dienen, sie sind vor allem in Form von Anwendungen aus dem Play Store & App Store bekannt.

Dabei ist es wichtig so eine Technologie richtig einordnen zu können. Jedoch wissen die wenigsten wie Chatbots funktionieren, was dazu führt kann, dass deren Fähigkeiten oft falsch eingeschätzt werden. Darum soll in dieser Arbeit der Frage nachgegangen werden, wie das Bild von KI Chatbots in der Gesellschaft aussieht, bzw. was an diesem Bild problematisch ist.

Mit meiner Arbeit verfolge ich das Ziel aufzuzeigen, welcher Eindruck in den Medien von Chatbots vermittelt wird. Daraus sollen Schlüsse gezogen werden, was die meisten Menschen über diese Form der KI denken. Außerdem will ich in meiner Arbeit gut verständlich erklären, warum dieses Bild der Chatbots bzw. KI oft fehlerhaft ist. Dazu wird näher auf die Funktionsweise dieser Programme eingegangen. Nach dem Lesen der Arbeit soll man ein grundlegendes Verständnis von Chatbots haben und mit Informationen diesbezüglich aufgeschlossener umgehen können.

1.2 Methodik und Aufbau

Im ersten Teil meiner Arbeit werde ich auf einige Beispiele eingehen, die zeigen, dass leicht ein falscher Eindruck von KI, also Chatbots, vermittelt wird. Als erstes werden über KI Chat Apps behandelt. Dabei gehe ich auf Versprechen ein, die von den Entwicklern der Apps gemacht werden. Außerdem will ich durch Bewertungen zeigen, wie die Chatbots in solchen Programme auf die Nutzer wirken. Danach will ich anhand einiger populärwissenschaftlichen Artikel zeigen, dass beispielsweise die Funktionsweise von KI des Öfteren falsch beschrieben wird oder dass es bei solchen Quellen leicht zu Missverständnissen kommen kann. Zuletzt gehen ich noch konkreter auf das Thema Chatbots und IQ-Tests ein. Hier erläutere ich erst ein mal, welche Bedeutung IQ-Tests im Bezug zu Chatbots haben und wozu deren Ergebnisse gerne verwendet werden.

Der zweite Teil der Arbeit behandelt zunächst die Funktionsweise von Chatbots. Erst ein mal müssen ein paar wichtige Begriffe geklärt werden und danach wird auf das LLM und den GPT eingegangen. Die grundlegende Funktionsweise von KI Chatbots wird hierbei leichtverständlich vermittelt. Mit diesem neuen Wissen wird dann auf das Thema AGI behandelt. Es soll dann anhand einiger Beispiele gezeigt werden, warum wir zum aktuellen Zeitpunkt noch keine AGI entwickelt haben. Das Thema AGI passt hier sehr gut, denn wie im ersten Teil der Arbeit gezeigt wird, passiert es schnell, dass aktuelle KI falsch dargestellt wird, nämlich gerne als AGI. Mit dem neuen Wissen sollen dann die Missverständnisse und Problematiken aus dem ersten Teil der Arbeit aufgeklärt werden. Zunächst soll verdeutlicht werden, dass KI noch nicht intelligent ist und auch keine menschlichen Gefühle oder Emotionen hat. Danach sollen Missverständnisse aus den vorher aufgezeigten Artikeln aufgeklärt werden. Zuletzt wird die Problematik mit IQ-Tests, die von Chatbots bearbeitet wurden, diskutiert.

2 Chatbots in der Gesellschaft

2.1 Chatbots mit Gefühlen

Eine kurze Erklärung vorab: Die meisten kennen vermutlich bekannte Chatbots wie ChatGPT, die dazu entwickelt wurden uns bei diversen Aufgaben zu unterstützen. Im Gegensatz dazu steht Character.AI, eines der Unternehmen, die „persönliche“ AEI¹ anbieten [3]. Das bedeutet, dass die KI dazu entworfen wurde, soziale Gespräche zu führen. Man kann zwischen verschiedenen Chatbots, die meist fiktive Charaktere repräsentieren, wählen. Sie sind darauf ausgelegt möglichst emotional und menschlich zu reagieren und oft stehen dabei sehr persönliche Themen im Vordergrund z. B. Einsamkeit, Freundschaft, Empathie oder sogar Liebesbeziehungen.

Das dabei eine emotionale Verbindung zu Chatbots entstehen kann ist kein Wunder und dafür reicht ein Chatbot aus, der nur begrenzte soziale Fähigkeiten hat [4]. KI ist für viele Menschen also nicht nur eine praktische Anwendung, die im Alltag hilfreich ist, sondern hat für einige Personen auch emotionalen Mehrwert. Ob und wieso das kritisch gesehen werden muss wird 3.4.1 diskutiert.

Character.AI ist nur einer von mehreren Unternehmen, die persönliche AEI anbieten, und das mit Erfolg. Seit September 2022 ist Character.AI frei verfügbar und hat zum jetzigen Zeitpunkt August 2024 bereits über 10 Millionen Downloads bei Google Play. Auf Grund der schnellen Weiterentwicklung von KI und des global stärker werdenden Gefühls von Einsamkeit ist demnach zu erwarten, dass die soziale Bedeutung von Chatbots zunehmen wird [4].

¹Artificial Emotional Intelligence

2.2 Mediale Berichterstattung

Die Medien spielen bei der Berichterstattung über KI eine entscheidende Rolle. Sie tragen nicht nur dazu bei, die Öffentlichkeit über aktuelle Entwicklungen und Innovationen zu informieren, sondern sind auch verantwortlich dafür, komplexe technische Zusammenhänge verständlich zu vermitteln. Doch genau hier kann es auch zu Problemen kommen, wodurch eine falsche Auffassung von KI Chatbots entsteht.

Ein Beispiel hierfür ist das Demonstrationsvideo von Google zu ihrem Chatbot Gemini [5]. Das Video zeigt eine Person, die mit dem Chatbot Gemini live interagiert und ihm Fragen stellt. Dabei zeichnet die Person beispielsweise ein Bild von einer Ente und stellt dem Chatbot immer wieder Fragen dazu. Später im Video deutet die Person durch handbewegungen an, dass sie mit dem Chatbot das Spiel „Schere Stein Papier“ spielen will. Der Chatbot antwortet, dass er die Herausforderung annimmt. Das Video betont damit vor allem die Interaktivität zwischen Gemini und der anwesenden Person. Außerdem zeigt das Video einen autonomen Chatbot, der aus einem bestimmten Context Schlüsse zieht, welche Antwort gefragt ist.

In vielen Artikeln zu Chatbots wird deren Funktionsweise und Training erklärt. Maria Gramsch erklärt in ihrem Artikel wie ChatGPT funktioniert [6]. Sie fasst sich hierbei kurz und erklärt unter anderem auch das Training von Chatbots. In dem Artikel ist zwei mal zu lesen, dass das Training von ChatGPT stets unvollendet ist und dass der Chatbot aus Konversationen mit den Nutzern immer weiter dazu lernt. Dadurch wird der Eindruck vermittelt, dass ChatGPT wie ein Mensch aus seinen Erfahrungen lernt.

Was genau an den genannten Beispielen für die Berichterstattung falsch ist wird in 3.4.2 verdeutlicht. Beide genannten Fälle sorgen dafür, dass KI besser dargestellt wird, als sie eigentlich ist. Dafür gibt es verschiedene Hintergründe, wie z. B. Marketing im Fall von Gemini oder mangelndes Wissen bezogen auf Chatbots. Problematisch ist, dass vor allem für die Personen, die nur wenig Verständnis von KI besitzen durch solche Informationsquellen leicht einen falschen Eindruck davon bekommen können, zu was Chatbots fähig sind.

2.3 Chatbots und IQ-Tests

Mit der steigenden Zahl an KI Modellen, wird KI-Benchmarking² immer wichtiger. Außerdem stellt sich die Frage, wie intelligent ein Chatbot verglichen mit einem Menschen ist. IQ-Tests bieten eine Möglichkeit das zu tun, sie sind dazu entworfen Intelligenz so gut wie möglich zu messen und das funktioniert auch bei KI. Jedoch ist es wichtig die Ergebnisse, welche KI in IQ-Tests erzielt, kritisch zu sehen und richtig zu deuten.

Eka Roivainen schreibt in ihrem Artikel, wie er mit dem damals neu veröffentlichten ChatGPT-4 einen IQ-Test durchgeführt hat [7]. Sie schreibt zu Beginn des Artikels, dass sie sich dafür interessiere, wie schlau ChatGPT im Vergleich zu menschlichen Standards ist. Im verbalen Teil des verwendeten WAIS-III Tests erklärt sie, der Chatbot hat einen Wert von 155 erzielt und ist damit besser als 99,9% der menschliche Teilnehmer. Der Artikel wurde mehreren anderen Artikeln zitiert, wie beispielsweise von Sandra Blütermann in ihrem Artikel [8]. In einem Interview von Sky News Australia erklärt Mo Gawdat ein ehemaliger Chief Business Officer von Google ebenfalls, dass ChatGPT einen IQ von 155 habe und damit fast so intelligent wie Albert Einstein sei [9].

Betrachtet man das Ergebnis, was Roivainen mit ChatGPT-4 in ihrem IQ-Test erzielt hat könnte man meinen, dass KI dem Menschen bereits überlegen ist. Mo Gawdat sagt das sogar schon fast wörtlich in seinem Interview: „Der IQ von ChatGPT-4 wird auf 155 geschätzt, das ist viel schlauer als der durchschnittliche Mensch“ [9]. In 3.4.3 werden die Ergebnisse, von ChatGPT-4 im WAIS-III Test genau betrachtet. Darüber hinaus wird die Aussage von Mo Gawdat über die Intelligenz von ChatGPT-4, diskutiert und widerlegt.

²Vergleich von KI-Systemen

3 Funktionsweise von Chatbots

3.1 Begriffe und Grundlagen

Um das nächste Kapitel verstehen zu können, müssen zunächst einige wichtige Begriffe und Sachverhalte erklärt werden. Natürlich kann im Rahmen dieser Arbeit nicht eine komplette Erklärung rund um LLM und GPT bis ins Detail gewährleistet werden, darum beziehe ich mir hier auf das Nötigste, um die folgenden Teile der Arbeit gut verstehen zu können:

- **Neuronale Netze:** Man kann ein Neuronales Netz als eine Mathematische Funktion sehen, die eine Anzahl n an Werten annimmt und eine Anzahl m an Werten ausgibt. Diese Werte kann man auch als Eingabe- und Ausgabevektor sehen:

$$f(\vec{v}) = \vec{w}, \quad \text{wobei } \vec{v} \in \mathbb{R}^n \text{ und } \vec{w} \in \mathbb{R}^m \quad (1)$$

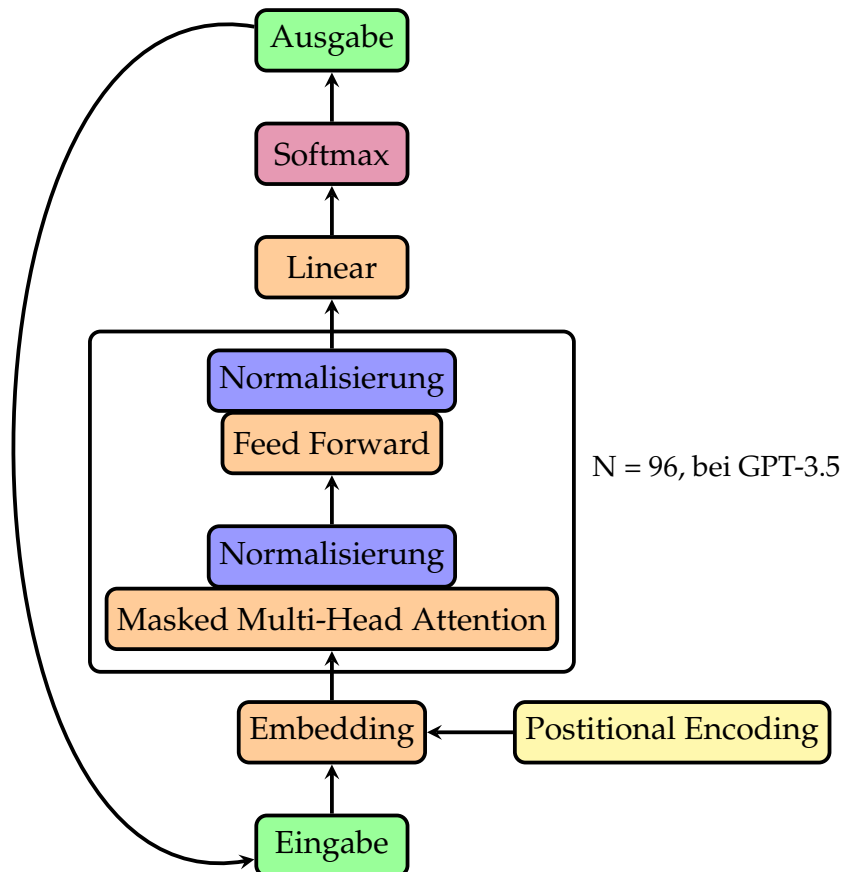
Die Funktion f hat zahlreiche Parameter (oft mehrere Milliarden), die man genau so anpassen will, dass aus unterschiedlichen Eingabevektoren die gewünschten Ausgabevektoren werden und genau hier kommt das Training ins Spiel. Durch Training ist es möglich die Zahlreichen Parameter des Netzwerks so einzustellen, dass sich die Werte der Ausgabevektoren möglichst gut an die gewünschten Werte annähern [10, S. 4f].

- **NLP:** Natural Language Processing ist ein Feld der KI, dass darauf spezialisiert ist natürliche Sprache zu analysieren, transformieren oder zu generieren.
- **LLM:** Ein Large Language Model ist, wie der Name schon sagt ein großes KI Model unserer Sprache. LLMs sind auf NLP spezialisiert und funktionieren unter der Verwendung von Neuronalen Netzen.
- **Transformer:** Die Transformerarchitektur ist ein neuronales Netzwerk-Modell, das erstmals im bekannten Forschungsartikel „Attention is all You Need“ von Forschern bei Google vorgestellt wurde [11]. Transformer in Verbindung mit Chatbots werden in 3.1 genauer erklärt.

- **GPT:** Ein Generative Pre-Trained Transformer ist ein LLM das sich die Transformerarchitektur zu Nutze macht um ein deutlich besseres Ergebnis bei der Sprachverarbeitung zu erzielen. Diese GPTs sind der Grund für den Fortschritt im Bereich NLP und sie werden derzeit in fast allen bekannten Chatbot Modellen, wie ChatGPT und Gemini verwendet [12, S. 2], [13, S. 1].
- **Token:** Wenn ein Chatbot eine Nachricht verarbeitet, dann wird diese zunächst in mehrere kleinere Tokens zerlegt. Tokens helfen bei der Sprachverarbeitung, indem sie Textblöcke in einzelne Zeichenfolgen zerlegen. Zur vereinfachung kann man sich auch vorstellen, dass ein Token ein Wort repräsentiert.

3.2 Transformer und Chatbots

Im Folgenden wird der schematische Aufbau eines Transformers erklärt. Wie schon in 3.1 erwähnt wurde ist ein Transformer ein neuronales Netzwerk-Modell. Das heißt wie in 1 beschrieben nimmt dieser Transformer einen Vektor an rationalen Zahlen als Eingabe und gibt ebenso einen Vektor aus.



Die erste Aufgabe eines Chatbots ist es also Text zunächst in einzelne Tokens aufzuteilen und dann diesen Tokens einen individuellen Vektor zuzuordnen. Das wird auch als Embedding bezeichnet. Durch das Umwandeln von Tokens in einen Vektor kann das Netzwerk Text mathematisch verarbeiten. Weil die einzelnen Tokens nicht der Reihe nach sondern unabhängig von einander und parallel zueinander verarbeitet werden, muss deren Position im Satz mit in den Eingabevektor encodiert werden [11, S. 2f]. Das liegt daran, dass bei vielen Wörtern die Position im Satz eine entscheidende Rolle dafür spielt, was der Satz aussagt:

Der Mann jagt den Hund. \neq Der Hund jagt den Mann.

Die grundlegende Funktionsweise besteht darin, dass der Transformer diese Tokens in Vektoren umwandelt, verarbeitet und dann als Ausgabe errechnet, welches Token mit welcher Wahrscheinlichkeit als nächstes folgt. Natürlich wäre ein Chatbot der nur ein Token als Antwort geben könnte ziemlich nutzlos. Das Problem ist aber einfach zu lösen, indem man dem Transformer die Tokens der vorherigen Eingabe und dazu die neu generierten Tokens gibt. Das wird auch als Selbstregression bezeichnet und bedeutet, dass der Transformer seine Ausgabe wieder als Eingabe verwendet [11, S. 2].

Nach dem Embedding folgt die eine Masked Multi-Head Attention Schicht³. Im fall eines generativen Transformers wird sich hier der Self-Attention Mechanismus zunutze gemacht [14, S. 2f]. Dabei wird zunächst errechnet wie stark der Zusammenhang zwischen zwei Tokens ist [11, S. 4]. Hat man zu Beispiel die Frage, *Wie nennt man ein großes Haus?* als Eingabe, könnten die errechneten Zusammenhänge für das Wort Haus so aussehen:

Zusammenhang mit Haus:	
<i>Wie</i>	0,046
<i>nennt</i>	0,08
<i>man</i>	0,023
<i>ein</i>	0,12
<i>großes</i>	0,7
<i>haus</i>	0,031
<i>?</i>	0,001

Weil sich das Adjektiv *großes* auf das Nomen *Haus* bezieht erkennt ein trainierter Transformer hier einen starken Zusammenhang. Durch diese errechneten Zusammenhänge zwischen den Tokens kann das Netzwerk abwägen, welches Token auf welches andere Token wie viel einfluss nehmen soll.

Nach der Multi-Head Attention folgt eine Normalisierung, welche dabei hilft die Vektoren zu verarbeiten. Normalisierungen sind zwar Teil der Transformerarchitektur aber nicht entscheidend für ein schematisches Verständnis und werden hier nicht genauer beschrieben[11, S.3].

³Teil eines neuronalen Netzwerks

Der zweite bedeutende Teil eines Transformers folgt direkt auf den attention Block und nennt sich Feed Forward Block. Im Feed Forward Block befinden sich c.a. zwei drittel der Parameter [15]. Parameter bestimmen bekanntlicher Weise die Ausgabewerte einer Funktion oder eines neuronalen Netzwerks. Im Fall eines Chatbots bestimmen diese Parameter, wie auf einen Satz den der Transformer als Eingabe bekommt geantwortet wird. Bei diesen Antworten reicht es meistens aber nicht aus, einfach nur Zusammenhänge zwischen Wörtern zu erkennen, wie oben erklärt, sondern man benötigt Hintergrundwissen zum Thema für eine konstruktive Antwort.

Im obigen Satz muss man zum Beispiel nicht nur verstehen, dass es um ein *großes Haus* geht sondern auch wissen welche alternativen Wörter es dafür gibt um die Frage zu beantworten. Dafür braucht man den Feed Forward Block, er kann als Gedächtnis des Transformers gesehen werden. Da sich hier ein Großteil der Parameter befindet ist hier auch ein Großteil des Wissens bzw. der Daten, die ein neuronales Netzwerk gelernt hat gespeichert.

Im Attention Block werden also Zusammenhänge zwischen den Tokens erkannt und dann Informationen zwischen den Tokens übertragen und im Feed Forward Block wird dann noch auf die eingehenden Tokens mit bestimmten Wissen, welches das Netzwerk besitzt reagiert.

Wichtig ist, dass dieser Ablauf von Attention und Feed Forward mehrmals wiederholt wird, damit auch längere komplexe Texte mit komplizierten Zusammenhängen sinnvoll verarbeitet werden können.

Am Ende des Transformer befinden sich noch eine Lineare und eine Softmax Schicht. Sie sind dazu da die Ausgabe des Transformers in einen Vektor zu überführen, der allen bekannten Tokens eine Wahrscheinlichkeit zuweist. Darunter wählt man dann meist das Token mit der höchsten Wahrscheinlichkeit aus und fügt es dem Text hinzu. Diesen Prozess wiederholt man dann so lange, bis der Transformer ein EOS-Token⁴ generiert und damit den Abschluss seiner Antwort signalisiert.

⁴End-of-Sequence-Token

3.3 Warum ist das noch nicht AGI?

3.4 Was folgt daraus?

3.4.1 Immer noch nur Algorithmen

3.4.2 Das falsche Bild in den Medien

3.4.3 IQ-Tests mit wenig Aussagekraft

4 Fazit: KI kann leicht missverstanden werden

Literaturverzeichnis

- [1] Manfred Spitzer. *Künstliche Intelligenz, Dem Menschen Überlegen - wie KI uns rettet und bedroht*. Bd. 1. München: Droemer Verlag, Nov. 2023. ISBN: "978-3-426-44850-2".
- [2] Dr. A. Shaji George, A. S. Hovan George und A. S. Gabrio Martin. „ChatGPT and the Future of Work: A Comprehensive Analysis of AI’s Impact on Jobs and Employment“. In: *Partners Universal International Innovation Journal* 1.3 (Juni 2023), S. 154–186. DOI: 10.5281/zenodo.8076921. URL: <https://www.puiij.com/index.php/research/article/view/57>.
- [3] Noam Shazeer und Daniel De Freitas. *Character.AI*. Auch als mobile App verfügbar. Aug. 2024. URL: <https://character.ai/>.
- [4] Angelina Chen u. a. „Feels Like Empathy: How “Emotional” AI Challenges Human Essence“. In: *Australasian Conference on Information Systems*. 2024.
- [5] Google. *The capabilities of multimodal AI | Gemini Demo*. Youtube. 2023. URL: <https://www.youtube.com/watch?v=UIZAiXYceBI>.
- [6] Maria Gramsch. „Künstliche Intelligenz: Wie funktioniert eigentlich ChatGPT?“. In: *BASIC thinking* (Mai 2023).
- [7] Eka Roivainen. „I Gave ChatGPT an IQ Test. Here’s What I Discovered“. In: *Scientific American* (März 2023).
- [8] Sandra Blütermann. „ChatGPT besteht IQ-Test“. In: *Computer Bild* (März 2023).
- [9] Sky News Australia. *Chat-GPT ‘much smarter’ than the average human: Mo Gawdat*. Youtube. 2023. URL: <https://www.youtube.com/watch?v=BKf7qrcN0X8>.
- [10] Charu C. Aggarwal. *Neural Networks and Deep Learning*. Second Edition. Springer, 2018. URL: https://scholar.google.de/citations?view_op=view_citation&hl=de&user=x_wsduUAAAAJ&citation_for_view=x_wsduUAAAAJ:1N037PikXpQC.

- [11] Ashish Vaswani. „Attention is all you need“. In: *Advances in Neural Information Processing Systems* (2017). URL: <https://user.phil.hhu.de/~cwurm/wp-content/uploads/2020/01/7181-attention-is-all-you-need.pdf>.
- [12] Google Gemini Team. *Gemini 1.5: Unlocking multimodal understanding across millions of tokens of context*. Techn. Ber. Version 4. arXiv Identifier: 2403.05530v4. Aug. 2024. URL: <https://arxiv.org/abs/2403.05530v4>.
- [13] OpenAI. *GPT-4 Technical Report*. Techn. Ber. Version 6. arXiv Identifier: 2303.08774v6. März 2024. URL: <https://arxiv.org/abs/2303.08774v6>.
- [14] Richard E. Turner. „An Introduction to Transformers“. Version 5. In: (Feb. 2024). arXiv Identifier: 2304.10557v5. URL: <https://arxiv.org/abs/2304.10557v5>.
- [15] Mor Geva. „Transformer Feed-Forward Layers Are Key-Value Memories“. Version 2. In: (Sep. 2021). arXiv Identifier: 2012.14913v2. URL: <https://arxiv.org/abs/2012.14913v2>.

Abbildungsverzeichnis

Tabellenverzeichnis

Anhang