Hochschule Ruhr West

Energie- und Wassermanagement, Campus Mülheim a.d. Ruhr

Bachelorarbeit

„Geschäftsfeld-Entwicklung mit Hilfe von LLMs – eine Fallstudie aus dem petro-chemischen Commodity-Handel“

Eingereicht bei:

Prof. Dr. Michael Römmich

Prof. Dr. Simone Roth

Abgabedatum: 16.12.2024

Luis Follmer

Matrikelnummer: 10015677

Paradijslaan 123A,

3034 SK Rotterdam, Niederlande

Energie- und Wassermanagement

9. Semester

Abgabedatum: 16.12.2024

Inhaltsverzeichnis

Abbildungsverzeichnis ..... **III**
Abkürzungsverzeichnis ..... **IV**
Symbolverzeichnis ..... **V**
1. Einleitung **1**
1.1 Thematischer Einstieg 1
1.2 Zielstellungen 2
1.3 Forschungsfrage und Methode 3
2. Einsteigende Definitionen **4**
3. Hoffnungen und Rückschläge: Geschichte der künstlichen Intelligenz **5**
3.1 Geburtsstunden in den 1940er Jahren 5
3.2 The hardware is the limit 6
3.3 Von Booms und Busts 8
4. Neuronale Netze und technische Funktionsweisen moderner KI **11**
4.1 Funktionsweisen einzelner Neuronen 11
4.2 Allein stark, gemeinsam stärker: Neuronale Netze 14
4.3 Selbstlernende Systeme 17
4.4 Mangelndes Verständnis und große Potenziale 19
5. Wie Maschinen Sprache meistern: Large Language Models **21**
6. Aktuell verfügbare Sprach-KI Anwendungen **23**
6.1 ChatGPT 23
6.2 ClaudeAI 24
6.3 Gemini 25
7. LLMs in der Arbeitswelt: Anpassung als Überlebensstrategie **26**
7.1 Kommende Veränderungen und wie auf diese reagieren 26
7.2 Effekte auf menschliche Arbeit 31
8. Vertrauen ist gut, Kontrolle ist besser: Richtiger Umgang mit Sprach-KI **34**
8.1 Schwächen moderner LLMs 34
8.2 Umgang mit Fehleranfälligkeit von Sprach-KI 35
9. Prompt-engineering: LLMs effektiver und effizienter nutzen **39**
9.1 Grundsätze des Prompt-engineering 39
9.2 Prompt-engineering Techniken 41
10. Fallstudie zu KI-unterstütztem Business Development **44**
10.1 Unternehmensbeschreibung 44
10.2 Versuchsaufbau 45
10.3 Erster Versuch 47
10.4 Zweiter Versuch 47
11. Fazit **53**
Literaturverzeichnis **56**

Abbildungsverzeichnis

Abbildung 1: IBM-Stretch 6
Abbildung 2: Grafisch dargestelltes Perceptron 11
Abbildung 3: Ergebnisfunktionen von Perceptrons (links) und Sigma-Neuronen (rechts) 13
Abbildung 4: Beispielhafte Struktur eines neuronalen Netzes 14
Abbildung 5: Beispielhafte Zahlenfolge 15
Abbildung 6: Strukturdarstellung eines neuronalen Netzes zur Identifikation handschriftlicher Ziffern 16
Abbildung 7: Arbeitsschritte nN 16
Abbildung 8: Ludditen zerstören Maschinen zur Textilverarbeitung 27
Abbildung 9: Auswirkungen von LLMs auf verschiedene Tätigkeiten 28
Abbildung 10: Berufe mit dem höchsten Potenzial für augmentation 29
Abbildung 11: Durch LLMs wenig betroffene Berufe 30
Abbildung 12: Berufe mit dem höchsten Potenzial für Automatisierung 30
Abbildung 13: Benötigte Zeit zur Lösung der Aufgabenstellungen 31
Abbildung 14: Durchschnittsnote beider Aufgabenstellungen 31
Abbildung 15: Notenverteilung der zweiten Aufgabenstellung 32
Abbildung 16: Gemessene Befriedigung durch erledigte Arbeit 32
Abbildung 17 35
Abbildung 18: Grafische Darstellung zero-shot-Prompting 42
Abbildung 19: Grafische Darstellung few-shot-Prompting 42
Abbildung 20: Grafische Darstellung chain-of-thought-Prompting 43
Abbildung 21: Ergebnisse des ersten Prompts (Blasengröße repräsentiert Anzahl maximal erkennbarer Produkte) 49
Abbildung 22: Ergebnisse des zweiten Prompts (Blasengröße repräsentiert Anzahl maximal erkennbarer Produkte) 51

Abkürzungsverzeichnis

KI	künstliche Intelligenz
AI	artificial intelligence
nN	neuronales Netz(werk)
LLM	Large Language Model

Symbolverzeichnis

x	Input
w	Gewichtung
b	Bias
C	gesamte Kosten
n	Anzahl an Datenpunkten
a	gewünschtes Ergebnis
y	tatsächliches Ergebnis
σ	Sigma
e	eulersche Zahl

1. Einleitung

1.1 Thematischer Einstieg

“OpenAI value surges to $157bn in funding deal”[1]
“AI Can Only Do 5% of Jobs, Says MIT Economist Who Fears Crash”[2]
“‘In awe’: scientists impressed by latest ChatGPT model o1”[3]

Dies waren Schlagzeilen, die zwischen dem 01. und 05. Oktober 2024 in bekannten Zeitungen und Magazinen zu lesen waren. Sie zeigen, dass der Begriff der künstlichen Intelligenz weit in die Mitte des Bewusstseins unserer Gesellschaften gerückt ist, sei es durch Anwendungen wie ChatGPT, die Presse oder die Popkultur. Es wird in Deutschland nur noch wenige Menschen geben, die den Begriff der künstlichen Intelligenz (KI) nicht kennen.
Allerdings zeigen die drei mit den Überschriften verbundenen Artikel auch die Diversität und Komplexität der Thematik. Wie passen Meldungen über die milliardenschwere Wertsteigerung eines Softwareunternehmens, erzielte wissenschaftliche Meilensteine und die Warnung eines respektierten Ökonomen vor einem möglichen Crash zusammen? Wie umgehen mit der neuen Technologie und den Veränderungen, die diese insbesondere für unser Arbeitsleben bedeuten wird? Muss ich mir Sorgen um meinen Arbeitsplatz machen oder handelt es sich bei KI nur um einen Hype, der bald wieder verschwinden wird? Dies ist nur ein Bruchteil der Fragen, die sich viele Menschen zu künstlicher Intelligenz stellen und bei deren Beantwortung die vorliegende Arbeit unterstützen soll.

[1] BBC (2024), o.S.
[2] Bloomberg (2024), o.S.
[3] Nature (2024), o.S.

1.2 Zielstellungen

Das Gesamtziel der vorliegenden Abschlussarbeit ist es, einen als Orientierungshilfe dienenden Leitfaden für die Auseinandersetzung mit und den Gebrauch von künstlicher Intelligenz zu schaffen. Der enorme Umfang der Thematik macht jedoch eine inhaltliche Fokussierung erforderlich, die im Rahmen dieser Arbeit auf generativer Sprach-KI liegt und durch die folgenden sieben Unterziele genauer definiert wird:

1) Erläuterung und Kontextualisierung von wichtigen Schlüsselbegriffen. Durch diese soll dem Leser das für eine weitere Auseinandersetzung mit der Thematik essenzielle Vokabular an die Hand gegeben werden.
2) Analyse der Geschichte von künstlicher Intelligenz. Durch Verständnis vergangener, soll die Interpretation zukünftiger Entwicklungen im Feld der KI ermöglicht und erleichtert werden.
3) Schaffung eines grundlegenden Verständnisses der technischen Funktionsweisen von künstlicher Intelligenz und den zugrundeliegenden neuronalen Netzen (nN).
4) Vorstellung von zum Zeitpunkt des Verfassens verfügbarer Sprach-KI Modelle in Bezug auf deren Funktionen, Schwächen und Stärken.
5) Beantwortung der Fragen, wie sich generative Sprach-KI auf unsere Arbeitswelt auswirken wird und wie die richtige Reaktion auf die kommenden Veränderungen, die eigene Relevanz auf dem Arbeitsmarkt sichern kann.
6) Das Erarbeiten einer gesunden Einstellung im Umgang mit Sprach-KI, die auf einer Balance zwischen Skepsis und Vertrauen sowie dem Wissen über die Fähigkeiten und Schwächen der Technologie beruht.
7) Erläuterung und Illustrierung von Techniken zur Effektivitäts- und Effizienzsteigerung im Umgang mit künstlicher Intelligenz.

1.3 Forschungsfrage und Methode

Die Forschungsfrage für die vorliegende Arbeit lautet:
„Lässt sich generative Sprach-KI auf dem heutigen Stand der Technik in unterstützenden Funktionen in die Geschäftsfeldentwicklungsprozesse eines modernen Handelsunternehmens sinnvoll einbinden?“

Die vorliegende Arbeit stützt sich zur Beantwortung der oben genannten Forschungsfrage und Erreichung der formulierten Ziele auf die folgenden wissenschaftlichen Methoden:

1) Literaturarbeit: Weite Teile der Arbeit basieren auf bereits vorhandener Literatur. Die Analyse, Reduzierung und Verknüpfung der maßgeblichen Inhalte aus verschiedenen Werken, erzeugt einen Mehrwert für den Leser. Komplexe Themen werden dabei durch eigene veranschaulichende Beispiele und Erklärungen ergänzt. Aus Gründen der Aktualität und hohen Dynamik der Thematik konnte nicht ausschließlich auf peer-reviewed Quellen zurückgegriffen werden; die geforderten Seriositätsmaßstäbe wurden jedoch trotzdem an alle verwendeten Veröffentlichungen angelegt.
2) Experimente: Im Zuge der Arbeit werden an verschiedenen Stellen Konzepte anhand des Erstellungsprozesses des vorliegenden Textes erläutert. Da hier bereits KI zum Einsatz kam, kann dieser selbst als ein abstraktes Experiment angesehen werden.
3) Fallstudie: Teile der zuvor auf Basis von Literatur erarbeiteten Inhalte fließen im vorletzten Kapitel in eine Fallstudie ein. Diese verdeutlicht beispielhaft die zuvor bereits erklärten Ideen durch die praktische und realitätsbezogene Anwendung.

2. Einsteigende Definitionen

Was ist eigentlich künstliche Intelligenz oder artificial intelligence (AI)? Als erstes gilt es klarzustellen, dass beide Begriffe dieselbe Bedeutung haben und daher im Zuge dieser Arbeit synonym verwendet werden. Um die gestellte Frage einsteigend zu beantworten und ein Basisverständnis zu entwickeln, ist es hilfreich mit der Betrachtung verschiedener Definitionen zu beginnen.

Die amerikanische Consulting-Firma McKinsey & Company definiert AI wie folgt: "Artificial intelligence is a machine's ability to perform some cognitive functions we usually associate with human minds."[4]

Eine passende Ergänzung lässt sich aus der Enzyklopädie Britannica ableiten: "(…) the ability of a digital computer or computer-controlled robot to perform tasks commonly associated with intelligent beings."[5] Denn es sind stets Computer, die mit ihrer Rechenleistung die Existenzgrundlage für künstliche Intelligenz schaffen. Ohne diese wäre KI nicht denkbar, ähnlich wie ohne das Vorhandensein unserer Gehirne menschliche Intelligenz wortwörtlich nicht vorstellbar wäre.

Die International Organization for Standardization führt mit: "(…)the engineered system that generates outputs such as content, forecasts, recommendations or decisions for a given set of human-defined objectives."[6], Beispiele für konkrete Aufgaben an, die mit Hilfe von KI durch Computer bearbeitet werden sollen.

Diese Informationen zusammenfassend, lässt sich AI wie folgt definieren: Künstliche Intelligenz ist die Fähigkeit eines Computers, Aufgaben zu erledigen, die jene kognitiven Kapazitäten und Fähigkeiten erfordern, die normalerweise nur mit Menschen in Verbindung gebracht werden und zu denen das Erstellen von Content, Vorhersagen oder Handlungsempfehlungen gehört.

[4] McKinsey (2023), o.S.
[5] Copeland (2024), o.S.
[6] International Organization for Standardization (o.J.), o.S.

3. Hoffnungen und Rückschläge: Geschichte der künstlichen Intelligenz

3.1 Geburtsstunden in den 1940er Jahren

Bewertungen von zukünftigen Entwicklungen im Feld der KI und AI-Economy sind nur möglich, wenn die wichtigsten Zusammenhänge aus der Geschichte der Technologie bekannt sind. Wie bereits erläutert, sind Computer die Maschinen, ohne die KI nicht existieren könnte. Daher verwundert es nicht, dass die Entwicklungsgeschichte von künstlicher Intelligenz, stets eng mit Innovationen im Feld der Computerwissenschaften verbunden war und bis heute ist. So liegen zwischen den Erfindungen der ersten Digitalcomputer von Konrad Zuse, John Presper Eckert und John William Mauchly Anfang der 1940er Jahre und den Veröffentlichungen grundlegender Werke und Forschungen zu KI nur wenige Jahre.[7]

Laut Russel/Norvig (2021), S. 35ff. wurden diese wissenschaftlichen Grundsteine unter anderem im Jahr 1943 von den amerikanischen Professoren Warren McCulloch und Walter Pitts gelegt. Ihnen gelang es, das erste Modell eines künstlichen neuronalen Netzwerks zu kreieren, das auf den damaligen Kenntnissen der Computer- und Neurowissenschaft beruhte. Damit schufen sie unbewusst die Grundlage für die heute modernsten und fähigsten AI-Anwendungen, die zusammen mit den genannten nN später noch genauer beleuchtet werden. Sie zeigten zudem, dass alle berechenbaren mathematischen Formeln von einem derartigen System, zumindest in der Theorie, gelöst werden konnten. Inspiriert wurden sie dabei unter anderem von den Arbeiten Alan Turings, einem weiteren wichtigen Vordenker und dem späteren Erfinder des sog. Turing-Tests. Dieser beschreibt die Fähigkeit eines Computers, Fragen eines menschlichen Testers so gut zu beantworten, dass es für diesen nicht mehr möglich ist zu erkennen, ob sie von einem Computer oder einem anderen Menschen stammen.[8] Der Test galt als eine lange für unerreichbar gehaltene Hürde, die die neuesten KI-Anwendungen wie ChatGPT nach der Ansicht mancher Autoren heute bereits genommen haben.[9] Ebenso war es Turing, der als erstes den Vorschlag machte, dass es einfacher sei, einen lernenden Algorithmus zu schaffen, als einem Computer jegliches Wissen händisch zu vermitteln. Eine ebenfalls recht simple und logische Idee, die jedoch von essenzieller Wichtigkeit für spätere Entwicklungen des Feldes war und noch immer ist.

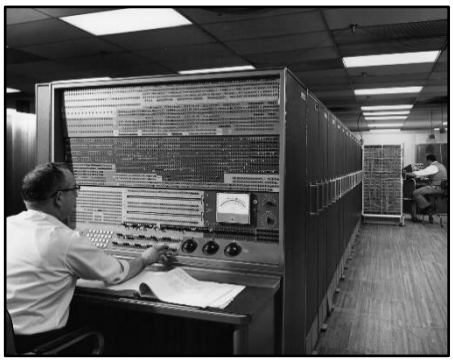
[7] Vgl. BMU Verlag (o.J.), o.S.
[8] Vgl. Oppy/Dowe, (2021), o.S.
[9] Vgl. Jackson (2024), o.S.

Das erste dann in der Praxis funktionale neuronale Netzwerk wurde 1950 von Marvin Minsky und Dean Edmonds geschaffen und enthielt 40 Neuronen. Zur Einordnung, neueste Modelle, wie das genannte ChatGPT, verfügen über bis zu ca. 175.000.000.000 Neuronen.[10] An dieser Stelle ist es noch vollkommen ausreichend, die Anzahl der Neuronen als die Fähigkeit von KI, komplexe Zusammenhänge zu verarbeiten, zu verstehen. An Größenordnungen wie die vorgenannte war aufgrund technischer Limitierungen in den 1950er Jahren allerdings noch nicht zu denken. Der enorme Größenunterschied verdeutlicht somit eines der Kernprobleme, von dem die KI-Forschung seit ihren Geburtsstunden und letztlich bis ins 21. Jahrhundert immer wieder verlangsamt und aufgehalten wurde; leistungsstarke Computer.

3.2 The hardware is the limit

Rechner zur Mitte des 20. Jahrhunderts hatten nach heutigem Standard minimale Rechenleistungen, waren in der Regel riesige Maschinen und so teuer, dass ihre Nutzung nur einem sehr kleinen Kreis von Menschen vorbehalten war. Hinzu kam die Problematik, dass es schlichtweg nur sehr wenige Computer gab und deren Einsatzzeiten eine hart umkämpfte Ressource zwischen verschiedenen Forschern und Interessengruppen war. Ein gutes Beispiel für den damaligen Stand der Technik ist der für seine Zeit Supercomputer IBM-Stretch der gleichnamigen Firma International Business Machines Corporation. Dieser Rechner, dargestellt in Abb. 1, hatte einen Arbeitsspeicher von 6.000.000 Bit, was umgerechnet ca. 0,75 Megabyte entspricht.[11] Der Kostenpunkt für eine Einheit lag nach Inflationsadjustierung (Umrechnungsdatum: 04.11.2024) bei ca. 83.000.000 USD in heutiger Währung.[12] Um den Unterschied zu aktuellen Computern zu verdeutlichen, muss lediglich ein Blick auf das neueste Flagship-Model des Elektronik-Herstellers Samsung geworfen werden. Das Smartphone S24 Ultra, das passenderweise stark in Verbindung mit Galaxy AI beworben wird, kostet ca. 1.800 USD, bietet dem Nutzer Zugriff auf eine Rechenleistung von zwölf Gigabyte und lässt sich in jeder üblichen Hand- oder Hosentasche transportieren.[13] Zwölf

Abbildung 1: IBM-Stretch



Quelle: IBM (o.J.), o.S.

[10] Vgl. Benzon (2023), o.S.
[11] Vgl. IBM (o.J.), o.S.
[12] Vgl. Hardy (o.J.), o.S.
[13] Vgl. Samsung (2024), o.S.

Gigabyte entsprechen 12.000 Megabyte oder 96.000.000.000 Bit, also der 16.000-fachen Leistung des IBM Stretch, zu einem Preis, der bei gerade einmal 0,0022% der Kosten der Hochtechnologie aus den 1960er Jahren liegt. Im Kern dieser rapiden Kostenreduktion und Leistungssteigerung steht das sog. Moore's Law, das die exponentielle Entwicklung in der Halbleitertechnologie beschreibt. Laut diesem halbieren sich die Stückkosten für eine Einheit Computerrechenleistung im Mittel alle zwölf bis 24 Monate.[14] Dies erlaubt unter anderem den Bau immer leistungsfähigerer Supercomputer, die eine fokussierte Spitzenforschung ermöglichen und somit die Entwicklungen der künstlichen Intelligenz vorantreiben.

Ebenso wichtig ist es, dass immer bessere Computertechnologie für eine breitere Nutzerzahl verfügbar ist, die damit auf zwei entscheidende Arten zum technischen Fortschritt beiträgt. Zum einen erlaubt dieser Zugang mehr und mehr interessierten Menschen selbst mit AI zu experimentieren, das Feld für sich beruflich zu entdecken und es im Rahmen einer Karriere voranzutreiben. Zum anderen setzen Unternehmen wie Microsoft, Alphabet oder Samsung ihre enormen Ressourcen nur dann für die Entwicklung einer Technologie ein, wenn sich ihrer Ansicht nach hieraus ein Mehrwert gewinnen lässt. Die Tatsache, dass ein großer Teil der kaufkraftstarken Weltbevölkerung Zugriff auf Computer hat (in Deutschland über 90% aller Haushalte)[15], macht die Entwicklung von KI-basierten Anwendungen für große Technologiekonzerne demnach erst wirtschaftlich sinnvoll, wenn man sich erneut vor Augen führt, dass KI nur dank Computern existieren kann.

1950 war das Kerngeschäft des heutigen AI-Entwicklungstreibers und Tech-Riesen Samsung allerdings noch der Handel mit Lebensmitteln.[16] Microsoft und Alphabet waren noch nicht gegründet und die KI-Forschung war auf Universitäten und sehr wenige Unternehmen beschränkt und abhängig von staatlicher Unterstützung. Eine AI-Economy, im Vergleich zur heutigen, existierte schlichtweg noch nicht. Das hielt viele Vordenker trotzdem nicht davon ab, fantastische technische Durchbrüche vorherzusagen. Allerdings wurde diesen anfänglichen Illusionen relativ schnell die Realität der äußeren Umstände und technischen Begrenztheiten vor Augen geführt. Der großen Euphorie aus den 1950ern folgte ein erstes Tief in der Entwicklung und damit eine Phase, in der immer weniger Ressourcen in KI-Projekte investiert wurden, weshalb diese letztlich ins Stocken gerieten.

[14] Vgl. Nordhaus (2015), S. 4f.
[15] Vgl. Statista (2024), o.S.
[16] Vgl. Bondarenko (2024), o.S.

3.3 Von Booms und Busts

An dieser Stelle ließen sich nun kleinteilig alle Innovationen in der KI-Forschung, die zwischen den 1960er und den 2020er Jahren gemacht wurden aufzählen und beschreiben, dies würde für die Arbeit jedoch keinen Mehrwert bieten. Stattdessen beschränkt sich der Rest des Kapitels auf die Erläuterung größerer und bedeutenderer Paradigmenwechsel und Meilensteine mit dem Ziel, daraus einen Entwicklungstrend abzuleiten, der Aufschlüsse über die mögliche Zukunft des Feldes geben kann.

Die meisten der frühen KI-Modelle der 50er Jahre basierten auf weak methods. Dabei handelt es sich um Algorithmen und Problemlösungsstrategien, die auf die Lösung einer breiten Spannweite von Aufgabenstellungen ausgelegt sind, da sie über kein einprogrammiertes Spezialwissen verfügen. Sie sind nicht grundsätzlich negativ zu bewerten und durchaus zur Bearbeitung sehr simpler Aufgaben geeignet, bringen jedoch verschiedene entscheidende Schwächen mit sich. Die wohl größte ist der Umgang mit der sog. combinatorial explosion, die die exponentielle Zunahme von Lösungsmöglichkeiten und Variablen mit steigender Komplexität eines Problems beschreibt. Um sich dies zu visualisieren, hilft die Vorstellung eines Wahrscheinlichkeitsbaumes, dessen Astzahl sich mit jedem Schritt verdoppelt. Weak methods haben das Ziel durch einen einfachen Vergleich aller Äste zu einer optimalen Lösung zu kommen, anstelle bestimmte Möglichkeiten von Beginn an strategisch auszuschließen. Sie werden dadurch sehr ineffizient und ungeeignet für realitätsbezogene Anwendungen, die sich oft durch enorme Komplexität, Dynamik und Unsicherheit auszeichnen.

Die Erkenntnis über diese und weitere Probleme führte nach dem ersten KI-Bust zu der Entwicklung von sog. expert systems. Diese waren zwar in der Lage, durch die Integration von problemspezifischem Wissen wesentlich komplexere Probleme als ihre Vorgänger zu lösen, aber nur in einem sehr engen Korridor vorbestimmter Anwendungsfälle. Ein prominentes Beispiel ist das in den späten 1960er Jahren entwickelte expert system DENDRAL, das Chemikern dabei half, molekulare Strukturen organischer Stoffe zu bestimmen. Das System basierte, wie beschrieben, auf spezifischem Wissen, aus dem es passende Lösungsschlüsse ziehen konnte. Allerdings taten sich auch hier neue Hürden auf. Jegliches Wissen musste in der Regel händisch programmiert werden und konnte nicht auf andere Problemstellungen übertragen werden, was die Entwicklung neuer, bzw. die Anpassung alter Systeme sehr zeit- sowie ressourcenintensiv und damit langfristig unwirtschaftlich machte. Trotzdem führten die frühen Erfolge der expert systems nach dem anfänglichen der 50er Jahre zunächst zu einem zweiten AI-Boom, der KI

zum ersten Mal von einer Millionen- zu einer Milliardenindustrie machte. Große Teile dieser kollabierten jedoch gegen Ende der 1980er wieder, als die beschriebenen Schwierigkeiten und Beschränkungen deutlicher und letztlich unübersehbar wurden. Der Kollaps und die darauffolgende Periode ist bis heute unter dem Begriff des großen AI-Winter bekannt und steht für eine weitere Phase der Desillusionierung und drastischen Verlangsamung des technischen Fortschritts.

Für eine zweite Wiederbelebung des Feldes sorgte die erneute Auseinandersetzung mit den bereits genannten neuronalen Netzen. Ein signifikanter Entwicklungsschritt war dabei die Erfindung von backpropagation durch unter anderem Geoffrey Hinton. Dabei handelt es sich um Algorithmen, die es nN erlauben, einzelne Neuronen durch rückwärtige Ergebnisanalyse anzupassen und damit Alan Turings bereits vorgestellte Idee des selbstlernenden Computers zu verwirklichen. Diese und weitere Fortschritte, wie die Erfindung der transformer-Architektur in den 2000er und 2010er Jahren sorgten für einen dritten AI-Boom, der spätestens seit der Veröffentlichung von ChatGPT im Jahre 2022 kaum noch zu bestreiten ist und dieses Kapitel damit in unserer Zeit münden lässt.

Aus dieser zusammengefassten Entwicklungsgeschichte lassen sich zwei zentrale Beobachtungen ableiten. Die erste ist, dass Innovationen im Feld von KI grundsätzlich von der Entwicklung der zugrundeliegenden Computertechnologie abhängen. Dies erschwert Vorhersagen über die Richtung, in die sich die Technologie bewegen wird und sollte immer als eine zusätzlich Variable mit bedacht werden, wenn man Versprechungen über die zukünftigen Fähigkeiten von KI zu bewerten versucht.

Die zweite beschreibt den klaren Trend eines ständigen Boom-Bust-Zyklus. Während eines Booms, der in der Regel durch eine fundamentale Idee und der sich daraus ergebenden neuen technischen Möglichkeiten entsteht, herrscht vornehmend große Euphorie. Es werden viele Ressourcen aus staatlicher und oder privatwirtschaftlicher Hand investiert und Hoffnungen in die Technologie geweckt. Versprechen über die baldigen technischen Möglichkeiten und kurz bevorstehenden Durchbrüche wirken dabei wie Brandbeschleuniger, die die Aufbruchsstimmung nochmals verstärken. Bisher wurden diese Perioden jedoch immer von Bust-Phasen abgelöst, in denen die Illusionen den wirtschaftlichen und oder technischen Realitäten weichen müssen. In diesen Phasen, wie beispielsweise dem großen AI-Winter, gingen Förderung und Forschung zurück, bis eine neue Idee den nächsten Boom einleitete. Aktuell befinden wir uns zweifelsohne in einer solchen Boom-Phase, wodurch noch ein weiterer Aspekt des Trends be-

legt wird. Die Entwicklung von AI ist bisher nie vollständig zum Erliegen gekommen. Die Faszination für die Technologie ist schlichtweg zu groß, um Menschen davon abzubringen, weiter an ihr zu forschen. Eine Fortsetzung des aktuellen Trends ist zwar nicht selbstverständlich, dennoch lassen sich zukünftige Entwicklungen und die anfängliche Warnung vor einem erneuten AI-Winter nun besser einordnen. Sollten sich aktuelle Erwartungen an die Technologie und ihre Fähigkeiten aus welchen Gründen auch immer langfristig nicht erfüllen, ist eine Verlangsamung der Forschung genauso wenig auszuschließen, wie ein darauffolgender erneuter Boom.
Moderne KI, insbesondere Large Language Models (LLMs) wie ChatGPT, mögen mit ihren Fähigkeiten schon fast magisch anmuten und die Zahl der kritischen Stimmen mag zurzeit noch kleiner als die der Hoffnungsträger sein. In den folgenden Kapiteln wird jedoch aufgezeigt, dass AI auch nach dem heutigen Stand der Technik durchaus Schwächen hat und man sich von den enormen gebotenen Möglichkeiten nicht blenden lassen darf.

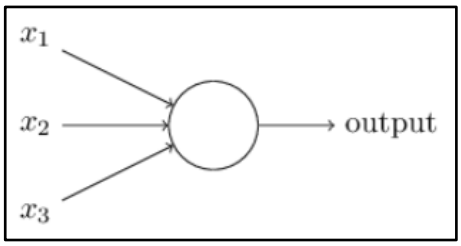
4. Neuronale Netze und technische Funktionsweisen moderner KI

Für den richtigen Umgang mit künstlicher Intelligenz ist es zunächst wichtig, ein grundlegendes Verständnis der unterliegenden Prozesse zu entwickeln, auf denen diese basiert. Da ein zu tiefer Einstieg in die Details allerdings den Rahmen der Arbeit sprengen würde, werden an dieser Stelle nur die wichtigsten Grundzusammenhänge beleuchtet und gleichzeitig viele der in den ersten Abschnitten gefallen Fachbegriffe genauer erklärt. Diese Vereinfachung steht dem allgemeinen Ziel der Arbeit, einen Leitfaden für die Nutzung von Sprach-KI zu entwickeln, allerdings nicht entgegen. Denn wie bei einem Auto, ist ein maximal rudimentäres Verständnis von dessen technischen Funktionsweise nötig, um es verkehrssicher zu bewegen und an ein gewünschtes Ziel zu gelangen.

4.1 Funktionsweisen einzelner Neuronen

Zunächst gilt es eine wichtige Sache klarzustellen. Auch wenn die Fähigkeiten scheinbar denkender Computer faszinierend sein mögen, ist künstliche Intelligenz keine Zauberei. Daher darf man sich beim Versuch diese zu verstehen, nicht von vermeintlicher Komplexität oder kompliziert klingenden Fachbegriffen abschrecken lassen. Moderne KI-Systeme basieren in der Regel auf den bereits erwähnten neuronalen Netzen bzw. Netzwerken, die wiederum schlichtweg mathematische Modelle sind, die für die Lösung bestimmter Aufgaben konstruiert und optimiert wurden. Sie sind nach den Neuronenstrukturen in menschlichen Gehirnen benannt. Das ist naheliegend, wenn man bedenkt, dass nN technisch gesehen digitale Kopien desselben Konzeptes sind. Alle nun folgenden Erklärungen des Kapitels werden sich primär auf Nielsen (2019), o.S. beziehen, das die wichtigsten Zusammenhänge einsteigerfreundlich darlegt. Die Inhalte werden zusätzlich durch eigene Beispiele und Erklärungen ergänzt. Eine breitere Quellenauswahl wurde in diesem Fall bewusst vermieden, da unterschiedliche Erklärungsversuche derselben Sache zu unnötiger Verwirrung führen können.

Abbildung 2: Grafisch dargestelltes Perceptron



Quelle: Nielsen (2019), o.S.

Um nN zu verstehen, empfiehlt es sich, mit den Grundlagen zu beginnen; in diesem Fall damit, was einzelne Neuronen sowie deren Funktionsweisen und Aufgaben sind. Zum besseren Verständnis ist ein sog. Perceptron mit dessen In- und Outputs auf Abb. 2 visualisiert. Es lässt sich

erkennen, dass ein Neuron verschiedene Daten-Inputs aufnehmen und zu einem einzelnen Ergebnis verarbeiten kann. Dahinter versteckt sich die folgende Formel:

$$Output = \begin{cases} 0\ wenn\ w \cdot x + b \leq 0 \\ 1\ wenn\ w \cdot x + b > 0 \end{cases} \quad (1)$$

Ein Perceptron ist also ein Neuron, das in der Lage ist, auf Basis von verschiedenen Inputs (vereinfacht ausgedrückt) eine Ja-Nein-Entscheidung zu treffen. Im ersten Teil der Formel findet sich *w*x*, wobei *w* eine beliebig gewählte Gewichtung darstellt, während *x* der eigentliche Input in die Formel ist. Im Falle multipler Inputs, wird eine Summe der Produkte gebildet, wie hier dargestellt; $\sum w1*x1 + w2*x2 + wn*xn$. Das ermöglicht in der Theorie beliebig viele Inputs und Gewichtungen. *b* ist der Bias des Neurons, der neben den Gewichtungen eine weitere Möglichkeit darstellt, das Ergebnis der Formel zu beeinflussen. Als ein Beispiel für die Entscheidungsfindungsprozesse mit Hilfe von Percetrons kann eine Planung der eigenen Abendgestaltung herangezogen werden.
Die zu beantwortende Frage lautet: „Soll ich heute Abend mit meinen Freunden ins Kino gehen?“ Alle Gewichtungen und der Bias wurden zufällig gewählt und dienen lediglich der Veranschaulichung.

Input 1:	Läuft ein Film, der mir gefällt?	(Ja; 1, Nein; -2)	(Gewichtung: 1,2)
Input 2:	Kommt mindestens ein Freund mit?	(Ja; 1, Nein; 0)	(Gewichtung: 0,7)
Input 3:	Bin ich in der Stimmung zu gehen?	(Ja; 1, Nein; -1)	(Gewichtung: 0,5)

Bias: -0,7

Sollten Input 1 und 2 positiv, 3 hingegen negativ sein, ergibt sich folgende Rechnung:

$$0{,}7\ (1) = \sum 1{,}2 \cdot 1 + 0{,}7 \cdot 1 + 0{,}5 \cdot (-)1 + (-)0{,}7 \quad (2)$$

Das Ergebnis ist positiv, was in der Folge zu einem positiven Output durch das Neuron führt. Die Empfehlung lautet entsprechend „Gehe ins Kino“, da die Summe aller dem Neuron bekannten Faktoren dafürspricht.
Diese einfachste Form der Neuronen hat jedoch den großen Nachteil allein die Ergebnisse Positiv oder Negativ erzeugen zu können, die in Binärcode normalerweise durch eins und null

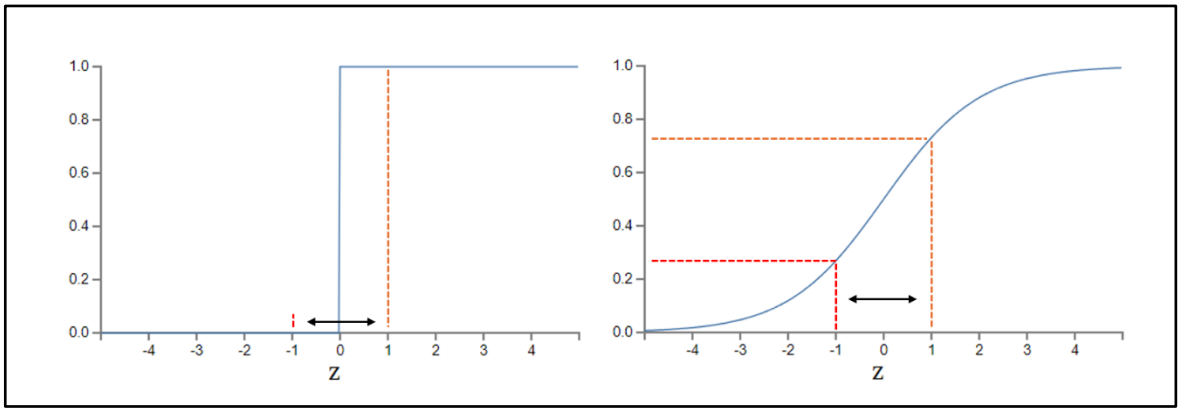
repräsentiert werden. Was für die Klärung der Freizeitpläne noch nützlich erscheinen mag, führt in den später erläuterten Netzwerkstrukturen allerdings zu Problemen. Denn eine leichte Anpassung der Gewichtungen oder des Bias kann entweder eine radikale oder gar keine Veränderung im Ergebnis zur Folge haben, wenn nur eine Entscheidung zwischen zwei absoluten Ausgängen möglich ist. Das ist in größeren Netzen kritisch, da nachgeschaltete Neuronen durch die Veränderungen des Outputs betroffen sind, was das Fine-Tuning des Zusammenspiels dieser extrem erschwert, wenn nicht gar unmöglich macht.

In der Praxis ist daher die Anwendung von sog. Sigma-Neuronen, auf Englisch sigmoid neurons, üblich, die der beschriebenen Problematik zuvorkommen. Wie der Name bereits vermuten lässt, hängen sie mit der unten dargestellten klassischen Sigma-Funktion zusammen.

$$\sigma(x) = \frac{1}{1 + e^{-x}} \tag{3}$$

Dieses statistische Instrument wird genutzt, um mehrere Werte auf einen Ergebniswert zu skalieren, der sich stets auf dem Intervall zwischen null und eins befindet. Damit ist der Output des Sigma-Neurons nicht wie bei einem Perceptron null oder eins, sondern ein dazwischenliegender Wert. Das bedeutet in der Folge, dass eine Anpassung von Gewichtungen und Bias deutlich feiner kalkulierbare Folgen hat. Der Unterschied zwischen den Funktionen der beiden kennengelernten Neuronen ist in der Abb. 3 grafisch illustriert.

Abbildung 3: Ergebnisfunktionen von Perceptrons (links) und Sigma-Neuronen (rechts)



Quelle: Eigene Darstellung nach: Nielsen (2019), o.S

Die grundlegende mathematische Funktion hinter einem Sigma-Neuron ist also:

$$Output = \frac{1}{1 + e^{-((\sum w \cdot x) - b)}} \tag{4}$$

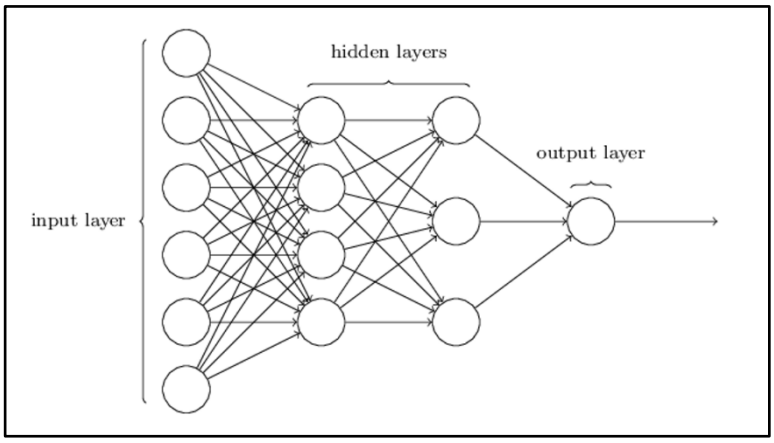
Diese lässt sich Anhand des oben verwendeten Beispiels demonstrieren. Werden die identischen Eingabewerte genutzt, ergibt sich 0,67 als Output des Neurons.

$$0{,}67 = \frac{1}{1 + e^{-((\sum 1{,}2 \cdot 1 + 0{,}7 \cdot 1 + 0{,}5 \cdot (-)1) - 0{,}7)}} \tag{5}$$

4.2 Allein stark, gemeinsam stärker: Neuronale Netze

Aber wie gelingt es einem Computer mithilfe dieser einfachen Formel die Fähigkeiten einer modernen KI zu erlangen? Mit einem einzelnen Neuron ist dies in der Tat nicht möglich, weshalb für eine komplexere Entscheidungsfindung mehrere Neuronen zu den bereits angesprochenen neuronalen Netzwerken oder Netzen zusammengesetzt werden. Diese folgen häufig einer Struktur, wie in Abb. 4 dargestellt. Was zunächst kompliziert aussehen mag, ist bei der Betrachtung im Detail nicht viel komplexer, als das Verstehen einzelner Neuronen. Auf der ersten Ebene, der sog. input layer, finden sich erste Neuronen, die alle nach der oben erläuterten Sigma-Methode aus gemachten Eingaben Werte zwischen null und eins berechnen. Sie werden in der Folge als Inputs an die dahinterliegende Ebene weitergegeben, die daraus wieder neue Werte errechnet. Dieser Vorgang wiederholt sich auf jeder der hidden layers, bis auf der letzten Ebene ein Output-Neuron ein finales Ergebnis berechnet. Durch die Akkumulierung von Bias und Gewichtungen über die verschiedenen Ebenen hinweg, ist es dem Netzwerk möglich immer komplexere Zusammenhänge auf hinteren Schichten

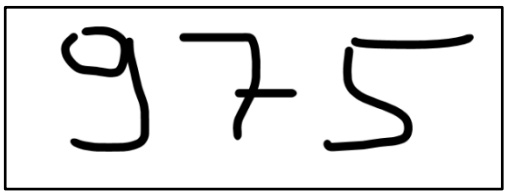
Abbildung 4: Beispielhafte Struktur eines neuronalen Netzes



Quelle: Nielsen (2019), o.S.

mathematisch darzustellen und damit zumindest statistisch zu verstehen. Um Missverständnissen in Bezug auf die Darstellung vorzubeugen, muss an dieser Stelle noch darauf hingewiesen werden, dass verschiedene Outputpfeile immer dasselbe Ergebnis repräsentieren, das an mehr als ein Folgeneuron weitergegeben wird. Auch ist die Anzahl der hidden layers und der Neuronen pro Ebene nicht festgelegt und kann je nach Aufgabentyp des Netzwerks und den gestellten Anforderungen variieren. Zudem handelt es sich bei den beschriebenen Netzen um die sog. feed-forward-Architektur. Wie erläutert, geben hier Neuronen die Ergebnisse nur an nachgeschaltete Partner weiter. Andere Architekturen, wie zum Beispiel recurrent, erlauben auch eine rückwärtige Weitergabe, in die Details dieser tiefer einzusteigen, würde an dieser Stelle aber zu weit gehen.

Abbildung 5: Beispielhafte Zahlenfolge



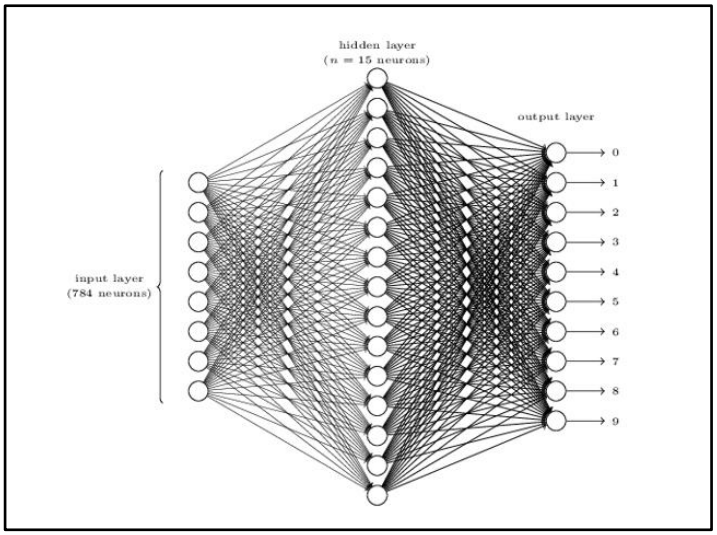
Quelle: Eigene Darstellung

Da nun die grundsätzlichen Funktionsweisen eines nN bekannt sind, stellt sich als nächstes die Frage, wie diese hochkomplexe Probleme, wie zum Beispiel das Verstehen von Spracheingaben und das Geben sinnvoller Antworten meistern. Dafür bietet es sich an, mit einer einfacheren Aufgabe zu beginnen, die sich häufig als ein Beispiel in der Literatur findet, nämlich dem Erkennen einer handschriftlichen Zahlenfolge, wie sie in Abb. 5 zu sehen ist. Während es für uns Menschen eine geringe Herausforderung ist, die Ziffern 975 zu erkennen, ist dies für Computer deutlich schwieriger. Denn gerade handschriftliche Zahlen können stark in den angenommen Formen variieren, aber zeitgleich eine Bedeutung teilen. Um diese Herausforderung trotzdem zu meistern, findet ein in der Auseinandersetzung mit neuronalen Netzwerken unabdingbares und im Rahmen der Arbeit ab jetzt immer wieder auftauchendes Prinzip Anwendung: Das Herunterbrechen großer Aufgabenstellungen in immer einfacher zu lösende, kleinere Stückchen. Diesem folgend ist es für das spezifische Beispiel also ratsam, zunächst mit dem Auswählen einer einzelnen Ziffer zu beginnen, beispielsweise der Neun. In einem Folgeschritt müssen die Daten zunächst für den Computer verarbeitbar gemacht werden. Während diese Kommunikation mit der Außenwelt gerade das Feld der Robotik vor eine große Herausforderung stellt, ist die Lösung für die vorliegende Problematik zum Glück recht simpel. Es erfordert lediglich die Übertragung der Ziffer auf ein digitales Feld, das aus 28 mal 28 Pixeln besteht, die Graustufen in Werten zwischen null und eins darstellen. Weiß (Hex #FFFFFF) kann hierbei beispielsweise null sein, während schwarz (Hex #000000) dem logisch folgend durch eins repräsentiert wird. Die so erhaltenen Werte pro Pixel stellen die in für Computer verarbeitbare

Werte codierten Inputs für das genutzte nN dar. Im Beispiel von Nielsen (2019) wird mit einem dreischichtigen Netzwerk gearbeitet, vereinfacht zu sehen in Abb. 6.

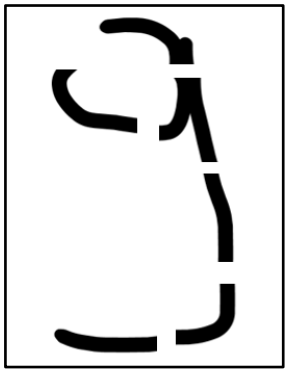
Erneut: Wie viele Ebenen und Neuronen pro Ebene ein Netzwerk hat, hängt ganz von der Aufgabe und den Präferenzen des Architekten ab. Die 784 Neuronen auf der input layer ergeben sich aus der Anzahl an Pixeln (*28*28=784*). Auf der letzten Ebene finden sich die zehn Neuronen, die die Ziffern 0,1,2,3,4,5,6,7,8,9, also die möglichen Lösungen, repräsentieren. Was auf der mittleren hidden layer geschieht, lässt sich aus der Perspektive der output layer als das Herunterbrechen der Ziffern in kleinere Bestandteile, die durch Outputwerte repräsentiert werden, beschreiben. Aus dem Blickwinkel der input layer würde man den Prozess als ein Zusammensetzen der einzelnen Pixelwerte zu kleinen Ziffernstücken beschreiben. In dieser Perspektive verbleibend, fließen diese Stücke dann in die letzte Ebene ein, was den dort angesiedelten Neuronen erlaubt, eine Entscheidung über die finale Antwort zu treffen. Ist das Netzwerk richtig kalibriert, sollte diese letzte Schicht aus einem beispielhaften Input, wie in Abb. 7 dargestellt, in der Lage sein, die Ziffer 9 zu erkennen, ähnlich wie wir dies selbst können. Es spielt sich also genau der bereits beschriebene Prozess ab. Einfache Inputs der Farbwerte in individuellen Pixeln, werden durch nachgeschaltete Neuronen zu komplexeren Konzepten vernetzt, die wiederum auf einer weiteren Ebene zu der handschriftlichen Neun zusammengesetzt werden.

Abbildung 6: Strukturdarstellung eines neuronalen Netzes zur Identifikation handschriftlicher Ziffern



Quelle: Nielsen (2019), o.S.

Abbildung 7: Arbeitsschritte nN



Quelle: Eigene Darstellung

Wie ein neuronales Netzwerk Entscheidungen trifft, ist nun also geklärt. Aber wie lernt ein nN Zahlen tatsächlich richtig zu erkennen? Hierfür kann erneut das anfängliche Beispiel des Kinoabends zur Illustrierung herangezogen werden.

4.3 Selbstlernende Systeme

Das beschriebene Perceptron kann bei der Entscheidungsfindung über die Abendgestaltung unterstützen, dies aber nur dann sinnvoll tun, wenn die Gewichtungen und der Bias den tatsächlichen Vorlieben entsprechen.
Ist es vollkommen egal, welcher Film gezeigt wird und man möchte einfach nur einen Abend mit seinen Freunden verbringen, macht es wenig Sinn, diesem Input das höchste Gewicht zuzuteilen. Dies kann dazu führen, dass das Neuron im schlimmsten Fall die Empfehlung ausspricht, ins Kino zu gehen, auch wenn keiner der Bekannten Zeit hat mitzukommen. Dieses Prinzip lässt sich direkt von einzelnen Neuronen auf den Output ganzer Netzwerke übertragen. Das Ziel für ein gut funktionierendes nN ist also immer die optimale Kalibrierung der Gewichtungen und des Bias. Eine händische Anpassung aller Parameter ist bei riesigen Netzen, man denke an die genannten 175.000.000.000 Neuronen von ChatGPT, aber nicht praktikabel. Es muss also ein Weg gefunden werden, den Prozess mithilfe von Algorithmen zu optimieren und Turings Idee des selbstlernenden Systems umzusetzen. In der mathematischen Ausdruckweise kann für das spezifisch vorliegende Problem die untenstehende Kostenfunktion genutzt werden.

$$C(w,b) = \frac{1}{2n}\sum(y(x) - a)^2 \tag{6}$$

C repräsentiert hierbei die Kosten oder anders ausgedrückt, die Genauigkeit des nN. Hinter dem Summenzeichen steht der Term der als SE oder squared error, auf Deutsch quadratischer Fehler, bekannt ist. Dieser ist nichts anderes als der Abgleich des tatsächlichen Ergebnis eines Datenpunktes mit dem vorgegebenen, beziehungsweise gewünschten Ergebnis zum Quadrat. *a* stellt dabei das geforderte Ergebnis dar, während *y* das tatsächliche Resultat wiedergibt. Die Quadrierung verfolgt den Zweck, die Sensitivität der Messbarkeit in den Veränderungen zu erhöhen. Zudem stellt sie sicher, dass die Funktion immer positiv bleibt. Das zweite Element in der Formel ist *1/2n*. *n* beschreibt die Anzahl der Datenpunkte, wodurch sich mit *1/n* errechnen lässt, wie groß der quadratische Fehler durchschnittlich ist, während die Multiplikation mit 2 lediglich das spätere Ableiten erleichtern soll. Es kann zusammenfassend festgehalten werden, dass das Ergebnis der Kostenfunktion vom durchschnittlichen quadrierten Fehler abhängt und ein Netzwerk besonders gut arbeitet, wenn das Resultat gegen Null geht.

Um sich in diese Richtung zu bewegen und die Kostenfunktion zu minimieren, müssen die jeweiligen Gewichte und Bias in einem iterativen Prozess so lange neu angepasst werden, bis ein Minimum erreicht wird. Zur mathematischen Optimierung eignet sich dafür die Verwendung des gradient descent, auf Deutsch Gradientenabstieg. Um verständlich zu machen, was dabei grundsätzlich passiert, bietet sich die folgende starke Vereinfachung an:
In einem ersten Schritt muss der namensgebende Gradient bestimmt werden. Bei diesem handelt es sich um einen Vektor, der die steilste Steigung einer Funktion, in diesem Fall C, anzeigt. Da dies nicht sofort hilfreich ist, wird der negative Gradient verwendet, um die Richtung des steilsten Abstiegs der Funktion zu erkennen. Daraufhin werden Gewichte und Bias in Richtung des Minimums der Funktion angepasst. Dies geschieht schrittweise, wobei die Größe der Schritte durch die Lernrate repräsentiert wird, die in der Literatur häufig als η bezeichnet wird. Sie darf nicht zu groß sein, um die Genauigkeit des Algorithmus zu garantieren, aber auch nicht zu klein, da sonst zu viele Schritte bis hin zur Optimierung nötig sind und der Algorithmus in Folge an Effizienz verliert.
Um sich diesen Prozess besser vorzustellen, hier ein visualisierender Vergleich; man steht auf einem Hang, der die Kostenfunktion repräsentiert und schaut sich in alle Richtungen um, um zu erkennen, in welcher Richtung es am steilsten zum Gipfel geht. Sich daran orientierend, macht man dann einen Schritt in die entgegengesetzte Richtung, was hier beispielhaft für die Anpassung der Gewichte und Bias steht. Diesen Prozess wiederholt man so oft, bis man das Ende des Hangs und den tiefsten Punkt des dahinterliegenden Tals erreicht hat. Ist dies geschehen, wurde die Kostenfunktion optimiert und das Netzwerk arbeitet mit der höchstmöglichen Genauigkeit. Da es sich allerdings häufig um gewaltige Datenmengen handelt, auf die ein KI-System trainiert werden muss, wird in der Praxis normalerweise der stochastische Gradientenabstieg verwendet, der an dieser Stelle noch kurz erläutert werden soll. Er dient dem Zweck, den eben beschriebenen Optimierungsprozess weiter zu beschleunigen. Das Problem des klassischen Gradientenabstiegs ist, dass dieser die Optimierung über alle an das System gefütterten Daten berechnet. Der stochastische Gradientenabstieg tut dies jeweils nur für stichprobenartige Mini-Batches und nicht das gesamte Datenset. So kann während eines Epochs (ein vollständiger Trainingsdurchlauf) enorm viel Rechenleistung eingespart und die Effizienz des Lernprozesses spürbar erhöht werden.

4.4 Mangelndes Verständnis und große Potenziale

An dieser Stelle sind noch zwei weitere wichtige Punkte zu nennen:
Erstens, durch die beschriebene algorithmische und damit automatische Bestimmung der Gewichtungen und Bias wird es bei steigender Anzahl der Neuronen in einem Netzwerk faktisch unmöglich nachzuvollziehen, warum sich bestimmte Parameter so ergeben, wie sie sich letztlich ergeben. Es lässt sich also die These aufstellen, dass die Menschheit mit hochkomplexen nN zum ersten Mal eine Technologie geschaffen hat, deren Vorgänge in der Entscheidungsfindung von uns selbst nicht zu 100% verstanden werden.
Zweitens, die Erkennung einfacher Ziffern oder von Ziffernfolgen ist bereits mit einem sehr simplen und verhältnismäßig kleinem neuronalen Netzwerk möglich. Es ist damit auch nur der Beginn der Bilderkennungsfähigkeiten, zu denen KI theoretisch und zum Stand der Abgabe der Arbeit auch praktisch bereits jetzt in der Lage ist. Jeder Leser mit einem Internetzugang, kann den Hyperlink (https://chatgpt.com/) zur freiverfügbaren Anwendung ChatGPT öffnen und dort via Drag&Drop ein beliebiges Bild hochladen. Sei es ein Haus, ein Tier oder ein anderer Gegenstand, das System wird mit sehr hoher Wahrscheinlichkeit in der Lage sein, das Objekt grundsätzlich zu identifizieren. Mit etwas Glück kann die KI, beim Beispiel des Autos bleibend, sogar den Hersteller, das Modell, den Jahrgang und weitere Details erfolgreich erkennen. Dies ist möglich, da die Architektur von ChatGPT dem beschriebenen Prinzip folgt, komplexe Aufgaben in für Neuronen lösbare Schritte herunterzubrechen und auf die entsprechenden Daten trainiert wurde. Der Kreativität der Architekten sind nur die technischen Grenzen der Hardware gesetzt. Strukturen mit Milliarden von Neuronen, verschiedenen Ebenen oder Sub- und Subsubnetzwerken sind, wie an dem kleinen Selbstexperiment gezeigt, sogar bereits für Privatverbraucher mit einem Internetzugang frei verfügbar.
Während es häufig zunächst beeindruckend ist, diese Fähigkeiten selbst zu testen und damit zu experimentieren, gibt es auch Stimmen, die die Entwicklungen weniger positiv sehen. Gerade Algorithmen zur Gesichtserkennung oder massenhaften Auswertung von Chat-Nachrichten stehen in diesem Kontext häufig in der Kritik, die Privatsphäre von Bürgern westlicher Länder einzuschränken und Autokraten zur Überwachung der Bevölkerung zu dienen.[17] So darf trotz aller Faszination nie außer Acht gelassen werden, welchen Schaden Technologien im Allgemeinen und KI im Speziellen anrichten können, wenn sie in die falschen Hände geraten, gerade

[17] Vgl. CNBC (2024), o.S.

weil die reine Bilderkennung nur einen kleinen Teil der Fähigkeiten moderner AI-Systeme ausmacht. Sei es das Verarbeiten und Generieren von Sprache, das Erstellen von Bildern und Videos, das Komponieren neuer Lieder bereits verstorbener Interpreten oder das Steuern selbstfahrender Autos, es gibt enorm viele Einsatzmöglichkeiten für moderne künstliche Intelligenz. Insbesondere von generativer Sprach-KI werden große Einschnitte in unsere Leben, sowohl privat als auch beruflich, erwartet, weshalb sich die folgenden Kapitel gezielt mit diesen, als Large Language Models bekannten Modellen oder abgekürzt LLMs, befassen.

5. Wie Maschinen Sprache meistern: Large Language Models

„Große Sprachmodelle (LLMs) sind ein Schlüsselelement moderner künstlicher Intelligenz, das natürliche Sprache verarbeitet und generiert. Sie basieren auf neuronalen Netzen und werden mit umfangreichen Daten trainiert. Dieses Kapitel beleuchtet ihre Funktionsweise und Anwendungsbereiche."

Das ist die Antwort, die das Programm ChatGPT auf die recht wage formulierte Befehlseingabe: „Schreibe eine kurze Einleitung (3 Sätze) für das Kapitel LLMs einer Bachelorarbeit über KI", innerhalb weniger Sekunden generiert. Was den Vorstellungen von Sciencefiction-Autoren aus dem vergangenen Jahrhundert gleicht, ist für uns heute bereits ein Teil der Realität. Computer, die menschliche Sprache beherrschen und von Erklärungen zu komplexen Themen der Teilchenphysik bis hin zu Cocktailrezepten auf scheinbar alles eine Antwort geben können, beruhen, wie in der erstellten Einleitung zu lesen, auf der Erfindung von Large Language Models.

„Large Language Models (kurz: LLM und auf Deutsch: Große Sprachmodelle) sind leistungsstarke Modelle, die darauf ausgelegt sind, menschliche Sprache zu verstehen und zu generieren. Sie können Text analysieren und verstehen, kohärente Antworten generieren und sprachbezogene Aufgaben ausführen."[18] lautet die Definition des Fraunhofer Institutes, die einen Überblick über die Fähigkeiten moderner LLMs gibt. Aber wie erlernen nN diese Funktionsweisen? Auch für die Beantwortung dieser Frage ließe sich tief ins Detail einsteigen, allerdings wird sich das Kapitel aus vorgenannten Gründen auf die wichtigsten Zusammenhänge beschränken.

Johri (2023), o.S. folgend, muss man sich dafür erneut das zentrale Prinzip des Herunterbrechens großer Aufgaben in kleinere Teile vor Augen rufen. Ein Text besteht möglicherweise aus mehreren Absätzen, die wiederum aus Sätzen bestehen, die letztlich in Zusammenhang gebrachte Wörter sind. Wörter können dann als Repräsentationen verschiedener Gegenstände, logischer Funktionen, Zustände und vielem mehr verstanden werden. Auf genau dieser Wortebene arbeiten LLMs, indem sie, vereinfacht gesagt, bei der Satz- und Textbildung versuchen, das jeweils nächstfolgende Wort richtig vorherzusagen. Um bei diesem Vorgang die inhaltliche Richtig- und Sinnhaftigkeit mit einer hohen Wahrscheinlichkeit zu ermöglichen, ist vor allem ein Faktor relevant; eine riesige Menge an Trainingsdaten und -durchläufen. Das Wort riesig ist hierbei gewusst gewählt und keine Übertreibung. Laut Layton (2023), o.S. wurden für das Training des älteren Models ChatGPT 3 von OpenAi vermutlich ca. 570 Gigabyte an Textdaten für

[18] Kelbert/Siebert/Jöckel (2023), o.S.

das sog. pre-training verwendet. Um dies in ein Verhältnis zu setzen, hilft der Vergleich mit der vorliegenden Arbeit. Bei Abgabe wird das Dokument eine Größe von ungefähr 5.500 Kilobyte haben (inklusive aller Abbildungen). Damit entspricht es in etwa 0,00001% der vermuteten Menge an Trainingsdaten. Diese gewaltigen Textmengen sind jedoch erforderlich, um das zugrundeliegende LLM auf grundsätzliche Sprachstrukturen zu trainieren. Zu diesen gehören unter anderem Grammatik, Rechtschreibung und weitere sprachliche Regeln. Während Menschen diese Sprachregeln normalerweise in Unterrichtseinheiten lernen und dann anwenden, lernen LLMs lediglich aus dem Kontext einer Vielzahl von Beispielen.

Neben der sprachlichen Korrektheit verfolgt dieses Training den weiteren Zweck, der KI zu helfen, die eigenen Parameter auf allgemein erkennbare inhaltliche Muster in der Sprache, außerhalb der genannten Regeln, zu kalibrieren. Auf diese Art und Weise lernen KI-Systeme ihr vermeintliches Wissen. Im Unterschied zum Wissen von Menschen basiert dieses allerdings allein auf statistischen Wahrscheinlichkeiten, die in den Gewichtungen und Bias eingeprägt wurden. Menschliches Wissen hingegen beruht auf einem Konzept einer größeren Realität sowie dem Verständnis von Fakten. Einmal gelernt, nutzen LLMs diese Sprachmuster, um eingegebene Eingabebefehle zu entschlüsseln und schlussendlich, durch das Aneinanderreihen der passenden Worte, zu beantworten. Beim Erstellen der obigen Einleitung hat das LLM die darin enthaltenen Informationen also nicht, wie wir es tun würden, als Fakten abgerufen, sondern lediglich errechnet, dass die gewählten Worte die statistisch wahrscheinlichste Antwortmöglichkeit repräsentieren. Das pre-training mit Textmassen ist demnach die Grundlage für die Sprachfähigkeit von nN. Je nach Bedarf und Anforderungen, kann eine Sprach-KI noch weiter auf bestimmte Fähigkeiten trainiert werden, wie zum Beispiel dem Schreiben von Code für Computerprogramme. Hierfür werden dem System weitere Daten zugefüttert, die sich auf den spezifischen Anwendungsfall beziehen. Dieses Detailtraining kann (muss aber nicht) unter der Aufsicht menschlicher Tester erfolgen, die dem LLM-Feedback zu den Lernergebnissen geben, um deren Genauigkeit zu erhöhen. LLMs greifen zudem auf eine bestimmte Architektur zurück, die als transformer bekannt ist. Diese an dieser Stelle weiter zu beleuchten, würde jedoch zu weit gehen.

Wichtiger für die Zielstellungen der Arbeit sind die Fragen, welche LLMs es bereits auf dem Markt gibt, wie sie unsere Arbeitswelt zukünftig beeinflussen werden und wie man sie optimal für sich nutzen kann. Hiermit befassen sich die nachfolgenden Kapitel.

6. Aktuell verfügbare Sprach-KI Anwendungen

Die Liste der verschiedenen LLM-Anwendungen, die für Privatverbraucher verfügbar sind, wird immer länger. Sie können sich dabei sowohl als ein eigenständiger Chat-Bot präsentieren, wie im Falle von ChatGPT, allerdings auch in andere Plattformen integriert sein, wie das Beispiel von xAI und GROK zeigt. Hierbei ist das LLM Teil der Social-Media-Plattform X (früher Twitter). Im Folgenden wird allerdings allein auf eigenständige Systeme eingegangen. Hier gilt vor allem: KI ist nicht gleich KI. Modelle wie Replika, eine AI, die das Ziel hat, einen virtuellen Partner zu simulieren, mag diese Aufgabe sehr gut erfüllen, ist für eine unterstützende Anwendung am Arbeitsplatz, die im Kern dieser Arbeit stehen soll, jedoch eher nicht geeignet. Zahlungspflichtige Modelle stellen für viele Nutzer eine Hemmschwelle dar, gerade wenn man sich erstmalig mit der Technologie vertraut machen möchte. Deshalb werden nun drei KI-Modelle vorgestellt, die einsteigerfreundlich sind und deren Gebrauch mit keinen oder höchstens optionalen, monetären Kosten verbunden ist.

6.1 ChatGPT

ChatGPT ist dabei das wohl bekannteste Modell auf dem Markt. Das von OpenAI entwickelte System war lange ein Magnet medialer Aufmerksamkeit und wurde von manchen fast schon als ein Synonym für künstliche Intelligenz genutzt. ChatGPT basiert, wie der Name bereits verrät, auf der generative pre-trained transformer Architektur und verfügt über 175.000.000.000 individuelle Neuronen, was es zu einem sehr großen und fähigen System macht. Die neuesten Funktionen und Versionen sind leider nicht kostenlos. Der Zugriff auf die Premiumversion kostet den Nutzer 20 USD pro Monat. Allerdings ist eine Vielzahl von Basisfunktionen und ältere Versionen kosten- und werbefrei uneingeschränkt nutzbar. Neben dem Lösen sprachbasierter Aufgaben gehört zu den Fähigkeiten des Modells unter anderem das Generieren und Erkennen von Bildern und die Auswertung von Dokumenten oder Informationen auf Webseiten.
ChatGPT hat dabei verschiedene Schwächen, die von der Entwicklerfirma OpenAI selbst genannt werden. Zu diesen gehört das Generieren richtig klingender, aber inhaltlich falscher Antworten, ein als Halluzination bekanntes Phänomen, das an späterer Stelle noch genauer erklärt wird. Außerdem kann die Formulierung einer Anfrage dazu führen, dass das Modell diese als nicht umsetzbar ansieht, obwohl der Eingabebefehl unter Verwendung einer anderen Wortwahl durchaus möglich gewesen wäre. Manchmal werden Befehle nicht ausgeführt, wenn sie gegen

die Nutzungsregeln des Produktes verstoßen, was in bestimmten Fällen zu einer Überzensur führen kann, wenn das System eine Nutzeranfrage falsch aufnimmt.[19] In der Bewertung externer Nutzer erreicht ChatGPT auf der Bewertungsplattform Capterra einen Punkteschnitt von 4,5 aus 5 möglichen Sternen. Die Punkte, die Usern am positivsten auffallen, sind die unkomplizierte Nutzung und Anmeldung, die Option durch Plug-Ins noch mehr Funktionen und Integrationen freizuschalten und die Tatsache, dass auch die frei verfügbaren Versionen den Nutzern sehr viele Anwendungsmöglichkeiten bieten. Negativ machen sich hingegen die oben bereits erwähnten, inhaltlich falschen Antworten bemerkbar, die häufig ohne Hinweis auf mögliche Unsicherheiten gegeben werden. Zudem gibt es Beschwerden und Studien über einen politisch links lehnenden Bias in den erhaltenen Antworten.[20]

6.2 ClaudeAI

ClaudeAI ist der schlichte Name der KI-Modelle der US-Firma Anthropic, die ebenfalls auf einer GPT-Architektur basieren. Deren Besonderheit liegt in dem angewandten constitutional AI Ansatz. Dieser verfolgt das Ziel, eine Sprach-KI zu kreieren, die hilfreich in der Lösung menschlicher Probleme ist, sich dabei allerdings an einer bindenden Verfassung orientiert, die deren Einsatz für schädliche Zwecke verhindern soll. An dieser Stelle muss jedoch erwähnt werden, dass diese Verfassung und damit ein Verständnis darüber, was schädlich ist und was nicht, in letzter Konsequenz allein durch die Firma Anthropic bestimmt wird.[21] Zu den größten Stärken gehört die hohe Anzahl an aufnehmbaren Token und die damit verbundene Fähigkeit des Systems, lange Texte verlässlich zu verstehen und zu analysieren. Ein Token entspricht etwa vier Zeichen und das Limit für Claude 3 liegt bei bis zu 200.000 Token. ChatGPT 3 bietet mit ca. 8.000 Token eine deutlich geringere Aufnahmekapazität. Die Schwächen von Claude sind vergleichbar mit denen anderer Modelle. Nutzer geraten laut eigener Aussage durch den beschriebenen constitutional AI Ansatz allerdings deutlich häufiger in Situationen, in denen das Modell Lösungsvorschläge verweigert, da die dazugehörigen Anfragen vermeintlich gegen Nutzungsregeln verstoßen. Positiv fällt hingegen die Nutzung im Rahmen der Softwareentwicklung auf, bei der sich Claude durch eine sehr geringe Fehlerquote in generiertem Code auszeichnen soll. Auf der Ratingplattform Capterra erhält die Anwendung von Nutzern 4,8 von

19 Vgl. OpenAI (2022), o.S.
20 Vgl. Janssen (2023), o.S.
21 Vgl. Anthropic (2022), o.S.

5 möglichen Sternen. Auch für Claude gibt es Premiumpläne, die den Einzelnutzer ca. 22 EUR pro Monat kosten. Sie erlauben unter anderem priorisierten Zugang in Zeiten hoher Serverauslastung und Zugriff auf neuere und bessere Modelle der Claude Serie.[22]

6.3 Gemini

Das letzte vorgestellte LLM ist Gemini, eine KI des Google bzw. Alphabet Konzerns. Ursprünglich als Bard bekannt, war der Launch eine Antwort auf das 2022 erschienene ChatGPT und den damit verbundenen medialen Aufruhr. Gemini ist wie Chat-GPT ein multimodales LLM, was bedeutet, dass Inputs nicht allein Text, sondern auch Bilder, Sprachaufzeichnungen oder Dateien sein können. In Tests wird die Schnelligkeit des Modells lobend erwähnt, vor allem, wenn es um die Generierung von Grafiken und Bildern geht. Allerdings wird auch Kritik über Situationen, in denen das System Antworten verweigert, verschiedene Fehler darin einbaut oder mit Anfragen schlichtweg überfordert wirkt und sehr verallgemeinernde und nicht wirklich hilfreiche Outputs erzeugt, geäußert. Auf der für die anderen Modelle genutzten Rating Plattform Capterra, ist zum Zeitpunkt der Erstellung erst eine Rezension vorhanden. Deshalb wird die Bewertung anhand von Sternen, aus Mangel an Repräsentativität an dieser Stelle nicht aufgeführt. Wie bei den beiden anderen vorgestellten Modellen, gibt es neben der kostenfreien Nutzung Premiumoptionen, die unter dem Namen Gemini Advanced ca. 22 EUR pro Monat kosten. Das Besondere daran ist, dass neben einer Verbesserung der Geschwindigkeit bspw. auch die Integration in Gmail und andere Google Produkte ermöglicht wird. Für Nutzer des breiteren Google Ökosystems ist Gemini Advanced also besonders attraktiv.[23] [24]

Wie am Anfang des Kapitels bereits erwähnt, gibt es neben den drei vorgestellten noch viele weitere Anwendungen, die auf LLMs basieren. Welche am besten geeignet ist oder sind, hängt letztlich von persönlichen Präferenzen und den Anwendungsfällen ab, in denen man das Modell nutzen möchte. Es empfiehlt sich also, verschiedene Angebote auszuprobieren, um eine Alltagslösung für sich selbst zu finden.

[22] Vgl. Capterra (2024), o.S.
[23] Vgl. Google (2023), o.S.
[24] Vgl. PCMAG (2024), o.S.

7. LLMs in der Arbeitswelt: Anpassung als Überlebensstrategie

Die Auswirkungen neuer Technologien auf unsere Arbeitswelt sind stets komplex und das Ausmaß der mit ihnen einhergehenden Veränderungen nur schwer vorherzusagen. Bei künstlicher Intelligenz ist dies noch schwieriger, denn es handelt es sich um eine Schlüsseltechnologie, die Maschinen bisher völlig ungekannte Fähigkeiten verleiht. In Kombination mit Innovationen im Feld der Robotik besitzt KI das Potenzial, die Arbeitsweisen in einer Vielzahl von Branchen massiv zu beeinflussen. In diesem Zusammenhang sind berechtigt auftauchende Fragen, wie stark der eigene Beruf betroffen sein wird, wie es um die eigene Jobsicherheit steht und wie als Arbeitnehmer mit der neuen Technologie am besten umzugehen ist. Diese Fragen sollen im aktuellen Kapitel beantwortet werden, allerdings mit einer inhaltlichen Fokussierung auf den Einfluss von Large Language Models und den Anwendungen, auf denen sie basieren und nicht KI im Allgemeinen.

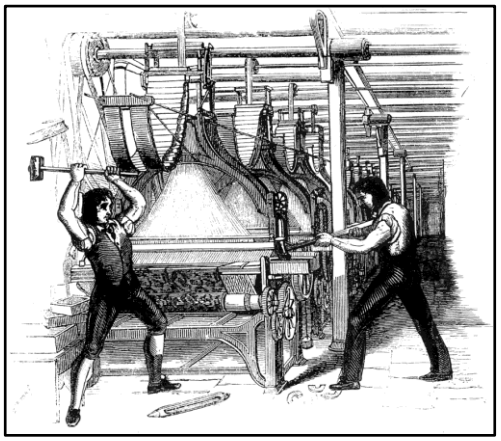
7.1 Kommende Veränderungen und wie auf diese reagieren

Wie lässt sich überhaupt feststellen, dass KI in der Zukunft spürbare Einflüsse auf unser Arbeitsleben haben wird? Zur Beantwortung dieser Frage können die Ergebnisse aus WEF (2023), S. 5 herangezogen werden. In diesem Report geben 75% der befragten Unternehmen an, AI im Allgemeinen und 62% generative und sprachbasierte KI im Speziellen in der näheren Zukunft in die Arbeitsabläufe integrieren zu wollen. Sicherlich ist es so, dass Vorhaben nicht immer in die Tat umgesetzt werden. Die Tatsache, dass laut Schlude et al. (2023), o.S. 23% der das Internet nutzenden deutschen Arbeitnehmer, LLMs schon heute für die Ausübung ihrer Jobs nutzen, spricht allerdings dafür, dass die Technologie bereits in vielen Betrieben angekommen ist. Das macht eine zukünftige offizielle Implementierung von Arbeitgeberseite nur wahrscheinlicher. Es ist also ratsam, den kommenden technologischen Wandel unserer Arbeitswelt in der eigenen Karriereplanung zu berücksichtigen, auch wenn KI noch nicht in der Lage sein mag, eigenständig ganze Berufe zu übernehmen. Dies gilt gerade für junge Menschen, die einen großen Teil ihres Arbeitslebens noch vor sich haben und damit am meisten von den anstehenden Veränderungen betroffen sein werden.

Um ein Verständnis von diesen zu erlangen und damit richtig umzugehen, hilft die Betrachtung der Auswirkungen vergangener technischer Innovationen. Um genau zu sein jenen, die zur Automatisierung von Arbeitsschritten beigetragen haben. Denn ob über Roboterarme gesprochen

wird, die hochpräzise Schweißnähte für Autos setzen, KI, die in der Zukunft Call Center obsolet machen könnte oder die verhältnismäßig simple Erfindung der Spinning Jenny, im Kern geht es immer um das gleiche Prinzip; den Ersatz menschlicher Arbeitskraft durch Maschinen, unter der Voraussetzung, dass dies aus wirtschaftlichen Gesichtspunkten sinnvoll ist. Die Implementierung von Automatisierungstechnologien führt in einem logischen Folgeschritt stets zur Vernichtung von alten und durch diese obsolet gewordenen Arbeitsplätzen, wie in Acemoglu/Robinson (2012), S. 246ff. anschaulich dargestellt und schöpferische Zerstörung getauft wird. Da es sich bei dieser um Prozesse auf verschiedenen gesellschaftlichen Ebenen handelt, die von einer Vielzahl von Menschen mitgetragen werden, ist es keine ratsame Strategie, sich neuen Technologien und dem Wandel, den sie mit sich bringen, entgegenzustellen. Ein historisches Beispiel hierfür bieten die Ludditen, Textilarbeiter, die im vereinigten Königreich der 1810er Jahre Maschinen zur automatisierten Textilverarbeitung in ausufernden Protesten zerschlugen. Die Motive waren Existenzängste durch den Wegfall von Arbeitsplätzen in der Textilindustrie und für alle weiter bestehenden Stellen ein damit einhergehendes, sinkendes Lohnniveau. Auch wenn bis heute berüchtigt, konnte der Wandel der Branche durch die Aktionen nicht aufgehalten werden, sondern das Gegenteil trat ein. Maschinen wurden in der Folgezeit bis zum heutigen Tag nur noch besser und leistungsfähiger und den Ludditen blieb nichts als die Anpassung an die neuen Gegebenheiten. Neben den britischen Textilarbeitern gibt es in der Geschichte viele weitere Beispiele, die zeigen, dass neue Technologien, wenn sie mit einem tatsächlichen Effizienzvorteil verbunden sind, nach ihrer ersten Implementierung nicht mehr verschwinden (außer sie werden selbst durch bessere Technologien ersetzt). Eine empfehlenswertere Strategie, um allem Wandel zum Trotz auf dem Arbeitsmarkt bestehen zu können, ist daher die proaktive Anpassung.

Abbildung 8: Ludditen zerstören Maschinen zur Textilverarbeitung



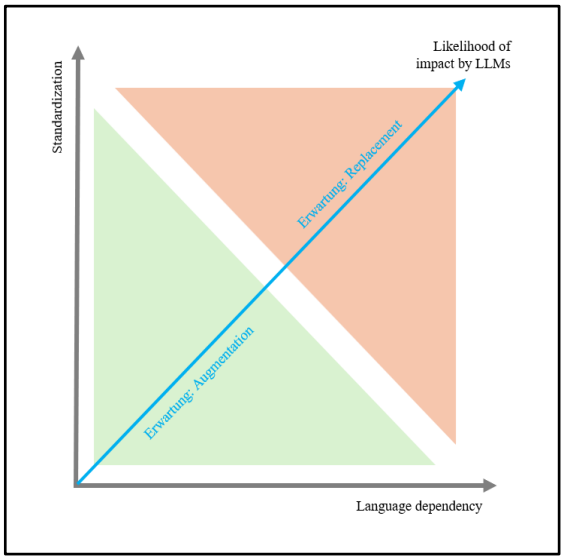
Quelle: Unbekannter Künstler (o.J.)

Im Rahmen von KI stehen dafür grundsätzlich zwei Wege offen. Der erste ist eine Spezialisierung auf die neue Technologie und eine Karriere in dem Feld. Dieser Pfad ist allerdings mit einem großen Investment von Zeit und anderen Ressourcen verbunden und nicht für jeden erstrebenswert, besonders wenn man sich bereits in einem anderen Sektor oder Beruf spezialisiert hat.

Der zweite verbleibende Weg ist die Neuorientierung innerhalb des eigenen Feldes, die eine Auseinandersetzung mit der Technologie und die Vorbereitung auf kommende Veränderungen erfordert. Eine Möglichkeit ist beispielsweise, zwar in der Energiewirtschaft tätig zu bleiben, sich aber auf Positionen zu bewerben, die voraussichtlich nur geringfügig durch die Einführung von KI betroffen sein werden.

Eine essenzielle Frage bei Wahl des letzteren Pfads ist daher, welche Arbeitsplätze auf welche Art und Weise von der beschriebenen schöpferischen Zerstörung oder weniger dramatisch ausgedrückt, der wirtschaftlichen Transformation, betroffen sein werden. Um dies für LLMs zu beantworten, ist es hilfreich damit zu beginnen, Berufe als Summen verschiedener Tätigkeiten zu betrachten. Die so entstandenen einzelnen Aufgaben lassen sich dann auf die Anfälligkeit für eine Übernahme oder Veränderung durch künstliche Intelligenz untersuchen. Laut WEF (2023), S. 5ff. sind dafür zwei primäre Faktoren unmittelbar ausschlaggebend. Der erste ist die sog. language dependency, also die Wichtigkeit, die die Beherrschung einer Sprache für die ausgeübte Tätigkeit einnimmt. Eine hohe language dependency trifft zum Beispiel auf Autoren, Kolumnisten oder Übersetzer zu, während für Dachdecker, Gärtner oder Reinigungskräfte das Gegenteil der Fall ist. Der zweite Faktor ist der Standardisierungsgrad der zu erledigenden Aufgaben. Sind Tätigkeiten sich stets wiederholend und ähnelnd, ist die Wahrscheinlichkeit, dass diese langfristig durch künstliche Intelligenz übernommen werden, deutlich höher als bei Aufgaben, die jedes Mal ein unterschiedliches Anforderungs- und Ergebnisprofil aufweisen. Abb. 9 illustriert diese beschriebene Dynamik in einem Koordinatensystem. Berufe, die sich in der orangenen Zone befinden, werden mit größerer Wahrscheinlichkeit langfristig von LLMs vollständig übernommen. Berufe im grünen Bereich werden sich in der absehbaren Zukunft wahrscheinlich nur in einzelnen Tätigkeiten verändern, der Kern der Arbeit wird jedoch in menschlicher Hand bleiben. In der Literatur wird für diese Fälle der englische Begriff der augmentation genutzt. Im selben Report werden verschiedene Tätigkeiten zudem in vier mögliche Kategorien unter-

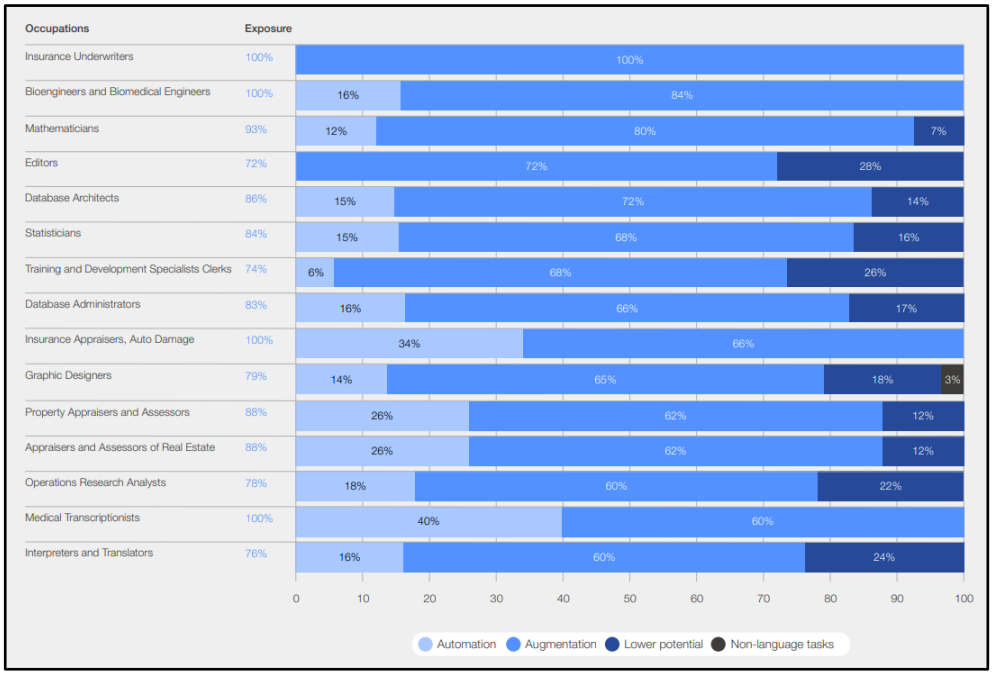
Abbildung 9: Auswirkungen von LLMs auf verschiedene Tätigkeiten



Quelle: Eigene Darstellung nach: WEF (2023), S. 7.

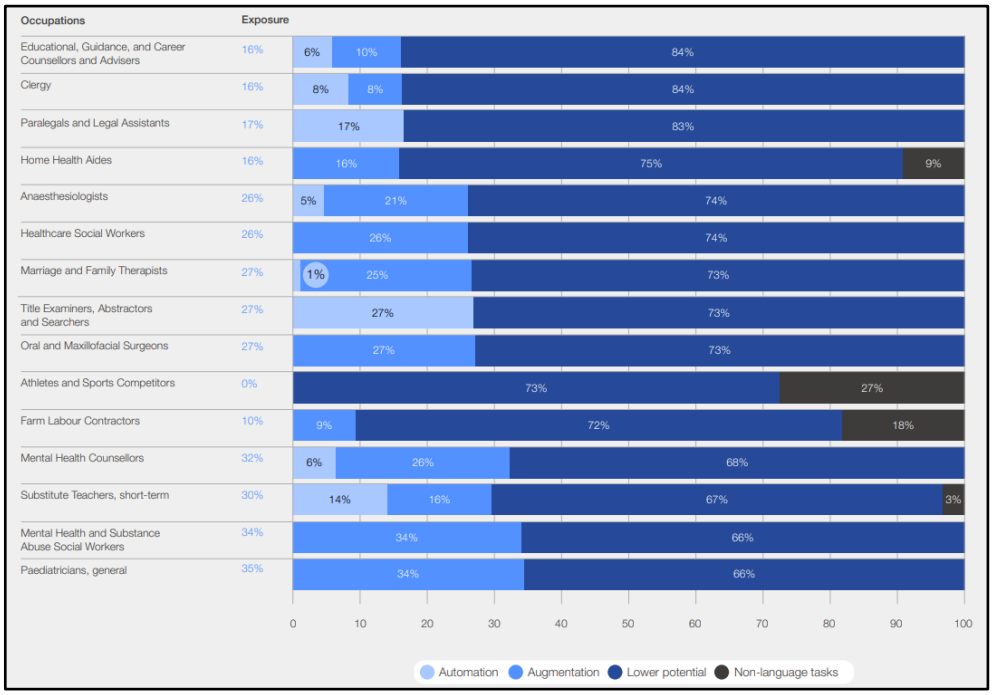
teilt. Diese sind: higher potential for automation, higher potential for augmentation, lower potential for exposure und non-language tasks. In die erste fallen unter anderem Analysetätigkeiten sowie administrative Aufgaben. In der zweiten Kategorie finden sich viele Aufgaben, die mit der Bewertung von Mitarbeitern, Kundenzufriedenheit und dem Testen von Systemen zu tun haben. Wenig Potenzial für augmentation oder Automatisierung bieten Vertrags- oder andere Verhandlungen, die Koordinierung von Ressourcen und Arbeitskräften sowie Qualitätskontrollen. In der letzten Kategorie sind Aufgaben angesiedelt, die primär mechanische Fähigkeiten erfordern. Zu diesen gehören Handwerkstätigkeiten sowie Arbeiten in der Gastronomie oder Landwirtschaft. In den Abb. 10, 11 und 12 werden die Ergebnisse der oben beschriebenen Analyse für verschiedene Berufe dargestellt.

Abbildung 10: Berufe mit dem höchsten Potenzial für augmentation



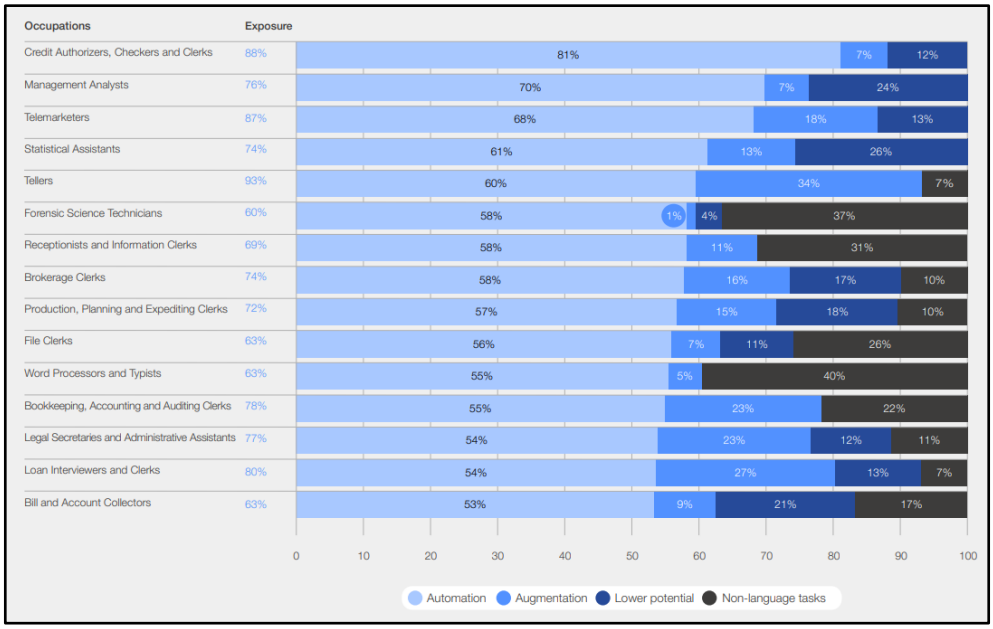
Quelle: WEF (2023), S.11.

Abbildung 11: Durch LLMs wenig betroffene Berufe



Quelle: WEF (2023), S.12

Abbildung 12: Berufe mit dem höchsten Potenzial für Automatisierung



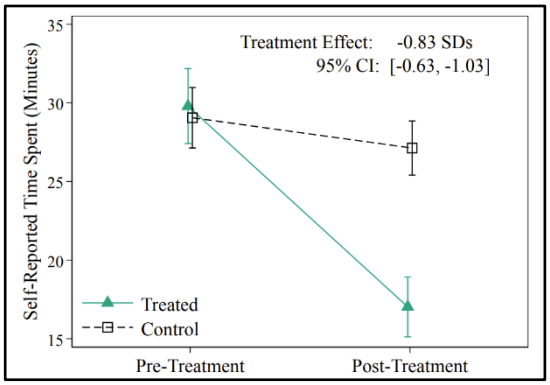
Quelle: WEF (2023), S.10

7.2 Effekte auf menschliche Arbeit

Aber hat der Einsatz von LLMs tatsächlich positive Auswirkungen auf die Qualität von Ergebnissen und lassen sich Steigerungen von Effizienz und Effektivität in der Praxis nachweisen? Diese Frage wurde in Noy/Zhang (2023), S. 2ff. untersucht. Zur Beantwortung wurde ein Experiment mit über 400 Teilnehmern, die alle in Berufen mit einer mindestens mittelhohen language depency tätig waren, durchgeführt. Im Rahmen des Versuchsaufbaus wurden sie in zwei Gruppen aufgeteilt und erhielten den Auftrag jeweils zwei dem Berufsalltag ähnliche Aufgaben zu erledigen. Die erste Aufgabe musste von allen Teilnehmern ohne die Hilfe von KI erledigt werden, während für die zweite ChatGPT von einer der beiden Gruppen unterstützend genutzt werden konnte. Anschließend wurden die Bearbeitungsgeschwindigkeit, die Qualität der Ergebnisse sowie die Qualitätsunterschiede zwischen den Teilnehmern und die job-satisfaction, auf Deutsch Befriedigung durch die eigene Arbeit, untersucht.

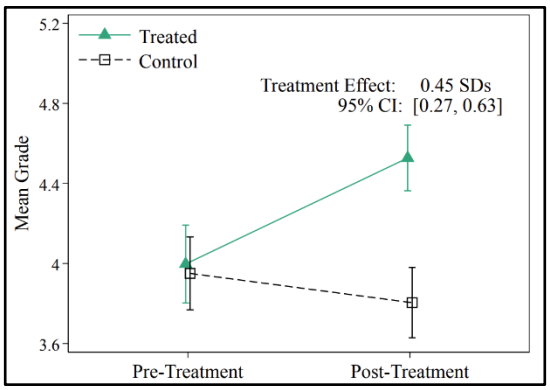
Wie in Abb. 13 (Treatment = mit KI-Unterstützung, Control = ohne Einsatz von KI) zu erkennen ist, konnte der durchschnittliche Zeitaufwand durch den Einsatz von künstlicher Intelligenz spürbar reduziert werden. Mitglieder der Treatmentgruppe lagen mit einer gemessenen mittleren Zeitersparnis von 37% für die zweite Aufgabe deutlich unter der Kontrollgruppe. Das Aufgaben mit Hilfe von LLMs schneller erledigt werden können, ist also erwiesen. Aber auch in Bezug auf die Qualität der Ergebnisse konnten merkliche Verbesserungen durch den Einsatz von ChatGPT erzielt werden. Zur Qualitätsmessung wurden für die Resultate von unabhängigen Juroren Noten auf einer Skala von eins bis sieben vergeben, wobei sieben die bestmögliche Bewertung war. Während sich die Kontrollgruppe in den Ergebnissen sogar verschlechterte, konnte sich die Treatmentgruppe im Durschnitt um 0,75 Notenpunkte verbessern, wie in Abb. 14 zu erkennen ist. In Abb. 15 lässt sich

Abbildung 13: Benötigte Zeit zur Lösung der Aufgabenstellungen



Quelle: Noy/Zhang (2023), S. 4

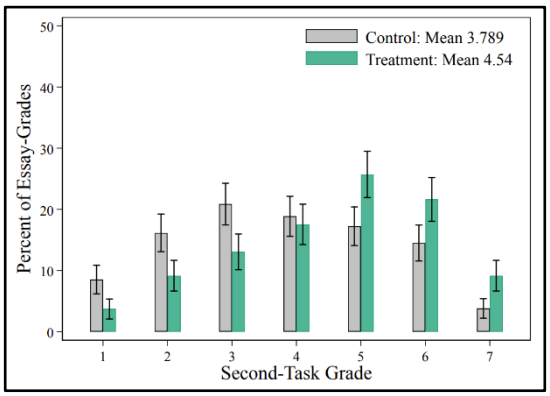
Abbildung 14: Durchschnittsnote beider Aufgabenstellungen



Quelle: Noy/Zhang (2023), S. 4

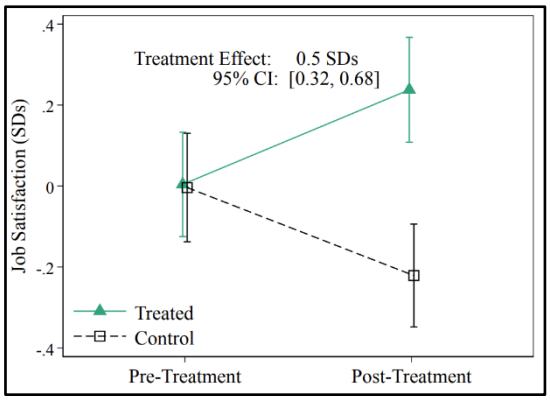
zudem ablesen, dass die Notenverteilung im Falle der Nutzung von KI linksschief ist, während die Kontrollgruppe eine rechtsschiefe Kurve erzielt. Die dritte untersuchte Dimension ist die Schwankung der Ergebnisse zwischen den Teilnehmern der jeweiligen Gruppen. Daraus lässt sich erkennen, dass der Einsatz von ChatGPT vor allem für Teilnehmer mit schlechten initialen Ergebnissen zu Leistungssteigerungen führt. Unterschiede zwischen menschlichen Fähigkeiten und Talenten können also durch den Einsatz von KI ausgeglichen werden.

Abbildung 15: Notenverteilung der zweiten Aufgabenstellung



Quelle: Noy/Zhang (2023), S. 4

Abbildung 16: Gemessene Befriedigung durch erledigte Arbeit



Quelle: Noy/Zhang (2023), S. 10

Neben Effizienz, Effektivität und Gleichheit der Teilnehmer wurde auch die Zufriedenheit mit bzw. Freude an der Arbeit bewertet. Sie ließ sich durch den Einsatz des LLM merklich steigern. Ein Grund hierfür kann die neu gewonnene Möglichkeit sein, insbesondere lästige Aufgaben durch die KI erledigen zu lassen. In Berufen, die besonders language dependent sind, ist der Einsatz von LLMs in ihrer Grundform also nachweisbar sinnvoll, sogar trotz der Tatsache, dass die KI nicht gezielt in die Arbeitsanwendungen eingebettet war, also der Informationstransfer von der Plattform der Aufgabenstellung zu ChatGPT und wieder zurück, manuell durch die Teilnehmer erledigt werden musste. Noch bessere Ergebnisse sind denkbar, wenn beispielsweise Emailpostfächer, PowerPoint, Excel, Word, Teams oder andere Anwendungen direkt mit einem LLM verknüpft sind und Zeitverluste durch den genannten händischen Informationstransfer vermieden werden können.

Selbst wenn noch nicht vollständig absehbar ist, welche weiteren Fähigkeiten LLMs in der Zukunft erlangen werden, nehmen diese bereits bei aktuellem Stand der Technik Einfluss auf unser Arbeitsleben und werden schon heute genutzt. Eine tiefere Integration ist also auch dann sehr wahrscheinlich, wenn sich in einigen Jahren ein erneuter AI-Winter einstellen sollte und sich die technische Weiterentwicklung von LLMs in einer Sackgasse verläuft. Sich der Technologie

gegenüber zu verschließen oder zu versuchen, diese einfach ignorieren, ist aus den oben beschriebenen Gründen demnach keine langfristig valide Strategie für die eigene Karriereplanung. Anpassung an die neuen Gegebenheiten ist hingegen entscheidend, um auch auf dem Arbeitsmarkt der Zukunft bestehen zu können. Dies erfordert nicht nur Wissen über die Existenz der neuen Technologie, sondern auch ein Verständnis von deren Fähigkeiten und den Auswirkungen einer breiten Implementierung. Besonders der zweite Punkt lässt sich am besten durch praktische Anwendung von künstlicher Intelligenz erlangen. Dazu kommt, dass die Fähigkeit, die Technologie richtig zu beherrschen ebenfalls eine Maßnahme ist, die eigenen beruflichen Überlebenschancen zu erhöhen. Hierfür ist es unausweichlich, eine gesunde Einstellung im Umgang mit dieser zu finden. Zudem können bestimmte Techniken dazu beitragen, die Arbeit mit KI in Bezug auf die Effektivität und Effizienz zu verbessern. Diese und weitere Punkte, die in der praktischen Arbeit mit LLMs zu beachten sind, werden in den folgenden Kapiteln behandelt.

8. Vertrauen ist gut, Kontrolle ist besser: Richtiger Umgang mit Sprach-KI

Die Nutzung von modernen LLMs ist für jeden Menschen mit einem Computer und Internetzugang nicht viel schwieriger als das reguläre Surfen im Internet. Man öffnet die Webseite der gewünschten Anwendung, registriert sich als Nutzer und kann beginnen, die ersten Befehle in das Textfeld einzutippen. Die Hürden LLMs im Alltag, sei es auf der Arbeit, für Schulaufgaben oder zur Informationsbeschaffung zu nutzen, sind also gering. Insbesondere in Kombination mit den bereits illustrierten Fähigkeiten, ist eine immer weitere Einbindung von AI in das eigene Leben attraktiv. Selbst bei der Erstellung dieser Arbeit wurden LMMs sehr erfolgreich in unterstützender Funktion eingesetzt. Es würde jedoch ein unvollständiges Bild entstehen, wenn nicht auch die Probleme, die selbst die größten Modelle wie ChatGPT plagen, Erwähnung finden.

8.1 Schwächen moderner LLMs

Während Halluzinationen Phänomene sind, die in der Regel mit Menschen in Verbindung gebracht werden, können auch LLMs darunter leiden. Selbst wenn sie nicht unbedingt mit denen vergleichbar sind, von denen manche Mitglieder unserer Spezies betroffen sind, ist der Begriff der Halluzination trotzdem im Rahmen der KI-Forschung anzutreffen. Von Ji et al. (2023), S. 4ff. stammt folgende Definition: „models (…) generate text that is nonsensical, or unfaithful to the provided source input". Es kann also vorkommen, dass ein Output inhaltlich falsch ist. Dem aufmerksamen Leser mag dies auch bereits im Kapitel zur technischen Funktionsweise von LLMs aufgefallen sein. Denn aus dem Satz „Um bei diesem Vorgang die inhaltliche Richtig- und Sinnhaftigkeit mit einer hohen Wahrscheinlichkeit zu ermöglichen (…)" geht bereits hervor, dass keine Garantie für die Richtigkeit der Ergebnisse von Sprach-KI besteht. Ji et al. weiter folgend, kann grundsätzlich zwischen zwei Arten der LLM-Halluzination unterschieden werden.

Die erste ist die intrinsische Halluzination, bei der ein Output konträr zu dem zugrunde liegenden Quellenmaterial ist. Dieser könnte also lauten, dass McDonalds seine Ursprünge in Deutschland hat, obwohl laut tatsächlicher Datengrundlage das erste Restaurant in San Bernadino, Kalifornien eröffnet wurde. Bei der extrinsischen Halluzination kann es vorkommen, dass

ein LLM-Fakten erfindet und diese dem Output hinzufügt. Das bedeutet nicht, dass diese inhaltlich falsch sein müssen, allerdings gehen die Informationen nicht aus den genutzten Quellen hervor und sind damit nicht verifizierbar. Um das Beispiel der bekannten Fast-Food Kette fortzuführen und zu verstehen was gemeint ist, ist es empfohlen an dieser Stelle folgenden Link zu öffnen (https://www.mcdonalds.com/gb/en-gb/help/faq/where-did-the-first-ever-mcdonalds-open.html) oder den QR-Code zu scannen und den kurzen Inhalt zu lesen. Das Ergebnis der KI kann basierend darauf beispielsweise lauten, dass die glücklich verheirateten Brüder Richard und Maurice McDonald das erste Restaurant eröffneten. Es mag zwar sein, dass die beiden glücklich verheiratet waren, allerdings gibt es dafür keine Quellensicherheit. Diese zweite Form der Halluzination ist die häufiger auftretende.

Abbildung 17

Quelle: eigene Darstellung

Aber selbst, wenn sich eine Aussage vollständig auf die Trainingsdaten stützt und deren Inhalte richtig wiedergibt, bedeutet dies nicht, dass darauf Verlass ist. Denn dem Prinzip „Garbage in – Garbage out“ folgend, ist jegliche von einem LLM gegebene Information selbst im allerbesten Fall nur so gut, wie die zur Verfügung gestellten Trainingsdaten. Enthalten diese Falschinformation oder Bias, kann das Modell noch so gut sein, das Ergebnis wird von der Realität abweichen und diese nur verzerrt wiederspiegeln. Sich an dem kleinen Experiment orientierend, wäre dies so, als ob man erst gar keinen Zugriff auf die Quelle hätte und damit deren Seriosität nicht einschätzen kann, noch weiß, was der tatsächliche Inhalt ist. Zudem erhält man die Informationen von jemandem, von dem man weiß, dass er sich irren kann und kann dabei auch nicht genau vorhersagen, wann seine Aussagen richtig und wann falsch sind, da die Vorgänge zwischen genutzten Quellen und Ergebnissen des LLM nicht nachvollziehbar sind. Trotz dieser großen Unsicherheiten, lässt sich eine grundsätzliche Faustregel für die Fehleranfälligkeit von LLMs aufstellen.

8.2 Umgang mit Fehleranfälligkeit von Sprach-KI

Die Fehlerquote von nN nimmt laut Marcus (2017), S. 5ff. immer dann zu, wenn diese komplexere Aufgaben bewältigen müssen, die sich weiter von den Datengrundlagen entfernen, auf die sie trainiert wurden. Was logisch klingt, ist besonders dann problematisch, wenn es für Nutzer nicht eroierbar ist, auf welche Daten das Modell nun tatsächlich trainiert wurde.

Um dies zu verdeutlichen, lässt sich das in Kapitel 4 vorgeschlagene Selbstexperiment zur Bilderkennung erneut nutzen. Wenn man beispielweise das Bild eines Hauses hochlädt und ChatGPT den Auftrag erteilt, das Objekt zu identifizieren, wird das Programm damit kein großes Problem haben. Vielleicht kann sogar der Baustil identifiziert werden und damit die Epoche, aus der es stammt. Beim Test mit Bildern des Hauses, in dem der Autor aufgewachsen ist, zeigte sich wie erwartet, dass je spezifischer die Aufgabe wurde, die Genauigkeit der Ergebnisse abnahm, bis das Modell schließlich an der Frage, in welcher Stadt das Haus liegt, scheiterte. Die Vermutung liegt nahe, dass es für diesen sehr spezifischen Fall einfach nicht trainiert war und daher keine sinnvolle Antwort geben konnte. Immerhin gab das Modell an dieser Stelle zu, keine sinnvolle Antwort zu haben. Dies ist aber bei weitem nicht immer der Fall und selbst während der Erstellung dieser Arbeit, kam es nicht selten zur Konfrontation mit inhaltlich falschen Outputs von KI, die jedoch als richtig dargestellt wurden. Bei der Nutzung nicht spezifisch trainierter LLMs kann man sich also an folgender Faustregel orientieren: Je unwahrscheinlicher es ist, dass das Modell auf ein Themenfeld trainiert wurde, umso wahrscheinlicher werden Halluzinationen. Je spezifischer eine Anfrage ist, je aktueller das Thema und je schwerer es ist, dieses objektiv, beziehungsweise rational zu bewerten, umso so stärker wird die KI gezwungen, sich von verlässlichen Datengrundlagen zu entfernen. Damit wird die Fehleranfälligkeit unweigerlich erhöht und die Ergebnisqualität im Mittel sinken.

Es zeigt sich also, dass auf KI des aktuellen Entwicklungsstandes trotz der beeindruckenden Fähigkeiten nicht unweigerlich Verlass ist. Im Umgang mit dieser ist es daher ratsam, ein gesundes und ausbalanciertes Mindset zu entwickeln, dass es erlaubt, von den Vorteilen der Technologie zu profitieren und die Nachteile und Fallstricke zu vermeiden. Wie zuvor am Beispiel der Ludditen aus England gezeigt, ist es definitiv nicht empfehlenswert, sich der Technologie zu verschließen und den Versuch zu unternehmen, deren Einzug in das eigene Leben zu ignorieren. Wie gerade dargelegt, ist es aber auch nicht empfehlenswert, sich zu 100% auf AI zu verlassen, diese für alles nur Mögliche zu nutzen und deren Ergebnisse unkritisch zu übernehmen. Die Lösung liegt auf einem Mittelweg zwischen den beiden Extremen.

Eines der entscheidenden Probleme im Umgang mit LLMs ist, dass es auch mit der erarbeiteten Faustregel nicht möglich ist, definitiv zu wissen, wann diese Fehler machen und wann nicht. Denn wie von Ji et al. (2023), S. 5ff. gezeigt, ist der Forschung bisher noch nicht bekannt, warum LLMs halluzinieren. Was zunächst nach einem allgemeinen Problem für die Frage der richtigen Nutzung von LLMs klingt, kann zeitgleich ein Teil der Lösung sein, wenn man es als einen Eckpunkt für die Ableitung von zwei zentralen Nutzungsempfehlungen ansieht:

1) Wenn die Richtigkeit einer LLM-generierten Lösung von kritischer Wichtigkeit ist, ist der blinde Verlass auf KI fehl am Platz. Es empfiehlt sich, in diesen Fällen alle faktenbasierten Ergebnisse händisch gegenzuprüfen.
2) Bei der Arbeit mit KI ist es sinnvoll, vor deren Einsatz zu überlegen, welche Fehler entstehen könnten, ob und wie sich diese vermeiden lassen und welche Auswirkungen auf das Gesamtarbeitsergebnis sich aus den möglichen Fehlern ergeben.

Um zu zeigen, wie die Anwendung dieser Regeln in der Praxis aussehen kann, wird im Folgenden aufgezeigt, wie und für welche Aufgaben LLMs für die Erstellung der vorliegenden Arbeit genutzt wurden.

Die Quellenrecherche ist eine zeitraubende Aufgabe, zumindest, wenn rein händisch erledigt. Hierfür bietet sich die Nutzung eines LLM sehr gut an. Die Möglichkeit, einer Sprach-KI, in diesem Fall ChatGPT, genau beschreiben zu können, zu welchen Themen Quellen gesucht werden und innerhalb von Sekunden eine große Auswahl an Werken hierzu zu erhalten, erwies sich als äußerst zeitsparend. Ein Risiko dabei ist das Ignorieren von Quellen, die dem System nicht bekannt sind und jenen, die aufgrund möglicher Bias nicht vorgeschlagen werden. Zudem besteht eine erhöhte Gefahr eines Confirmation Bias. Ersteres lässt sich durch eine dann rein ergänzende händische Recherche vermeiden. Letzteres entsteht aus der Fähigkeit, einfach nach Quellen suchen zu können, die bestimmte geforderte Aussagen enthalten. Um dies zu vermeiden, kann nur an die Integrität des Nutzers und dessen Fähigkeit zur kritischen Quellenreflektion appelliert werden.

Neben dem eigentlichen Auffinden der Quellen, war auch die Fähigkeit von ChatGPT PDF-Dokumente zu scannen und zusammenzufassen, sehr hilfreich, um initial zu bewerten, ob eine tiefere Auseinandersetzung mit einer Quelle überhaupt sinnvoll ist. Tatsächlich fand dabei eine Konfrontation mit dem genannten Halluzinieren statt, da mehr als einmal Inhalte in eine Quelle interpretiert wurden, die in dieser nicht vorkamen. Mögliche Risiken für das Endergebnis der Arbeit lassen sich aber vermeiden, indem man sich nicht allein auf das LLM verlässt, sondern den tatsächlichen Inhalt der Dokumente liest und versucht, diesen vollständig zu verstehen. Das dann noch verbleibende Restrisiko, sich mit sich mit einer Quelle nicht weiter zu befassen, weil die KI Inhalte in der Zusammenfassung außen vor lässt, lässt sich leider nicht komplett auflösen. Da es aber nur um die Frage einer sehr groben Orientierung ging, wurde die Problematik als weniger relevant eingeschätzt und nach Abwägung in Kauf genommen.

KI wurde ebenfalls für eine Erläuterung schwierig zu verstehender Konzepte genutzt, die sich nicht direkt aus den Erklärungen in den Originalquellen erschließen ließen. Zwar verbleibt bei dieser Art der Nutzung grundsätzlich das Risiko, dass ein LLM eine falsche Erklärung liefert. Dieses lässt sich jedoch durch das Gegenprüfen der Originalerklärung vermeiden. Die Fähigkeit von LLMs ein Konzept auf verschiedene Arten und mit variierenden Graden der Komplexität zu erläutern, stellte sich als besonders hilfreich heraus. Zudem dient diese Fähigkeit als eine weitere Möglichkeit der Kontrolle, da verschiedenen Erklärungen der KI miteinander abgeglichen und auf Abweichungen überprüft werden können.

Die Bandbreite an Aufgaben, die von LLMs übernommen werden können oder in denen diese eine unterstützende Rolle einnehmen, ist deutlich breiter als die drei hier präsentierten Beispiele. Durch das proaktive Einstellen auf die Schwächen der Technologie und die Anwendung der beiden eingangs genannten Grundsätze, kann man sich aber in nahezu allen anderen Anwendungsfällen gut vor den bestehenden Risiken schützen.

Zusammenfassend lässt sich ein gesundes Mindset für die Arbeit und den Umgang mit KI an folgender Analogie darstellen:

Ein gutes LLM ist wie ein brillanter Freund, den man jederzeit zu allem befragen kann, womit auch immer man sich gerade beschäftigt. In sehr vielen Fällen kann er mit seiner Arbeit das eigene Leben auf verschiedenste Art und Weise erleichtern. Darum sollte man, wenn es sich anbietet, nie davor zurückschrecken, seine Dienste in Anspruch zu nehmen. Aber wie ein Mensch, kann dieser Freund auch Fehler machen und sich irren. Deshalb empfiehlt es sich gerade bei faktenbasierten Aufgaben dem Grundsatz „Trust but verify“ zu folgen und sich gegenüber den möglichen Fehlern durch bedachtes Handeln präventiv abzusichern.

9. Prompt-engineering: LLMs effektiver und effizienter nutzen

Bei der Rückkehr zur Analogie des Autos, kann das gerade erarbeitete Mindset im Umgang mit LLMs wie eine Bescheinigung angesehen werden, ein Kraftfahrzeug verkehrssicher bewegen zu können. In diesem Kapitel werden nun noch Techniken und grundsätzliche Prinzipien erläutert, die dabei helfen, die eigene Effektivität und Effizienz im Umgang mit Sprach-KI zu erhöhen. Denn es ist zwar grundsätzlich möglich, sein Ziel im ersten Gang zu erreichen. Wenn man das Schalten beherrscht, wird man jedoch in der Regel schneller sein und eine komfortablere Fahrt haben.

Genauer gesagt geht es um das sog. Prompt-engineering. Prompt ist der aus dem Englischen stammende Fachbegriff für einen Befehl, den man einer KI gibt, der in der Folge zu einem entsprechenden Output führt.[25] Das Prompt-engineering hat dabei zum Ziel, bestmögliche Inputs zu entwickeln, die mit geringstem Aufwand zu optimalen Ergebnissen führen. Vor der weiteren inhaltlichen Auseinandersetzung, muss allerdings noch der Hinweis gegeben werden, dass die Informationslage zu dem Thema sehr dynamisch ist. Es erscheinen nahezu täglich neue Videos, Artikel, Blogeinträge, etc., die sich mit Prompt-engineering befassen. Um ein Mindestmaß an Aktualität zu wahren, wird sich dieses Kapitel in erster Linie auf Internetquellen stützen, an die jedoch die geforderten Qualitätsmaßstäbe angelegt wurden.

9.1 Grundsätze des Prompt-engineering

Um dem Thema gerecht zu werden, bietet sich eine inhaltliche Aufspaltung in wichtige Prinzipien des Prompting bzw. Prompt-engineering und konkrete Techniken an. Beginnend mit ersteren, wurden aus dem Quellenmaterial und eigener Erfahrungen während der Erstellung der Arbeit, die folgenden sieben Richtlinien erarbeitet:

1) Für sich häufig wiederholende Aufgaben empfiehlt es sich, sich die nötige Zeit zu nehmen, einen verlässlich funktionierenden Prompt zu entwickeln und ihn ausführlich zu testen. Dieses anfängliche Investment wird sich garantiert auszahlen, da man in der Zukunft nur noch kleine fallspezifische Anpassungen vornehmen muss, um ein gewünschtes Ergebnis zu erzielen.[26]

[25] Vgl. Bendel (2024), o.S.
[26] Vgl. OpenAI (o.J.), o.S.

2) Das bekannte Sprichwort „Probieren geht über Studieren“ ist zu einem gewissen Grad auch für das Prompt-engineering gültig. Natürlich ist es sinnvoll, sich theoretisches Grundwissen anzueignen und zu versuchen, auf dem Stand der Dinge zu bleiben, aber am besten lässt sich aus der Praxis lernen. Daher lautet die Empfehlung, möglichst viel mit KI zu arbeiten und jedes Mal zu versuchen, das eigene Prompting zu verbessern.
3) Viele Wege führen nach Rom und es gibt im Prompt-engineering kein absolutes Richtig oder Falsch. Letztlich sind die einzigen Grenzen der Kreativität die, die man sich selbst setzt. Man kann zum Beispiel ein LLM befragen, wie ein bestimmter Prompt verbessert werden kann, um neue Anregungen zu erhalten. „Outside the box“ zu denken und nie scheu zu sein, etwas Neues auszuprobieren, sollte immer ein Teil des eigenen Mindsets für die Arbeit mit Sprach-KI sein.[27]
4) Je genauer die Informationen über den Auftrag sind, umso besser wird die KI ihn ausführen können. Man kann das gut mit einem Arbeitsauftrag, den man selbst erhält, vergleichen. Je unklarer dieser ist, umso schwieriger ist es, ein zufriedenstellendes Ergebnis zu erzeugen, da die Bedingungen dafür schlichtweg unbekannt sind. Dabei geht es zum einen um den Inhalt, der die Aufgaben und Zielstellungen genau definieren soll. Zum anderen spielt das Format eine wichtige Rolle. Soll der Output Stichpunkte, ein Fließtext oder etwas ganz anderes sein? Zusammenfassend gilt: Je genauer der Prompt, umso präziser das Ergebnis, was in der Folge bedeutet, dass für Aufgaben, deren Ergebnisse nicht genau sein müssen, auch nur weniger präzise Prompts erforderlich sind.[28]
5) Auch hier trifft man erneut auf ein bereits bekanntes Prinzip: Das Herunterbrechen komplexer Aufgaben in besser zu bewältigende Teilstücke. Wenn man eine große Aufgabe mit Hilfe von LLM-Unterstützung erledigen will, empfiehlt es sich, diese vorher selbst aufzuteilen, um dem System Arbeit abzunehmen, damit dieses die eigenen Kapazitäten für optimale inhaltliche Ergebnisse nutzen kann. Die Vorgabe von einzelnen Lösungsschritten kann ebenfalls zu einer Verbesserung von Ergebnissen führen.
6) Bei regelmäßiger Arbeit mit LLMs kann es nützlich sein, auf externe Hilfe für das Organisieren und Verbessern der eigenen Prompts zurückzugreifen. Denn es gibt neben Agenta, Prompt Engine und PromptLayer noch viele weitere Beispiele für PlugIns, Bibliotheken, Foren und Programme, die dabei helfen, die eigenen Prompts zu entwickeln,

[27] Vgl. OpenAI (o.J.), o.S.
[28] Vgl. IBM (2023), o.S.

zu speichern und immer leicht zugänglich zu halten. Die Vielzahl der verfügbaren Werkzeuge erlaubt es, für nahezu jeden Bedarf ein passendes Tool zu finden.[29]

7) Prompt-engineering ist stets ein iterativer Prozess. Beim ersten Versuch den perfekten Prompt zu finden ist zwar nicht unmöglich, allerdings nicht sehr wahrscheinlich. Daher ist es sinnvoll, auch an vermeintlich fertigen Prompts weiterzuarbeiten und zu versuchen immer bessere Ergebnisse zu erzielen.[30]

9.2 Prompt-engineering Techniken

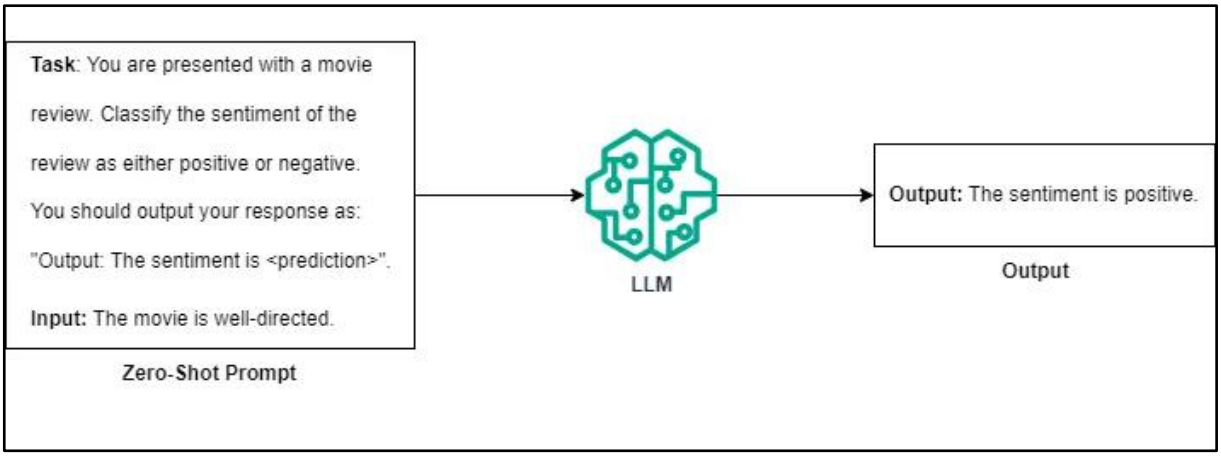
Während die obigen Richtlinien als eine grundsätzliche Orientierung beim Arbeiten und Experimentieren mit Eingabebefehlen helfen, werden nun konkrete Techniken erläutert, die bei der Arbeit mit LLMs genutzt werden können. In der englischsprachigen Literatur sind diese häufig unter dem Namen Prompting strategies bekannt. Da dieser Begriff jedoch mit der klassischen Definition von Strategien nicht übereinstimmt, wird im Rahmen dieser Arbeit weiterhin der Terminus Techniken Verwendung finden.

Eine der unkompliziertesten Techniken ist das sog. zero-shot-Prompting. Bei diesem erfolgt eine Befehlseingabe ohne das weitere Zurverfügungstellen von Beispielen, an denen sich das LLM orientieren kann. Der Prozess eines typischen zero-shot-Prompts ist in Abb. 18 dargestellt. Diese Art der Befehle eignet sich besonders dazu, ein schnelles Ergebnis zu erzielen, bei dem Genauigkeit und Qualität eine untergeordnete Rolle einnehmen. Die Schwäche dieser Methode liegt darin, dass das LLM nur mit den Instruktionen aus dem Befehl arbeiten muss und keine weiteren Referenzpunkte hat.

Für sich wiederholende Aufgaben, deren Ergebnisse eine höhere Genauigkeit erfordern, bietet es sich an, auf das few-shot-Prompting zurückzugreifen. Hierbei folgt die Befehlseingabe grundsätzlich demselben Muster wie beim zero-shot-Prompting, allerdings werden dem LLM noch eine beliebige Anzahl an Beispielen zur Verfügung gestellt, die die gewünschte Lösung der Aufgabe demonstrieren. Dies bedeutet zwar einen größeren Aufwand in der Befehlsentwicklung, kann dafür jedoch die Genauigkeit der Ergebnisse und die Nähe dieser zu einer tatsächlichen Wunschlösung erhöhen.

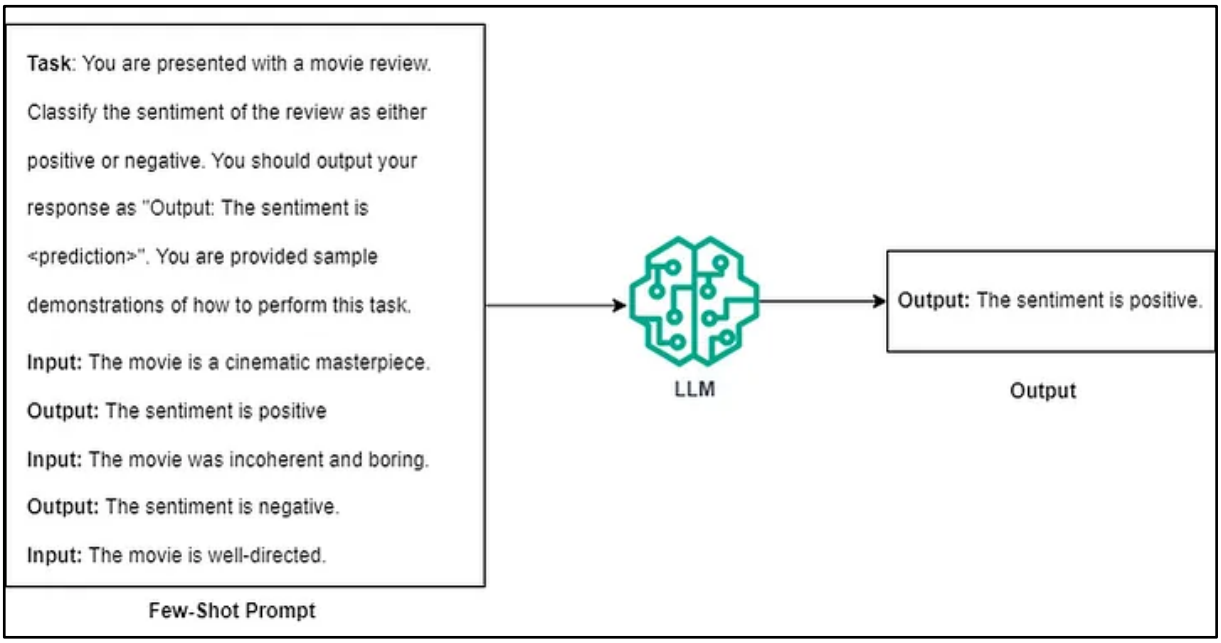
[29] Vgl. Lam (2024), o.S.
[30] Vgl. Lam (2024), o.S.

Abbildung 19: Grafische Darstellung zero-shot-Prompting



Quelle: Subramanian (2024), o.S.

Abbildung 19: Grafische Darstellung few-shot-Prompting

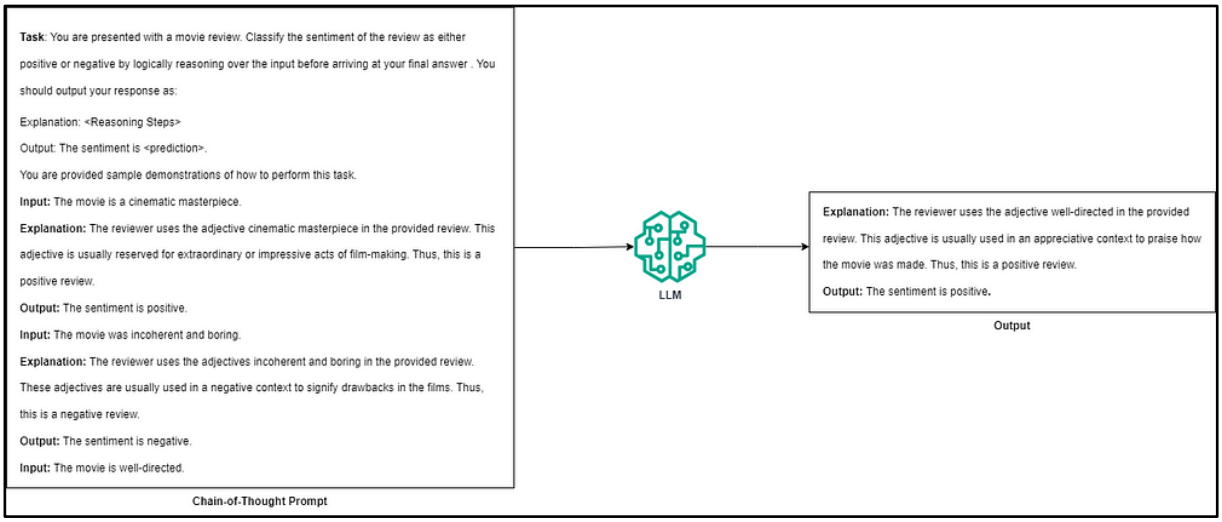


Quelle: Subramanian (2024), o.S.

Die dritte und letzte an dieser Stelle vorgestellte Methode ist das COTP oder chain-of-thought-Prompting. Diese Art der Befehlseingabe verfolgt das Ziel, die KI zu ermutigen, komplexe Aufgaben durch ein Herunterbrechen in Teilaufgaben (auch hier findet sich das Prinzip wieder) besser zu lösen. Dies geschieht durch das Verlangen einer Beschreibung des gewählten Lösungsweges wie in Abb. 20 dargestellt. Bei Tests der Technik an verschiedenen Modellen konnte in allen Fällen eine Verbesserung der Ergebnisse festgestellt werden. Allerdings waren diese nicht immer signifikant. Ein Problem der Methode liegt darin, dass ein deutlich größerer

Output erstellt werden muss. Dies führt zwangsläufig zu einer langsameren Lösungsfindung und mehr Daten, die letztlich vom Nutzer ausgewertet werden müssen.[31]

Abbildung 20: Grafische Darstellung chain-of-thought-Prompting



Subramanian (2024), o.S.

Die drei vorgestellten Techniken stellen in Kombination mit den erarbeiteten Prinzipien einen Werkzeugkasten dar, der bei richtiger Anwendung, für jeden Nutzer eine gute Hilfestellung im effektiven und effizienten Umgang mit LLMs leisten kann.

[31] Vgl. Subramanian (2024), o.S.

10. Fallstudie zu KI-unterstütztem Business Development

Nach der Erläuterung der Funktionsweisen, dem Ausarbeiten eines gesunden Mindsets für die Nutzung und dem Kennenlernen verschiedener Techniken zur Effektivitäts- und Effizienzsteigerung in der Arbeit mit KI, soll dieses Wissen nun in einem praktischen Beispiel Anwendung finden.

10.1 Unternehmensbeschreibung

Als Basis für diese Fallstudie dient die Untersuchung von Business-Development-Prozessen auf Möglichkeiten der Effizienzsteigerungen durch den Einsatz von LLMs. Das Unternehmen KH Chemicals B.V. aus den Niederlanden hat diesen Test ermöglicht und wird zur Kontextualisierung kurz vorgestellt. Es handelt sich dabei um ein Handelsunternehmen im Sektor der petro-chemischen Commodities. Neben dem Hauptsitz in der Nähe von Rotterdam, verfügt das Unternehmen über Zweigstellen in Dubai, Schanghai, New York, Mumbai, Kiew und Brasilien. Mehr als 80 Mitarbeiter sind über den Globus in einem Netzwerk aus Logistikern, Ein- und Verkäufern verteilt. Das Hauptgeschäft ist der Import von petro-chemischen Rohstoffen aus dem Nahen Osten und Asien nach Europa. Der Fokus liegt dabei auf den Produktgruppen der Weichmacher, Lösemittel und Acrylate, die an industrielle Endverbraucher weiterverkauft werden. Nach der Weiterverarbeitung finden die gehandelten Materialien in verschiedenen Endprodukten wie Farben, Lacken, Beschichtungen, Klebstoffe, Folien, Schäumen, Verpackungen oder synthetischen Harzen Verwendung. Die ökonomische Lebensberechtigung für das Unternehmen besteht unter anderem darin, dass in Europa Produktionsdefizite für viele der gehandelten Rohstoffe bestehen und die Marktregion damit auf Importe angewiesen ist. Zudem führen Faktoren, wie geringere Energie-, Rohstoff- oder Personalkosten zu einer Kostendifferenz zu heimisch hergestellten Materialien, die den interkontinentalen Warenverkehr erst rentabel macht. Zwar findet auch direkter Handel zwischen globalen Produzenten und europäischen Endnutzern statt. Allerdings ist dieser dadurch limitiert, dass es für die handelnden Akteure kostengünstiger ist, Distributoren zu bezahlen, die Wissensvorteile über die Markteilnehmer auf der Ein- und Verkaufsseite besitzen, anstelle dieses Wissen selbst aufzubauen. Um diesen Kostenvorteil zu wahren, ist es für KH Chemicals also von strategischer Relevanz, stets so effizient wie möglich zu arbeiten. Denn auch die beschriebenen Produktionsdefizite in Europa

sind nicht gleichbedeutend mit garantierten Gewinnen, da auch andere Handelsunternehmen sich im Wettbewerb um diese Importe befinden und versuchen auf den Markt zu drängen.
Im Rahmen der Fallstudie soll nun eine Möglichkeit der Effizienzsteigerung im Ausbau und Erhalt des so essenziellen Wissensvorteils auf der Verkaufsseite untersucht werden. Konkret geht es um die Frage, ob sich die Geschäftsfeld-Entwicklung mithilfe von LLMs und passend entwickelten Prompts beschleunigen und vereinfachen oder zumindest unterstützen lässt.

10.2 Versuchsaufbau

Dazu muss zunächst der Prozess des Customer-Developments von KH Chemicals analysiert werden. In diesem durchläuft ein Unternehmen und potenziell späterer Kunde drei Phasen. In der ersten wird das Unternehmen von einem Sales Manager entdeckt und grob bewertet, ob es zu einem künftigen Kunden werden könnte. Ein Farbhersteller würde umgehend weiteruntersucht werden, während einem Lebensmitteleinzelhandel keine weitere Beachtung geschenkt wird. Sollte dieser Test bestanden werden, ist das Unternehmen ein Suspect. In einem nächsten Schritt gilt es herauszufinden, ob es sich um einen End-User der von KH Chemicals geführten Produkte handelt. Der in fast allen Fällen übliche Weg, ist die Analyse der Suspect-Webseite. Oft lassen sich aus Produktbeschreibungen wie „Coatings auf Acrylatbasis“ erste Hinweise auf bestimmte Produkte oder Produktgruppen finden. Ein genaueres Ergebnis ergibt die Suche nach Sicherheitsdatenblättern. Dies sind standardisierte Dokumente, auf denen bestimmte Inhaltsstoffe mit Gefahrenpotenzial für Mensch, Tier und Umwelt aufgelistet werden. Da die allermeisten der von KH Chemicals gehandelten Rohstoffe unter diese Kategorie fallen, ist bei deren Vorhandensein eine Analyse der Sicherheitsdatenblätter unumgänglich. Sind aus diesen oder über andere Wege passende Produkte identifiziert, kann das Unternehmen kontaktiert werden und der Versuch beginnen, eine Lieferantenbeziehung aufzubauen. Nach der ersten Bestellung ist die dritte Phase erreicht und das anfänglich nur beobachtete Unternehmen wird offiziell als Kunde kategorisiert.
Ein Arbeitsschritt aus der zweiten Phase, der sich besonders für die Nutzung von LLMs anbietet, ist die Analyse der Unternehmenswebseite. Das Ziel ist dabei, die Materialverbräuche erfolgreich durch ein LLM identifizieren zu lassen, was den Sales Managern viel Arbeitszeit ersparen kann. Konkret wird für das Experiment ChatGPT zum Einsatz kommen, da es die Fähigkeit besitzt, die Inhalte von Webseiten automatisch in Inputs zu verwandeln und entspre-

chend auszuwerten. Weiterhin ist es wichtig zu erwähnen, dass der Versuch völlig ohne Vorerfahrungen vorgenommen wurde, der Ausgang zu Beginn also komplett offen und unvorhersehbar war.
Nachfolgend werden der Versuchsaufbau und die Ziele des Experiments beschrieben:
Es werden Internetseiten von zehn verschiedenen Bestandskunden untersucht, deren Materialverbräuche gesichert und bekannt sind. Die Richtigkeit der Verbrauchsdaten ist durch regelmäßige Kundengespräche im Laufe eines Geschäftsjahres sichergestellt. Diese dienen so als eine Vergleichsdatenbank, mit der sich die Ergebnisse der KI abgleichen lassen und die für den Versuch optimale Resultate darstellen. In der Bewertung der Outputs des LLM sind zwei Kriterien ausschlaggebend, die aus den eigenen Erfahrungen des Autors als Sales Manager abgeleitet wurden:

1) Vollständigkeit: Wie viele der bereits bekannten Produkte kann ChatGPT selbständig identifizieren, bzw. wie viele werden übersehen oder nicht erkannt? Diese Metrik ist essenziell, denn sollte allein ein Rohstoff identifiziert werden können, ist dies grundsätzlich bereits genug, um ein Verkaufsgespräch zu beginnen. Damit kann diese Zahl als eine absolute Untergrenze angesehen werden, um das Experiment in dieser Hinsicht als einen Erfolg verbuchen zu können. Allerdings ist es auch wichtig, die Anzahl der durch das LLM erkannten Produkte in ein Verhältnis zu den gesichert bekannten Produkten zu setzen, da bei einer zu geringen Erkennungsquote wichtige Chancen auf neue Umsatzquellen keine Beachtung finden. Anhand dieser lässt sich dann die Notwendigkeit einer nachträglichen händischen Untersuchung der Webseite bewerten.
2) Richtigkeit: Entsprechen alle erkannten Materialien den tatsächlichen Verbräuchen oder stellt die KI verschiedene Bedarfe fest, die sich nicht mit der Realität decken? Diese zweite Dimension ist wichtiger als die tatsächliche Vollständigkeit. Zusätzliche Materialverbräuche lassen sich im Zweifel durch die nachträgliche Untersuchung der Webseite ermitteln oder ergeben sich manchmal sogar direkt im Rahmen erster Verkaufsgespräche. Die Erwähnung falscher Materialien, die dem einkaufenden Unternehmen völlig fremd sind, ist hingegen deutlich kritischer. Denn durch diese wird ein oft negativer und uninformierter Eindruck vermittelt, der, wenn irgendwie möglich, vermieden werden sollte.

Mithilfe dieser beiden Achsen lässt sich relativ verlässlich bewerten, ob der Einsatz eines LLM in dem beschriebenen Anwendungsfall grundsätzlich geeignet ist oder nicht.

10.3 Erster Versuch

Direkt zu Beginn des Experimentes trat bereits ein größeres Problem auf, dass die Anzahl möglicher zukünftiger Anwendungsfälle spürbar beschränkt hat. Während ChatGPT zwar in der Lage war, die auf einer Webseite direkt als Text verfügbaren Informationen auszuwerten, war es dem Programm auch nach längerem Probieren und Reprompting nicht möglich, auf die auf einer Webseite verfügbaren Sicherheitsdatenblätter zuzugreifen. Der Grund hierfür lag darin, dass es sich bei diesen in nahezu allen Fällen um herunterladbare PDF-Dateien handelt. Zum Zeitpunkt des Experimentes (November 2024) waren der selbstständige Download und das Öffnen dieser Dateien dem LLM nicht möglich. Die anderen vorgestellten Systeme hatten ähnliche Probleme. Gemini scheiterte ebenfalls am Herunterladen der Dateien und war nicht in der Lage sinnvolle Ergebnisse aus der Webseite abzuleiten, während Claude erst gar keinen Zugriff auf externe Webseiten und deren Informationen hatte. Damit war der Einsatz der vorgestellten Modelle für den angedachten Hauptzweck deutlich eingeschränkt.
Wie bereits erwähnt, werden Sicherheitsdatenblättern soweit vorhanden immer geprüft, da sich hierüber, ohne großes Wissen über die Produkte und Prozesse der Kunden, Materialverbräuche direkt identifizieren lassen. Eine Hoffnung vor Beginn des Experimentes lag darin, dass ein LLM in der Lage sein könnte, die Datenblätter vollautomatisiert auszuwerten und mit einer relativ hohen Genauigkeit die relevanten Produkte zu identifizieren. Dies erschien als eine grundsätzlich machbare Aufgabe, da lediglich die auf den Datenblättern vorkommende Materialien mit dem Verkaufsportfolio von KH Chemicals abzugleichen gewesen wären. Es ist zwar so, dass die Datenblätter einzeln heruntergeladen und der KI zur Auswertung zugefüttert werden können, allerdings ist im Ergebnis die händische Prüfung dann zeitsparender und weniger aufwendig und die Verwendung eines LLM somit nicht gerechtfertigt.

10.4 Zweiter Versuch

Trotz dieses initialen Rückschlags haben sich noch andere Anwendungsmöglichkeiten für den Einsatz von Sprach-KI ergeben. Ein nicht insignifikanter Anteil Kundenwebseiten enthält nämlich keine Sicherheitsdatenblätter. In diesen Situationen wird der Kunde normalerweise nur auf

Basis einer sehr groben Voreinschätzung durch einen Sales Manager kontaktiert. Die genauen Verbräuche ergeben dann sich im Laufe der Zeit aus einer je nach Fall variierenden Anzahl an Verkaufsgesprächen. Das Problem besteht hierbei darin, dass die Chance irrelevante Unternehmen zu kontaktieren deutlich größer ist. Zudem ist keine gute Basis für ein Erstgespräch gegeben, was den Aufbau einer Geschäftsbeziehung oft schwieriger und langwieriger macht. In diesen Fällen besteht die Möglichkeit für einen unterstützenden Einsatz von ChatGPT, indem man dem Programm den Auftrag erteilt, mögliche Verbräuche allein basierend auf den Inhalten der Webseite zu erkennen. Sollte dies mit zufriedenstellenden Ergebnissen der Parameter Vollständigkeit und Genauigkeit gelingen, wäre dies eine sinnvolle Unterstützung in der täglichen Arbeit als Sales Manager. Der Rahmen des Experimentes wurde also leicht angepasst, während der Versuchsaufbau nicht grundlegend verändert wurde.
Der erste dann praktisch angewandte längere Prompt zur Problemlösung war:

„
1) Analyze the following website: [*insert web-address*]
2) Identify the products the company is producing
3) Which of the below mentioned chemical raw materials are likely to be compounds of the identified products:
1.4 Butanediol; 2 Ethyl Hexanol; 2 Ethyl Hexyl Acrylate; 2-Hydroxy Ethyl Acrylate;2-Hydroxy Ethyl Methacrylate; 2-Hydroxy Propyl Acrylate; 2-Hydroxy Propyl Methacrylate; Acetic Acid Glacial Tech Grade; Acetone; Acetonitrile; Acrylic Acid; Acrylic Acid Glacial; (…); White Spirit D40; White Spirit D60; White Spirit K; Xylene
4) Create a list sorted by the likelihood
“

Dieser Prompt verfolgt einen Mittelweg zwischen dem chain-of-thought und zero-shot-Prompting. In der Grundstruktur lässt sich auch das Prinzip des Herunterbrechens komplexer Aufgaben wiederfinden. Der abstrakte Auftrag, die für KH Chemicals relevanten Materialverbräuche eines willkürlichen Unternehmens zu entschlüsseln, wird durch die Vordefinierung eines Lösungsweges erleichtert. Zudem werden dem System alle nötigen Inputs und die gewünschten Outputformate vorgegeben. Die aufgeführten Produkte sind nichts anderes, als eine Liste der von KH Chemicals gehandelten Rohstoffe, die hier verkürzt dargestellt wurde, da eine vollständige Aufführung keinen Mehrwert für die Arbeit hätte.

Die Ergebnisse waren für diesen Anwendungsfall allerdings verhältnismäßig ernüchternd. Die zu dem Schluss führende Bewertung stützt sich primär auf die drei unten aufgeführten Kennzahlen:

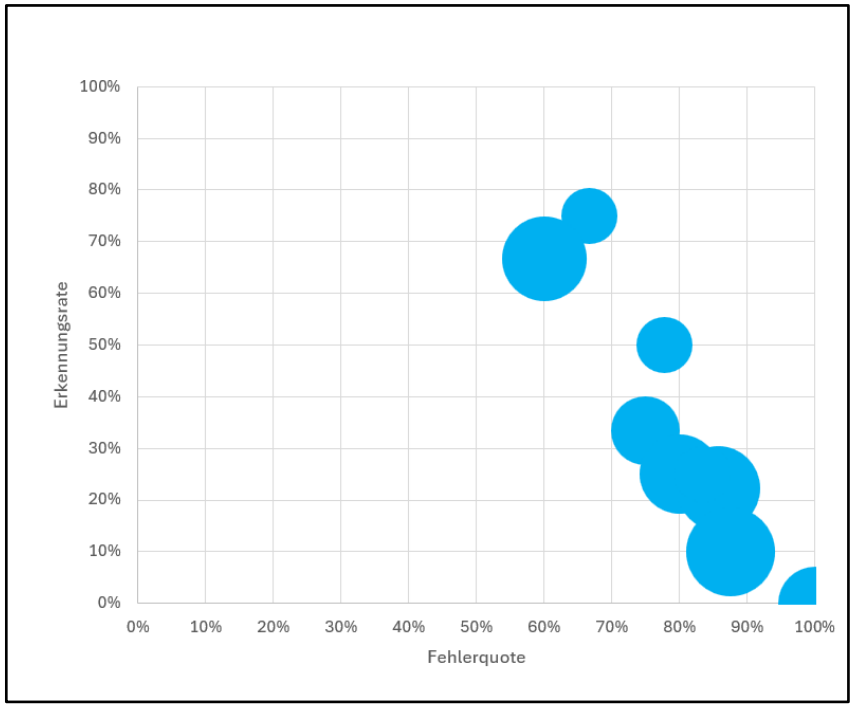
$$Erkennungsrate = \frac{Anzahl\ richtig\ erkannter\ Verbräuche}{Anzahl\ maximal\ erkennbarer\ Verbräuche} \tag{7}$$

$$Nichterkennungsrate = 1 - Erkennungsrate \tag{8}$$

$$Fehlerquote: \frac{Anzahl\ richtig\ erkannter\ Verbräuche}{Gesamtanzahl\ erkannter\ Verbräuche} \tag{9}$$

Für die Erkennungsrate ergab sich der Mittelwert von 32%, mit einem Peak von 75% und einer Bodenschwelle von 0%, die zweimal auftrat. In 20% aller Fälle wurde also kein einziger Verbrauch durch das LLM richtig identifiziert, während im Schnitt 68% der vorhandenen Verbräuche nicht erkannt wurden.

Abbildung 21: Ergebnisse des ersten Prompts (Blasengröße repräsentiert Anzahl maximal erkennbarer Produkte)

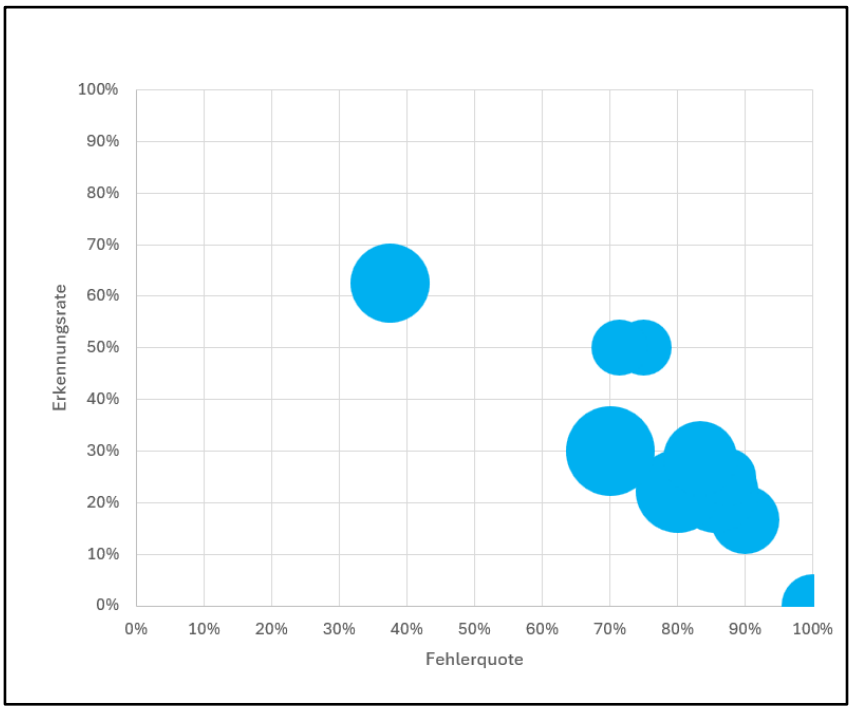


Quelle: Eigene Darstellung

Während diese Ergebnisse eine Anwendung des obigen Prompts im operativen Geschäft also bereits nahezu ausschließbar machen, wird die Entscheidung durch die Betrachtung der Fehlerquote unumgänglich. Im Schnitt lag diese bei 82%, mit einem Peak von 100% und einem Minimum von 60%. Es wurden demnach nicht nur viele richtige Materialien nicht erkannt, es war auch auf die vorhandenen Ergebnisse kein Verlass. Eine weitere Auffälligkeit war, dass das Output-Format nicht regelmäßig war und stets in Umfang und Aufbau variierte, obwohl dem LLM eine Vorgabe zum gewünschten Ergebnis vorlag.
Daraufhin wurde derselbe Prompt erneut angewendet, dabei allerdings mit folgendem Beispiel versehen und minimal in der Formulierung angepasst:
"
Example:
Input:
https://www.mipa-paints.com/en/
Output:
MIPA SE is a producer of different coatings, paints and varnishes.
Relevant products in could be:
1) Ethyl Acetate
2) Butyl Acetate
3) Xylene
4) Methoxy Propyl Acetate
"
Die Intention war es, den Prompt mit einem few-shot Element zu versehen, in der Hoffnung, dass sich dadurch die Ergebnisse in die gewünschte Richtung verbessern. Die Erkennungsrate blieb dabei im Mittelwert stabil bei 32%, wobei sich die Fälle, in denen kein einziges Produkt erkannt wurde, von zwei auf null reduzierten. Dies kann als eine zumindest leicht positive Entwicklung verbucht werden. Auch bei der Fehlerrate gab es eine leichte Verbesserung. Diese veränderte sich im statistischen Mittel um 4% von 82% auf 78%.

Trotzdem ist eine Anwendung im Alltagsgeschäft mit den genannten Werten immer noch unrealistisch. Ebenfalls interessant ist, dass die oben erwähnten, sich stetig ändernden Formate der Outputs weiterhin auftraten, obwohl dem System zusätzlich zur Vorgabe ein Beispiel für das gewünschte Ergebnis vorgelegt wurde.

Abbildung 22: Ergebnisse des zweiten Prompts (Blasengröße repräsentiert Anzahl maximal erkennbarer Produkte)



Quelle: Eigene Darstellung

Einige zuvor durchgeführte Versuche mit einfacher strukturierten Befehlen, die sich am zero-shot-Prompting orientierten, führten erst gar nicht zu verwertbaren Ergebnissen, weshalb diese hier nicht weiter aufgeführt wurden. Die Tatsache, dass trotz mehrfacher Versuche (Prompts) und verschiedenen Beispielunternehmen keine auch nur im Ansatz zufriedenstellenden Outputs erzeugt werden konnten, führt zu der Schlussfolgerung, dass die aktuell auf dem Markt verfügbaren LLMs noch nicht in der Lage sind, die gestellte Aufgabe auszuführen. Der naheliegendste Erklärungsversuch ist, dass das LLM schlichtweg nicht auf das erforderliche Wissen trainiert wurde. Das ist umso wahrscheinlicher, da es sich um sehr spezielles Branchenwissen handelt. Bei der zu erwartenden zukünftigen Entwicklung von LLMs, kann sich dies und die Fähigkeit Sicherheitsdatenblätter auszuwerten jedoch jederzeit ändern. In Kombination mit der genannten aktuellen Mangelhaftigkeit der Ergebnisse, führte dies zu dem Entschluss, das Experiment erst wieder mit der Einführung neuer ChatGPT, beziehungsweise Gemini oder Claude Versionen fortzuführen, in der Hoffnung, dann bessere Resultate erzielen zu können.

Trotzdem war das Experiment, obwohl die gewünschten Ergebnisse ausblieben, sehr hilfreich für das Unternehmen. Denn es konnte nicht nur gelernt werden, was noch nicht funktioniert. Während der Durchführung ergaben sich auch einige neue Ideen für den Einsatz von KI im operativen Geschäft. Zu diesen gehört die Recherche nach neuen Suspects, welche ähnlich gut wie im Fall der Quellen für die Bachelorarbeit funktioniert. Auch erwies sich das Beantworten vieler Kundenanfragen und Emails als sehr gut umsetzbar, solange dafür kein fallspezifischer Kontext von Nöten war. Auch konnte das Programm einige verwendbare Anregungen für die strategische Entwicklung der Marktregion D/A/CH liefern, wie zum Beispiel die Analyse des Marktes nach Produktgruppen und deren noch offenen Absatzpotenzialen. Die in der Einleitung gestellte Forschungsfrage „Lässt sich generative Sprach-KI auf dem heutigen Stand der Technik in unterstützenden Funktionen in die Geschäftsfeldentwicklungsprozesse eines modernen Handelsunternehmens sinnvoll einbinden?“ kann also trotz des gescheiterten Versuchs mit einem klaren Ja beantwortet werden.

11. Fazit

KI ist eine faszinierende Schlüsseltechnologie, die Maschinen bisher ungekannte kognitive Fähigkeiten verleiht, die es ihnen erlauben, Aufgaben zu erledigen, zu denen wir Menschen bisher nur uns selbst in der Lage sahen.

Die grundlegenden Ideen, auf denen moderne KI-Systeme beruhen, sind dabei überraschen alt. Geprägt von den Pionieren der 1940er und 50er Jahre, hängt die Entwicklung künstlicher Intelligenz dabei bis heute unweigerlich mit dem Fortschritt in der Computertechnologie zusammen, da Computer jene Maschinen sind, die die Existenz von KI erst ermöglichen. Deren Fähigkeiten stellen somit immer die Grenze dessen dar, zu dem AI selbst in der Lage war und ist. Dies reflektiert sich auch in dem ständigen Boom-Bust-Zyklus, der die Geschichte der KI seit ihren Anfängen geprägt hat.

Dabei sind die grundlegenden Prozesse, auf denen Anwendungen wie ChatGPT basieren, letztlich nichts anderes mathematische Modelle. Innerhalb dieser werden Neuronen genutzt, die ähnlich denen in menschlichen Gehirnen in der Lage sind, verschiedene Inputs zu einem Output zu wandeln. Wenn diese zu größeren Netzen verknüpft werden, können durch die Weitergabe der Outputs an andere Neuronen immer komplexere Konzepte mathematisch dargestellt werden. Diese Strukturen erlauben den Modellen dann Probleme „zu verstehen“ und Lösungsvorschläge für diese zu errechnen. Durch gigantische Strukturen mit bis zu 175.000.000.000 individuellen Neuronen und spezielle Architekturen, erlangt KI sogar die Fähigkeit menschliche Sprache nicht nur zu verstehen, sondern basierend auf den Trainingsdaten auch zu beherrschen. Dafür müssen jedoch die Parameter aller Neuronen optimal angepasst werden, was händisch schlichtweg nicht möglich ist. Daher setzt moderne AI auf Algorithmen, die diese Aufgabe übernehmen und Gewichtungen und Bias automatisch kalibrieren. Das führt wiederum dazu, dass es für uns Menschen nicht mehr nachvollziehbar ist, warum welche Parameter wie gewählt werden. Wir haben demnach eine Technologie geschaffen, deren Prozesse wir selbst nicht mehr zu 100% nachvollziehen können. Dieser Tatbestand erscheint vielen Menschen als eine Bedrohung, er sollte jedoch nicht dazu führen, dass das große Potenzial das KI bietet, nicht ausgenutzt wird.

Dass insbesondere LLMs bereits ein unweigerlicher Teil unserer Realität geworden sind, zeigt die große Auswahl an Anwendungen, die heute schon frei verfügbar sind. Die Tatsache, dass knapp ein Viertel der deutschen Arbeitnehmer bereits Anwendungen wie ChatGPT bei der Ausübung seines Berufes genutzt hat, gibt einen Hinweis darauf, dass die Technologie spürbare

Veränderungen in unserem Berufsalltag bewirken wird. Selbst wenn nicht so dramatisch wie von manchen dargestellt, wird KI als eine Automatisierungstechnologie bestimmte Berufe langfristig obsolet machen. Besonders sprachintensive und Tätigkeiten mit einem hohen Grad an Standardisierung werden absehbar davon betroffen sein. Eine geeignete Strategie, um diesem anstehenden Wandel unserer Arbeitswelt zu begegnen und langfristig auf dem Arbeitsmarkt relevant zu bleiben, ist die proaktive Anpassung.

Um für die zukünftige Veränderung unserer Gesellschaft gut aufgestellt zu sein, ist es zudem notwendig, sich tiefer mit der Technologie und den sehr großen Potentialen, die sich durch alle Lebensbereiche ziehen, zu befassen, um deren Fähigkeiten und Grenzen zu verstehen und die sich daraus ergebenden Risiken zu vermeiden und Chancen zu nutzen.

Dies erfordert eine unvoreingenommene Einstellung in der Arbeit mit KI. Sich dem technologischen Wandel zu verschließen ist dabei nicht ratsam, da man ihn als Individuum ohnehin nicht verhindern kann. Sich voll auf AI zu verlassen und diese bei jeder passenden und unpassenden Gelegenheit zum Einsatz zu bringen, ist allerdings genauso ungeeignet. Denn während gerade moderne Sprach-KI über beeindruckende Fähigkeiten verfügt, ist diese trotzdem fehleranfällig. Halluzinationen können zu falschen Ergebnissen führen und man sollte sich daher gerade bei Faktenfragen nie blindlings auf LLMs verlassen. Es ist ratsamer, Informationen gegenzuprüfen und sich vor dem Einsatz von KI gut zu überlegen, welche Fehlerrisiken existieren und wie man negative Auswirkungen dieser vermeiden kann.

Hat man dies jedoch verinnerlicht, steht einem eine faszinierende Welt offen, in der ein digitaler Begleiter jederzeit Abhilfe bei der Lösung kreativer Aufgaben schaffen kann. Um die Beziehung mit diesem Weggefährten noch zu verbessern, ist es zudem ratsam sich damit zu befassen, wie man ihn am effektivsten und effizientesten nutzen kann. Durch das Befolgen bestimmter Grundsätze und die Anwendung verschiedener Techniken, können Befehle oder Prompts entwickelt werden, die es erlauben, immer bessere Ergebnisse von einer KI zu erhalten.

Ebenso wichtig ist es, keine Angst vor dem Scheitern zu haben. Denn selbst wenn die im Rahmen der Fallstudie durchgeführten Experimente keinen direkten Erfolg hatten, ergaben sich dadurch viele interessante Lernmöglichkeiten und neue Anwendungsfälle.

Das Thema KI ist groß und sehr komplex. Sei es die Technik selbst, deren Auswirkungen auf unsere Arbeitswelt, globale Sicherheitsrisiken und Technologiewettrennen, die größere AI-Economy, etc., die Liste der behandlungswürdigen Unterthemen ist lang. Der Versuch, alle diese im Rahmen einer Bachelorarbeit unterzubringen, würde nur dazu führen, dass keinem Themenkomplex die ihm angemessene Aufmerksamkeit zukommt. Der Fokus auf Sprach-KI (und deren

Auswirkungen) wurde bewusst gewählt, da diese, wie oben aufgezeigt, bereits Einzug in unser Leben gehalten hat. Durch frei verfügbare Anwendungen wie ChatGPT, Gemini oder ClaudeAI steht jedem Leser zudem die Möglichkeit offen, die behandelten Inhalte anzuwenden und selbst damit zu experimentieren. Daher sind LLMs der wohl beste Einstiegspunkt in die umfangreiche Welt einer faszinierenden Technologie, die das Potenzial besitzt, unsere Leben langfristig fundamental zu verändern.

Die persönlichen Ansprüche des Autors an diese Bachelorarbeit sind dann erfüllt, wenn auf Seiten des Lesers ein grundsätzliches Interesse für und ein Gefühl des einsteigenden Verständnisses von künstlicher Intelligenz sowie Mut zu deren Anwendung geschaffen werden konnte.

Literaturverzeichnis

Acemoglu, Daron / Robinson, James A. (2012), *Warum Nationen scheitern. Die Ursprünge von Macht, Wohlstand und Armut.*, Berlin: FISCHER-Verlag.

Anthropic (2022), *Constitutional AI: Harmlessness from AI Feedback*,
Quelle: https://www.anthropic.com/research/constitutional-ai-harmlessness-from-ai-feedback, Zugriffsdatum: 12.11.2024.

Bendel, Oliver, Gabler Wirtschaftslexikon, *Prompt*,
Quelle: https://wirtschaftslexikon.gabler.de/definition/prompt-125087/version-390114, Zugriffsdatum: 23.11.2024.

Benzon, William, 3 Quarks Daily, (2023), *ChatGPT is a miracle of rare device.*,
Quelle: https://3quarksdaily.com/3quarksdaily/2023/02/chatgpt-is-a-miracle-of-rare-device-heres-how-im-thinking-about-it.html, Zugriffsdatum: 07.09.2024.

BMU Verlag (o.J.), *Der erste Computer der Welt: Wer war der Erfinder des Computers?*,
Quelle: https://bmu-verlag.de/der-erste-computer-der-welt/, Zugriffsdatum: 07.09.2024.

Bondarenko, Peter, Britannica (2024), *Samsung*,
Quelle: https://www.britannica.com/money/Samsung-Electronics, Zugriffsdatum: 01.10.2024.

Capterra (2024), *Reviews of ChatGPT*,
Quelle: https://www.capterra.com/p/10009334/ChatGPT/reviews/, Zugriffsdatum: 12.11.2024.

Capterra (2024), *Reviews of Claude*,
Quelle: https://www.capterra.com/p/10011218/Claude/reviews/, Zugriffsdatum: 12.11.2024.

Copeland, Brian Jack, Britannica (2024), *artificial intelligence*,
Quelle: https://www.britannica.com/technology/artificial-intelligence, Zugriffsdatum: 25.09.2024.

Field, Hayden, CNBC (2024), *How Walmart, Delta, Chevron and Starbucks are using AI to monitor employee messages*,
Quelle: https://www.cnbc.com/2024/02/09/ai-might-be-reading-your-slack-teams-messages-using-tech-from-aware.html, Zugriffsdatum: 02.11.2024.

Google (2023), *Introducing Gemini: our largest and most capable AI model*,
Quelle: https://blog.google/technology/ai/google-gemini-ai/#capabilities, Zugriffsdatum: 11.12.2024.

Hardy, Norman (o.J.), *The IBM 7030 (STRETCH),*
Quelle: https://computer-history.info/Page4.dir/pages/IBM.7030.Stretch.dir/#2, Zugriffsdatum: 07.09.2024.

IBM (2023), *Prompt-Engineering*,
Quelle: https://www.ibm.com/think#:~:text=Prompt%20engineering%20plays%20a%20role,assist%20developers%20in%20coding%20tasks, Zugriffsdatum 26.11.2024.

IBM (o.J.), *The IBM 7030, aka Stretch*,
Quelle: https://www.ibm.com/history/stretch#:~:text=The%20system's%20shortcomings%20forced%20IBM,the%20plug%20on%20the%20system, Zugriffsdatum: 10.09.2024.

International Organization for Standardization (o.J.), *What is artificial intelligence (AI)?*,
Quelle: https://www.iso.org/artificial-intelligence/what-is-ai, Zugriffsdatum: 25.09.2024.

Jackson, Matthew, Stanford (2024), *Study finds ChatGPT's latest bot behaves like humans, only better*, Quelle: https://humsci.stanford.edu/feature/study-finds-chatgpts-latest-bot-behaves-humans-only-better#:~:text=In%20the%20study%2C%20ChatGPT's%20version,have%20won%20itself%20many%20friends, Zugriffsdatum: 23.11.2024.

Jamali, Lily, BBC (2024), *OpenAI value surges to $157bn in funding deal*,
Quelle: https://www.bbc.com/news/articles/c8rd0jd1g6xo, Zugriffsdatum: 07.10.2024.

Janssen, Tom, University of Leiden (2023), *ChatGPT has left-wing bias in Stemwijzer voting advice application*, Quelle: https://www.universiteitleiden.nl/en/news/2023/03/chatgpt-has-left-wing-bias-in-stemwijzer-voting-quiz, Zugriffsdatum: 02.11.2024.

Ji, Ziwei et al. (2022), *Survey of Hallucination in Natural Language Generation*, Hongkong: ACM Computing Surveys.

Jones, Nicola, nature (2024), *'In awe': scientists impressed by latest ChatGPT model o1*,
Quelle: https://www.nature.com/articles/d41586-024-03169-9, Zugriffsdatum: 07.10.2024.

Johri, Shreya (2023), *The Making of ChatGPT: From Data to Dialogue*,
Quelle: https://sitn.hms.harvard.edu/flash/2023/the-making-of-chatgpt-from-data-to-dialogue/, Zugriffsdatum: 01.12.2024.

Kelbert, Patricia / Siebert, Julien / Jöckel, Lisa (2023), *Was sind Large Language Models? Und was ist bei der Nutzung von KI-Sprachmodellen zu beachten?*,
Quelle: https://www.iese.fraunhofer.de/blog/large-language-models-ki-sprachmodelle/, Zugriffsdatum: 16.11.2024.

Lam, Lina, Helicone (2024), *Prompt Engineering: Tools, Best Practices, and Techniques*,
Quelle: https://www.helicone.ai/blog/prompt-engineering-tools, Zugriffsdatum: 26.11.2024.

Layton, Dennis, Medium (2023), *ChatGPT – Show me the Data Sources*,
Quelle: https://medium.com/@dlaytonj2/chatgpt-show-me-the-data-sources-11e9433d57e8, Zugriffsdatum: 10.11.2024.

Marcus, Gary (2017), *Deep Learning: A Critical Appraisal*, New York: Departments of Psychology and Neural Science, New York University.

McKinsey (2023), *What is AI (artificial intelligence)?*,
Quelle: https://www.mckinsey.com/featured-insights/mckinsey-explainers/what-is-ai, Zugriffsdatum: 23.09.2024.

Nielsen, Michael (2019), *Neural Networks and Deep Learning*,
http://neuralnetworksanddeeplearning.com/chap1.html, Zugriffsdatum: 26.11.2024.

Nordhaus, William D.(2015), *Are we approaching economic singularity? Information technology and the future of economic growth*., Cambridge: National Bureau of economic research.

Noy, Shakked / Zhang, Whitney (2023), *Experimental Evidence on the Productivity Effects of Generative Artificial Intelligence*, Cambridge: MIT.

OpenAI (2022), *Introducing ChatGPT*,
Quelle: https://openai.com/index/chatgpt/, Zugriffsdatum: 12.11.2024.

OpenAI (o.J.), *Prompt engineering*,
Quelle: https://platform.openai.com/docs/guides/prompt-engineering/six-strategies-for-getting-better-results, Zugriffsdatum: 26.11.2024.

Oppy, Graham / Dowe, David, Stanford (2021), *The Turing Test*,
Quelle: https://plato.stanford.edu/entries/turing-test/#Tur195ImiGam, Zugriffsdatum: 15.09.2024.

Russel, Stuart / Norvig, Peter (2021), *Artificial Intelligence: A Modern Approach, Global Edition*, 4. Auflage, Hoboken: Pearson.

Samsung (2024), *S24 Ultra*,
https://www.samsung.com/nl/smartphones/galaxy-s24/, Zugriffsdatum: 07.09.2024.

Schlude, Antonia et al., BIDT (2023), *Verbreitung und Akzeptanz generativer KI in Deutschland und an deutschen Arbeitsplätzen*.,
Quelle: https://www.bidt.digital/publikation/verbreitung-und-akzeptanz-generativer-ki-in-deutschland-und-an-deutschen-arbeitsplaetzen/, Zugriffsdatum: 25.10.2024.

Statista (2024), *Statistiken zum Thema Computer*,
Quelle: https://de.statista.com/themen/159/computer/, Zugriffsdatum: 11.12.2024.

Subramanian, Anand, Medium (2024), *An Introduction to Prompting for LLMs*,
Quelle: https://towardsdatascience.com/an-introduction-to-prompting-for-llms-61d36aec2048, Zugriffsdatum: 26.11.2024.

Stobing, Chris, PCMAG (2024), *Google Gemini 1.0 Ultra Review*,
Quelle: https://www.pcmag.com/reviews/google-gemini-1-ultra, Zugriffsdatum: 11.12.2024.

Unbekannter Künstler (o.J.), *Ludditen zerschlagen Maschinen zur Textilherstellung*,
Quelle: https://www.worldhistory.org/Luddite/, Zugriffsdatum: 12.11.2024.

WEF (2023), *Jobs of Tomorrow: Large Language Models and Jobs*, Auflage September 2023, WEF: Cologny/Geneva.

Wittenstein, Jeran, Bloomberg (2024), *AI Can Only Do 5% of Jobs, Says MIT Economist Who Fears Crash*, Quelle: https://www.bloomberg.com/news/articles/2024-10-02/ai-can-only-do-5-of-jobs-says-mit-economist-who-fears-crash, Zugriffsdatum: 07.10.2024.

Eidesstattliche Erklärung

„Hiermit erkläre ich, dass ich die vorliegende Arbeit selbstständig, ohne fremde Hilfe und nur unter Benutzung der angegebenen Quellen und Hilfsmittel angefertigt habe. Alle Textstellen, die wörtlich oder sinngemäß aus veröffentlichten oder nicht veröffentlichten Quellen entnommen wurden, sind als solche kenntlich gemacht. Die Arbeit hat in gleicher oder in ähnlicher Form keiner anderen Prüfungsbehörde vorgelegen."

Rotterdam, 16.12.24

Ort, Datum

Unterschrift