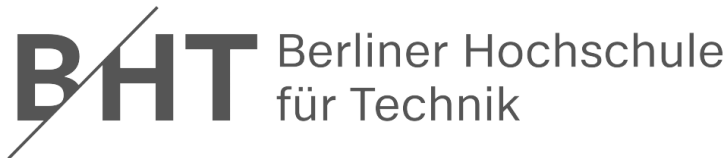


# Evaluierung und Optimierung von Large Language Models für die Entwicklung von Webanwendungen

Ein Ansatz zur Verbesserung des Entwicklungsprozesses bei Softwareprojekten



Masterthesis  
für den Master of Science Studiengang Medieninformatik

eingereicht von: Wilfried Pahl  
Matrikelnummer: 901932  
Studiengang: Online Medieninformatik  
Berliner Hochschule für Technik

betreut durch Prof. Dr. S. Edlich  
Berliner Hochschule für Technik

Temmen-Ringenwalde, der 7. Oktober 2024

## Eidesstattliche Erklärung

Hiermit erkläre ich, dass ich die vorliegende Arbeit mit dem Titel „Evaluierung und Optimierung von Large Language Models für die Entwicklung von Webanwendungen (*Ein Ansatz zur Verbesserung des Entwicklungsprozesses bei Softwareprojekten*)“ selbstständig und ohne unerlaubte Hilfe verfasst habe. Alle benutzten Quellen und Hilfsmittel sind vollständig angegeben und wurden entsprechend den wissenschaftlichen Standards zitiert.

Ich versichere, dass alle Passagen, die nicht von mir stammen, als Zitate gekennzeichnet wurden und dass alle Informationen, die ich aus fremden Quellen übernommen habe, eindeutig als solche kenntlich gemacht wurden. Insbesondere wurden alle Texte und Textpassagen anderer Autoren sowie die Ergebnisse von Sprachmodellen wie OpenAI's GPT-3 entsprechend den wissenschaftlichen Standards zitiert und referenziert.

Ich versichere weiterhin, dass ich keine anderen als die angegebenen Quellen und Hilfsmittel verwendet habe und dass ich keine Teile dieser Arbeit in anderer Form für Prüfungszwecke vorgelegt habe.

Mir ist bewusst, dass eine falsche eidesstattliche Erklärung strafrechtliche Konsequenzen haben kann.

Temmen-Ringenwalde, den 7. Oktober 2024

Unterschrift

# ABSTRACT

Abstract in Englisch.

Entwurf

## ZUSAMMENFASSUNG

Zusammenfassung in Deutsch.

Entwurf

# INHALTSVERZEICHNIS

<b>Abstract</b>	<b>i</b>
<b>Abbildungsverzeichnis</b>	<b>v</b>
<b>Tabellenverzeichnis</b>	<b>vi</b>
<b>Listings</b>	<b>viii</b>
<b>Abkürzungsverzeichnis</b>	<b>x</b>
<b>1 Einleitung</b>	<b>1</b>
1.1 Hintergrund und Kontext	1
1.2 Problemstellung	2
1.3 Zielsetzung und Forschungsfragen	2
1.4 Aufbau der Arbeit	2
1.5 Abgrenzung	3
<b>2 Grundlagen</b>	<b>5</b>
2.1 Künstliche Intelligenz	5
2.1.1 Historisches	6
2.1.2 Maschinelles Lernen	7
2.1.3 Lernparadigmen des ML	8
2.1.4 Theoretische Grundlagen	9
2.1.5 Neuronale Netze	10
2.1.6 Deep Learning	10
2.1.7 Natural Language Processing	10
2.2 Large Language Model	11
2.2.1 Grundlagen	11
2.2.2 Historie der LLM	11
2.3 Orchestrierung von LLMs	11

2.4	Multi-Agenten-Systeme . . . . .	11
2.5	Prompt Engineering . . . . .	12
2.6	Grundlagen bei der Entwicklung von Webanwendungen . . . . .	12
<b>3</b>	<b>Stand der Forschung</b>	<b>13</b>
3.1	Methoden und Ansätze . . . . .	13
3.2	Forschungslücken und zukünftige Forschung . . . . .	13
3.2.1	Identifikation von Forschungslücken . . . . .	13
3.2.2	Zukünftige Forschungsrichtungen . . . . .	13
<b>4</b>	<b>Methodik</b>	<b>15</b>
4.1	Auswahl der LLM . . . . .	15
4.2	Prompt-Engineering . . . . .	15
<b>5</b>	<b>Implementierung</b>	<b>17</b>
<b>6</b>	<b>Evaluation</b>	<b>19</b>
6.1	Einfache HTML Seite . . . . .	19
6.1.1	ChatGPT 3.5 . . . . .	19
<b>7</b>	<b>Anwendungsszenarien</b>	<b>21</b>
<b>8</b>	<b>Diskussion und Ausblick</b>	<b>23</b>
<b>9</b>	<b>Fazit</b>	<b>25</b>
	<b>Literatur</b>	<b>27</b>
	<b>Glossar</b>	<b>27</b>
	<b>Anhang</b>	<b>29</b>

# ABBILDUNGSVERZEICHNIS

2.1	Einordnung LLMs	5
2.2	KI Winterzyklen	7
2.3	Lernparadigmen des maschinellen Lernens	8

Entwurf



# TABELLENVERZEICHNIS

Entwurf

Entwurf

## LISTINGS

Entwurf

Entwurf

# ABKÜRZUNGSVERZEICHNIS

**k-NN** k-Nearest Neighbors.

**KI** Künstliche Intelligenz.

**LLM** Large Language Model.

**ML** Maschine Learning.

## 1.1 Hintergrund und Kontext

Durch die zunehmende Globalisierung und Digitalisierung wird die Gesellschaft der Gegenwart und Zukunft geprägt. Der Ausbau von Hochgeschwindigkeitsnetze und die globale Corona-Pandemie haben diese Entwicklung noch einmal beschleunigt. Immer mehr Unternehmen erkennen die Potenziale der Digitalisierung und stellen ihre Prozesse um. Ganze Wertschöpfungsketten werden auf cloudbasierte Umgebungen umgestellt. Angefangen bei der Kommunikation, über Beschaffung und Produktion bis zum Verkauf der Waren und Dienstleistungen. In allen Stufen der Prozesse kommen webbasierte Anwendungen zum Einsatz, um die Kommunikation der Anwender mit den Systemen zu ermöglichen oder Schnittstellen für die Datenübertragung zwischen den Systemen zu gewährleisten. Durch wachsende Anzahl von Web-Anwendungen wächst auch der Druck an die Entwicklungsfirmen, ihre Anwendungen den schnell wechselnden Kundenanforderungen anzupassen (Beweis fehlt).

Durch diesen Prozess getrieben, müssen Entwicklungsfirmen in immer kürzeren Release-Zyklen Softwarekomponenten hinzufügen und vorhandene erweitern. Gleichzeitig wachsen aber auch die Anforderungen an Stabilität und Sicherheit der cloudbasierten Anwendungen, sowie der Bedarf an kostengünstigeren IT-Abläufen (Beweis fehlt). Ein weiteres Problem ist der wachsende Fachkräftemangel in der Wirtschaft und die damit verbundenen steigenden Gehälter der Entwickler (Beweis fehlt).

Die Verwendung künstlicher Intelligenz bei der Programmierung gewinnt immer mehr an Bedeutung. Eine Technologie die im besonderen Maße an dieser Entwicklung beteiligt ist, sind die Large Language Models. Insbesondere mit der Veröffentlichung vom ChatGPT wurde hier ein regelrechter Hype um die LLMs ausgelöst. Diese Modelle erlauben eine Softwareentwicklung mit natürlicher

Sprache. Tiefe Kenntnisse der verwendeten Programmiersprache sind nicht mehr in dem Maße erforderlich, wie ohne LLMs.

### 1.2 Problemstellung

So groß der Hype um Künstliche Intelligenz auch sein mag, zurzeit kann KI noch nicht alles. Dies sollte auch bei der Verwendung von KI generiertem Inhalten und Code beachtet werden.

*KI denkt nicht, KI trifft keine Entscheidungen. Eine KI antwortet auf eine Eingabe nicht mit der besten Antwort, sondern mit der Wahrscheinlichsten.*

VATTENFALL ONLINE , KI für Unternehmen – die Grenzen der KI

Test Der Mensch muss die Ergebnisse prüfen, ehe generierte Programmcodestücke in vorhandene Programme eingefügt und in Produktionsumgebungen implementiert werden.

Viele Entwickler setzen auf ChatGPT zur Generierung von Code, wie eine Umfrage von stackoverflow vom Mai 2024 zeigt [1]. Gleichzeitig wachsen auch die technischen Schulden bei Softwareprojekten, da diese Modelle nicht für die Entwicklung von Software optimiert sind (Beweis fehlt).

### 1.3 Zielsetzung und Forschungsfragen

Diese Arbeit soll eine Auswahl von Modellen evaluieren und dessen Brauchbarkeit für die Softwareentwicklung aufzeigen.

Des Weiteren soll gezeigt werden, ob die automatisierte Verwendung beider Techniken eine Effizienz und Effektivität des Entwicklungsprozesses gesteigert werden kann.

### 1.4 Aufbau der Arbeit

Ein paar Worte zum Aufbau dieser Arbeit. Im Kapitel 3 wird der aktuelle Stand der Forschung vorgestellt und Erkenntnisse anderer Arbeiten diskutiert. Die in dieser Arbeit verwendeten Methoden, werden im Kapitel 4 behandelt. Die Implementierung der Test LLMs wird in Kapitel 5 besprochen und in Kapitel 6 die Ergebnisse evaluiert. Bevor in Kapitel 9 auf mögliche Folgearbeiten eingegangen wird, gibt es in Kapitel 7 Anwendungsszenarien, die zu den Ergebnissen dieser Arbeit geführt haben.

## 1.5 Abgrenzung

Ausschluss anderer Anwendungsbereiche.

Rechtliche und ethische Überlegungen werden nur am Rande berücksichtigt.

Entwurf



Entwurf

Die hier besprochenen Grundlagen gehen nicht in eine Tiefe, um alle evtl. Fragen zu klären. Jedes einzelne Gebiet könnte eine Arbeit füllen. Stattdessen sollen lediglich einen kleinen Einblick geben.

## 2.1 Künstliche Intelligenz

Die Grafik 2.1 soll die Einordnung von LLMs im Bereich der künstlichen Intelligenz zeigen.

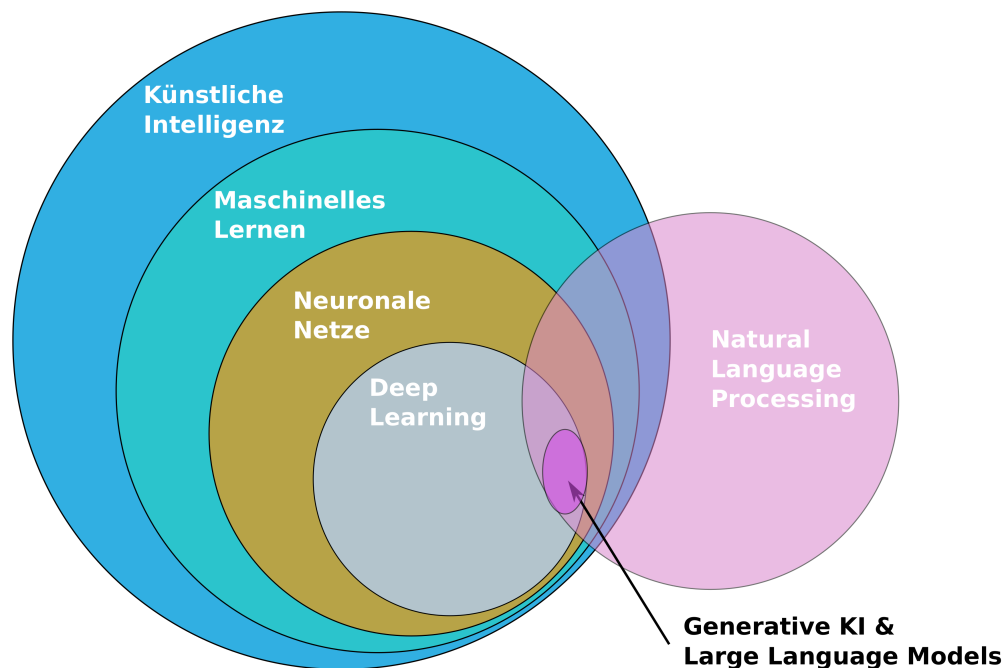


Abbildung 2.1: Einordnung LLMs

In den folgenden Kapiteln werden die wichtigsten Technologien und Begriffe erläutert.

Eine explizite Definition für künstliche Intelligenz ist zurzeit noch nicht erfolgt. Geschuldet ist diese Tatsache, da der Begriff Intelligenz nicht eindeutig definiert ist. Somit finden sich viele Versuche eine Definition für künstliche Intelligenz herzuleiten. In dieser Arbeit wird die Definition aus [2, 6 ff.] verwendet.

*Systeme der künstlichen Intelligenz (KI-Systeme) sind vom Menschen entwickelte Softwaresysteme (und gegebenenfalls auch Hardwaresysteme)<sup>3</sup>, die in Bezug auf ein komplexes Ziel auf physischer oder digitaler Ebene handeln, indem sie ihre Umgebung durch Datenerfassung wahrnehmen, die gesammelten strukturierten oder unstrukturierten Daten interpretieren, Schlussfolgerungen daraus ziehen oder die aus diesen Daten abgeleiteten Informationen verarbeiten, und über das bestmögliche Handeln zur Erreichung des vorgegebenen Ziels entscheiden. KI-Systeme können entweder symbolische Regeln verwenden oder ein numerisches Modell erlernen, und sind auch in der Lage, die Auswirkungen ihrer früheren Handlungen auf die Umgebung zu analysieren und ihr Verhalten entsprechend anzupassen.*

(Wirtschaftliche Bedeutung ..., 26 ff.).

### 2.1.1 Historisches

An dieser Stelle eine kleine historische Exkursion in der Entwicklung der künstlichen Intelligenz.

**1966: Rückschläge bei der maschinellen Übersetzung**

**1969: Auswirkungen der Perzeptron-Kritik**

**1971-75: Die Finanzierungsherausforderungen der DARPA**

**1973: Der Fallout des Lighthill-Berichts**

**1973-74: Kürzung der DARPA-Mittel**

**1987: Zusammenbruch der LISP-Maschine**

**1988: Kürzungen bei der strategischen Datenverarbeitung**

**1990er Jahre: Niedergang von Expertensystemen**

**1990er Jahre: Ende des Projekts der fünften Generation**

**KI-Frühling des frühen 21. Jahrhunderts**

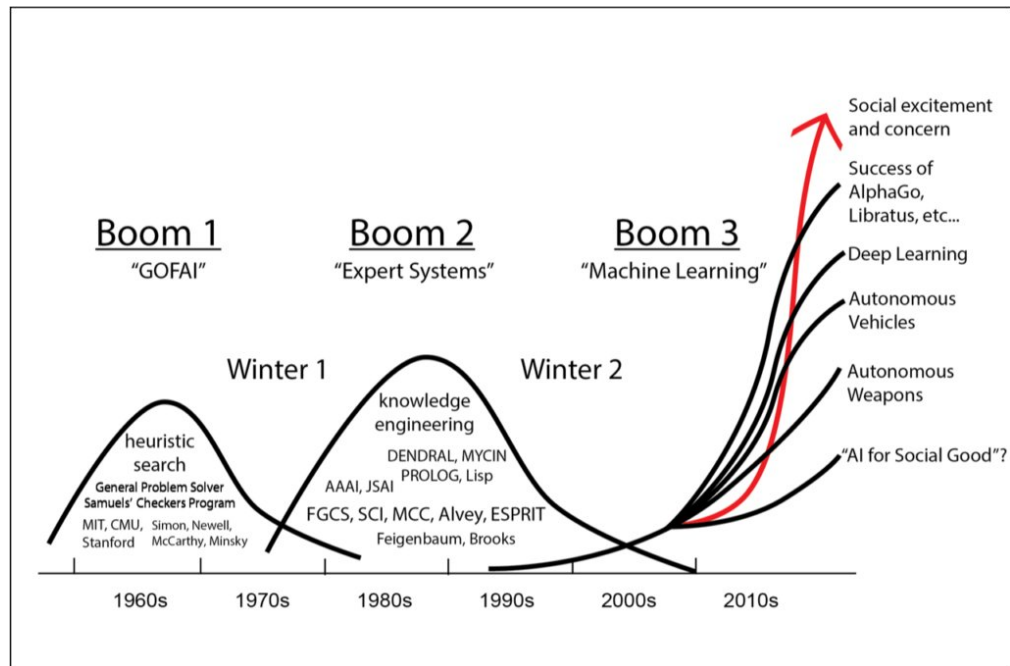


Abbildung 2.2: KI Winterzyklen

### 2.1.2 Maschinelles Lernen

Das Gebiet um maschinelles Lernen (eng. Machine Learning) ist ein Teilgebiet der Künstlichen Intelligenz. Für das maschinelle Lernen wird in dieser Arbeit die allgemein gültige Definition nach Tom M. Mitchell verwendet.

*A computer program is said to learn from experience  $E$  with respect to some class of tasks  $T$  and performance measure  $P$ , if its performance at tasks in  $T$ , as measured by  $P$ , improves with experience  $E$ .*

Ein Computerprogramm lernt aus Erfahrung  $E$  in Bezug auf eine Klasse von Aufgaben  $T$  und ein Leistungsmaß  $P$ , wenn sich seine Leistung bei Aufgaben in  $T$ , gemessen an  $P$ , mit der Erfahrung  $E$  verbessert.

Mitchell, Tom M

MACHINE LEARNING

Bei dieser Definition ist,

**E** die **Erfahrung**, die Daten aus denen das System lernt,

**T** beschreibt die **Aufgabe**, die das System erledigen soll und

**P** ist die **Leistungskennzahl** an dem der Erfolg des Systems die Aufgabe zu lösen gemessen wird.

Maschinelles Lernen (ML) und Künstliche Intelligenz (KI) sind nicht wirklich in der Lage selbstständig zu lernen oder denken, sie imitieren dies lediglich nach. Das maschinelle Lernen ist aber wohl in der Lage komplexe Muster und Funktionen in großen Datenmengen zu erkennen. Durch die Unfähigkeit zu lernen kann KI keine neuen Inhalte schaffen.

Es ist auch egal wie gut die Modelle trainiert werden, eine 100% Fehlerfreiheit gibt es nicht. Für die Praxis bedeutet dies, die Ausgaben von Modellen müssen durch Menschen immer wieder evaluiert werden, um dessen Richtigkeit sicherzustellen. Zudem sollten keine Modelle verwendet werden, wenn die Lösung eines Problems durch einen konkreten Algorithmus erfolgen kann. Hier können die Ergebnisse nachvollzogen werden und sind für den Menschen besser zu verstehen.

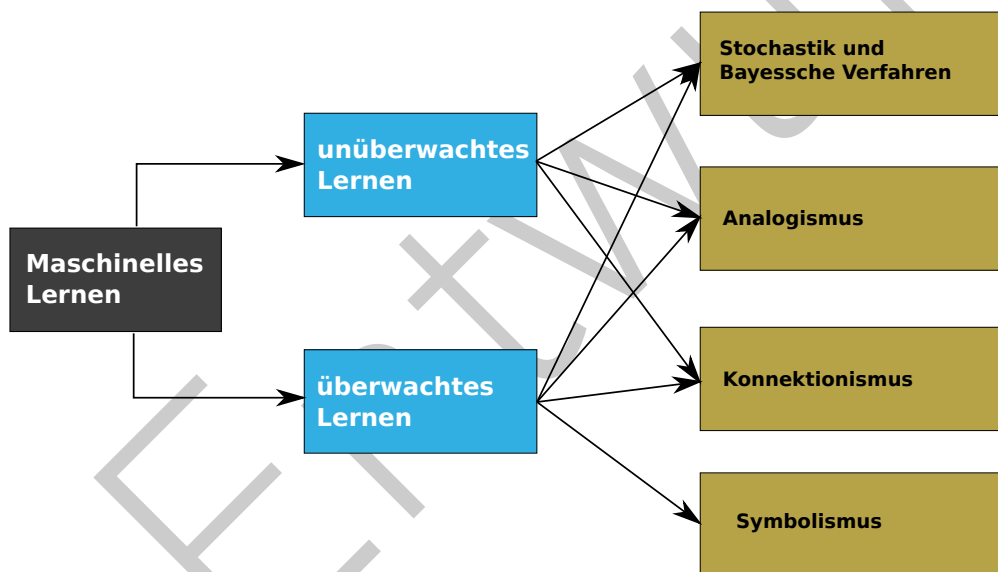


Abbildung 2.3: Lernparadigmen des maschinellen Lernens

### 2.1.3 Lernparadigmen des ML

Das maschinelle Lernen wird in zwei wichtige Formen der Lernparadigmen unterteilt. Dem überwachten und dem unüberwachten Lernen. Abbildung 2.3 zeigt die wichtigsten Lernparadigmen des maschinellen Lernens.

### Überwachtes Lernen

Bei überwachtem Lernen sind für die Eingaben der Trainingsdaten dazugehörige Ausgaben (Labels) definiert. Das Ziel ist es eine Funktion zu trainieren um künftige Eingaben korrekt klassifizieren oder vorhersagen zu können.

### Unüberwachtes Lernen

Bei unüberwachtem Lernen sind die gelabelten Ausgaben nicht vorhanden. Hierbei wird beispielsweise durch Clustering oder Dimensionsreduktion versucht Muster und Strukturen zu erkennen.

### 2.1.4 Theoretische Grundlagen

In den folgenden Kapiteln werden vier wichtige Konzepte zum maschinellen Lernen besprochen.

#### Stochastik und Bayessche Verfahren

Als Teilgebiet der Mathematik befasst sich die Stochastik mit Wahrscheinlichkeiten und zufälligen Prozessen. Auf dem Gebiet des maschinellen Lernens werden mithilfe der Stochastik Prognosen erstellt.

Bei den Bayesschen Verfahren handelt es sich um stochastische Methoden, die auf dem Bayes-Theorem basieren. Hierbei werden neue Daten berücksichtigt, um die Wahrscheinlichkeit zu aktualisieren.

Beide Methoden finden im überwachtem und unüberwachtem Lernen Anwendung.

#### Analogismus

Dieser Lernansatz sucht nach Ähnlichkeiten in den Daten. Er basiert auf der Annahme, dass ähnliche Daten ähnliche Vorhersagen oder Klassifizierungen besitzen. Dieses Modell lernt, indem neue Daten mit bekannten vergleicht und nach ähnlichen Strukturen und Mustern sucht. Ein bekanntes Verfahren für diesen Lernansatz ist der k-Nearest Neighbors (k-NN).

Der Analogismus wird im überwachtem Lernen als auch im unüberwachtem Lernen angewandt, um Muster und Strukturen zuerkennen.

#### Konnektionismus

Der Lernansatz des Konnektionismus beruht auf kleine Einheiten die miteinander verbunden sind. Diese werden als Neuronen bezeichnet, die den Nervenzellen von Organismen nachempfunden sind. Die

Künstlichen neuronalen Netze sind die bekanntesten Vertreter auf denen auch Deep Learning Modelle basieren.

Auch dieser Lernansatz wird im unüberwachten und überwachten Lernen angewandt.

### Symbolismus

Anders als beim Konnektionismus arbeiten die Einheiten beim Symbolismus mit explizite formale Regeln und Symbole, um das Wissen darzustellen. Der Symbolismus ist weniger flexible im Umgang mit unvollständigen Datensätzen und Unsicherheiten. Daher hat dieser Lernansatz weniger Relevanz als der Konnektionismus.

Dieser Lernansatz findet im überwachten Lernen Anwendung, als Beispiel sind hier Entscheidungsbäume zu nennen.

### 2.1.5 Neuronale Netze

KNN

#### Neuronen im neuronalen Netz

Neuronen

#### Arten der neuronalen Netzen

KNN Arten

#### Lernprozess im neuronalen Netz / Training

Training

### 2.1.6 Deep Learning

DL

### 2.1.7 Natural Language Processing

NPL

## 2.2 Large Language Model

Large Language Model

### 2.2.1 Grundlagen

Grundlagen

#### Tokenisierung

Token

#### Embedding

Embedding

#### Vorhersage

Transformer

#### Dekodierung

Dekodierung

### 2.2.2 Historie der LLM

Historie

## 2.3 Orchestrierung von LLMs

Orchestrierung

## 2.4 Multi-Agenten-Systeme

Multi-Agent-System



## 2.5 Prompt Engineering

Prompt

## 2.6 Grundlagen bei der Entwicklung von Webanwendungen

Webanwendung

Entwurf

# STAND DER FORSCHUNG

## 3.1 Methoden und Ansätze

## 3.2 Forschungslücken und zukünftige Forschung

### 3.2.1 Identifikation von Forschungslücken

### 3.2.2 Zukünftige Forschungsrichtungen

Entwurf

# METHODIK

## 4.1 Auswahl der LLM

## 4.2 Prompt-Engineering

Entwurf

Entwurf

# IMPLEMENTIERUNG

Entwurf

Entwurf

# EVALUATION

## 6.1 Einfache HTML Seite

### 6.1.1 ChatGPT 3.5

Entwurf



Entwurf

## ANWENDUNGSSZENARIEN

7

Entwurf

Entwurf

## DISKUSSION UND AUSBLICK

Entwurf

Entwurf

FAZIT

9

Entwurf

Entwurf

# LITERATUR

- [1] Erin Yepis. *Developers want more, more, more: the 2024 results from Stack Overflow's Annual Developer Survey*. 24. Juli 2024. URL: <https://stackoverflow.blog/2024/07/24/developers-want-more-more-more-the-2024-results-from-stack-overflow-s-annual-developer-survey/> (besucht am 09.08.2024).
- [2] Pekka Ala-Pietilä u. a. *Eine Definition der KI: Wichtigste Fähigkeiten und Wissenschaftsgebiete*. 5. März 2019. URL: [https://elektro.at/wp-content/uploads/2019/10/EU\\_Definition-KI.pdf](https://elektro.at/wp-content/uploads/2019/10/EU_Definition-KI.pdf) (besucht am 10.09.2024).



Entwurf

ANHANG

Entwurf