



Hochschule Neu-Ulm  
University of Applied Sciences

Bachelorarbeit im Studiengang

**Betriebswirtschaft**

an der Hochschule für angewandte Wissenschaften Neu-Ulm

**Untersuchung von Gestaltungsmöglichkeiten für Online-Prüfungsfragen zur  
Prävention von Betrug durch sprachverarbeitende KI-Systeme wie ChatGPT**

Erstkorrektor/-in: Prof. Dr. Sonja Köppl

Zweitkorrektor/-in: Prof. Dr. Marlon Füller

Verfasser/-in: Danny Barretta (Matrikel-Nr.: 276500)

Thema erhalten: 15.06.2023

Arbeit abgegeben: 15.10.2023

# Inhaltsverzeichnis

Abbildungsverzeichnis.....	i
Abkürzungsverzeichnis.....	ii
1. Einleitung .....	1
1.1. Problemstellung und Zielsetzung.....	2
1.2. Forschungsfrage .....	2
1.3. Methodik .....	2
2. Grundlagen .....	3
2.1 Einführung Künstliche Intelligenz.....	3
2.1.1 Künstliche Intelligenz.....	3
2.1.2 Maschinelles Lernen.....	4
2.1.3 Neuronale Netzwerke.....	5
2.1.4 Natural Language Processing .....	7
2.2 Automatisiertes KI-Modell: ChatGPT .....	7
2.2.1 Funktionsweise .....	7
2.2.2 Transformer Architektur.....	9
2.2.3 Training von ChatGPT.....	10
2.2.4 Stärken und Limitierungen .....	11
3. Technische Analyse.....	12
3.1 Beantwortung der Online-Prüfungsfragen.....	12
3.2 Bewertungsschema für ChatGPT-Prompts in Online-Prüfungen.....	20
4. Maßnahmen zur Betrugsprävention .....	24
4.1 Verwendung von Bildern zur Wissensabfrage.....	24
4.1.1 Bildinterpretation.....	24
4.1.2 Bildassoziation.....	25
4.1.3 Schrittweise Lösung mittels Bilder.....	26
4.1.4 Bild-basierte Fallstudien.....	27
4.2 Verwendung von Videomaterial zur Wissensabfrage.....	27
5. Fazit .....	29
6. Literaturverzeichnis.....	iii

## Abbildungsverzeichnis

Abbildung 2-1 Beispiel Modell eines Neuronalen Netzwerks ( <a href="https://datasolut.com/neuronale-netzwerke-einfuehrung/">https://datasolut.com/neuronale-netzwerke-einfuehrung/</a> ).....	6
Abbildung 2-2 Beispiel Modell Transformer Struktur (Vaswani 2017, S. 3).....	9
Abbildung 3-1 Startseite ChatGPT / Eingabeaufforderung ( <a href="https://chat.openai.com/">https://chat.openai.com/</a> ).....	12
Abbildung 3-2 Prüfungsfrage Moodle Kurs IT-Management II.....	13
Abbildung 3-3 ChatGPT / Beantwortung der Prüfungsfrage ( <a href="https://chat.openai.com/">https://chat.openai.com/</a> )....	13
Abbildung 3-4 Prüfungsfrage Moodle Kurs IT-Management II.....	14
Abbildung 3-5 ChatGPT / Beantwortung der Prüfungsfrage ( <a href="https://chat.openai.com/">https://chat.openai.com/</a> )....	15
Abbildung 3-6 Prüfungsfrage Moodle Kurs IT-Management II.....	16
Abbildung 3-7 ChatGPT / Beantwortung der Prüfungsfrage ( <a href="https://chat.openai.com/">https://chat.openai.com/</a> )....	17
Abbildung 3-8 Zwischenprompt zur Lösung der Aufgabe ( <a href="https://chat.openai.com/">https://chat.openai.com/</a> ).....	17
Abbildung 3-9 ChatGPT / Beantwortung der Prüfungsfrage ( <a href="https://chat.openai.com/">https://chat.openai.com/</a> )....	18
Abbildung 3-10 Prüfungsfrage Moodle Kurs IT-Management II.....	18
Abbildung 3-11 ChatGPT / Beantwortung der Prüfungsfrage ( <a href="https://chat.openai.com/">https://chat.openai.com/</a> )..	19
Abbildung 3-12 Prüfungsfrage Moodle Kurs IT-Management II.....	19
Abbildung 3-13 ChatGPT / Beantwortung der Prüfungsfrage ( <a href="https://chat.openai.com/">https://chat.openai.com/</a> )..	20
Abbildung 3-14 Bewertungsschema ( <a href="https://chat.openai.com/">https://chat.openai.com/</a> ).....	28
Abbildung 4-1 Schrittweise Darstellung einer möglichen Video Prüfungsfrage ( <a href="https://www.vyond.com/">https://www.vyond.com/</a> ).....	28
Abbildung 4-2 Schrittweise Darstellung einer möglichen Video Prüfungsfrage ( <a href="https://www.vyond.com/">https://www.vyond.com/</a> ).....	28
Abbildung 4-3 Schrittweise Darstellung einer möglichen Video Prüfungsfrage ( <a href="https://www.vyond.com/">https://www.vyond.com/</a> ).....	28
Abbildung 4-4 Schrittweise Darstellung einer möglichen Video Prüfungsfrage ( <a href="https://www.vyond.com/">https://www.vyond.com/</a> ).....	28

## Abkürzungsverzeichnis

GPT	Generative Pretrained Transformer
KI	Künstliche Intelligenz
NLP	Natural Language Processing
ReLU	Rectified Linear Unit
CNN	Convolutional Neural Network
RNN	Rekurrentes Neuronales Netzwerk
NER	Named-Entity recognition
SCF	Single Choice Fragen

## 1. Einleitung

In den letzten Jahren hat die Digitalisierung des Bildungswesens, insbesondere an Universitäten und Hochschulen, tiefgreifende Auswirkungen auf die Struktur und den Ansatz traditioneller Bildungssysteme gehabt. Diese Transformation wurde durch technologische Innovationen beschleunigt und während der Covid-Pandemie durch Plattformen wie Moodle deutlich verstärkt. Der rasche Übergang zu solchen Online-Lernplattformen hat es ermöglicht, dass Wissen nicht nur in physischen Klassenzimmern, sondern auch in virtuellen Räumen vermittelt wird, und gewährte Schülern und Pädagogen weiterhin einen qualitativ hochwertigen Bildungszugang trotz physischer Distanzierungsmaßnahmen.

Allerdings hat dieser schnelle Wechsel auch Herausforderungen in Bezug auf die Integrität von Online-Prüfungen mit sich gebracht (vgl. Aleksei Egorov 2020). Insbesondere werden die Glaubwürdigkeit und die Validität der Ergebnisse von Online-Prüfungen, die über Moodle abgelegt wurden, von vielen in Frage gestellt (vgl. Hernandez, Sorokina 2021). Hinzu kommt, dass mit dem Aufstieg von Technologien im Bereich der Künstlichen Intelligenz, insbesondere des Natural Language Processing, Tools wie ChatGPT existieren, die Antworten auf eine Vielzahl von Fragestellungen bereitstellen können.

Es ist daher unerlässlich, die Problematik rund um die Glaubwürdigkeit von Online-Assessments mit einem kritischen und wissenschaftlichen Ansatz zu betrachten und Lösungsansätze zu finden, um die Integrität des Bildungsprozesses in einer digitalisierten Welt sicherzustellen.

## **1.1 Problemstellung und Zielsetzung**

Die zunehmende Leistungsfähigkeit von KI-Systemen, insbesondere ChatGPT, stellt die Authentizität und Glaubwürdigkeit von Online-Prüfungen in Frage. Es wird befürchtet, dass Studierende diese Werkzeuge nutzen könnten, um Prüfungsantworten zu erhalten, was die Integrität des Prüfungsprozesses untergraben würde.

Das Hauptziel dieser Arbeit ist es, die möglichen Auswirkungen von ChatGPT auf Online-Prüfungen zu untersuchen. Es wird auch versucht, Strategien zu identifizieren und vorzuschlagen, um die Integrität von Online-Prüfungen im Zeitalter fortgeschrittener KI-Modelle zu gewährleisten.

## **1.2 Forschungsfrage**

In Anbetracht der Fähigkeiten von ChatGPT und ähnlichen KI-Modellen lautet die zentrale Forschungsfrage: "Macht es in der heutigen Zeit, in der fortschrittliche KI-Modelle weit verbreitet sind, noch Sinn, traditionelle Moodle Online-Prüfungen durchzuführen? Und wenn ja, wie können Prüfungsfragen gestaltet werden, um die Wahrscheinlichkeit zu verringern, dass sie von solchen Modellen beantwortet werden können?"

## **1.3 Methodik und Vorgehensweise**

Literaturrecherche: Zunächst wird eine umfassende Literaturrecherche durchgeführt. Hierbei wird ein tiefgreifender Blick in existierende wissenschaftliche Publikationen, Studien und Artikel geworfen, die sich mit der fortgeschrittenen Entwicklung von KI-Modellen und deren weitreichenden

Auswirkungen auf den Bildungsbereich und Prüfungsverfahren auseinandersetzen. Dies soll einen soliden Grundstein legen und das bisherige Wissen im Forschungsfeld zusammenfassen.

Technische Analyse: Anschließend wird eine gründliche technische Analyse von ChatGPT vorgenommen. Hierbei steht die Erkundung der technischen Fähigkeiten, Potenziale und Grenzen dieser spezifischen KI im Vordergrund. Ziel ist es, ein detailliertes Verständnis darüber zu gewinnen, inwieweit ChatGPT in verschiedenen Prüfungsszenarien Anwendung finden könnte und welche spezifischen Vorteile oder Herausforderungen sich daraus ergeben.

## **2. Grundlagen**

### **2.1 Einführung Künstliche Intelligenz**

#### **2.1.1 Künstliche Intelligenz**

Künstliche Intelligenz basiert auf Algorithmen die von Computern ausgeführt werden. Ein Hauptmerkmal der KI, das maschinelle Lernen, verwendet bestimmte Algorithmen, die Computern das Lernen aus Daten ermöglichen. Beispielsweise nutzt der "Deep Learning"-Ansatz neuronale Netzwerke, die nach dem Vorbild der menschlichen Gehirnstruktur gestaltet sind, um komplexe Zusammenhänge in umfangreichen Datenbeständen zu identifizieren. Diese Netzwerke sind aus mehreren Ebenen von Verknüpfungspunkten (oder "Neuronen") aufgebaut, die in der Lage sind, Daten sowohl linear als auch nicht-linear zu transformieren, um bestimmte Eigenschaften und Zusammenhänge zu erkennen. Mit umfangreichen Datenmengen werden diese Modelle "geschult", um ihre Präzision bei der Prognose oder Einordnung von Daten zu steigern.

Heutzutage erstreckt sich die Bandbreite der KI-Forschung und -Anwendung über diverse Bereiche (vgl. Ashish Vaswani 2017) (vgl. Richard S. 1998) (vgl. Mariusz Bojarski 2016). Neben dem maschinellen Lernen beschäftigt sich die Robotertechnik mit der Schaffung autonomer Maschinen; die Bildverarbeitung

ermöglicht Maschinen das "Erkennen" von Bildern; und die Sprachverarbeitung (NLP) versetzt Maschinen in die Lage, menschliche Sprache zu dekodieren und zu produzieren.

### **2.1.2 Maschinelles Lernen**

Statt sich auf festgelegte Algorithmen zu stützen, erlaubt das maschinelle Lernen den Maschinen, eigenständig Muster in den Daten zu identifizieren und basierend darauf Prognosen abzugeben (vgl. Bendel 2019). Es gibt verschiedene Hauptansätze im Bereich des maschinellen Lernens (vgl. Alpaydin 2022, S. 4–15)

#### Überwachtes Lernen

Beim überwachten Lernen wird ein Algorithmus darauf geschult, Vorhersagen zu treffen, indem er ein Datenset verwendet, das sowohl Eingabe- als auch Ausgabedaten enthält. Man spricht von "markierten" Daten, da jede Eingabe mit einem zugehörigen Ausgabewert versehen ist. Das primäre Ziel ist es, ein Modell zu erstellen, das fähig ist, Ausgaben für bisher nicht gesehene Daten zu ermitteln. Der Algorithmus arbeitet stetig daran, die Abweichung zwischen seinen Schätzungen und den realen Werten zu minimieren, um die bestmögliche Vorhersagemethode zu finden (vgl. Andreas Müller 2016, S. 21-83). Es existieren verschiedene Ansätze für das überwachte Lernen, darunter Techniken wie lineare Regression, logistische Regression und Support-Vektor-Maschinen. Ein anschauliches Beispiel hierfür ist die Bewertung von Hauspreisen anhand von Merkmalen wie Größe, Lage und Anzahl der Räume. Wenn der Algorithmus mit einem Datensatz von Häusern mit festgelegten Preisen trainiert wurde, kann er eingesetzt werden, um den Wert eines unbekanntes Hauses abzuschätzen. Ein kritischer Aspekt für die Genauigkeit solch eines Modells ist die Beschaffenheit und Menge der verwendeten Trainingsdaten. Je detaillierter und umfangreicher die Datenbasis, desto genauer sind die Prognosen, die das Modell liefert (vgl. Ian Goodfellow 2016, S.128-131).

## Unüberwachtes Lernen

Unüberwachtes Lernen zielt darauf ab, allgemeine statistische Muster aus Daten zu extrahieren, im Gegensatz zum überwachten Lernen, das spezifische Vorhersageaufgaben verfolgt. Ein zentrales Ziel des unüberwachten Lernens ist es, eine optimale Darstellung der Daten zu finden, die sowohl informativ als auch vereinfacht ist. Es gibt verschiedene Ansätze, um solche Darstellungen zu definieren, darunter niedrigdimensionale, spärliche und unabhängige Darstellungen. Diese Darstellungen können dazu beitragen, die zugrunde liegenden Strukturen und Beziehungen in den Daten besser zu verstehen. Deep Learning, ein Unterbereich des maschinellen Lernens, legt besonderen Wert auf die Entwicklung solcher Darstellungen, um komplexe Datenmuster effektiv zu modellieren auseinandersetzt. Es hilft, nicht offensichtliche Beziehungen und Strukturen in den Daten zu erkennen (vgl. Ian Goodfellow 2016, S.131-134) (vgl. Patel, A.A 2016, S. 1-53).

### **2.1.3 Neuronale Netzwerke**

Diese rechnerischen Strukturen der Neuronale Systeme setzen sich aus miteinander kommunizierenden Elementen, den sogenannten Knotenpunkten, zusammen. Ihre Konstruktion und Arbeitsweise orientieren sich, zumindest in der Grundidee, an den biologischen Nervenzellen.

Die Architektur solcher Systeme ist typischerweise in einer schichtweisen Struktur organisiert (siehe Abbildung 2-1). Es beginnt mit einer Eingangsebene, die Daten in Form von Vektoren empfängt. Diese Daten werden dann an die nachfolgenden Ebenen weitergegeben, wobei jede Ebene eine spezifische Übertragungsfunktion verwendet, um die Daten zu bearbeiten. Jeder Knotenpunkt in einer Ebene erhält eine Netzeingabe, die auf Basis der gewichteten Pfade von den vorherigen Knotenpunkten und einem Schwellenwert berechnet wird. Wenn diese Netzeingabe einen bestimmten Schwellenwert überschreitet, wird der Knotenpunkt aktiviert (vgl. Yoav Goldberg 2017, Kapitel 1.2).

Der Gradientenabstieg ist ein Optimierungsverfahren, das dazu verwendet wird, die Gewichtungen der Pfade so anzupassen, dass der Fehler zwischen den Vorhersagen des Netzwerks und den tatsächlichen Daten minimiert wird. Dabei wird in jedem Schritt in die Richtung des steilsten Abstiegs des Fehlergradienten gegangen, um schrittweise zum Minimum des Fehlerlandschafts zu gelangen. Dieser iterative Prozess wird solange fortgesetzt, bis ein lokales Minimum erreicht ist oder der Fehler unter einen bestimmten Schwellenwert fällt (vgl. Ian Goodfellow 2016, S.195-230).

Schließlich liefert die Ausgabeebene das Ergebnis, das entweder eine Kategorisierung oder eine kontinuierliche Vorhersage darstellen kann.

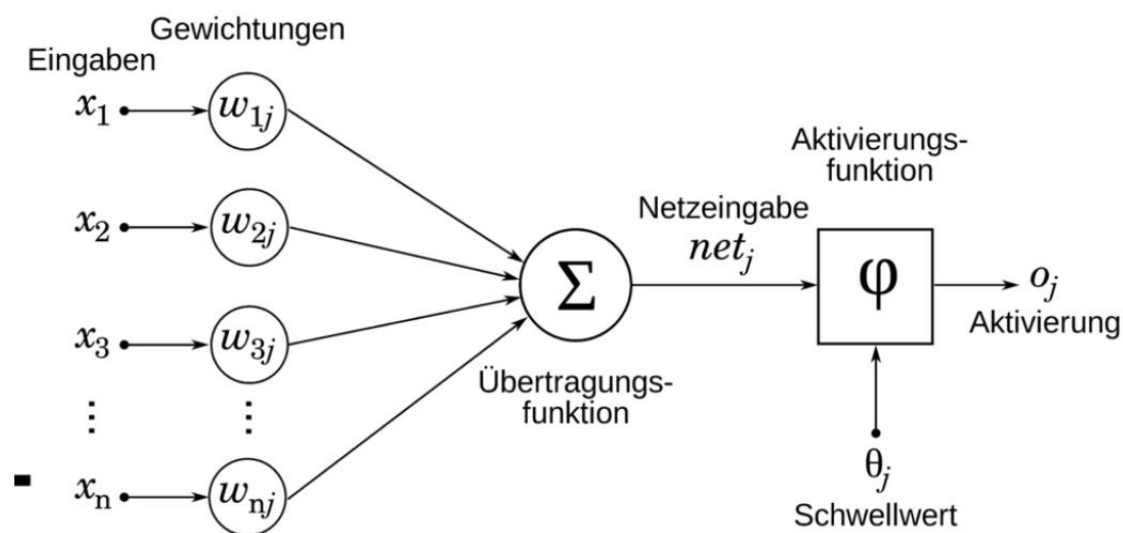


Abbildung 2-1 Beispiel Modell eines Neuronalen Netzwerks  
 (<https://datasolut.com/neuronale-netzwerke-einfuehrung/>)

Ein spezieller Sektor innerhalb der neuronalen Systeme ist das Tiefenlernen. Hierbei geht es um Systeme mit vielen internen Ebenen, die dazu fähig sind, zahlreiche Eigenschaften in den Daten zu identifizieren. Diese Tiefe erlaubt den Systemen, gestufte Dateninterpretationen zu erlernen. Während die Basis-Ebenen einfache Eigenschaften detektieren, sind die höheren Ebenen für ausgefeiltere Interpretationen zuständig. Dies zeigt sich besonders in spezialisierten Systemen wie den Faltungsnetzen (CNNs) für Bildanalysen oder den wiederkehrenden neuronalen Netzwerken (RNNs) für aufeinanderfolgende Daten (vgl. Michael Nielsen 2015, S. 20-45) (vgl. Ian Goodfellow 2016, S.195-230).

## **2.1.4 Natural Language Processing (NLP)**

NLP ist ein spezieller Zweig der künstlichen Intelligenz. Im Rahmen dieses umfangreichen Feldes strebt NLP danach, Computern das Verstehen und Erzeugen menschlicher Sprache beizubringen. Das übergeordnete Ziel von NLP ist die Schaffung einer nahtlosen Kommunikation zwischen Computern, die in binärem Code funktionieren, und Menschen, die in einer facettenreichen Sprache interagieren.

NLP nutzt die Tokenisierung und zerlegt Textpassagen in kleinere Segmente oder "Tokens", um eine tiefere Analyse zu ermöglichen. Das Stemming bringt Worte auf ihren Basiskern zurück, sodass verschiedene Varianten eines Begriffs als äquivalent betrachtet werden. Das Part-of-Speech-Tagging bestimmt die grammatikalische Funktion von Begriffen in einem Abschnitt, während die Erkennung benannter Entitäten (NER) konkrete Entitäten in einem Text herausstellt (vgl. Manning & Schütze 1999, S. 127-137).

## **2.2 Automatisiertes KI-Modell: ChatGPT**

### **2.2.1 Funktionsweise**

ChatGPT, eine Kreation von OpenAI, steht an der Spitze der KI-Entwicklung und verkörpert eine weiterentwickelte Version der GPT-Struktur (Generative Pre-trained Transformer). Bevor wir uns detailliert mit ChatGPT beschäftigen, sollten wir einen Blick auf die Grundtechnik werfen (vgl. OpenAI 2022).

Ein Transformer ist ein Modell, das in der natürlichen Sprachverarbeitung (NLP) Anwendung findet. Sein Hauptziel ist es, Verknüpfungen zwischen verschiedenen Worten oder Textsegmenten zu identifizieren, unabhängig von ihrer Stellung im gegebenen Text. Dadurch kann das Modell den Kontext begreifen und basierend darauf präzise Prognosen oder Antworten erstellen. Im Kern von ChatGPT arbeitet ein auf Transformern basierendes neuronales Netz (vgl. Ashish Vaswani 2017). Dank seiner erstaunlichen Kapazität, komplexe

Strukturen in Texten zu identifizieren und darauf aufzubauen, kann es Textreaktionen erzeugen, die dem menschlichen Ausdruck sehr nahekommen. Dieser Vorgang, bei dem das Modell Eingabeabfragen (engl. "prompt") analysiert und schlüssige Textantworten produziert, verdeutlicht sein tiefes Verständnis für menschliche Kommunikation.

ChatGPTs Trainingsverlauf besteht aus zwei Phasen: Vorabtraining und Feintuning. In der Vorabtrainingsphase wird das Modell mit enormen Mengen an Textdaten geschult, wodurch es eine Fülle von Informationen, Kontexten und stilistischen Nuancen aufnimmt. Es begreift die Struktur menschlicher Kommunikation, von einfachen grammatikalischen Strukturen bis zu komplizierten thematischen Verknüpfungen (vgl. OpenAI 2022).

Nach dieser intensiven Datensammlung tritt die Feintuning-Phase in Aktion. Hierbei wird ChatGPT gezielt für spezifische Funktionen oder Fachgebiete geschult. Es verwendet das während des Vorabtrainings gesammelte Wissen und verfeinert es, um in bestimmten Anwendungsgebieten präzise und relevante Antworten zu geben (vgl. OpenAI 2022).

## 2.4.2. Transformer Architektur

Während rekurrente neuronale Netze (RNNs) Daten sequenziell und lineare Neuronale Netze (CNNs) vorrangig für Bildverarbeitung konzipiert sind, bieten Transformer den Vorteil, Daten simultan zu bearbeiten. Dies ermöglicht es ihnen, mehrere Informationseinheiten gleichzeitig zu verarbeiten, was den Trainingsablauf erheblich beschleunigt.

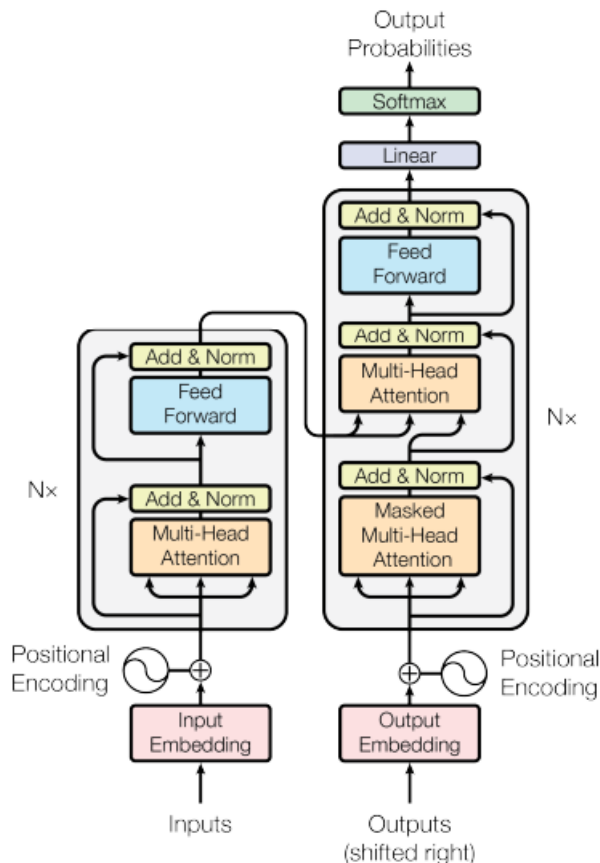


Abbildung 2-2 Beispiel Modell Transformer Struktur (Vaswani 2017, S. 3)

Ein zentrales Merkmal der Transformer-Struktur ist der Selbst-Aufmerksamkeits-Mechanismus. Er erlaubt dem System, unterschiedliche Abschnitte eines gegebenen Textes basierend auf ihrer Bedeutung für den gegebenen Kontext zu priorisieren. Man kann sich das wie einen Scheinwerfer vorstellen, der bestimmte Begriffe oder Ausdrücke in einem Satz betont, je nach ihrer Bedeutung für das Verständnis des gesamten Satzes. Dieses Feature ist besonders wertvoll, um sowohl direkte als auch entfernte Zusammenhänge in einem Text zu erkennen, was zu einer tiefgründigeren Textanalyse beiträgt (vgl. Ashish Vaswani 2017).

Die überwiegende Mehrheit der leistungsstarken Modelle für neuronale Sequenzumwandlungen setzt auf eine Struktur aus Encoder und Decoder [5, 2,

35]. Dabei konvertiert der Encoder eine gegebene Eingabesequenz von Symbolen  $(x_1, \dots, x_n)$  in eine fortlaufende Darstellungssequenz  $z = (z_1, \dots, z_n)$ . Basierend auf  $z$  erstellt der Decoder eine Ausgabesequenz  $(y_1, \dots, y_m)$  in einer schrittweisen Abfolge. In jedem dieser Schritte arbeitet das Modell autoregressiv, indem es bereits generierte Symbole zur Berechnung des nächsten Symbols heranzieht. Der Transformer nutzt diese Grundstruktur und setzt dabei auf übereinander geschichtete Selbst-Aufmerksamkeitsmechanismen und punktgenaue, vollvernetzte Ebenen für sowohl den Encoder als auch den Decoder (siehe Abbildung 2-2) (vgl. Ashish Vaswani 2017).

### **2.4.3. Training von ChatGPT**

Die Schulung von ChatGPT beruht auf einer geschickten Mischung aus unsupervised und supervised Lernmethoden. Zunächst wird das Modell durch unsupervised Learning anhand umfangreicher Textsammlungen geschult, um ihm ein breites Spektrum an Wissen und Kontext zu bieten. Diese Textsammlungen spiegeln ein vielfältiges Spektrum menschlicher Erkenntnisse und Erfahrungen wider, die aus verschiedenen Ecken des Internets gezogen werden. In Bezug auf die verwendeten Daten profitiert ChatGPT von einer beeindruckenden Vielfalt an Textressourcen. Dazu gehören nicht nur literarische Texte wie Romane und Poesie, sondern auch Informationswebsites, Blogs, Nachrichtenplattformen und diverse Diskussionsforen. Selbst Fachartikel aus verschiedenen Fachgebieten fließen ein, um dem Modell ein tiefgehendes und präzises Fachwissen zu verleihen. Dank dieser Datenreichtums kann ChatGPT eine Vielzahl von Anfragen bearbeiten, von einfachen Alltagsthemen bis hin zu anspruchsvollen wissenschaftlichen Fragestellungen. Zusätzlich wird das unsupervised Learning durch zielgerichtetes supervised Learning erweitert. In diesem Schritt wird ChatGPT mit speziell ausgewählten Daten geschult, oft in Form von Frage-Antwort-Duos oder bestimmten Gesprächen. Dieser Prozess hat das Ziel, dem Modell eine präzise Antwortfähigkeit und ein fundiertes Verständnis für menschliche Interaktionen zu vermitteln (vgl. OpenAI 2022).

#### **2.4.5. Stärken und Limitierungen**

Die erstaunliche Kapazität, menschliche Interaktionen zu simulieren, hat ChatGPT zum Mittelpunkt vieler Forschungen gemacht. Mit einem Training, das auf Milliarden von Texteinheiten basiert, kann ChatGPT auf eine breite Palette von Benutzeranfragen antworten. Seine bemerkenswerte Anpassungsfähigkeit erlaubt es ihm, an diversen Gesprächsthemen teilzunehmen und Wissen aus unterschiedlichsten Bereichen zu teilen. Ein besonderes Kennzeichen von ChatGPT ist seine Kompetenz, Inhalte zu erstellen, die erstaunlich menschenähnlich klingen. Dank modernster Technologien und Deep Learning kann es menschliche Sprachnuancen nachahmen.

Dennoch hat auch ChatGPT seine Grenzen. Manchmal kann es zu unpräzisen oder sogar irreführenden Antworten neigen, besonders wenn es mit vagen oder doppeldeutigen Fragen konfrontiert wird. Es ist ratsam, die von ihm gelieferten Daten stets kritisch zu hinterfragen. Ein weiterer wichtiger Punkt ist, dass ChatGPT, obwohl es menschliche Ausdrucksweisen gut nachahmen kann, nicht wirklich den tieferen Sinn oder den Kontext hinter den Ausdrücken versteht. Es agiert ohne Bewusstsein oder echte Erkenntnis und stützt sich lediglich auf die Daten, mit denen es gefüttert wurde. Obwohl es menschlich klingen kann, fehlen ihm Emotionen, Intuition und tiefere menschliche Perspektiven (vgl. OpenAI 2022).

### 3. Technische Analyse

#### 3.1 Beantwortung der Online-Prüfungsfragen

In der vorliegenden Untersuchung werden diverse Prüfungsfragetypen analysiert, indem versucht wird, mithilfe von ChatGPT die korrekten Antworten zu ermitteln. Die betrachteten Prüfungsfragetypen variieren in ihrer Beschaffenheit. Die Ergebnisse der jeweiligen Prüfungsfragetypen stehen repräsentativ für alle anderen Fragen des gleichen Typs, hierbei wurden bis zu 5 Fragen des gleichen Typs getestet. Zunächst wird die jeweilige Prüfungsfrage aus dem Moodle-System zusammen mit der korrekten Lösung präsentiert. Anschließend wird die Antwort, die von ChatGPT generiert wurde, dargestellt. Um die Antworten von ChatGPT zu erhalten, wurden die Fragen und zugehörigen Antwortoptionen als Eingabeaufforderung in das ChatGPT-System eingegeben.



Abbildung 3-1 Startseite ChatGPT / Eingabeaufforderung (<https://chat.openai.com/>)

## Prüfungsfragetyp: Single Choice

Unüberwachtes Lernen

Welches ist die richtige Antwort? Bitte kreuzen Sie an.

Welche der nachfolgenden Aufgabenstellungen kann mit dem k-means Algorithmus gelöst werden

Wählen Sie eine Antwort:

- a. Anhand des Nutzenverhaltens auf einer Webseite soll herausgefunden werden, welche unterschiedlichen Benutzergruppen die Webseite besuchen
- b. Die Vielzahl der eingehenden E-Mails soll nach Spam bzw. Kein-Spam unterschieden werden
- c. Ausgehend von Wetterdaten soll für morgen vorhergesagt werden, ob es sonnig, wolkig oder regnerisch sein wird
- d. Keine dieser Aufgaben kann mit dem k-means Algorithmus gelöst werden

Abbildung 3-2 Prüfungsfrage Moodle Kurs IT-Management II

Die richtige Antwort ist:

**a. Anhand des Nutzenverhaltens auf einer Webseite soll herausgefunden werden, welche unterschiedlichen Benutzergruppen die Webseite besuchen**

Abbildung 3-3 ChatGPT / Beantwortung der Prüfungsfrage (<https://chat.openai.com/>)

In der Untersuchung zur Effizienz von KI-Modellen bei der Beantwortung von SCF zeigte sich, dass das Modell in allen Fällen das richtige Ergebnis erzielte. Ein zentraler Faktor für diese Leistung ist die spezifische Architektur, auf der ChatGPT basiert. Es nutzt die Transformer-Architektur um komplexe Muster in großen Datenmengen zu identifizieren und kontextbasierte Vorhersagen zu generieren.

Die inhärente Struktur von SCF spielt ebenfalls eine entscheidende Rolle. Bei dieser Art von Fragen gibt es eine klare Antwort unter mehreren vorgegebenen Optionen. Dies reduziert die Komplexität des Auswahlprozesses für das Modell, da es lediglich die korrekte Antwort aus den gegebenen Optionen identifizieren muss.

Ein weiterer bedeutsamer Aspekt ist das umfassende Training, das ChatGPT durchlaufen hat. Durch den Zugriff auf eine vielfältige Datenbasis kann das Modell oft die korrekte Antwort extrahieren, indem es den Kontext der Frage mit den verfügbaren Antwortoptionen abgleicht. Es ist plausibel anzunehmen, dass

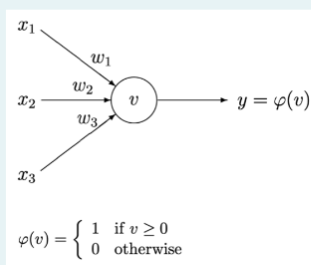
ChatGPT während seiner Trainingsphase auf eine Vielzahl ähnlicher Fragen gestoßen ist, was seine Effizienz bei der Beantwortung von SCF weiter steigert.

Die Fähigkeit den Kontext einer Frage zu erfassen, darf nicht übersehen werden. Bei SCF verwendet das Modell diesen Kontext, um die Antwortoptionen zu analysieren und die relevanteste auszuwählen. Dieser Prozess wird durch die tokenbasierte Analyse unterstützt, bei der das Modell jede Option basierend auf ihrer Struktur bewertet.

#### Überwachtes Lernen - Algorithmen

Welches ist die richtige Antwort? Bitte kreuzen Sie an.

Nachfolgendes neuronales Netz ist gegeben:



Die Gewichte werden wie folgt gewählt:  $w_1 = 2$   $w_2 = -4$   $w_3 = 1$ . Als Aktivierungsfunktion wird die Treppenfunktion mit  $t = 0$  verwendet

Wie lautet der Ausgabewert  $y$  für den Eingabe-Datensatz  $X [ 1 \ 1 \ 1 ]$ ?

Wählen Sie eine Antwort:

- a. 0
- b. 2
- c. 1
- d. 3

Abbildung 3-4 Prüfungsfrage Moodle Kurs IT-Management II

Berechnung:

$$y = w_1 * x_1 + w_2 * x_2 + w_3 * x_3$$

Mit den gegebenen Werten:

$$y = 2 * 1 + (-4) * 1 + 1 * 1$$

$$y = 2 - 4 + 1$$

$$y = -1$$

Da die Treppenfunktion mit  $t = 0$  verwendet wird, wird der Ausgabewert 0 sein, wenn die gewichtete Summe kleiner als 0 ist und 1, wenn sie größer oder gleich 0 ist.

In diesem Fall ist  $y = -1$ , also:

$$y = 0$$

Die richtige Antwort ist:

a. 0

Erneut generieren

Abbildung 3-5 ChatGPT / Beantwortung der Prüfungsfrage (<https://chat.openai.com/>)

Darüber hinaus ist die Fähigkeit von ChatGPT, Bilder zu interpretieren, zwar begrenzt, aber in Kombination mit den bereitgestellten textuellen Informationen kann das Modell oft genug Kontext extrahieren, um eine fundierte Entscheidung zu treffen. Die Bereitstellung von Details wie der Treppenfunktion mit  $t=0$ , den Gewichtungen und den Eingabe-Datensatz gibt dem Modell zusätzliche Hinweise, die es bei der Auswahl der richtigen Antwort verwenden kann.

Konfusionsmatrix - Evaluierung von Klassifikationsergebnissen:

Erstellen Sie das aus den gegebenen Labels und den Vorhersagen des Klassifikators die Konfusionsmatrix (Verkauf entspricht positiver Klasse, Nacharbeit entspricht negativer Klasse)

Datensatz	Wahres Label	Vorhergesagtes Label
1	Nacharbeit	Nacharbeit
2	Nacharbeit	Verkauf
3	Nacharbeit	Nacharbeit
4	Verkauf	Verkauf
5	Verkauf	Verkauf
6	Verkauf	Verkauf
7	Verkauf	Verkauf
8	Verkauf	Nacharbeit
9	Verkauf	Nacharbeit
10	Verkauf	Nacharbeit

Welche Aussage bezüglich der Konfusionsmatrix trifft zu?

Wählen Sie eine Antwort:

- a. Konfusionsmatrix  
TP: 3    FP: 2  
FN: 4    TN: 1
- b. Konfusionsmatrix  
TP: 4    FP: 1  
FN: 3    TN: 2
- c. Konfusionsmatrix  
TP: 1    FP: 4  
FN: 2    TN: 3
- d. Konfusionsmatrix  
TP: 1    FP: 4  
FN: 2    TN: 3

Abbildung 3-6 Prüfungsfrage Moodle Kurs IT-Management II

In der Untersuchung von Single-Choice-Fragen, bei welchen entscheidende Informationen als Bilddatei zur Fragestellung hinzugefügt wurde, wurde eine signifikante Einschränkung des Modells festgestellt. Obwohl ChatGPT über ein tiefes Verständnis des Konzepts der Konfusionsmatrix und der Evaluierung von Klassifikationsergebnissen verfügt, stieß es auf Schwierigkeiten, als es darum ging, Daten aus einem Bild zu extrahieren und zu interpretieren.

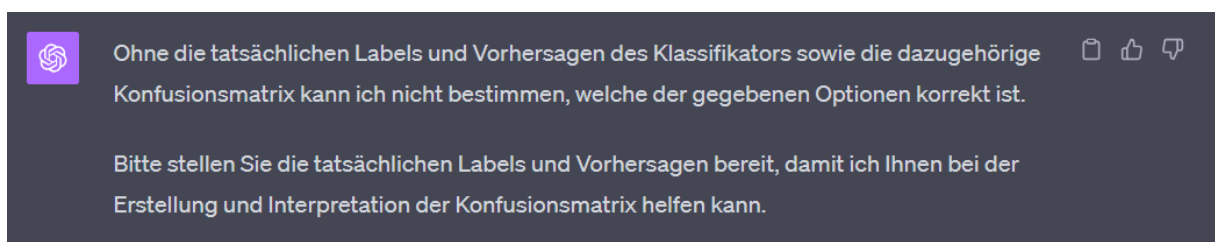
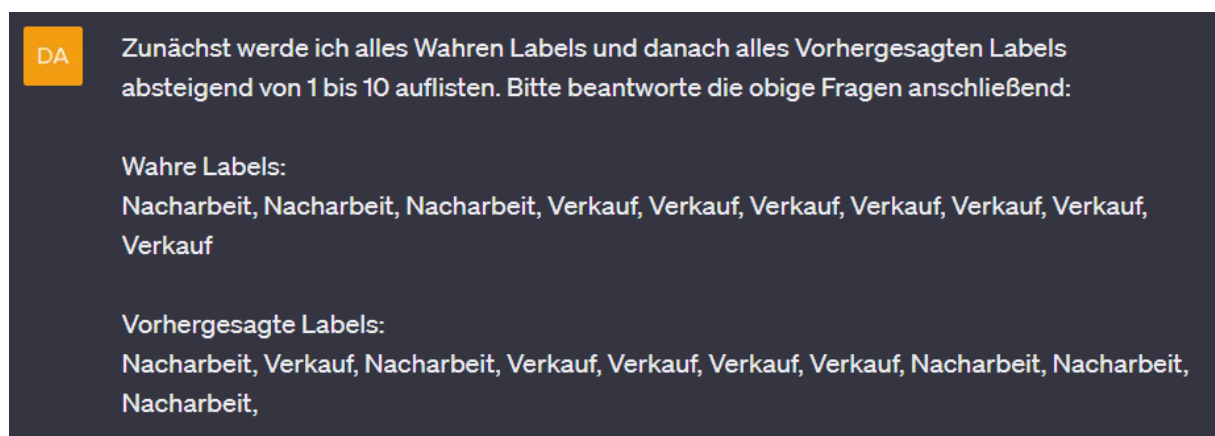


Abbildung 3-7 ChatGPT / Beantwortung der Prüfungsfrage (<https://chat.openai.com/>)

Ein Hauptgrund für diese Einschränkung liegt in der inhärenten Architektur und dem Design von ChatGPT. Das Modell wurde primär für die Textverarbeitung und -interpretation entwickelt und trainiert. Während es in der Lage ist, textbasierte Informationen effizient zu verarbeiten und darauf zu reagieren, fehlt ihm die Fähigkeit, Bildinhalte direkt zu interpretieren. In dem spezifischen Szenario, in dem die wahren Labels und die vorhergesagten Labels eines Klassifikators in einem Bild dargestellt wurden, konnte ChatGPT die notwendigen Daten nicht extrahieren, um die Konfusionsmatrix zu erstellen. Um die richtige Antwort zu erhalten war der prompt (siehe Abbildung 3-8) als Zwischenschritt notwendig.



DA Zunächst werde ich alles Wahren Labels und danach alles Vorhergesagten Labels absteigend von 1 bis 10 auflisten. Bitte beantworte die obige Fragen anschließend:

Wahre Labels:  
Nacharbeit, Nacharbeit, Nacharbeit, Verkauf, Verkauf, Verkauf, Verkauf, Verkauf, Verkauf, Verkauf

Vorhergesagte Labels:  
Nacharbeit, Verkauf, Nacharbeit, Verkauf, Verkauf, Verkauf, Verkauf, Nacharbeit, Nacharbeit, Nacharbeit,

Abbildung 3-8 Zwischenprompt zur Lösung der Aufgabe (<https://chat.openai.com/>)



Die richtige Antwort ist:

b. Konfusionsmatrix

TP: 4 FP: 1

FN: 3 TN: 2

Abbildung 3-9 ChatGPT / Beantwortung der Prüfungsfrage (<https://chat.openai.com/>)

## Prüfungsfragetyp: Multiple Choice

Ordnen Sie die nachfolgenden Aussagen entweder Deep Learning oder Supervised Learning zu.

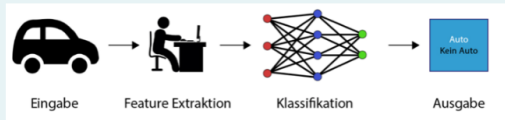
Deep Learning	Supervised Learning	
<input type="radio"/>	<input checked="" type="radio"/>	Das Training kann je nach Aufgabenstellung schnell gehen. Trotzdem können gute Vorhersagen und Entscheidungen getroffen werden.
<input checked="" type="radio"/>	<input type="radio"/>	Das menschliche Lernverhalten wird mittels sehr großen Datenmengen (über 100 Millionen Datenpunkte) imitiert.
<input checked="" type="radio"/>	<input type="radio"/>	Die historisch neuere und komplexere Technologie.
<input type="radio"/>	<input checked="" type="radio"/>	

Abbildung 3-10 Prüfungsfrage Moodle Kurs IT-Management II

Hier ist die Zuordnung der Aussagen:

- **Supervised Learning:**
  - Das Training kann je nach Aufgabenstellung schnell gehen. Trotzdem können gute Vorhersagen und Entscheidungen getroffen werden.
- **Deep Learning:**
  - Das menschliche Lernverhalten wird mittels sehr großen Datenmengen (über 100 Millionen Datenpunkte) imitiert.
  - Die historisch neuere und komplexere Technologie.

Abbildung 3-11 ChatGPT / Beantwortung der Prüfungsfrage (<https://chat.openai.com/>)

Im Umgang mit Multiple-Choice-Fragen wurde festgestellt, dass das Modell in der Lage ist, die Textbasierten Antwortoptionen in die richtige Kategorie einzuordnen. Jedoch stieß das Modell erneut auf eine signifikante Herausforderung, als eine der Auswahlmöglichkeiten in Form eines Bildes dargestellt wurde. Trotz seiner Textverarbeitungsfähigkeiten ist ChatGPT nicht in der Lage, Bildinhalte direkt zu interpretieren und kann somit keine Lösung angeben.

## Prüfungsfragetyp: Matching-Frage

Beim Cloud Computing geht es darum IT-Ressourcen nicht mehr zu besitzen sondern diese über ein Kommunikationsnetz bei Bedarf zu nutzen.  
Die wichtigsten Service Ebenen der Cloud sind Infrastructure as a Service (IaaS), Plattform as a Service (PaaS) und Software as a Service (SaaS). Ordnen Sie das aufgeführte Merkmal der passenden Serviceebene zu.  
Die Verfügbarkeit der Rechenleistung und des Speicherplatzes wird hier über Service Level Agreements zwischen Nutzer und Dienstleister geregelt.

Antwort:

Abbildung 3-12 Prüfungsfrage Moodle Kurs IT-Management II



Das Merkmal "Die Verfügbarkeit der Rechenleistung und des Speicherplatzes wird hier über Service Level Agreements zwischen Nutzer und Dienstleister geregelt" bezieht sich auf die Bereitstellung von grundlegenden IT-Ressourcen wie Rechenleistung, Speicherplatz und Netzwerkkapazitäten. Dies ist charakteristisch für:

### **Infrastructure as a Service (IaaS).**

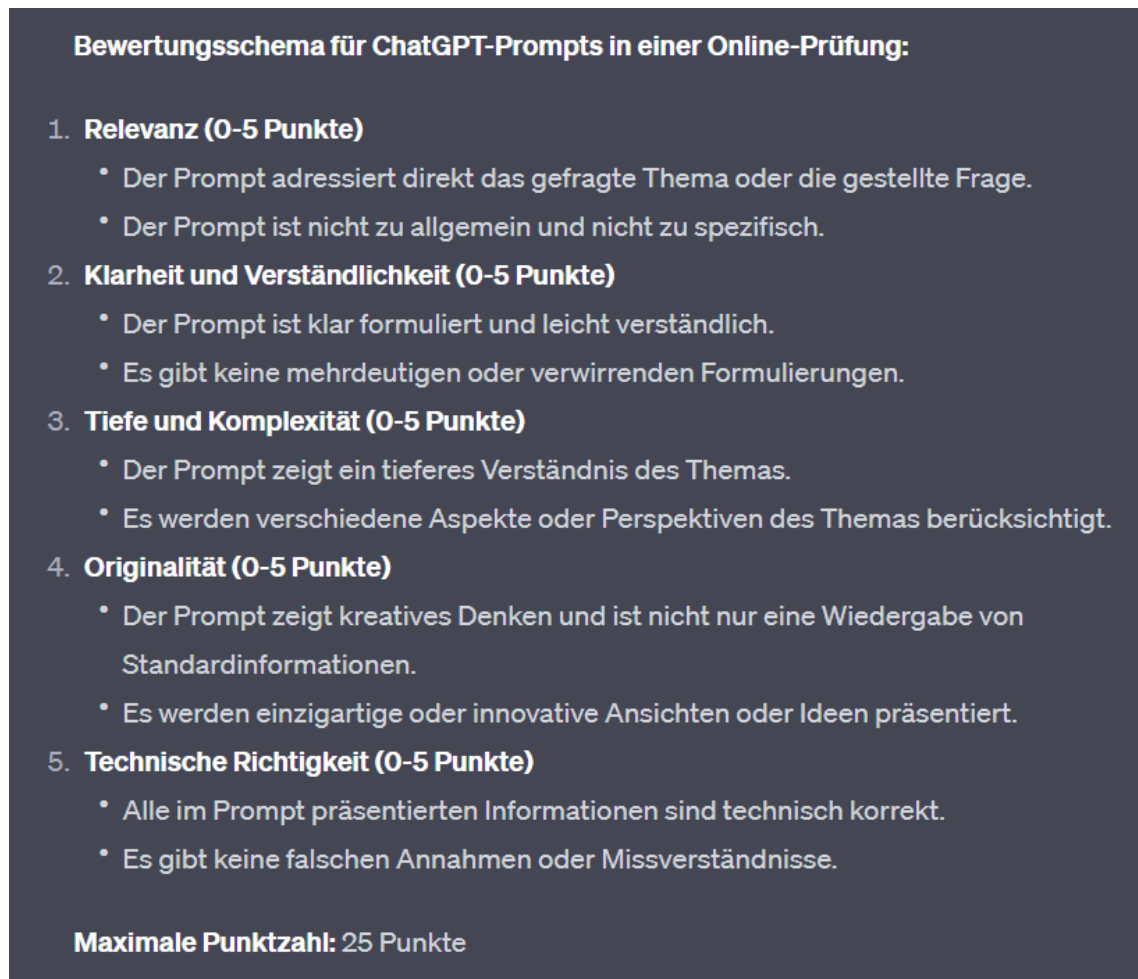
Bei IaaS mieten die Nutzer grundlegende IT-Infrastrukturen wie Server, Speicher und Netzwerke von einem Cloud-Anbieter, anstatt sie physisch zu besitzen. Die genauen Bedingungen, unter denen diese Ressourcen bereitgestellt werden, einschließlich ihrer Verfügbarkeit, werden in Service Level Agreements (SLAs) festgelegt.

Abbildung 3-13 ChatGPT / Beantwortung der Prüfungsfrage (<https://chat.openai.com/>)

Abschließend zeigt diese Untersuchung, dass die Kombination aus ChatGPTs umfangreichem Training und seiner Transformer-Architektur es ihm ermöglicht, Matching-Fragen zu technischen Themen wie den Serviceebenen des Cloud Computings effektiv zu beantworten. Dies unterstreicht das Potenzial von KI-Modellen in wissenschaftlichen Prüfungsumgebungen und betont ihre Rolle in der modernen Bildung.

## 3.2 Bewertungsschema für ChatGPT-Prompts in Online-Prüfungen

Da KI-Modelle wie ChatGPT nicht mehr wegzudenken sind wäre ein weiterer Ansatz diesen mit in den Prüfungsprozess einzubinden. Ein zentrales Element dieses Ansatzes ist die Bewertung von Prompts, die von den Prüflingen erstellt werden. Hierfür wurde ein spezifisches Bewertungsschema von ChatGPT entwickelt.



**Bewertungsschema für ChatGPT-Prompts in einer Online-Prüfung:**

- 1. Relevanz (0-5 Punkte)**
  - Der Prompt adressiert direkt das gefragte Thema oder die gestellte Frage.
  - Der Prompt ist nicht zu allgemein und nicht zu spezifisch.
- 2. Klarheit und Verständlichkeit (0-5 Punkte)**
  - Der Prompt ist klar formuliert und leicht verständlich.
  - Es gibt keine mehrdeutigen oder verwirrenden Formulierungen.
- 3. Tiefe und Komplexität (0-5 Punkte)**
  - Der Prompt zeigt ein tieferes Verständnis des Themas.
  - Es werden verschiedene Aspekte oder Perspektiven des Themas berücksichtigt.
- 4. Originalität (0-5 Punkte)**
  - Der Prompt zeigt kreatives Denken und ist nicht nur eine Wiedergabe von Standardinformationen.
  - Es werden einzigartige oder innovative Ansichten oder Ideen präsentiert.
- 5. Technische Richtigkeit (0-5 Punkte)**
  - Alle im Prompt präsentierten Informationen sind technisch korrekt.
  - Es gibt keine falschen Annahmen oder Missverständnisse.

**Maximale Punktzahl:** 25 Punkte

Abbildung 3-14 Bewertungsschema (<https://chat.openai.com/>)

Das vorgeschlagene Bewertungsschema für ChatGPT-Prompts in Online-Prüfungen basiert auf mehreren Schlüsselkriterien, die sowohl die Qualität der Kommunikation als auch das Verständnis des Themas berücksichtigen. Die

Kriterien sind: Relevanz, Klarheit und Verständlichkeit, Tiefe und Komplexität, Originalität und technische Richtigkeit. Jedes dieser Kriterien zielt darauf ab, verschiedene Aspekte des Prompts zu bewerten und sicherzustellen, dass die Antworten der Prüflinge sowohl korrekt als auch durchdacht sind. Es geht nicht nur darum, ob der Schüler die richtige Antwort kennt, sondern auch darum, wie er sein Wissen und seine Fähigkeiten anwendet, um zu dieser Antwort zu gelangen. Dies spiegelt einen Bildungsansatz wider, der den Schwerpunkt von reinem Auswendiglernen auf kritisches Denken und Anwendung verlagert.

Ein praktisches Beispiel für die Anwendung dieses Bewertungsschemas könnte eine Online-Klausur in einem Informatikkurs sein. Die Schüler könnten aufgefordert werden, einen Prompt zu erstellen, der ein neuronales Netzwerk erklärt. Die Antworten würden dann anhand des Bewertungsschemas bewertet, wobei besonderes Augenmerk auf die Tiefe des Verständnisses, die Klarheit der Erklärung und die technische Richtigkeit gelegt wird.

### **Prüfungsfrage:**

"Erklären Sie das Konzept des Deep Learning und wie ein neuronales Netzwerk mit mehreren versteckten Schichten (Deep Neural Network) funktioniert. Stellen Sie sich vor, Sie haben eine große Datenmenge von Bildern von Hunden und Katzen. Wie würden Sie ein tiefes neuronales Netzwerk trainieren, um diese Bilder automatisch zu klassifizieren? Bitte erläutern Sie Ihre Antwort schrittweise und beziehen Sie sich auf spezifische Aspekte und Mechanismen des Deep Learning."

### **Best Case Prompt:**

"Erkläre das Prinzip des Deep Learning und die Funktionsweise von tiefen neuronalen Netzwerken. Angenommen, es gibt eine Sammlung von Bildern, die entweder Hunde oder Katzen darstellen. Wie würde man ein Deep Neural Network aufbauen und trainieren, um diese Bilder effektiv zu unterscheiden?"

Welche spezifischen Schritte und Techniken im Deep Learning-Prozess sind notwendig, um eine hohe Genauigkeit in der Bildklassifikation zu erreichen?"

Dieser Prompt ist so formuliert, dass er:

- Relevant ist, da er sich direkt auf das Deep Learning und die Bildklassifikation bezieht.
- Klar und verständlich formuliert ist, wobei er spezifische Details und Kontext bietet.
- Tiefe und Komplexität aufweist, da er nicht nur eine einfache Definition verlangt, sondern auch eine schrittweise Erklärung und Anwendung.
- Originalität zeigt, indem er den Kontext der Bildklassifikation von Hunden und Katzen verwendet.
- Technisch korrekt ist, da er sich auf die richtigen Mechanismen und Techniken des Deep Learning bezieht.

### **Worst Case Prompt:**

"Was ist Deep Learning? Wie unterscheidet man Hunde und Katzen?"

Dieser Prompt ist so formuliert, dass er:

- Relevanz verfehlt, da er nur eine sehr oberflächliche und allgemeine Frage zum Thema stellt.
- Klarheit und Verständlichkeit fehlt, da er nicht spezifisch oder detailliert genug ist.
- Tiefe und Komplexität vermisst, da er keine tiefgehenden Informationen oder Erklärungen verlangt.
- Originalität fehlt, da er sehr allgemein und nicht kreativ ist.
- Technische Richtigkeit ist zwar gegeben, aber der Prompt kratzt nur an der Oberfläche des Themas.

Dieser Ansatz hat den Vorteil, dass er die Schüler dazu anregt, über das bloße Wissen hinauszugehen und ihre Fähigkeiten in einem realen Kontext anzuwenden. Es fördert die Problemlösung, da die Schüler nicht nur Informationen wiedergeben, sondern sie auch interpretieren und in einem gegebenen Kontext anwenden müssen.

Bei näherer Betrachtung ergeben sich einige kritische Punkte, die berücksichtigt werden müssen.

Erstens erfordert dieses Schema, dass die Lernenden gründlich in den Kriterien geschult werden, nach denen ihre Antworten bewertet werden. Dies stellt eine erhebliche Umstellung in der Art und Weise dar, wie Online-Prüfungen bisher durchgeführt wurden. Die Einführung eines solchen Bewertungsschemas würde bedeuten, dass wertvolle Unterrichtszeit darauf verwendet werden muss, den Schülern die spezifischen Kriterien und Erwartungen beizubringen, anstatt sich auf den eigentlichen Lehrstoff zu konzentrieren.

Zweitens, und vielleicht noch kritischer, ist die Tatsache, dass, wenn das Bewertungsschema den Lernenden bekannt ist, es potenziell ausgenutzt werden könnte in dem man das Bewertungsschema in seinen prompt mit einfügt und ChatGPT befiehlt, die zu bewertende Frage so zu formulieren, dass die volle Punktzahl erreicht wird.

Schließlich besteht die Gefahr, dass das Bewertungsschema zu mechanisch wird und die tatsächliche Qualität oder Tiefe des Verständnisses eines Schülers nicht genau widerspiegelt. Während die Kriterien objektiv erscheinen mögen, könnten sie in der Praxis dazu führen, dass Schüler Antworten "nach Rezept" erstellen, anstatt kritisches Denken und echtes Verständnis zu demonstrieren.

## **4. Maßnahmen zur Betrugsprävention**

Aus der detaillierten technischen Analyse ergaben sich klare Handlungsempfehlungen. Diese zielen darauf ab, in zukünftigen Online-Prüfungen auf Plattformen wie Moodle das Risiko betrügerischer Handlungen durch den Einsatz von KI-Systemen, insbesondere solchen wie ChatGPT, signifikant zu minimieren. Es ist von entscheidender Bedeutung, diese Empfehlungen in Betracht zu ziehen, um die Integrität des Prüfungsprozesses in einem digitalen Bildungsumfeld zu gewährleisten. Hierbei wurden bereits erlangte Erkenntnisse einer Studie miteinbezogen (vgl. Teo Susnjak 2022, S. 16).

### **4.1 Verwendung von Bildern zur Wissensabfrage**

Die Integration von Bildmaterial in Prüfungen kann die Komplexität von Fragen erhöhen und stellt eine Hürde für die automatische Beantwortung durch KI-Modelle wie ChatGPT dar. Die technische Analyse hat ergeben, dass die Einbeziehung von Bildmaterial die Fähigkeit von KI-Modellen, Fragen korrekt zu beantworten, tatsächlich erschwert oder sogar unmöglich macht. Es gibt unterschiedliche Wege, wie Bilder in Prüfungen eingesetzt werden können:

#### **4.1.1 Bildinterpretation**

Die Fähigkeit, visuelle Informationen zu analysieren und zu interpretieren, ist ein wesentlicher Bestandteil des menschlichen Lernprozesses. In Prüfungsszenarien wird diese Fähigkeit oft durch die Einbindung von Bildmaterial getestet, um den Prüfling herauszufordern und seine Fähigkeiten in der Bildinterpretation zu bewerten.

Ein gängiger Ansatz in solchen Prüfungen ist die Verwendung von Bildern, Diagrammen oder Grafiken, die spezifische Informationen oder Konzepte

darstellen. Der Prüfling wird dann aufgefordert, Fragen zu beantworten, die sich direkt auf das präsentierte visuelle Material beziehen. Ein klassisches Beispiel hierfür könnte ein Diagramm eines neuronalen Netzwerks sein, das die verschiedenen Schichten und Knotenpunkte zeigt. Anstatt den Prüfling einfach nach den Funktionen oder Eigenschaften des neuronalen Netzwerks zu fragen, wird er mit einem visuellen Hilfsmittel konfrontiert und muss Fragen wie "Identifizieren Sie die abgebildete Schicht des Netzwerks" oder "Wie funktioniert der abgebildete Backpropagation-Prozess?" beantworten.

Dieser Ansatz hat mehrere Vorteile. Erstens testet er nicht nur das reine Wissen des Prüflings, sondern auch seine Fähigkeit, Informationen aus einem Bild zu extrahieren und zu verarbeiten. Zweitens simuliert er reale Szenarien, in denen Fachleute oft mit visuellen Darstellungen komplexer Modelle konfrontiert werden und diese interpretieren müssen, insbesondere in technischen und wissenschaftlichen Bereichen wie dem Deep Learning.

#### **4.1.2 Bildassoziationen**

Die Verwendung von Bildassoziationen in Prüfungen stellt eine innovative Methode dar, um das Verständnis und die kritische Denkfähigkeit von Prüflingen zu bewerten. Das Hauptziel dieser Methode ist es, den Prüfling dazu anzuregen, tiefe und bedeutungsvolle Verbindungen zwischen einem präsentierten Bild und einem zugrunde liegenden Konzept oder Thema herzustellen. Anstatt sich nur auf textbasierte Informationen zu verlassen, fordert die Bildassoziation den Prüfling heraus, visuelle Hinweise zu interpretieren und diese Interpretationen in einen breiteren Kontext zu stellen.

Ein exemplarisches Verfahren könnte so aussehen: Ein Bild wird dem Prüfling ohne jeglichen Kontext oder begleitenden Text präsentiert. Der Prüfling muss dann das Bild analysieren und es mit einem bestimmten Thema, Konzept oder einer bestimmten Idee in Verbindung bringen. Als Beispiel könnte ein Bild eines neuronalen Netzwerkschemas vorgelegt werden, gefolgt von der Aufforderung:

"Erläutern Sie die Schlüsselkomponenten und Prinzipien des Deep Learning, die in diesem Bild dargestellt werden." Hierbei wird nicht nur das Wissen des Prüflings über Deep Learning und neuronale Netzwerke getestet, sondern auch seine Fähigkeit, visuelle Informationen zu interpretieren und in einen größeren, oft komplexeren, thematischen Rahmen zu setzen.

Diese Methode der Bildassoziation kann besonders wertvoll sein, um die Fähigkeiten der Prüflinge in den Bereichen kritisches Denken, Analyse und Interpretation zu bewerten. Es ermutigt sie, über das Offensichtliche hinauszusehen und tiefere Bedeutungen und Zusammenhänge zu erkennen. In einer wissenschaftlichen Umgebung kann dies dazu beitragen, ein umfassenderes Verständnis für ein Thema zu fördern und die Fähigkeit zur interdisziplinären Analyse zu stärken.

#### **4.1.3 Schrittweise Lösung mittels Bildern**

Ein anschauliches Beispiel für diesen Ansatz ist die Darstellung eines Programmierprozesses oder einer Code-Abfolge durch eine Serie von Bildern. Nehmen wir zum Beispiel die Implementierung eines Sortieralgorithmus in Python: Anstatt den Prüfling zu bitten, den Algorithmus in Worten oder Code zu beschreiben, könnten Bilder, die verschiedene Phasen der Implementierung zeigen, in zufälliger Reihenfolge präsentiert werden. Die Aufgabe des Prüflings wäre es dann, diese Bilder in die korrekte chronologische Reihenfolge zu bringen. Dies testet nicht nur das Wissen des Prüflings über den spezifischen Algorithmus, sondern auch sein allgemeines Verständnis für Programmierkonzepte und seine Fähigkeit, visuelle Hinweise im Kontext von Code zu interpretieren.

Diese Methode hat den zusätzlichen Vorteil, dass sie den Prüfling dazu anregt, kritisch zu denken und Code-Konzepte zu synthetisieren, anstatt sich nur auf das Auswendiglernen von Syntax zu verlassen. Es fördert auch die Entwicklung von Fähigkeiten wie Mustererkennung und logisches Denken im Kontext der Programmierung.

#### **4.1.4 Bild-basierte Fallstudien**

Ein gängiges Verfahren bei dieser Methode ist die Präsentation eines Bildes oder einer Serie von Bildern, die eine bestimmte Situation oder ein Szenario darstellen. Dies wird dann von Fragen begleitet, die eine tiefgehende Analyse oder die Anwendung bestimmter Theorien oder Konzepte erfordern. Als Beispiel könnte ein Scatterplot mit Datenpunkten präsentiert werden, der eine lineare Beziehung zwischen zwei Variablen zeigt. Die begleitende Frage könnte lauten: "Basierend auf dem gezeigten Scatterplot, würden Sie eine lineare Regression oder eine Klassifikation als geeignetes Modell wählen, um die Beziehung zwischen den Variablen zu beschreiben?" Hier wird nicht nur das Wissen des Prüflings über Regression und Klassifikation getestet, sondern auch seine Fähigkeit, dieses Wissen auf eine konkrete visuelle Darstellung anzuwenden.

#### **4.2 Verwendung von Videomaterial zu Wissensabfrage**

Ein vielversprechender Ansatz, ähnlich wie bei Bildern, ist die Integration von vorher aufgenommenen Videos in den Prüfungsprozess.

In einer möglichen Visualisierung durch den Service der Webseite <https://www.vyond.com/> (siehe Abbildung 4-1 bis 4-4), könnte die Prüfungsfrage, z.B. zum Thema Big Data Frameworks, zusammen mit den zugehörigen Antwortoptionen präsentiert werden. Bei einer SCF würden die dann auswählbaren Antwortmöglichkeiten durch ihre Kennzeichnungen a, b, c oder d repräsentiert. Durch diese Methode, insbesondere in Kombination mit einem festgelegten Zeitlimit für die Beantwortung, könnte die Anwendung von KI-Systemen wie ChatGPT erheblich eingeschränkt werden. Ein solches Zeitlimit würde die Nutzung solcher Systeme während der Prüfung weniger attraktiv machen und somit die Integrität der Online-Prüfung weiterhin gewährleisten.



Abbildung 4-1



Abbildung 4-2



Abbildung 4-3

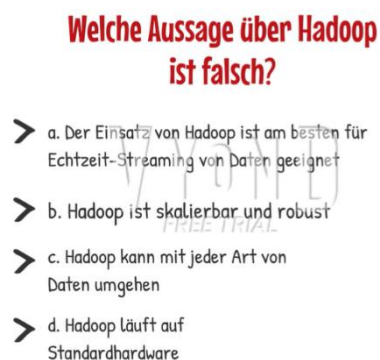


Abbildung 4-4

Abbildung 4-1 bis 4-4 Schrittweise Darstellung einer möglichen Video Prüfungsfrage (<https://www.vyond.com/>)

Zusätzlich wäre die Ultima Ratio eine Rückkehr zu mündlichen Prüfungen. Mündliche Prüfungen, ob online oder vor Ort, erfordern von den Studierenden, ihr Wissen in Echtzeit verbal zu demonstrieren. Dies minimiert die Möglichkeit, externe Hilfsmittel oder Technologien zur Beantwortung von Fragen zu verwenden und stellt sicher, dass das gezeigte Wissen tatsächlich dem Prüfling gehört.

## 5. Fazit

Die voranschreitende Digitalisierung im Bildungsbereich hat zu einer Vielzahl von Innovationen geführt, von denen die Integration von KI-Modellen wie ChatGPT sicherlich eine der bemerkenswertesten ist. In der hier vorgestellten wissenschaftlichen Arbeit wurde ein tiefgreifendes Dilemma untersucht: "Wie können Prüfungsfragen gestaltet werden, um die Wahrscheinlichkeit zu verringern, dass sie von KI-Modellen beantwortet werden können?"

Durch sorgfältige technische Analysen wurde festgestellt, dass die Einbindung von Bildern und Videos in Prüfungsfragen eine effektive Methode darstellt, um die Fähigkeiten von ChatGPT zu begrenzen. Dies liegt nicht nur an den aktuellen Beschränkungen von KI-Modellen bei der Interpretation visueller Daten, sondern auch daran, dass solche multimedialen Elemente oft eine tiefere, kontextbezogene Analyse erfordern, die über das hinausgeht, was reine Textmodelle bieten können.

Interessanterweise schlägt diese Arbeit nicht vor, KI vollständig aus dem Prüfungsprozess zu entfernen. Stattdessen wird ein innovativer Ansatz vorgeschlagen, bei dem ChatGPT als Werkzeug zur Bewertung der von den Lernenden erstellten Prompts genutzt wird. Dieser Paradigmenwechsel erkennt die unaufhaltsame Präsenz und den Einfluss von KI in der Bildung an und sucht nach Wegen, diese Technologie konstruktiv zu nutzen.

Das Herzstück dieses Ansatzes ist ein detailliertes Bewertungsschema, das darauf abzielt, nicht nur die Richtigkeit, sondern auch die Tiefe und Qualität des Verständnisses zu bewerten, das in den Prompts der Lernenden zum Ausdruck kommt. Es geht darum, einen Bildungsansatz zu fördern, der den Schwerpunkt auf kritisches Denken, Reflexion und Anwendung legt, anstatt sich nur auf das Wiederkäuen von Informationen zu konzentrieren.

Diese Methode wurde nicht zu den Maßnahmen der Betrugsprävention mitaufgenommen, da es hier zu einer unkonventionellen Umgestaltung des Lehransatz führt welche auf großen Hürden stoßen kann und eine Betrügerische Handlung des Prüflings dennoch nicht ausgeschlossen ist. Es bedarf hier weitere Forschung.

In der heutigen Zeit, in der Technologie und Bildung immer stärker miteinander verflochten sind, bietet diese Arbeit eine zeitgemäße Perspektive auf die Herausforderungen und Chancen an dieser Schnittstelle. Es ist unerlässlich, dass Bildungseinrichtungen und Pädagogen sich dieser Herausforderungen bewusst sind und proaktiv Strategien entwickeln, um die Vorteile der Technologie zu nutzen und gleichzeitig die Integrität des Bildungsprozesses zu wahren.

## 6 Literaturverzeichnis

Ian Goodfellow, Yoshua Bengio, Aaron Courville (2016): Deep Learning: MIT Press

Alpaydın, Ethem (2022): Maschinelles Lernen. 3. Berlin, Boston: De Gruyter Oldenbourg. Online verfügbar unter <https://www.degruyter.com/isbn/9783110740141>

Müller, A.C. & Guido, S. (2016). Introduction to Machine Learning with Python: A Guide for Data Scientists. O'Reilly Media. Seiten 21-83

Teo Susnjak (2022): ChatGPT: The End of Online Exam Integrity?, Auckland, New Zealand: School of Mathematical and Computational Sciences, Massey University

Patel, A.A. (2016). Unsupervised Machine Learning in Python, Packt Publishing. Seiten 1-53

Michael A. Nielsen (2015): Neural Networks and Deep Learning: A Textbook, Determination Press

Manning, C.D., & Schütze, H. (1999): Foundations of Statistical Language Processing, S. 127-137, MIT Press

Bendel, Oliver (2019): Maschine Learning. Definition: Was ist Maschine Learning? Springer. Online verfügbar unter <https://wirtschaftslexikon.gabler.de/definition/machine-learning-120982/version370915>

Aleksei Egorov (2020): Online-Prüfungen verleiten zum Schummeln, forschung-und-lehre, [online] <https://www.forschung-und-lehre.de/lehre/online-pruefungen-verleiten-zum-schummeln-3369/> [abgerufen am 16.08.2023]

Hernandez, Sorokina (2021): Onlineprüfungen luden zum Betrug ein, forschung-und-lehre, [online] <https://www.forschung-und-lehre.de/lehre/onlinepruefungen-luden-zum-betrug-ein-3786/> [abgerufen am 16.08.2023]

OpenAI (2022): Introducing ChatGPT, openai, [online] <https://openai.com/blog/chatgpt> [abgerufen am 17.08.2023]

Ashish Vaswani, Noam Shazzer, Niki Parmar (2017): Attention Is All You Need, proceedings-neurips, [online] <https://papers.neurips.cc/paper/7181-attention-is-all-you-need.pdf> [abgerufen am 17.08.2023]

Richard S. Sutton und Andrew G. Barto (1998): Reinforcement Learning, MIT Press

Mariusz Bojarski (2016): End to End Learning for Self-Driving Cars, Cornell University

## Eidesstattliche Erklärung

Hiermit erkläre ich, dass ich die vorliegende Arbeit selbstständig verfasst habe, dass ich sie zuvor an keiner anderen Hochschule und in keinem anderen Studiengang als Prüfungsleistung eingereicht habe und dass ich keine anderen als die angegebenen Quellen und Hilfsmittel benutzt habe. Alle Stellen der Arbeit, die wörtlich oder sinngemäß aus Veröffentlichungen oder aus anderweitigen fremden Äußerungen entnommen wurden, sind als solche kenntlich gemacht.

Neu-Ulm 15.10.2023

Ort, Datum



Unterschrift