

TH Köln

Institut für Informatik und Ingenieurwissenschaften

Verbundstudiengang Wirtschaftsinformatik

Abschlussarbeit

zur Erlangung

des Bachelorgrades

Bachelor of Science

In der Fachrichtung Informatik

Untersuchung über die Eignung von KI- Modellen als Tutor in der Mathematikausbildung

Erstprüfer: Prof. Dr. Wolfgang Konen

Zweitprüfer: Prof. Dr. Dietlind Zühlke

Vorgelegt am: 16.04.2024

Von cand. Jonathan Paul

aus Steegerstraße 7

51067 Köln

Tel.: 017655408314

E-Mail: jonathan.paul@smail.th-koeln.de

Matrikelnummer: 11132801

1. Inhaltsverzeichnis

1.	Inhaltsverzeichnis.....	2
2.	Abbildungsverzeichnis	3
3.	Abbildungsverzeichnis	4
4.	Abkürzungs- u. Symbolverzeichnis	5
5.	Einleitung	5
a.	Chatbots.....	7
6.	Forschungsstand.....	8
7.	Konzept des Experiments	10
a.	Aufbau.....	10
b.	Aufgaben.....	10
c.	Bewertung	13
8.	Resultate der Aufgaben.....	15
a.	ChatGPT 3.5	15
Exemplarische Aufgaben	16	
Weitere Bemerkungen	23	
b.	ChatGPT 4	25
Exemplarische Aufgaben	26	
Weitere Bemerkungen	32	
c.	Google Bard	33
Exemplarische Aufgaben	33	
Weitere Bemerkungen	40	
d.	Copilot.....	42
Exemplarische Aufgaben	42	
Weitere Bemerkungen	45	
e.	Auswertung	48
9.	Umfrage	50
a.	Aufbau der Umfrage.....	50

b.	Ergebnisse der Aufgaben.....	52
c.	Ergebnis der Umfrage	53
10.	Eigenschaften eines Tutors	57
11.	Fazit	58
a.	Einleitung des Fazits	58
b.	Umfrage	58
c.	Einbindung der Literatur	58
d.	Vergleich zwischen Chatbot und Tutor.....	59
e.	Abschluss.....	60
12.	Quellen.....	60
13.	Anhang.....	62

2. Abbildungsverzeichnis

Abbildung 1	Input Beispielaufgabe	11
Abbildung 2	Beispielaufgabe Output 1	12
Abbildung 3	Beispielaufgabe Output 2	12
Abbildung 4	Beispielaufgabe Output 3	13
Abbildung 5	ChatGPT 3.5 erste Aufgabe Input 1.....	16
Abbildung 6	ChatGPT 3.5 erste Aufgabe Output 1	17
Abbildung 7	ChatGPT 3.5 erste Aufgabe Output 2	18
Abbildung 8	ChatGPT 3.5 erste Aufgabe Output 3	18
Abbildung 9	ChatGPT 3.5 zweite Aufgabe Input 1	19
Abbildung 10	ChatGPT 3.5 zweite Aufgabe Output 1.....	20
Abbildung 11	ChatGPT 3.5 zweite Aufgabe Output 2	21
Abbildung 12	ChatGPT 3.5 zweite Aufgabe Output 3.....	21
Abbildung 13	ChatGPT 3.5 zweite Aufgabe Output 4.....	22
Abbildung 14	ChatGPT 3.5 weitere Bemerkungen Output 1	23
Abbildung 15	ChatGPT 3.5 weitere Bemerkung Output 2	24
Abbildung 16	ChatGPT 3.5 weitere Bemerkung Output 3	24
Abbildung 17	ChatGPT 3.5 weitere Bemerkung Output 4	25
Abbildung 18	ChatGPT 4 erste Aufgabe Input 1.....	26
Abbildung 19	ChatGPT 4 erste Aufgabe Output 1	26

Abbildung 20 ChatGPT 4 erste Aufgabe Output 2	27
Abbildung 21 ChatGPT 4 erste Aufgabe Output 3	28
Abbildung 22 ChatGPT 4 erste Aufgabe Output 4	28
Abbildung 23 ChatGPT 4 zweite Aufgabe Input 1.....	29
Abbildung 24 ChatGPT 4 zweite Aufgabe Output 1.....	30
Abbildung 25 ChatGPT 4 zweite Aufgabe Output 2.....	31
Abbildung 26 ChatGPT 4 weitere Bemerkungen Output 1	32
Abbildung 27 Google Bard erste Aufgabe Input 1	33
Abbildung 28 Google Bard erste Aufgabe Output 1.....	34
Abbildung 29 Google Bard erste Aufgabe Output 2.....	35
Abbildung 30 Google Bard erste Aufgabe Output 3.....	35
Abbildung 31 Google Bard erste Aufgabe Output 4.....	36
Abbildung 32 Google Bard zweite Aufgabe Input 1	37
Abbildung 33 Google Bard zweite Aufgabe Output 1.....	38
Abbildung 34 Google Bard zweite Aufgabe Output 2.....	39
Abbildung 35 Google Bard weitere Bemerkungen Output 1	41
Abbildung 36 Google Bard weitere Bemerkungen Output 2	41
Abbildung 37 Copilot erste Aufgabe Input 1	42
Abbildung 38 Copilot erste Aufgabe Output 1.....	43
Abbildung 39 Copilot zweite Aufgabe Output 1	44
Abbildung 40 Copilot dritte Aufgabe Output 1.....	45
Abbildung 41 Copilot weitere Bemerkungen Output 1	46
Abbildung 42 Copilot weitere Bemerkungen Input 1.....	46
Abbildung 43 Copilot weitere Bemerkungen Output 2	47
Abbildung 44 Copilot weitere Bemerkungen Input 2.....	47
Abbildung 45 Copilot weitere Bemerkungen Output 3.....	48
Abbildung 46 Umfrageergebnis	55
Abbildung 47 Umfrageergebnis groß 1.....	62
Abbildung 48 Umfrageergebnis groß 2.....	62
Abbildung 49 Umfrageergebnis groß 3.....	62
Abbildung 50 Umfrage Aufgabe 1 ChatGPT 3.5 groß.....	64

3. Abbildungsverzeichnis

Tabelle 1 Ergebnis Experiment ChatGPT 3.5	15
---	----

Tabelle 2 Ergebnis Experiment ChatGPT 4	25
Tabelle 3 Ergebnis Experiment Google Bard	33
Tabelle 4 Ergebnis Experiment Copilot	42
Tabelle 5 Umfrage Fragen 1	51
Tabelle 6 Umfrage Fragen 2	52
Tabelle 7 Umfrageergebnis 1	53
Tabelle 8 Umfrageergebnis 2	54
Tabelle 9 Umfrageergebnis 3	56

4. Abkürzungs- u. Symbolverzeichnis

KI <i>Siehe</i> Künstliche Intelligenz	LLM <i>Siehe Large Language Modell</i>
--	--

5. Einleitung

Diese Bachelorarbeit behandelt das Thema Künstlicher Intelligenz (KI) unter der Fragestellung, ob und inwiefern diese geeignet sein kann, um Studierende im Fachbereich Mathematik mit fachkundiger Expertise zu unterstützen. Im Verlauf der Untersuchung wird die Qualität der Antworten verschiedener KIs auf mathematische Fragestellungen beleuchtet und darauffolgend deren Wert für Nutzer erörtert. Mit der Entwicklung von funktioneller KI befasst sich die Menschheit seit vielen Jahren. Im Jahr 1956 fand hierzu die erste große Veranstaltung im Dartmouth College statt. Seitdem hat sich die KI in viele Richtungen weiterentwickelt und viele Unternehmen haben das Potential von KI für diverse Lebensbereiche erkannt. Hierzu zählen Suchassistenten oder auch Sprachsteuerungssoftware von den Firmen Google, Microsoft und Apple. KI steht zwar oft in der Kritik, wie zum Beispiel durch die Unterhaltungsindustrie, die fürchtet viele Jobs zu verlieren [7], oder auch in der Politik, da Besorgnis besteht, dass Wahlen mithilfe von durch KIs erstellte und echt aussehende, Falschaussagen beeinflusst werden können [8]. KI wird aber auch positiv in den Medien dargestellt. In der Wissenschaft zum Beispiel kann KI helfen Forschungen oder Studien mit arbeits- sowie zeitintensiven Aufgaben zu beschleunigen

[9]. Die Stadt Goslar beispielsweise nutzt KI zur Auswertung von Straßenaufnahmen mit Schlaglöchern, um zu ermitteln welche Schlaglöcher als erstes ausgebessert werden müssen [10]. An innovativen Neuprojekten wie diesen ist bereits zu erkennen wie vielseitig KIs einsetzbar sein können. Eine weitere positive Tendenz ist die Erleichterung des Berufsalltags in vielen Bereichen [15]. Ein großer Bereich der KIs sind die sogenannten Chatbots, die mithilfe von KI mit Menschen interagieren können. Einer der Bekanntesten Chatbots ist ChatGPT. Dieser wurde als Version 3.0 im November 2022 von der Firma OpenAI veröffentlicht. Der Dienst ist seit der Veröffentlichung kostenlos. OpenAI forscht an ChatGPT seit 2015 und hat vor ChatGPT 3.0 die Versionen 1.0 und 2.0 veröffentlicht. Der Durchbruch erfolgte aber erst mit der Version GPT 3. GPT steht für Generative Pre-trained Transformer. Später veröffentlichten auch andere Firmen wie Google und Microsoft ihre eigenen KI basierten Chatbots. Googles KI hieß zu Beginn Google Bard und mittlerweile Gemini. Die Adaption von Microsoft heißt Copilot. Google greift hierfür auf eine eigene KI zurück. Microsoft arbeitet mit OpenAI zusammen und nutzt deren GPT 4 Version.

Chatbots sind, bei einer bestehenden Internetverbindung, immer erreichbar und bieten den Vorteil, sich jedem individuellen Zeitplan und jedem beliebigen Thema anzupassen. Man kann Chatbots jede Frage stellen und die Antworten sind für Nutzer oft hilfreich. Mathematik ist in vielen Studiengängen eine Hürde für Studierende und der Bedarf an Nachhilfe in Mathematik ist groß. Viele Studierende greifen bereits auf Chatbots zurück, um schnell und unkompliziert Einzelnachhilfe zu konkreten Fragestellungen zu erhalten. Im Verlauf der globalen Nutzung von Chatbots ist allerdings aufgefallen, dass diese auch falsche Antworten geben können, welche erst beim genauen Hingucken als falsch erkennbar werden. In Anbetracht dessen schließt sich fast automatisch die Frage an, ob und inwiefern sich eine KI als Tutor bzw. Nachhilfelehrerkraft eignen kann. In der vorliegenden Untersuchung wurde exemplarisch die Tutor Eignung im Bereich Mathematik getestet. Um zu prüfen, ob sich ein Chatbot hierfür eignet sollte geprüft werden,

wie zuverlässig und fachlich korrekt die Antworten von Chatbots auf mathematische Fragen sind, denn Chatbots können bei mehrmaligem aufeinanderfolgendem Fragen mit unterschiedlichen Ergebnissen lösen. Chatbots sind keine Taschenrechner, die immer das richtige bzw. gleiche Ergebnis anzeigen. In der vorliegenden Arbeit soll daher die Frage beleuchtet werden, ob und inwiefern Chatbots eine Tutor-Rolle einnehmen und den Studierenden bei mathematischen Fragestellungen helfen können.

a. Chatbots

Die vier Chatbots, die in dieser Arbeit behandelt werden sind ChatGPT 3.5 [1], ChatGPT 4 [1], Google Bard [3] und Copilot [4]. Es gibt jedoch noch viele andere Chatbots, die untersucht werden könnten, dies würde allerdings den Rahmen dieser Arbeit sprengen.

ChatGPT 3.5 ist die kostenlose Version von OpenAIs Large Language Modell (LLM). ChatGPT 3.5 ist seit Januar 2022 nicht mehr weiter trainiert worden und besitzt daher keine Informationen zu aktuellen Themen und wird sich folglich auch nicht weiterentwickeln. Diese Information gibt ChatGPT 3.5 selbst, wenn man die KI fragt [1]. Von Studierenden kann sie aufgrund des barrierefreien Zugangs kostensparend genutzt werden.

ChatGPT 4 ist die kostenpflichtige Version von OpenAI. Diese kostet im Frühjahr 2024 23,80\$ im Monat was umgerechnet 22,57€ sind [1]. GPT 4 wird regelmäßig aktualisiert und mit neuen Datensätzen und Informationen trainiert. Für dieses Experiment wurde ChatGPT 4 ausgewählt, um zu prüfen, ob für die potenziellen Nutzer – den Studierenden – die Nutzung der kostenpflichtigen Variante Vorteile gegenüber der kostenlosen Variante (ChatGPT3.5) bietet. Für einen Studenten können 22,57€ im Monat zwar viel sein, aber ein Nachhilfelehrer kann mehr kosten.

Google Bard ist ein kostenloses LLM von Google. Während der laufenden Untersuchung wurde Google Bard zu Gemini umbenannt. In

der vorliegenden Arbeit wird weiterhin der ursprüngliche Name Google Bard verwendet. Google Bard ist eine Alternative zu ChatGPT 3.5, weil es ebenfalls kostenlos ist, jedoch ein anderes LLM benutzt. Zunächst wurde das von Google selbst entwickelte LaMDA (Language Model for Dialogue Applications) genutzt. Später wurde auf PaLM 2 (Pathways Language Model) gewechselt, welches ebenfalls von Google selbst entwickelt wurde [4][5]. Hier ist demnach zu prüfen, ob und inwiefern sich die Ergebnisse der kostenlosen KIs voneinander unterscheiden.

Copilot ist ein kostenloses LLM von Microsoft. Früher trug Copilot den Namen Bing Chat, weil es direkt in die Suchmaschine Bing integriert ist. Copilot benutzt dasselbe LLM als Basis wie ChatGPT 4, ist im Gegensatz dazu aber kostenlos. Es ist also interessant zu sehen, ob sich die kostenpflichtige KI ChatGPT 4 und die kostenlose KI Copilot ähneln oder sogar identische Antworten geben.

6. Forschungsstand

Da die Thematik um Chatbots noch recht neu ist, ist zu diesem Zeitpunkt nur wenig themenbezogene Literatur zu finden. Zum Thema KI gibt es viele Dokumentationen. Die Chatbots entwickeln sich schnell weiter, daher ist die Literatur in manchen Aspekten veraltet. So sagt Vagelis Pelvris zusammen mit seinen Forscher Kollegen [11], dass ChatGPT 4 keine Internetverbindung seit 2021 habe. Dies hat sich seitdem geändert und ChatGPT 4 ist mit dem Internet verbunden. Diese Quelle stammt aus dem Mai 2023. Die Literatur ist sich einig, dass die KI nicht allein zum Lernen oder zum Lösen von Problemen genutzt werden sollte. Forscher aus Nepal [12] sagen zu diesem Punkt, dass man stets mit Menschlichen Lehrern interagieren solle. Weiter sagen sie, dass Chatbots bis zu einem bestimmten Grad an Komplexität nützlich sein können, die KI jedoch ihre Limits habe. Laut ihnen sei die KI eine Bereicherung, weil es helfen, kann mehr Menschen zu erreichen und für Mathe zu begeistern, weil die Lernerfahrung viel personalisierter sein kann. Außerdem können Chatbots die Fragen und Probleme Schritt für Schritt zusammen mit dem Nutzer durchgehen. Es wird dadurch ein Lernen in der eigenen Geschwindigkeit geschaffen.

Zusätzlich dazu können Chatbots Beispielaufgaben geben, welche mit Lösungen ausgegeben werden können. Dies habe ich schon mit Google Bard erreicht. Horst Hirscher [13] arbeitet an der Universität des Saarlandes und behauptet, dass im Fachbereich der Mathematik keine negative Stimmung gegen die Entwicklung der Chatbots herrsche. Jedoch sind die Texte der Chatbots noch häufig fiktional und teilweise fehlerhaft. Vagelis Pelvris hat zusammen mit seinen Kollegen George Papazafeiropoulos und Alejandro Jiménez Rios [11] eine Studie aufgestellt. Sie ähnelt dem Experiment, welches in dieser Arbeit durchgeführt wurde. Sie haben ChatGPT 3.5, ChatGPT 4 und Google Bard getestet, indem sie ihnen Aufgaben aus dem mathematischen Bereich sowie logischen Bereich gestellt haben. Hierbei waren auch Textpuzzle, welche keinen mathematischen Bezug hatten. Es wurden den Chatbots 30 Aufgaben gestellt. Diese wurden drei Mal gestellt, um die Konstanz zu prüfen. Es wurde lediglich die Korrektheit geprüft. Seit der eben beschriebenen Prüfung, welche im Mai durchgeführt wurde, gab es grundlegende Veränderungen in der KI. Die Aufgaben wurden in zwei Sets unterteilt. Set A beinhaltet Aufgaben, welche sich die Forscher selbst überlegt haben und nicht veröffentlicht wurden. Set B sind Aufgaben, welche im Internet veröffentlicht wurden. Dieser Aspekt ist relevant, denn zu diesem Zeitpunkt hatte nur Google Bard eine Internetverbindung. Die anderen beiden Chatbots konnten lediglich auf ihre Datenbank zugreifen. Mittlerweile hat ChatGPT 4 auch die Möglichkeit Informationen aus dem Internet zu bekommen. Im ersten Set hat ChatGPT 4 am besten abgeschnitten mit 80% Erfolgsquote. Danach kamen ChatGPT 3.5 mit 53,3% und Google Bard mit 42,2% Erfolgsquote. Im zweiten Set, indem schwerere Aufgaben gestellt wurden als im ersten Set, hatte Google Bard die beste Quote mit 80%. Danach kamen ChatGPT 4 mit 62,2% und ChatGPT 3.5 mit 42,2% Erfolgsquote. Die Forscher haben von „AI hallucination“ [11], der KI-Halluzination, gesprochen. Hiermit meinen die Forscher, dass die KI eine sehr lange und ausführliche Erklärung zu einer Frage gibt, die auf den ersten Blick richtig aussieht, jedoch falsch ist. Die Literatur ist sich

also einig, dass die KI von einem Menschen weiterhin geprüft werden muss und man nicht alles glauben soll, was ein Chatbot sagt.

7. Konzept des Experiments

a. Aufbau

In der vorliegenden Untersuchung wurden alle vier Chatbots durch dieselben 15 Aufgaben jeweils drei Mal getestet. Insgesamt sind es daher 180 Anfragen an die verschiedenen Chatbots. Ihnen wurden die gleichen Aufgaben mehrfach gestellt, um die Konstanz zu testen und festzustellen, ob sich die Chatbots mit zeitlichem Abstand unterschiedlich verhalten würden. Die Aufgaben selbst werden im nachfolgenden Punkt behandelt. Aus jeder der Aufgaben wurde ein Prompt erstellt, welcher immer gleich bleibt. Dadurch ist sichergestellt, dass die Aufgaben der Chatbots und deren Eingabe identisch bleiben. Die Eingabe wurde jedes Mal mit einem Screenshot festgehalten. Die Ausgabe des jeweiligen Chatbots wurde ebenfalls mit einem oder mehreren Screenshots erfasst. Ergebnisse und der Verlauf der jeweiligen Testrunden wurden in einer Tabelle zusammengetragen und übersichtlich visualisiert. Zusätzlich wurde erfasst, ob die Aufgabe richtig oder falsch gelöst wurde. Das Datum wurde festgehalten, um die Konstanz zu überwachen. Mithilfe eines Notenschlüssels wurde jeder Lösung eine Note vergeben und diese erklärt. Die Konstitution des Notenschlüssels aussieht, wird an einem späteren Punkt der vorliegenden Arbeit erklärt.

b. Aufgaben

Den Chatbots wurden 15 Aufgaben aus 4 verschiedenen mathematischen Themenbereichen gestellt. Das erste Themengebiet ist die Analysis mit 4 Aufgaben. Dort wurden Aufgaben zur Integralrechnung und Extremwertbestimmung gewählt.

Das zweite Themengebiet ist die lineare Algebra mit 4 Aufgaben. In diesem Gebiet wurden Aufgaben zur Matrizenrechnung ausgesucht.

Das dritte Themengebiet ist die Statistik mit 4 Aufgaben. Hier wurden Aufgaben zum arithmetischen Mittel und zur empirischen Varianz gestellt.

Das letzte Themengebiet ist die Stochastik mit 3 Aufgaben. Für diesen Themenbereich wurden Aufgaben zur Wahrscheinlichkeitsrechnung gewählt.

Die genannten Themengebiete waren alle Teil des Studiengangs Wirtschaftsinformatik und haben vielen Studierenden in den von mir besuchten Vorlesungen und Seminaren Schwierigkeiten bereitet. Daher wurden sie für die Untersuchung ausgewählt. Teilweise sind diese Aufgaben Probeklausuren entnommen, zum Teil stammen sie auch von der Seite WMINT [6]. Die Aufgabe 15 wurde von ChatGPT 4 selbst erstellt und von mir auf Sinnhaftigkeit geprüft.

Zur besseren Veranschaulichung folgt hier eine Beispielaufgabe:

Zuerst ist die Eingabe zu sehen, welche ich, in diesem Fall ChatGPT 4, gestellt habe. Es ist die Aufgabe 11 aus dem ersten Durchlauf:

You

Datenreihe $x = 5,8,11,7,4,2,15,6,10,2$ ist gegeben. Erkläre mir das arithmetische Mittel der Datenreihe x als wärst du mein Tutor

Abbildung 1 Input Beispielaufgabe

Im Bild ist zusehen, dass am Ende der Eingabe „als wärst du mein Tutor“ steht. Dies hat den Grund, dass die KIs ohne diesen Zusatz die Aufgabe berechnen, jedoch nicht erklären würden. Daher ist dieser Teilsatz bei jeder Eingabe dabei.

Hiernach folgen die Ausgaben von ChatGPT 4 welche zeigen, wie der Chatbot vorgegangen ist und zu welchem Ergebnis die KI gekommen ist.



ChatGPT

Das arithmetische Mittel, oft einfach als "Durchschnitt" bezeichnet, ist eine grundlegende statistische Maßzahl, die dazu dient, die zentrale Tendenz einer Gruppe von Datenpunkten zu beschreiben. Es wird berechnet, indem die Summe aller Werte in der Datenreihe genommen und dann durch die Anzahl dieser Werte geteilt wird. Das arithmetische Mittel bietet einen einzigen Wert, der repräsentativ für die gesamte Datenmenge ist und hilft, die allgemeine Tendenz der Daten zu verstehen.

Berechnung des arithmetischen Mittels

Für die Berechnung des arithmetischen Mittels einer Datenreihe folgen wir dieser Formel:

$$\bar{x} = \frac{1}{n} \sum_{i=1}^n x_i$$

Dabei ist:

- $\bar{x}$ das arithmetische Mittel,
- n die Anzahl der Datenpunkte in der Datenreihe,
- $\sum$ das Summenzeichen, das anzeigt, dass wir alle folgenden x_i addieren,
- x_i jeder einzelne Datenpunkt in der Datenreihe.

Beispiel

Für die gegebene Datenreihe $x = 5, 8, 11, 7, 4, 2, 15, 6, 10, 2$, berechnen wir das arithmetische Mittel, indem wir die Summe aller Datenpunkte bilden und diese dann durch die Anzahl der Datenpunkte teilen. In diesem Fall haben wir 10 Datenpunkte.

Lass uns das arithmetische Mittel für diese spezifische Datenreihe berechnen.

Das arithmetische Mittel der gegebenen Datenreihe $x = 5, 8, 11, 7, 4, 2, 15, 6, 10, 2$ beträgt 7.0.

Abbildung 2 Beispielaufgabe Output 1

Es ist jedoch wichtig zu beachten, dass das arithmetische Mittel durch Ausreißer (sehr hohe oder sehr niedrige Werte im Vergleich zum Rest der Daten) beeinflusst werden kann. In solchen Fällen kann der Median (der mittlere Wert in einer geordneten Liste von Datenpunkten) manchmal ein besseres Maß für die zentrale Tendenz sein. [↗]

Abbildung 3 Beispielaufgabe Output 2

Erklärung

Das bedeutet, dass der Durchschnittswert aller Datenpunkte in dieser Reihe 7.0 ist. Wenn wir uns vorstellen, dass wir die Gesamtsumme der Werte gleichmäßig auf alle Datenpunkte verteilen würden, dann hätte jeder Datenpunkt den Wert 7.0. Das arithmetische Mittel hilft uns, ein Gefühl für die "Mitte" oder den "Durchschnitt" der Daten zu bekommen, und bietet eine einfache Möglichkeit, verschiedene Datensätze zu vergleichen.

Abbildung 4 Beispielaufgabe Output 3

Hiernach würde ein Text folgen, der erklärt, warum diese Aufgabe von mir gewählt wurde, was die KI richtig und falsch gemacht hat, welche Note die KI bekommen hat sowie eine Erklärung für die Benotung.

c. Bewertung

Die Aufgaben wurden mit einer Skala von 1 bis 5 bewertet. Dabei ist eins das Beste und fünf das Schlechteste. Die Bewertung ergibt sich aus sieben Kriterien, die sich wie folgt zusammensetzten.

Das erste Kriterium ist, ob eine Lösung vorhanden ist. Während des Experimentes kam es immer wieder zu Situationen, in denen die Chatbots nicht bis zur Lösung vordringen konnten. Ohne eine Lösung zur gestellten Aufgabe gibt es keinen Nutzen für Studierende unter der bekannten Fragestellung, da sie ihre eigenen Lösungen nicht mithilfe des Chatbots abgleichen können.

Das zweite Kriterium ist, ob die Aufgabe richtig gelöst wurde. Dieser Umstand ist sehr wichtig, da die Chatbots Sprachmodelle und keine Taschenrechner sind und daher viele mathematische Rechenfehler machen.

Das dritte Kriterium ist, ob die Berechnung vorhanden ist und nicht nur Ergebnisse angezeigt werden, da für Studierende der richtige Lösungsweg nachvollziehbar sein muss, um eine Aufgabe selbst zu verstehen.

Das vierte Kriterium ist, ob die Berechnung korrekte und nachvollziehbare Rechenschritte aufzeigt. Die Chatbots überspringen

oft Schritte, wodurch Nutzerinnen die Logik der Rechenschritte verborgen bleibt und die KI-Lösung nicht auf eigene Lösungsansätze übertragbar ist.

Das fünfte Kriterium ist, ob eine Erklärung vorhanden ist. Die Rechenschritte allein reichen meist nicht, um die Aufgabe zu verstehen. Ein Tutor erklärt zusätzlich zum Rechenschritt die nötigen kognitiven Operationen, daher war es wichtig zu prüfen, ob die Chatbots ihre Rechenwege mit Worten erklären können.

Das sechste Kriterium ist, ob die Erklärung ausreichend lang ist. Die Erklärung muss auch genügend Informationen in ganzen Sätzen vermitteln, die die Aufgabe Schritt für Schritt erklären.

Das siebte Kriterium ist, ob die Erklärung auch ausreichend verständlich ist. Wenn die Chatbots nur in Fachtermini schreiben und die Begriffe nicht erklären, kann die Erklärung schwer verständlich sein.

In der Bewertung gab es noch einen weiteren Punkt. Wenn die Chatbots etwas außerhalb der Bewertungsskala besonders gut oder besonders schlecht gemacht haben, ist dies in die Bewertung mit eingeflossen.

Teilweise wurden Aufgaben auch mit 1 bewertet, obwohl diese nicht perfekt gelöst wurden. Der Untersuchung liegt der Ansatz zugrunde, dass die Frage nach der möglichen Tutoren Schafft Vorrang vor inhaltlicher Perfektion hat. Ein Tutor kann eine Aufgabe verstanden haben und erklären können und sich dennoch Rechenfehler leisten. Die Fragestellung dieser Arbeit fragt nicht nach der Funktion von KI als verbesserte menschliche Tutoren, sondern nach der Vergleichbarkeit von menschlichen Tutoren und KI als Tutor. Demnach ist eine Fehlertoleranz vertretbar. Das heißt wenn eine Aufgabe gut erklärt ist, kann es eine eins geben, auch wenn die Aufgabe nicht komplett in Rechenschritten beschrieben wurde. Außerdem wurde nicht immer eine 5 vergeben, wenn das Ergebnis falsch oder nicht vorhanden war.

8. Resultate der Aufgaben

Alle Ergebnisse der Chatbots werden im Folgenden nicht angegeben, da es 180 Ergebnisse gibt. Die Chatbots werden in folgender Reihenfolge behandelt: ChatGPT 3.5, ChatGPT 4, Google Bard und Copilot. Dabei werden die statistischen Ergebnisse aufgezeigt. Die Kennzahlen sind dabei die Durchschnittsnote, welche die Chatbots erhalten haben. Sie zeigen, wie oft eine Aufgabe mit der Note 5 bewertet wurde, wie viele Aufgaben richtig gelöst wurden und wie viele Aufgaben falsch gelöst wurden. Dabei werden nochmal die Teilgebiete genauer behandelt. Wie oft eine Aufgabe mit der Note 5 bewertet wurde, ist relevant, weil diese Ausgaben der Chatbots unbrauchbar sind und einem Studierenden eher schaden als helfen. Außerdem wird anhand von zwei Beispielaufgaben gezeigt, wie die Chatbots vorgegangen sind. Dabei werden eine gut gelöste Aufgabe und eine schlecht gelöste Aufgabe vorgestellt. Diese Aufgaben können unterschiedlich sein, weil die Chatbots unterschiedliche Stärken und Schwächen in den verschiedenen Gebieten haben.

a. ChatGPT 3.5

In der folgenden Tabelle wird gezeigt, wie ChatGPT 3.5 die Aufgaben des Experiments gelöst hat.

Tabelle 1 Ergebnis Experiment ChatGPT 3.5

Thema	Anzahl Aufgaben	Korrekt	Falsch	Durchschnittsnote	Anzahl Note 5
Analysis	12	7	5	2,92	3
Lineare Algebra	12	2	10	4,17	7
Statistik	12	9	3	1,75	0
Stochastik	9	2	7	4,22	6
Insgesamt	45	20	25	3,2	16

Hierbei fällt auf, dass ChatGPT 3.5 die wenigsten Probleme mit Statistik hatte. Lineare Algebra und Stochastik waren ähnlich problematisch

wobei die Stochastik eine Aufgabe und damit drei Antworten weniger hat.

Exemplarische Aufgaben

Im Folgenden werden zwei Aufgaben exemplarisch für die Leistung von ChatGPT 3.5 gezeigt. Eine Aufgabe, die besonders gut gelöst wurde und eine Aufgabe die schlecht gelöst wurde.

Beispiel für eine gute Aufgabe

Eine gut gelöste Aufgabe ist die Aufgabe 12 im ersten Durchlauf. Sie gehört zu den Statistik Aufgaben. Im ersten Prompt ist die Aufgabenstellung für ChatGPT 3.5 zu sehen:

You

Datenreihe $x = 5,8,11,7,4,2,15,6,10,2$ ist gegeben. Erkläre mir die empirische Varianz s^2_x der Datenreihe x als wärst du mein Tutor

Abbildung 5 ChatGPT 3.5 erste Aufgabe Input 1

Diese Aufgabe stammt aus einer Probeklausur der vergangenen Jahre. Interessant bei dieser Aufgabe ist, dass der Chatbot erkennen muss, dass die Datenreihe nicht sortiert ist, zudem muss er erkennen, dass es keine Kommazahlen in der Reihe gibt. ChatGPT 3.5 hatte in einer vorherigen Aufgabe erklärt, dass er die Datenreihe in dieser Form angegeben bekommen möchte.

In den folgenden Prompts sieht man die Antwort von ChatGPT 3.5:

**ChatGPT**

Die empirische Varianz s_x^2 einer Datenreihe x ist ein Maß für die Streuung oder Variation der Daten um ihr arithmetisches Mittel. Sie gibt an, wie weit die einzelnen Beobachtungen von diesem Mittelwert entfernt sind.

Um die empirische Varianz s_x^2 einer Datenreihe x zu berechnen, führen wir die folgenden Schritte durch:

1. Berechnen Sie das arithmetische Mittel $\bar{x}$ der Datenreihe.
2. Berechnen Sie die quadratische Abweichung jedes Datenpunkts von diesem arithmetischen Mittel.
3. Addieren Sie alle quadratischen Abweichungen auf.
4. Teilen Sie diese Summe durch die Anzahl der Datenpunkte minus 1.

Die Formel zur Berechnung der empirischen Varianz lautet:

$$s_x^2 = \frac{1}{n-1} \sum_{i=1}^n (x_i - \bar{x})^2$$

Abbildung 6 ChatGPT 3.5 erste Aufgabe Output 1

wobei:

- n die Anzahl der Datenpunkte in der Datenreihe ist,
- x_i der i -te Datenpunkt in der Datenreihe ist,
- $\bar{x}$ das arithmetische Mittel der Datenreihe ist.

Für die gegebene Datenreihe $x = 5, 8, 11, 7, 4, 2, 15, 6, 10, 2$ gehen wir wie folgt vor:

1. Berechnen des arithmetischen Mittels $\bar{x}$:

$$\bar{x} = \frac{5+8+11+7+4+2+15+6+10+2}{10} = \frac{70}{10} = 7$$

1. Berechnen der quadratischen Abweichungen jedes Datenpunkts von $\bar{x}$:

$$(5 - 7)^2 = 4$$

$$(8 - 7)^2 = 1$$

$$(11 - 7)^2 = 16$$

$$(7 - 7)^2 = 0$$

$$(4 - 7)^2 = 9$$

$$(2 - 7)^2 = 25$$

$$(15 - 7)^2 = 64$$

$$(6 - 7)^2 = 1$$

$$(10 - 7)^2 = 9$$

$$(2 - 7)^2 = 25$$

Abbildung 7 ChatGPT 3.5 erste Aufgabe Output 2

1. Addieren Sie alle quadratischen Abweichungen auf:

$$4 + 1 + 16 + 0 + 9 + 25 + 64 + 1 + 9 + 25 = 154$$

1. Teilen Sie diese Summe durch die Anzahl der Datenpunkte minus 1:

$$s_x^2 = \frac{154}{10-1} = \frac{154}{9} \approx 17.1111$$

Die empirische Varianz s_x^2 der Datenreihe x beträgt also ungefähr 17.1111.

Die empirische Varianz ist eine wichtige Kennzahl für die Streuung der Daten um ihren Mittelwert. Sie wird oft verwendet, um die Varianz oder Streuung in Stichproben oder Beobachtungen zu quantifizieren.

Abbildung 8 ChatGPT 3.5 erste Aufgabe Output 3

Hervorzuheben ist an dieser Stelle, dass ChatGPT 3.5 auf das richtige Ergebnis gekommen ist. Gerade bei einer Aufgabe wie der empirischen

Varianz ist zu beobachten, dass die Chatbots Probleme beim Lösen der Aufgaben haben, weil diese Potenzrechnung beinhalten und die Chatbots Aufgaben dazu nicht immer richtig lösen. Die Formatierung in dieser Aufgabe ist gelungen, sowohl die Formel für die Varianz als auch der Bruch für die Berechnung ist gut erkennbar. Die Aufgabe wird erst in vier Punkte heruntergebrochen, um zu zeigen, wie sie zu lösen ist. Danach wird die Formel für die Varianz aufgestellt und sehr kleinschrittig erklärt. An dieser Stelle haben Chatbots die Formel nur angegeben und nicht erklärt, wie diese zu benutzen ist. In der Berechnung für das arithmetische Mittel ist der gesamte Bruch zu sehen. Dies ist auch nicht immer der Fall. Es zeigt sehr gut, wie zu rechnen ist und man kann prüfen, ob man selbst einen Fehler gemacht hat. Die Berechnung der quadratischen Abweichung für jeden Datenpunkt ist ebenfalls komplett angegeben. Ein weiterer positiver Punkt ist, dass ChatGPT 3.5 angegeben hat, dass die 17,1111 das ungefähre Ergebnis ist. Am Ende erfolgt noch eine kurze Erklärung, wozu man die empirische Varianz benötigt. Ein Kritikpunkt wäre, dass am Anfang zwar der Ablauf der Lösung der Aufgabe nummeriert ist, in der Lösung selbst jedoch jeder Punkt mit „1.“ anfängt. Für diese Antwort hat ChatGPT 3.5 die Note 1 bekommen. Ich habe mich für diese Benotung entschieden, weil sie korrekt gelöst wurde, eine strukturierte und nachvollziehbare Formatierung hat und ChatGPT 3.5 in vielen kleinen Schritten erklärt, wie vorgegangen wurde.

Beispiel für eine Schlechte Aufgabe

Eine schlecht gelöste Aufgabe ist die Aufgabe 5 im zweiten Durchlauf. Sie gehört zu den Aufgaben der linearen Algebra. Im ersten Prompt ist die Aufgabenstellung für ChatGPT 3.5 zusehen:

You

Die Matrix $A = \begin{bmatrix} 3 & 2 & 6 & -1 \\ 1 & 1 & 2 & 1 \\ 2 & 1 & 4 & -2 \\ 1 & 0 & 2 & -3 \end{bmatrix}$ und der Vektor $b = [6, 0, 6, 6]$ seien gegeben. Berechnen Sie die Lösungsmenge des linearen Gleichungssystems $Ax = b$ und erkläre es mir als wärst du mein Tutor

Abbildung 9 ChatGPT 3.5 zweite Aufgabe Input 1

Diese Aufgabe kommt aus einer Probeklausur der vergangenen Jahre. ChatGPT 3.5 hatte in einer vorherigen Aufgabe erklärt, dass die Matrix und den Vektor in dieser Form angegeben werden solle.

In den folgenden Prompts sieht man die Antwort von ChatGPT 3.5:



ChatGPT

Um die Lösungsmenge des linearen Gleichungssystems $Ax = b$ zu berechnen, wobei A die gegebene Matrix und b der gegebene Vektor sind, können wir den Gauss-Jordan-Eliminationsalgorithmus oder andere Methoden der linearen Algebra verwenden.

Zuerst schreiben wir das erweiterte Matrix-System:

$$[A|b] = \left[\begin{array}{cccc|c} 3 & 2 & 6 & -1 & 6 \\ 1 & 1 & 2 & 1 & 0 \\ 2 & 1 & 4 & -2 & 6 \\ 1 & 0 & 2 & -3 & 6 \end{array} \right]$$

Abbildung 10 ChatGPT 3.5 zweite Aufgabe Output 1

Dann führen wir Reihenoperationen durch, um die erweiterte Matrix in eine reduzierte Zeilenstufenform zu bringen. Ich werde die Reihenoperationen durchführen und die Schritte erklären:

1. $R_2 = R_2 - \frac{1}{3}R_1$
2. $R_3 = R_3 - \frac{2}{3}R_1$
3. $R_4 = R_4 - \frac{1}{3}R_1$

$$\left[\begin{array}{cccc|c} 3 & 2 & 6 & -1 & 6 \\ 0 & \frac{1}{3} & 0 & \frac{4}{3} & -2 \\ 0 & -\frac{1}{3} & 0 & -\frac{8}{3} & 4 \\ 0 & -\frac{2}{3} & 2 & -2 & 4 \end{array} \right]$$

1. $R_3 = R_3 + R_2$
2. $R_4 = R_4 + 2R_2$

$$\left[\begin{array}{cccc|c} 3 & 2 & 6 & -1 & 6 \\ 0 & \frac{1}{3} & 0 & \frac{4}{3} & -2 \\ 0 & 0 & 0 & \frac{4}{3} & 2 \\ 0 & 0 & 2 & 2 & 0 \end{array} \right]$$

1. $R_4 = R_4 - R_3$

$$\left[\begin{array}{cccc|c} 3 & 2 & 6 & -1 & 6 \\ 0 & \frac{1}{3} & 0 & \frac{4}{3} & -2 \\ 0 & 0 & 0 & \frac{4}{3} & 2 \\ 0 & 0 & 2 & 0 & -2 \end{array} \right]$$

1. $R_3 = \frac{3}{4}R_3$
2. $R_4 = \frac{1}{2}R_4$

Abbildung 11 ChatGPT 3.5 zweite Aufgabe Output 2

$$\left[\begin{array}{cccc|c} 3 & 2 & 6 & -1 & 6 \\ 0 & \frac{1}{3} & 0 & \frac{4}{3} & -2 \\ 0 & 0 & 0 & 1 & \frac{3}{2} \\ 0 & 0 & 1 & 0 & -1 \end{array} \right]$$

1. $R_2 = R_2 - \frac{4}{3}R_3$
2. $R_1 = R_1 - 6R_3 + R_4$

$$\left[\begin{array}{cccc|c} 3 & 2 & 0 & 0 & 0 \\ 0 & \frac{1}{3} & 0 & 0 & -4 \\ 0 & 0 & 0 & 1 & \frac{3}{2} \\ 0 & 0 & 1 & 0 & -1 \end{array} \right]$$

Abbildung 12 ChatGPT 3.5 zweite Aufgabe Output 3

$$1. R_1 = R_1 - 2R_2$$

$$\left[\begin{array}{cccc|c} 3 & 0 & 0 & 0 & 8 \\ 0 & \frac{1}{3} & 0 & 0 & -4 \\ 0 & 0 & 0 & 1 & \frac{3}{2} \\ 0 & 0 & 1 & 0 & -1 \end{array} \right]$$

Jetzt haben wir die reduzierte Zeilenstufenform der erweiterten Matrix erreicht. Diese Matrix repräsentiert das Gleichungssystem in einer Form, in der die Lösung leicht abgelesen werden kann.

Die Lösungsmenge des Gleichungssystems ist:

$$x_1 = 8$$

$$x_2 = -12$$

$$x_3 = \frac{3}{2}$$

$$x_4 = -1$$

Das bedeutet, dass die Lösungsmenge des Gleichungssystems $Ax = b$ die einzige Lösung ist, die durch den Vektor $[8, -12, \frac{3}{2}, -1]$ dargestellt wird.

Abbildung 13 ChatGPT 3.5 zweite Aufgabe Output 4

Diese Aufgabe wurde nicht korrekt gelöst. Ein positiver zu erwähnender Punkt ist, dass ChatGPT 3.5 wieder eine gute Formatierung hat. Außerdem werden die Rechenschritte erklärt, wodurch man die Rechenfehler erkennen kann. Ich habe diese Aufgabe gewählt, weil man beim schnellen Durchlesen nicht direkt merkt, dass die Aufgabe falsch ist. Beim ersten Rechenschritt treten schon einige Fehler auf. R_2 ist noch korrekt, aber R_3 und R_4 haben einige Fehler. ChatGPT 3.5 rechnet $R_3 = R_3 - \frac{2}{3} R_1$. Wenn man dann die Werte einfügt, erhält man für $x_3 = -2 - \frac{2}{3} * (-1)$ das ergibt $-\frac{4}{3}$ und nicht $-\frac{8}{3}$. Das gleiche Problem tritt noch häufiger im selben Schritt auf. Der nächste Wert, welcher eine 4 laut ChatGPT 3.5 ist, sollte eine 2 sein und in der nächsten Zeile ist die 2 eine 0 und die -2 ist $-\frac{8}{3}$. Diese Fehler sind nur ein kurzer Einblick. Die gesamte Aufgabe durchzugehen, würde zu lange dauern. Allgemein ist diese Aufgabe zwar in vielen Rechenschritten angegeben, diese sind jedoch kompliziert und schwer nachzuvollziehen. ChatGPT 3.5 erhielt für diese Aufgabe die Note 5, weil sie falsch gelöst wurde und viele Rechenfehler zeigt. Die Berechnung wurde nicht ausführlich genug

beschrieben, um die Note zu verbessern. Ich habe mich für diese Aufgabe entschieden, weil die Formatierung der Matrizen anschaulich gemacht wird und zeigt, dass die Aufgaben auf den ersten Blick korrekt aussehen, der Chatbot jedoch sehr viele Rechenfehler in der Berechnung gemacht hat.

Weitere Bemerkungen

Im Folgenden sind zwei Outputs von ChatGPT 3.5 zu sehen, welche zeigen, wie unterschiedlich die Ausgaben von ChatGPT 3.5 sein können. Die Ausgaben sind von Aufgabe 14 im ersten und zweiten Durchlauf.

Zuerst erstellen wir eine Kontingenztafel, die die beobachteten Häufigkeiten der Kombinationen von Studiengang und Lieblingsfarbe enthält:

	Rot	Grün	Blau
WI	12	8	30
MI	10	7	33
IT	13	13	24
AI	5	12	33

Als nächstes berechnen wir die erwarteten Häufigkeiten unter der Annahme der Unabhängigkeit von Studiengang und Lieblingsfarbe. Dazu multiplizieren wir die Randhäufigkeiten (Summen über die Zeilen und Spalten) und dividieren durch die Gesamtanzahl der Beobachtungen:

Die Gesamtanzahl der Beobachtungen beträgt $12 + 8 + 30 + 10 + 7 + 33 + 13 + 13 + 24 + 5 + 12 + 33 = 200$.

Die erwarteten Häufigkeiten wären:

	Rot	Grün	Blau
WI	$\frac{12 \times (12+8+30)}{200}$	$\frac{8 \times (12+8+30)}{200}$	$\frac{30 \times (12+8+30)}{200}$
MI	$\frac{10 \times (10+7+33)}{200}$	$\frac{7 \times (10+7+33)}{200}$	$\frac{33 \times (10+7+33)}{200}$
IT	$\frac{13 \times (13+13+24)}{200}$	$\frac{13 \times (13+13+24)}{200}$	$\frac{24 \times (13+13+24)}{200}$
AI	$\frac{5 \times (5+12+33)}{200}$	$\frac{12 \times (5+12+33)}{200}$	$\frac{33 \times (5+12+33)}{200}$

Abbildung 14 ChatGPT 3.5 weitere Bemerkungen Output 1

Hier ist zu sehen, dass die Ausgabe in übersichtlichen Tabellen gegeben wird. Die Untere Tabelle zeigt den gesamten Rechenschritt, mit dessen Hilfe die Werte berechnet werden. Eine Tabelle oder anschauliche Formate wie diese kann den Studierenden helfen,

Probleme schneller zu verstehen, weil es strukturiert, dargestellt und somit leichter nachvollzogen werden kann. Im Gegensatz dazu kann es Studierende verwirren, wenn erst die Formatierung verstanden werden muss, bevor die Aufgabe selbst verstanden werden kann.

Zunächst müssen wir die beobachteten und erwarteten Häufigkeiten in einer Kreuztabelle organisieren:

diff Copy code

	Rot	Grün	Blau
WI	12	8	30
MI	10	7	33
IT	13	13	24
AI	5	12	33

Abbildung 15 ChatGPT 3.5 weitere Bemerkung Output 2

Jetzt berechnen wir die erwarteten Häufigkeiten unter der Annahme der Unabhängigkeit. Dazu multiplizieren wir die Randsummen für jede Zelle und teilen durch die Gesamtzahl der Beobachtungen:

diff Copy code

	Rot	Grün	Blau
WI	11.3	11.3	27.3
MI	11.7	11.7	28.3
IT	11.2	11.2	27.6
AI	11.8	11.8	28.8

Abbildung 16 ChatGPT 3.5 weitere Bemerkung Output 3

Die zweite Ausgabe zeigt den großen und relevanten Unterschied. Hier wird die Tabelle über eine Fläche mit ASCII-Zeichen angezeigt. Im Vergleich zur ersten Ausgabe wird hier in der zweiten Tabelle nur das Ergebnis auf die erste Nachkommastelle gerundet angezeigt. Diese Tabellen sind weniger intuitiv. Dieses Beispiel kann man schnell

verstehen. Spätere Beispiele von anderen Chatbots zeigen noch besser, wie schlecht die Formatierung der Chatbots sein kann. Außerdem zeigt die Ausgabe, wie unterschiedlich die Lösungen der Chatbots sein können, obwohl dieselbe Aufgabenstellung vorliegt.

Das folgende Bild zeigt die Interaktion mit ChatGPT 3.5 von der Aufgabe 15 im zweiten Durchlauf.

Wenn du möchtest, kann ich die Berechnungen durchführen.

You

kannst du mir die werte berechnen?

ChatGPT

Natürlich, ich führe die Berechnungen für dich durch:

Abbildung 17 ChatGPT 3.5 weitere Bemerkung Output 4

Hier hat ChatGPT 3.5 die Aufgabe erst nicht bis zum Ende gerechnet und gefragt, ob die Berechnung gemacht werden soll. An dieser Stelle hat es funktioniert, dass der Chatbot weiter gerechnet hat. Dies hat bei Copilot nicht immer funktioniert wie später zusehen ist.

b. ChatGPT 4

In der folgenden Tabelle wird gezeigt, wie ChatGPT 4 die Aufgaben des Experiments gelöst hat.

Tabelle 2 Ergebnis Experiment ChatGPT 4

Thema	Anzahl Aufgaben	Korrekt	Falsch	Durchschnittsnote	Anzahl Note 5
Analysis	12	9	3	2,25	0
Lineare Algebra	12	12	0	2,25	0
Statistik	12	12	0	1,75	0
Stochastik	9	9	0	1,67	0
Insgesamt	45	42	3	2,04	0

Hierbei fällt auf, dass ChatGPT 4 keine Note 5 bekommen und die beste Note in der Stochastik bekommen hat.

Exemplarische Aufgaben

Beispiel für eine gute Aufgabe

Die Aufgabe, die ich als Beispiel für eine gute Aufgabe genommen habe, ist die Aufgabe 1 im ersten Durchlauf. Sie gehört zu den Analysis Aufgaben. Im ersten Prompt ist die Aufgabenstellung für ChatGPT 4 zu sehen:

You

Die Funktion $f: D \subset \mathbb{R} \rightarrow \mathbb{R}$ mit $f(x) = \sqrt{1+x}$ sei gegeben. Berechnen Sie den Inhalt der Fläche, die von dem Graphen Γ_f von f und den Geraden $y=2$ und $x=0$ in der (x,y) -Ebene eingeschlossen wird und erkläre es mir als wärst du mein Tutor

Abbildung 18 ChatGPT 4 erste Aufgabe Input 1

Diese Aufgabe kommt aus einer Probeklausur der vergangenen Jahre. Es ist eine der anspruchsvolleren Aufgaben aus der Aufgabenliste. Die Chatbots müssen hier die Definition von f verstehen und verarbeiten. Außerdem gibt es den Graphen von f und die Gerade, welche beide von den Chatbots miteinander verknüpft werden müssen.

In den folgenden Prompts sieht man die Antwort von ChatGPT 4:



ChatGPT

Um den Inhalt der Fläche zu berechnen, die vom Graphen Γ_f der Funktion $f(x) = \sqrt{1+x}$ und den Geraden $y = 2$ und $x = 0$ in der (x, y) -Ebene eingeschlossen wird, gehen wir schrittweise vor. Da $f(x)$ eine Wurzelfunktion ist, ist ihr Definitionsbereich $D = \{x \in \mathbb{R} | x \geq -1\}$, weil das Argument der Wurzel nicht negativ sein darf.

Schritt 1: Schnittpunkte bestimmen

Zunächst müssen wir den Schnittpunkt S der Kurve $y = \sqrt{1+x}$ mit der Geraden $y = 2$ finden, um die Grenzen des Integrals zu bestimmen, das wir verwenden werden, um die Fläche zu berechnen. Dafür setzen wir $\sqrt{1+x} = 2$ und lösen die Gleichung nach x auf.

Abbildung 19 ChatGPT 4 erste Aufgabe Output 1

Schritt 2: Integral aufstellen

Nachdem wir den Schnittpunkt gefunden haben, stellen wir das Integral auf, das den Flächeninhalt unter Γ_f von $x = 0$ bis zum gefundenen Schnittpunkt S und oberhalb der x -Achse berechnet. Da $y = 2$ die obere Grenze ist, berechnen wir die Differenz der Flächen, die von $y = 2$ und $y = \sqrt{1+x}$ begrenzt werden, von $x = 0$ bis zum Schnittpunkt.

Schritt 3: Integral berechnen

Das Integral, das wir berechnen müssen, ist:

$$A = \int_0^{x_S} (2 - \sqrt{1+x}) dx$$

wobei x_S der x -Wert des Schnittpunkts S ist.

Schritt 4: Erklärung

- **Warum diese Methode?** Die Berechnung des Flächeninhalts zwischen einer Kurve und einer Geraden (oder zwischen zwei Kurven) erfolgt durch Integration. Das Integral berechnet die "Summe" aller unendlich kleinen Flächenstreifen (Infinitesimale) zwischen der Kurve und der Geraden über das Intervall, das durch die Schnittpunkte gegeben ist.
- **Warum die Differenz $2 - \sqrt{1+x}$?** Da wir die Fläche zwischen $y = 2$ und $y = \sqrt{1+x}$ berechnen, subtrahieren wir die untere Funktion ($\sqrt{1+x}$) von der oberen Grenze ($y = 2$), um die Höhe jedes infinitesimalen Streifens zu bestimmen.
- **Warum von $x = 0$?** Die Gerade $x = 0$ ist die y -Achse, die als eine Grenze für die Fläche dient, die wir berechnen möchten. Sie begrenzt die Fläche von der linken Seite.

Lass uns jetzt mit der Berechnung beginnen, angefangen mit der Bestimmung des Schnittpunktes S .

Der Schnittpunkt S der Kurve $y = \sqrt{1+x}$ mit der Geraden $y = 2$ liegt bei $x = 3$. Das bedeutet, dass die Fläche, die wir berechnen möchten, zwischen $x = 0$ und $x = 3$ liegt.

Abbildung 20 ChatGPT 4 erste Aufgabe Output 2

Nun berechnen wir das Integral $A = \int_0^3 (2 - \sqrt{1+x}) dx$, um den Flächeninhalt zu ermitteln.

[-]

Der Inhalt der Fläche, die vom Graphen Γ_f von $f(x) = \sqrt{1+x}$, der Geraden $y = 2$, und $x = 0$ in der (x, y) -Ebene eingeschlossen wird, beträgt $\frac{4}{3}$ Quadrat-Einheiten.

Zusammenfassung des Vorgangs:

1. **Schnittpunkte bestimmen:** Wir haben herausgefunden, dass der Schnittpunkt von $f(x) = \sqrt{1+x}$ mit $y = 2$ bei $x = 3$ liegt.
2. **Integral aufstellen:** Das Integral wurde aufgestellt, um den Flächeninhalt zwischen den Kurven $y = \sqrt{1+x}$ und $y = 2$ von $x = 0$ bis $x = 3$ zu berechnen.
3. **Integral berechnen:** Durch Integration fanden wir heraus, dass die Fläche $\frac{4}{3}$ beträgt.

Abbildung 21 ChatGPT 4 erste Aufgabe Output 3

Diese Methode nutzt die Grundlagen der Integralrechnung, um den Flächeninhalt zwischen einer Kurve und einer Geraden (oder zwischen zwei Kurven) zu bestimmen. Es ist eine nützliche Technik in vielen Bereichen der Mathematik und Physik, um Flächen, Volumen und weitere Größen zu berechnen. [-]

🔊 📄 ↺ 🗑️

Abbildung 22 ChatGPT 4 erste Aufgabe Output 4

Diese Aufgabe hat ChatGPT 4 korrekt gelöst. Die Aufgabe ist übersichtlich durchgearbeitet worden. Der Vorgang ist nummeriert und in Abschnitte aufgeteilt, welche jeweils mit einem Titel versehen wurden, welcher den Inhalt der Durchführung beschreibt. Die Formatierung ist verständlich und einfach zu lesen. Die Definition von f und dabei der Definitionsbereich D wurden korrekt verstanden und auch verständlich abgebildet. Die Rechenschritte an sich, also die genaue Berechnung, ist nicht zusehen, dafür hat ChatGPT 4 alles sehr präzise beschrieben. ChatGPT 4 berechnet $\sqrt{1+x} = 2$ nicht und schreibt jedes Mal x_5 statt „3“, was die Lösung des Terms ist. Am Ende hat ChatGPT 4 noch eine kurze Zusammenfassung des Vorgangs abgebildet, welche genau beschreibt, wie vorgegangen wurde und wofür die Methode benötigt wird. ChatGPT hat für diese Aufgabe die Note 1 bekommen. Ich habe mich für diese Aufgabe und Benotung entschieden, weil sie das Paradebeispiel für die Ausgaben von ChatGPT 4 ist. In den meisten Aufgaben ist die Berechnung selbst nicht gezeigt, dafür sehr detailliert erklärt. Ich bin der Meinung, dass eine vorgerechnete Aufgabe

anschaulicher ist, daher hat ChatGPT 4 häufiger die Note 2 statt 1 bekommen, weil die Aufgabe nicht gut genug erklärt wurde. Die Formatierung sieht bei ChatGPT 4 immer gut aus und die Aufgaben haben eine verständliche Struktur.

Beispiel für eine Schlechte Aufgabe

Die Aufgabe, die ich als Beispiel für eine schlechte Aufgabe genommen habe, ist wieder die Aufgabe 1, jedoch dieses Mal im zweiten Durchlauf. Im ersten Prompt ist die Aufgabenstellung für ChatGPT 4 zu sehen:

You

Die Funktion $f: D \subset \mathbb{R} \rightarrow \mathbb{R}$ mit $f(x) = \sqrt{1+x}$ sei gegeben. Berechnen Sie den Inhalt der Fläche, die von dem Graphen Γ_f von f und den Geraden $y=2$ und $x=0$ in der (x,y) -Ebene eingeschlossen wird und erkläre es mir als wärst du mein Tutor

Abbildung 23 ChatGPT 4 zweite Aufgabe Input 1

Im Prompt ist zu sehen, dass es die identische Aufgabenstellung ist, wie im ersten Durchlauf. ChatGPT 4 hat dennoch eine andere Lösung erzeugt.

In den folgenden Prompts sieht man die Antwort von ChatGPT 4:



ChatGPT

Um die Fläche, die von dem Graphen Γ_f von $f(x) = \sqrt{1+x}$, der Geraden $y = 2$, und der Geraden $x = 0$ eingeschlossen wird, zu berechnen, folgen wir diesen Schritten:

1. Schnittpunkte bestimmen

Zuerst müssen wir die Schnittpunkte der Funktion $f(x) = \sqrt{1+x}$ mit der Geraden $y = 2$ finden, um die Grenzen des Integrals zu bestimmen. Dazu setzen wir $f(x)$ gleich 2:

$$\sqrt{1+x} = 2$$

2. Integralgrenzen

Nach dem Bestimmen der Schnittpunkte wissen wir, dass der Bereich, über den wir integrieren müssen, von $x = 0$ bis zu dem x -Wert des Schnittpunktes reicht.

3. Integral aufstellen

Um den Flächeninhalt zu berechnen, integrieren wir die Funktion $f(x) = \sqrt{1+x}$ von $x = 0$ bis zum x -Wert des Schnittpunktes und subtrahieren davon die Fläche unter der Geraden $y = 2$ im selben Intervall.

4. Berechnung durchführen

Wir führen die Integration durch, um den exakten Flächeninhalt zu bestimmen.

Lass uns nun mit Schritt 1 beginnen und den Schnittpunkt von $f(x)$ mit $y = 2$ berechnen. Der Schnittpunkt der Funktion $f(x) = \sqrt{1+x}$ mit der Geraden $y = 2$ liegt bei $x = 3$. Das bedeutet, dass der Bereich, über den wir integrieren müssen, von $x = 0$ bis $x = 3$ reicht.

Nun setzen wir das Integral auf, um den Flächeninhalt zu berechnen. Die Fläche A unter der Kurve von $f(x)$ von $x = 0$ bis $x = 3$ wird durch das Integral

$$A = \int_0^3 \sqrt{1+x} \, dx$$

berechnet. Da die Gerade $y = 2$ oberhalb des Graphen von $f(x)$ im gesamten Intervall liegt, müssen wir keine Fläche subtrahieren – das Integral gibt uns direkt den gesuchten Flächeninhalt.

Lass uns das Integral berechnen. [↗]

Der Inhalt der Fläche, die von dem Graphen Γ_f von $f(x) = \sqrt{1+x}$, der Geraden $y = 2$, und der Geraden $x = 0$ in der (x, y) -Ebene eingeschlossen wird, beträgt etwa 4.67 Flächeneinheiten.

Abbildung 24 ChatGPT 4 zweite Aufgabe Output 1

1. **Schnittpunkte bestimmen:** Wir haben herausgefunden, dass der Schnittpunkt von $f(x)$ mit $y = 2$ bei $x = 3$ liegt. Das heißt, ab $x = 3$ ist die Funktion $f(x)$ kleiner als 2, was uns die obere Grenze für unsere Berechnung gibt.
2. **Integralgrenzen:** Da die Fläche, die wir berechnen wollen, zwischen $x = 0$ und $x = 3$ liegt, verwenden wir diese beiden Werte als unsere Integrationsgrenzen.
3. **Flächenberechnung:** Durch das Integrieren der Funktion $f(x) = \sqrt{1+x}$ über das Intervall von $x = 0$ bis $x = 3$ haben wir den Flächeninhalt gefunden. Dieses Integral gibt uns direkt die Größe der Fläche, da wir die Fläche unterhalb der Kurve $f(x)$ und oberhalb der x -Achse messen, ohne dass eine Subtraktion nötig ist, weil die Gerade $y = 2$ konstant bleibt und nur als obere Begrenzung dient.
4. **Ergebnis:** Das Ergebnis von 4.67 Flächeneinheiten repräsentiert den gesuchten Flächeninhalt, der durch diese Kurven und Geraden eingeschlossen wird.

Falls noch Fragen zum Vorgehen oder zu den einzelnen Schritten bestehen, stehe ich gerne zur Verfügung! [↩]

Abbildung 25 ChatGPT 4 zweite Aufgabe Output 2

Diese Aufgabe ist nicht korrekt, obwohl ChatGPT 4 dieselbe Aufgabe 12 Tage vorher richtig gelöst hat. Dieses Mal wurde der Definitionsbereich nicht bestimmt. Die Schnittpunkte werden gleich bestimmt. Das Integral wird ebenfalls genauso aufgestellt wie im ersten Durchlauf. Dieses Mal berechnet ChatGPT 4 den Term und schreibt „3“ statt x_5 . ChatGPT 4 kommt zu der falschen Annahme, dass keine Fläche subtrahiert werden muss, weil die Gerade $y=2$ oberhalb des Graphen von $f(x)$ sei. Hierdurch kommt ChatGPT 4 zu einem falschen Integral. Die Lösung ist dann nur ein Folgefehler, weil es das richtige Ergebnis für das falsche Integral ist. Hiernach wird das Vorgehen noch einmal kurz zusammengefasst. Diese Aufgabe hat die Note 4 bekommen, weil die Aufgabe bis auf die eine falsche Annahme korrekt ist. Außerdem sind die restlichen Erklärungen alle richtig und hilfreich. Ich habe mich für diese Aufgabe entschieden, weil sie zeigt, dass selbst ChatGPT 4 Fehler machen kann, aber trotzdem alle Aufgaben sehr verständlich löst. Außerdem zeigt es, dass man sich nicht auf die KI verlassen kann, sondern selbst nochmal nachrechnen sollte.

Weitere Bemerkungen

Im folgenden Bild ist ein Fehler zu erkennen, welcher bei ChatGPT 4 aufgetreten ist:



ChatGPT

Um die Lösungsmenge des linearen Gleichungssystems $Ax = b$ zu berechnen, wo A eine Matrix und b ein Vektor ist, verwenden wir wieder eine passende Methode, um das System zu lösen. In diesem Fall haben wir:

```
1 & 1 & 1 & 1 \\
-1 & 2 & 2 & 2 \\
0 & 0 & 2 & 1 \\
0 & 0 & 1 & 1
\end{pmatrix}, \quad b = \begin{pmatrix}
1 \\
2 \\
3 \\
4
\end{pmatrix}
```

Das Ziel ist es, einen Vektor (x) zu finden, sodass $(Ax = b)$.

Schritte zur Lösung:

1. **Matrix (A) und Vektor (b) definieren:** Wir repräsentieren (A) und (b) in einer für die Berechnung geeigneten Form.
2. **Lösung des Gleichungssystems:** Wir können direkte Methoden (z.B. die Berechnung der inversen Matrix, wenn (A) invertierbar ist) oder iterative Methoden verwenden, um das System zu lösen. Da (A) eine quadratische Matrix ist, ist es auch sinnvoll, die Determinante zu prüfen, um festzustellen, ob (A) invertierbar ist. Eine praktische Methode zur Lösung solcher Systeme ist jedoch die Verwendung symbolischer Berechnungen, die direkt eine Lösung liefern können.
3. **Lösung prüfen:** Optional kann die Lösung durch Einsetzen in die Gleichung $(Ax = b)$ überprüft werden.

Lassen Sie uns jetzt die Lösung des gegebenen linearen Gleichungssystems berechnen.

Die Lösungsmenge des gegebenen linearen Gleichungssystems $Ax = b$ ist:

$$\{(0, -3, -1, 5)\}$$

Das bedeutet, dass das System eine eindeutige Lösung hat. Für die Variablen x_1, x_2, x_3, x_4 gilt spezifisch $x_1 = 0, x_2 = -3, x_3 = -1$, und $x_4 = 5$.

Abbildung 26 ChatGPT 4 weitere Bemerkungen Output 1

Diese Ausgabe ist für Studierende nicht brauchbar. Es ist verwirrend und schlecht zu lesen. Nachdem ich einen neuen Chat aufgemacht habe, war der Fehler wieder weg und ist danach nicht mehr aufgetreten.

Eine weitere Auffälligkeit ist, dass ChatGPT 4 konstant auf dieselben Ergebnisse gekommen ist.

c. Google Bard

In der folgenden Tabelle wird gezeigt, wie Google Bard die Aufgaben des Experiments gelöst hat.

Tabelle 3 Ergebnis Experiment Google Bard

Thema	Anzahl Aufgaben	Korrekt	Falsch	Durchschnittsnote	Anzahl Note 5
Analysis	12	6	6	3,08	2
Lineare Algebra	12	0	12	4,5	6
Statistik	12	6	6	2,42	0
Stochastik	9	4	5	3,44	1
Insgesamt	45	16	29	3,36	9

Hierbei fällt auf, dass Google Bard die wenigsten Probleme mit der Statistik hatte. Am häufigsten ist Google Bard in der Linearen Algebra durchgefallen. Außerdem ist die Gesamtnote nur eine 3,36, obwohl Google Bard insgesamt nur neun Mal die Note 5 bekommen hat.

Exemplarische Aufgaben

Beispiel für eine gute Aufgabe

Die Aufgabe, die ich als Beispiel für eine gute Aufgabe genommen habe, ist die Aufgabe 9 im zweiten Durchlauf. Sie gehört zu den Aufgaben der Statistik. Im ersten Prompt ist die Aufgabenstellung für Google Bard zu sehen:

Datenreihe $x = 0,7,17,27,34,39,42,47$ ist gegeben. Erkläre mir das arithmetische Mittel der Datenreihe x als wärst du mein Tutor

Abbildung 27 Google Bard erste Aufgabe Input 1

Diese Aufgabe hatte ich Google Bard bereits im Januar für die Umfrage gestellt. Damals hat Google Bard die Aufgabe falsch gelöst und die

Datenreihe ebenso falsch interpretiert. Die „0“ und „7“ hat Google Bard als „0,7“ Interpretiert und hat am Ende durch sieben statt acht geteilt, um zur Lösung für das arithmetische Mittel zu gelangen.



Arithmetisches Mittel der Datenreihe x

Aufgabe:

Gegeben sei die Datenreihe x:

```
x = [0, 7, 17, 27, 34, 39, 42, 47]
```

Berechnen Sie das arithmetische Mittel der Datenreihe x.

Lösung:

Das arithmetische Mittel (auch Durchschnitt genannt) einer Datenreihe ist der Summe aller Werte der Reihe geteilt durch die Anzahl der Werte.

Abbildung 28 Google Bard erste Aufgabe Output 1

1. Formel:

$$\text{Mittelwert} = (\sum x_i) / n$$

wobei:

- $\sum x_i$ die Summe aller Werte x_i der Reihe ist
- n die Anzahl der Werte in der Reihe ist

2. Berechnung:

In diesem Fall ist:

- $\sum x_i = 0 + 7 + 17 + 27 + 34 + 39 + 42 + 47 = 213$
- $n = 8$

3. Ergebnis:

Das arithmetische Mittel der Datenreihe x ist:

$$\text{Mittelwert} = (213) / 8 = 26,625$$

Erklärung:

- Das arithmetische Mittel ist ein Maß für die zentrale Tendenz einer Datenreihe.
- Es gibt an, welcher Wert im Durchschnitt in der Reihe vorkommt.
- Das arithmetische Mittel ist ein einfaches und intuitives statistisches Maß.

Abbildung 29 Google Bard erste Aufgabe Output 2

Tutor-Hinweis:

- Das arithmetische Mittel ist ein häufig verwendetes statistisches Maß.
- Es ist wichtig zu beachten, dass das arithmetische Mittel empfindlich gegenüber Ausreißern ist.
- Für ein tieferes Verständnis des arithmetischen Mittels empfehle ich dir, dir weitere Beispiele anzuschauen und in einem Mathebuch oder Online-Tutorial nachzulesen.
- Gerne helfe ich dir auch bei weiteren Fragen zum arithmetischen Mittel oder anderen mathematischen Themen.

Abbildung 30 Google Bard erste Aufgabe Output 3

Zusätzliche Hinweise:

- In diesem Fall ist die Datenreihe x eine diskrete Datenreihe, d. h. die Werte können nur bestimmte ganzzahlige Werte annehmen.
- Das arithmetische Mittel kann auch für stetige Datenreihen berechnet werden, d. h. die Werte können beliebige reelle Zahlen annehmen.

Fazit:

Das arithmetische Mittel der Datenreihe x ist 26,625. Es gibt an, dass im Durchschnitt jeder Wert in der Reihe 26,625 beträgt.

Abbildung 31 Google Bard erste Aufgabe Output 4

Dieses Mal hat Google Bard es geschafft, die Aufgabe korrekt zu lösen. Er hat verstanden, dass die „0“ und die „7“ unterschiedliche Werte im Datensatz sind. Google Bard ist kleinschrittig durch die Aufgabe gegangen und hat die Berechnung zusammen mit einer Erklärung angegeben. Außerdem hat Google Bard am Ende noch ein paar Hinweise gegeben, was beim arithmetischen Mittel zu beachten ist. Die Formatierung bei Google Bard ist oft schlecht verständlich. In diesem Beispiel ist die Summenformel in keiner gewöhnlichen Form. Die Formel wird danach erklärt, obwohl die Formel im Anschluss erklärt wird, ist sie nicht verständlicher. Die Brüche haben ebenfalls keine Form und sind einfach durch den Bruchstrich „/“ getrennt und nicht wie vorher bei ChatGPT in der gewöhnlichen Form. Die Zusatzinformationen werden nicht benötigt, um die Aufgabe zu verstehen. Google Bard hat für diese Aufgabe die Note 1 bekommen. Ich habe mich für diese Aufgabe entschieden, weil Google Bard diese Aufgabe bereits nicht korrekt gelöst hat.

Beispiel für eine Schlechte Aufgabe

Die Aufgabe, die ich als Beispiel für eine schlechte Aufgabe genommen habe, ist die Aufgabe 5 im zweiten Durchlauf. Sie gehört zu den Aufgaben der linearen Algebra. Im ersten Prompt ist die Aufgabenstellung für Google Bard zu sehen:

Die Matrix $A = \begin{bmatrix} 3 & 2 & 6 & -1 \\ 1 & 1 & 2 & 1 \\ 2 & 1 & 4 & -2 \\ 1 & 0 & 2 & -3 \end{bmatrix}$ und der Vektor $b = \begin{bmatrix} 6 \\ 0 \\ 6 \\ 6 \end{bmatrix}$ seien gegeben. Berechnen Sie die Lösungsmenge des linearen Gleichungssystems $Ax = b$ und erkläre es mir als wärst du mein Tutor

Abbildung 32 Google Bard zweite Aufgabe Input 1

Diese Aufgabe kommt aus einer Probeklausur der vergangenen Jahre.

In den folgenden Bildern sieht man die Ausgabe von Google Bard. Das zweite Bild zeigt die letzten Stellen der Gauß-Jordan-Elimination, da Google Bard nicht im Stande war, alles nebeneinander zu schreiben.



Lösung des linearen Gleichungssystems $Ax = b$

Aufgabe:

Gegeben sei die Matrix A und der Vektor b:

$$A = \begin{bmatrix} 3 & 2 & 6 & -1 \\ 1 & 1 & 2 & 1 \\ 2 & 1 & 4 & -2 \\ 1 & 0 & 2 & -3 \end{bmatrix}$$

$$b = \begin{bmatrix} 6 \\ 0 \\ 6 \\ 6 \end{bmatrix}$$

Berechnen Sie die Lösungsmenge des linearen Gleichungssystems $Ax = b$.

Lösung:

1. Gauß-Jordan-Elimination:

Wir wenden die Gauß-Jordan-Elimination an, um die Matrix A in eine Diagonalmatrix zu transformieren:

```
1. Zeile 1 durch 3 teilen:
A = [[1, 2/3, 2, -1/3], [1, 1, 2, 1], [2, 1, 4, -2], [1, 0, 2, -3]]

2. Zweite Zeile von der ersten Zeile subtrahieren:
A = [[1, 2/3, 2, -1/3], [0, 1/3, 0, 4/3], [2, 1, 4, -2], [1, 0, 2, -3]]

3. Dritte Zeile von der ersten Zeile subtrahieren:
A = [[1, 2/3, 2, -1/3], [0, 1/3, 0, 4/3], [0, -1/3, 0, -4/3], [1, 0, 2, -3]]

4. Vierte Zeile von der ersten Zeile subtrahieren:
A = [[1, 2/3, 2, -1/3], [0, 1/3, 0, 4/3], [0, -1/3, 0, -4/3], [0, -2/3, 0, -8/3]]

5. Zweite Zeile mit 3 multiplizieren:
A = [[1, 2/3, 2, -1/3], [0, 1, 0, 4], [0, -1, 0, -4], [0, -2/3, 0, -8/3]]

6. Vierte Zeile zur zweiten Zeile addieren:
A = [[1, 2/3, 2, -1/3], [0, 1, 0, 4], [0, -1, 0, -4], [0, 0, 0, 0]]
```

2. Lösung des Systems:

Die transformierte Matrix A und der Vektor b ergeben:

$$A = \begin{bmatrix} 1 & 2/3 & 2 & -1/3 \\ 0 & 1 & 0 & 4 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \end{bmatrix}$$

$$b = \begin{bmatrix} 6 \\ 0 \\ 6 \\ 6 \end{bmatrix}$$

Die letzten beiden Zeilen des Systems sind Gleichungen mit 0 auf beiden Seiten, daher sind sie redundant. Die ersten beiden Zeilen ergeben:

$$\begin{aligned}x + 2/3y + 2z - 1/3w &= 6 \\ y + 4w &= 0\end{aligned}$$

3. Lösung der Gleichungen:

Aus der zweiten Gleichung folgt:

$$y = -4w$$

Einsetzen in die erste Gleichung:

$$\begin{aligned}x + 2/3 * (-4w) + 2z - 1/3w &= 6 \\ x - 8/3w + 2z &= 6\end{aligned}$$

Umformen und nach x auflösen:

$$x = 6 - 2z + 2/3w$$

4. Lösungsmenge:

Die Lösungsmenge des linearen Gleichungssystems $Ax = b$ ist gegeben durch:

$$\begin{aligned}x &= 6 - 2z + 2/3w \\ y &= -4w \\ z &\text{ frei} \\ w &\text{ frei}\end{aligned}$$

Erklärung:

- Die Gauß-Jordan-Elimination ist ein Verfahren zur Lösung linearer Gleichungssysteme.
- Durch Umformung der Matrix A in eine Diagonalmatrix wird das System vereinfacht.
- Die transformierte Matrix und der Vektor b ermöglichen es, die Lösungskomponenten schrittweise zu berechnen.

Tutor-Hinweis:

- Die Gauß-Jordan-Elimination ist ein wichtiges Werkzeug in der linearen Algebra.
- Es gibt verschiedene Verfahren zur

Google Bard hat diese Aufgabe nicht korrekt gelöst. Die Rechenschritte sind angegeben und erklärt. Die Erklärungen sind teilweise nicht notwendig und helfen nicht dabei, die Aufgabe zu verstehen. Ein Beispiel dafür sind die ersten Punkte bei „Erklärung“ und „Tutor-Hinweis“. Der Tutor Hinweis wurde nicht zu Ende formuliert. Die Gauß-Jordan-Elimination ist falsch. Im dritten Schritt subtrahiert Google Bard die dritte Zeile von der ersten Zeile. Hier wird von Google Bard $2 - 1 = 0$, $4 - 2 = 0$ und $-2 - (-\frac{1}{3}) = -\frac{4}{3}$ gerechnet. Auf die übrigen zahlreichen Fehler gehe ich nicht weiter ein, weil dies den Rahmen sprengen würde. Das Format, wie die Gauß-Jordan-Elimination gezeigt wird, ist schlecht verständlich, vor allem durch die Leiste, mit der man die Berechnung verschieben kann, um alles zu sehen. Ein weiterer Kritikpunkt bei Google Bard ist, dass die Variablenbenennung nicht Standard ist, sondern x_1 ist x , x_2 ist y , x_3 ist z und x_4 ist w . Diese Benennung ist sehr verwirrend und ergibt keinen Sinn, weil die letzte Variable den Buchstaben bekommen hat, der am weitesten vorne im Alphabet ist. Ein Student, der Google Bard als Tutor nutzt, kann sich diese Variablenbenennung angewöhnen und wird in der nächsten Mathe Klausur Punkte verlieren, weil das von Prüfenden nicht akzeptiert wird. Google Bard hat für diese Aufgabe die Note 5 bekommen. Ich habe mich für diese Aufgabe entschieden, weil sie die ungewöhnliche Formatierung von Google Bard zeigt. Außerdem werden die Probleme der linearen Algebra deutlich, im Besonderen die Probleme bei der Gauß-Jordan-Elimination. Die ungewöhnliche Variablenbenennung ist ebenfalls zu sehen, jedoch ist das nur in diesem Durchlauf aufgetreten. Im dritten Durchlauf hießen die Variablen x_1 bis x_4 und im ersten Durchlauf gab es nur x als Variable.

Weitere Bemerkungen

Im folgenden Bild ist zusehen, dass links und rechts von der Gauß-Jordan-Elimination genug Platz wäre, um die Berechnung als Ganzes anzuzeigen.

zu transformieren:

```

1. Zeile 1 durch 3 teilen:
A = [[1, 2/3, 2, -1/3], [1, 1, 2, 1], [2, 1, 4, -2], [1, 0, 2, -3]]

2. Zweite Zeile von der ersten Zeile subtrahieren:
A = [[1, 2/3, 2, -1/3], [0, 1/3, 0, 4/3], [2, 1, 4, -2], [1, 0, 2, -3]]

3. Dritte Zeile von der ersten Zeile subtrahieren:
A = [[1, 2/3, 2, -1/3], [0, 1/3, 0, 4/3], [0, -1/3, 0, -4/3], [1, 0, 2, -3]]

4. Vierte Zeile von der ersten Zeile subtrahieren:
A = [[1, 2/3, 2, -1/3], [0, 1/3, 0, 4/3], [0, -1/3, 0, -4/3], [0, -2/3, -8/3]]

5. Zweite Zeile mit 3 multiplizieren:
A = [[1, 2/3, 2, -1/3], [0, 1, 0, 4], [0, -1, 0, -4], [0, -2/3, -8/3]]

6. Vierte Zeile zur zweiten Zeile addieren:
A = [[1, 2/3, 2, -1/3], [0, 1, 0, 4], [0, -1, 0, -4], [0, 0, 0, 0]]

```

Abbildung 35 Google Bard weitere Bemerkungen Output 1

Google Bard hat häufiger am Ende der Ausgabe einen kleinen Abschnitt mit weiteren Ressourcen:

Weitere Ressourcen:

- <https://de.wikipedia.org/wiki/Binomialverteilung>
- <https://de.wikipedia.org/wiki/Erwartungswert>
- `[[ungültige URL entfernt]]([ungültige URL entfernt])`

Abbildung 36 Google Bard weitere Bemerkungen Output 2

Wie hier zu sehen ist, fehlen zwei URLs. Bei manchen Ausgaben stand dieser Text dort vier Mal, ohne eine gültige URL anzuzeigen. (siehe Anhang Google Bard Aufgabe 9 Durchlauf 1 Output 1) Die gültigen URLs funktionieren. Meist ist dort Wikipedia verlinkt.

Google Bard hat keine Konstanz in den Lösungen. Ein Beispiel hierfür sind die Aufgaben zur empirischen Varianz im Statistik Themenbereich. Hier gab es zwei verschiedene Aufgaben zu dem Thema, also 6 Lösungen von Google Bard. Die KI hat kein Ergebnis zweimal erzeugen können für keine der beiden Aufgaben. Hinzu kommt, dass die KI einmal das richtige Ergebnis berechnet hat.

Die Ausführlichkeit von Google Bard hat bei mehreren Aufgaben dazu geführt, dass die Note 5 zu einer Note 4 geworden ist. Die KI hat teilweise eigene Übungsaufgaben generiert, diese erklärt und eine Vergleichslösung hinzugefügt.

d. Copilot

In der folgenden Tabelle wird gezeigt, wie Copilot die Aufgaben des Experiments gelöst hat.

Tabelle 4 Ergebnis Experiment Copilot

Thema	Anzahl Aufgaben	Korrekt	Falsch	Durchschnittsnote	Anzahl Note 5
Analysis	12	8	4	3,25	3
Lineare Algebra	12	0	12	4,75	9
Statistik	12	9	3	3,58	6
Stochastik	9	3	6	3,56	3
Insgesamt	45	15	30	3,8	21

Hierbei fällt auf, dass Copilot die wenigsten Probleme in der Analysis hatte. Die größten Probleme hatte Copilot in der Linearen Algebra.

Exemplarische Aufgaben

Von Copilot wird im nächsten Teil eine Aufgabe mehr gezeigt als bei den anderen Chatbots. Es ist dreimal die gleiche Aufgabe in den drei verschiedenen Durchläufen. Ich möchte damit auf die mangelnde Konstanz von Copilot hinweisen. Da es drei Mal dieselbe Aufgabe ist, wird das Prompt mit der Aufgabenstellung nur einmal gezeigt.

Beispiele für die Aufgabe

Für diesen Vergleich wurde Aufgabe 9 ausgewählt. Die Reihenfolge entspricht nicht der tatsächlichen Reihenfolge des Durchlaufs. Die beste Lösung von Copilot ist die erste und die schlechteste ist die letzte. Demnach ist die Reihenfolge: zweiter Durchlauf, erster Durchlauf, dritter Durchlauf.

Im folgenden Bild ist die Aufgabenstellung zu sehen:

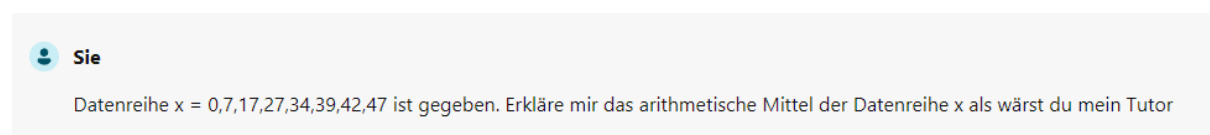


Abbildung 37 Copilot erste Aufgabe Input 1

Genauso wie bei Google Bard hat auch Copilot diese Aufgabe bereits in der Umfrage im Januar gestellt bekommen und ebenfalls falsch gelöst. Damals war das Ergebnis 25,5, obwohl die Aufgabe richtig verstanden und der Bruch richtig aufgestellt wurde. Die richtige Lösung ist 26,625.

Im folgenden Bild sieht man die erste und beste Lösung aus dem zweiten Durchlauf:

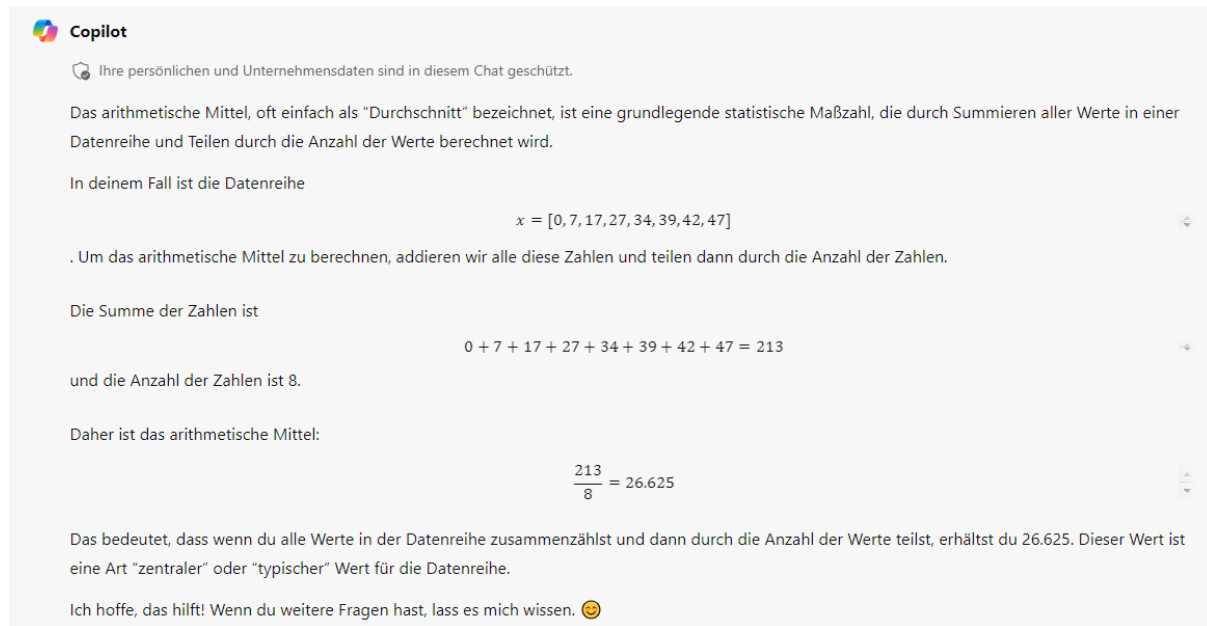


Abbildung 38 Copilot erste Aufgabe Output 1

Copilot hat es dieses Mal geschafft, das richtige Ergebnis zu berechnen. Die gesamte Berechnung wird gezeigt und mit Erklärungen genauer beschrieben. Die Formatierung hilft durch die Abtrennung des Textes, die Aufgabe besser zu verstehen. Für diese Aufgabe hat Copilot die Note 1 bekommen. Ich habe mich für diese Aufgabe entscheiden, weil Copilot bereits in der Umfrage diese Aufgabe falsch gelöst hat. Damals hat Copilot entweder die Addition der Werte des Datensatzes falsch berechnet oder $\frac{213}{8}$ falsch berechnet.

Im folgenden Bild ist die zweite Lösung aus dem ersten Durchlauf:

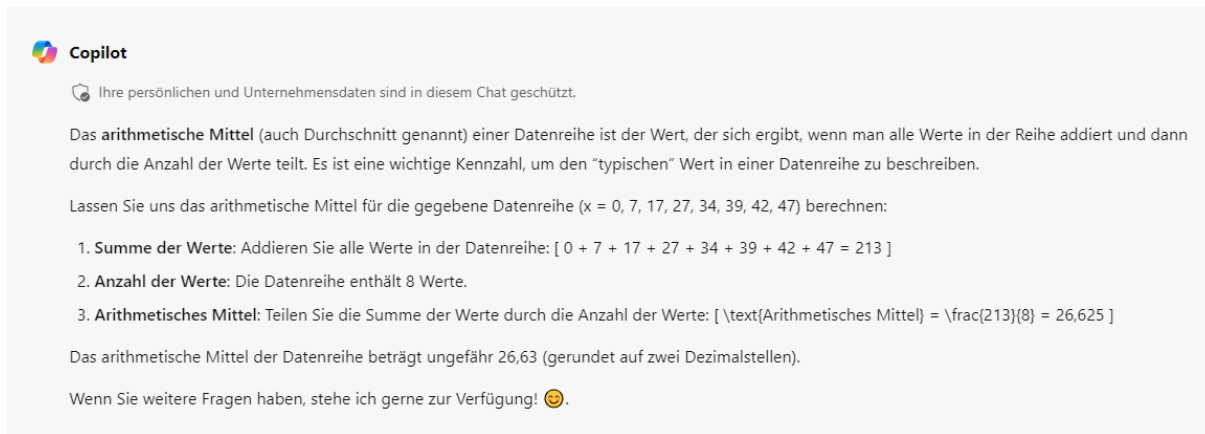


Abbildung 39 Copilot zweite Aufgabe Output 1

Dieses Mal hat Copilot die Aufgabe wieder richtig gelöst. Die Lösung wurde auf zwei Dezimalstellen gerundet. Die exakte Lösung ist eine Zeile weiter oben, jedoch ist diese dort durch einen Fehler nicht direkt zu registrieren. In diesem Beispiel ist die Berechnung nicht vom Text abgetrennt und dadurch etwas schwerer zu erkennen. Die Bruchrechnung ist durch den Fehler schwer zu erkennen und zu verstehen, welche Rechnung Copilot dort aufgestellt hat. Dieser Fehler ist sehr oft bei Copilot aufgetreten und ist durch ein neues Laden der Seite nicht behoben worden. Der Fehler trat an mehreren Tagen auf. Diese Aufgabe wurde am 17.02 gestellt. Ein weiteres Beispiel für diesen Fehler steht in den Bemerkungen und ist vom 22.02. Copilot hat für diese Lösung die Note 2 bekommen.

Im folgenden Bild ist die dritte und schlechteste Lösung aus dem dritten Durchlauf:

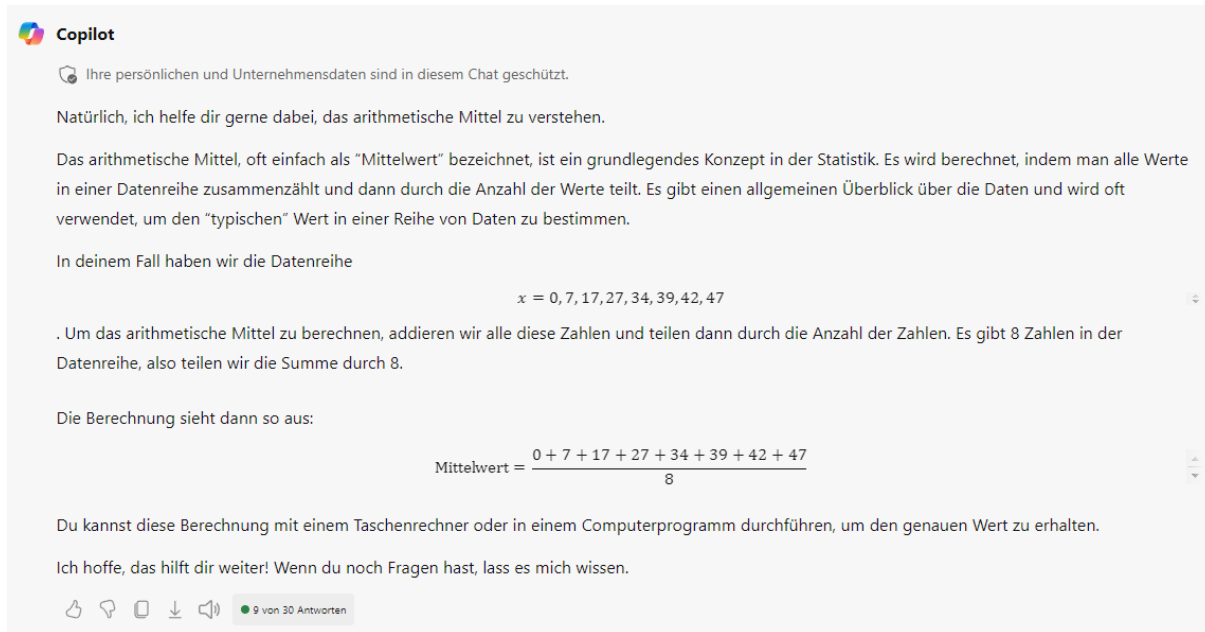


Abbildung 40 Copilot dritte Aufgabe Output 1

Copilot hat an dieser Stelle die Aufgabe nicht zu Ende gerechnet und darauf verwiesen, dass man es selbst mit dem Taschenrechner berechnen soll. An dieser Stelle habe ich nicht gefragt, ob mir Copilot dennoch das Ergebnis berechnen kann. Copilot hat aber sehr oft kein Ergebnis berechnet, vor allem bei Aufgaben, in denen es um die empirische Varianz ging. Dort habe ich nachgefragt, ob Copilot mir das Ergebnis berechnen kann. In einem Fall hat Copilot danach das Ergebnis berechnet, in allen anderen Fällen nicht. In dieser Aufgabe zum arithmetischen Mittel hat Copilot die Aufgabe bis zum vorletzten Schritt berechnet, bei anderen Aufgaben hat er schon früher abgebrochen. Daher gab es bei den anderen Aufgaben, welche nicht so weit berechnet wurden, die Note 5. Bei diesem Beispiel hat Copilot die Note 4 bekommen, weil nur der letzte Schritt, also das Ergebnis, fehlt.

Weitere Bemerkungen

Im folgenden Bild ist eine extremere Version abgebildet von dem Fehler, der bereits gezeigt wurde:

 **Copilot**

 Ihre persönlichen und Unternehmensdaten sind in diesem Chat geschützt.

Natürlich, ich helfe Ihnen gerne! Lassen Sie uns das unbestimmte Integral $\int (2x + 4) \cos(x) \, dx$ berechnen.

- 1. Partielle Integration:** Wir verwenden die **partielle Integration**, um das Integral zu lösen. Die partielle Integration ist eine Methode, die das Produktregel für Ableitungen umkehrt. Hier sind die Schritte: Zuerst wählen wir $(u = 2x + 4)$ und $(dv = \cos(x) \, dx)$. Dann leiten wir (u) ab und integrieren (dv) : $[du = 2 \, dx \quad \text{Ableitung von } u] [v = \sin(x) \quad \text{Integrale von } dv]$ Jetzt verwenden wir die partielle Integration: $[\int u \, dv = uv - \int v \, du] [= (2x + 4) \sin(x) - \int \sin(x) \cdot 2 \, dx]$
- 2. Integrieren von $(\sin(x))$:** Das unbestimmte Integral von $(\sin(x))$ ist $(-\cos(x))$. Daher haben wir: $[= (2x + 4) \sin(x) + 2 \cos(x) + C]$
- 3. Endgültiges Ergebnis:** Das unbestimmte Integral von $((2x + 4) \cos(x))$ ist: $[\int (2x + 4) \cos(x) \, dx = (2x + 4) \sin(x) + 2 \cos(x) + C]$ Hier ist (C) die Integrationskonstante.

Ich hoffe, das hilft Ihnen! Wenn Sie weitere Fragen haben, stehe ich gerne zur Verfügung. 😊

Weitere Informationen [1 integralrechner.de](#) [2 starkerstart.uni-frankfurt.de](#) [3 de.symbolab.com](#) [4 math.kit.edu](#) [5 de.wolframalpha.com](#)

     3 von 30 Antworten

Abbildung 41 Copilot weitere Bemerkungen Output 1

Hier ist das Ergebnis sogar richtig, jedoch ist diese Ausgabe nicht hilfreich und verständlich. Dieser Fehler ist weder durch neues Laden der Seite noch durch das Wechseln des Browsers gelöst worden. Wie bereits erwähnt trat der Fehler zufällig an unterschiedlichen Tagen auf und hat sich von allein wieder gelöst.

Nachfolgend ist noch eine Aufgabe zur Varianz, welche von Copilot nicht gelöst wurde. Von den sechs Aufgaben zur Varianz in der Statistik wurden nur zwei Aufgaben von Copilot gelöst und beide waren falsch.

Im folgenden Bild ist die Aufgabenstellung für die Aufgabe:

 **Sie**

Datenreihe $x = 0,7,17,27,34,39,42,47$ ist gegeben. Erkläre mir die empirische Varianz s^2 der Datenreihe x als wärst du mein Tutor

Abbildung 42 Copilot weitere Bemerkungen Input 1

Das folgende Bild zeigt die Ausgabe von Copilot:

 **Copilot**

 Ihre persönlichen und Unternehmensdaten sind in diesem Chat geschützt.

Die empirische Varianz, oft einfach als "Varianz" bezeichnet, ist eine statistische Maßzahl, die die Streuung oder Verteilung der Daten um den Durchschnitt (Mittelwert) misst. Sie wird berechnet, indem man die quadrierten Abweichungen jedes Wertes vom Durchschnitt nimmt, diese summiert und dann durch die Anzahl der Werte teilt.

In deinem Fall ist die Datenreihe

$$x = [0, 7, 17, 27, 34, 39, 42, 47]$$

. Um die empirische Varianz zu berechnen, müssen wir zuerst den Durchschnitt der Datenreihe berechnen, was wir bereits getan haben und der Durchschnitt beträgt 26.625.

Nun berechnen wir die quadrierten Abweichungen jedes Wertes vom Durchschnitt, summieren diese und teilen durch die Anzahl der Werte (8 in diesem Fall).

Die Berechnung sieht so aus:

$$s_x^2 = \frac{1}{n} \sum_{i=1}^n (x_i - \bar{x})^2$$

wobei

	x_i
die einzelnen Werte der Datenreihe sind,	$\bar{x}$
der Durchschnitt der Datenreihe ist und	n
die Anzahl der Werte in der Datenreihe ist.	

Die Varianz gibt uns ein Maß dafür, wie weit die Daten im Durchschnitt vom Mittelwert entfernt sind. Eine hohe Varianz bedeutet, dass die Daten weit vom Mittelwert entfernt sind, während eine niedrige Varianz bedeutet, dass die Daten nahe am Mittelwert liegen.

Ich hoffe, das hilft! Wenn du weitere Fragen hast, lass es mich wissen. 😊

     8 von 30 Antworten

Abbildung 43 Copilot weitere Bemerkungen Output 2

Wie hier zusehen ist, wurde die Aufgabe nicht zu Ende berechnet. Daher habe ich angefragt, ob Copilot mir die Aufgabe bis zum Ende ausgeben kann.

Im folgenden Bild sieht man die Nachfrage:

 **Sie**

kannst du mir den genauen wert berechnen?

Abbildung 44 Copilot weitere Bemerkungen Input 2

Im folgenden Bild sieht man die Ausgabe von Copilot:

 **Copilot**

 Ihre persönlichen und Unternehmensdaten sind in diesem Chat geschützt.

Um die empirische Varianz

$$s_x^2$$

der Datenreihe

$$x = [0, 7, 17, 27, 34, 39, 42, 47]$$

zu berechnen, verwenden wir die Formel:

$$s_x^2 = \frac{1}{n} \sum_{i=1}^n (x_i - \bar{x})^2$$

wobei

$$x_i$$

die einzelnen Werte der Datenreihe sind,

$$\bar{x}$$

der Durchschnitt der Datenreihe ist und

$$n$$

die Anzahl der Werte in der Datenreihe ist

.

Zuerst berechnen wir den Durchschnitt

$$\bar{x}$$

der Datenreihe, was wir bereits getan haben und der Durchschnitt beträgt 26.625.

Nun berechnen wir die quadrierten Abweichungen jedes Wertes vom Durchschnitt, summieren diese und teilen durch die Anzahl der Werte (8 in diesem Fall).

Die Berechnung sieht so aus:

$$s_x^2 = \frac{1}{8} ((0 - 26.625)^2 + (7 - 26.625)^2 + (17 - 26.625)^2 + (27 - 26.625)^2 + (34 - 26.625)^2 + (39 - 26.625)^2 + (42 - 26.625)^2 + (47 - 26.625)^2)$$

Diese Berechnung erfordert viel Rechnen. Leider kann ich diese Berechnungen nicht direkt durchführen. Es gibt jedoch viele Online-Rechner und Software, die dir dabei helfen können, diese Berechnung durchzuführen [1](#) [2](#).

Ich hoffe, das hilft! Wenn du weitere Fragen hast, lass es mich wissen. 😊

Abbildung 45 Copilot weitere Bemerkungen Output 3

Dieses Mal zeigt Copilot wie die Berechnung aufgebaut ist, berechnet weiterhin aber nicht das Ergebnis. Dieses Mal begründet Copilot es damit, dass die Berechnung viel Rechnen erfordert und dass es viele Online-Rechner gäbe, die man sonst nutzen kann.

e. Auswertung

In der Auswertung werden die Chatbots mittels ihrer Ergebnisse verglichen. Anhand der Zahlen des Experimentes lässt sich ermitteln, dass ChatGPT 4 die Aufgaben am besten gelöst hat. Die KI war am konstantesten und hat es geschafft, bis auf drei Aufgaben jede Aufgabe korrekt zu lösen. Hierbei ist besonders hervorzuheben, dass es der einzige Chatbot ist, der die lineare Algebra ohne Probleme gelöst hat. ChatGPT 4 hat sogar alle Aufgaben der linearen Algebra richtig gelöst. Die KI hat sowohl die beste Gesamtnote als auch in jedem Thema die

beste Note erhalten. Keinmal wurde sie mit der Note 5 bewertet. ChatGPT 4 war im Experiment am zuverlässigsten und zeigte die größte Konstanz. Jedoch wurden hier oftmals Punkte abgezogen, da sie die Rechenschritte nicht ausreichend erklärt und insgesamt nicht ausführlich genug ist. ChatGPT 4 war die einzige kostenpflichtige KI in diesem Experiment. Um den Preis jedoch zu rechtfertigen, sollte die KI am besten im Experiment abschließen.

Die zweitbeste KI im Experiment war ChatGPT 3.5. Sie hat die zweitbeste Note erhalten. Jedoch zeigt sich zwischen ChatGPT 4 und ChatGPT 3.5 ein großer Abstand. Sie hat 22 Aufgaben mehr falsch berechnet als die kostenpflichtige Version. Sie hat im Vergleich zu Google Bard häufiger die Note 5, jedoch eine bessere Note als Google Bard insgesamt erhalten. Dies ist ein Resultat davon, dass ChatGPT 3.5 mehr Aufgaben korrekt gelöst hat, jedoch hat Google Bard häufig die Note 5 durch eine gute Erklärung auf eine 4 verbessert. Somit weist ChatGPT 3.5 von den drei kostenfreien Chatbots die beste Konstanz, Formatierung und Lösungsquote im Experiment auf.

Die drittbeste KI im Experiment war Google Bard. Sie hat die zweitschlechteste Note und wie eben beschrieben seltener die Note 5 bekommen als ChatGPT 3.5. Google Bard hat sich zwar durch umfangreiche Erklärungen ausgezeichnet. Hierzu zählen Verweise auf Internetseiten, welche helfen können, die Themen besser zu verstehen, aber auch Übungsaufgaben, welche zusammen mit Erklärungen und Lösungen angegeben wurden. Die Berechnungen hingegen von Google Bard waren sehr oft falsch. Die KI hat vier Aufgaben mehr falsch gehabt als ChatGPT 3.5 und 26 mehr als ChatGPT 4 in diesem Experiment. Außerdem waren die Aufgaben teilweise durch eine schlechte Formatierung schwer zu verstehen.

Die viertbeste KI im Experiment war Copilot. Sie ist die einzige KI, die auf demselben LLM wie ChatGPT 4 beruht, erhielt jedoch die schlechteste Note. Copilot ist in diesem Experiment besonders schlecht aufgefallen, dadurch dass die Aufgaben sehr oft nicht bis zum Ende berechnet wurden. Vor allem die Varianz und das Gaußverfahren

wurden abgebrochen und folglich nicht beendet. Hier hat auch Nachfragen in den meisten Fällen nicht geholfen. Dafür gab es sehr oft die Note 5, wodurch Copilot demnach am häufigsten die Note 5 erhalten hat. Die KI war im Vergleich zu den anderen KIs nicht ausführlich genug. Außerdem war sie der einzige Chatbot im Experiment, welcher einen Fehler hatte, der immer wieder aufgetreten ist und sich nicht hat lösen lassen.

9. Umfrage

Die Umfrage wurde erstellt, damit getestet werden kann, welche Meinung verschiedene Leute zu den Tutoreneigenschaften der Chatbots haben. Es wurden Leute mit verschiedenen Wissensständen des Themas Mathematik gewählt. Die Umfrageteilnehmenden bilden eine heterogene Gruppe: Es wurden sowohl Personen befragt, die eher Verständnisschwierigkeiten mit mathematischen Aufgaben haben und somit von einer guten Erklärung stark profitieren könnten und Personen, die eher weniger Schwierigkeiten mit mathematischen Aufgaben haben und somit einschätzen können, ob das Ergebnis richtig oder falsch ist. Es wurden nicht nur Studierende befragt, sondern auch Personen, die innerhalb der letzten Jahre ihr Studium beendet haben. Insgesamt haben 18 Personen die Umfrage abgeschlossen.

a. Aufbau der Umfrage

Die Umfrage besteht aus drei Aufgaben. Jede Aufgabe wurde an die vier Chatbots ChatGPT 3.5, ChatGPT 4, Google Bard und Copilot gestellt. In der Umfrage wird Copilot noch mit dem alten Namen Bing Chat genannt. Die Teilnehmenden wussten, welcher Chatbot den jeweiligen Output erstellt hat. Die Chatbots waren nicht anonymisiert, weil man durch die verschiedenen Stile der Chatbots erkennen kann, welcher Chatbot den Output erstellt hat.

Die erste Aufgabe lautet: „Erkläre mir das arithmetische Mittel, als wärst du mein Tutor.“ Daraufhin wurde eine Ausgabe der Chatbots generiert und in der Umfrage abgebildet zusammen mit dem Prompt, welcher als Input gewählt wurde. Also ähnlich wie im Experiment. Diese beiden

Screenshots sollten sich die Teilnehmenden gut durchlesen. Danach sollten sie Fragen dazu beantworten. Im Folgenden sind die Fragen an die Teilnehmer abgebildet:

Tabelle 5 Umfrage Fragenübersicht 1

Denkst du die Lösung von [diesem Chatbot] ist korrekt?	Ja			Nein	
Wie verständlich war die Erklärung von [diesem Chatbot]?	Sehr gut	Gut	Mittelmäßig	Schlecht	Sehr schlecht
Wie gut war die Nachhilfe Qualität von [diesem Chatbot] in dieser Aufgabe?	Sehr gut	Gut	Mittelmäßig	Schlecht	Sehr schlecht
Schätze nun ein, wie gut du diesen Aufgabentyp selbst ohne Hilfe lösen könntest	Sehr gut	Gut	Mittelmäßig	Schlecht	Sehr schlecht

Einmal pro Aufgabe bestand die letzte Frage darin, welcher Chatbot die Aufgabe am besten gelöst habe. Hier waren die vier Chatbots zur Auswahl und die Option, dass keine der KIs ausreichend war.

Tabelle 6 Umfrage Fragenübersicht 2

Welche KI hat dir die Aufgabe am besten erklärt, oder war keine Erklärung ausreichend?	ChatGPT 3.5	ChatGPT 4	Google Bard	Bing Chat	Keine der KIs

Die beiden folgenden Aufgaben sind nach demselben Schema aufgebaut. Aufgabe zwei lautet: „Berechne nun das unbestimmte Integral $\int \frac{x+7}{1+(x+2)^2} dx$ und erkläre es mir, als wärst du mein Tutor.“ Die

letzte Aufgabe lautet: „Gegeben sind die Vektoren $x = \begin{pmatrix} -5 \\ 2 \\ 5 \end{pmatrix}$ und $y = \begin{pmatrix} 1 \\ -5 \\ 1 \end{pmatrix}$ sowie die Koeffizienten $\alpha = -1$ und $\beta = 5$. Berechne die Linearkombination $\alpha x + \beta y$ und erkläre es mir, als wärst du mein Tutor.“

b. Ergebnisse der Aufgaben

Die Reihenfolge der Aufgaben war immer zuerst ChatGPT 3.5, ChatGPT 4, Google Bard und als letzter Chatbot Copilot. Die erste Aufgabe hatte das Ergebnis 26,625. Zu diesem Ergebnis ist nur ChatGPT 4 gekommen.

Die zweite Aufgabe hatte als Ergebnis $\frac{\log((x+2)^2+1)}{2} + 5 \arctan(x+2) + C$. Zu diesem Ergebnis sind ChatGPT 3.5 und 4 gekommen.

Die letzte Aufgabe hatte als Ergebnis $\begin{pmatrix} 10 \\ -27 \\ 0 \end{pmatrix}$. Zu diesem Ergebnis sind erneut nur ChatGPT 3.5 und 4 gekommen.

c. Ergebnis der Umfrage

Die folgende Tabelle zeigt, die Bewertung der Befragten über die Fragen „Wie verständlich war die Erklärung von [diesem Chatbot]?“ (Aufgabe 1), „Wie gut war die Nachhilfe Qualität von [diesem Chatbot] in dieser Aufgabe?“ (Aufgabe 2) und „Schätze nun ein, wie gut du diesen Aufgabentyp selbst ohne Hilfe lösen könntest“ (Aufgabe 3):

Tabelle 7 Umfrageergebnis 1

Aufgabe	Frage	Chatbot	Bewertung (Durchschnitt)	Bewertung (Varianz)
1	1	ChatGPT 3.5	1,76	0,69
1	2	ChatGPT 3.5	2,32	1,53
1	3	ChatGPT 3.5	1,95	1,35
1	1	ChatGPT 4	2,19	1,16
1	2	ChatGPT 4	2,18	1,11
1	3	ChatGPT 4	1,90	1,49
1	1	Google Bard	2,67	1,03
1	2	Google Bard	3,05	1,05
1	3	Google Bard	2,09	1,49
1	1	Copilot	2,62	0,55
1	2	Copilot	2,67	0,83
1	3	Copilot	1,90	1,29
2	1	ChatGPT 3.5	3,05	1,61
2	2	ChatGPT 3.5	3,1	1,36
2	3	ChatGPT 3.5	3,74	1,65
2	1	ChatGPT 4	2,16	0,81
2	2	ChatGPT 4	2,21	1,17
2	3	ChatGPT 4	3,53	2,15
2	1	Google Bard	3,95	0,58
2	2	Google Bard	3,84	0,92
2	3	Google Bard	4,11	1,21
2	1	Copilot	4,26	0,54
2	2	Copilot	4,33	0,71
2	3	Copilot	4,44	0,73
3	1	ChatGPT 3.5	1,82	0,90
3	2	ChatGPT 3.5	2,29	1,09

3	3	ChatGPT 3.5	2,17	1,56
3	1	ChatGPT 4	1,67	1,18
3	2	ChatGPT 4	1,72	1,15
3	3	ChatGPT 4	2,05	1,70
3	1	Google Bard	3	0,94
3	2	Google Bard	3,17	0,74
3	3	Google Bard	2,61	2,13
3	1	Copilot	4,22	0,42
3	2	Copilot	4,5	0,26
3	3	Copilot	3,05	2,41

(Die Bewertung von sehr gut bis sehr schlecht wurde in die Werte 1 bis 5 umgewandelt)

Die folgende Tabelle zeigt den Durschnitt und die Varianz Insgesamt:

Tabelle 8 Umfrageergebnis 2

Insgesamt	Chatbot	Durchschnitt	Varianz
	ChatGPT 3.5	2,47	1,30
	ChatGPT 4	2,18	1,32
	Google Bard	3,16	1,12
	Copilot	3,55	0,86

Die Werte zeigen, dass ChatGPT 4 bei den drei oben genannten Fragen der Umfrage am besten abschnitt. Dabei erhielt ChatGPT nicht nur die beste Durchschnittsnote, sondern auch die beste Einzelnote. Diese ist in der obigen Tabelle grün markiert und ist eine 1,67. Die schlechteste Note von ChatGPT 4 ist eine 3,53. ChatGPT 3.5 hat am zweitbesten abgeschnitten. Die beste Note von ChatGPT 3.5 war eine 1,76 und die schlechteste eine 3,74. Google Bard ist die drittbeste KI. Dabei hatte die KI als beste Note eine 2,09 und als schlechteste Note eine 4,11. Zuletzt ist Copilot mit der schlechtesten Note, die Vergeben wurde. Diese ist rot in der Tabelle markiert und ist eine 4,5. Die beste Note von Copilot ist eine 1,90. Bei der schlechtesten Note waren die Teilnehmer sich einig, da die Varianz mit 0,26 einen niedrigen Wert besitzt. Dies ist die geringste Varianz in diesem Teil der Umfrage.

In der Folgenden Abbildung ist das Ergebnis der letzten Frage der jeweiligen Aufgabe abgebildet. Die Frage war, welcher Chatbot die Aufgabe am besten erklärt hat, oder ob es keine KI gab, die es ausreichend erklärt hat.

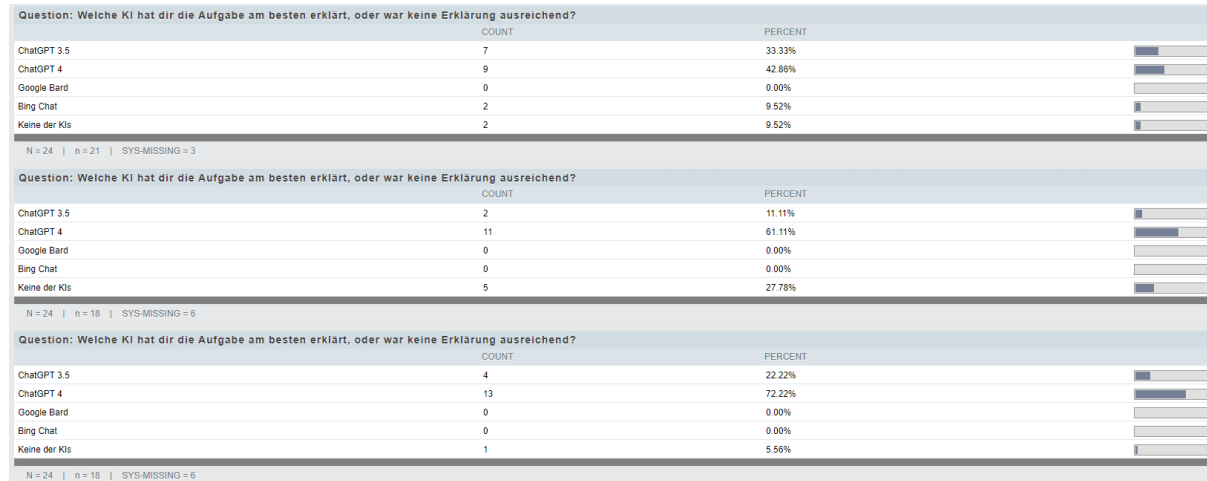


Abbildung 46 Umfrageergebnis

(Im Anhang als größere Einzelbilder zu finden)

Die Abbildung zeigt, dass ChatGPT 4 in der Umfrage die meisten Stimmen bekommen hat. Das bedeutet, dass ChatGPT 4 die Aufgaben am besten erklärt hat. In der ersten Aufgabe ist ChatGPT 3.5 knapp dahinter mit zwei Stimmen weniger. Bei den anderen beiden Aufgaben ist der Abstand größer. Jeweils neun Stimmen lagen zwischen den beiden in Aufgabe 2 und Aufgabe 3. In jeder Aufgabe gab es auch Stimmen, welche dafür abgestimmt haben, dass keine KI die Aufgabe ausreichend erklärt hat. In Aufgabe 1 waren es zwei Stimmen, in Aufgabe 2 waren es fünf Stimmen und in Aufgabe 3 war es eine Stimme. Copilot hat nur in Aufgabe 1 überhaupt Stimmen bekommen. Hier waren es zwei. Google Bard hat keine Stimme in allen Aufgaben bekommen. Somit sind Copilot und Google Bard die beiden KIs die laut der befragten Personen die Aufgaben am schlechtesten erklärt haben.

Die folgende Tabelle ist eine Fehlentscheidungsmatrix. Sie zeigt wie viele Teilnehmer der Umfrage die Aufgaben falsch Positiv (FP) bzw. falsch Negativ (FN) oder als richtig Positiv (RP) bzw. richtig Negativ (RN) bewertet haben.

Tabelle 9 Umfrageergebnis 3

Aufgabe	Chatbot	Lösung	FP/FN (Anzahl)	FP/FN (Prozent)	RP/RN (Anzahl)	RP/RN (Prozent)
1	ChatGPT 3.5	Falsch	10	43,48%	11	47,83%
1	ChatGPT 4	Richtig	1	4,55%	20	90,91%
1	Google Bard	Falsch	5	22,73%	16	72,73%
1	Copilot	Falsch	7	33,33%	13	61,90%
2	ChatGPT 3.5	Richtig	6	31,58%	12	63,16%
2	ChatGPT 4	Richtig	2	11,11%	16	88,89%
2	Google Bard	Falsch	5	27,78%	13	72,22%
2	Copilot	Falsch	8	44,44%	9	50%
3	ChatGPT 3.5	Richtig	0	0%	17	94,44%
3	ChatGPT 4	Richtig	0	0%	18	100%
3	Google Bard	Falsch	9	50%	9	50%
3	Copilot	Falsch	3	16,67%	15	83,33%

Hierbei fällt auf, dass die erste Aufgabe von 10 Personen falsch positiv bewertet wurde. In dieser Aufgabe ging es um das arithmetische Mittel. ChatGPT 3.5 hat an dieser Stelle die Addition des Datensatzes komplett abgebildet. Der Wert, den ChatGPT 3.5 errechnet hat ist jedoch falsch, daher werden viele der falsch positiven Bewertungen daran liegen, dass sie diese Addition ChatGPT 3.5 geglaubt haben und diese nicht selbst nachgerechnet haben. Die Aufgabe ist als Abb. 4 im Anhang abgebildet.

Eine weitere Auffälligkeit ist die Aufgabe 3. Das Ergebnis ist der Vektor $\begin{pmatrix} 10 \\ -27 \\ 0 \end{pmatrix}$. Diesen haben beide Versionen von ChatGPT korrekt berechnet. Dies haben fast alle befragten Personen erkannt. Google Bard hat diese Aufgabe nicht korrekt gelöst. Die Lösung dieser KI lautet $\begin{pmatrix} 5 \\ -27 \\ 0 \end{pmatrix}$. Diese Ausgabe war wohl so nah an den Lösungen von ChatGPT 3.5 und 4, dass die Hälfte der Befragten hier ebenfalls abgestimmt haben, dass dieser Wert richtig sei.

Die Umfrage zeigt somit, dass ChatGPT 4 für die meisten befragten Personen die Aufgaben am besten erklärt hat. Außerdem waren sich die meisten befragten Personen sicher, dass ChatGPT 4 die richtige Lösung ausgegeben hat.

10. Eigenschaften eines Tutors

Um die Leitfrage lösen zu können, muss erst die Frage beantwortet werden, was einen Tutor ausmacht. Wikipedia sagt hierzu, dass der Tutor „im akademischen Bereich eine Person, die [...] mit der Begleitung, Unterrichtung und Leitung anderer Personen beauftragt ist.“ [14] In der Definition von Wikipedia heißt es weiter, dass „der Tutor in der Regel selbst auch (noch) Student [ist]“ [14]. Für diese Arbeit wird dieser Part nicht mit in die Definition mit einbezogen. Wichtiger jedoch folgt danach, dass ein Tutor nicht unbedingt direkt mit dem Unterrichtsstoff arbeiten muss und dass er „auch in anderer Hinsicht Betreuer, Ansprechpartner oder eine Vertrauensperson sein [kann]“ [14]. Des Weiteren benötigt ein Tutor genügend Fachwissen. Dieses sollte so weit reichen, dass Studierende vielseitige Fragen stellen können und diese vom Tutor beantwortet werden können, ohne lange recherchieren zu müssen. Ein weiterer Punkt wäre die pädagogische Kompetenz. Sie zeigt, dass der Tutor nicht nur das Fachwissen besitzt, sondern dieses den Studierenden auch beibringen kann. Hierzu zählen ausführliche Erklärungen und die Anwendung verschiedener Lehrmethoden. Eine weitere wichtige Eigenschaft ist, dass der Tutor zuverlässig sein sollte. Hierzu zählt bei Tutoren, die keine KI sind, dass

Veranstaltungen nicht ausfallen sollten, weil Studierende teilweise auf Tutorien angewiesen sind. Außerdem zählt zur Zuverlässigkeit, dass sich der Studierende darauf verlassen können sollte, dass der Tutor Aufgaben richtig rechnen kann und seinen Rechenweg erklären kann. Dieser Punkt ist besonders wichtig, weil Studierende sich auf Tutoren verlassen können müssen und im schlimmsten Fall Themen falsch lernen und deswegen durch eine Klausur fallen.

11. Fazit

a. Einleitung des Fazits

Im Fazit werden die Oberthemen der Arbeit zusammengefasst und abschließend die Leitfrage versucht zu beantworten.

b. Umfrage

Die Ergebnisse der Umfrage zeigen ähnliche Ergebnisse wie das Experiment. ChatGPT 4 ist der Chatbot, welcher von den meisten befragten ausgewählt wurde in der Frage welcher Chatbot die Aufgaben am besten erklärt hat. Im Experiment hat ChatGPT 4 ebenfalls das beste Ergebnis erreicht. Auf Platz zwei im Experiment und in der Umfrage ist ChatGPT 3.5. Jedoch sind die Zahlen aus der Umfrage ähnlich, wie im Experiment und zeigen, dass es einen großen Abstand zwischen ChatGPT 3.5 und ChatGPT 4 gibt. Zuletzt sind Google Bard und Copilot. Copilot konnte mehr Stimmen als Google Bard in der Umfrage erlangen, jedoch hatte Google Bard ein besseres Ergebnis im Experiment.

c. Einbindung der Literatur

Wie bereits dargestellt, ist die Meinung der Autoren, dass KI eine Bereicherung ist. Jedoch sind sich die Autoren auch einig, dass die KI aktuell noch keine Lehrkräfte ersetzt [12][13]. Hierzu können die Nutzer den Antworten der KI noch nicht genug vertrauen. Dies zeigte sich im Experiment, welches in dieser Arbeit durchgeführt wurde, durch die vielen falschen Antworten der KIs. In der Studie von Vagelis Pelvris und seinen Forscherkollegen [11] war ChatGPT 4 im ersten Teil die KI,

welche die meisten korrekten Ergebnisse ausgegeben hat. Diese Zahlen konnten im Experiment bestätigt werden. Jedoch war Google Bard im zweiten Teil besser als ChatGPT 3.5 und sogar besser als ChatGPT 4. Dies konnte in diesem Experiment nicht gezeigt werden. Hinzu kommt jedoch, dass ChatGPT 4 damals keine Internetverbindung hatte. Dies hat sich inzwischen geändert und Google Bard hat daher keinen Vorteil mehr. In der Studie aus 2023 [11] sind alle drei Chatbots nicht über eine Quote von 80% richtigen Ergebnissen gekommen. Dieses Ergebnis ist für einen Tutor zu niedrig, weil die Studierenden sich darauf verlassen können sollten, dass ein Tutor nahezu jedes Ergebnis richtig berechnet. Meiner Meinung nach sollte eine Quote von über 90% dafür gegeben sein.

d. Vergleich zwischen Chatbot und Tutor

Im dritten Teil des Fazits soll die Definition des Tutors mit den Chatbots abgeglichen werden. Chatbots können im akademischen Bereich Studierende begleiten und unterrichten. Außerdem können Chatbots ein Ansprechpartner für Studierende sein. Hierbei ist jedoch die Frage, inwieweit ein Chatbot auf die Uni – und Hochschuldaten zugreifen kann, um organisatorische Fragen zu beantworten. Dies ist ebenfalls eine Kompetenz, welche in diesen Bereich fällt. Chatbots besitzen eine große Datenbank und außer im Fall von ChatGPT 3.5 das Internet als Fachwissen, jedoch haben diese beiden Datensätze nicht dazu geführt, dass die Chatbots alle Fragen im Experiment richtig gelöst haben. Ein Tutor besitzt in der Regel eine pädagogische Kompetenz. Die Chatbots haben ebenfalls die Fähigkeit ihre Lösungen zu erklären, womit sie in dieser Kategorie punkten können. Außerdem können Chatbots auf weitere Fragen eingehen, um tiefere Verständnisfragen zu lösen. Hierbei ist jedoch nicht gegeben, dass die Erklärungen des Chatbots korrekt sind. Womit die letzte Tutor Eigenschaft gemeint ist. Die Chatbots sind nicht zuverlässig genug, weil die Lösung oft falsch ist oder gar keine Lösung ausgegeben wird.

e. Abschluss

Die Leitfrage, ob sich Chatbots als Tutor für die Mathematikausbildung eignen, kann in dieser Arbeit nicht beantwortet werden, weil das Gebiet der Mathematikausbildung zu groß ist, um es definitiv testen zu können. Das Experiment in dieser Arbeit ist lediglich eine Stichprobe. Jedoch kann es eine Tendenz aufzeigen. Diese Tendenz geht eher in die Richtung, dass sich die kostenlosen Chatbots aktuell noch nicht als Tutor eignen. Der einzige kostenpflichtige Chatbot ist der, welcher in den Tests und auch in der Umfrage am besten abgeschnitten hat. Dieser ist ChatGPT 4. Jedoch waren selbst bei dieser KI die Erklärungen teilweise nicht ausreichend, um einen richtigen Tutor zu ersetzen. ChatGPT 4 hat zusätzlich die Berechnung selten in Rechenschritten angezeigt. Genau das ist aber immens wichtig für Studierende, um den Rechenweg nachvollziehen zu können und dann in der Klausur selbstständig die Lösung errechnen zu können. Daher ist es problematisch, wenn ChatGPT 4 diese Rechenschritte vorenthält. Der Chatbot, welcher von den kostenfreien Chatbots am besten im Experiment und in der Umfrage abgeschlossen hat, ist ChatGPT 3.5. Der Abstand zwischen ChatGPT 4 und ChatGPT 3.5 ist allerdings groß, denn ChatGPT 3.5 hat 22 Aufgaben mehr falsch gerechnet als ChatGPT 4. Studierende sollte sich auf den Tutor verlassen können, dass dieser selten einen Fehler macht. ChatGPT 3.5 hat mehr als die Hälfte der Aufgaben falsch gerechnet und ist somit nicht vertrauenswürdig genug. Google Bard hat sowohl im Experiment als auch in der Umfrage schlechter abgeschnitten als ChatGPT 3.5. Copilot hat, obwohl es das gleiche LLM wie ChatGPT 4 nutzt, ebenfalls in Experiment und Umfrage schlechtere Ergebnisse als beide KIs von OpenAI.

12. Quellen

- [1] OpenAI: „ChatGPT 3.5 und 4“, <https://chat.openai.com/>, 21.02.2024

- [2] Google: „Google Bard (Gemini)“, <https://gemini.google.com/app>, 21.02.2024
- [3] Microsoft: „Copilot“, <https://copilot.microsoft.com/>, 22.02.24
- [4] Wikipedia: „Google Gemini“, https://de.wikipedia.org/wiki/Google_Gemini, 07.03.24
- [5] TheVerge: „Google announces PaLM 2 AI language model, already powering 25 Google services“, <https://www.theverge.com/2023/5/10/23718046/google-ai-palm-2-language-model-bard-io>, 07.03.24
- [6] Bernhard Halbig, Klaus Giebertmann: „WMINT“, <https://wmint.de/mathematik/>, 14.03.2024
- [7] Georg Laurenz Dittlbacher, DerStandard: „Unetisch, gefährlich: Wie KI-Tools die Unterhaltungsindustrie bedrohen“, <https://www.derstandard.de/story/3000000211273/the-ai-revolution-is-televised>, 18.03.24
- [8] Ole Reißmann, Spiegel.de: „Die große Angst vor den KI-Fakes“, <https://www.spiegel.de/netzwelt/netzpolitik/wahlkampf-in-den-usa-die-grosse-angst-vor-den-ki-fakes-a-fd7afd75-d225-4627-8032-826023389978>, 18.03.24
- [9] Stephen Wolfram, heise.de: „Missing Link: Stephen Wolfram über die Rolle der KI in der Forschung (Teil 2)“, <https://www.heise.de/hintergrund/Missing-Link-Stephen-Wolfram-ueber-die-Rolle-der-KI-in-der-Forschung-Teil-2-9650643.html>, 18.03.24
- [10] NDR: „Schlaglöcher: Stadt Goslar setzt bei Erfassung auf KI“, https://www.ndr.de/nachrichten/niedersachsen/braunschweig_harz_goettingen/Schlagloecher-Stadt-Goslar-setzt-bei-Erfassung-auf-KI,aktuellbraunschweig12470.html, 18.03.24
- [11] Vagelis Plevris, George Papazafeiropoulos and Alejandro Jiménez Rios: „Chatbots Put to the Test in Math and Logic Problems: A Comparison and Assessment of ChatGPT-3.5, ChatGPT-4, and Google Bard“, <https://www.mdpi.com/2673-2688/4/4/48>, 18.01.24

[12] Niroj Dahal, Basanta Raj Lamichhane, Bal Chandra Luitel, Binod Prasad Pant: „Chatbots as Math Algorithm Problem Solvers: A Critical Evaluation of Its Capabilities and imitations“, <https://atcm.mathandtech.org/EP2023/regular/30102.pdf>,

18.01.24

- [13] Horst Hirscher: „ChatGPT und Mathematikunterricht – eine didaktische Herausforderung?“, <https://www.math.uni-sb.de/preprints/preprint409.pdf>, 10.04.24
- [14] Wikipedia: „Tutor“, <https://de.wikipedia.org/wiki/Tutor>, 12.04.24
- [15] Spiegel: „Mehrheit der Studierenden erwartet, dass KI ihr Berufsleben erleichtert“, <https://www.spiegel.de/start/ki-im-berufsleben-mehrheit-der-studierenden-erwartet-dass-ki-ihr-berufsleben-erleichtert-a-289d087a-531d-4273-852e-907d1a1a9181>, 13.04.24

13. Anhang

Question: Welche KI hat dir die Aufgabe am besten erklärt, oder war keine Erklärung ausreichend?		
	COUNT	PERCENT
ChatGPT 3.5	7	33.33%
ChatGPT 4	9	42.86%
Google Bard	0	0.00%
Bing Chat	2	9.52%
Keine der KIs	2	9.52%

Abbildung 47 Umfrageergebnis groß 1

Question: Welche KI hat dir die Aufgabe am besten erklärt, oder war keine Erklärung ausreichend?		
	COUNT	PERCENT
ChatGPT 3.5	2	11.11%
ChatGPT 4	11	61.11%
Google Bard	0	0.00%
Bing Chat	0	0.00%
Keine der KIs	5	27.78%

Abbildung 48 Umfrageergebnis groß 2

Question: Welche KI hat dir die Aufgabe am besten erklärt, oder war keine Erklärung ausreichend?		
	COUNT	PERCENT
ChatGPT 3.5	4	22.22%
ChatGPT 4	13	72.22%
Google Bard	0	0.00%
Bing Chat	0	0.00%
Keine der KIs	1	5.56%

Abbildung 49 Umfrageergebnis groß 3

Abb. 4:

ChatGPT

Das arithmetische Mittel (auch Durchschnitt genannt) für eine Reihe von Werten wird berechnet, indem man die Summe aller Werte durch die Anzahl der Werte teilt.

Für die Datenreihe:

$$x : 0, 7, 17, 27, 34, 39, 42, 47$$

$$y : 0, 8, 19, 28, 40, 48, 53, 58$$

Berechnen wir die arithmetischen Mittel:

Für x :

$$\text{Mittelwert von } x = \frac{0+7+17+27+34+39+42+47}{8}$$

$$\text{Mittelwert von } x = \frac{253}{8}$$

Für y :

$$\text{Mittelwert von } y = \frac{0+8+19+28+40+48+53+58}{8}$$

$$\text{Mittelwert von } y = \frac{254}{8}$$

Daher beträgt der Mittelwert für x $\frac{253}{8}$ und für y $\frac{254}{8}$. Sie können diese Werte auch als Dezimalzahlen ausdrücken, wenn gewünscht.

You

der mittelwert für x ist falsch bitte berechne diesen erneut

ChatGPT

Entschuldigen Sie bitte den Fehler. Lassen Sie uns den Mittelwert für x erneut berechnen:

$$\text{Mittelwert von } x = \frac{0+7+17+27+34+39+42+47}{8}$$

$$\text{Mittelwert von } x = \frac{253}{8}$$

Der korrekte Mittelwert für x beträgt also $\frac{253}{8}$.

Ich versichere an Eides Statt, die von mir vorgelegte Arbeit selbständig verfasst zu haben. Alle Stellen, die wörtlich oder sinngemäß aus veröffentlichten Arbeiten anderer entnommen sind, habe ich als entnommen kenntlich gemacht. Sämtliche Quellen und Hilfsmittel, die ich für die Arbeit benutzt habe, sind angegeben. Die Arbeit hat mit gleichem Inhalt bzw. in wesentlichen Teilen noch keinen anderen Prüfungsbehörde vorgelegen.

16.04.2024, Köln, Jonathan Paul

