



Fachhochschule Nordwestschweiz
Hochschule für Angewandte Psychologie

Bedingungen und Massnahmen für eine erfolgreiche Interaktion mit Large Language Models

BACHELOR THESIS

Juni 2024

Autorin

Christe, Mathilde

Betreuungsperson

Theiler, Sven

Praxispartner

Institut Mensch in komplexen Systemen (MikS)

Prof. Dr. Yanik Sterchi

Abstract

Diese Bachelorarbeit untersucht, ob eine Instruktion die Fähigkeiten und das Nutzungsverhalten der Studierenden im Umgang mit ChatGPT verbessern kann. Zudem wird der Zusammenhang zwischen dem Umgang mit ChatGPT und der Selbstwirksamkeit erforscht. Im Rahmen dieser Studie wurden Studierende mittels Fragebogen gebeten, ihre Selbstwirksamkeit sowie ihre Nutzungsabsicht einzuschätzen und anschliessend 15 ChatGPT-Aufgaben zu beurteilen. Die Konstrukte Selbstwirksamkeit sowie die Nutzungsabsicht der Studierenden wurden mit wissenschaftlich validierten Items erfasst. Mittels t-Test und Spearman-Korrelation wurden die formulierten Hypothesen überprüft. Die Stichprobe (N=69) umfasst Studierende aus diversen Hochschulen sowie unterschiedlichen Alters und Geschlechts. Eine Instruktion beeinflusst den Umgang mit ChatGPT sowie das Nutzungsverhalten der Studierenden nicht signifikant. Zudem wurde festgestellt, dass eine höhere Selbstwirksamkeit keinen signifikanten Zusammenhang mit dem Umgang mit ChatGPT hat. Anhand der Analyse der Daten konnten Handlungsempfehlungen abgeleitet werden, um die Implementierung von Large Language Models (LLM`s) an Hochschulen zu unterstützen. Die daraus resultierenden Handlungsempfehlungen umfassen das Anbieten von Schulungen, um Studierende im Umgang mit LLM`s wie ChatGPT zu bilden, einschliesslich ethischer Aspekte und Datenschutzfragen. Darüber hinaus sollten klare Richtlinien für die Nutzung von KI-Technologien erstellt werden, um Transparenz zu gewährleisten und Studierende im kritischen Umgang mit KI-generierten Texten zu schulen.

Schlüsselwörter: Large Language Models (LLM`s), ChatGPT, Selbstwirksamkeit, Nutzungsverhalten, Instruktion, kritischer Umgang, Richtlinien

Anzahl Wörter Abstract: 190

Anzahl Zeichen Bachelorarbeit (inkl. Leerzeichen, exkl. Anhang): 105`543

Inhaltsverzeichnis

1	Einleitung.....	1
1.1	<i>Zielsetzung und Ableitung der Fragestellungen</i>	2
1.2	<i>Praxispartner</i>	3
1.3	<i>Aufbau der Arbeit</i>	3
2	Theoretische Grundlagen.....	4
2.1	<i>Künstliche Intelligenz</i>	4
2.2	<i>Large Language Models</i>	5
2.3	<i>ChatGPT.....</i>	6
2.4	<i>Mensch-Computer-Interaktion</i>	9
2.5	<i>Lernen.....</i>	10
2.6	<i>Selbstwirksamkeit.....</i>	12
2.7	<i>Nutzungsabsicht.....</i>	12
3	Methodik.....	13
3.1	<i>Forschungsansatz und Begründung</i>	13
3.2	<i>Hypothesenbildung.....</i>	13
3.3	<i>Instrument zur Erhebung der Daten und Testverfahren</i>	14
3.4	<i>Stichprobe und Akquise</i>	15
3.5	<i>Aufbau und Inhalt des Fragebogens</i>	15
3.6	<i>Pretest</i>	21
3.7	<i>Bereinigung des Datensatzes</i>	21
4	Auswertung der Datenerhebung / Ergebnisse.....	22
4.1	<i>Analyse der Stichprobe</i>	22
4.2	<i>Explorative Datenanalyse</i>	23
4.2.1	<i>Nutzung von LLM`s der Studierenden.....</i>	23
4.2.2	<i>Weitere Variablen in Bezug auf die Bedingungen und Massnahmen für eine erfolgreiche Interaktion mit LLM`s.....</i>	25
4.3	<i>Überprüfung der Hypothese HS1</i>	27
4.4	<i>Überprüfung der Hypothese HS2</i>	29
4.5	<i>Überprüfung der Hypothese HS3</i>	30
5	Diskussion	32
5.1	<i>Beantwortung der ersten Fragestellung</i>	32
5.2	<i>Beantwortung der zweiten Fragestellung</i>	33
5.3	<i>Beantwortung der dritten Fragestellung.....</i>	34

5.4	<i>Handlungsempfehlungen</i>	35
5.5	<i>Limitationen dieser Arbeit</i>	36
6	Fazit und weiterführende Forschungsansätze	37
7	Literaturverzeichnis	38
8	Tabellenverzeichnis	46
9	Abbildungsverzeichnis	46
10	Anhang	47

1 Einleitung

Gemäss Seidel und Macht (2023) hält die Künstliche Intelligenz (KI) «zunehmend Einzug in viele Lebensbereiche von Menschen. KI soll das Leben der Menschen vereinfachen und sie in einer Vielzahl von Situationen unterstützen» (S.124). Die KI kommt bereits in vielen Prozessen zum Einsatz, ohne dass uns dessen bewusst ist. Mittlerweile existieren zahlreiche KI-Tools, welche dafür entwickelt wurden, Aspekte aus unserem modernen Leben zu erleichtern, darunter persönliche Assistenten, wie Google Home, Alexa und Siri. Mit ihren vielfältigen Anwendungsmöglichkeiten scheint die KI gemäss Rainsberger (2021) wie eine Art Wunderland zu wirken, denn vieles erscheint gleichzeitig möglich und unmöglich, nah und doch fern zu sein. Gemäss Schallmo, Williams und Tidd (2022), zählen zu den bedeutendsten gesellschaftlichen und wirtschaftlichen Entwicklungen unserer Zeit der Einsatz von menschenähnlichen und menschengleichen Maschinen sowie die digitale Transformation. Die Interaktion zwischen Menschen und Computern ist heutzutage alltäglich geworden und kaum wegzudenken. Ausserdem hat sie sich gemäss Fuchs (2018) tief in unsere physische und soziale Umgebung integriert. Sie befasst sich mit der benutzer- und aufgabenorientierten Gestaltung sowie der Bewertung von interaktiven Systemen. Ausserdem gewinnt die Mensch-Computer-Interaktion gemäss Gross (2013) sowohl in der Wissenschaft als auch in der Praxis zunehmend an Bedeutung. Seit der Einführung von ChatGPT-3.5, einer verbesserten Version von GPT-3, durch OpenAI, hat das Programm weltweit einen regelrechten Boom in der Nutzung erfahren, insbesondere um menschenähnliche Interaktionen mit Nutzern zu ermöglichen. ChatGPT ist ein grosses Sprachmodell, ein sogenanntes Large Language Model (LLM), welches von OpenAI entwickelt wurde. ChatGPT verwendet Deep-Learning-Algorithmen, um das nächste Wort in einem Text vorherzusagen. Dazu greift es auf einen grossen Datensatz zurück (Kim, Chua, Rickard & Lorenzo 2023). «Seitdem im November 2022 ChatGPT öffentlich und kostenfrei zugänglich wurde, häufen sich Artikel, Berichte und Tutorials darüber, wie ChatGPT Schulabschlüsse und Zertifizierungen erlangen sowie Auswahlverfahren bestehen kann» (Römer, Kleinmanns & Jobmann, 2023, S. 2). Gemäss Seemann (2023) sollte der Umgang mit Large Language Models (LLM`s) nicht unterschätzt werden, da sie sowohl positive wie auch negative Eigenschaften mit sich bringen. Weshalb es wichtig ist nicht nur ihre Funktionsweise zu verstehen, sondern auch praktische Erfahrungen zu sammeln, bevor man LLM`s einsetzt. Denn erst mit der Zeit erkennt man wofür LLM`s geeignet sind, wo sie Fehler machen und in welchen Bereichen, sie vorsichtig eingesetzt werden sollten und wann nicht auf die selbstsicheren Aussagen der LLM`s vertraut werden darf. Die KI-basierten Tools wie beispielsweise ChatGPT verändern aktuell gemäss Lew (2023) die Hochschullandschaften.

Gemäss Aldosari (2020) soll die KI zukünftig die Entwicklung der Hochschulen prägen. Damit der Einsatz von KI erfolgreich in Hochschulen integriert werden kann, müssen gemäss Büching, Mah, Otto, Paulicke und Hartman (2019) wesentliche Probleme geklärt werden. In den Hochschulbereichen ergeben sich gemäss Schmohl, Watanabe und Schelling (2023) durch den Einsatz der KI mehrere wichtige Ansätze. Die Studierenden können durch den Einsatz der KI an Hochschulen neue Erkenntnisse gewinnen. Daher ist es zunehmend entscheidend, die KI-Kompetenzen an Hochschulen zu integrieren. Die KI trägt dazu bei, dass die Studierenden kognitive Prozesse und Fähigkeiten genauer untersuchen können, was wiederum neue Möglichkeiten für die Forschung im Bereich Lehren und Lernen eröffnet. Ausserdem kann gemäss Schmohl et al. (2023) die KI in der Hochschulbildung als Unterstützung in verschiedenen Bereichen angesehen werden, wie beispielsweise als eine Lernhilfe für Studierende. Die Problemstellung legt nahe, einen genaueren Blick auf die Bedingungen und Massnahmen für eine erfolgreiche Interaktion mit LLM's zu werfen.

1.1 Zielsetzung und Ableitung der Fragestellungen

In Anbetracht der Auswirkungen von KI, insbesondere ChatGPT an Hochschulen und der Argumentation, dass es zunehmend entscheidend ist, die KI-Kompetenzen an Hochschulen zu integrieren, verfolgt diese Arbeit mehrere Ziele.

Zum einen soll untersucht werden, wie sich eine Bereitstellung einer Instruktion von Large Language Models auf den Umgang der Studierenden mit ChatGPT auswirkt. Als zweites Ziel, soll die Beeinflussung des Nutzungsverhalten durch eine Instruktion untersucht werden. Weiter soll der Zusammenhang zwischen der Selbstwirksamkeit der Studierenden und ihrem Umgang mit ChatGPT untersucht werden. Folgende Fragestellungen, lassen sich aufgrund dieser Ziele ableiten:

1. Führt die Bereitstellung einer Instruktion zur Nutzung von Large Language Models, wie ChatGPT, zu besseren Fähigkeiten der Studierenden im Umgang mit ChatGPT im Vergleich zu jenen Studierenden, welche keine spezifische Instruktion erhalten haben?
2. Wie wird das Nutzungsverhalten der Studierenden durch eine Instruktion beeinflusst?
3. Gibt es einen Zusammenhang zwischen Selbstwirksamkeit und dem Umgang mit ChatGPT?

1.2 Praxispartner

Das Institut Mensch in komplexen Systemen (MikS) ist Teil der Hochschule für Angewandte Psychologie an der Fachhochschule Nordwestschweiz (FHNW). Das MikS beschäftigt sich mit Erforschung und Gestaltung des Zusammenspiels zwischen Mensch, Technik und Organisationen. Das Ziel des MikS ist es, die Interaktion in komplexen Systemen zu verstehen und zu optimieren, besonders im Bereich der Automatisierung, Digitalisierung und technische Vernetzung. Die Forschungs- und Tätigkeitsfelder des MikS sind: Arbeit und Gesundheit, Eignung und Personalentwicklung sowie Zuverlässigkeit und Sicherheit.

Dabei verfolgt das MikS folgende Ziele:

- Förderung der Gesundheit und Leistungsfähigkeit von Menschen und Organisationen
- Optimierung der Passung von Anforderungen und Personenmerkmalen
- Erhöhung von Zuverlässigkeit und Sicherheit in soziotechnischen Systemen

Die MikS trägt durch ihre interdisziplinäre Forschung und praxisorientierte Ansätze dazu bei, die Interaktion zwischen Mensch und Technik zu verbessern (FHNW, n.d.)

1.3 Aufbau der Arbeit

Das Kapitel eins (1) soll, nachdem die Einleitung in die Thematik erfolgt, die Fragestellungen festgelegt und der Praxispartner erläutert wurde, mit dem Aufbau der Arbeit abgeschlossen werden. Das nachfolgende Kapitel zwei (2) dient den theoretischen Grundlagen, um ein gemeinsames Verständnis der Schlüsselbegriffe und Grundlagen zu schaffen. Das detaillierte Vorgehen der Datenerhebung sowie die Grundlagen für die Datenauswertung wird in Kapitel drei (3) erläutert. Nachfolgend werden im vierten (4) Kapitel die Ergebnisse der Datenerhebung dargestellt. In Kapitel fünf (5) folgt die Diskussion der Ergebnisse, gefolgt von den Handlungsempfehlungen und den Limitationen dieser empirischen Untersuchung. Das abschliessende Kapitel sechs (6) endet mit einem Fazit und Ausblick für weiterführende Forschungsansätze.

2 Theoretische Grundlagen

In den nachfolgenden Kapiteln 2.1-2.7 werden Grundlagen und Schlüsselbegriffe erläutert, die für das Verständnis dieser vorliegenden Arbeit von Bedeutung sind. Insbesondere wird auf die Bedeutung und den Kontext eingegangen, um deren Relevanz für den Anwendungsbereich im Untersuchungsansatz im darauffolgenden Kapitel drei (3) aufzuzeigen.

2.1 Künstliche Intelligenz

Die Künstliche Intelligenz (KI) befasst sich mit der Automatisierung intelligenten Verhaltens und gehört der Informatik als Teilgebiet an (Sieber, 2019).

Künstliche Intelligenz, wurde als erstes 1955 von John McCarthy wie folgt definiert, KI ist die Wissenschaft und Technologie, welche sich damit beschäftigt, intelligente Maschinen zu entwickeln besonders intelligente Computerprogramme. Das Ziel dabei ist es Computer so zu programmieren, dass sie menschliche Intelligenz verstehen können (McCarthy, 1996).

Der Begriff Künstliche Intelligenz (KI) wird auf viele verschiedene Arten interpretiert, was es schwierig macht, die Künstliche Intelligenz klar und verständlich zu definieren (Ertel, 2021). Gemäss Harwardt und Köhler (2023) ist die Künstliche Intelligenz ein sogenanntes «Kofferwort», darunter zu verstehen ist, dass jeder Künstliche Intelligenz so definiert, wie es für jeden persönlich richtig empfunden wird. Gemäss Paass und Hecker (2020) lässt sich sagen, dass von Künstlicher Intelligenz gesprochen wird, wenn die Fähigkeiten, wie geistige Wahrnehmung und kognitive Fähigkeiten von Computern oder Maschinen, denen von Menschen ähneln. Im Allgemein kann die KI als System betrachtet werden, welches menschenähnliches Denken oder Verhalten aufweist und rational handelt (Russell, Norvig & Davis, 2010). Obwohl der Begriff der Künstlichen Intelligenz nicht einheitlich definiert wird, nimmt die KI mittlerweile einen sehr grossen Stellenwert ein (Sieber, 2019). Der Kern der KI liegt darin, Funktionen zu übernehmen, welche menschliche Handlungen bisher ausmachten. Ausserdem erschliessen sich zahlreiche Anwendungsmöglichkeiten, Dank der Fähigkeit von KI-Systemen, welche umfangreiche Datensätze analysieren, Muster erkennen und neue Inhalte generieren (Kreutzer, 2023). Gemäss Russel et al. (2010) umfasst die KI eine Vielzahl von Teilbereichen, welche von allgemeinen Fähigkeiten, wie Lernen und Wahrnehmung bis hin zu spezifischen Tätigkeiten reichen. Die Künstliche Intelligenz ist ein vielschichtiges Forschungsfeld, welches zunehmend nicht nur von Fachleuten erforscht und genutzt wird. Sie findet immer mehr Anwendung und Interesse in der breiten Öffentlichkeit sowie in unterschiedlichen Alltagsbereichen. Wir begegnen KI in vielen Formen, welche häufig nicht direkt als solche erkennbar sind. Was uns bei der Nutzung der KI auffällt, ist die

Einfachheit und Effizienz bei der Durchführung bestimmter Aufgaben im Vergleich zu früher. Ein Beispiel dafür sind unsere persönlichen digitalen Assistenten, wie zum Beispiel Google Home, Alexa und Siri, welche verschiedene Funktionen per Sprachbefehl übernehmen können (Kreutzer, 2023). Die Akzeptanz der Menschen gegenüber der KI hat durch die Sprachassistenten, wie beispielsweise Siri auf dem Smartphone zugenommen (Sieber, 2019).

2.2 Large Language Models

Nachdem die Künstliche Intelligenz erläutert wurde, werden in diesem Kapitel Large Language Models (LLM's) betrachtet. Gemäss Paass und Hecker (2020) sind Large Language Models auf Deutsch übersetzt grosse Sprachmodelle, ein neuronales Netzwerk, welches in der Lage ist, das nächste Wort in einem Satz vorherzusagen. Sprachmodelle sind sogenannte Deep-Learning Modelle, welche unbeaufsichtigt auf Textmengen mit grosser Anzahl von Parametern trainiert werden (Valmeekam, Marquez, Olmo, Sreedharan & Kambhampati, 2023). Unter unbeaufsichtigtem Lernen ist zu verstehen, dass grosse Datenmengen von Menschen nicht manuell überprüft und verarbeitet werden können und somit Sprachmodelle eigenständige Sprachstrukturen erschliessen können (Paass und Hecker, 2020). Hierdurch erreichen die Sprachmodelle, «die Fähigkeit, selbst komplexe Aufgabenstellungen in Sekundenschnelle zu bearbeiten und detaillierte Antworten auf hohem sprachlichem Niveau in textbasierte Form auszugeben» (Fink, 2023, S.665). Unter Deep Learning «werden Methoden bezeichnet, mit denen Berechnungen nicht durch das Abarbeiten herkömmlicher Programmcodes erzielt werden, sondern durch das Zusammenwirken vieler vergleichsweise einfacher Elemente, die in vager Anlehnung an die Bauweise unseres Gehirns «Neuronen» genannt werden» (Geissler, 2017, S.383). Wichtig zu beachten ist, dass beim Deep-Learning alle Informationen, welche das neuronale Netzwerk aus den Rohdaten lernt, bereits in den enthaltenden Daten vorhanden sein müssen (Amigo, 2021). Zusammenfassend lässt sich gemäss Fink (2023) sagen, dass LLM's durch ihre Vielseitigkeit und Leistungsfähigkeit vielversprechende Werkzeuge sind. Allerdings sind noch grundlegende Fragen wie zum Beispiel zu den Themen Datenschutz, Zuverlässigkeit und den Limitationen der Modelle ungeklärt, um eine effektive und sichere Anwendung gewährleisten zu können. Der Bekanntheitsgrad von LLM's basierend auf Künstlicher Intelligenz erhöhte sich Ende November 2022 durch die Einführung von ChatGPT (Fink, 2023).

2.3 ChatGPT

Gemäss Spitzer (2023) ist ChatGPT ein Chatbot, welcher in den letzten Monaten viel Aufsehen erregt hat. Unter Chatbot ist ein technisches System zu verstehen, mit dem man sich unterhalten kann. Die Abkürzung ChatGPT steht für «generative pretrained transformer», was auf Deutsch mit «hervorbringender vortrainierter Umformer» übersetzt werden kann. ChatGPT wurde in den USA durch die Firma Fa. OpenAI (artificial intelligence) entwickelt (Stavridis & Wacker, 2023). ChatGPT ist ein auf KI gestütztes Computerprogramm für menschenähnliche Konversationen. Man spricht von einem sogenannten Chatbot. GPT-3 mittlerweile GPT-3.5 bietet eine Echtzeitkommunikation und ist für alle kostenlos zugänglich (Krettek, 2023). ChatGPT wurde mit Millionen von Daten aus dem Internet trainiert, um gestellte Aufgaben mithilfe dieser Informationen verarbeiten zu können. Daraus generiert die KI kontinuierlich auf einer Vielzahl von Fragestellungen passende Antworten. Wichtig zu beachten ist jedoch, dass ChatGPT-3.5 einen limitierten Informationsstand hat und nur bis zum Jahr 2021 aktualisiert ist und deshalb keine Kenntnisse über neuere Informationen hat (Stenzel, 2024). Ausserdem besteht eine grosse Menge der Trainingsdaten aufgrund von Menschen erstellter Texte aus Quellen wie Büchern, Sozialen Medien und viele andere (Krettek, 2023). Dadurch hat ChatGPT gelernt, mit welcher Wahrscheinlichkeit welche Wörter aufeinanderfolgen und dies auch über mehrere Wörter (Spitzer, 2023). Die Anwendungsmöglichkeiten von ChatGPT sind umfangreich. Mit ChatGPT lassen sich freie Texte, wie beispielsweise Bewerbungen, E-Mails, Briefe sowie Zusammenfassungen erstellen. Des Weiteren können Texte in anderen Sprachen übersetzt werden. Ausserdem erstellt ChatGPT Multiple-Choice-Prüfungsfragen und überprüft die Antworten. ChatGPT versteht auch komplexe mathematische Zusammenhänge und den Satz des Pythagoras. Zudem kann ChatGPT in verschiedenen Programmiersprachen Programmcode schreiben, Fehler erkennen und Verbesserungsvorschläge bringen (Krettek, 2023). Gemäss Spitzer (2023) werden die Texte nach Anweisungen sehr präzise geschrieben (z.B. beschreibe mir eine Frucht, mit 50 Wörter), dass diese kaum von geschriebenen Texten von Studenten zu unterscheiden sind. Wird die gleiche Anweisung mehrfach an ChatGPT gegeben, so wird meistens ein neuer Text produziert, wie in den untenstehenden Abbildungen 1 und 2 zu sehen ist.

**You**

beschreibe mir eine Frucht, mit 50 Wörter

**ChatGPT**

Die Ananas, eine tropische Frucht, zeichnet sich durch ihre stachelige, grün-gelbe Schale und ihr süß-saures, saftiges Fruchtfleisch aus. Ihr intensiver Duft und ihr exotischer Geschmack machen sie zu einem beliebten Genuss. Die Ananas ist reich an Vitamin C und Enzymen und vielseitig in der Küche verwendbar.

Abbildung 1. Beschreibung der Frucht Ananas selbst erstellter Prompt mit ChatGPT

**You**

beschreibe mir eine Frucht, mit 50 Wörter

**ChatGPT**

Die Mango, eine exotische Frucht mit glatter, goldener Haut und saftigem, süßem Fruchtfleisch, ist eine wahre Delikatesse. Ihr Aroma ist betörend und ihr Geschmack ein Zusammenspiel von süßen und leicht säuerlichen Noten. Die Mango ist reich an Vitaminen und Antioxidantien und ein Genuss für die Sinne.

Abbildung 2. Beschreibung der Frucht Mango selbst erstellter Prompt mit ChatGPT

«Noch vor der Benutzung von ChatGPT poppt jedoch eine Nachricht auf, die jedem Nutzenden zu denken geben sollte: „It is not intended to give advice.“ („Es sollte nicht als Ratgeber verwendet werden“») (Spitzer 2023, S. 194). Wie bereits erwähnt, basiert ChatGPT auf Wahrscheinlichkeiten, was bedeutet, dass der Wahrheitsgehalt der Informationen nicht garantiert ist. Wenn die Aufgabe von ChatGPT unsinnig oder falsch ist, aber für die Nutzenden als sehr wahrheitsgemäss erscheint, spricht man von Halluzinationen. Dies führt dazu, dass die Texte nicht immer zuverlässig und wahrheitsgemäss sind. Diese Halluzinationen können aus den Trainingsdaten stammen, welche Duplikationen und Informationslücken enthalten und deshalb zu inkorrekten Ergebnissen führen können (Artificial intelligence a modern approach, 2022). Möglicherweise ist das unbedeutend für Texte, bei welchen die Wahrheit keine Rolle spielt.

Bei Texten, die einen Anspruch auf Wahrheit haben, ist dies jedoch essenziell. Aus diesem Grund ist ChatGPT in der Wissenschaft je nach Kontext sorgfältig anzuwenden. Selbst wenn zukünftige Softwares versuchen, den Wahrheitsgehalt zu maximieren, wird es keine vollständige Gewissheit dafür geben. Nicht zu vernachlässigen ist, dass LLM's mit Sprache trainiert wird, die Unwahrheiten, Vorurteilen sowie veraltete Meinungen und Fakten enthalten kann (Spitzer, 2023).

Nach intensiver Auseinandersetzung mit ChatGPT, stellt sich laut Dale (2021) eine entscheidende Frage: Ist einem Programm wie ChatGPT zu vertrauen?

Gemäss Dale (2021) haben verschiedene Beobachtende folgendes festgestellt:

- Die produzierten Texte können unverständlich sein sowie Mängel in inhaltsmässigen Zusammenhängen aufweisen, was dazu führt, dass die Texte, je länger sie sind, umso unsinniger werden.
- Alle Texte, die von ChatGPT erstellt werden, können Vorurteile und falsche Informationen enthalten, welche das Programm in den Trainingsdaten gelernt hat. Wird beispielsweise nach Texten gesucht, die rassistische Ansichten unterstützten, wird ChatGPT immer wieder solche Texte erzeugen.
- Die Ergebnisse können Aussagen enthalten, welche nicht wahrheitsgemäss sind

Basierend auf diesen Schwächen empfiehlt Dale (2021), die Aufforderungen mehrmals mit der gleichen Eingabe auszuführen und anhand der verschiedenen Ausgaben das beste Ergebnis für sich selbst auszuwählen, damit weniger überzeugende und unsinnige Ergebnisse vermieden werden können.

Zusammenfassend lässt sich festhalten, dass ChatGPT-3.5 einen zeitlich limitierten Informationsstand bis ins Jahr 2021 hat (Stenzel, 2024). «Wie bereits allgemein festgestellt wurde, hat die Technologie ihre Grenzen» (Dale, 2021, S.116). Ausserdem kann gemäss Brown et al. (2020) gesagt werden, dass ChatGPT für die Bewältigung einer Aufgabe viel mehr Trainingsdaten braucht, als ein Mensch jemals verarbeiten könnte. Konkret bedeutet dies, dass die Menge an Daten, die es benötigt, um eine Aufgabe zu bewältigen enorm gross ist und über das hinausgeht, was ein einzelner Mensch aufnehmen oder verarbeiten könnte. Deshalb ist es mit den derzeitigen Ansätzen nicht machbar, eine KI zu entwickeln, welche menschenähnlich ist. Das Können eines Menschen ist nicht mit dem Können einer Maschine wie beispielsweise ChatGPT zu vergleichen.

2.4 Mensch-Computer-Interaktion

Die Digitalisierung hat in den vergangenen Jahren rasch zugenommen. Besonders relevant ist dabei die Künstliche Intelligenz geworden, wie zuvor in Kapitel 2.1 erläutert wurde. Die tägliche Interaktion mit Computern ist zu einem festen Bestandteil unseres Alltags geworden und kaum weg zu denken. Dabei bilden besonders die humanorientierten Gestaltungen Mensch-Computer-Interaktion (MCI) Herausforderungen und spielen bei der Interaktion eine wichtige Rolle (Reuter, 2021). Wie Menschen mit Computersystemen interagieren, damit beschäftigt sich die MCI (Niels, 2019). Trotz der grossen Relevanz bleibt der Begriff Mensch-Computer-Interaktion keiner kompakten und eindeutigen Definition zugeordnet (Jetter, 2022). Zunehmend wird die Unterscheidung zwischen Mensch-Maschine-Interaktion (MMI) und Mensch-Computer-Interaktion (MCI) schwieriger. Der Grund dafür ist, dass technische Geräte immer mehr Komponenten wie Touchscreens, Tastaturen und Bildschirme enthalten, welche zuvor nur in Computern zu finden waren (Heinecke, 2012). Weshalb in dieser Arbeit nur der Begriff MCI verwendet wird. Gemäss Koch, Ziegler, Reuter, Schlegel und Prilla (2020) befasst sich die MCI mit den benutzer- und kontextgerechten Gestaltungen der Digitalisierung. Die neuen Formen der Mensch-Computer-Interaktion sowie die fortschreitende Digitalisierung verändern die menschlichen Kompetenzen erheblich. Besonders durch die hohe Zunahme komplementärer Technologien erfordert die Nutzung von Computern eine Anpassung an deren Funktionsweise (D'Mello und Graesser, 2013). Weshalb es laut De Witt, Gloerfeld und Wrede (2023) wichtig ist, die jeweiligen Stärken von Menschen und Computern zu identifizieren. Betrachtet man die Interaktion am Beispiel der Künstlichen Intelligenz, kann gesagt werden, dass die KI viele zeitaufwändige Aufgaben, wie das Suchen oder Zusammenfassen von Informationen übernehmen kann. In der Zusammenarbeit zwischen Menschen und KI kann dies gemäss De Witt et al. (2023) zur Qualitätssteigerung führen, da vorhandenes Wissen effizienter genutzt werden kann. Die Interaktion zwischen Computer und Mensch sollte gemäss Kellogg, Valentine und Christin (2020) unter Berücksichtigung ihrer Stärken und Schwächen eingesetzt werden. Wobei das jeweilige digitale Hilfsmittel wie beispielsweise die KI gut verstanden werden muss, um die KI gezielt und reflektiert, unter Berücksichtigung ihrer Stärken und Einschränkungen einsetzen zu können. Um eine erfolgreiche Mensch-Computer-Interaktion zu gestalten, ist es wichtig eine klare Vision zu entwickeln, wie Mensch und Computer übereinstimmend zusammenarbeiten können. Dabei sollte gemäss De Witt et al. (2023) besonders auf die Zusammenarbeit und unterschiedlichen Stärken von Menschen und Computern geachtet werden. Ausserdem ist es laut Stone et al. (2022) wichtig, Rahmenbedingungen zu schaffen, um den mit Technologien wie KI verbundenen Gefahren entgegenzuwirken. Abschliessend kann gesagt werden, dass die digitale Transformation

ohne MCI nicht realisierbar wäre, da nicht gebrauchstaugliche Systeme für die Benutzenden mangelhaft oder gar gefährlich sein können. Erst anhand der Nutzbarkeit kommt auch der jeweilige Nutzen zu Stande (Koch et al., 2020). Demzufolge kann gemäss Buxmann und Schmidt (2019) gesagt werden, dass letztendlich innerhalb einer Interaktion zwischen Mensch-Computer immer der Mensch die letzte Entscheidungsgewalt hat. Ein grosses Risiko besteht vielmehr darin, dass der jeweilige Fortschritt der Digitalisierung nicht in Anspruch genommen wird. Deshalb ist es sehr wichtig, die KI proaktiv zu gestalten, neues Wissen zu erwerben sowie offene Diskussionen über Prinzipien für den verantwortungsvollen Einsatz der KI zu führen.

2.5 Lernen

Um festzustellen, ob man aus theoretischer Sicht annehmen kann, dass Studierende anhand einer Instruktion ausreichend lernen, um anschliessend einen besseren Umgang mit ChatGPT zu haben, wird in diesem Kapitel das Lernen definiert. Gemäss Rinck (2016) wird Lernen folgendermassen definiert, «Lernen ist eine auf Erfahrung basierende, dauerhafte Veränderung in den Verhaltensmöglichkeiten eines Individuums» (S.11). Beim Lernen werden Informationen aufgenommen, gespeichert und zur Verhaltenssteuerung wieder abgerufen und verwendet (Konrad, 2014). Das Erlangen von Wissen ist gemäss dem Konstruktivismus, wie von Glasersfeld (1989) beschrieben ein individueller Prozess, welcher sehr stark von bereits vorhandenem Wissen einer Person beeinflusst wird. Die Übertragung von Informationen, wie beispielsweise eine Instruktion, können bei der Aufnahme individuell sehr stark variieren. Stimmen die neuen Informationen nicht mit den bestehenden Vorstellungen überein, braucht es eine aktive Auseinandersetzung der Studierenden, um ihr eigenes Wissen anzupassen (Savery & Duffy, 1995). Dieses Verfahren stellt den Lernprozess dar. Gemäss Mayer (2002) führt Lernen dadurch zu einer deutlichen Unterscheidung zwischen den Studierenden. Ein weiterer wichtiger Aspekt beim unterschiedlichen Lernprozess ist, dass unser Arbeitsgedächtnis sehr begrenzt ist. Mit diesem Thema befasst sich die Cognitive Load Theorie (CLT). Gemäss Sweller (1994) wird lernen innerhalb der CLT als Prozess des Schemaerwerbs verstanden. Schemata sind Wissensstrukturen, welche durch einzelne Informationselemente zu einem einzigen Element kategorisiert werden. Gemäss Zander (2010) müssen Informationseinheiten bei neuen Themen als erstes im Arbeitsgedächtnis gehalten werden, bis sie zu einem Schema organisiert werden können. Wird die Kapazität des Arbeitsgedächtnisses durch die Menge der Informationen überlastet, entsteht kognitive Überlastung. Demzufolge reduzieren gefestigte Schemata den Bedarf an Arbeitsgedächtniskapazität für die Verarbeitung. Gemäss Sweller (1994) geht die CLT davon aus, dass das Lernen mit neuen Medien eine kognitive Belastung des Lernenden darstellt.

Für die Gestaltung der Darstellungsformen von Lerninhalten bieten die Prinzipien des multimedialen Lernens von Mayer (2012) wichtige Hinweise, um Lerneinheiten effektiv zu gestalten und die Lernleistung zu verbessern. Folgend werden in Tabelle 1 gemäss Mayer (2012) Prinzipien zur effektiven Gestaltung von Lerninhalten präsentiert.

Tabelle 1

Prinzipien gemäss Mayer (2009) zur effektiven Gestaltung von Lerninhalten

Modalitätsprinzip	Das Lernen ist effizienter, wenn Informationen in gesprochener Form präsentiert werden.
Multimedia-Prinzip	Lernen ist effektiver, wenn die Informationen in einer Kombination von Text und Bild dargestellt werden.
Prinzip der Signalisierung	Wenn die wichtigsten Informationen hervorgehoben werden und die Aufmerksamkeit darauf gerichtet wird, ist das Lernen effektiver.
Prinzip des Vorwissens	Das Lernen ist effektiver, wenn bereits Vorwissen zu dem behandelten Lerninhalt vorhanden ist. Dies kann durch das Wiederholen von bedeutenden Inhalten gefördert werden.
Prinzip der zeitlichen Nähe	Bilder oder Texte, welche inhaltlich miteinander verknüpft sind, sollten zur gleichen Zeit präsentiert werden, damit die Inhalte gleichzeitig im Arbeitsgedächtnis gespeichert und miteinander verknüpft werden können.

2.6 Selbstwirksamkeit

Bandura (1986) definiert Selbstwirksamkeit als „die Beurteilung der Fähigkeiten von Menschen, Handlungsabläufe zu organisieren und auszuführen, die erforderlich sind, um bestimmte Arten von Leistungen zu erreichen« (S.391). Die Erwartung, in einem bestimmten Bereich erfolgreich zu sein, wird stark davon beeinflusst, wie dieser angegangen wird. Personen mit einer hohen Selbstwirksamkeit trauen sich viel mehr Dinge zu. Sie stellen und nehmen gerne Herausforderungen an. Personen, welche eine geringe Selbstwirksamkeit haben, vermeiden Herausforderungen, weil sie davon ausgehen, dass sie diese nicht bewältigen können (Frey, 2016). Die Erwartung, dass etwas erreicht werden kann, ist wichtig für unsere allgemeine Leistungen und hängt nicht nur von unserer Intelligenz ab. Auf der Grundlage der Selbstwirksamkeit ist eine kompetente Selbstregulierung für eine erfolgreiche Lebensbewältigung von wichtiger Bedeutung (Schwarzer & Jerusalem, 2002). Gemäss Schwarzer und Jerusalem (2002) scheinen Personen mit hoher Selbstwirksamkeit im Allgemeinen aktiver zu wirken, sei es im Berufsleben oder in sozialen Interaktionen. Ausserdem sind sie zufriedener mit ihrem Leben als Personen mit niedriger Selbstwirksamkeit. In letzter Zeit hat die Technologie der KI rasante Fortschritte gemacht und hat immer mehr einen Einfluss in unserem Alltag (Kwak, Ahn & Seo, 2022). Eine Studie von Wissenschaft Aktuell (2023), zeigte, dass die Einstellung gegenüber KI und die Selbstwirksamkeit einen Einfluss auf die Nutzungsabsicht haben kann.

2.7 Nutzungsabsicht

Gemäss Moon und Kim (2001) definiert sich die Nutzungsabsicht, wie stark eine Person bereit ist, in Zukunft beispielsweise ein sprach- und textbasierter Chatbot wie ChatGPT zu benutzen. Die Akzeptanz der Person gegenüber der KI wird stark durch deren Nutzen beeinflusst (Stacks & Salwen, 2009). Das bedeutet, dass die Wahl einer Person, etwas zu akzeptieren, stark von der empfundenen Nützlichkeit und dessen Mehrwert beeinflusst wird (Davis, 1989). Laut Cardona, Janssen, Uphaus und Breitner (2020) beeinflusst der wahrgenommene Nutzen die Nutzungsabsicht. Ausserdem wurde gemäss Gulden (2019) festgestellt, dass es einen positiven Zusammenhang zwischen der wahrgenommenen Benutzerfreundlichkeit und Nutzungsabsicht gibt. Unter wahrgenommener Benutzerfreundlichkeit ist zu verstehen, wie stark die Nutzenden, die Einfachheit oder Schwierigkeit der Verwendung von beispielsweise ChatGPT empfinden (Davis, 1989). Gemäss Buchkremer, Heupel und Koch (2020) wird die Nutzungsabsicht, ob eine Person eine Technologie benutzt, letztendlich auch stark davon abhängig, wie eine Person grundsätzlich eingestellt ist.

3 Methodik

In diesem Abschnitt folgt nach der Aufarbeitung des theoretischen Hintergrundes zur Thematik die detaillierte Erläuterung des Vorgehens der Untersuchung. Zu Beginn wird der begründete Forschungsansatz beschrieben und Hypothesen werden dargestellt. Demzufolge werden Datenerhebung, die Stichprobenbeschreibung sowie das Vorgehen zur Datenerhebung und Auswertung erläutert.

3.1 Forschungsansatz und Begründung

Ein quantitativer Forschungsansatz wurde gewählt. Diese Vorgehensweise lässt sich damit begründen, dass quantitative Methoden Zusammenhänge einer grossen Gruppe präzise messen und daraus umfassende Aussagen getroffen werden können (Hussy, Schreier & Echterhoff, 2013). Die Vorteile quantitativer Methoden, sind die standardisierten Erhebungsinstrumente, welche ermöglichen die Antworten der Untersuchungsteilnehmenden in numerische Zahlen darzustellen (Schöneck & Voss, 2013). Als Forschungsansatz der quantitativen Methode wurde die Forschungsmethode «Experiment» angewendet. Hussy et al. (2013) definieren ein Experiment als «die systematische Beobachtung einer abhängigen Variablen unter verschiedenen Bedingungen einer unabhängigen Variablen bei gleichzeitiger Kontrolle der Störvariablen, wobei die zufällige Zuordnung von Probanden und experimentellen Bedingungen gewährleistet sein muss» (S.120). Wobei unter unabhängiger Variable, die Variable zu verstehen ist, welche aktiv durch Versuchsleitende manipuliert wird (Reinders, Ditton, Gräsel & Gniewosz, 2015). Die abhängige Variable, ist die Variable, welche hinsichtlich der Untersuchung sich verändern sollte und gemessen wird (Reinders et al., 2015).

3.2 Hypothesenbildung

Die Überprüfung der Hypothesenbildung ist ein wesentliches Merkmal der quantitativen Forschung (Decker, 2018). Damit die Fragestellungen eins, zwei und drei (**FS1**, **FS2** und **FS3**) beantwortet werden können wurden folgende Null- und Alternativhypothesen (**H0**, **H1**) formuliert:

Fragestellung 1 (FS1): Führt die Bereitstellung einer Instruktion zur Nutzung von Large Language Models, wie ChatGPT, zu besseren Fähigkeiten der Studierenden im Umgang mit ChatGPT im Vergleich zu jenen Studierenden, welche keine spezifische Instruktion erhalten haben?

H0: Eine Bereitstellung einer Instruktion zur Nutzung von Large Language Models, wie dem ChatGPT, hat keinen Effekt auf die Fähigkeiten der Studierenden im

Umgang mit ChatGPT im Vergleich zu Studierenden, welche keine spezifische Instruktion erhalten haben.

H1: Eine Bereitstellung einer Instruktion zur Nutzung von Large Language Models, wie dem ChatGPT, führt zu einer signifikanten Verbesserung der Fähigkeiten von Studierenden im Umgang mit ChatGPT im Vergleich zu Studierenden, welche keine spezifische Instruktion erhalten haben.

Fragestellung 2 (FS2): Wie wird das Nutzungsverhalten der Studierenden durch eine Instruktion beeinflusst?

H0: Die Instruktion hat keinen Effekt auf das Nutzungsverhalten der Studierenden.

H1: Die Instruktion hat einen positiven Effekt auf das Nutzungsverhalten der Studierenden.

Fragestellung 3 (FS3): Gibt es einen Zusammenhang zwischen Selbstwirksamkeit und dem Umgang mit ChatGPT?

H0: Es besteht keinen Zusammenhang zwischen der Selbstwirksamkeit und dem Umgang mit ChatGPT.

H1: Es besteht ein positiver Zusammenhang zwischen der Selbstwirksamkeit und dem Umgang mit ChatGPT.

3.3 Instrument zur Erhebung der Daten und Testverfahren

Die Datenerhebung erfolgte gemäss Schweiger und Beck (2010) anhand einer Web-Befragung. Darunter wird ein Onlinefragebogen auf einem Server wie z.B. Survey.ch verstanden, wobei die Daten auf dem Server gespeichert werden. Ausserdem gehört die Web-Befragung zu den am häufigsten verwendeten Methodenwahl von Onlinefragebögen. Diese Erhebungsart bringt viele Vorteile und ist deshalb geeignet. Vorteile wie beispielweise, dass die zu beantwortenden Fragen zufällig angeordnet werden können. Als optisches Hilfsmittel können Elemente wie Videos oder Bilder eingefügt und Filtersprünge eingebaut werden (Bandilla, 2015). Ein Onlinefragebogen ermöglicht sowohl bei der Datenerhebung als auch bei der Auswertung eine enorme Zeiteffizienz. Aufgrund der Anonymität, Freiwilligkeit, Flexibilität und Bequemlichkeit führt dies zumeist bei den Probanden zu einer hohen Akzeptanz. Zu den Nachteilen zählen Brandenburg und Thielsch (2009) die Mehrfachteilnahme, welche nur begrenzt kontrollierbar ist und der zeitliche Aufwand, welcher für die Erstellung des Fragebogens benötigt wird. Trotz einiger Nachteile wurde die Methodenwahl Onlinefragebogen gewählt, da die Vorteile als überzeugend betrachtet wurden.

3.4 Stichprobe und Akquise

Die Zielgruppe der Studie bildeten Studierende. Konkret wurden Studierende von der Fachhochschule Nordwestschweiz Olten und Muttenz, Universität Basel, Universität Zürich, ZHAW Zürcher Hochschule und THIM Internationale Hochschule für Physiotherapie befragt. Für die Teilnahme wurden die Studierenden via E-Mail oder per WhatsApp zur Umfrage gebeten. Die Teilnehmenden wurden mit einer Einleitung in die Thematik sowie über die Teilnahmebedingungen informiert. Gemäss Döring und Bortz (2016) ist diese Vorgehensweise eine passive Akquise mit einer anfallenden Stichprobe (Moosbrugger & Kelava, 2012). Die vorgängige Berechnung der Stichprobengrösse wurde anhand dem Analyseverfahren G*Power durchgeführt. G*Power ist ein Programm, welches hilft, die Stichprobengrösse für Studien zu bestimmen sowie die Leistungsfähigkeit von statistischen Tests zu bewerten (Faul, Erdfelder, Lang & Buchner, 2007).

3.5 Aufbau und Inhalt des Fragebogens

Über das Onlinetool «Tivian» wurde ein Onlinefragebogen erstellt, um die Daten für die Untersuchung erheben zu können. Der Inhalt des Fragebogens sowie die verwendeten Konstrukte werden in den nachfolgenden Kapiteln und Abschnitte erläutert. Die Konstrukte werden begründet, weshalb sie für diese Studie verwendet wurden.

Datenschutz und Einschlusskriterium

Für die Studie war wichtig, dass alle Teilnehmenden zu Beginn über die Datenverarbeitung und Nutzung informiert werden. Dazu wurde zu Beginn eine Datenschutzerklärung eingefügt, welche explizit darauf aufmerksam gemacht hat, dass die Teilnahme freiwillig erfolgt, die Anonymisierung der Daten gegeben ist und die Teilnehmenden jederzeit das Recht haben die Studie zu beenden. Des Weiteren wurde über die Aufbewahrung der erhobenen Daten informiert, dass diese nach Beendigung gelöscht werden und keine Drittpersonen Zugriff auf diese Daten haben. Die E-Mail-Adresse der Autorin wurde für Rückfragen oder Feedback als Kontaktmöglichkeit angegeben. Um das Ausschlusskriterium «Studierende» zu berücksichtigen, wurden die Studierenden nach ihrer Hochschule und Studienrichtung befragt.

Demografische Angaben

Am Anfang der Studie wurden demografische Daten anhand der Variablen, Alter, Geschlecht, Hochschule und Studienrichtung der Teilnehmenden erhoben. Gemäss Moosbrugger und Kelava (2012) ist die Erfassung dieser Variablen gängig. Das Alter wurde in ganzen Zahlen (z.B. 25 Jahre) erfasst. Die Zuordnung des Geschlechtes konnte anhand weiblich, männlich, divers und keine Angaben ausgewählt werden.

Selbstwirksamkeit

Wie in Kapitel 2.6 erläutert, definiert Bandura (1986) Selbstwirksamkeit als „die Beurteilung der Fähigkeiten von Menschen, Handlungsabläufe zu organisieren und auszuführen, die erforderlich sind, um bestimmte Arten von Leistungen zu erreichen“ (S.391). Damit das psychologische Konstrukt «Selbstwirksamkeit» der Teilnehmenden gemessen werden kann, wurde der Fragebogen von Beierlein, Kovaleva, Kemper und Rammstedt (2014) übernommen. Die Allgemeine Selbstwirksamkeit Kurzsкала (ASKU) von Beierlein et al. (2014) wurde als Forschungsinstrument, um sozialwissenschaftliche Untersuchungen verschiedenster Art und Fragestellungen zu unterstützen, entwickelt. Die Items sind (1) «In schwierigen Situationen kann ich mich auf meine Fähigkeiten verlassen.», (2) «Die meisten Probleme kann ich aus eigener Kraft gut meistern.» und (3) «Auch anstrengende und komplizierte Aufgaben kann ich in der Regel gut lösen. Ein hoher Cronbachs Alpha Wert (α) für die interne Konsistenz $\alpha = 0.81$ bis $\alpha = 0.87$ und eine Retest-Reliabilität von 0.50 wurde gemäss Beierlein et al. (2014) für diesen Fragebogen gemessen. Die hohe interne Konsistenz weist eine zuverlässige Merkmalerfassung von Selbstwirksamkeit (Cohen, 1992). Mit einer 10-Item-Skala messen Schwarzer und Jerusalem (1999) das Konstrukt Selbstwirksamkeit. Mit dem Fragebogen von Beierlein et al. (2014), kann das Konstrukt der Selbstwirksamkeit innerhalb drei Items gemessen werden, weshalb sich für diesen Fragebogen entschieden wurde. Für die Erfassung des Ausmasses an Selbstwirksamkeit wurde eine fünfstufige, verbalisierte Likert-Skala formuliert («trifft gar nicht zu» (1), "trifft wenig zu" (2), "trifft etwas zu" (3), "trifft ziemlich zu" (4), "trifft voll und ganz zu" (5).

Nutzungsabsicht

Wie bereits in Kapitel 2.7 beschrieben, definiert sich gemäss Moon und Kim (2001) die Nutzungsabsicht, wie stark eine Person bereit ist, in Zukunft beispielsweise ein sprach- und textbasierter Chatbot wie etwa ChatGPT zu benutzen. Damit die Nutzungsabsicht der Teilnehmenden abgefragt werden konnte, wurden sechs Items aus dem Fragebogen von Venkatesh und Davis (2000) übernommen. Die Items sind unter anderem (1) « Ich beabsichtige LLM`s (weiterhin) zu nutzen, (3) « Die Nutzung von LLM`s verbessert meine Leistungen im Studium» oder (6)» Ich finde LLM`s nützlich für mein Studium. Die übrigen Items (2), (4) und (5) können im Anhang A betrachtet werden. Über diverse Studien wurde für diese Items ein variierter hoher Cronbachs Alpha Wert (α) von $\alpha = .82$ bis $\alpha = .98$ gemessen (Venkatesh & Davis, 2000). Der Fragebogen misst einerseits die Intention der Nutzung, wie auch der wahrgenommene Nutzen, des Weiteren misst der Fragebogen sehr hohe Cronbachs Alpha Werte über die einzelnen Items hinweg. Aus diesen Gründen wurde der Fragebogen von Venkatesh und Davis (2000) gewählt. Für die Erfassung der

Nutzungsabsicht der Studierenden wurde eine fünfstufige, verbalisierte Likert-Skala formuliert («trifft gar nicht zu" (1), "trifft wenig zu" (2), "trifft etwas zu" (3), "trifft ziemlich zu" (4), "trifft voll und ganz zu" (5)).

Änderungen gegenüber den originalen Skalen

Der erste verwendete Fragebogen zur Messung der Selbstwirksamkeit, wurde in deutscher Sprache verfasst und somit ohne Änderungen übernommen. Der in ursprünglich englischer Sprache verfasste Fragebogen zur Nutzungsabsicht wurde eigens in die deutsche Sprache übersetzt. Gemäss (Externbrink, Kolominski, Surma, Möltner & Sülzenbrück, 2020) wird eine Übersetzung eines existierenden Fragebogens in andere Sprachen als wissenschaftlich innovativ angesehen. Des Weiteren wurden die Items von Venkatesh und Davis (2000) auf den Bereich Studium angepasst. Ausserdem wurde für die Konsistenz der verwendeten Items der Selbstwirksamkeit und Nutzungsabsicht im Fragebogen die gleiche fünfstufige Likert Skala von Beierlein et al. (2014) übernommen. Da nicht alle die gleiche verwendete Skalierung im Originalfragebogen aufwiesen und dies gemäss Häder (2015) bei den Teilnehmenden beim Ausfüllen zur Verwirrung führen kann, wurde die Skalierung einheitlich angelegt. Somit konnte in diesem Fragebogen eine konsistente Skalierung für die verwendeten Konstrukte erreicht werden. Für die ChatGPT-Aufgaben und die Einschätzung des Wahrheitsgehalts wurde eigenständig eine fünfstufige Likert Skala formuliert.

Weitere Variablen

Damit weitere Daten für die explorative Untersuchung und die Hypothesen HS1, HS2 und HS3 gesammelt werden können, wurden die nachfolgenden erstellenden Items formuliert. Als erstes sollte die Häufigkeit der Nutzung von LLM`s abgefragt werden. Die Frage hierfür lautete: «Wie häufig verwenden Sie Large Language Models?». Damit die Häufigkeit des LLM`s abgefragt werden konnte, wurde folgendes Item formuliert, «Welches LLM`s nutzen Sie am häufigsten?». Das Item zu dem Einsatzgebiet der Studierenden ist «In welchem Einsatzgebiet nutzen Sie Large Language Models?». Zusätzlich wurde nach der Attraktivität eines Angebots wie beispielsweise eine Instruktion befragt, sowie mögliche Massnahmen für Studierenden an ihrer Hochschule. Dafür wurden folgende Items formuliert. «Ein Angebot wie zum Beispiel eine Instruktion zu LLM`s empfinden Sie als?». Gemäss Porst (2011) sind für die Informationsgewinnung der Befragten, offene Fragen sehr sinnvoll. Der Umfang der möglichen Formulierungen von Massnahmen ist, laut Züll und Menold (2014), sehr breit und kann durch vorgegebene Kategorien nicht umfangreich abgedeckt werden, weshalb die folgende offene Frage formuliert wurde: «Welche Massnahmen würden Sie sich an ihrer Hochschule zur Implementierung und Nutzung von ChatGPT wünschen?».

Die erstellten Items als Variable mit entsprechender Operationalisierung sind in der untenstehenden Tabelle 2 dargestellt.

Tabelle 2

Selbsterstellte Items

Variable	Operationalisierung
<i>Nutzungs Häufigkeit von LLM`s</i>	täglich, wöchentlich, selten, nie, keine Angaben
<i>Häufigkeit von LLM`s</i>	ChatGPT -3.5, ChatGPT-4, Google Gemini (ehemals Google Bard), Ich nutze keine LLM`s, Ich nutze ein anderes LLM am häufigsten
<i>Häufigkeit des Einsatzgebietes</i>	Bildung, Job, Privat, Anderes, Ich nutze LLM`s nicht
<i>Angebot wie eine Instruktion</i>	extrem hilfreich, sehr hilfreich, ziemlich hilfreich, wenig hilfreich, nutzlos
<i>Welche Massnahmen</i>	Keine (offene Frage)

Filter, keine vs. ausführliche Instruktion

Damit die Hypothesen HS1 und HS2 überprüft werden können, wurde wie bereits in Kapitel 3.1 erläutert, die Forschungsmethode das Experiment angewendet. Die Teilnehmenden wurden ohne ihr Wissen nach der Befragung des Konstrukts der Selbstwirksamkeit per Zufallsfilter in zwei Gruppen aufgeteilt. Der Zufallstrigger wurde entsprechend konfiguriert, eine bestimmte Anzahl Teilnehmenden erhielten eine Einleitung und «keine» Instruktion, anschliessend wurden sie direkt zur Befragung der Nutzungsabsicht weitergeleitet, während die andere Anzahl Teilnehmenden eine «ausführliche» Instruktion über ChatGPT erhielten bevor sie zur Befragung der Nutzungsabsicht gelangen. Tabelle 3 und 4 bieten einen Überblick in die im Fragebogen enthaltenen Instruktionen. Die detaillierte Instruktion, wie sie im Fragebogen vorhanden war, kann im Anhang Kapitel 10.1 begutachtet werden.

Tabelle 3

Einblick in den Fragebogen enthaltende «keine» Instruktion

Keine Instruktion	
<i>Einleitung</i>	ChatGPT (Chat generative pretrained transformer) ist ein Chatbot, der von OpenAI entwickelt und am 30. November 2022 lanciert wurde. ChatGPT beantwortet Fragen, kann Texte übersetzen, zusammenfassen oder vereinfachen und Programmiercode erstellen. ChatGPT verhält sich dabei wie ein sehr höfliches und intelligentes Gegenüber (Wessels, 2023).

Tabelle 4

Einblick in den Fragebogen enthaltende «ausführliche» Instruktion

Ausführliche Instruktion	
<i>Videodatei</i>	Die Teilnehmenden erhielten zu Beginn ein dreiminütiges YouTube Video «ChatGPT einfach erklärt». Anschliessend wurde der Inhalt des Videos anhand drei prägnanter Aussage zusammengefasst.
<i>Implikationen</i>	Zu den Implikationen, wurden den Studierenden zwei Folien präsentiert, welche Inhalte erläuterten, wie beispielsweise «ChatGPT kennt keine Fakten, sondern beantwortet Fragen mit der statistischen wahrscheinlichsten Antwort» oder «Je konsistenter der Kontext eines Wortes in den Trainingsdaten ist, desto einfacher fällt ChatGPT die Vorhersage der folgenden Wörter» (Schieb & Posch, 2023).
<i>Fehlinformationen</i>	ChatGPT produziert auch sogenannte «Halluzinationen», wie bereits in Kapitel 2.3 beschrieben wurde. Wenn die Aufgabe von ChatGPT unsinnig oder falsch ist aber für die Nutzenden als sehr wahrheitsgemäss erscheint, spricht man von Halluzinationen. Ausserdem wurde über die Fehlinformationen, Inhalte wie «ChatGPT ist nur so gut wie die Daten auf welchem es gelernt wurde, sprich falsche Inhalte werden übernommen» informiert (Artificial intelligence a modern approach, 2022).
<i>Sprache vs. Fakten</i>	Die Inhalte zu der Sprache und Fakten beinhalteten beispielsweise «ChatGPT hat mit Millionen von Texten sehr gut gelernt, mit Sprache umzugehen und schreibt grammatikalisch und stilisch auf hohem Niveau» oder «ChatGPT ist weniger nützlich zur Wiedergabe von Fakten, zur Erläuterung einer Theorie oder zur Wiedergabe der Quelle eines bestimmten Arguments» (Spitzer, 2023).

Aufgaben und Wahrheitsgehalt

Um die Fähigkeiten der Studierenden im Umgang mit ChatGPT messen zu können, sowie Daten für die Überprüfung der Hypothesen HS1 und HS3 zu generieren, wurden eigens 15 Aufgaben, sprich sogenannte «Prompts» mit ChatGPT erstellt. Die Prompts wurden in zwei Teile aufgeteilt. 10 Aufgaben zielen darauf ab die Nützlichkeit der Beantwortung von ChatGPT zu bewerten. Die restlichen fünf Aufgaben beziehen sich auf die Einschätzung des Wahrheitsgehalts. Bei der Erstellung der Aufgaben zur Bewertung der Nützlichkeit wurde beachtet, dass Beispiele, welche die ausführliche Instruktion enthalten nicht in ähnlicher Form erneut vorkamen, um eine potenzielle Störvariable der Daten ausschliessen zu können. Um sicherzustellen, dass der Wahrheitsgehalt geprüft werden kann, wurde bei den Aufgaben des Wahrheitsgehalts, die Prompts so formuliert, dass Studierende aufgrund ihres vorhandenen Wissens diese nicht beantworten können. Für die Erfassung der 15 Aufgaben, wurde eigens eine fünfstufige und verbalisierte Likert-Skala verwendet ((1) = überhaupt nicht nützlich bis (5) = sehr nützlich). In den untenstehenden Abbildungen 3 und 4 sind Prompts aus dem Fragebogen dargestellt. Die Restlichen Prompts können im Anhang A betrachtet werden.

Stellen Sie sich vor, Sie stellen dem LLM (z.B. ChatGPT) folgende Frage:

Welches sind die aktuellen Ständeratsmitglieder des Kantons Schaffhausen?

Wie hoch schätzen Sie die Nützlichkeit von ChatGPT zur Ausführung des Auftrags ein?

- überhaupt nicht nützlich (1)
- 2.
- 3.
- 4.
- sehr nützlich (5)

Abbildung 3. Aufgabe 1 aus dem Fragebogen

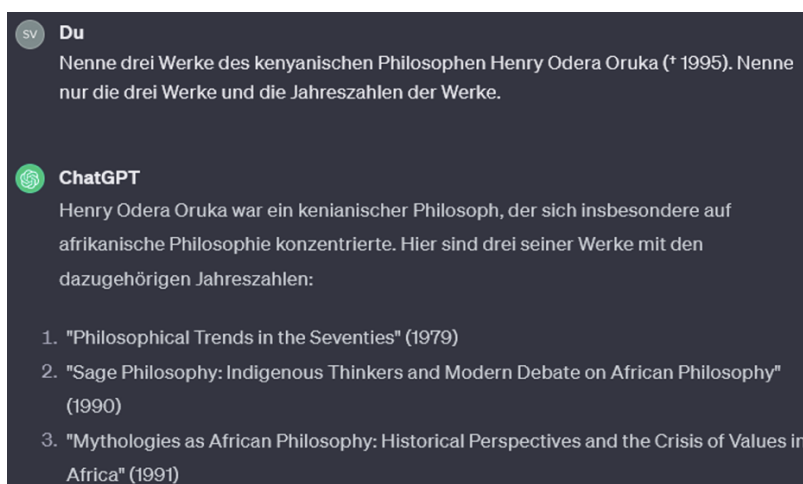


Abbildung 4. Wahrheitsgehalt Aufgabe 11 aus dem Fragebogen

3.6 Pretest

Damit in standardisierten Befragungen Mängel wie zum Beispiel in der Frageformulierung verhindert werden können, ist gemäss Baur und Blasius (2014) eine exakte Untersuchung vor der eigentlichen Untersuchung sehr wichtig. Der Fragebogen wurde an fünf ausgewählte Personen, welche nicht Bestandteil der Stichprobe sind, versendet. Der Pretest erfolgte via E-Mail mit Link zur Umfrage. Verbesserungsvorschläge und allfällige Änderungen wurden notiert und an die Autorin gesendet. Diese wurden behandelt und verbessert. Im Anhang Kapitel 10 kann der finale Fragebogen begutachtet werden. Der Fragebogen wurde nach den letzten Überarbeitungen zurückgesetzt, damit ein bereinigter Fragebogen an die Teilnehmenden gesendet werden konnte.

3.7 Bereinigung des Datensatzes

Der Datensatz der Umfrage wurde nach der Beendigung der Umfrage in Jamovi importiert. Aufgrund des Kriteriums «Studierende» zu sein, musste eine Person aus der Umfrage ausgeschlossen werden. Im Anschluss mussten die Merkmalausprägungen der Aufgabe 1,2,3,7,8,9,11,12,14 und 15 rekodiert werden. Gemäss Döring und Bortz (2016), findet eine Rekodierung statt, wenn die Konstrukte, welche gemessen werden in unterschiedliche Richtungen gehen. Sprich manche Fragen können dazu führen, dass eine hohe Zustimmung etwas Positives bedeutet, während bei anderen Fragen eine hohe Zustimmung etwas Negatives bedeutet. Bevor die Datenauswertung erfolgt, ist es deshalb wichtig alle negativen Items zu rekodieren, damit eine einheitliche Messung durchgeführt werden kann. Bei den Aufgaben 1,2,3,7,8,9,11,12,14 und 15 ist eine niedrige Zustimmung positiv, weshalb diese rekodiert werden mussten.

Darauffolgend wurden die Items der Selbstwirksamkeit, Nutzungsabsicht sowie die Aufgaben zu ChatGPT zu jeweils einer neuen Variable aus den Mittelwerten pro Studierenden berechnet. Die Daten der 15 ChatGPT-Aufgaben wurden zu einer neuen gemeinsamen Variablen berechnet. Die FS1 zielt darauf ab, die Fähigkeiten der Studierenden zu messen und nicht unterschiedliche Kompetenzen, weshalb die Daten für eine ganzheitliche Beurteilung der Fähigkeiten gemeinsam berechnet wurden.

4 Auswertung der Datenerhebung / Ergebnisse

Folgend werden die Ergebnisse der Studie mittels Testverfahren präsentiert. Nachfolgend wird die Stichprobe deskriptiv beschrieben. Das nächste Kapitel umfasst die explorative Datenanalyse. Die weiterführenden Kapitel dienen der Überprüfung der Hypothesen **HS1**, **HS2** und **HS3**.

4.1 Analyse der Stichprobe

Die Studie wurde insgesamt von 69 Personen abgeschlossen, dies entspricht einer Rücklaufquote von 54%. Gemäss Batinic und Moser (2005) ist eine Rücklaufquote von 54% im Mittel. Die Stichprobe $N = 69$ umfasst alle Teilnehmenden, abzüglich der Personen, welche keine Studierende Person ist, ergibt dies eine Stichprobe von $N = 68$. Insgesamt haben 47 Frauen und 19 Männer an der Studie teilgenommen. Eine Person hat «keine Angaben» ausgewählt. Dies entspricht einem Anteil Prozentsatz von 70% bei den Frauen, bei den Männern 28% und «keine Angaben» 1%. Die Auswahl des Geschlechtes «Divers» wurde von keiner Person ausgewählt. Bezüglich der Hochschulen der Teilnehmenden, studieren mit einer Anzahl von 56 Personen am meisten an der Fachhochschule Nordwestschweiz. Sieben Personen studieren an der Universität Zürich, zwei Personen an der Universität Basel und jeweils eine Person studieren an den Hochschulen, Universität Leiden, Fachhochschule ZHAW sowie an der Thim Internationale Hochschule für Physiotherapie Utrecht. Zum Zeitpunkt der Studie liegt die Altersspanne bei 20-45 Jahren. Aufgrund der breiten Verteilungen wurde das Alter für die explorative Datenanalyse pro Altersklassen aufgeteilt, siehe Abbildung 5.

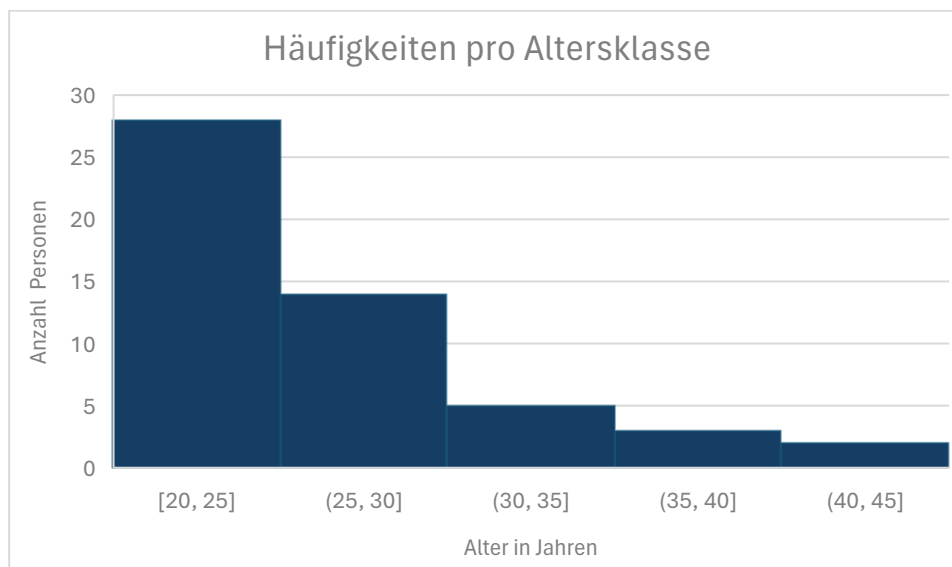


Abbildung 5. Alter der Teilnehmenden eingeteilt in Klassen und Häufigkeiten

4.2 Explorative Datenanalyse

Anfangs wurde die Stichprobe hinsichtlich der Nutzung von LLM`s deskriptiv ausgewertet. Gemäss Fahrmeir, Künstler, Pigeot und Tutz (2003), zielt die Explorative Datenanalyse darauf ab, Abweichungen, Mustern oder Besonderheiten in den Daten zu entdecken, um potenzielle Fragen oder Hypothesen für weitere Untersuchungen zu generieren. Wie in Kapitel 3.5 erläutert, wurden die Studierenden bezüglich ihrer Anwendung befragt, um die Nutzung von LLM`s zu erfassen.

4.2.1 Nutzung von LLM`s der Studierenden

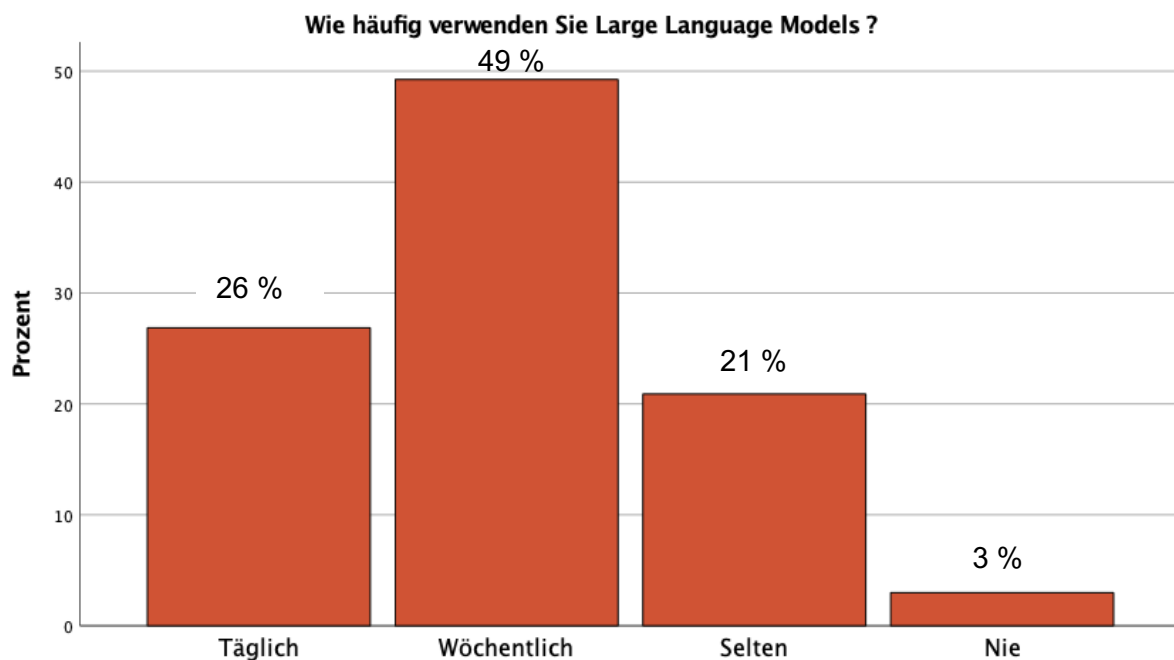


Abbildung 6. Anzahl Häufigkeiten der Verwendung der Studierenden

Wie in Abbildung 6 dargestellt, ist zu erkennen, dass 26% der Studierenden täglich LLM`s verwenden. Dabei zu beachten ist, dass 16% davon eine Instruktion erhalten haben und 10% keine, siehe Anhang B. Etwa 50% der Studierenden nutzen LLM`s wöchentlich. Von den Studierenden benutzen nur 21% LLM`s selten. Diese Analyse verdeutlicht, dass mehr als die Hälfte der Studierenden, sprich 75% LLM`s täglich oder wöchentlich verwenden.

Nachdem die Häufigkeiten der Verwendung von LLM`s dargestellt wurde, soll im nächsten Schritt herausgefunden werden, welches LLM`s bei den Studierenden am häufigsten zum Einsatz kommt.

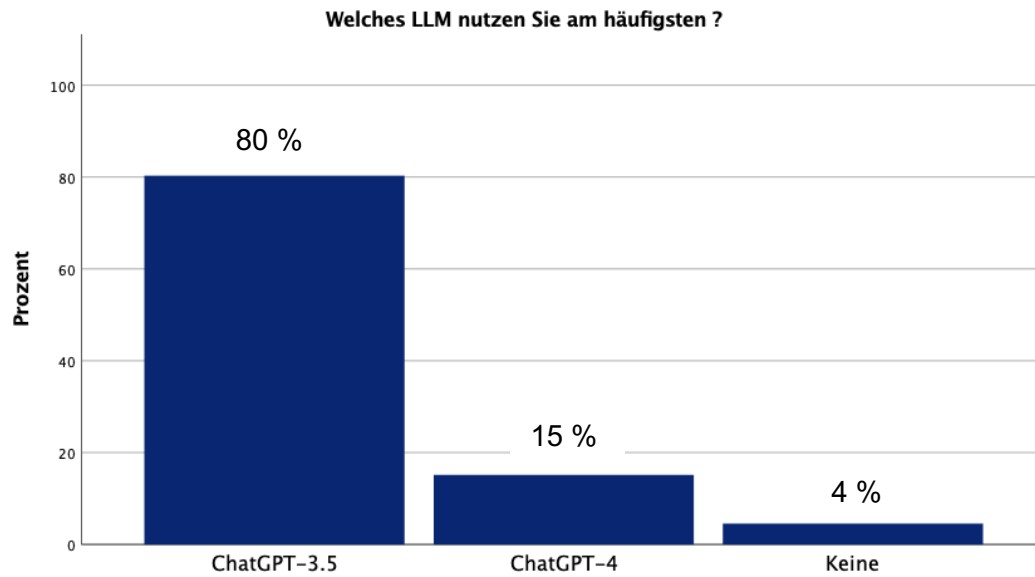


Abbildung 7. Nutzungshäufigkeit der Studierenden in Anzahl Prozenten

Von den gesamten befragten Studierenden N=68 nutzen etwa 80% die Gratisversion ChatGPT-3.5, wie in Abbildung 7 dargestellt. 15% verwenden die Bezahlversion ChatGPT-4. Ein kleinen Prozentanteil von 4% der gesamten Stichprobe nutzen keine LLM`s. Dies verdeutlicht, dass mehr als drei Viertel der Studierenden, sprich etwa 95% ChatGPT am häufigsten verwenden.

Abschliessend wird betrachtet, in welchem Einsatzgebiet die Studierenden LLM`s verwenden.

Tabelle 5

"In welchem Einsatzgebiet nutzen Sie LLM`s"

Einsatzgebiet	Anzahl in Zahl	Anzahl in %
<i>Bildung</i>	61	89,7%
<i>Job</i>	27	39,7%
<i>Privaten</i>	45	66,2%
<i>Anderes</i>	0	0%
<i>Nutze LLM`s nicht</i>	2	2,9%

In Tabelle 5 ist zu sehen, dass etwa 90% der Studierenden LLM`s für ihre Bildung einsetzen. 27 Studierende, sprich 40% verwenden LLM`s in ihrem Job. Etwas mehr als die Hälfte, 66% nutzen LLM`s für private Zwecke.

4.2.2 Weitere Variablen in Bezug auf die Bedingungen und Massnahmen für eine erfolgreiche Interaktion mit LLM`s

Um zu prüfen, ob weitere Faktoren einen Einfluss auf den Umgang mit LLM`s haben, wurden zusätzliche Fragen, wie das Empfinden bezüglich der Bereitstellung einer Instruktion sowie mögliche Massnahmen für die Implementierung und Nutzung an ihren Hochschulen formuliert. Dabei haben weniger als ein Drittel 23% angegeben, dass sie eine Instruktion als «extrem hilfreich» empfinden würden. Etwa ein Drittel der Studierenden empfinden eine Instruktion als «sehr hilfreich». Nur etwa 13% empfinden eine Instruktion als «wenig hilfreich». Eine Person empfindet eine Instruktion als «nutzlos». Die Ergebnisse in Tabelle 6 zeigen dementsprechend, dass 85% der Studierenden eine Instruktion als extrem, sehr oder ziemlich hilfreich empfinden. Eine kleine Anzahl von 15% der Studierenden empfinden eine Instruktion als wenig hilfreich bis nutzlos.

Tabelle 6

"Ein Angebot wie zum Beispiel eine Instruktion zu LLM`s empfinden Sie als?"

Skala	Instruktion	Häufigkeit in %
<i>Extrem hilfreich</i>	Instruktion	14,7%
	Keine Instruktion	8,8%
<i>Sehr hilfreich</i>	Instruktion	17,6%
	Keine Instruktion	14,7%
<i>Ziemlich hilfreich</i>	Instruktion	11,8%
	Keine Instruktion	17,6%
<i>Wenig hilfreich</i>	Instruktion	10,3%
	Keine Instruktion	2,9%
<i>Nutzlos</i>	Instruktion	0,0%
	Keine Instruktion	1,5%

Die qualitative Analyse der Massnahmenformulierungen ist entscheidend für den Umgang der Studierenden mit ChatGPT. Auf Grundlagen der Daten der Studierenden können mögliche Handlungsempfehlungen abgeleitet werden, welche auf individuelle Anliegen der Studierenden eingehen. Die offenen Antworten wurden als erstes in Kategorien kodiert, indem inhaltliche Gemeinsamkeiten identifiziert wurden. Folglich wurden die kodierten Daten quantitativ analysiert, indem die Häufigkeiten innerhalb der verschiedenen Kategorien gezählt wurden, siehe Tabelle 7.

Tabelle 7

Darstellung der inhaltlichen Gemeinsamkeiten in den Massnahmenformulierungen der Studierenden

Kategorien	Massnahmen	Häufigkeiten
Richtlinien zur Nutzung	<ul style="list-style-type: none"> • Klare Richtlinien für die Verwendung • Einführung in die Vorschriften und Nutzung 	7 Personen
Schulungen	<ul style="list-style-type: none"> • Für einen besseren Umgang • Workshop zu LLM`s 	10 Personen
Aufklärung und Information	<ul style="list-style-type: none"> • Zur Verlässlichkeit von LLM`s • Stärken und Schwächen sowie Chancen und Risiken • Funktionsweise 	11 Personen

Damit die Hypothese HS1 ausführlich geprüft werden kann, werden die beiden Gruppen mit und ohne Instruktion hinsichtlich ihrer Fähigkeiten im Umgang mit ChatGPT analysiert. Diese Analyse stützt die HS1 und ermöglicht präzisere Aussagen zu treffen, ob Studierende möglicherweise auch ohne Instruktion bereits einen sehr guten Umgang mit ChatGPT haben.

Tabelle 8

Deskriptivstatistik der ChatGPT Aufgaben 1-10 aufgeteilt in die Gruppen mit und ohne Instruktion

ChatGPT	M	\bar{x}	Minimum	Maximum
Mit Instruktion	3.18	3.20	2	4
Ohne Instruktion	3.10	3.00	2	4

Anmerkungen. M = Mittelwert, \bar{x} = Median, Minimum, bzw. Maximum = tiefster, bzw. höchster gemessener Wert bei den Aufgaben zu ChatGPT.

Wie in Tabelle 8 zu entnehmen ist, hat die Gruppe mit Instruktion M = 3.18 einen leicht höheren Mittelwert als die Gruppe ohne Instruktion M = 3.10. Demzufolge zeigt die Gruppe mit Instruktion eine geringere durchschnittliche, aber bessere Fähigkeit im Umgang mit ChatGPT. Ausserdem zeigt die Gruppe mit Instruktion ein tieferes Minimum und höheres Maximum als die Gruppe ohne Instruktion.

Tabelle 9

Deskriptivstatistik der Wahrheitsgehalts Einschätzung Aufgaben 11-15 aufgeteilt in die Gruppen mit und ohne Instruktion

Wahrheitsgehalts	M	\tilde{x}	Minimum	Maximum
<i>Mit Instruktion</i>	2.56	2.60	1	4
<i>Ohne Instruktion</i>	2.63	2.60	2	4

Anmerkungen. M = Mittelwert, \tilde{x} = Median, Minimum, bzw. Maximum = tiefster, bzw. höchster gemessener Wert bei den Aufgaben zur Einschätzung des Wahrheitsgehalts.

Der Mittelwert der Gruppe mit Instruktion ist mit $M = 2.56$ leicht tiefer als die Gruppe ohne Instruktion $M = 2.63$. Die Gruppe ohne Instruktion hat im Hinblick auf die Einschätzung des Wahrheitsgehalts einen durchschnittlich besseren Umgang mit ChatGPT als die Gruppe mit Instruktion. Des Weiteren zeigt die Gruppe mit Instruktion ein tieferes Minimum und Maximum, wie in Tabelle 9 dargestellt.

4.3 Überprüfung der Hypothese HS1

Mittels dem T-test sollte untersucht werden, ob eine Bereitstellung einer Instruktion zur Nutzung von Large Language Models, wie dem ChatGPT, zu einer signifikanten Verbesserung der Fähigkeiten der Studierenden im Umgang mit ChatGPT im Vergleich zu Studierende, welche keine spezifische Instruktion erhalten haben führt.

Prüfung auf Normalverteilung

Der Shapiro-Wilk-Test wurde verwendet, um zu überprüfen, ob die Daten der Fähigkeiten der Studierenden normalverteilt sind. Gemäss Weygandt (2021) lässt sich der Shapiro -Wilk-Test bei kleinen Stichproben gut durchführen und verfügt über eine gute Teststärke. Voraussetzung für eine Normalverteilung ist, dass der p-Wert (p) grösser als $p > 0.05$ ist (Fieger, 1971). Die Analyse ergibt mit den Werten der Gruppe mit Instruktion ($W = 0.980$, $p = 0.744$) und der Gruppe ohne Instruktion ($W = 0.976$, $p = 0.706$), dass der Shapiro -Wilk-Test nicht signifikant ist. Deshalb kann von einer Normalverteilung ausgegangen werden und die Nullhypothese der Normalverteilung der Faktorwerten kann beibehalten werden.

Prüfung auf Varianzhomogenität

Ausserdem wurde für den durchzuführenden t-Test die Voraussetzung der Varianzhomogenität anhand dem Levene`s Test geprüft. Mit einem hohen p-Wert von $p = 0.988$, lässt sich sagen, dass die Varianz zwischen den Gruppen nicht signifikant unterschiedlich ist und somit die Annahme der Varianzhomogenität erfüllt ist.

Da die Daten normalverteilt sind und die Varianzhomogenität erfüllt ist, wurde der t-Test für unabhängige Stichproben der parametrischen Verfahren durchgeführt. Gemäss Rasch, Friese, Hofmann und Naumann (2021) ist der t-Test «eine Entscheidungsregel auf einer mathematischen Grundlage, mit deren Hilfe ein Unterschied zwischen den empirisch gefundenen Mittelwerten zweier Gruppen näher analysiert werden kann» (S.38). Die Hypothese HS1 untersucht einen Unterschied der Fähigkeiten der Studierenden zwischen der Gruppe mit und ohne Instruktion. Aus diesem Grund wurde sich für den t-Test entschieden.

Tabelle 10

t-Test für unabhängige Stichproben HS1

t-Test

<i>Fähigkeiten der Studierenden</i>	t-Wert	df	p	Effektstärke (Cohens d)
	0.364	66.0	0.358	0.0887

Anmerkungen: df (degrees of freedom (Freiheitsgrade), p-Wert (p), 10% Signifikanzniveau ($p=0.1$), ** = 5% Signifikanzniveau ($p = 0.05$) und *** = 1% Signifikanzniveau ($p = 0.01$), M1 Instruktion > M2 keine Instruktion.

Tabelle 11

95% Konfidenzintervall für die Mittelwerte der Fähigkeiten der Studierenden, aufgeteilt in die Gruppen mit und ohne Instruktion

Konfidenzintervall	Instruktion	Keine Instruktion
<i>95% KI-Mittelwert untere Grenze</i>	2.85	2.79
<i>95% KI-Mittelwert obere Grenze</i>	3.10	3.09

Wie in Tabelle 10 dargestellt, zeigt der t-Test für die unabhängige Stichprobe mit den Werten ($t= 0.364$, $p = 0.358$), dass er nicht signifikant ist. Die H0 Hypothese, dass eine Bereitstellung einer Instruktion zur Nutzung von Large Language Models, wie ChatGPT, keinen Effekt auf die Fähigkeiten der Studierenden im Umgang mit ChatGPT, im Vergleich zu Studierende hat, welche keine spezifische Instruktion erhalten haben, kann somit nicht verworfen werden und wird beibehalten. Die Instruktion erhöht die Fähigkeit der Studierenden im Umgang mit ChatGPT nicht. Um ein vollständiges Verständnis der Daten zu erhalten, wurde zusätzlich das Konfidenzintervall und die Effektstärke betrachtet. Gemäss Bender und Lange (2007) lassen sich aus einem Konfidenzintervall Rückschlüsse auf die statistische Signifikanz ableiten. Die Konfidenzintervalle für die beiden Gruppen (2.85, 2.79 & 3.10, 3.09) siehe Tabelle 11, überlappen sich stark. Eine Überlappung deutet darauf hin, dass es keinen präzisen Unterschied zwischen den Mittelwerten der Gruppen mit oder ohne Instruktion gibt, was wiederum die fehlende Signifikanz unterstützt.

Ausserdem deutet eine Effektstärke von 0.0887 auf einen kleinen Effekt hin, sprich dass der Unterschied der Gruppen mit oder ohne Instruktion gering ist. Auch in Kombination mit dem p-Wert von 0.717 deutet eine Effektstärke von 0.0887 darauf hin, dass selbst wenn es einen signifikanten Unterschied zwischen den Gruppen gäbe, der Unterschied, nicht von praktischer Bedeutung ist.

4.4 Überprüfung der Hypothese HS2

Um zu überprüfen, ob die Instruktion einen positiven Effekt auf das Nutzungsverhalten der Studierenden hat, wurde der t-Test für unabhängige Stichproben als Methode beigezogen.

Prüfung auf Normalverteilung

Wie in Kapitel 4.3 erwähnt wurde auch hier mittels Shapiro -Wilk-Test die Annahme auf Normalverteilung geprüft. Gemäss Fieger (1971) ist die Voraussetzung für eine Normalverteilung, dass der p-Wert (p) grösser als $p < 0.05$ ist. Die Analyse ergibt mit den Werten der Gruppe mit Instruktion ($W = 0.929$, $p = 0.021$) und der Gruppe ohne Instruktion ($W = 0.889$, $p = 0.004$), dass der Shapiro -Wilk-Test signifikant ist. Einen signifikanten Shapiro -Wilk-Test deutet darauf hin, dass die Daten nicht normalverteilt sind. Deshalb kann die Nullhypothese der Normalverteilung der Faktorwerte verworfen werden.

Der t-Test für unabhängige Stichproben erfordert eine Normalverteilung. Die Daten sind wie oben beschrieben nicht normalverteilt, weshalb sich für ein nichtparametrischen t-Test Mann-Whitney-U-Test entschieden wurde. Mann-Whitney-U-Test testet Unterschiede von zwei Gruppen ohne spezifische Verteilungen (McKnight & Najab, 2010). Die HS2 untersucht einen Unterschied zwischen dem Nutzungsverhalten der Studierenden der Gruppen mit und ohne Instruktion. Aufgrund dessen wurde der Mann-Whitney-U-Test angewendet.

Tabelle 12

t-Test für unabhängige Stichproben HS2

Mann-Whitney-U-Test	U	p	Effektstärke
<i>Nutzungsabsicht</i>	569	0.478	0.00872

Anmerkungen: 10% Signifikanzniveau ($p=0.1$), ** = 5% Signifikanzniveau ($p = 0.05$) und *** = 1% Signifikanzniveau ($p = 0.01$), M1 Instruktion > M2 keine Instruktion.

Tabelle 13

95% Konfidenzintervall für die Mittelwerte der Nutzungsabsicht der Studierenden, aufgeteilt in die Gruppen mit und ohne Instruktion

Konfidenzintervall	Instruktion	Keine Instruktion
95% KI-Mittelwert untere Grenze	3.53	3.55
95% KI-Mittelwert obere Grenze	4.12	4.12

Wie in Tabelle 12 dargestellt, zeigt der U-Test für die unabhängige Stichprobe mit den Werten ($U = 569$, $p = 0.478$), dass er nicht signifikant ist. Eine Effektstärke von 0.00872 deutet auf einen sehr geringen Effekt hin, konkret bedeutet dies, dass der Unterschied des Nutzungsverhalten der Gruppen mit oder ohne Instruktion sehr klein und nicht besonders stark ausgeprägt ist. Damit das Ergebnis des U-Tests bestätigt werden kann, wurde das Konfidenzintervall und die Effektstärke berechnet. Wie in Tabelle 13 ersichtlich ist, sind die Konfidenzintervalle von beiden Gruppen mit oder ohne Instruktion beinahe identisch, dies bestätigt den U-Wert von 569. Das bedeutet, dass kein bedeutsamer Unterschied zwischen den Gruppen mit oder ohne Instruktion besteht. Auch die Effektstärke von 0.00872 deutet darauf hin, dass selbst wenn es einen signifikanten Unterschied zwischen den Gruppen gäbe, der Unterschied sehr gering wäre. Demzufolge wird die H_0 , dass die Instruktion keinen Effekt auf das Nutzungsverhalten der Studierenden hat, beibehalten und nicht verworfen.

4.5 Überprüfung der Hypothese HS3

Die Hypothese, dass Selbstwirksamkeit und der Umgang mit ChatGPT einen positiven Zusammenhang haben, soll geprüft werden. Das Ziel von Korrelationsstudien besteht darin, den Zusammenhang von zwei Variablen zu analysieren (Kuckartz, Rädiker, Ebert & Schehl, 2013). Weshalb sich für den Korrelationskoeffizient nach Pearson entschieden wurde.

Prüfung auf Normalverteilung

Mittels Shapiro -Wilk-Test wurde die Annahme auf Normalverteilung geprüft. Die Voraussetzung für eine Normalverteilung ist, dass der p-Wert (p) grösser als $p > 0.05$ ist (Fieger, 1971). Die Auswertung ergibt bei der Variable Selbstwirksamkeit die Werte ($W = 0.929$, $p = <.001$) und bei den ChatGPT Aufgaben ($W = 0.970$, $p = 0.105$). Die Daten der Selbstwirksamkeit sind mit einem p-Wert von $<.001$ nicht normalverteilt, hingegen sind die Daten der ChatGPT Aufgaben mit einem p-Wert von 0.105 normalverteilt.

Der Korrelationskoeffizient nach Pearson bedingt eine Voraussetzung der Normalverteilung, weshalb sich für das nicht-parametrische Verfahren, die Spearman-Korrelation entschieden wurde. Die Spearman-Korrelation setzt mindestens ordinal skalierte Daten voraus.

Tabelle 14

Spearman-Korrelation HS3

Selbstwirksamkeit		
ChatGPT Aufgaben	<i>Spearman`s Rho</i>	p-Wert
	0.068	0.584

Anmerkungen: * = Signifikanzniveau ($p=0.5$), ** = Signifikanzniveau ($p = 0.01$) und *** = Signifikanzniveau ($p = 0.001$).

Wie in Tabelle 14 dargestellt, ergibt die Spearman-Korrelation mit den Werten (Spearman-Rho-Wert = 0.068, p-Wert = 0.584), dass es keinen signifikanten Zusammenhang zwischen der Selbstwirksamkeit und dem Umgang von ChatGPT gibt. Ein Spearman-Rho-Wert von 0.068 deutet darauf hin, dass der Zusammenhang zwischen der Selbstwirksamkeit und dem Umgang mit ChatGPT positiv sehr schwach ist. Der p-Wert von über 0.05, ist sehr hoch, weshalb die Nullhypothese $H_0 = \text{«Es besteht kein Zusammenhang zwischen der Selbstwirksamkeit und dem Umgang mit ChatGPT»}$, nicht abgelehnt werden kann und beibehalten wird. Man kann davon ausgehen, dass der beobachtete Korrelationskoeffizient auf rein zufällige Variationen zurückzuführen sein könnte.

5 Diskussion

In den nachfolgenden Kapiteln werden die Ergebnisse interpretiert, damit die ursprünglichen Fragestellungen beantwortet werden können. Daraufhin werden mögliche Handlungsempfehlungen formuliert, um einen erfolgreichen Umgang an Hochschulen mit LLM's zu gestalten. Zudem werden Limitationen der vorliegenden Studie dargelegt.

Diese Arbeit verfolgte das Ziel zu untersuchen, ob Studierende mit einer Instruktion einen besseren Umgang mit ChatGPT haben, wie jene ohne Instruktion. Des Weiteren sollte aufgeklärt werden, ob das Nutzungsverhalten der Studierenden durch eine Instruktion beeinflusst wird. Zusätzlich sollte untersucht werden, ob ein Zusammenhang zwischen der Selbstwirksamkeit und dem Umgang mit ChatGPT besteht.

Aufbauend auf diesen Zielen wurden Hypothesen HS1, HS2 und HS3 abgeleitet, siehe in Kapitel 3.2.

Die Hypothesen HS1 und HS2 wurden anhand dem t-Test für unabhängige Stichproben ausgewertet. Die Auswertung erfolgte bei der Hypothese HS3 anhand der Spearman-Korrelation.

5.1 Beantwortung der ersten Fragestellung

Die Ergebnisse der Fragestellung eins:

«Führt die Bereitstellung einer Instruktion zur Nutzung von Large Language Models, wie ChatGPT, zu besseren Fähigkeiten der Studierenden im Umgang mit ChatGPT im Vergleich zu jenen Studierenden, welche keine spezifische Instruktion erhalten haben?»

wurden mittels t-Test für unabhängige Stichproben untersucht. Die Ergebnisse des t-Test zeigen einen nicht signifikanten Unterschied zwischen den Fähigkeiten der Studierenden mit und ohne Instruktion. Somit weisen Studierende, die eine Instruktion erhalten haben, keine besseren Fähigkeiten im Umgang mit ChatGPT auf. Die Ergebnisse der explorativen Untersuchungen ergaben hingegen, dass Studierende, die eine Instruktion erhalten haben, eine gering bessere Fähigkeit in den 10 ChatGPT Aufgaben aufweisen, welche auf die Nutzungsabsicht abzielen. Dementgegen zeigen Studierende ohne Instruktion im Hinblick auf die Einschätzung des Wahrheitsgehalts einen durchschnittlichen, minimalen besseren Umgang mit ChatGPT als Studierende mit Instruktion. Ein möglicher Grund, für den nicht signifikanten Unterschied könnte sein, dass das Erlangen von neuem Wissen ein individueller Prozess ist, welcher sehr stark von bereits vorhandenem Wissen einer Person beeinflusst wird (Glaserfeld, 1989). Gemäss Mayer (2002) zeigen Studierende unterschiedliche Lernprozesse, welches ein weiterer Einflussfaktor, für den nicht signifikanten Unterschied erklären könnte. Ausserdem hat gemäss Mayer (2012) ein effektives Lernen einen Einfluss auf die Gestaltung von Lerninhalten. Demzufolge kann

interpretiert werden, dass die Instruktion, welche den Studierenden in dieser Studie präsentiert wurde, nicht effektiv genug gestaltet war, um einen ausreichenden Lerneffekt zur Verbesserung der Fähigkeiten im Umgang mit ChatGPT zu gewährleisten. Die Ergänzung von visuellen Darstellungen zu den jeweiligen Inhalten, sowie eine gleichzeitige Präsentation von inhaltlich miteinander verknüpften Texten, hätte gemäss dem Prinzip der zeitlichen Nähe von Mayer (2012) bei den Studierenden zu einem wirksameren Lernen des Inhalts der Instruktion führen können. Zusätzlich könnte gemäss dem Prinzip des Vorwissens, wie in Kapitel 2.5 erläutert, das Vorwissen der Studierenden ohne Instruktion bereits sehr gut gewesen sein. Dadurch könnte es sein, dass die erhaltenen Informationen, welche die Studierenden ohne Instruktion durch die Einleitung, wie in Tabelle 3 dargestellt, erhalten haben, dazu geführt haben, ihre Fähigkeiten im Umgang mit ChatGPT zu verbessern. Zudem könnte aufgrund der Tatsache, dass die befragten Probanden dieser Studie mehrheitlich «junge» Studierende im Alter von 20-30 Jahren sind, die tendenziell bereits ein grosses Vorwissen (auch ohne Instruktion) zu ChatGPT haben einen Einfluss haben. Diese Interpretationen könnten weitere Argumente für den nicht signifikanten Unterschied sein. Obwohl der t-Test einen nicht signifikanten Unterschied zwischen den Studierenden mit oder ohne Instruktion ergab, zeigen die Ergebnisse in Tabelle 6, dass 85% der Studierenden eine Instruktion als extrem, sehr oder ziemlich hilfreich empfinden. Basierend auf diesen Ergebnissen, könnte eine Implementierung einer strukturierten Instruktion dennoch für die Studierenden hilfreich sein, um deren Umgang mit ChatGPT zu verbessern. Allerdings fehlen gemäss Schmohl, Watanabe und Schelling (2023) umfassende empirische Studien zur Lernwirksamkeit von KI. Ausserdem mangelt es an systematischen theoretischen Ansätzen und praktische Anwendungen. Zusätzlich soll es gemäss Bates, Cobo, Mariño und Wheeler (2020) nur wenige Studien geben, die den Erfolg von bereits umgesetzten Implementierungen messen. Demzufolge soll es noch weniger Studien geben, welche eine Implementierung in der bildungswissenschaftlichen Forschung untersuchen.

5.2 Beantwortung der zweiten Fragestellung

Die Fragestellung zwei:

«Wie wird das Nutzungsverhalten der Studierenden durch eine Instruktion beeinflusst?»

wurde mittels Mann-Whitney-U-Test untersucht. Die Ergebnisse zeigen einen nicht signifikanten U-Test. Dementsprechend kann gesagt werden, dass die Studierenden mit oder ohne Instruktion hinsichtlich ihrem Nutzungsverhalten keinen bedeutsamen Unterschied aufzeigen. Dementsprechend kann die Fragestellung folgend beantwortet werden, dass das Nutzungsverhalten der Studierenden durch eine Instruktion nicht bemerkenswert beeinflusst

wird. Gemäss Buchkremer et al. (2020) soll die Nutzungsabsicht stark davon abhängig sein, wie eine Person grundsätzlich eingestellt ist. Folglich könnte daraus interpretiert werden, dass die Studierenden, welche keine Instruktion erhalten haben, grundsätzlich schon ein sehr gutes Nutzungsverhalten gegenüber ChatGPT aufzeigen und deshalb in dieser Studie kein bedeutsamer Unterschied aufgezeigt werden konnte. Jedoch konnten Alalwan, Dwivedi und Rana (2017) empirisch belegen, dass es einen signifikanten Zusammenhang zwischen der Nutzungsabsicht und der tatsächlichen Nutzung gibt. Der Mehrwert, welcher daraus erzielt werden kann, beeinflusst stark die Akzeptanz gegenüber der jeweiligen Technologie. Trotz zahlreicher Studien, welche verschiedene Zusammenhänge mit der Nutzungsabsicht bestätigen konnten, ist gemäss Seidel und Macht (2023) wenig darüber bekannt, wie die individuelle Nutzungsabsicht eine Technologie mit KI wie beispielsweise ChatGPT zu verwenden beeinflusst wird.

5.3 Beantwortung der dritten Fragestellung

Die letzte Fragestellung drei:

«Gibt es einen Zusammenhang zwischen Selbstwirksamkeit und dem Umgang mit ChatGPT?»

wurde mittels Spearman-Korrelation geprüft. Die gerichtete Hypothese bildete einen positiven Zusammenhang, dass die Selbstwirksamkeit einen positiven Zusammenhang mit dem Umgang von ChatGPT hat. Die Resultate ergaben einen nicht signifikanten Zusammenhang zwischen Selbstwirksamkeit und dem Umgang der Studierenden mit ChatGPT. Dieses Resultat widerspricht der Studie von Kwak, Ahn und Seo (2022), welche anhand einer Regressionsanalyse bestätigen konnten, dass eine positive Einstellung gegenüber KI und ein hohes Mass an Selbstwirksamkeit die zwei wichtigsten Faktoren waren, welche das Verhalten der Studierenden beeinflussten. Demzufolge empfehlen Kwak et al. (2022) gezielte Massnahmen zu formulieren und diese auch auszuführen, welche den Studierenden hilft besser mit der KI-Technologie wie beispielsweise ChatGPT umzugehen. Gemäss Mischel (2015) wird in westlichen Gesellschaften die Fachrichtung Technik tendenziell als «männlich» angesehen. Konkret bedeutet dies, dass Männer schon in jungen Jahren oft Zugang zu technischen Berufen, Aufgaben oder Aktivitäten erhalten, während Mädchen eher ermutigt werden, sich in sozialen oder kreativen Bereichen zu engagieren. Folglich könnte interpretiert werden, dass der Mangel an Erfahrung von Frauen im Bereich der Technik sich negativ auf ihre Selbstwirksamkeit auswirken könnte. Demgegenüber könnte dies ein Interpretation für den nicht signifikanten Zusammenhang zwischen der Selbstwirksamkeit und dem Umgang mit ChatGPT sein, da diese Studie einen Anteil von 70% Frauen umfasste.

5.4 Handlungsempfehlungen

In Bezug auf die Beantwortung der Fragestellungen in den vorherigen Kapiteln lassen sich Empfehlungen für die Hochschulen ableiten. Bezüglich der Argumentation, dass Hochschulen sich noch nicht ausreichend mit der Künstlichen Intelligenz befasst haben, sollten konkrete Vorschläge formuliert werden. Gemäss Gimpel, Jung, Utz und Wöhl (2023) könnten basierend auf den Handlungsempfehlungen, folgende konkrete Schritte zur Integration und Anwendung von KI-Modellen an Hochschulen wie beispielsweise ChatGPT eingeführt werden. Um Studierende über die Nutzung von ChatGPT und andere LLM's zu informieren, könnten Hochschulen Workshops und Schulungen für die Studierenden anbieten. Somit könnten auch ethische Aspekte, Datenschutzfragen sowie Methoden zur effektiven Anwendung von KI-Technologien behandelt werden. Damit sich Studierende mit LLM's wie beispielsweise ChatGPT vertraut machen, könnte eine Instruktion für Studierenden gemäss den Prinzipien von Mayer (2012) formuliert werden, sowie beispielsweise eine Schritt für Schritt Anleitung wie ChatGPT angewendet werden kann, wie Gimpel et al. (2023) sie formuliert haben, eine ergänzende Möglichkeit für Studierende darstellen. Zusätzlich würden die formulierten Handlungsempfehlungen die Anliegen der Studierenden, wie in Tabelle 7 dargestellt, abdecken. In einem Interview zwischen Dilger (2023) und ChatGPT, ergaben sich auf die Frage wie ChatGPT in Hochschulen eingesetzt werden kann folgende Einsatzmöglichkeiten:

- Als Unterstützung um grosse Datenmengen zu analysieren und komplexe Modelle zu erstellen.
- Als personalisierte Lernunterstützung für Studierende
- Für die Prüfungsvorbereitung, indem ChatGPT automatische Übungsfragen generiert, damit Studierende ihr Wissen überprüfen können

Angesichts der zunehmenden Verbreitungen von KI in wissenschaftlichen Arbeiten, sollten gemäss Vogelgesang, Bleher, Krupitzer, Stein und Jung (2023) Hochschulen den Einsatz der KI nicht verbieten. Vielmehr sollten Hochschulen klare Richtlinien formulieren, auf welche Weise Studierende KI verwenden dürfen und wie deren Nutzung transparent gemacht werden muss. Zusätzlich betonen Vogelgesang et al. das Studierende einen angemessenen Umgang mit KI erlernen müssen sowie Kompetenzen entwickeln, KI-generierte Texte auch kritisch zu analysieren. Gemäss Leschke (2023) spielt der Einfluss der KI beispielsweise in wissenschaftlichen Arbeiten keine Rolle, solange diese auch wissenschaftlich fundiert sind. Ausserdem betont Salden (2023), dass der Einsatz der Künstlichen Intelligenz Studierende unterstützen kann, die KI aber gelenkt und kontrolliert werden muss. Konkret bedeutet dies das Studierende, welche keine grundlegenden Kompetenzen im wissenschaftlichen Schreiben haben, auch keine wissenschaftliche Arbeit erbringen können.

Zusammenfassend kann festgehalten werden, dass es wichtig ist, dass Hochschulen den Einsatz von KI zulassen und klare Richtlinien für den Umgang mit der KI an Hochschulen formulieren und an die Studierenden kommunizieren. Zusätzlich sollten die Studierenden bezüglich LLM's aufgeklärt und informiert werden, sodass Studierende sich in diesem Bereich Kompetenzen aneignen können.

5.5 Limitationen dieser Arbeit

Im Folgenden werden mögliche Limitationen bei der Erhebung und Auswertung dieser Forschungsarbeit diskutiert. Die Fragebögen zur Erhebung der Konstrukte der Selbstwirksamkeit von Beierlein et al. (2019) und die Nutzungsabsicht von Venkatesh und Davis (2000), wiesen beide eine hohe interne Konsistenz auf. Dennoch besteht die Einschränkung, dass eine soziale Erwünschtheit die Selbsteinschätzung der Studierenden beeinflusst haben könnte. Zudem wurden die Items der Nutzungsabsicht von Venkatesh und Davis (2000) eigens auf den Bereich Studium angepasst. Des Weiteren wurde die Skalierung von einer siebenstufigen Likert Skala auf eine fünfstufige Likert Skala von Beierlein et al. (2019) übernommen. Die Fähigkeiten der Studierenden wurden anhand selbst erstellter Aufgaben erhoben, wodurch eine wesentliche Einschränkung dieser Studie bestehen kann. Die Bewertungskriterien der 15 ChatGPT-Aufgaben wurden eigens festgelegt, was eine zusätzliche Einschränkung darstellt, die berücksichtigt werden sollte. Eine weitere Einschränkung könnte die mangelnde Validität und Reliabilität dieser Studie sein, welchen Einfluss auf die Ergebnisse haben könnte (Möhring & Schlütz, 2013).

Bezüglich der Auswertungsmethoden lassen sich weitere Limitationen feststellen. Die Fähigkeiten der Studierenden wurden als eine Gesamtvariable berechnet. Konkret bedeutet dies, dass die 10 Aufgaben der Nutzungsabsicht und die fünf Aufgaben des Wahrheitsgehaltes für die Datenauswertung zusammen berechnet wurden. Eine weitere Einschränkung ist, dass die Mehrheit der Studierenden, welche befragt wurden an der Fachhochschule Nordwestschweiz studieren und diese Ergebnisse deshalb nur begrenzt nach aussen getragen werden können. Ausserdem unterschied sich die Altersklasse der Probanden sehr gering, die Mehrheit der Studierenden waren eher jüngere Personen im Alter von 20-30 Jahren. Abschliessend wird erwähnt, dass diese Studie eine Querschnittsanalyse war. Gemäss Bierhoff und Petermann (2013), bringen solche Analysen Nachteile wie, dass eine einmalige erfasste Wahrnehmung bestimmter Konstrukte oftmals mit dem aktuellen Zustand verglichen wird und nicht über eine bestimmte Zeitspanne neutral betrachtet wird.

6 Fazit und weiterführende Forschungsansätze

Zusammenfassend lässt sich festhalten, dass die gewonnenen Ergebnisse der quantitativen Datenerhebung ein breites Forschungsfeld an Möglichkeiten und Erkenntnisse eröffnet. Durch die Erhebung wurde ein detaillierter Einblick in den Umgang von Studierenden mit LLM`s gewonnen. Die gewonnenen Ergebnisse deuten darauf hin, dass weder eine Instruktion noch eine höhere Selbstwirksamkeit einen signifikanten Einfluss auf den Umgang mit LLM`s haben. Demzufolge muss bei der Interpretation der Ergebnisse beachtet werden, dass durch die nicht signifikanten Ergebnisse der Einfluss auf den Umgang mit ChatGPT ungewiss ist.

Daraus ergibt sich für die Hochschulen folgende zentrale Frage, welche Massnahmen ergriffen werden müssen, damit LLM`s in Hochschulen implementiert werden können? Zur Beantwortung dieser Frage wird vorgeschlagen Schulungen und Workshops für Studierende anzubieten sowie als ergänzende Möglichkeit eine Bereitstellung einer Instruktion wie beispielsweise «eine Schritt für Schritt Anleitung für Studierende» in Betracht zu ziehen.

Diese Studie deckt aber auch neue Forschungslücken auf: Was sind Ursachen, dass Studierende mit oder ohne Instruktion keinen Unterschied auf den Umgang mit ChatGPT aufzeigen? Hängt die Antwort allenfalls mit der Qualität und Klarheit der Instruktion zusammen, dass die Instruktion nicht ausreichend klar oder detailliert formuliert war, damit ein Unterschied erkannt werden konnte? Oder haben interpersonale Unterschiede einen Einfluss? Könnte das Alter einen Einfluss haben? Hatten die Studierenden schon ein sehr gutes Vorwissen, sodass die Instruktion keinen Einfluss hatte? Es empfiehlt sich daher, die Thematik über den Umgang mit ChatGPT tiefgründiger zu analysieren. Ein Vorschlag wäre die Durchführung einer qualitativen Studie. Durch die Planung von Interviews mit Studierenden könnte die individuelle Erfahrung und Wahrnehmung der Instruktion erfasst werden. Diese Interviews könnten Aufschluss darüber geben, ob und wie die Instruktion von den Studierenden wahrgenommen werden. Ein weiterer Vorschlag könnte eine Längsschnittstudie sein, indem Studierende über einen längeren Zeitraum befragt werden, um zu sehen, ob die Bereitstellung einer Instruktion über längere Zeit das Verhalten verändern würde und somit langfristige Unterschiede sichtbar werden. Durch eine qualitative Studie und einer Längsschnittstudie könnten somit tiefere Einblicke gewonnen werden und Massnahmen zum Umgang mit ChatGPT abgeleitet werden. Die Ergebnisse legen dar, dass zusätzlich ein anderer Blickwinkel auf diese Thematik nötig ist. LLM`s sowie die KI sind komplexe Themen und erfordern die Prüfung weiterer Faktoren, damit den Studierenden ein besserer Umgang mit LLM`s ermöglicht werden kann.

7 Literaturverzeichnis

- Alalwan, A. A., Dwivedi, Y. K. & Rana, N. P. (2017). Factors influencing adoption of mobile banking by Jordanian bank customers: Extending UTAUT2 with trust. *International Journal of Information Management*, 37(3), 99–110. <https://doi.org/10.1016/j.ijinfomgt.2017.01.002>
- Aldosari, S. A. M. (2020). The Future of Higher Education in the Light of Artificial Intelligence Transformations. *International Journal of Higher Education*, 9(3), 145. <https://doi.org/10.5430/ijhe.v9n3p145>
- Amigo, J. M. (2021). Data Mining, Machine Learning, Deep Learning, Chemometrics. Definitions, common points and Trends (Spoiler Alert: VALIDATE your models!). *Brazilian Journal of Analytical Chemistry*, 8(32), 45–61. <https://doi.org/10.30744/brjac.2179-3425.AR-38-2021>
- Artificial intelligence a modern approach. (2022). (Fourth edition, global edition.). Boston: Pearson.
- Bandilla, W. (2015). *Online-Befragungen (GESIS Survey Guidelines) Online-Befragungen (GESIS Survey Guidelines)*. [object Object]. https://doi.org/10.15465/GESIS-SG_003
- Bandura, A. (1986). *Social foundations of thought and action: A social cognitive theory*. Englewood Cliffs, NJ: Prentice Hall.
- Bates, T., Cobo, C., Mariño, O., & Wheeler, S. (2020). Can Artificial Intelligence Transform Higher Education? *International Journal of Educational Technology in Higher Education*, (S. 1–12).
- Batinic, B. & Moser, K. (2005). Determinanten der Rücklaufquote in Online-Panels. *Zeitschrift für Medienpsychologie*, 17(2), 64–74. <https://doi.org/10.1026/1617-6383.17.2.64>
- Baur, N. & Blasius, J. (Hrsg.). (2014). *Handbuch Methoden der empirischen Sozialforschung*. Wiesbaden: Springer Fachmedien Wiesbaden. <https://doi.org/10.1007/978-3-531-18939-0>
- Beierlein, C., Kovaleva, A., Kemper, C. J. & Rammstedt, B. (2014). Allgemeine Selbstwirksamkeit Kurzsкала (ASKU). *Zusammenstellung sozialwissenschaftlicher Items und Skalen (ZIS)*. [object Object]. <https://doi.org/10.6102/ZIS35>
- Bender, R. & Lange, S. (2007). Was ist ein Konfidenzintervall? *DMW - Deutsche Medizinische Wochenschrift*, 132(S 01), e17–e18. <https://doi.org/10.1055/s-2007-959031>
- Bierhoff, H. W & Petermann, F. (2013). *Forschungsmethoden der Psychologie*. Bern: Hogrefe Verlag GmbH & Company KG.
- Brandenburg, T. & Thielsch, M. T. (Hrsg.). (2009). *Praxis der Wirtschaftspsychologie: Themen und Fallbeispiele für Studium und Anwendung. 1 (MV-Wissenschaft)*. Münster: Verl.-Haus Monsenstein und Vannerdat.
- Brown, T. B., Mann, B., Ryder, N., Subbiah, M., Kaplan, J., Dhariwal, P., Neelakantan, A., et al. (2020). Language Models are Few-Shot Learners. arXiv. Zugriff am 2.5.2024. Verfügbar unter: <http://arxiv.org/abs/2005.14165>
- Büching, C., Mah, D-K., Otto, S., Paulicke, P., & Hartman, E. (2019). Learning Analytics an Hochschulen. In Wittpahl, V. (Hrsg.). (2019). *Künstliche Intelligenz: Technologie |*

- Anwendung | Gesellschaft* (S. 142–160). Berlin, Heidelberg: Springer Berlin Heidelberg.
<https://doi.org/10.1007/978-3-662-58042-4>
- Buchkremer, R., Heupel, T. & Koch, O. (Hrsg.). (2020). *Künstliche Intelligenz in Wirtschaft & Gesellschaft: Auswirkungen, Herausforderungen & Handlungsempfehlungen* (FOM-Edition). Wiesbaden: Springer Fachmedien Wiesbaden. <https://doi.org/10.1007/978-3-658-29550-9>
- Buxmann, P. & Schmidt, H. (Hrsg.). (2019). *Künstliche Intelligenz: Mit Algorithmen zum wirtschaftlichen Erfolg*. Berlin, Heidelberg: Springer Berlin Heidelberg.
<https://doi.org/10.1007/978-3-662-57568-0>
- Cardona, D. R., Janssen, A., Uphaus, J. & Breitner, M. H. (2020). Nutzerakzeptanz von Robo-Advisor Systemen für das digitale Investitionsmanagement in Deutschland.
- Cohen J. A. (1992). Quantitative methods in psychology: A power primer. *Psychol. Bull.*, 112, 1155–1159.
- Dale, R. (2021). GPT-3: What's it good for? *Natural Language Engineering*, 27(1), 113–118.
<https://doi.org/10.1017/S1351324920000601>
- Davis, F. D. (1989). Perceived usefulness, perceived ease of use, and user acceptance of information technology. *MIS Quarterly*, 13, 319–340.
- De Witt, C., Gloerfeld, C. & Wrede, S. E. (Hrsg.). (2023). *Künstliche Intelligenz in der Bildung*. Wiesbaden: Springer Fachmedien Wiesbaden. <https://doi.org/10.1007/978-3-658-40079-8>
- Decker, O. (Hrsg.). (2018). *Sozialpsychologie und Sozialtheorie*. Wiesbaden: Springer Fachmedien Wiesbaden. <https://doi.org/10.1007/978-3-531-19564-3>
- Dilger, A. (2023). Interview mit ChatGPT über KI an Hochschulen. University of Muenster, Institutional Repository MIAMI. <https://doi.org/10.17879/50099607072>
- D'Mello, S. & Graesser, A. (2013). Design of Dialog-Based Intelligent Tutoring Systems to Simulate Human-to-Human Tutoring. In A. Neustein & J.A. Markowitz (Hrsg.), *Where Humans Meet Machines* (S. 233–269). New York, NY: Springer New York. https://doi.org/10.1007/978-1-4614-6934-6_11
- Döring, N. & Bortz, J. (2016). *Forschungsmethoden und Evaluation in den Sozial- und Humanwissenschaften*. Springer Lehrbuch. Berlin, Heidelberg: Springer.
<https://doi.org/10.1007/978-3-642-41089-5>
- Ertel, W. (2021). *Grundkurs Künstliche Intelligenz: Eine praxisorientierte Einführung* (Computational Intelligence). Wiesbaden: Springer Fachmedien Wiesbaden. <https://doi.org/10.1007/978-3-658-32075-1>
- Externbrink, K., Kolominski, S., Surma, S., Möltner H. & Sülzenbrück, S. (2020). *Wege zur empirischen Abschlussarbeit in der Wirtschaftspsychologie*. iwv Schriftenreihe No 3. Essen: MA Akademie Verlags- und Druck-Gesellschaft mbH. <http://hdl.handle.net/10419/224461>
- Fachhochschule Nordwestschweiz (FHNW), Hochschule für Angewandte Psychologie. (n.d.). Institut Mensch in komplexen Systemen. Abgerufen am 24. Mai 2024, von

<https://www.fhnw.ch/de/die-fhnw/hochschulen/psychologie/institut-mensch-in-komplexen-systemen>

- Fahrmeir, L., Künstler, R., Pigeot, I. & Tutz, G. (2003). *Statistik* (Springer-Lehrbuch). Berlin, Heidelberg: Springer Berlin Heidelberg. <https://doi.org/10.1007/978-3-662-22657-5>
- Faul, F., Erdfelder, E., Lang, A.-G. & Buchner, A. (2007). G*Power 3: A flexible statistical power analysis program for the social, behavioral, and biomedical sciences. *Behavior Research Methods*, 39(2), 175–191. <https://doi.org/10.3758/BF03193146>
- Fieger, W. (1971). Eine statistische Charakterisierung der Normalverteilung. *Zeitschrift für Wahrscheinlichkeitstheorie und Verwandte Gebiete*, 19(4), 330–344. <https://doi.org/10.1007/BF00535838>
- Fink, M. A. (2023). Große Sprachmodelle wie ChatGPT und GPT-4 für eine patientenzentrierte Radiologie. *Die Radiologie*, 63(9), 665–671. <https://doi.org/10.1007/s00117-023-01187-8>
- Frey, D. (Hrsg.). (2016). *Psychologie der Werte: Von Achtsamkeit bis Zivilcourage – Basiswissen aus Psychologie und Philosophie*. Berlin, Heidelberg: Springer Berlin Heidelberg. <https://doi.org/10.1007/978-3-662-48014-4>
- Fuchs, C. (2018). *Soziale Medien und Kritische Theorie: Eine Einführung* (1. Auflage). Stuttgart, Deutschland: utb GmbH. <https://doi.org/10.36198/9783838550817>
- Geissler, S. (2017). Ob als Fachleute, aber auch als Konsumenten oder Bürgerinnen und Bürger: Es erscheint angeraten und nötig, sich ein kompetentes Bild von den neuen Verfahren, seinen Möglichkeiten und Grenzen zu machen.
- Gimpel, H., Hall, K., Decker, S., Eymann, T., Lämmermann, L., Mädche, A. et al. (2023). Unlocking the Power of Generative AI Models and Systems such as GPT-4 and ChatGPT for Higher Education.
- Gimpel, H., Jung, C., Utz, L. & Wöhl, M. (2023). Von Null auf ChatGPT: Eine Schritt-für-Schritt-Anleitung, um sich mit der künstlichen Intelligenz vertraut zu machen. Universität Hohenheim.
- Glaserfeld, E. V. (1989). Facts and the self from a constructivist point of view. *Poetics*, 18(4–5), 435–448. [https://doi.org/10.1016/0304-422X\(89\)90041-7](https://doi.org/10.1016/0304-422X(89)90041-7)
- Gross, T. (2013). Mensch-Computer-Interaktion in Wissenschaft und Praxis. *HMD Praxis der Wirtschaftsinformatik*, 50(6), 6–15. <https://doi.org/10.1007/BF03342064>
- Gulden, J. (2019). *Automatisierte Geldanlage: Determinanten und Einflussbedingungen der Akzeptanz von Investment Management FinTechs*. Wiesbaden: Springer Fachmedien Wiesbaden. <https://doi.org/10.1007/978-3-658-24054-7>
- Häder, M. (2015). *Empirische Sozialforschung* (3., überarbeitete Auflage). Wiesbaden: Springer Press.

- Harwardt, M. & Köhler, M. (2023). *Künstliche Intelligenz entlang der Customer Journey: Einsatzpotenziale von KI im E-Commerce (essentials)*. Wiesbaden: Springer Fachmedien Wiesbaden. <https://doi.org/10.1007/978-3-658-39109-6>
- Heinecke, A. M. (2012). *Mensch-Computer-Interaktion: Basiswissen für Entwickler und Gestalter* (X.media.press). Berlin, Heidelberg: Springer Berlin Heidelberg. <https://doi.org/10.1007/978-3-642-13507-1>
- Hussy, W., Schreier, M. & Echterhoff, G. (2013). *Forschungsmethoden in Psychologie und Sozialwissenschaften für Bachelor* (Springer-Lehrbuch). Berlin, Heidelberg: Springer Berlin Heidelberg. <https://doi.org/10.1007/978-3-642-34362-9>
- Jetter, H.-C. (2022). D 3 Mensch-Computer-Interaktion, Usability und User Experience. In R. Kuhlen, D. Lewandowski, W. Semar & C. Womser-Hacker (Hrsg.), *Grundlagen der Informationswissenschaft* (S. 525–534). De Gruyter. <https://doi.org/10.1515/9783110769043-045>
- Kellogg, K. C., Valentine, M. A. & Christin, A. (2020). Algorithms at Work: The New Contested Terrain of Control. *Academy of Management Annals*, 14(1), 366–410. <https://doi.org/10.5465/annals.2018.0174>
- Kim, J. K., Chua, M., Rickard, M. & Lorenzo, A. (2023). ChatGPT and large language model (LLM) chatbots: The current state of acceptability and a proposal for guidelines on utilization in academic medicine. *Journal of Pediatric Urology*, 19(5), 598–604. <https://doi.org/10.1016/j.jpuro.2023.05.018>
- Koch, M., Ziegler, J., Reuter, C., Schlegel, T. & Prilla, M. (2020). Mensch-Computer-Interaktion als zentrales Gebiet der Informatik – Bestandsaufnahme, Trends und Herausforderungen. *Informatik Spektrum*, 43(6), 381–387. <https://doi.org/10.1007/s00287-020-01299-8>
- Konrad, K. (2014). *Lernen lernen – allein und mit anderen: Konzepte, Lösungen, Beispiele*. Wiesbaden: Springer Fachmedien Wiesbaden. <https://doi.org/10.1007/978-3-658-04986-7>
- Krettek, C. (2023). ChatGPT: Milestone-Text-KI mit Game-Changer-Potenzial. *Die Unfallchirurgie*, 126(3), 252–254. <https://doi.org/10.1007/s00113-023-01296-y>
- Kreutzer, R. T. (2023). *Künstliche Intelligenz verstehen: Grundlagen – Use-Cases – unternehmenseigene KI-Journey*. Wiesbaden: Springer Fachmedien Wiesbaden. <https://doi.org/10.1007/978-3-658-42598-2>
- Kuckartz, U., Rädiker, S., Ebert, T. & Schehl, J. (2013). *Statistik: Eine verständliche Einführung*. Wiesbaden: VS Verlag für Sozialwissenschaften. <https://doi.org/10.1007/978-3-531-19890-3>
- Kwak, Y., Ahn, J.-W. & Seo, Y. H. (2022). Influence of AI ethics awareness, attitude, anxiety, and self-efficacy on nursing students' behavioral intentions. *BMC Nursing*, 21(1), 267. <https://doi.org/10.1186/s12912-022-01048-0>
- Leschke, J. (2023). *Educational and legal perspectives on AI-supported writing in higher education*. (S. 614 KB, 39 pages). Ruhr-Universität Bochum. <https://doi.org/10.13154/294-10284>

- Lew, R. (2023). ChatGPT as a COBUILD lexicographer. *Humanities and Social Sciences Communications*, 10(1), 704. <https://doi.org/10.1057/s41599-023-02119-6>
- Mayer, R. E. (2002). Rote Versus Meaningful Learning. *Theory Into Practice*, 41(4), 226–232. https://doi.org/10.1207/s15430421tip4104_4
- Mayer, R. E. (2012). *Multimedia learning*. (Second edition.). New York: Cambridge University Press.
- McCarthy, J. (1996). Defending AI research: a collection of essays and reviews. CSLI lecture notes: no. 49. Center for the Study of Language and Information. distributed by Cambridge University Press
- McKnight, P. E. & Najab, J. (2010). Mann-Whitney U Test. In I.B. Weiner & W.E. Craighead (Hrsg.), *The Corsini Encyclopedia of Psychology* (1. Auflage, S. 1–1). Wiley. <https://doi.org/10.1002/9780470479216.corpsy0524>
- Mischel, W. (2015). A social learning view of sex differences in behavior. In V. Burr (Ed.), *Gender and psychology* (pp. 108–129). Routledge/Taylor & Francis Group.
- Möhring, W. & Schlütz, D. (Hrsg.). (2013). *Handbuch standardisierte Erhebungsverfahren in der Kommunikationswissenschaft*. Wiesbaden: Springer Fachmedien Wiesbaden. <https://doi.org/10.1007/978-3-531-18776-1>
- Moon, J.-W. & Kim, Y.-G. (2001). Extending the TAM for a World-Wide-Web context. *Information & Management*, 38(4), 217–230. [https://doi.org/10.1016/S0378-7206\(00\)00061-6](https://doi.org/10.1016/S0378-7206(00)00061-6)
- Moosbrugger, H. & Kelava, A. (2012). *Testtheorie und Fragebogenkonstruktion* (2. Auflage.). Berlin: Springer. <https://doi.org/10.1007/978-3-531-18776-1>
- Niels, A. (2019). *Attributionen in der Mensch-Computer-Interaktion: Einfluss auf die Bewertung und Gestaltung interaktiver Produkte*. Wiesbaden: Springer Fachmedien Wiesbaden. <https://doi.org/10.1007/978-3-658-25596-1>
- Paass, G. & Hecker, D. (2020). *Künstliche Intelligenz: Was steckt hinter der Technologie der Zukunft?* Wiesbaden: Springer Fachmedien Wiesbaden. <https://doi.org/10.1007/978-3-658-30211-5>
- Porst, R. (2011). *Fragebogen: Ein Arbeitsbuch*. Wiesbaden: VS Verlag für Sozialwissenschaften. <https://doi.org/10.1007/978-3-531-92884-5>
- Rainsberger, L. (2021). *KI - die neue Intelligenz im Vertrieb: Tools, Einsatzmöglichkeiten und Potenziale von Artificial Intelligence*. Wiesbaden [Heidelberg]: Springer Gabler. <https://doi.org/10.1007/978-3-658-31773-7>
- Rasch, B., Friese, M., Hofmann, W. & Naumann, E. (2021). *Quantitative Methoden 1: Einführung in die Statistik für Psychologie, Sozial- & Erziehungswissenschaften*. Berlin, Heidelberg: Springer Berlin Heidelberg. <https://doi.org/10.1007/978-3-662-63282-6>
- Reinders, H., Ditton, H., Gräsel, C. & Gniewosz, B. (Hrsg.). (2015). *Empirische Bildungsforschung: Strukturen und Methoden*. Wiesbaden: Springer Fachmedien Wiesbaden. <https://doi.org/10.1007/978-3-531-19992-4>

- Reuter, C. (Hrsg.). (2021). *Sicherheitskritische Mensch-Computer-Interaktion: Interaktive Technologien und Soziale Medien im Krisen- und Sicherheitsmanagement*. Wiesbaden: Springer Fachmedien Wiesbaden. <https://doi.org/10.1007/978-3-658-32795-8>
- Rinck, M. (2016). *Lernen: ein Lehrbuch für Studium und Praxis* (Kohlhammer-Standards Psychologie). Stuttgart: Verlag W. Kohlhammer.
- Römer, L. V., Kleinmanns, A. & Jobmann, D. A.-L. (2023). Wie verändert ChatGPT die Zukunft der Personalauswahl?
- Russell, S. J., Norvig, P. & Davis, E. (2010). *Artificial intelligence: a modern approach* (Prentice Hall series in artificial intelligence) (3rd ed.). Upper Saddle River: Prentice Hall.
- Salden, P. (2023). *Didaktische und rechtliche Perspektiven auf KI-gestütztes Schreiben in der Hochschulbildung*. (S. 1119 KB, 41 pages). Ruhr-Universität Bochum. <https://doi.org/10.13154/294-9734>
- Savery, J. R. & Duffy, T. M. (1995). Problem based learning: An instructional model and its constructivist framework. *Educational Technology*, 35(5), 31–38.
- Schallmo, D., Williams, C. A. & Tidd, J. (2022). THE ART OF HOLISTIC DIGITALISATION: A META-VIEW ON STRATEGY, TRANSFORMATION, IMPLEMENTATION, AND MATURITY. *International Journal of Innovation Management*, 26(03), 2240007. <https://doi.org/10.1142/S1363919622400072>
- Schieb, J. & Posch, P. N. (2023). *Der Digitalschock: was vom Hype bleiben wird - so verändern ChatGPT, Bard & Co. unseren Alltag* (1. Auflage.). München: Redline Verlag.
- Schmohl, T., Watanabe, A. & Schelling, K. (Hrsg.). (2023). *Künstliche Intelligenz in der Hochschulbildung: Chancen und Grenzen des KI-gestützten Lernens und Lehrens* (Hochschulbildung: Lehre und Forschung) (1. Auflage, Band 4). Bielefeld, Germany: transcript Verlag. <https://doi.org/10.14361/9783839457696>
- Schöneck, N. M. & Voss, W. (2013). *Das Forschungsprojekt: Planung, Durchführung und Auswertung einer quantitativen Studie*. Wiesbaden: Springer Fachmedien Wiesbaden. <https://doi.org/10.1007/978-3-531-19502-5>
- Schwarzer, R. & Jerusalem, M. (Hrsg.). (1999). *Skalen zur Erfassung von Lehrer- und Schülermerkmalen. Dokumentation der psychometrischen Verfahren im Rahmen der Wissenschaftlichen Begleitung des Modellversuchs Selbstwirksame Schulen*. Berlin: Freie Universität Berlin.
- Schwarzer, R. & Jerusalem, M. (2002). Das Konzept der Selbstwirksamkeit. [object Object]. <https://doi.org/10.25656/01:3930>
- Schweiger, W. & Beck, K. (Hrsg.). (2010). *Handbuch Online-Kommunikation* (1. Aufl.). Wiesbaden: VS Verlag für Sozialwissenschaften.

- Seemann, Michael (2023): Künstliche Intelligenz, Large Language Models, ChatGPT und die Arbeitswelt der Zukunft, Working Paper Forschungsförderung, No. 304, HansBöckler-Stiftung, Düsseldorf
- Seidel, M. & Macht, J. (Hrsg.). (2023). *China & Innovation: Was der deutsche Mittelstand von China lernen kann* (FOM-Edition). Wiesbaden: Springer Fachmedien Wiesbaden.
<https://doi.org/10.1007/978-3-658-40440-6>
- Sieber, A. (2019). *Dialogroboter: Wie Bots und künstliche Intelligenz Medien und Massenkommunikation verändern*. Wiesbaden: Springer Fachmedien Wiesbaden.
<https://doi.org/10.1007/978-3-658-24393-7>
- Spitzer, M. (2023). ChatGPT: Nur ein weiterer Trend oder eine Revolution? *Nervenheilkunde*, 42(04), 192–199. <https://doi.org/10.1055/a-1948-8785>
- Stacks, D. W. & Salwen, M. B. (Hrsg.). (2009). *An integrated approach to communication theory and research* (Communication series Communication theory and methodology) (2. ed.). New York, NY: Routledge.
- Stavridis, D. & Wacker, M. (2023). ChatGPT und künstliche Intelligenz – Die Zukunft ist jetzt! *Zeitschrift für Herz-, Thorax- und Gefäßchirurgie*, 37(5), 266–272.
<https://doi.org/10.1007/s00398-023-00593-3>
- Stenzel, S. (2024). ChatGPT kann auch „coachen“!? – Rechenpower und Algorithmen statt Sinn oder Bedeutung. *Organisationsberatung, Supervision, Coaching*, 31(1), 97–112.
<https://doi.org/10.1007/s11613-023-00858-8>
- Stone, P., Brooks, R., Brynjolfsson, E., Calo, R., Etzioni, O., Hager, G. et al. (2022). Artificial Intelligence and Life in 2030: The One Hundred Year Study on Artificial Intelligence. [object Object]. <https://doi.org/10.48550/ARXIV.2211.06318>
- Sweller, J. (1994). Cognitive load theory, learning difficulty, and instructional design. *Learning and Instruction*, 4(4), 295–312. [https://doi.org/10.1016/0959-4752\(94\)90003-5](https://doi.org/10.1016/0959-4752(94)90003-5)
- Valmeekam, K., Marquez, M., Olmo, A., Sreedharan, S. & Kambhampati, S. (2023, November 25). PlanBench: An Extensible Benchmark for Evaluating Large Language Models on Planning and Reasoning about Change. arXiv. Zugriff am 30.4.2024. Verfügbar unter:
<http://arxiv.org/abs/2206.10498>
- Venkatesh, V. & Davis, F. D. (2000). A Theoretical Extension of the Technology Acceptance Model: Four Longitudinal Field Studies. *Management Science*, 46(2), 186–204.
- Vogelgesang, J., Bleher, J., Krupitzer, C., Stein, A. & Jung, R. (2023). Nutzung von ChatGPT in Lehre und Forschung – eine Einschätzung der AIDAHO-Projektgruppe.
- Wessels, D. (2023). ChatGPT – ein Meilenstein der KI-Entwicklung. *Mitteilungen der Deutschen Mathematiker-Vereinigung*, 31(1), 17–19. <https://doi.org/10.1515/dmvm-2023-0007>
- Weygandt, B. (2021). *Mathematische Weltbilder weiter denken: Empirische Untersuchung des Mathematikbildes von Lehramtsstudierenden am Übergang Schule–Hochschule sowie*

dessen Veränderungen durch eine hochschuldidaktische Mathematikvorlesung. Wiesbaden: Springer Fachmedien Wiesbaden. <https://doi.org/10.1007/978-3-658-34662-1>

Wissenschaft Aktuell. (2023). *Pflegezeitschrift*, 76(4), 54–55. <https://doi.org/10.1007/s41906-023-2020-y>

Zander, S. (2010). *Motivationale Lernvoraussetzungen in der Cognitive Load Theory: zwei Studien zum Einfluss motivationaler Lernvoraussetzungen auf die kognitive Belastung beim Lernen mit unterschiedlichen Instruktionsdesigns (Wissensprozesse und digitale Medien)*. Berlin: Logos-Verl.

Züll, C. & Menold, N. (2014). Offene Fragen. In N. Baur & J. Blasius (Hrsg.), *Handbuch Methoden der empirischen Sozialforschung* (S. 713–719). Wiesbaden: Springer Fachmedien Wiesbaden. https://doi.org/10.1007/978-3-531-18939-0_52

8 Tabellenverzeichnis

Tabelle 1 Prinzipien gemäss Mayer (2009) zur effektiven Gestaltung von Lerninhalten	11
Tabelle 2 Selbsterstellte Items	18
Tabelle 3 Einblick in den Fragebogen enthaltende «keine» Instruktion	19
Tabelle 4 Einblick in den Fragebogen enthaltende «ausführliche» Instruktion	19
Tabelle 5 "In welchem Einsatzgebiet nutzen Sie LLM`s"	24
Tabelle 6 "Ein Angebot wie zum Beispiel eine Instruktion zu LLM`s empfinden Sie als?"	25
Tabelle 7 Darstellung der inhaltlichen Gemeinsamkeiten in den Massnahmenformulierungen der Studierenden	26
Tabelle 8 Deskriptivstatistik der ChatGPT Aufgaben 1-10 aufgeteilt in die Gruppen mit und ohne Instruktion	26
Tabelle 9 Deskriptivstatistik der Wahrheitsgehalts Einschätzung Aufgaben 11-15 aufgeteilt in die Gruppen mit und ohne Instruktion	27
Tabelle 10 t-Test für unabhängige Stichproben HS1	28
Tabelle 11 95% Konfidenzintervall für die Mittelwerte der Fähigkeiten der Studierenden, aufgeteilt in die Gruppen mit und ohne Instruktion	28
Tabelle 12 t-Test für unabhängige Stichproben HS2	29
Tabelle 13 95% Konfidenzintervall für die Mittelwerte der Nutzungsabsicht der Studierenden, aufgeteilt in die Gruppen mit und ohne Instruktion	30
Tabelle 14 Spearman-Korrelation HS3	31

9 Abbildungsverzeichnis

Abbildung 1. Beschreibung der Frucht Ananas selbst erstellter Prompt mit ChatGPT	7
Abbildung 2. Beschreibung der Frucht Mango selbst erstellter Prompt mit ChatGPT	7
Abbildung 3. Aufgabe 1 aus dem Fragebogen	20
Abbildung 4. Wahrheitsgehalt Aufgabe 11 aus dem Fragebogen	20
Abbildung 5. Alter der Teilnehmenden eingeteilt in Klassen und Häufigkeiten	22
Abbildung 6. Anzahl Häufigkeiten der Verwendung der Studierenden	23
Abbildung 7. Nutzungshäufigkeit der Studierenden in Anzahl Prozenten	24

10 Anhang

10.1 Anhang A- Fragebogen

1 Demografische Angaben

Wie alt sind Sie? Bitte geben Sie Ihr Alter in ganzen Jahren an (Beispiel: 25 Jahre).

Welchem Geschlecht ordnen Sie sich zu?

- Weiblich
- Männlich
- Divers
- Keine Angaben

An welcher Hochschule studieren Sie?

Welche Studienrichtung studieren Sie?

2 Fragen zur Anwendung von ChatGPT

Wie häufig nutzen Sie LLM's?

LLM's (Large Language Models) wie ChatGPT oder Google Gemini (ehemals Google Bard) sind Sprachmodelle die auf Deep-Learning-Algorithmen basieren

- Täglich
- Wöchentlich
- Selten
- Nie
- Keine Angaben

Welches LLM nutzen Sie am häufigsten?

- ChatGPT-3.5 (Gratisversion von ChatGPT)
- ChatGPT-4 (Bezahlversion von ChatGPT)
- Google Gemini (ehemals Google Bard)
- Ich nutze keine LLM's
- Ich nutze ein anderes LLM am häufigsten:

In welchem Einsatzgebiet nutzen Sie LLM's?

- Bildung
- Job
- Privat
- Anderes
- Ich nutze LLM's nicht

Welche Massnahmen würden Sie sich an Ihrer Hochschule zur Implementierung und Nutzung von LLM's wie ChatGPT wünschen?

Ein Angebot wie zum Beispiel eine Instruktion zu LLM's empfinden Sie als?

- Extrem hilfreich

- Sehr hilfreich
 - Ziemlich hilfreich
 - Wenig hilfreich
-
- Nutzlos

3 Selbstwirksamkeit

Folgend werden Ihnen drei Items präsentiert, die darauf abzielen, Ihre individuelle Selbstwirksamkeit zu erfassen.

Die folgenden Aussagen können mehr oder weniger auf Sie zutreffen. Bitte geben Sie bei jeder Aussage an, inwieweit diese auf Sie persönlich zutrifft.

trifft gar nicht zu (1) trifft wenig zu (2) trifft etwas zu (3) trifft ziemlich zu (4)

trifft voll und ganz zu (5)

- In schwierigen Situationen kann ich mich auf meine Fähigkeiten verlassen.
- Die meisten Probleme kann ich aus eigener Kraft gut meistern.
- anstrengende und komplizierte Aufgaben kann ich in der Regel gut lösen.

4 Filter Filter, Kurzinstruktion

c_0001 Benutzerdefinierte Variable	Benutzerdefinierte Variable - Benutzerdefinierte Variable (von Seite : System)	gleich 0
------------------------------------	---	----------

Keine Instruktion

Folgend erhalten Sie eine kurze Beschreibung von LLM's am Beispiel von ChatGPT. Lesen Sie die Beschreibung bitte aufmerksam durch.

ChatGPT (Chat Generative Pre-trained Transformer) ist ein Chatbot, der von OpenAI entwickelt und am 30. November 2022 lanciert wurde. ChatGPT beantwortet Fragen, kann Texte übersetzen, zusammenfassen oder vereinfachen und Programmiercode erstellen. ChatGPT verhält sich dabei wie ein sehr höfliches und intelligentes Gegenüber.

Filter Filter, ausführliche Instruktion

c_0001 Benutzerdefinierte Variable	Benutzerdefinierte Variable - Benutzerdefinierte Variable (von Seite : System)	gleich 1
------------------------------------	---	----------

Ausführliche Instruktion

Im Folgenden erwartet Sie ein rund 3-minütiges Video "ChatGPT einfach erklärt", sowie anschliessend 5 Folien mit zusammenfassenden und ergänzenden Informationen zu ChatGPT.

Es ist wichtig, dass Sie die Tonspur des Videos hören, stellen Sie nun bitte sicher, dass Sie die Möglichkeit dazu haben. Sehen Sie sich das Video und die Folien bitte aufmerksam an und versuchen Sie den Inhalt zu verstehen.

YouTube Video: <https://www.youtube.com/watch?v=fYUaTA9dTI>

Zusammenfassung Video

- ChatGPT wurde mit hunderten Millionen im Internet verfügbarer Texte trainiert
- ChatGPT lernte dabei die «Bedeutung» von Wörtern anhand des Kontextes der Wörter
- Bei einer Eingabe (Prompt) errechnet ChatGPT die Wahrscheinlichkeiten von Wörtern, die auf die Wörter der Eingabe folgen könnten und antwortet mit den wahrscheinlichsten Wörtern

Implikationen 1/2

- ChatGPT kennt keine Fakten, sondern beantwortet Fragen mit der statistisch wahrscheinlichsten Antwort
- Je häufiger ein Wort in den Trainingsdaten vorhanden ist, desto besser kann ChatGPT mit diesem Wort umgehen
 - Englische Wörter oder Ziffern kommen sehr häufig in den Trainingsdaten vor, entsprechend kann ChatGPT gut mit Englisch und Ziffern umgehen
 - neue Wortschöpfungen («rizz»), seltene Zahlen («7'482») oder aktuelle Daten («18. März 2024») kennt ChatGPT nicht, was die Berechnung folgender Wörter erschwert

Implikationen 2/2

- Je konsistenter der Kontext eines Wortes in den Trainingsdaten ist, desto einfacher fällt ChatGPT die Vorhersage der folgenden Wörter
 - Formale Sprachen wie Programmiersprachen und Mathematik weisen eine klare Syntax auf, ChatGPT hätte folglich keine Probleme damit vorherzusagen, dass in «5 * (x + 2)» eine Klammer fehlt
 - Anreden und Grussformeln in E-Mails sind oft ähnlich, ChatGPT würde es nicht schwer fallen das Wort, welches in einem E-Mail auf «Freundliche» folgt («Grüsse») zu berechnen

Fehlinformationen

- ChatGPT ist nur so gut wie die Daten auf welchem es gelernt wurde, sprich falsche Inhalte werden übernommen
- ChatGPT kann keine aktuelle Informationen aus dem Internet abrufen, sondern hat nur die Daten, mit welchen es trainiert wurde
- ChatGPT kann Unsicherheiten kommunizieren, wenn es etwas nicht weiss, es kann aber auch eine überzeugend klingende Antwort liefern, die falsch ist (Halluzination)

Sprache vs. Fakten

- ChatGPT hat mit Millionen von Texten sehr gut gelernt, mit Sprache umzugehen und schreibt grammatikalisch und stilistisch auf hohem Niveau
- Texte von ChatGPT wirken aufgrund der hohen sprachlichen Qualität sehr gut, was aber nicht bedeutet, dass die Informationen und Fakten im Text korrekt sind
- ChatGPT ist demnach nützlich für Zusammenfassungen, alternative Formulierungen oder zur Verringerung der Komplexität eines Textes
- ChatGPT ist weniger nützlich zur Wiedergabe von Fakten, zur Erläuterung einer Theorie oder zur Wiedergabe der Quelle eines bestimmten Arguments

5 Nutzungsabsicht

Nutzungsabsicht und wahrgenommener Nutzen

Bitte beantworten Sie folgende Items anhand der gezeigten Skala von 1 (trifft gar nicht zu) bis 5 (trifft voll und ganz zu). Denken Sie beim Begriff "LLM" an jenes LLM, das Sie am häufigsten nutzen (z.B. ChatGPT).

trifft gar nicht zu (1), trifft wenig zu (2), trifft etwas zu (3), trifft ziemlich zu (4), trifft voll und ganz zu (5)

- Ich beabsichtige LLM's (weiterhin) zu nutzen.
- Ich sage voraus, dass ich LLM's (weiterhin) nutzen werde.
- Die Nutzung von LLM's verbessert meine Leistungen im Studium.

- Die Nutzung von LLM's erhöht meine Produktivität im Studium.
- Die Nutzung von LLM's steigert meine Effizienz im Studium.
- Ich finde LLM's nützlich für mein Studium.

6 Einleitung Fragen

Bewertung von Nutzungsszenarien

Folgend werden Ihnen 10 Szenarien gezeigt. Bitte stellen Sie sich jeweils vor, dass Sie ein LLM wie ChatGPT zur Beantwortung der jeweiligen Frage resp. zur Ausführung des jeweiligen Auftrags nutzen.

Bewerten Sie bei jeder Frage resp. bei jedem Auftrag, wie hoch Sie die Nützlichkeit vom LLM zur Beantwortung der jeweiligen Frage resp. des jeweiligen Auftrags empfinden. Bewerten Sie die Nützlichkeit vom LLM dabei jeweils anhand der gezeigten Skala von 1 (überhaupt nicht nützlich) bis 5 (sehr nützlich).

Nehmen Sie sich bitte ausreichend Zeit für eine sorgfältige Bewertung.

Aufgabe 1:

Stellen Sie sich vor, Sie stellen dem LLM (z.B. ChatGPT) folgende Frage:

Welches sind die aktuellen Ständeratsmitglieder des Kantons Schaffhausen?

Wie hoch schätzen Sie die Nützlichkeit von ChatGPT zur Ausführung des Auftrags ein?

- überhaupt nicht nützlich (1)
- 2.
- 3.
- 4.
- sehr nützlich (5)

Aufgabe 2:

Stellen Sie sich vor, Sie geben dem LLM (z.B. ChatGPT) folgenden Auftrag:

Bitte übersetze den folgenden Satz aus dem Estnischen ins Deutsche: "Okkad torkavad aeglaselt, kui kõnnite aeglaselt."

Wie hoch schätzen Sie die Nützlichkeit von ChatGPT zur Ausführung des Auftrags ein?

- überhaupt nicht nützlich (1)
- 2.
- 3.
- 4.
- sehr nützlich (5)

Aufgabe 3:

Stellen Sie sich vor, Sie stellen dem LLM (z.B. ChatGPT) folgende Frage:

Was ist das Resultat der folgenden Aufgabe: $12'437 \times 74'893 = ?$

Wie hoch schätzen Sie die Nützlichkeit von ChatGPT zur Beantwortung Ihrer Frage ein?

- überhaupt nicht nützlich (1)
- 2.
- 3.
- 4.

• sehr nützlich (5)

Aufgabe 4:

Stellen Sie sich vor, Sie geben dem LLM (z.B. ChatGPT) folgenden Auftrag:

Formulieren Sie den folgenden Satz so um, dass er einfacher verständlich ist: „Handle nur nach derjenigen Maxime, durch die du zugleich wollen kannst, dass sie ein allgemeines Gesetz werde.“

Wie hoch schätzen Sie die Nützlichkeit von ChatGPT zur Ausführung des Auftrags ein?

- überhaupt nicht nützlich (1)
- 2.
- 3.
- 4.

• sehr nützlich (5)

Aufgabe 5:

Stellen Sie sich vor, Sie geben dem LLM (z.B. ChatGPT) folgenden Auftrag:

Mir gefällt die Idee, Entlohnungen so anzupassen, dass sie dem gesellschaftlichen Nutzen der jeweiligen Tätigkeit entspricht und möchte wissen, was dafür und dagegen sprechen könnte. Formuliere je 3 Gründe die für und gegen diese Idee sprechen.

Wie hoch schätzen Sie die Nützlichkeit von ChatGPT zur Ausführung des Auftrags ein?

- überhaupt nicht nützlich (1)
- 2.
- 3.
- 4.

• sehr nützlich (5)

Aufgabe 6:

Stellen Sie sich vor, Sie stellen dem LLM (z.B. ChatGPT) folgende Frage:

Ich lerne die Programmiersprache Python und möchte wissen, was der folgende Python-Code macht?

```
print(pd.Series([1] * len(df["ProportionTrue"])))
```

Wie hoch schätzen Sie die Nützlichkeit von ChatGPT zur Beantwortung Ihrer Frage ein?

- überhaupt nicht nützlich (1)
- 2.
- 3.
- 4.

• sehr nützlich (5)

Aufgabe 7:

Stellen Sie sich vor, Sie geben dem LLM (z.B. ChatGPT) folgenden Auftrag:

Ersetze zwei Buchstaben im Wort "backen" durch zwei andere Buchstaben, so dass ein neues, bekanntes Wort entsteht.

Wie hoch schätzen Sie die Nützlichkeit von ChatGPT zur Ausführung des Auftrags ein?

- überhaupt nicht nützlich (1)
- 2.
- 3.
- 4.

• sehr nützlich (5)

Aufgabe 8:

Stellen Sie sich vor, Sie geben dem LLM (z.B. ChatGPT) folgenden Auftrag:

Ich schreibe eine wissenschaftliche Arbeit zur Beziehung zwischen Persönlichkeitsmerkmalen und Stress am Arbeitsplatz. Welche Literatur könnte mir bei dieser wissenschaftliche Arbeit helfen? Bitte gebe mir 5 Quellen, jeweils mit Titel, Autor(en) und Jahreszahlen, an.

Wie hoch schätzen Sie die Nützlichkeit von ChatGPT zur Ausführung des Auftrags ein?

- überhaupt nicht nützlich (1)
- 2.
- 3.
- 4.

• sehr nützlich (5)

Aufgabe 9:

Stellen Sie sich vor, Sie stellen dem LLM (z.B. ChatGPT) folgende Frage:

Aus welchem Werk stammt das folgende Zitat? "How is it that anything so remarkable as a state of consciousness comes about as a result of irritating nervous tissue, is just as unaccountable as the appearance of the Djin, when Aladdin rubbed his lamp."

Wie hoch schätzen Sie die Nützlichkeit von ChatGPT zur Beantwortung Ihrer Frage ein?

- überhaupt nicht nützlich (1)
- 2.
- 3.
- 4.

• sehr nützlich (5)

Aufgabe 10:

Stellen Sie sich vor, Sie geben dem LLM (z.B. ChatGPT) folgenden Auftrag:

Folgend sende ich dir zwei Seiten aus einem Fachbuch zur Gesundheitspsychologie. Bitte fasse den Text in 3-4 Sätzen zusammen. [...2-seitiger Text...]

Wie hoch schätzen Sie die Nützlichkeit von ChatGPT zur Beantwortung Ihrer Frage ein?

- überhaupt nicht nützlich (1)
- 2.
- 3.
- 4.

• sehr nützlich (5)

- **Wahrheitsgehalt**

Bewertung des Wahrheitsgehalts

Folgend werden Ihnen 5 Prompts samt Antwort von ChatGPT-3.5 gezeigt. Bitte bewerten Sie den Wahrheitsgehalt (Richtigkeit) der Antworten von ChatGPT für jede Aufgabe auf einer Skala von 1 ('sehr niedrig') bis 5 ('sehr hoch').

Im Normalfall kennen Sie die korrekte Antwort nicht, versuchen Sie einfach anhand der Frage und der Antwort abzuschätzen, wie korrekt die Antwort von ChatGPT-3.5 sein könnte.

Aufgabe 11

Du
 Nenne drei Werke des kenyanischen Philosophen Henry Odera Oruka († 1995). Nenne nur die drei Werke und die Jahreszahlen der Werke.

ChatGPT
 Henry Odera Oruka war ein kenianischer Philosoph, der sich insbesondere auf afrikanische Philosophie konzentrierte. Hier sind drei seiner Werke mit den dazugehörigen Jahreszahlen:

1. "Philosophical Trends in the Seventies" (1979)
2. "Sage Philosophy: Indigenous Thinkers and Modern Debate on African Philosophy" (1990)
3. "Mythologies as African Philosophy: Historical Perspectives and the Crisis of Values in Africa" (1991)

Wie hoch schätzen Sie den Wahrheitsgehalt (Richtigkeit) der Antwort von ChatGPT ein:

- sehr niedrig (1)
- 2
- 3
- 4

sehr hoch (5)

Aufgabe 12

Du
 Nenne die fünf Formen von "Guidance" gemäss dem Paper "Guided Search 6.0" von Jeremy M. Wolfe.

ChatGPT
 Im Paper "Guided Search 6.0" von Jeremy M. Wolfe werden die fünf Formen von "Guidance" wie folgt beschrieben:

1. **Spatial guidance:** Dies bezieht sich auf die Orientierung der Aufmerksamkeit aufgrund von räumlichen Merkmalen wie Farbe, Helligkeit oder Bewegung.
2. **Feature guidance:** Hier wird die Aufmerksamkeit aufgrund spezifischer Merkmale wie Farbe, Form oder Orientierung gelenkt.
3. **Temporal guidance:** Diese Form der Lenkung erfolgt basierend auf zeitlichen Merkmalen wie die Reihenfolge, in der Reize präsentiert werden oder die Dauer eines Reizes.
4. **Semantic guidance:** Die Aufmerksamkeit wird aufgrund semantischer Eigenschaften oder Bedeutung gelenkt, wie zum Beispiel das Suchen nach einem bestimmten Objekt in einer Szene.
5. **Conjunctive guidance:** Hierbei wird die Aufmerksamkeit aufgrund einer Kombination verschiedener Merkmale gelenkt, wie beispielsweise das Suchen nach einem roten Kreis in einer Szene, die sowohl rote als auch nicht-rote Kreise enthält.

Diese verschiedenen Formen von Guidance spielen eine wichtige Rolle bei der Steuerung und Lenkung der visuellen Aufmerksamkeit während der Wahrnehmungsaufgabe.

Wie hoch schätzen Sie den Wahrheitsgehalt (Richtigkeit) der Antwort von ChatGPT ein:

- sehr niedrig (1)
 - 2
 - 3
 - 4
-
- sehr hoch (5)

Aufgabe 13

Du
 Nenne die Hauptstädte aller Inselstaaten im pazifischen Ozean, die im Jahr 2010 weniger als 100'000 Einwohner hatten.

ChatGPT
 Im Jahr 2010 hatten die Hauptstädte der folgenden Inselstaaten im Pazifischen Ozean jeweils weniger als 100.000 Einwohner:

1. Funafuti - Tuvalu
2. Ngerulmud - Palau
3. Majuro - Marshallinseln
4. Palikir - Mikronesien

Bitte beachten Sie, dass sich die Bevölkerungszahlen im Laufe der Zeit ändern können und ich Ihnen daher die Daten von 2010 gegeben habe, die möglicherweise nicht mehr aktuell sind.

Wie hoch schätzen Sie den Wahrheitsgehalt (Richtigkeit) der Antwort von ChatGPT ein:

- sehr niedrig (1)
 - 2
 - 3
 - 4
-
- sehr hoch (5)

Aufgabe 14

Du
 Übersetze die folgende Frage aus dem Hindi ins Englische: क्या सोच रहे हो?

ChatGPT
 Die Frage auf Hindi bedeutet auf Englisch: "What are you thinking?"

Wie hoch schätzen Sie den Wahrheitsgehalt (Richtigkeit) der Antwort von ChatGPT ein:

- sehr niedrig (1)
 - 2
 - 3
 - 4
-
- sehr hoch (5)

Aufgabe 15

Du
 Ich möchte wissen wie man das folgende Wort "kainan", was Party in Tagalog bedeutet, ausspricht. Bitte übersetze das Wort ins internationale phonetische Alphabet (IPA).

ChatGPT
 Das Wort "kainan" (Party in Tagalog) wird im Internationalen Phonetischen Alphabet (IPA) wie folgt transkribiert:

/kai'nan/

Die Transkription gibt an, wie das Wort ausgesprochen wird, wobei jede Buchstabenkombination eine bestimmte Lautfolge darstellt.

Wie hoch schätzen Sie den Wahrheitsgehalt (Richtigkeit) der Antwort von ChatGPT ein:

- sehr niedrig (1)
- 2
- 3
- 4
- sehr hoch (5)

7 Abschluss

Besten Dank für Ihre Teilnahme an dieser Umfrage!

Falls Sie an der Verlosung von einem von zwei Coop-Gutscheinen im Wert von CHF 50.- teilnehmen möchten, können Sie unten Ihre E-Mail-Adresse hinterlassen.

Selbstverständlich ist die Teilnahme an der Verlosung freiwillig.

Filter Filter, ausführliche Instruktion

c_0001 Benutzerdefinierte Variable	Benutzerdefinierte Variable - Benutzerdefinierte Variable (von Seite : System)	gleich 0
------------------------------------	---	----------

Ausführliche Instruktion

Sie waren Teil der Kontrollgruppe und haben demnach nur eine kurze Instruktion zu LLM's erhalten. Falls sie die ausführliche Instruktion sehen möchte können Sie auf "weiter" klicken und es erwartet Sie ein rund 3-minütiges Video "ChatGPT einfach erklärt" mit Tonspur, sowie anschliessend 5 Folien mit zusammenfassenden und ergänzenden Informationen zu ChatGTP.

Falls Sie die ausführliche Instruktion nicht sehen möchten, dann danken wir Ihnen vielmals für die Teilnahme und Sie dürfen den Fragebogen an dieser Stelle verlassen.

10.2 Anhang B Explorative Datenanalyse

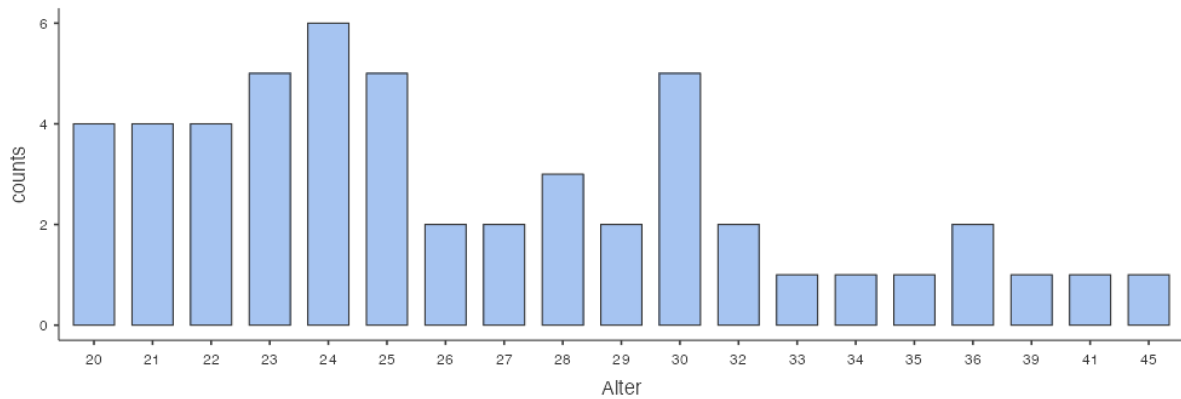


Abbildung 8 Alter der Studierenden, aufgeteilt in Anzahl Personen

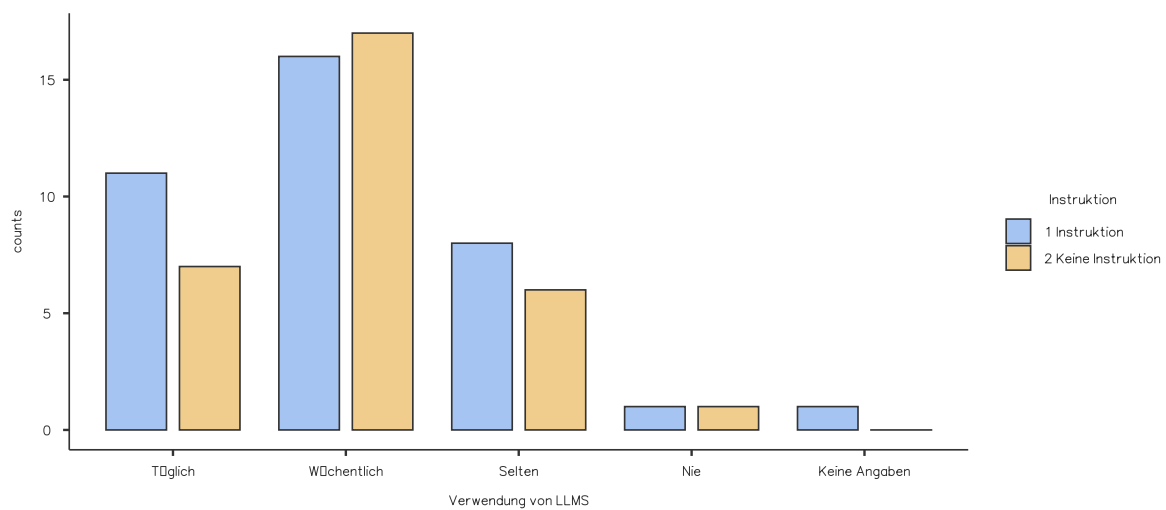


Abbildung 9 Nutzungshufigkeit in Gruppen mit und ohne Instruktion aufgeteilt

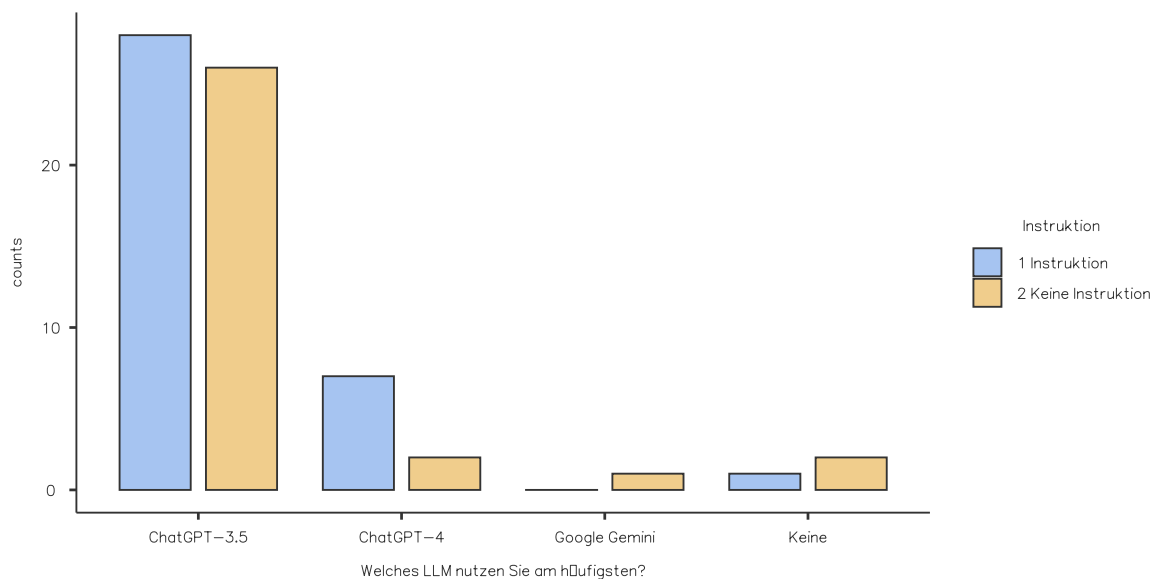


Abbildung 10 Hufigkeit in Gruppen mit und ohne Instruktion aufgeteilt

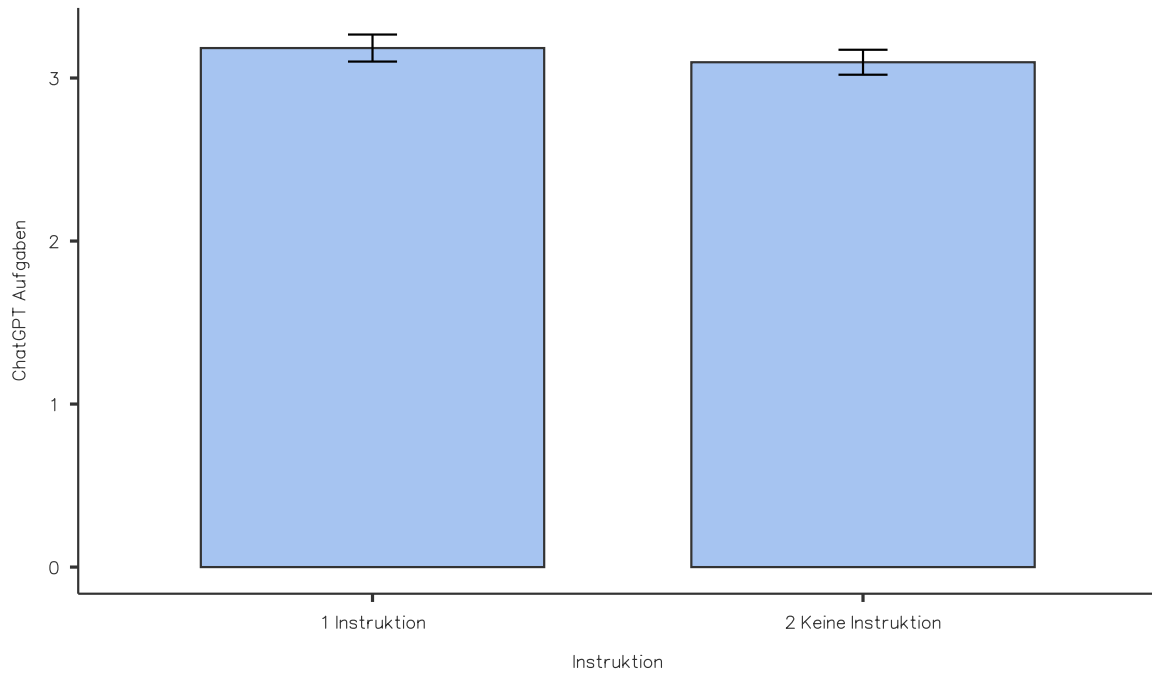


Abbildung 11 ChatGPT-Aufgaben Box-Plot Mittelwert vergleich der Gruppen mit und ohne Instruktion

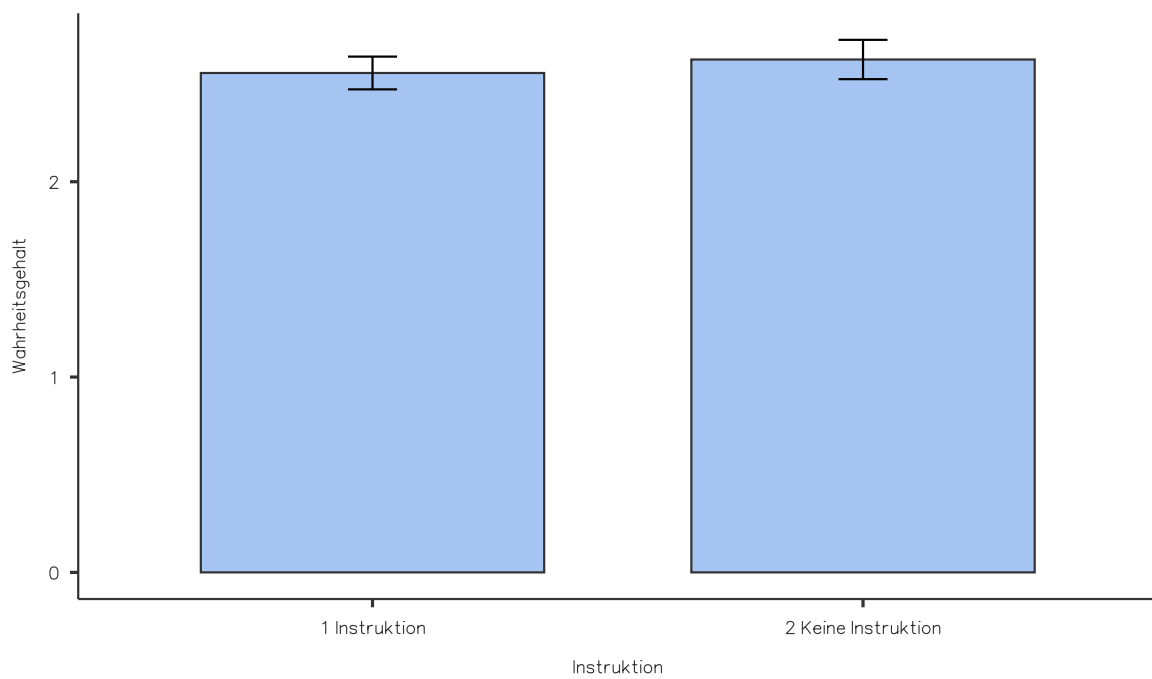


Abbildung 12 Wahrheitsgehalts Ausgaben Box-Plot Mittelwert vergleich der Gruppen mit und ohne Instruktion

Deskriptivstatistik											
	Instruktion	Aufgabe 1 R	Aufgabe 2 R	Aufgabe 3 R	Aufgabe4	Aufgabe5	Aufgabe6	Aufgabe 7 R	Aufgabe 8 R	Aufgabe 9 R	Aufgabe10
N	1 Instruktion	37	37	37	37	37	37	37	37	37	37
	2 Keine Instruktion	31	31	31	31	31	31	31	31	31	31
Mittelwert	1 Instruktion	3.32	2.27	2.22	4.14	3.68	3.73	1.97	3.89	2.65	3.97
	2 Keine Instruktion	3.10	1.94	2.00	4.13	3.58	4.06	2.03	3.35	2.61	4.16

Abbildung 13 Deskriptivstatistik ChatGPT-Aufgaben