



Studienabschlussarbeiten

Sozialwissenschaftliche Fakultät

Steer, Anna:

Vertrauen ist gut, Kontrolle ist besser?
Eine empirische Untersuchung der
Studierendenperspektive auf Vertrauen in KI mit der
reflexiven Grounded Theory

Masterarbeit, Sommersemester 2025

Gutachter*in: von Unger, Hella

Sozialwissenschaftliche Fakultät
Institut für Soziologie

Ludwig-Maximilians-Universität München

<https://doi.org/10.5282/ubm epub.131048>

Abstract

Mit meiner Masterarbeit untersuche ich in einer qualitativen Studie die Studierendenperspektive auf Vertrauen in Künstliche Intelligenz. Dazu verorte ich den Vertrauensbegriff Soziologisch und stelle den Forschungsstil der reflexiven Grounded Theory und den Forschungsprozess dar. Mit den Ergebnissen zeige ich drei Dimensionen des Vertrauens sowie einen Vertrauensprozess, den die Studierenden im Umgang mit KI durchlaufen und entlang dem sich ihr Vertrauen verändert. Deutlich wird ein Spannungsverhältnis und ein individuelles Aushandeln zwischen Vertrauen und Misstrauen und die zentrale Rolle der Eigenverantwortung für die Studierenden bei der Gestaltung ihres Vertrauens.

In my master's thesis, I examine students' perspectives on trust in artificial intelligence through a qualitative study. To this end, I situate the concept of trust within a sociological framework and present the research style of reflexive Grounded Theory as well as the research process. The findings reveal three dimensions of trust and a trust process that students undergo when engaging with AI, along which their trust evolves. The results highlight a tension and individual negotiation between trust and mistrust, as well as the central role of personal responsibility in shaping students' trust.

Inhalt

1.	Relevanz der Studierendenperspektive	1
2.	Theoretische Sensibilisierung	3
2.1.	Soziologische Annäherung an Vertrauen	3
2.2.	Vertrauen zwischen Mensch und Technik	5
2.3.	Vertrauen zwischen Menschen und automatischen Systemen	7
2.4.	Vertrauen in KI.....	9
2.4.1.	Contractual trust.....	9
2.4.2.	<i>Human-like</i> und <i>functionality</i> Vertrauen.....	11
2.4.3.	Paradoxes Vertrauensverhältnis.....	12
3.	Forschungsfrage.....	13
4.	Methodologische Rahmung.....	14
4.1.	Methodisches Vorgehen	14
4.1.1.	Reflexive Grounded Theory	14
4.1.2.	Peerforschung	17
4.2.	Problemzentrierte Interviews	19
4.3.	Datenauswahl und Auswertung	20
4.3.1.	Theoretisches Sampling.....	20
4.3.2.	Auswertung.....	22
4.4.	Forschungsethische Reflexion	23
4.4.1.	Reflexion der Interviewkontakte	23
4.4.2.	Reflexion der Datenverarbeitung	24
4.4.3.	Selbstreflexion	25
5.	Empirische Ergebnisse	26
5.1.	Dimensionen des Vertrauens	29
5.1.1.	Für-wahr-halten und Anvertrauen	30
5.1.2.	Grundhaltung	31
5.2.	Vertrauensprozess.....	33
5.2.1.	Grundhaltung einschätzen.....	34
5.2.2.	Einflüsse auf das Vertrauen.....	36
5.2.3.	Strategien zum Umgang mit eingeschränktem Vertrauen.....	44
6.	Diskussion	48
6.1.	Diskussion zentraler Ergebnisse	48
6.2.	Diskussion der Limitationen	50
7.	Fazit.....	52
	Literaturverzeichnis.....	53
	Anhang	I
	Leitfäden.....	I
	Informierte Einwilligung	II

1. Relevanz der Studierendenperspektive

Mit der Veröffentlichung von ‚Chat GPT‘ im November 2022 (Open AI, 2025a) konnte die breite Öffentlichkeit erstmals generative Künstliche Intelligenz¹ in Form von Large Language Models (LLMs) nutzen. Diese Form der KI ermöglicht vielseitige Einsatzmöglichkeiten – von einfacher dialogischer Interaktion bis hin zur Bearbeitung komplexer Problemstellungen – bedeutete aber auch einen technologischen Umbruch für viele. Der präzise Output von Chat GPT und die Leistungsfähigkeit brachten nicht nur Faszination und Begeisterung mit sich, sondern auch Skepsis und Sorgen. Deswegen steht die Frage nach den Grenzen der Anwendung, je nach spezifischer Personengruppe und Kontext im Raum. Bereiche, in denen Autorenschaft, Nachvollziehbarkeit und personenbezogene Leistungen von besonderer Relevanz sind, werden durch den Einsatz von KI vor neue Herausforderungen gestellt. Dies betrifft in besonderem Maße den wissenschaftlichen Betrieb und in diesem Zusammenhang auch Hochschulen und Universitäten, die neben der Wissensproduktion auch die Aufgabe haben, Studierende wissenschaftlich auszubilden und zu bewerten.

Der Bezugspunkt für KI in dieser Arbeit sind LLMs, die natürliche Sprache nutzen, um Befehle und Output zu generieren (vgl. Banh & Strobel, 2023, S. 3). Mit der Veröffentlichung des LLM GPT-3.5 von Open AI im November 2022 hat der Begriff KI eine neue Bedeutung im alltagssprachlichen Gebrauch bekommen. Die KI-Modelle von Open AI, genannt GPTs (Generative Pretrained Transformer), gibt es mittlerweile in unterschiedlichen Versionen. Die KI-Modelle sind als Grundgerüst in diversen KI-Anwendungen, von Chat Bot, Recherche Tool über Schreibassistenten, eingebaut. Ist also in dieser Arbeit die Rede von KI, geht es nicht um ein spezifisches KI-Modell sondern vergleichbare, hoch entwickelte, frei verfügbare textgenerierende KI-Anwendungen, die auf LLMs beruhen. Beispiele hierfür sind Chat GPT, Perplexity oder Copilot.

Die unmittelbare Verfügbarkeit von Chat GPT im Herbst 2022 führte dazu, dass weder Hochschulen noch Studierende adäquat auf eine sinnvolle Nutzung vorbereitet waren und fehlende verbindliche Richtlinien maßgeblich den Einbezug von KI in wissenschaftliche Aufgaben erschweren. Wie mehrere Meinungsumfragen herausfanden (vgl. Der Spiegel, 2023; vgl. Murašov, 2023) äußerten Studierende den Bedarf an klaren Orientierungshilfen, um KI reflektiert und verantwortungsvoll nutzen zu können. Mittlerweile haben einige Fakultäten und Lehrstühle eigene Richtlinien zur erlaubten Nutzung von KI im Studium erlassen – mit dem Ergebnis, dass nun ein Flickenteppich an Regelungen

¹ Im Folgenden als KI bezeichnet

und Verboten existiert und es kaum übersichtlicher geworden ist. Die Studierenden bewegen sich nach wie vor in einem Spannungsfeld zwischen Neugier, verschiedenen Möglichkeiten und Unsicherheit, welche Anwendungen von KI im Studium zulässig sind und wie sie KI sinnvoll nutzen können. Es besteht kein etablierter, geschweige denn kontrollierter Umgang mit KI. Der Aushandlungsprozess im Studium wird zur individuellen Aufgabe der Studierenden. Woran orientieren sich Studierende also bei der Frage, ob und wie sie KI wissenschaftlich einsetzen können? Über den wissenschaftlichen Kontext hinaus stellt sich die Frage: Lässt sich eine verantwortungsvolle Anwendung von KI etablieren? Besonders für eine junge, technikaffine Generation sind neue technische Entwicklungen von Interesse. Viele sind motiviert, KI auszuprobieren und suchen Wege KI gezielt für sinnvolle Anwendungen einzusetzen. Der Studiumskontext der Teilnehmenden ist relevant, da diese Gruppe mit den technischen Möglichkeiten der KI konfrontiert wird, während sie in ihrem Studium die Ausbildung zum wissenschaftlichen Arbeiten durchlaufen. Das Studium ist also ein zentraler Zugang zur Anwendung von KI. Für Studierende stellt sich dabei die Herausforderung durch die Nutzungsmöglichkeiten die KI selbst einzuordnen. Das Spannungsfeld zwischen Chancen und Risiken macht die Studierendenperspektive dabei zu einem interessanten Untersuchungsfeld. Für die Aushandlung von verantwortungsvoller Verwendung spielt das Vertrauen in die KI eine Schlüsselrolle. Ein reflektierter Umgang mit KI setzt voraus, das eigene Vertrauen in diese Technologie sowie die zugrundeliegende Einschätzung und Bewertung kritisch zu hinterfragen. Mit der Untersuchung des Vertrauens kann ein differenziertes Bild des Zusammenspiels aus Verantwortung und Verwendung der KI hergestellt werden. Ziel dieser Arbeit ist es also zu zeigen, wie Studierende ihr Vertrauen in KI wahrnehmen, begründen und reflektieren. Kann ein verantwortungsbewusster Umgang mittels der Aushandlung von Vertrauen von den Studierenden geschaffen werden? Durch leitfadengestützte problemzentrierte Interviews wird der Fokus der Studierenden erfasst und es wird analysiert, wie die Befragten ihr Vertrauen in KI beschreiben und einordnen.

In meiner Masterarbeit stelle ich zuerst meine sensibilisierenden Konzepte vor. Grundlage ist eine soziologische Annäherung an das Thema Vertrauen in Technik ergänzt durch aktuelle Forschung zu Vertrauen in KI. Aus dieser Auseinandersetzung ergibt sich die Forschungsfrage: Wie äußern sich Studierende zu Vertrauen in KI? Anschließend stelle ich meinen methodischen Rahmen und meinen Forschungsstil vor. Dieses Kapitel umfasst außerdem die Beschreibung der Datenerhebung und -auswahl sowie eine forschungsethische Reflexion. Im Ergebniskapitel fasse ich die zentralen Auswertungsergebnisse zusammen und präsentiere sie. Die Einteilung des Vertrauens in drei Dimensi-

onen und der Vertrauensprozess werden dabei anhand ausgewählter Interviewausschnitte veranschaulicht. Zum Abschluss diskutiere ich die Ergebnisse meiner Arbeit in Hinblick auf die zu Beginn vorgestellten theoretischen Konzepte und gehe auf die Limitationen der durchgeführten Forschung ein.

2. Theoretische Sensibilisierung

Im Folgenden erläutere ich verschiedene Vertrauensdefinitionen und theoretische Konzepte, die ich im Rahmen des Forschungsprozess zur theoretischen Sensibilisierung herangezogen habe. Ein Verständnis für die Eigenschaften und Besonderheiten von Vertrauen in KI erarbeite ich, indem ich, beginnend bei einer allgemeinen Vertrauensdefinition über Konzepte des Vertrauens in Technik und automatische Systeme, ein Verständnis für Vertrauen in KI herleite. Anschließend stelle ich theoretische Ansätze und empirische Arbeiten vor, die insbesondere die studentische Perspektive des Vertrauens in KI aufgreifen.

2.1. Soziologische Annäherung an Vertrauen

Bei der Auseinandersetzung mit dem Forschungsinteresse, dem Vertrauen von Studierenden in KI, möchte ich zunächst einen Überblick über zentrale Vertrauenskonzepte geben. Um mich dem Vertrauen in KI anzunähern, ist es notwendig, mich zunächst mit allgemeinen und schließlich spezifischen Definitionen von Vertrauen auseinanderzusetzen.

Zu Beginn zeigt sich, dass es keine allgemeingültige Definition von Vertrauen gibt. Ein erster Blick richtet sich auf den entsprechenden Beitrag aus den Grundbegriffen der Soziologie. So definiert Endreß den Begriff dort wie folgt:

„Generell steht der Begriff des Vertrauens in der Soziologie für eine Beschreibung und Erklärung der Genese und der Form von sozialen Verhältnissen unter dem Blickwinkel der Ausgestaltung wechselseitigen Aufeinander-verlassen-Könnens. Typischerweise verweist Vertrauen aufgrund vergangener Erfahrungen in den Konstellationen einer Gegenwart auf Zukünftiges in Form mehr oder weniger realisierter, reflexiv verfügbarer Erwartungen.“ (Endreß, 2024, S. 552)

Hier werden mehrere Eigenschaften deutlich: Vertrauen ist eine soziale Angelegenheit, betrifft dadurch mindestens zwei in Beziehung zueinander stehende Akteure. Deutlich wird auch, dass Vertrauen etwas mit Erfahrungen aus der Vergangenheit und Erwartungen an die Zukunft zu tun hat.

Es findet ein Prozess des ‚sich auf etwas verlassen‘ statt. Eine Besonderheit in der Mensch-KI-Interaktion ist, dass die hier beschriebenen Eigenschaften und Akteure nicht wechselseitig sind. Die Untersuchung des Vertrauens in diesem Zusammenhang fokussiert stets den Menschen als vertrauenden Akteur, der sich auf die KI als Objekt des Vertrauens bezieht. Diese Rollen werden nicht umgekehrt und es kommt auch nicht zu einem gegenseitigen Vertrauen. Um sich also dem Vertrauen von Studierenden in KI anzunähern, ist es erforderlich, Vertrauensdefinitionen zu betrachten, die nicht-menschliche Akteure als mögliche Vertrauensobjekte berücksichtigen.

An dieser Stelle ist hervorzuheben, dass unterschiedliche theoretische Ansätze jeweils eigene Strömungen und Vertrauenskonzepte hervorgebracht haben. Neben den hier vorgestellten Definitionen existieren zahlreiche weitere Perspektiven auf Vertrauen in Technik. Dazu zählen unter anderem Zugänge der Strukturierungstheorie (vgl. Giddens, 1995), sowie systemtheoretische Ansätze (vgl. Luhmann, 2014; vgl. Lukyanenko et al., 2022), die im Folgenden nicht vertieft werden, aber dennoch wertvolle theoretische Perspektiven bieten und zur Sensibilisierung bei der Untersuchung von Vertrauen in KI geeignet sind. Die hier verwendete zentrale Definition von Mayer et al. (1995) stammt aus der Organisationssoziologie, ist jedoch richtungsweisend in der Techniksoziologie für anschließende Forschung zu Vertrauen in Technik und automatischen Systemen.

Mit dem Vertrauenskonzept von Mayer et al. (1995) lässt sich Vertrauen nicht nur zwischenmenschlich (*interpersonelles Vertrauen*) definieren. Vertrauen wird darüber hinaus von ihnen wie folgt definiert:

„the willingness of a party to be vulnerable to the actions of another party based on the expectation that the other will perform a particular action important to the trustor, irrespective of the ability to monitor or control that other party.“ (Mayer et al., 1995, S. 712)

Das Interessante an dieser Definition ist, dass sie die Perspektive der Vertrauenden widerspiegelt und dabei nicht automatisch eine wechselseitige Vertrauensbeziehung annimmt. Das Vertrauen ergibt sich aus der Erwartungshaltung der Vertrauenden, unabhängig der Möglichkeit der Kontrolle. Der zu vertrauenden Seite wird die Erwartung einer bestimmte Handlung zu vollziehen zugeteilt, ohne diese zu kontrollieren. Damit wird jedes Subjekt, das in der Lage ist eine Handlung zu vollziehen oder am Zutun oder Scheitern einer Handlung beteiligt ist, zum Objekt des Vertrauens. Ein wichtiger Punkt dieser Definition ist außerdem das Wechselspiel aus Erwartung und Kontrolle. Das Vertrauen tritt für die vertrauende Person sowieso erst dann ein, wenn nicht kontrolliert wird, ob die erwartete Handlung eintritt, egal ob die Kontrolle möglich wäre oder nicht. Indem sie vertraut, zeigt sie eine

Bereitschaft das Risiko einzugehen, dass ihre Erwartungen nicht erfüllt werden. Die Vertrauende muss die Verletzlichkeit ihrer Position annehmen.

Anhand ihres Vertrauenskonzepts argumentieren Mayer et al. allerdings auch, dass Vertrauen nicht nur durch Eigenschaften der Vertrauenden betrachtet werden kann. Um die Vielschichtigkeit von Vertrauen zu begreifen, müssen auch die Eigenschaften des Vertrauensobjekts miteinbezogen werden. Diese Sichtweise stellen die Autoren als Konzept der Vertrauenswürdigkeit vor (vgl. ebd., S. 716). Allerdings ist zu beachten, dass die Perspektive stets von den Vertrauenden ausgeht. Die Vertrauenswürdigkeit des Vertrauensobjekts ist eine Interpretationsleistung bzw. Zuschreibung der Vertrauenden, basierend auf bestimmten Eigenschaften des Objekts. Als Eigenschaften für gesteigerte Vertrauenswürdigkeit nennen Mayer et al. zuerst den Eindruck fähig zu sein, eine erwartete Handlung zu vollziehen: „Ability“ (ebd., S.717). Außerdem nennen sie die Einschätzung, wie wohlwollend das Vertrauensobjekt dem Vertrauenden gegenüber ist: „Benevolence“ (ebd., S. 718). Und schließlich die Eigenschaft „Integrity“ (ebd., S. 719), die dem Vertrauensobjekt zugesprochen wird, dass es sich an bestimmte Regeln hält.

Das Objekt des Vertrauens und seine Eigenschaften der Vertrauenswürdigkeit bei Mayer et al. sind stets allgemein formuliert. Dieses Vertrauenskonzept lässt sich, wie bereits beschrieben, auf jegliche menschliche, aber auch nicht-menschliche Akteure beziehen. Aufgrund eben dieser Besonderheit, nämlich der nicht exklusiv menschlichen Vertrauensbeziehung, wird die Definition von Mayer et al. oftmals für Untersuchungen von Vertrauen zwischen menschlichen und nicht-menschlichen Akteuren herangezogen.

2.2. Vertrauen zwischen Mensch und Technik

Die Vertrauensbeziehung zwischen Menschen und Technik haben McKnight et al. (2011) in ihrem Beitrag untersucht. Sie bauen dabei auf dem Vertrauenskonzept von Mayer et al. (1995) auf und beziehen es konkret auf die Mensch-Technik Beziehung. So entwickeln sie die drei Eigenschaften der Vertrauenswürdigkeit weiter angepasst auf technische Vertrauensobjekte. Aus *Ability* wird „Functionality“ (McKnight et al., 2011, S. 12:5). Die Nutzenden bewerten die Vertrauenswürdigkeit von Technik, anders als Vertrauenswürdigkeit von Menschen, anhand der vorhandenen Funktionen, und ob diese geeignet sind eine Aufgabe zu erfüllen. Die Eigenschaft *Benevolence* wird von den Autor*innen in „Helpfulness“ (ebd. 2011, S. 12:5) umbenannt. Von Technik erwarten Nutzende keine emotionale oder moralische Verpflichtung, sondern einen klaren Nutzen. So bezieht sich die Bewer-

tung der Vertrauenswürdigkeit auf den Aspekt, wie hilfreich die Technologie ist. Die dritte Eigenschaft *Integrity* bezieht sich darauf, dass sowohl Menschen als auch Technik als Vertrauensobjekt verlässlich sind. Bei Technik entsteht die Verlässlichkeit dadurch, dass sie bei Nutzung konsistente und vorhersagbare Reaktionen zeigt. Deshalb wird diese Eigenschaft umbenannt in „Reliability“ (ebd., S. 12:5).

Auf dieser theoretischen Grundlage haben die Autor*innen eine quantitative Studie mit 359 Studierenden durchgeführt (vgl. ebd., S. 12:10). McKnight et al. konnten mit ihrer Studie eine Unterscheidung zwischen anfänglichem und wissensbasiertem Vertrauen feststellen (vgl. ebd., S. 12:13). Anfängliches Vertrauen in Technologie begründet sich demnach durch Annahmen und Kosten-Nutzen Abschätzungen, wohingegen wissensbasiertes Vertrauen auf individuellen Erfahrungen mit spezifischen Technologien beruht (vgl. ebd., S. 12:13). Anschließend an bestehende Forschung konnten McKnight et al. bestätigen, dass anfängliches Vertrauen gegenüber unbekannten Technologien eintritt, wenn diese mit institutionellen Mechanismen oder bereits Bekanntem assoziiert werden (vgl. ebd., S. 12:13). Handelt es sich allerdings um wissensbasiertes Vertrauen, werden weniger institutionelle Überzeugungen herangezogen, sondern Vertrauensentscheidungen anhand Charakteristika der Technologie selbst getroffen (vgl. ebd., S. 12:13).

Die Mensch-KI-Beziehung kann als spezielle Form der Mensch-Technik-Beziehung betrachtet werden. Die Unterscheidung in anfängliches und wissensbasiertes Vertrauen könnten auch für das Vertrauen in KI Anwendungen eine Rolle spielen. Die zum Zeitpunkt der Datenerhebung noch relativ neuen KI-Chatbots könnten bei unerfahrenen Nutzenden anfängliches Vertrauen hervorrufen, indem sie Assoziationen mit bereits bekannten Chatbots herstellen. Gleichzeitig könnte bei erfahrenen Nutzenden wissensbasiertes Vertrauen entstehen, da diese bereits praktische Erfahrungen mit der KI gesammelt haben.

Dennoch ist festzuhalten, dass die KI spezifische Merkmale in die Mensch-Technik-Beziehung mitbringt und es daher einen genaueren Blick auf die Besonderheit in der Vertrauensbeziehung bedarf. Als weitere Annäherung von Technik zu KI, geht es um Vertrauen in Automatisierung. Anders als bei allgemeiner Technik wird hier das Vertrauen in automatisch ablaufende Prozesse betrachtet. Die Besonderheit dabei ist die Selbststeuerung der Technik, die ohne menschliches Zutun einen Prozess vollzieht. Der Mensch ist im Wesentlichen weniger involviert in das Gelingen oder Scheitern einer Handlung oder eines Outcomes.

2.3. Vertrauen zwischen Menschen und automatischen Systemen

Lee und See (2004) gehen einen Schritt weiter, indem sie Vertrauen in Automatisierung genauer untersuchen. Sie entwickeln eine Vertrauensdefinition aufbauend auf Mayer et al. (1995) wie folgt: „[Trust] is the attitude that an agent will help achieve an individual's goal in a situation characterized by uncertainty and vulnerability“ (Lee & See, 2004, S. 54).

Diese Definition wird von Hoff und Bashir (2015) in Kombination mit den von Mayer et al. (1995) entwickelten Eigenschaften der Vertrauenswürdigkeit - *Ability*, *Benevolence* und *Integrity* - herangezogen, um Vertrauen als dynamisches Zusammenspiel zwischen Automaten und Menschen zu konzeptualisieren (Hoff & Bashir, 2015, S. 410). Darüber hinaus argumentieren sie mit dem Hinweis auf die Definition von Lee und See (2004), dass Vertrauen hauptsächlich von Emotionen bestimmt wird und sich daher nicht immer rational erklären lässt (Hoff & Bashir, 2015, S. 410). Auf Basis dieser theoretischen Einordnung des Vertrauenskonzepts führen Hoff und Bashir (2015) eine Literaturanalyse mit 127 Studien zum Thema Vertrauen in automatische Systeme durch.

Die Ergebnisse der Analyse lassen sich in ein dreischichtiges Vertrauensmodell einordnen. Das ursprünglich von Marsh und Dibben (2003) entwickelte Modell entstand aus einer multidisziplinären Kategorisierung von Vertrauen. Eine Kombination aus philosophischen, soziologischen und psychologischen Erklärungsansätzen für Vertrauen ermöglichte die Entwicklung der dreischichtigen Vertrauensebenen (Marsh & Dibben, 2003, S. 468). Die Autoren unterscheiden in ihrem Modell in „dispositional“, „situational“ und „learned trust“ (ebd., S. 469).

Dispositional trust bezeichnet die individuelle Tendenz, einem automatischen System zu vertrauen, unabhängig vom spezifischen System und der spezifischen Situation. Dabei variiert die Veranlagung zu Vertrauen u.a. je nach kultureller Zugehörigkeit, Alter, Geschlecht und Persönlichkeit (Hoff & Bashir, 2015, S. 414–415). Hoff und Bashir weisen darauf hin, dass Unterschiede in veranlagtem Vertrauen bestehen, z.B. zwischen Generationen, auch wenn sich einzelne Merkmale, z.B. das Alter einer Person, mit der Zeit verändern. Für eine Person bleibt das veranlagte Vertrauen innerhalb einer Interaktion stabil (vgl. ebd., S. 415). *Dispositional trust* spielt also als situationsunabhängige Dimension des Vertrauens eine große Rolle, bleibt (relativ) stabil über den Zeitraum der Interaktion und kann als persönlich verankertes ‚Grundvertrauen‘ bezeichnet werden.

Situational trust bezeichnet die Entwicklung von Vertrauen, abhängig von der konkreten Situation. Als zwei gegenüberstehende Einflüsse auf das situationsabhängige Vertrauen haben Hoff und Bashir das externe Umfeld und den internen, Kontext abhängigen Charakter herausgearbeitet (vgl. ebd., S.

415). Die externen Einflüsse auf das Vertrauen umfassen die Komplexität des automatischen Systems und den Schwierigkeitsgrad der zu lösenden Aufgabe. Außerdem beeinflusst das Umfeld die Einschätzung potenzieller Risiken, die wiederum die Entwicklung von Vertrauen prägt (vgl. ebd., S. 416). Zu den externen Einflüssen zählen Hoff und Bashir auch das Setting der Interaktion, wobei die Selbstpräsentation der Automatisierung und der Ergebnisse, sowie Verteilung der Verantwortung für die Ergebnisse eine große Rolle für das Vertrauen spielen (vgl. ebd., S. 417). Neben den externen Einflüssen formulieren Hoff und Bashir auch interne Eigenschaften, die die Mensch-Automat-Interaktion und das Vertrauen beeinflussen. Die spezifischen Charakterzüge wurden zwar schon mit dem veranlagten Vertrauen abgedeckt, jedoch werden hier kontextabhängige Charaktereigenschaften aufgegriffen. Das Selbstbewusstsein der Nutzenden variiert möglicherweise: Expertenwissen in einem Feld oder die generelle Stimmung in einer Situation beeinflussen das Vertrauen in die Automatisierung (vgl. ebd., S. 417–418). Zusammenfassend lässt sich das situationsabhängige Vertrauen durch Einflüsse aus dem externen Umfeld und den internen, kontextabhängigen Charakteristika der Nutzenden beschreiben.

Learned trust beschreibt Vertrauen basierend auf Erfahrungen. Dabei spielt sowohl existierendes Wissen als auch Erfahrungen mit den Ergebnissen eines spezifischen automatischen Systems eine Rolle (vgl. Hoff & Bashir, 2015, S. 420). Das anfänglich gelernte Vertrauen bezieht sich dabei auf Vertrauen, das sich durch Vorannahmen noch vor der Interaktion mit dem System etabliert hat; dynamisches Vertrauen dagegen entsteht und verändert sich während der Interaktion (vgl. ebd., S. 420). Einflüsse auf das gelernte Vertrauen gehen allerdings über Erfahrungswerte hinaus und schließen Wissen zu Prozessen hinter automatischen Systemen und deren Transparenz mit ein (vgl. ebd., S. 421–423). Ebenso spielen Assoziationen in der Interaktion, Vermenschlichung oder Eigenschaften, die Vertrauen durch vermeintliche Erfahrungen hervorbringen sollen, eine Rolle in der Entstehung des gelernten Vertrauen (vgl. ebd., S. 423).

Zusammenfassend liefert die Betrachtung der drei Ebenen eine klare Struktur, um das komplexe Phänomen Vertrauen in Automatisierung zu erfassen. Jede Ebene birgt spezifische Einflussfaktoren auf die individuelle, situations- und kontextabhängige Vertrauensbeziehungen zwischen den Nutzenden und dem automatischen System. Diese differenzierte Darstellung von Vertrauen in automatischen Systemen ist ebenfalls eine Möglichkeit Vertrauen in KI zu kategorisieren, da KI ein hochentwickeltes, spezifisches automatisches System ist.

2.4. Vertrauen in KI

Nicht jedes automatische System ist eine KI. Die Besonderheit der KI liegt in der begrenzten Transparenz der internen Prozesse, der scheinbaren Menschlichkeit und der breiten thematischen Einsatzmöglichkeiten. Die Aspekte Transparenz und Verstehen der KI gewinnen durch das Black Box-Phänomen an Bedeutung. Es beschreibt den Zustand der KI, dass nur begrenztes Wissen über die Zusammenhänge der internen Prozesse bekannt sind (vgl. Kuhn & Hartmann, 2023, S. 5). Damit ist ein vollständiges Nachvollziehen der KI unmöglich. Eine weitere Besonderheit der KI ist die Vermenschlichung der Interaktion. Indem die KI mit menschlicher Sprache kommuniziert, entsteht ein anthropomorpher Eindruck. Das bedeutet, dass durch die menschenähnliche Kommunikation der KI menschenähnliche Eigenschaften zugesprochen werden, ohne dass sie dies jedoch tatsächlich besitzt (vgl. Felder & Kückelhaus, 2025). Daran knüpft auch der Aspekt der thematischen Breite der KI. Indem sich das Anwendungsfeld der KI nicht auf ein spezifisches Gebiet oder eine Aufgabe begrenzt, kann KI vermeintlich jegliches Thema zusammenfassen, kritisch diskutieren oder darüber philosophieren. Außerdem kann die KI unterschiedliche feldspezifische Ausdruckseisen verwenden (z.B. wissenschaftlich, journalistisch, alltagsweltlich) und ist damit flexible anwendbar. Eine zunehmend entscheidende Rolle im Vertrauensverhältnis mit der KI spielt also nicht nur die eingeschränkte Transparenz, sondern auch die scheinbar menschliche Intelligenz und suggerierte Universalität des automatischen Systems KI.

Im Folgenden stelle ich eine theoretische Annäherung und empirische Studien vor, die sich konkret mit dem Vertrauen in KI auf unterschiedliche Weise und mit unterschiedlichem Fokus beschäftigt haben. Ich werde die relevanten Ergebnisse darstellen und im Anschluss daraus die Forschungsfrage meiner Arbeit herausarbeiten.

2.4.1. Contractual trust

Jacovi et al. (2021) gehen in ihrem Beitrag auf Vertrauen in KI ein, stellen ihre Definition von „contractual trust“ (ebd., S. 625) vor und arbeiten am Konzept der Vertrauenswürdigkeit unterschiedliche Vertrauensgrundlagen heraus. Die Autor*innen nähern sich in ihrem Beitrag an Vertrauen ebenfalls mit der Definition von Mayer et al. (1995) an. Mit dieser Perspektive gehen sie auf die ebenfalls in Kapitel 2.1 besprochenen Eigenschaften der Vertrauensbeziehung, Vulnerabilität, Erwartbarkeit und Risiko (vgl. Jacovi et al., 2021, S. 625) ein. Mit diesem Ausgangspunkt blicken sie auf die Nutzung von KI-Anwendungen und leiten ab, dass Vertrauen in KI mit Erwartungen der Nutzenden in richtige, bzw. korrekte Modelle und ihrem Output einhergeht. Sie argumentieren:

„a more nuanced and accurate view of what ‘trust in model correctness’ refers to: it is in fact not trust in the general performance ability of the model, but that the patterns that distinguish the model’s correct and incorrect cases are available to the user.” (ebd., S. 626)

Das Vertrauen der Nutzenden basiert also auf den Möglichkeiten, die eigene Erwartbarkeit gegenüber der KI an einem erkennbaren Muster von richtigem und falschem Output festzumachen. Die Autor*innen greifen in diesem Zug das aus der Philosophie stammende Konzept des ‚trust with commitment‘ von Hawley (2014) auf: Darin beschreibt sie Vertrauen in jemanden, als Glaube daran, dass diese Person eine Verpflichtung hat und sich an diese Verpflichtung hält (Hawley, 2014, S. 10). Diesen Aspekt der Verpflichtung in der Vertrauensbeziehung verwenden Jacovi et al. nun in ihrer Argumentation für ihre Definition von Vertrauen in KI, sprechen hier aber von *contractual trust* (vgl. Jacovi et al., 2021, S. 626). *Contractual trust* im Kontext der KI bezieht sich also auf die Erwartung, dass ein bestimmtes Muster zur Zuordnung der richtigen und falschen Leistungen eingehalten wird. Sie formulieren ihre Vertrauensdefinition wie folgt:

„Trust, as mentioned, aims to enable the ability to anticipate intended behavior through the belief that a contract will be upheld. Further, (...) the ability to anticipate does not necessarily manifest with the existence of trust; it is possible for a user to trust a model despite their inability to anticipate its behavior.“ (Jacovi et al., 2021, S. 627)

Der vorliegenden Definition zufolge muss die Vereinbarung zwischen KI und Nutzenden explizit sein (vgl. ebd., S. 626). Da KI-Anwendungen, wie Chat GPT, erst nach der Publikation von Jacovi et al. erschienen sind, muss diskutiert werden, ob die vielseitig einsetzbaren KI-Systeme über eine explizite Funktions- bzw. Leistungsabmachung mit den Nutzenden verfügen. In ihrer Definition betonen die Autor*innen, dass Vertrauen auch ohne Abmachung und Erwartbarkeit entstehen kann. Folglich weist die KI damit zwar keine inhärente Vertrauenswürdigkeit auf, dennoch kann Vertrauen entstehen, das als „Unwarranted Trust“ (ebd., S. 627) bezeichnet wird. Wie Jacovi et al. hervorheben, muss diese Art des Vertrauens explizit evaluiert und verhindert oder minimiert werden (ebd., S. 628). Wichtig ist, dass die Unterscheidung in gerechtfertigtes und ungerechtfertigtes Vertrauen, eine Perspektive auf Vertrauen in KI bietet, und diese anhand der wahrgenommenen Vertrauenswürdigkeit der KI bestimmt wird. Die Einschätzung eines vertrauenswürdigen KI-Modells klassifizieren Jacovi et al. außerdem anhand des „Intrinsic Trust“ (ebd., S. 628) und „Extrinsic Trust“ (ebd., S. 629). *Intrinsic Trust* beschreibt das Vertrauen in die inneren Prozesse der KI, also sozusagen der Weg, wie sie zu

ihren Ergebnissen gelangt. Dagegen beruht *Extrinsic Trust* auf der Bewertung der Ergebnisse im Hinblick auf deren Erwartbarkeit, wie sie aus bisherigen Erfahrungen abgeleitet wird. Inwiefern die teilnehmenden Studierenden in der von mir durchgeföhrten Untersuchung die Vertrauenswürdigkeit von KI bewerten, welche Rolle *Intrinsic Trust* und *Extrinsic Trust* dabei spielen und wie sie ihr Vertrauen begründen, soll anhand der Ergebnisse in Kapitel 6.1 diskutiert werden.

2.4.2. *Human-like* und *functionality* Vertrauen

Eine empirische Untersuchung zum Vertrauen in KI wurde von Choung et al. (2022) durchgeführt. Die Autor*innen möchten nicht nur das Vertrauen in KI, sondern auch die Rolle, die Vertrauen bei der Akzeptanz der KI spielt, mit einbeziehen. Dabei nutzen sie das Technology Acceptance Model (TAM), ein theoriegestütztes Modell, dass die Verwendung und Akzeptanz von diversen Technologien in unterschiedlichen Populationen untersucht (vgl. Venkatesh & Davis, 2000 zitiert nach Choung et al., 2022, S. 3). Choung et al. (2022) erarbeiten in ihrem Beitrag die Rolle von Vertrauen als holistisches Konstrukt und betrachten dabei zentrale Vertrauensdimensionen sowie deren Beziehungen zur Akzeptanz von KI (vgl. Choung et al., 2022, S. 3). Aufbauend auf der Definition von Vertrauen in Technologie von Mayer et al. (1995) und den drei Dimensionen des Vertrauens von McKnight et al. (2011), liefern Choung et al. (2022) ein zweiteiliges Vertrauenskonzept. Sie argumentieren, dass KI zwar als Technologie gilt und daher bestehende Definitionen von Vertrauen in Technologie relevant sind, jedoch die Besonderheit von KI, insbesondere die Fähigkeit menschliche Denkaufgaben zu übernehmen und Entscheidungen zu treffen, berücksichtigt werden müssen (vgl. Choung et al., 2022, S. 10). Dieses hohe Maß an Automatismus und die Annäherung an menschliches Handeln sollen im Konzept des Vertrauens in KI explizit mitgedacht werden. Vertrauen in KI wird daher in zwei Dimensionen differenziert: dem „*human-like trust*“ und dem „*functionality trust*“ (Choung et al., 2022, S. 11). Das *human-like trust* bezieht sich auf die sozialen und kulturellen Eigenschaften der KI sowie die zugrundeliegenden ethischen Werte. Das *functionality trust* dagegen bezieht sich auf die Kompetenz, Zuverlässigkeit und Leistung der KI.

Zur Untersuchung haben Choung et al. (2022) zwei quantitative Umfragen mit Online-Fragebögen zu Vertrauen in KI durchgeführt. In der ersten Studie wurden 312 Studierende befragt, in der zweiten Studie wurden die Ergebnisse mit 640 Teilnehmenden US-Bürger*innen reproduziert. Mit ihren Ergebnissen konnten sie ihre Unterscheidung in *human-like trust* und *functionality trust* in Bezug auf KI bestätigen. Obwohl das Vertrauen in die Funktionalität der KI einen stärkeren Einfluss auf

Wahrnehmung der Nützlichkeit, Einstellung und Intention der Nutzung hat, ist die ‚human-like‘ Dimension des Vertrauens besonders bei KI mit menschlichen Eigenschaften und emotionalem Bezug relevant (vgl. ebd., S.20).

Für die Untersuchung des Vertrauens von Studierenden in KI sind beide KI-spezifischen Dimension interessant. Zum einen muss das Vertrauen, das aus der wahrgenommenen Nützlichkeit und der funktionalen Leistungsfähigkeit der KI entspringt, betrachtet werden. Zum anderen spielt aber auch das Vertrauen eine Rolle, das sich auf die menschlichen Eigenschaften der KI bezieht und somit die soziale Dimension einer menschlichen Vertrauensbeziehung adressiert.

2.4.3. Paradoxes Vertrauensverhältnis

Auch Amoozadeh et al. (2024) haben zu Vertrauen von Studierenden in KI geforscht. Ihr Vertrauensverständnis lässt sich mit den zuvor dargestellten Definitionen vereinen. Vertrauen spielt in ihrer Betrachtung eine zentrale Rolle in der Interaktion mit KI und tritt insbesondere in Situationen auf, die von Unsicherheit und Verletzlichkeit geprägt sind. In solchen Momenten wird das Vertrauen durch bestehende Erwartungen sowie die Beurteilung der KI bestimmt (vgl. Amoozadeh et al., 2024, S. 68).

Amoozadeh et al. (2024) legen in ihrer Untersuchung einen besonderen Fokus auf die Verbindung von Vertrauen in KI und dem Nutzen im Studium. Die Befragung konzentriert sich dabei vor allem auf spezifischen Anwendungsbereiche, wie etwa die Unterstützung beim Programmieren. Grundlage der Studie ist eine quantitative Umfragestudie mit 253 Computer Science Studierenden aus den USA und Indien. Die Ergebnisse zeigen einen positiven Zusammenhang zwischen Vertrauen in KI und der Produktivität. Studierende, die ein höheres Maß an Vertrauen in die KI haben, haben auch mehr Motivation im Studium und ein höheres Selbstbewusstsein bezüglich ihrer eigenen Kompetenzen (vgl. Amoozadeh et al., 2024, S. 72). Im Fragebogen war außerdem eine offene Frage enthalten. Diese wurde von insgesamt 114 Studierenden beantwortet. Die Frage richtete sich an ihre generelle Meinung zu KI beim Programmieren. Die häufigsten Themen waren dabei Misstrauen und genereller Nutzen. Besonders interessant ist dabei, dass genau jene Studierende, die über Misstrauen und Skepsis gegenüber der KI berichtet haben, auch diejenigen waren, die die KI am nützlichsten und besonders hilfreich empfanden (vgl. ebd., S. 71). Dieses scheinbare Paradox zwischen Skepsis und Nutzen wirft Fragen nach weiteren Untersuchungen auf.

3. Forschungsfrage

Die vorgestellten theoretischen und empirischen Arbeiten haben mir ermöglicht, mich für das Thema Vertrauen in KI zu sensibilisieren und darauf eine Forschungsfrage zu erarbeiten. Ausgangspunkt war die möglichst allgemeine Vertrauensdefinition von Endreß (2024). Von dort aus habe ich mich spezifischeren Zugängen zu Vertrauen, entwickelt von Mayer et al. (1995), zugewandt. Diese schließen das Vertrauen auch gegenüber nicht-menschlichen Objekten ein. Die Arbeiten von McKnight et al. (2011) und Hoff und Bashir (2015) boten die Grundlage, Vertrauen in Technik und anschließend Vertrauen in automatische Systeme präziser zu definieren. Durch die Abgrenzung von KI zu allgemeiner Technik und Automaten konnte ich drei Eigenschaften identifizieren, die in der Vertrauensbeziehung zu berücksichtigen sind: die eingeschränkte Nachvollziehbarkeit, die vermenschlichte Kommunikation und die thematische Universalität der KI. Einen theoretischen Ansatz zur Beschreibung von Vertrauen in KI, der allerdings vor der Veröffentlichung von Chat GPT etc. erschien, liefern Jacovi et al. (2021). Bisherige empirische Studien zu Vertrauen in KI mit dem Fokus auf die Studierendenperspektive bilden interessante Ergebnisse ab. Choung et al. (2022) untersuchten das Vertrauen in KI. Sie befragten zwar in einer Studie ausschließlich Studierende, bezogen ihre Ergebnisse allerdings nicht auf diese konkrete Perspektive. Anders haben Amoozadeh et al. (2024) den Fokus ihrer Untersuchung auf eine klar abgegrenzte Gruppe an Studierenden einer Fachrichtung gelegt und die Ergebnisse stark auf das Vertrauen für spezifische Aufgaben im Computer Science Studium bezogen.

In dieser Arbeit verfolge ich daher das Ziel, mit der Forschungsfrage eine breitere Perspektive auf das Vertrauen von Studierenden in KI zu erfassen. Das Vertrauen soll als Schlüssel zwischen Reflexion und Nutzen untersucht werden. Es gibt keine Fokussierung auf fachspezifische Aufgaben, sondern es geht um eine allgemeine Einordnung des Vertrauens in KI. Als Ausgangspunkt dient die soziale Position der Studierenden im Feld der Wissenschaft. Die Verbindung der sozialen Position der Studierenden, dem Lernprozess des wissenschaftlichen Arbeitens und die Reflexion neuer Technologien machen das Vertrauen zu einem wesentlichen Indikator für einen verantwortungsbewussten Umgang mit KI. Dieser Umgang beschränkt sich nicht auf den universitären Kontext, sondern betrifft die Studierenden auch in ihrem Alltag oder Arbeitskontext. Außerdem wird nur die Perspektive der Teilnehmenden auf sich selbst und ihre eigene Beschreibung berücksichtigt. Es kann also nur eine Aussage darüber getroffen werden, wie die Studierenden selbst über ihr Vertrauen in KI sprechen. Deshalb lautet die Forschungsfrage dieser Arbeit wie folgt: Wie äußern sich Studierende zu ihrem Vertrauen in KI?

4. Methodologische Rahmung

In diesem Kapitel werde ich den methodischen Rahmen der Forschung zu Vertrauen von Studierenden in KI erläutert. Ziel ist es, das methodische Vorgehen sowie Forschungsentscheidungen nachvollziehbar darzustellen. Im Folgenden wird zuerst der Forschungsstil der reflexiven Grounded Theory (RGTM) nach Breuer et al. (2019) beschrieben und anschließend auf die Datenerhebung mit leitfadengestützten, problemzentrierten Interviews nach Witzel und Reiter (2022) eingegangen. Daraufhin erkläre und begründe ich die Datenauswahl und den Auswertungsprozess. Abschließend folgen forschungsethische Reflexionen der Arbeit.

Da diese Arbeit an das Lehrforschungsprojekt ‚Mit KI über KI qualitativ forschen‘ anschließt, werde ich einen Teil der dort erhobenen Daten in dieser Arbeit verwenden. Im Lehrforschungsprojekt haben Kommiliton*innen und ich KI-Tools für qualitative Sozialforschung getestet und die Einstellungen von Studierenden zu KI erforscht. Das Forschungsinteresse des Projekts lag sowohl auf der Anwendung der KI-Tools als auch auf den subjektiven Eindrücken über, und Perspektiven zu KI von Studierenden. Verschiedene KI-Tools wurden im Rahmen des Projekts in jedem Forschungsschritt eingebunden. Für diese Arbeit habe ich allerdings keine mit KI erhobenen Daten verwendet, und auch in der Auswertung habe ich auf die Unterstützung von KI-Tools verzichtet, um eine vollständig eigenständige und methodisch nachvollziehbare Analyse zu ermöglichen. Zum Transkribieren der Aufzeichnungen der Face-to-Face Interviews, haben alle Forschenden das KI-Tool noScribe verwendet, wobei die Audiodaten mithilfe der KI lokal am Rechner verschriftlicht und alle Transkripte korrekturgehört wurden (vgl. Dröge, 2024). Außerdem habe ich zur sprachlichen Überarbeitung dieser Arbeit ChatGPT auf Basis von GPT-4o genutzt. Die KI wurde dabei zur grammatischen Korrektur und sprachlichen Verfeinerung angewendet, es wurden keine Inhalte mit KI generiert.

4.1. Methodisches Vorgehen

Bei dieser Arbeit habe ich nach dem Forschungsstil der RGTM von Breuer et al. (2019) gearbeitet. Die Daten wurden mit leitfadengestützten, problemzentrierten Interviews erhoben. Das Vorgehen beschreibe ich im Folgenden näher.

4.1.1. Reflexive Grounded Theory

Die Entscheidung mit der Grounded Theory (GTM) zu arbeiten, beruht auf der damit verbundenen Offenheit, den Forschungsprozess flexibel an das empirischen Feldes sowie an die verfügbaren Res-

sourcen anzupassen (vgl. Strübing, 2018, S. 124). Angesichts der dynamischen Entwicklung im Bereich Kl erweist sich ein flexibler Forschungsstil wie das der Grounded Theory als besonders geeignet. Darüber hinaus bietet das theoriegenerierende Vorgehen und der explorative Grundsatz der GTM die Möglichkeit, auf soziale Veränderung durch Kl zu reagieren und bisher unbekannte Phänomene zu entdecken (vgl. ebd., S. 125). Die erkenntnis- und wissenschaftstheoretische Ausrichtung der GTM positioniert die Datenerhebung im Interaktionismus und Pragmatismus, wodurch es möglich ist besonders nuancierte, durch Interaktion prozesshaft hergestellte und perspektivgebundene Ergebnisse zu generieren (vgl. ebd., S. 123).

Methodologie

Um besonders die Insider Position im Feld zu reflektieren und zu nutzen, sowie meine subjektiven Interessen zu integrieren, bietet sich die RGTM als Forschungsstil an. Auf die Insider Position und die Aspekte der Peerforschung gehe ich in Kapitel 4.1.2 genauer ein. Die Reflexiven Grounded Theory (RGTM) von Breuer et al. (2019) ist eine Weiterentwicklung der Grounded Theory von Glaser und Strauss (1967). Die erkenntnistheoretischen Grundlage erweitert die RGTM: Ergebnisse entstehen nicht von sich aus durch die Daten, sondern werden aktiv und unter Einfluss der Forschenden hervorgebracht (vgl. Breuer et al., 2019, S. 9 f). Die RGTM deckt sich in ihrer erkenntnistheoretischen Verankerung weitestgehend mit der konstruktivistischen Strömung der Grounded Theory von Kathy Charmaz (2014) (vgl. Breuer et al., 2019, S. 40 ff). Der Fokus liegt auf der Interaktion zwischen Forschenden und Feld und bezieht damit die Position und Subjektivität der Forschenden in den Prozess des Erkenntnisgewinns ein (vgl. ebd. S. 38). Durch diese enge Verknüpfung der Ergebnisse mit der konstitutiven Subjektivität der Forschenden und den daraus resultierenden Voraussetzungen für den Forschungsprozess leitet sich der Kern der Theorie ab: „Selbst-/Reflexion bzw. Dezentrierung“ (ebd., S. 10) der Forschenden ist das Ziel der Forschungspraxis. Der Forschungsprozessentwickelt sich entlang und unter Berücksichtigung der individuellen Forschenden, deren Aufgabe es ist, ihren Einfluss zu reflektieren und nachvollziehbar zu machen. Der Erkenntnisgewinn mit der RGTM entsteht in der gemeinsamen Interpretation der subjektiven Bedeutungen der Teilnehmenden. Dabei ist das Verstehen nicht nur vom persönlichen Kontext der Forschenden abhängig, sondern ist auch an ihr Vorverständnis gebunden (vgl. ebd., S. 39). Der Analyseprozess zeichnet sich durch einen zirkulären Vorgang aus Einordnung, Reflexion der eigenen Deutungsvoraussetzungen aus (vgl. ebd., S. 46). Dieser Versuch Sinnzusammenhänge nachzuvollziehen und dabei eine Balance zwischen dem

Vorverständnis und der Dezentrierung eigener Annahmen zu finden, wird als iterativer, hermeneutischer Zirkel bezeichnet (vgl. ebd., S. 55). Das Ziel ist es sowohl die theoretische Sensibilisierung zu nutzen als auch theoretische Offenheit zu bewahren.

Forschungsablauf

Der Forschungsprozess mit der RGTM gliedert sich in methodische Schritte, die hier zwar linear aufgeführt werden, in der Forschungsarbeit allerdings nicht stringent aufeinanderfolgend, sondern iterativ zyklisch als Umkreisung der Daten stattfinden. Der Ausgangspunkt des Projekts ist zunächst ein ungefähres thematisches Anliegen. Spezifisches Vorwissen zu KI und Vertrauen aus dem Alltag oder persönliche biografischen Erfahrungen werden dabei als „Präkonzepte“ (Breuer et al., 2019, S. 131) reflektiert. In dieser Arbeit umfasst dieser Punkt die eigene Einstellung und das eigene Vertrauen gegenüber KI, die ich in Kapitel 4.4.3 im Rahmen der Selbstreflexion aufgreife. Zusätzlich zur eigenen Reflexion wird auch wissenschaftliche Literatur zur theoretischen Sensibilisierung herangezogen. Die in Kapitel 2 vorgestellten theoretischen Konzepte stammen aus der Fachliteratur und dienen einerseits einer aktuellen Forschungsübersicht, andererseits auch der Einordnung der eigenen Ergebnisse in den Forschungsstand (vgl. ebd., S. 146). Eine vollständiger Literaturüberblick zur theoretischen Sensibilisierung ist dabei nicht der Anspruch der RGTM (vgl. ebd., S.148). Eine Balance aus theoretischer Offenheit und theoretischer Sensibilität kann durch Differenzierung und Reflexion gewährleistet werden und hat einen hohe Priorität (vgl. ebd., S. 160 f). Die (Selbst-) Reflexion, sowie die Auseinandersetzung mit theoretischen Konzepten und die Arbeit mit den Daten habe ich in einem Forschungstagebuch festgehalten (vgl. Breuer et al., 2019, S. 169). Dort habe ich informelle Texte verfasst, die zur Erarbeitung von Kategorien und dem Sortieren und Kombinieren verschiedener Gedankenzusammenhänge beitragen. Eine Weiterführung und Präzisierung des Forschungstagebuchs stellen Memos dar. Auch hier entstehen informelle Schriften, die zum Beispiel aus Forschungstagebucheinträgen herausgearbeitet werden (vgl. ebd., S.170). Die Forschungsfrage habe ich in einer iterativen Auseinandersetzung mit meiner eigenen Positionalität im Feld, der Aushandlung des Forschungsinteresses, der theoretischen Sensibilisierung und den empirischen Daten entwickelt und präzisiert (vgl. ebd., S.151).

Theoretisches Sampling

Mit der RGMT können unterschiedliche Daten erhoben werden. Auf die hier verwendeten Methoden gehe ich in den folgenden beiden Kapiteln 4.2 und 4.3 ein. Der sowohl vor, während als auch nach der Datenerhebung stattfindende Prozess des theoretischen Samplings bezieht sich auf die

Auswahl der Untersuchungseinheiten (vgl. Breuer et al., 2019, S. 156). Absichtsvolle Entscheidungen bezüglich der zu erhebenden Daten und der Auswahl des zu analysierenden Materials werden flexibel und unter Berücksichtigung der Kontrastierung getroffen. Bei der Kontrastierung wird auf eine Datenauswahl geachtet, die gegensätzliche Aspekte vereint und besonders Kontraste in den Daten beleuchtet, um eine robuste Theorie zu entwickeln. Die Auswahl der Teilnehmenden wurde für das Lehrforschungsprojekt auf Studierende in München beschränkt, und dabei auf eine Mischung von den Studierende aus technikaffinen und anderen Studiengängen geachtet. Ziel war es damit die durch das Studium erworbenen Vorkenntnisse der Teilnehmenden zu berücksichtigen und auf fächerspezifische Anforderungen verschiedener Studiengänge zu reagieren. Die Samplingentscheidungen der Materialauswahl habe ich in Kapitel 4.3.1 ausgeführt. Im iterativen Forschungsprozess der RGTM spielt das theoretische Sampling auch in der Wiederholung der Datenerhebung eine Rolle. Durch den zyklischen Prozess kann entsprechend der Erkenntnisse aus den ersten Daten im weiteren Verlauf das theoretische Sampling fortgeführt werden. Im Rahmen des Lehrforschungsprojekts geschah dies in Kombination mit der Weiterentwicklung der Forschungsfrage, indem mehrere Phasen der Datenerhebung durchlaufen wurden. Zu Beginn standen die eigenen Erfahrungen und das Ausprobieren der KI-Tools im Vordergrund, in der zweiten Phase wurden dann studierende Teilnehmer*innen zu ihren Einstellungen befragt. Die hier verwendeten Transkripte stammen ausschließlich aus der zweiten Phase. Obwohl ich das in der RGTM vorgeschlagene theoretische Sampling mit einer zyklisch wiederholende Datenerhebung nicht in dieser Form durchführen konnte, habe ich mich diesem Verfahren durch die nachträgliche Auswahl des Sample aus dem ursprünglich sehr großen Datenkorpus angenähert (vgl. Breuer et al., 2019, S. 159). Eine Beschreibung des theoretischen Samplings und der Datenauswahl findet in Kapitel 4.3.1 statt.

4.1.2. Peerforschung

Ein zentraler Aspekt der Datenerhebung besteht darin, dass das Vorgehen als Peerforschung konzipiert wurde. Die Daten im Rahmen des Lehrforschungsprojekt wurden von uns Studierenden mit Teilnehmenden erhoben, die ebenfalls Studierende sind, maximal 30 Jahre alt sind und in München studieren. Ziel war es, die Vorteile der Peerforschung für das Forschungsdesign zu nutzen.

Als *Insider* im Forschungsfeld beschreibt Hockney (1993) die Position der Forschenden in der Peerforschung. Die Interviewenden und Forschenden, mich eingeslossen, sind selbst Teil der sozialen Gruppe, die wir untersuchen. Da wir denselben Status als Student*innen mit ähnlichen Tagesabläu-

fen und Ansprüchen haben und zusätzlich zur selben Generation gehören, teilen wir eine vergleichbare Perspektive und gehören als Forschende selbst zum Feld (vgl. ebd., 200). Im Lehrforschungsprojekt wurde diese *Insider Position* besonders durch die zusätzliche Erhebung von Daten aus eigenen Erfahrungen von uns Forschenden genutzt. Durch diese Perspektive nehme ich an, dass die Teilnehmenden einen ähnlichen Bezug zu neuen technischen oder digitalen Entwicklungen haben und ein grundlegendes Verständnis von unterschiedlichen Einstellungen zu KI besteht. Das bedeutet jedoch nicht, dass automatisch eine gemeinsame Meinung oder einheitlicher Zugang zu KI besteht. Vielmehr ermöglicht die *Insider Position* ein vertieftes Verständnis für die jeweiligen Haltungen und Kontexte, die sich aus vergleichbaren Lebenssituationen ergeben (vgl. ebd., 204).

Hockey (1993) betont in seinem Artikel zu Peerforschung einen weiteren Vorteil der *Insider Position*: „A further advantage of the insider role is that on occasion access to the research site is achieved on the back of relationships or similar experiences already established prior to the research commencing“ (ebd., S. 204). Mit *access to the research site* ist dabei nicht nur der tatsächliche Zugang zu Kommiliton*innen und potenziellen Interviewteilnehmenden gemeint, sondern auch der gewährte Zugang zu persönlichen Einstellungen, Meinungen oder Ängsten. In beiden Fällen ergeben sich Vorteile für den Forschungsprozess durch gemeinsame oder ähnliche Erfahrungen im Studium oder im Umgang mit KI.

Die Rollenverteilung im Forschungsprozess gestaltet sich in qualitativen Erhebungen häufig komplex und muss bewusst reflektiert werden. Auch die Peerforschung hat Einfluss auf die Rollenverteilung zwischen Teilnehmenden und Interviewenden. Wie Hockey (1993) treffend formuliert: „The role(s) of mutual friends influences the research scenario, so there is liable to be a blurring of formal (research) and informal (friendship) roles“ (ebd., S. 212). Um diese Vermischung formeller und informeller Rollen vorzubeugen, haben wir möglichst Interviewteilnehmende außerhalb des engen Freundes- oder Familienkreis ausgewählt. Auf diese Weise konnten wir Rollenkonflikte weitestgehend vermeiden, unsere Rolle als Forscher*in wahrnehmen und dennoch Personen aus unserem sozialen Umfeld befragen. An dieser Stelle muss ich darauf hinweisen, dass sich der Feldzugang für einzelne Forschende durch kurzfristige Absagen oder Krankheit erschwert hat und dadurch auf Studierende aus dem näheren Umfeld als Teilnehmende zurückgegriffen wurde.

Auch die Art der Kommunikation wurde durch die Peerforschung positiv beeinflusst. Die sprachlichen und nonverbalen kulturellen Eigenschaften der Kommunikation erleichtern nicht nur den Prozess des Verstehens, sondern fördern darüber hinaus gegenseitige Vertrautheit und die Möglichkeit

das Gegenüber besser einzuschätzen (vgl. Hockey, 1993, S. 204). Die Entscheidung, eine Peerforschung durchzuführen, trägt also maßgeblich zur Optimierung der Datenerhebung bei. Die *Insider Position* erweist sich als besonders produktiv für das Erheben von Daten und das Verstehen des Feldes. Dennoch muss die *Insider Position* ebenso kritisch reflektiert werden und die Einflüsse auf den Forschungsprozess können nicht nur Vorteile mit sich bringen (vgl. Hockey, 1993, S. 200). Auf die Einordnung und den Umgang mit möglichen Problemen der Peerforschung gehe ich im Kapitel 4.4.2 zur Reflexion der Datenverarbeitung genauer ein.

4.2. Problemzentrierte Interviews

Die hier verwendeten Daten stammen aus Interviews mit Studierenden zu ihren Erfahrungen und Einstellungen mit KI. Die Face-to-Face Interviews wurden nach Witzel und Reiter (2022) als problemzentrierte Interviews (PZI) konzipiert.

Die Interviews haben sich um ein bestimmtes Problem bzw. eine bestimmte Thematik zentriert: die Erfahrungen und Einstellungen der Studierenden zu KI. Passend für das Forschungsprojekt und die Arbeit mit der reflexiven Grounded Theory ist auch, dass sich das PZI an das theoriegenerierende Verfahren der GTM anlehnt (vgl. Witzel & Reiter, 2022, S. 77). Die Voraussetzung für das PZI ist, dass ein wechselseitiges Interesse der Beteiligten am Thema besteht und beide Parteien bereit sind, sich auf die jeweils andere Perspektive einzulassen (vgl. ebd., S. 54). Dennoch ist die Forschungsbeziehung asymmetrisch, denn die Vermittlung des wissenschaftlichen Interesses an der Alltagswelt ist nicht das Ziel. Im Fokus steht das Verstehen und gemeinsame Erarbeiten der Perspektive der Teilnehmenden. Das zentrale Thema des PZI wird bereits im Vorfeld des Interviews mitgeteilt. Durch die Studieninformation und informierte Einwilligung haben wir mit den Teilnehmenden das Thema des Interviews besprochen sowie ihre Bereitschaft darüber zu reden, eingeholt, um auf die Forschungsproblematik zu zentrieren (vgl. ebd., S. 55).

Zur Vorbereitung des Interviews zählte neben der Auswahl der Teilnehmenden, der Vorbereitung der Studieninformation und der informierten Einwilligung auch die eigene inhaltliche und methodische Vorbereitung auf das Interview. Gesammeltes Vorwissen wurde systematisch in einen Leitfaden integriert, um während des Interviews angemessen und methodisch planvoll darauf zurückgreifen zu können (vgl. Witzel & Reiter, 2022, S. 108). Der Leitfaden diente dabei nicht als starrer Fragenkatalog, sondern gab vielmehr eine programmatische Orientierung des Gesprächsverlaufs vor. Unterstützt durch geeignete Kommunikationsstrategien konnte so Offenheit gewährleistet und auf individuelle Erzählungen der Teilnehmenden eingegangen werden, etwa durch aktives Zuhören und

gezielte Nachfragen (vgl. ebd., S. 135). Im PZI nehmen die Forschenden eine Doppelrolle an. Zum einen als fachkundige, informierte Personen mit thematischem Vorwissen, andererseits als Lernende, offen für Neues und Unerwartetes (vgl. ebd., S. 79). Mit dem Leitfaden konnte diese Gleichzeitigkeit als Forschende umgesetzt werden.

Da die Teilnehmenden die Wahl hatten, zunächst einen KI-Interviewbot zu testen oder direkt das Face-to-Face zu beginnen, wurden zwei unterschiedliche Leitfadenvarianten für das PZI entwickelt. Sie unterscheiden sich in ihrer Einstiegsfrage und in Bezug auf das (nicht) getesteten KI-Tool, enthalten im Übrigen jedoch identische Fragen. Beide Versionen des Leitfadens befinden sich im Anhang. Der Leitfaden gliedert sich in einen allgemeinen Abschnitt, der sich auf Erfahrungen mit KI fokussiert, sowie einen Abschnitt, der von den jeweiligen Forschungsgruppen im Projekt eigenständig entwickelt wurde und entsprechend unterschiedliche thematische Schwerpunkte aufweist. Jedes Interview dreht sich dennoch um das zentrale Kernthema. Der Austausch und die Wissensreproduktion im Interview erfolgt dialogisch, wobei sich durch das gemeinsame Erarbeiten und Validieren von Interpretationen eine Form von „Arbeitsbündnisses“ (Witzel & Reiter, 2022, S. 128) zwischen Forschenden und Teilnehmenden entwickelt.

4.3. Datenauswahl und Auswertung

Die Daten wurden, wie bereits erwähnt, im Rahmen des Lehrforschungsprojekts ‚Mit KI über KI qualitativ forschen‘ erhoben. Am Projekt konnten sowohl BA-Studierende als auch MA-Studierende teilnehmen. Über den Zeitraum des Sommersemesters 2024 und des Wintersemesters 2024/2025 wurden in Forschungsgruppen Daten erhoben und ausgewertet. Durch die doppelte Zielsetzung des Projekts wurden unterschiedliche Daten in zwei Phasen erhoben und diverse Themen in den Forschungsgruppen fokussiert und analysiert. Die erste Phase der Datenerhebung bestand aus eigenen Erfahrungen von uns Forschenden mit dem KI-Interviewbot Qualia. Dabei haben wir uns selbst von Qualia zu unserer Einstellung und Nutzung zu KI interviewen lassen und unsere Erfahrungen mit dem Tool in Feldnotizen festgehalten. Anschließend daran haben wir in einer zweiten Datenerhebungsphase KI-Interviews mit Qualia und Face-to-Face Interviews mit anderen Studierenden durchgeführt. Für diese Arbeit beziehe ich mich ausschließlich auf erhobene Daten der Face-to-Face Interviews aus der zweiten Phase des Projekts.

4.3.1. Theoretisches Sampling

Der gesamte zur Verfügung stehende Datenkorpus besteht aus 127 Feldnotizen, 79 Post- und Transkripten der Face-to-Face Interviews und 88 Qualia-Transkripten. Den Prozess der Datenauswahl

habe ich damit begonnen alle Daten zu sichten. Durch das Lesen und Analysieren der Daten im Lehrforschungsprojekt entstand eine Familiarität mit den Inhalten und Themen. Im Zuge dessen bin ich auf das wiederkehrende Thema Vertrauen vor allem in den Face-to-Face Interviews gestoßen. Die Frage nach dem Vertrauen in KI wurde von zwei Forschungsgruppen im selbstgestalteten Teil des Leitfadens thematisiert. So wurde beispielsweise im Interview aus dem Bachelorkurs mit einer Informatikstudentin gefragt: „Zu der KI hast du da irgendeine Vertrauensgrundlage? Würdest du sagen, du hast Vertrauen in die KI?“ (TS_033, 42-43). Im Interview aus dem Masterkurs wurde eine Architekturstudentin interviewt und folgendes gefragt: „Jetzt würde mich noch interessieren, in Bezug auf KI, inwiefern vertraust du dann künstlicher Intelligenz?“ (TS_046, 212-213). In manchen Interviews wurde die Frage nach dem Vertrauen auch als Nachfrage gestellt. Generell diente der Leitfaden als Orientierungshilfe und nicht als vorgeschriebene Frage. Die Fragen nach dem Vertrauen sieht also in den Interviews unterschiedlich aus, die beiden Zitate hier dienen als Beispiele.

Die Qualia-Transkripte habe ich aufgrund der weniger ausführlichen Antworten der Teilnehmenden sowie der unspezifischen Nachfragen von Qualia aus der weiteren Analyse ausgeschlossen. Zwar wurde auch in den Qualia-Interviews vereinzelt Vertrauen in KI angesprochen, jedoch nur in wenigen Fällen und ohne inhaltliche Tiefe. Die Feldnotizen aus der ersten Phase zu Qualia habe ich auch nicht zur weiteren Analyse herangezogen, da es in diesen Daten ausschließlich um die Anwendung des KI-Tools in der qualitativen Sozialforschung ging. Neben den Face-to-Face Transkripten habe ich außerdem die dazugehörigen Postskripte gelesen. Durch das Lesen der Postskripte habe ich ein beseres Gefühl für die Interviewsituation bekommen, allerdings wurde vereinzelt in Interviews im Kontext der informierten Einwilligung von persönlichem Vertrauen in die Universität oder die Interviewende Person gesprochen, ohne dass dabei weitere Ausführungen zum Vertrauen in KI aufgegriffen wurden. Dementsprechend habe ich auch die Postskripte aus der weiteren Analyse ausgeschlossen.

Aus den Face-to-Face Transkripten habe ich eine Auswahl getroffen und die 14 Transkripte offen kodiert, die auch konkret das Vertrauen in KI im Interview thematisiert haben. Die Datenauswahl für diese Arbeit umfasst dementsprechend insgesamt 14 Transkripte und 14 Postskripte. Die Hälfte davon wurde von Bachelor- und die andere von Masterstudierenden durchgeführt. Unter den Interviewteilnehmenden sind zehn Frauen und vier Männer, sowie neun unter 25 Jährige und fünf im Alter zwischen 25 und 30. Vier der Interviewten sind Studierende der Soziologie im Hauptfach, außerdem gab es eine Person die ein duales soziales Studium absolviert. Weitere Studienfächer der Teilnehmenden waren Psychologie, Medizin und Sportmanagement, sowie Informatik, Statistik, Maschinenbau und Architektur.

Im Sinne des theoretischen Samplings der RGTM habe ich nach Inspektion der ausgewählten Daten besonders interessante Interviewabschnitte fokussiert (vgl. Breuer et al., 2019, S. 260). Dabei handelt es sich um eine reversible und flexible Forschungsentscheidung. Im Laufe des Forschungsprozesses und der Analyse wurde die Fokussierung der Ausschnitte angepasst und ausgeweitet. Darüber hinaus ist es wichtig die ausgewählten Abschnitte in den Kontext des Interviews einzubetten und dafür das gesamte Transkript zu kennen.

4.3.2. Auswertung

Die Datenauswertung mit der RGTM gestaltet sich nach dem Prinzip der Grounded Theory in drei Schritten: dem offenen, axialen und selektiven Kodieren (vgl. Breuer et al., 2019, S. 9 f). Ziel der Analyse ist es die Kodes in Zusammenhang zu bringen, Kategorien zu erarbeiten und eine Theorie zu entwickeln (vgl. ebd., S. 265). Begleitet werden die drei Schritte durch den Prozess des ständigen Vergleichens (vgl. ebd., S. 273).

Die erste Phase des offenen Kodierens hat einen kreativen Charakter und folgt dem Prinzip der Abduktion (vgl. Breuer et al., 2019, S. 269). Kleinteilige Datenausschnitte werden mit Kodes versehen die sich aus abstrahierten Benennungen von Ideen, Deutungen und Assoziationen zu den Inhalten der Datensequenzen zusammensetzen. Das Datenmaterial steht im Mittelpunkt und empirische Fälle und Kodes werden ständig untereinander verglichen (vgl. ebd., S. 270 ff). Ich habe aus den 14 Interviewtranskripten die Datenausschnitte ausgewählt, die sich im weiteren Sinne mit Vertrauen und der Einstellung zu KI beschäftigt haben und diese offen kodiert. Im zweiten Schritt, dem axialen Kodieren, habe ich Kodes in Beziehung zueinander gesetzt und systematisch Angeordnet (vgl. Breuer et al., 2019, S. 280). Der Fokus liegt dabei auf den Relationen zwischen den Kodes (vgl. ebd., S. 281). Dadurch habe ich erste Kategorievorschläge gebildet, die aus den Daten entstanden sind und mit den empirischen Fällen abgeglichen werden müssen (vgl. ebd., S. 273). An diesem Punkt ist zum einen die Unterscheidung des Vertrauens in eine Grundhaltung, Vertrauen in den Output und Vertrauen in den Datenschutz aufgefallen. Aber auch den Bezug zum Vertrauen, den Erfahrungen mit der KI und Strategien im Umgang für die Studierenden herstellen. Der dritte Schritt, das selektive Kodieren dient der theoretischen Integration aller Kategorien (vgl. Breuer et al., 2019, S. 284). Ich habe meine Perspektive auf den Gesamtzusammenhang der Daten, meines Vorwissens und der theoretischen Konzepte gerichtet und ein zusammenhängendes Konzept entwickelt. Die beim axialen

Kodieren erstellten Kategorien werden miteinander verglichen und in das Theoriekonzept eingeordnet (vgl. ebd., S. 273). Die dadurch entstandene Darstellung des Vertrauens in KI umfasst drei Dimensionen und einen Vertrauensprozess die ich ausführlich in Kapitel 5 darstelle.

4.4. Forschungsethische Reflexion

In diesem Kapitel geht es um die forschungsethische Reflexion dieser Arbeit. Die forschungsethische Reflexion steht nicht am Ende der Arbeit, sondern ist ein durchgängiger Prozess, der in jedem Schritt des Forschungsprozesses mitgedacht wird. Damit ist gemeint, dass Konsequenzen der Forschung für alle Beteiligten in jeder Forschungsentscheidung reflektiert abgewogen werden (vgl. von Unger, 2014, S. 16). Das Kapitel gliedere ich zuerst in die Reflexion der Datenerhebung und den Kontext der Interviews, anschließend werde ich auf die Reflexion in der Datenverarbeitung und Analyse eingehen und schließlich entsprechend der reflexiven Grounded Theory meine Position als Person und als Forscherin in diesem Projekt und in Bezug zum Feld darstellen.

4.4.1. Reflexion der Interviewkontakte

Die Daten wurden im Rahmen des Lehrforschungsprojekts erhoben und umfassen Interviews, die von Bachelor- und Masterstudierenden der Soziologie geführt wurden. Zwischen den beiden Gruppen bestanden Unterschiede in der sozialwissenschaftlichen Forschungserfahrung: Während die Vorbereitung auf die Interviews in beiden Kursen vergleichbar war, verfügten die Masterstudierenden tendenziell über umfassendere Vorerfahrung und Übung in der Führung von Interviews. Dies wirkte sich zwar nicht grundsätzlich auf die Qualität der Interviews aus, führte jedoch in einzelnen Fällen zu Unterschieden in der Präzision der Nachfragen und der Gesprächsführung. Genauso wie es auf Seiten der Bachelorstudierenden Interviewer*innen mit Vorerfahrung gab, haben auch Masterstudierende ohne einschlägige Interviewerfahrung teilgenommen. Für die Auswahl der für diese Arbeit relevanten Interviews stand in erster Linie der inhaltliche Fokus auf das Vertrauen in die KI. In der Analyse konnten jene Interviews mit ausführlicheren Antworten natürlich intensiver ausgewertet werden. Kein Interview der Auswahl wurde aber als unbrauchbar eingestuft.

Neben der unterschiedlichen Erfahrung in der Interviewführung spielt ein weiterer Aspekt des Lehrforschungsprojekts eine zu reflektierende Rolle: Die Interviews wurden nicht mit dem primären Ziel geführt, das Vertrauen von Studierenden in KI zu untersuchen. Auch jene Interviews, die im Leitfaden Fragen zum Vertrauen enthalten und dieses Thema im Interview aufgreifen, fokussieren vorrangig die allgemeine Haltung, Erfahrung und Nutzung mit KI. Für die Interpretation der Daten ist es daher wesentlich, den Entstehungskontext der Interviews zu berücksichtigen. Die Daten wurden

nicht explizit zur Beantwortung der hier untersuchten Forschungsfrage erhoben, was jedoch nicht ausschließt, dass sie für diese Analyse geeignet sind. Vielmehr möchte ich betonen, dass der Fokus der Datenerhebung breiter angelegt war und das Thema Vertrauen in den ausgewählten Interviews als Teilaспект behandelt wurde.

Die Durchführung der informierten Einwilligung stellt einen wichtigen Schritt im Forschungsprozess dar. Sie hat das Ziel, möglichst ausführlich über die Forschung zu informieren und eine freie Entscheidung der Teilnehmenden zur Beteiligung einzuholen (vgl. von Unger, 2014, S. 25 f.). Im Rahmen des Lehrforschungsprojekts wurden Einwilligungen für die Teilnahme am Forschungsprojekt und an anschließenden Abschlussarbeiten eingeholt. Die informierte Einwilligung wurde dialogisch durchgeführt und mit Zustimmung und Unterschrift der Teilnehmenden bestätigt. Dabei wurde dokumentiert, welchen Interviews (KI und/oder Face-to-Face) und welchen Formen der Datenauswertung mit KI die Teilnehmenden zustimmen. Insbesondere die Integration von KI-Tools in verschiedenen Phasen des Forschungsprozess wurde mit den Teilnehmenden besprochen. Sie hatten die Möglichkeit und Verpflichtung selbst über die Verwendung von KI zu bestimmen. Im Zuge der informierten Einwilligung wurden die Teilnehmenden zudem auch über das übergeordnete Forschungsprojekt informiert und über die problemzentrierte Ausrichtung der Interviews aufgeklärt (vgl. Witzel & Reiter, 2022, S. 84). Ziel der informierten Einwilligung war es, die Teilnehmenden in die Lage zu versetzen, eine eigenverantwortliche und informierte Entscheidung über die Verarbeitung ihrer Daten zu treffen, besonders im Hinblick auf den Einsatz von KI-Tools und die darin enthaltene Intransparenz der Datenverarbeitung der KI-Betreibenden. Da ich für die vorliegende Arbeit keine KI-Anwendungen zur Auswertung herangezogen habe und auch nicht auf die Transkripte des Interview-Bots zurückgegriffen habe, konnte ich alle Daten in die Auswertung einbeziehen, auch jene, die einer Auswertung mit KI nicht zugestimmt haben.

4.4.2. Reflexion der Datenverarbeitung

Im weiteren Forschungsprozess haben ebenso forschungsethische Überlegungen stattgefunden. Allem voran sind für die Datenverarbeitung Fragen nach der Anonymisierung und Vertraulichkeit von Relevanz. In sozialwissenschaftlicher Forschung bezieht sich das Prinzip der Schadensvermeidung auf die Antizipation und den Ausschluss von möglichen Schäden für die Beteiligten durch Veröffentlichung von Informationen (vgl. von Unger, 2014, S. 24). Dazu sollen die Daten in solcher Form veröffentlicht werden, dass keine Benachteiligung für die Beteiligten zu erwarten ist. Die Anonymisierung der Daten geht dabei über die Pseudonymisierung personenbezogener eindeutiger Merkmale,

wie z.B. Namen, hinaus. Auch kontextbezogene Informationen können den Rückschluss auf eine Person möglich machen und müssen daher reflektiert anonymisiert werden. Die Anonymisierung steht damit im Gegensatz zur kontextabhängigen Analyse der Daten (vgl. ebd. 2014, S. 25). Es muss abgewogen werden, welche Informationen für den Kontext notwendig sind, aber gleichzeitig keinen Rückschluss auf die Person ermöglichen. Die Lösung dieses Problems in dieser Arbeit sieht folgendermaßen aus: Die Daten wurden in einem ersten Schritt von den Interviewenden anonymisiert. Dabei wurden Namen und eindeutige persönliche Nennungen aus den Transkripten gestrichen, ersetzt und als solche gekennzeichnet. Die Teilnehmenden haben entsprechend der Anzahl der Interviews eine Nummer zugewiesen bekommen. Weitere Informationen über die Teilnehmenden - Geschlecht, Altersspanne und Studienfach - wurden für alle Interviews mit der Interviewnummer zusammengefasst. In einem zweiten Anonymisierungsschritt habe ich die Daten nach der Analyse, zur Veröffentlichung nochmals auf Anonymität geprüft. Dadurch sind keine Rückschlüsse auf die teilnehmenden Personen von Außenstehenden möglich, der Kontext der Interviews und der Zitate ist allerdings erhalten und kann nachvollzogen werden (vgl. Von Unger, 2018, S. 96).

Ein weiterer Aspekt, der den Forschungsprozess maßgeblich beeinflusst und somit reflektiert werden muss, ist der Aufbau als Peerforschung. Die in Kapitel 4.1.2 dargestellten Gründe für Peerforschung verdeutlichen das Potenzial, erfordern jedoch zugleich eine kritische Auseinandersetzung mit dem Einfluss, den die Nähe zum Feld und die *Insider Position* bewirken. Im Rahmen der Datenauswertung muss darauf geachtet werden, dass aufgrund der Nähe zum Feld keine relevanten Aspekte übersehen oder als selbstverständlich vorausgesetzt werden (vgl. Hockey, 1993, S. 202). Der Einfluss der Peerforschung auf die erhobenen Daten muss sorgfältig reflektiert werden. So wie es möglich ist, dass die Teilnehmenden durch die Peerforschung vertraulicher und offener kommunizieren, kann es auch den gegenteiligen Effekt haben: Die soziale Nähe zwischen Interviewer*in und Teilnehmer*in kann Unsicherheit auslösen, zu Zurückhaltung führen oder normative, sozial erwünschte Antworten generieren (vgl. ebd., S. 203). Diese Einflüsse auf den Forschungsprozess müssen reflektiert in der Auswertung der Daten beachtet werden.

4.4.3. Selbstreflexion

In diesem Kapitel gehe ich auf meine Selbstreflektion in diesem Forschungsprojekt als Forscherin und als Person im Feld ein. Besonders für die Nachvollziehbarkeit der Ergebnisse ist es wichtig meinen eigenen Zugang zu KI und Vertrauen darzustellen. Gerade auch deswegen, weil die eigene Reflexion ein zentraler Bestandteil des Forschungsstils RGTM ist.

Ich habe mein Soziologie Studium bereits vor der Veröffentlichung von Chat GPT begonnen. Zum Zeitpunkt meiner Bachelorarbeit im Frühjahr 2023 waren bereits erste KI-Anwendungen wie Chat GPT verfügbar. Dadurch, dass es allerdings keine Richtlinien und meinerseits Skepsis über den Nutzen gab, habe ich in meinem Studium keine KI verwendet. Erst im Laufe der Zeit und mit zunehmender Popularität von Chat GPT habe ich im privaten Rahmen eigene Anwendungserfahrungen mit verschiedenen KIs gemacht. Dabei habe ich sehr gegensätzliche Gefühle entwickelt. Zum einen ist meine Grundskepsis gegenüber der KI gewachsen, durch fehlerhaften Output und Halluzinationen der KI haben sich meine Bedenken gegenüber dem Nutzen und meine ohnehin skeptische Haltung zu Datenschutz bestätigt. Andererseits habe ich auch eine Neugierde entwickelt und den Wunsch zu lernen, wie die KI sinnvoll und verantwortungsvoll auch im universitären Kontext eingesetzt werden kann. In Zuge dieser Motivation habe ich das Lehrforschungsprojekt im Sommersemester 2024 besucht. Darin ging es um das Forschen mit und über KI. Durch mein Nebenfach Statistik im Bachelorstudium habe ich auch Vorlesungen zu Maschine Learning besucht und hatte daher ein Grundverständnis von einfachen maschinellen Algorithmen. Weitere Informationen über die internen Prozesse der KI sowohl in der Datenverarbeitung als auch im Output habe ich im Lehrforschungsprojekt durch eigene Recherche aber vor allem durch Beiträge der Dozierenden bekommen. In dieser Zeit ist also mein Verständnis der KI gewachsen, ebenso aber auch meine Bedenken durch die Intransparenz der Datenverarbeitung und der Funktionsweise. Mein Spannungsverhältnis zwischen Skepsis und Vertrauen habe ich auch in den erhobenen Interviewdaten wiedergefunden und mich daraufhin intensiv mit dieser Thematik beschäftigt, um einen Weg zu finden Vertrauen in die KI zu erforschen.

Abgesehen von meinem persönlichen Zugang zum Thema, hat natürlich auch meine Sozialisation und mein Privileg mich mit der KI auseinanderzusetzen meinen Zugang und meine Interpretation beeinflusst. In kleinstem weise war ich jemals darauf angewiesen die KI zu verwenden, gleichzeitig hatte ich über ein Jahr Zeit mich mit der KI im Projekt und jetzt in der Masterarbeit zu befassen und die KI auszuprobieren.

5. Empirische Ergebnisse

Im Zentrum der nun folgenden Analyse steht das Vertrauen der Studierenden in die KI. Um das Vertrauen von Studierenden zu verstehen, nehme ich Bezug auf die eigenen Beschreibungen der Teilnehmenden in den Interviews. Ich kann also nur eine Aussage über das kommunikativ vermittelte Vertrauen treffen, nicht jedoch über das implizite Vertrauen. Wie Hitzler (2016) in seinem Artikel zu interpretativer Sozialforschung festhält, findet die Rekonstruktion des subjektiv gemeinten Sinns

des Gegenübers nicht unmittelbar statt, sondern kann nur gefiltert durch die Sprechenden und Hörenden wahrgenommen werden (vgl. ebd. 2016, S. 176). Was im Interview über einen Zustand gesagt wird, kann sich vom tatsächlichen Zustand unterscheiden ebenso wie das von mir Gesagte anders verstanden werden kann, als es von mir gemeint ist. Dennoch gehe ich davon aus, dass die Ausführungen der Studierenden einen Wahrheitsgehalt haben und man durch die Selbstreflektion der Studierenden hindurch etwas über ihr Vertrauen in die KI lernen kann. Mit dieser Arbeit verfolge ich also nicht das Ziel, das implizite Vertrauen der Studierenden in die KI herauszuarbeiten, sondern ich versuche die Frage zu beantworten, wie die Studierenden in ihren eigenen Worten Vertrauen in KI äußern. An dieser Stelle weise ich darauf hin, dass ich direkte Zitate aus den Interviewtranskripten verwende. Zur verbesserten Lesbarkeit habe ich sie sprachlich minimal überarbeitet und dabei Inhalt und Charakter der Wortbeiträge erhalten.

In den Interviews wird deutlich, dass die Studierenden der KI nur bedingt vertrauen. Das bedeutet, dass Vertrauen in die KI vorhanden ist, aber durch bestimmte Faktoren eingeschränkt wird. Die Studierenden setzen sich (selbst-)kritisch mit ihrem eigenen Vertrauen in KI auseinander und machen deutlich, an welche Bedingungen und Eigenschaften der KI, aber auch an welchen äußeren Einflüssen sie ihr Vertrauen knüpfen.

Das Vertrauen in die KI lässt sich in drei miteinander verknüpfte Dimensionen unterscheiden. Das Vertrauen in den Output der KI (Für-wahr-halten) und das Anvertrauen von Informationen an die KI. Die dritte Dimension ist ein Grundvertrauen, das unabhängig der anderen beiden Dimensionen eine intrinsische Haltung gegenüber der KI beschreibt. Zusammen ergeben diese drei Dimensionen ein Gesamtbild des Vertrauens in die KI. Die Dimensionen werden nicht unabhängig voneinander beurteilt, sondern als Aspekte eines Ganzen beschrieben.

Die Studierenden reflektieren die Einflüsse und Bedingungen auf das Vertrauen in die Datensicherheit und den Output und begründen dies. Die Ergebnisse möchte ich in einer prozesshaften Annäherung an das Vertrauen in die KI darstellen. Damit ordne ich das beschriebene Vertrauen in einen Ablauf, den ich aus den Interviewdaten herausgearbeitet habe. Die Struktur gliedert sich in drei Schritte, die durch verschiedene Faktoren Einfluss auf die Entwicklung des Vertrauens nehmen.

Der erste Schritt des Prozesses betrifft die Dimension der Grundhaltung. Eine Grundhaltung gegenüber der KI steht an erster Stelle und zeichnet sich durch einen generellen Optimismus/Pessimismus gegenüber Technik oder einer Affinität/Antipathie aus. Wie aus der Literatur bekannt, ist diese Grundhaltung biographisch und charakterlich bedingt und wird durch Vorwissen und Sozialisation

beeinflusst (vgl. Hoff & Bashir, 2015, S. 414). Da diese implizite Haltung der Teilnehmenden nicht mit den Interviewdaten erhoben wurde, kann ich über die Entstehung und Bedingung der Grundhaltung keine Aussage treffen. Dass sie allerdings existiert und relevant für das Vertrauen in die KI ist, berichten die Studierenden in ihren Interviews (vgl. Kapitel 5.2.1). Aus der Grundhaltung entwickelt sich also ein bestehendes, aber im Prozess veränderliches Grundvertrauen/-misstrauen.

Für den nächsten Schritt habe ich eine Annäherung an die KI herausgearbeitet. Die Studierenden reflektieren ihre Vertrauensgrundlage anhand neu dazu gewonnener Eindrücke. Sie machen eigene Erfahrungen und informieren sich über die internen Prozesse der KI und deren Funktionsweise. Dies betrifft sowohl die Dimension des Anvertrauens von Informationen als auch die Dimension des Für-wahr-haltens. Aus meiner Perspektive spielt die Nachvollziehbarkeit der KI dabei die größte Rolle für die Einschätzung des Vertrauens der Studierenden. Es entsteht ein differenziertes Bild mit unterschiedlichen Reaktionen auf Prozesse der KI. Außer ihren eigenen Erfahrungen und ihrem eigenen Verständnis der KI ziehen Studierende auch Einschätzungen von Expert*innen oder von Institutionen, denen sie bereits vertrauen, heran. Durch das angeeignete Wissen und die Erfahrungen beurteilen die Studierenden das Vertrauen in die KI neu und entwickeln dabei kontextabhängiges Vertrauen (vgl. Hoff & Bashir, 2015, S. 417). Das Vertrauen unterscheidet sich für die jeweilige Verwendung der KI und das zu bearbeitende Thema und wird entsprechend der Komplexität und der Verantwortung angepasst. Mit steigender Verantwortung steigt auch die Skepsis gegenüber der KI, genauso wie beim Anvertrauen zunehmend persönlicherer Daten weniger Vertrauen in die KI oder bei komplexeren Themen weniger Vertrauen in den Output besteht.

Im dritten Schritt habe ich die Strategien der Studierenden, wie sie die KI verantwortungsvoll und mit eingeschränktem Vertrauen verwenden, zusammengefasst. Es zeigt sich, dass die Studierenden dabei ihre eigene Verantwortung an oberste Stelle setzen. Der entscheidende Maßstab ist die individuelle Einschätzung der Vertrauenswürdigkeit der KI, die durch Kontextwissen, Plausibilität, Kontrolle und Anonymisierung hergestellt wird. Diese Einschätzung des Grads des Vertrauens wird abhängig von Grundeinstellung, Misstrauen durch Nachvollziehbarkeit, Thema und Kontext in der Interaktion mit der KI neu ausgehandelt.

Das Vertrauensspektrum der Studierenden in die KI reicht von gering bis mittel, es ist in keinem Fall ein vollumfängliches Vertrauen. Genau dieser gemischte Zustand aus Vertrauen und Misstrauen gegenüber der KI wird in einem andauernden Lernprozess geformt. In den Interviews wird deutlich,

dass der Aushandlungsprozess des Vertrauens immer wieder neu durchlaufen und durch neue oder geänderte Faktoren beeinflusst wird.

5.1. Dimensionen des Vertrauens

Wenn die Studierenden über ihr Vertrauen in KI sprechen, unterscheiden sie zwischen Vertrauen in den Output der KI und Vertrauen in den Datenschutz der KI. Die Dimension der Grundhaltung wird dabei in den Interviews weniger explizit als Vertrauensdimension erwähnt. Sie steht im Vertrauensprozess noch vor der aktiven Annäherung an die KI. Die beiden anderen Dimensionen dagegen entwickeln sich in der aktiven Auseinandersetzung mit der KI. Die drei Dimensionen sind nicht eindeutig voneinander getrennt und bilden für die Studierenden zusammen das Vertrauen in die KI. Die Vertrauensdimensionen, vor allem das Anvertrauen und das Für-wahr-halten, werden durch aktive Einschätzung selbst bestimmt. Die Einflussfaktoren der Nachvollziehbarkeit und des Kontexts ist für beide Dimensionen relevant, hingegen spielen die Erfahrungen mit der KI vor allem für das Für-wahr-halten des Outputs eine Rolle.

Im Laufe des Interviews mit einer Soziologiestudentin geht es um Vertrauen und Nutzen von KI im akademischen Bereich. Die Teilnehmerin zeigt eine deutliche Skepsis und Vorsicht bei der Verwendung von KI. Bei der Beschreibung ihrer Skepsis wird deutlich, dass sich diese für die Studentin vor allem aus der Intransparenz der KI ergibt. In anderen Interviews wird beispielsweise auch die Fehleranfälligkeit des Outputs als Grund für mangelndes Vertrauen genannt. Die Abläufe in der KI sind nicht nachvollziehbar und dadurch weniger vertrauenswürdig. Die Intransparenz beeinflusst sowohl das Anvertrauen, als auch das Für-wahr-halten. Zum einen bezieht sich die Unsicherheit auf den Datenschutz des Inputs in die KI, zum anderen auf die internen Prozesse, wie der Output der KI zustande kommt. So führt die Soziologiestudentin im Face-to-Face Interview aus:

„Ich weiß ja nicht wo die Info landet oder generell ist es für mich irgendwie so eine große Blackbox, woher er die Informationen hat und wie das Ganze also funktioniert.“ (TS_070, 219-222).

Es wird deutlich, wie die beiden Vertrauensdimensionen Datenschutz („wo die Info landet“) und Output („woher er die Informationen hat“) das gesamte („das Ganze“) Vertrauen in die KI beschreiben. Neben dieser Verbindung nennt die Teilnehmerin auch den für sie entscheidenden gemeinsamen Einflussfaktor auf beide Dimensionen des Vertrauens: Die mangelnde Nachvollziehbarkeit (vgl. Kapitel 5.2.2).

5.1.1. Für-wahr-halten und Anvertrauen

Sehr deutlich formuliert auch ein Informatikstudent in seinem Interview die beiden Aspekte, die das Vertrauen in die KI für ihn ausmachen:

„Ich vertraue der KI aber es ist schon so, dass ich ihr auch in gewissen Maßen nicht vertraue, weil ich nicht weiß, passen die Antworten, die ich da zurückbekomme oder was passiert mit den Daten, die ich eingebe.“ (TS_065, 94-97)

Der Teilnehmer äußert beides: Ein gewisses Maß an Vertrauen und Misstrauen in die KI. Das Vertrauen ist also bedingt und setzt sich zusammen aus Vertrauen in die Antworten – dem Output der KI und dem Anvertrauen von Informationen an die KI sowie damit verbundene Datensicherheit. Im weiteren Interviewverlauf geht er noch genauer auf die beiden Dimensionen ein. Zunächst beschreibt er Datenschutzbedenken. Hierbei bezieht er sich konkret auf Chat GPT – eine KI, die er selbst nutzt:

„Ich hab bisschen Angst, in Anführungsstrichen, davor, dass die Daten gespeichert werden natürlich, und zum anderen, Chat GPT ist ja glaube ich ein amerikanisches Unternehmen. Ja, das ist mit dem Datenschutz ist auf jeden Fall anders wie in Deutschland, das weiß ich, aber es ist auch so ein bisschen, dass ich's einfach nicht wirklich weiß. Also das ist, ich denke, einfach so ein komplexes System und Datenschutzerklärungen kennt man ja, ich glaube es gibt wenige, die die komplett durchgelesen haben.“ (TS_065, 105-112)

Hier spricht er zum einen Angst vor Datenmissbrauch an, zum anderen die unklaren rechtlichen Datenschutzbedingungen, die für das amerikanische Unternehmen Open AI gelten. Das Vertrauen in europäische Datenschutzvorgaben ist größer, aber eine differenzierte Unterscheidung der tatsächlichen Datenschutzvorschriften ist nicht bekannt. Zudem sind die Vorgaben, Bedingungen und internationalen Unterscheidungen ein „komplexes System“, das die Nachvollziehbarkeit erschwert. Auf die Thematik dieser Verständlichkeitsprobleme geht er auf Nachfrage der Interviewerin noch einmal genauer ein und macht deutlich, wie wichtig es für sein Vertrauen ist, die Datenverarbeitung transparent und vor allem verständlich zu gestalten:

„Ich fände es gut, wenn es verständlicher erklärt wird (...), das wäre tatsächlich so ein, das Thema, was mir das erleichtern würde, wobei ich jetzt auch sagen muss, das ist, wie ich schon davor gesagt habe, einfach ein komplexes Thema, einfach, weil es auch ein globales, also vor allem Chat GPT jetzt ein globales Produkt ist, das kann ich mir vorstellen, dass man das nicht

in fünf Sätze runterbrechen kann aber wenn es mehr transparent ist und wenn mehr offen kommuniziert wird, was wird denn mit diesen Daten eigentlich gemacht.“ (TS_065, 124-132)

Der Teilnehmer reflektiert auch, dass es schwierig ist, komplexe Prozesse, wie sie hier vorliegen, nachvollziehbar darzustellen. Dennoch ist die Nachvollziehbarkeit der Datenverarbeitung eine zentrale Bedingung für das Anvertrauen von Informationen an die KI, die in diesem Zustand nicht erfüllt wird und damit das Vertrauen in die KI negativ bedingt.

Auf die Dimension des Vertrauens in den Output, kommt der Informatikstudent anschließend zu sprechen. Auch hier herrscht für ihn ein Mischverhältnis aus Vertrauen und Misstrauen. Die Bedingung dafür ordnet er, ähnlich wie bei der Datenverarbeitung, der mangelnden Nachvollziehbarkeit zu. Auf die Nachfrage zu den Bedingungen des Vertrauens führt er weiter aus: „Ich weiß nicht, welche Datenquellen ChatGPT oder auch andere KI-Systeme benutzen und wie diese gewichtet werden“ (TS_065, 170-171). Dass die internen Prozesse der KI intransparent sind, spielt für ihn also nicht nur für das Vertrauen in die Datensicherheit, sondern auch für das Vertrauen in den Output und die Datenquellen der KI eine Rolle. Zum Vertrauen in den Output nennt er außerdem schlechte Erfahrungen mit fehlerhaften Antworten, die das Vertrauen negativ beeinflussen:

„Ich hatte es auch schon so, dass ich mir versucht habe Code ausgeben zu lassen der dann einfach nicht funktioniert hat. Also da war dann auch der Output letztendlich nicht richtig.“
(TS_065, 223-226)

Zusammenfassend lässt sich feststellen, dass die beiden Dimensionen Anvertrauen und Für-wahr-halten durch mangelnde Nachvollziehbarkeit geprägt sind. Für das Vertrauen in den Output spielen allerdings außerdem Erfahrungswerte mit dem Output eine Rolle. Vor allem fehlerhafter Output hat dabei eine negative Auswirkung auf die Vertrauenswürdigkeit.

5.1.2. Grundhaltung

Die dritte Dimension, die Grundhaltung, stellt eine übergeordnete Haltung gegenüber der KI dar. Sie beeinflusst sowohl die Dimension des Anvertrauens, als auch die des Für-wahr-haltens, und steht, aus meiner Perspektive, im Vertrauensprozess an erster Stelle. Die Einstellung der Teilnehmenden zu KI kann aus einer bereits etablierten Haltung zu (neuer) Technik abgeleitet werden oder unvoreingenommen sein. Die Dimension der Grundhaltung kann sich auch weiterentwickeln und durch den Vertrauensprozess verändern. Jedoch stellt sie eine träge und persönlich bedingte Hal-

tung dar und wurde in der Literatur, wie bereits erwähnt, von Hoff und Bashir (2015) als *Dispositional trust* zusammengefasst. Über eine positive Grundeinstellung des Vertrauens in KI berichtet beispielsweise eine Psychologiestudentin. Auf die Nachfrage, woher ihr Vertrauen in KI kommt, antwortet sie im Face-to-Face Interview:

„Ich weiß gar nicht. Ich hab das Gefühl, ich vertrau dem Ganzen irgendwie. Ich weiß gar nicht, ob das so eine naive Unvoreingenommenheit ist. Ich finde man hat ja trotzdem immer noch in der Hand, was gibt man preis. Und das habe ich aber bei einem Menschen auch und im Internet auch.“ (TS_025, 207-210)

Dem „Ganzen irgendwie“ zu vertrauen, weist auf ein inhärentes Grundvertrauen hin – eine Haltung, die die grundsätzliche Einstellung gegenüber der KI beschreibt. Die Studentin beschreibt ihre Haltung als unvoreingenommen und naiv, gibt aber im weiteren Interview Gründe für diese Vertrauensbasis an. Dadurch, dass sie der KI also nicht grundsätzlich skeptisch, sondern offen gegenüber tritt, zeichnet sich ein Grundvertrauen gegenüber der KI ab. Die Studentin argumentiert mit ihrer eigenen Verantwortung und eigenen Handlungsmöglichkeiten, Gefahren der KI zu entgehen („man hat ja trotzdem immer noch in der Hand, was gibt man preis“). Damit benennt sie auch eine Strategie, wie sich das Vertrauen nicht blind, sondern verantwortungsbewusst etabliert: Nämlich durch die eigene Einschätzung als zentraler Mechanismus, die das Vertrauen einschränkt. Auf diesen Aspekt werde ich in Kapitel 5.2.3 zurückkommen. Im weiteren Interviewverlauf spricht sie auch an, dass ihr Vertrauen nicht durch „schlechte Erfahrungen“ (TS_025, 213) gestört wurde und die Erfahrungen mit KI eher zu mehr Vertrauen beigetragen haben. Das Grundvertrauen wird also wie die anderen beiden Dimensionen, Anvertrauen und Für-wahr-halten, durch Erfahrungen und Gelerntes gefestigt. Hätte die Studentin schlechte Erfahrungen gemacht, wäre ihr Grundvertrauen möglicherweise gesunken oder hätte sich in eine skeptischere Richtung entwickelt. Durch das Zitat im vorherigen Abschnitt wird außerdem deutlich, dass die Studentin ihre Grundhaltung der KI gegenüber an bereits Bekanntem festmacht und vergleicht. Dabei zieht die Teilnehmerin ihr Vertrauen in andere, für sie vergleichbare Beziehungen heran: Ihr Vertrauen in das Internet oder in Menschen. Aus ihrer persönlichen Einschätzung ähnelt die Beziehung mit der KI in dieser Hinsicht derer in andere Vertrauensobjekte. Daher beeinflussen ihre Erfahrungen mit anderen Vertrauensbeziehungen ihre Grundhaltung gegenüber der KI. Durch die Entwicklung des Vertrauens als Prozess wird deutlich, dass die Grundhaltung durch Erfahrungen und Gelerntes schließlich beeinflusst wird und sich in eine bestimmte Richtung entwickeln kann. Dieser Prozess wird im folgenden Kapitel 5.2 genauer erläutert.

Eine zweite Soziologiestudentin hingegen äußert eine eher skeptische Grundhaltung gegenüber der KI in ihrem Face-to-Face Interview. Für sie ist es sehr wichtig, die KI zu verstehen (vgl. TS_009, 184-185), um sie nutzen zu können und nicht in eine Abhängigkeit zu geraten (vgl. TS_009, 173-175). Auf die Anmerkung der interviewenden Person, dass eine KI, bspw. Chat GPT, nicht gänzlich verstanden werden kann, und die Frage, ob sich trotzdem Vertrauen entwickeln könnte, entwickelt sich folgender Dialog:

„TN: Also ich sehe auf jeden Fall Grenzen. Das zeigt sich ja auch in meinem Konsumverhalten. Also ich benutze Chat GPT eigentlich selbst nie. (...) manchmal werde ich so drauf gebracht, dafür habe ich jetzt keine E-Mail² und so, da denke ich, ja stimmt, könnte ich eigentlich mal benutzen. Aber ich habe das so nicht in mir drin. Ich weiß gar nicht warum.“

I: Also du hast da gar nicht die Möglichkeit auf dem Schirm, weil du da zu kritisch eingestellt bist?

TN: Ja, wahrscheinlich das. Und weil ich einfach, ich habe auch nicht das Bedürfnis danach, mir das so einfach zu machen über KI.“ (TS_009, 191-200)

Aus diesem Zitat lese ich heraus, dass diese Teilnehmerin die KI nicht aus eigener Motivation verwendet und sie durch ihre Skepsis gehemmt ist, eigene Erfahrungen mit der KI zu machen. Auf die Nachfrage bestätigt die Studentin diese Einschätzung und weist darauf hin, dass sie den Aufwand, die KI zu verstehen, um ihr zu vertrauen, nicht bereit ist zu leisten. Sie hat kein eigenes Bedürfnis eigene Erfahrungen mit der KI zu machen und zu Lernen. Trotzdem scheint sie nicht vollständig von der KI abgeneigt zu sein, da sie sowohl davon berichtet, Chat GPT bereits genutzt zu haben, als auch dem Interview mit KI im Vorfeld zugestimmt hat.

5.2. Vertrauensprozess

Die Entstehung von Vertrauen der Studierenden in die KI lässt sich aus meiner Perspektive als Prozess in drei Phasen beschreiben. Ich stelle dar, wie die Studierenden ihr Vertrauen in die KI einordnen, welche Faktoren Einfluss auf das Vertrauen haben und ordne das Ganze in eine zeitliche Ebene ein. Die Schritte des Prozesses können dabei nicht strikt voneinander getrennt werden, ebenso wie

² Eine Anmeldung per E-Mail ist bei Chat GPT nicht zwingend notwendig, um die KI zu nutzen. Mit der Anmeldung können aber Vorteile der KI, wie z.B. ein höher entwickeltes GPT-Modell oder Zugriff auf vorherige Chats mit der KI genutzt werden. Im Interview weist die Studentin damit darauf hin, dass sie sich nicht bei Chat GPT angemeldet hat und über keinen Account verfügt.

der Prozess nicht ausschließlich linear verläuft. In den Interviews wird deutlich, dass die Studierenden ihr Vertrauen auch auf einer zeitlichen Ebene betrachten. Das Vertrauen in KI ist also nicht unveränderlich, sondern wird immer wieder neu beurteilt. Nach den Fragen zur Erfahrung mit dem KI-Interview geht es im Face-to-Face Interview mit einer dritten Soziologiestudentin um das Vertrauen in KI. Auf die Frage, in welchen Bereichen sie KI vertraut, äußert sie sich aufgrund von fehlender Transparenz der KI und eigenen Erfahrungen skeptisch. Zu Kontexten und Situationen, in denen sie der KI vertrauen würde, sagt die Teilnehmerin folgendes:

„Das Ding ist, wenn halt mehr darüber bekannt wäre woher genau die KI ihre Daten schöpft und worauf alles sie wirklich Zugriff hat und ja, welche Mengen an Daten und welche Qualität von Daten auch eingespeist wurden vielleicht, dann, glaube ich, würde das mein Vertrauen in einigen Bereichen in die KI auf jeden Fall erhöhen und dann könnte ich mir auch vorstellen, das in mehr Bereichen anzuwenden. Aber bisher finde ich es relativ undurchsichtig woher genau die Informationen genommen werden und, dass halt auch die KI teilweise sich Antworten auf gewisse Dinge ausdenkt wenn sie keine Antwort hat. Das ist zumindest meine Erfahrung mit Chat GPT und insofern, so zu 100 Prozent Vertrauen wüsste ich jetzt noch nicht. Dafür müsste es transparenter sein, wie es funktioniert und woher die Daten kommen.“ (TS_010, 60-72)

Zum einen benennt sie Bedingungen, an denen sie ihr Vertrauen in KI misst: Ihre eigenen Erfahrungen und die Nachvollziehbarkeit der KI. Zum anderen bezieht sie sich auf den aktuellen Stand der Technik. Sie spricht vom bisherigen Zustand und benennt Eigenschaften der KI, die sich ändern müssten, damit sie mehr vertrauen würde. Dazu zählen mehr Transparenz der internen Prozesse, als es aktuell der Fall ist. Dem entnehme ich, dass die Einschätzung des Vertrauens in die KI kein abgeschlossener Prozess ist, sondern immer wieder mit sich verändernden Eigenschaften neu evaluiert und angepasst wird.

5.2.1. Grundhaltung einschätzen

Der erste Zugang zu Vertrauen in die KI findet aus meiner Perspektive über die Dimension der Grundhaltung statt. Im Folgenden beziehe ich mich auf die bereits aufgeführten Zitate zur Dimension des Grundvertrauens aus dem Interview mit der zweiten Soziologiestudentin und der Psychologiestudentin (Interview 09 und 25). Die Grundhaltung wird dabei zuerst wahrgenommen und beeinflusst dann die weitere Herangehensweise an das Vertrauen. Für die Psychologiestudentin herrscht eine positive Grundhaltung, vor allem Begründet mit vorherigen, positiven Erfahrungen

und Strategien, mit denen sie ihr Vertrauen gezielt einordnet. Durch diese Einstellung hat sie bereits ein Grundvertrauen in die KI, das durch die weiteren Schritte der Annäherung verändert werden kann.

Für die zweite Soziologiestudentin steht ebenfalls die Grundhaltung an erster Stelle und beeinflusst, wie sie weiterhin Vertrauen zur KI aufbaut. In ihrem Fall kann man eher von einem Grundmisstrauen sprechen, denn ihre Grundhaltung gegenüber der KI ist skeptisch. In ihrem Interview wird deutlich, dass ihre Skepsis und ihr Misstrauen sie dahingehend beeinflussen, aus eigener Motivation keine weiteren Annäherungsschritte an die KI zu unternehmen. Für sie ist der Aufwand sich ausreichend über die KI zu informieren zu groß, noch dazu sieht sie keinen Bedarf für Vertrauen und Nutzen der KI.

Dass sich die Grundhaltung verändern kann und durch weitere Schritte im Vertrauensprozess beeinflusst wird, beschreibt die vierte Soziologiestudentin in ihrem Face-to-Face Interview. Sie antwortet auf die Frage nach ihrem Vertrauen in KIs, wie z.B. Chat GPT, folgendes:

„Also zu 100 Prozent vertraue ich dem Ganzen nicht tatsächlich. Wenn ich nach Korrekturen oder so frage, lese ich tatsächlich nochmal durch. Also vergleiche, ist was rausgeschnitten worden, ist alles drin was ich brauche. Also ich vertraue nicht 100 Prozent darauf, aber ich würde schon sagen, so ein Grundvertrauen habe ich und je öfter quasi der Chat GPT mich ‚nicht enttäuscht‘, also verlässlich ist oder allgemein, desto mehr vertraue ich dem. Das ist wahrscheinlich halt so ein menschliches Ding“ (TS_014, 82-88)

Hier führt sie aus, inwiefern sie ein Grundvertrauen in die KI hat und dass dieses ihre Grundhaltung beeinflusst, allerdings kein unhinterfragtes Vertrauen bedeutet. Außerdem wird deutlich, dass sich ihr Vertrauen in die KI durch ihre Erfahrungen weiterentwickeln. Macht sie gute Erfahrungen und wird nicht von der KI enttäuscht, wächst ihr Vertrauen. Durch die Anwendung der KI wird dazugelernt und das Vertrauen in die KI aufgebaut. Auffallend ist auch, dass sie diesen Prozess der Vertrauensaushandlung als „menschliches Ding“ beschreibt. Die Einschätzung und Aushandlung, ob jemand oder etwas vertrauenswürdig ist oder ihm/ihr vertraut werden kann, ist individuell und menschlich, findet aber in jeglicher Vertrauensbeziehung, nicht nur zur KI, statt. Die Besonderheit dieser Situation, dass die KI ein neues Vertrauensobjekt ist und zu ihr durch ihre Neuheit und Dynamik noch keine verankerten Zugänge und Beziehungen herrschen, gleichen die Teilnehmenden durch bekannte Annäherungen an andere Vertrauensobjekte an und nutzen bewährte und etablierte

menschliche Prozesse, um das Vertrauen in die KI zu bestimmen. Anschließend an die Einschätzung des eigenen Grundvertrauens folgen also weitere Schritte, um das Vertrauen in die KI zu bestimmen.

5.2.2. Einflüsse auf das Vertrauen

Durch die Analyse der Daten habe ich mehrere Faktoren herausgearbeitet, die Einfluss auf das Vertrauen in KI nehmen und im Prozess an nächster Stelle stehen. Die Teilnehmenden berichten von Aspekten, die sie nutzen, um das Vertrauen in KI durchgehend zu bewerten. Nachdem sie also ihre Grundhaltung gegenüber der KI wahrgenommen und daraus ein Grundvertrauen bzw. -misstrauen abgeleitet haben, wird dieses durch eigene Initiative ergänzt. Es gibt keine Verpflichtung, die KI zu nutzen, und es liegt an der Neugier und der Motivation der Studierenden selbst, ihr Vertrauen in die KI zu ergründen und zu reflektieren.

Erfahrungen

Der erste Faktor ist bereits in einigen Zitaten angeklungen: Es ist die eigene Erfahrung der Studierenden. Die eigenen Erfahrungen mit der KI beeinflussen die Grundhaltung, aber auch die Dimension des Für-wahr-haltens des Outputs. Wie aus den bereits vorgestellten Interview 25 mit der Psychologiestudentin und Interview 14 mit der Soziologiestudentin zeigt sich, dass die Erfahrungen mit der KI als direkte Einflussfaktoren auf das Vertrauen genannt werden. Darüber hinaus berichten beide Studentinnen davon, dass ihr Grundvertrauen nicht durch schlechte Erfahrungen negativ beeinflusst wurde (vgl. TS_014, 86-87; TS_025, 213).

Erfahrungen mit dem Output der KI werden von vielen Teilnehmenden als Zuverlässigkeitmaß verwendet. Basierend auf diesen Erfahrungen bewerten sie danach ihr eigenes Vertrauen in die KI. Eine Public Social Management Studentin äußert sich in ihrem Face-to-Face Interview eher skeptisch und mit einem Grundmisstrauen gegenüber der KI (vgl. TS_064, 72-73). Neben Datenschutzbedenken schildert sie außerdem schlechte Erfahrungen mit fehlerhaftem Output mit Chat GPT:

„Wenn ich jetzt zum Beispiel mir irgendwas anschau oder lerne und ich die Definition von irgendeinem Begriff nicht so verstanden habe, dann schaue ich eben noch mal in Chat GPT, ob ich mir das besser oder einfacher erklären kann. Und da habe ich dann teilweise schon gemerkt, dass es nicht so ganz richtig war und dann habe ich doch nochmal gegoogelt und dann war das da auch anders erklärt als jetzt über Chat GPT. Also teilweise ist es dann doch nicht so richtig.“ (TS_064, 95-101)

Die Studentin hat die Erfahrung gemacht, dass die KI einen fehlerhaften Output ausgibt. Sie kann sich nicht ohne weiteres auf die KI verlassen, sondern muss den Output kontrollieren und korrigieren. Sie stellt also eine Inkonsistenz in die Vertrauenswürdigkeit des Outputs fest und leitet daraus ein eingeschränktes Vertrauen ab.

Über ihr Vertrauen in die KI spricht eine Architekturstudentin im Face-to-Face Interview sehr selbst-reflektiert. Sie erklärt ausführlich, weshalb ihr der Output der KI vertrauenswürdig erscheint, welche Erfahrungen sie dahingehend mit der KI gemacht hat und welche Konsequenzen sie daraus zieht.

Auf die Frage, inwiefern sie KI vertraue, antwortet sie:

„Ja, das ist schwierig. (...) Ich glaube, es ist so ein bisschen das Typische, dass ich der leider auch blind vertraue. Weil man einfach, wenn man das durchliest, sich das so seriös irgendwie und wissenschaftlich fundiert anhört, dass man sehr dankbar darüber ist, so einen scheinbar zuverlässigen Text zu bekommen. Deswegen, wenn man es durchliest, denk ich mir perfekt, ich muss nichts mehr machen und vertrau dem blind. Aber wenn ich, ich frag meistens immer noch irgendwie zwei, dreimal nach oder lass mir ein Beispiel nochmal erläutern oder nochmal erklären und nachdem sich das dann immer wiederholt oder mal was ganz was anderes dann genannt wird, bin ich da immer skeptisch und Google meistens immer noch nach, was die KI mir vorgeschlagen hat, dass ich das nochmal überprüfe mit Google und vertraue dem nicht ganz.“ (TS_046, 214-228)

Die Studentin behauptet zuerst, der KI blind zu vertrauen, relativiert die Aussage aber sofort, indem sie die genauen Gründe für ihr Vertrauen schildert und es am Eindruck des Outputs festmacht, „einen scheinbar zuverlässigen Text“ ausgegeben zu bekommen, das seriös und wissenschaftlich wirkt. Es sind also Eindrücke des Outputs der KI, die Einfluss auf das Vertrauen nehmen. Daran anschließend erklärt die Studentin allerdings, dass sie sich trotzdem auf weitere Erfahrungen mit der KI stützt und ihr Vertrauen nicht nur von oberflächlichen Eindrücken abhängig macht. Ihre weiteren Erfahrungen mit der KI sind ebenso geprägt von inkonsistenten Antworten und Unzuverlässigkeit. Ihr Vertrauen ist dadurch eingeschränkt und kann eigentlich nicht als blindes Vertrauen bezeichnet werden. Durch dieses Zitat möchte ich zeigen, dass die Erfahrungen mit der KI vielseitig sind und von den Studierenden reflektiert werden. Im weiteren Interviewverlauf führt die Architekturstudentin außerdem ihre Strategie genauer aus, um der KI dennoch aus eigener Verantwortung zu vertrauen. Auf diesen Schritt im Vertrauensprozess werde ich im Kapitel 5.2.3 über Strategien zum Umgang zurückkommen.

Auch die zuallererst zitierte Soziologiestudentin berichtet von ihren Erfahrungen mit dem Output der KI. Sie schildert ebenfalls ihre Erfahrungen mit fehlerhaftem Output und die daraus abgeleitete Notwendigkeit, den Output der KI zu prüfen:

„Manchmal sagt er dann irgendwie Sachen die irgendwie nicht ganz passen oder erfindet auch tatsächlich manchmal einfach Infos und da muss man auf jeden Fall nochmal abchecken und skeptisch sein.“ (TS_070, 157-159)

Sie spricht an, dass der Output der KI unpassende und erfundene Informationen enthält und es daher notwendig ist, der KI eine Skepsis entgegenzubringen und nicht vollständiges Vertrauen zu haben. Auch sie deutet die Strategie der Kontrolle an, die dazu beiträgt, dass die KI trotz eingeschränktem Vertrauen genutzt werden kann.

Die geschilderten Erfahrungen in den Interviews zielen alle entweder auf den Einfluss auf die Dimension der Grundhaltung oder die Dimension des Für-wahr-haltens des Outputs ab. Die beiden Vertrauensdimensionen sind durch eigene Erfahrungen geprägt. Die Dimension des Anvertrauens hingegen wird hier nicht genannt. Daraus erschließt sich für mich, dass Datenschutz weniger nachvollzogen werden kann als die Korrektheit des Outputs. Bekommt man einen Output von der KI, kann man mit eigenen Recherchen den Wahrheitsgehalt der Aussagen prüfen. Gibt man allerdings Informationen in die KI hinein, kann man nicht ohne weiteres nachvollziehen, wo diese Daten (zwischen-)gespeichert oder für Trainingszwecke verwendet werden. Man muss an der Stelle des Inputs mehr Aufwand betreiben, um herauszufinden, ob Datenschutzversprechen eingehalten werden. Die Dimension des Anvertrauens wird weniger durch eigene Erfahrungen und mehr durch Nachvollziehbarkeit, Wissen und Verstehen von internen Prozessen beeinflusst. Beide Aspekte, Erfahrung und Verstehen als Einfluss auf das Vertrauen, finden sich in der Literatur im Konzept des *learned trust* von Hoff und Bashir (2015) wieder.

Nachvollziehbarkeit

Wissen über die internen Prozesse der KI wird am häufigsten als Einflussfaktor auf das Vertrauen der Studierenden beschrieben. Die Teilnehmenden führen dabei aus, wie das (Nicht-)Verstehen oder Nachvollziehen von Funktionen der KI die Vertrauenswürdigkeit verändert. Die Interviews zeigen, dass die Nachvollziehbarkeit sowohl auf die Grundhaltung, als auch auf das Anvertrauen sowie auf das Für-wahr-halten des Outputs Einfluss nimmt.

Ein Beispiel für den Einfluss von Nachvollziehbarkeit auf die unterschiedlichen Vertrauensdimensionen liefert eine Informatikstudentin. Sie hat sich gegen ein KI-Interview entschieden und deshalb nur das Face-to-Face Interview geführt. Darin antwortet sie auf die Nachfrage, warum sie ungern mit der KI kommunizieren möchte, folgendes:

„Also zum Teil hat das bestimmt auch so was mit der Datenverarbeitung zu tun, auch wenn es alles geklärt und sicher ist und es Datenschutzgrundlinien gibt, so ist es halt trotzdem irgendwie das Gefühl, dass man diese Informationen in einen Computer reintut, die man halt sonst nicht in der digitalen Welt sehen möchte. Das ist, glaube ich, so vom Gefühl her so ein bisschen weird, Dinge in den Computer zu tun die so supersensibel für einen sind.“ (TS_033, 25-31)

Ihr Misstrauen gegenüber der KI begründet sie mit einem unguten Gefühl. Sie spricht über die expliziten Datenschutzvereinbarungen, dennoch hat sie Hemmungen, sensible Informationen an die KI zu geben. Das Vertrauen in die KI geht für sie über die explizite Datenschutzbestimmung hinaus. Trotz der Vereinbarungen hat die Studentin Misstrauen, dass sich die Betreiber*innen der KI nicht an diese halten. Das Risiko, dass durch die KI sensible Informationen gesammelt und missbräuchlich verwendet werden, ist der Studentin zu groß. Auf die Frage nach der Vertrauensgrundlage mit der KI antwortet die Informatikerin folgendes:

„Ich würde sagen eher, also ja und nein. Ich hätte die Vertrauensgrundlage, dass wenn ich jetzt mit einer KI ein faktenbasiertes Gespräch führen würde, also Gespräche in Anführungszeichen, dann hätte ich, glaube ich, schon Vertrauen, dass eben wenn das eine KI ist die halt gut trainiert ist und wo man sich so ein bisschen darauf verlassen kann, dass die halt mit tatsächlichen Fakten und nicht Fake News und so gefüttert wurde.“ (TS_033, 44-49)

Ihr Vertrauen knüpft sie also ebenfalls an Wissen über die Quellen und Entstehung des Outputs. Die Studentin legt dabei Wert auf die Entstehung der Antworten, dass diese faktenbasiert sind und verlässliche Quellen genutzt werden. Die Bedingungen für eine vertrauenswürdige KI scheinen für sie allerdings nicht vollständig erfüllt zu sein. Die Entscheidung gegen ein KI-Interview und die Beschreibung und Formulierung der Einflussfaktoren auf ihre Vertrauensgrundlage, ordne ich so ein, dass diese nicht ihren jetzigen Voraussetzungen für Vertrauenswürdigkeit entsprechen. Dabei zeigt die Studentin, dass sie unter der Bedingung der Nachvollziehbarkeit – d.h. wüsste sie mehr über die internen Prozesse – mehr Vertrauen gegenüber der KI hätte. Anschließend geht sie noch auf eine

weitere Bedingung für ihre Vertrauensgrundlage ein. Dabei beschreibt sie ihre „Gefühls-ebene“ (TS_033, 52), die auch an nonverbale Kommunikation und Zwischenmenschlichkeit geknüpft ist (vgl. ebd., 51 ff) und von der KI nicht geleistet werden kann.

Mehrere Teilnehmer*innen haben die KI aufgrund von Intransparenz als „undurchsichtig“ (TS_010, 66) oder „spooky“ (TS_045, 135) beschrieben. Der Medizinstudent, der die KI als „spooky“ bezeichnet, führt auf Nachfrage der Interviewerin im Face-to-Face Interview aus, was er damit meint:

„Ich versteh halt nicht genau wie KI funktioniert. Und man weiß ja nie, was sind die Trainingsdaten und was wird da jetzt verwendet und was wird mit dem, was ich der KI gebe, gemacht. Und der Mechanismus an sich, weiß nicht, ich bin kein Informatiker oder Physiker, sondern ich weiß nicht wie's funktioniert, deswegen ist mir das ein bisschen unheimlich.“ (TS_045, 160-164)

Dabei wird deutlich, dass generelle Unwissenheit über die Funktionsweise der KI zu Unbehagen führt. Der Student bezieht sich dabei einerseits auf die Dimension des Anvertrauens („was wird mit dem, was ich der KI gebe, gemacht“) als auch auf die Dimension des Für-wahr-haltens („man weiß ja nie, was sind die Trainingsdaten“). Für ihn spielt es eine zentrale Rolle, dass er die KI nicht verstehen und nachvollziehen kann. Im Vertrauensprozess ordnete ich seine Situation so ein, dass er Vertrauen durch Verstehen aufbauen will, dies aber für ihn nicht möglich ist. Er nennt konkret Verständnisprobleme, da ihm Wissen für den Zugang fehlt. Er vermutet, dass Personen aus anderen Fachbereichen wie Informatik oder Physik mehr von den internen Prozessen der KI verstehen und dadurch mehr vertrauen können.

Über Datenschutzbedenken spricht die vierte Soziologiestudentin in ihrem Face-to-Face Interview. Sie bezeichnet sich selbst als uninformiert und beantwortet die Frage zuerst zögerlich (vgl. TS_014, 417-418) führt ihre Bedenken gegenüber der KI dann allerdings aus:

„Ich glaube, dass man den Aspekt „Vertrauen“ so ein bisschen im Hinterkopf behalten muss. Also wie stark vertraue ich auf sowas? Ich meine die Tools und alles, das wird immer besser. Aber es gibt ja auch, es hört sich jetzt irgendwie doof an, aber es gibt ja so aus einem Grund Cyber-Security und wenn Chat GPT zum Beispiel oder Künstliche Intelligenz oder was mir doch jetzt einfällt, dass Chat GPT z.B. Sachen die man gefragt hat speichert, was passiert mit den Daten, wer kann darauf zugreifen? Zum Beispiel auch was ich mir oft denke, wenn ich den was Korrektur lesen lasse, das ist ja meine Leistung was ich getan habe. Oder was ich

vollbracht habe. Inwiefern kann der da noch darauf zugreifen oder können andere darauf zugreifen oder dann wissen, was ich da mache.“ (TS_014, 418-429)

In diesem Zitat erkenne ich den zeitlichen, wiederholenden Vertrauensprozess. Das Vertrauen muss „im Hinterkopf behalten“, also immer wieder neu evaluiert werden. Zur Einordnung und Begründen ihres Vertrauens zieht sie die Datensicherheit bei der Nutzung der KI heran. Nachzuvollziehen, was mit den Daten passiert und wer darauf zugreifen kann, sind Aspekte, die ihr Vertrauen in die KI steigern. Da sie darüber aber in ihren eigenen Worten nicht ausreichend informiert ist, ist sie in dieser Hinsicht eher misstrauisch.

Über den Einfluss von Nachvollziehbarkeit und Verstehen der KI auf das Vertrauen des Outputs, spricht eine weitere Psychologiestudentin in ihrem Face-to-Face Interview. Sie beschreibt, welche Annahmen über die Funktionsweise der KI und die Entstehung des Outputs mehr oder weniger Vertrauenswürdigkeit bei ihr hervorrufen. In einem Abschnitt in dem es um die Glaubwürdigkeit von KI geht, sagt sie:

„Also wenn du jetzt weißt, okay das greift auf Internetquellen zurück und halt auf alle gleichzeitig oder halt alles, was so einfach zugänglich ist, gleichzeitig, und kann dir dann daraus die beste Antwort geben oder irgendwie eine Kombi daraus, dann finde ich ist es noch ein bisschen vertrauenswürdiger. Also wenn du dann weißt, ja okay, aber eigentlich macht er einfach nur irgendeinen random Satz der irgendwie zum Thema passt, und was er da so aus den Quellen hat, finde ich, wird es ein bisschen weniger vertrauensvoll.“ (TS_021, 371-378)

Zuerst beschreibt die Studentin eine Annahme, die ihr Vertrauen in die KI steigern würde. Durch das Wissen über die Quellen und Funktionsweise, wie die KI zu Informationen für ihren Output kommt, wirkt die KI vertrauenswürdiger, weil sie nachvollziehbaren Output produziert. Hingegen die Annahme und das Wissen, dass die Antworten der KI wahrscheinlichkeitsbasierte, zufällige Sätze sind, lassen die Vertrauenswürdigkeit sinken. Was in der ersten Annahme zu mehr Vertrauen führt, ist die Systematik der Antworten. Was in der zweiten Annahme wiederum zu weniger Vertrauen führt, ist die Willkürlichkeit der Antworten. Hier wird deutlich, wie unterschiedlich die Vertrauenswürdigkeit eingeschätzt werden kann, je nachdem wie tief das Verständnis über die Funktionsweise der KI geht und auch wie dieses Wissen eingeordnet wird.

Die Studierenden versuchen ihr Vertrauen durch Verständnis der KI aufzubauen. Es wird deutlich, dass das Verstehen, oder Nichtverstehen alle drei Vertrausdimensionen betrifft. Die Herangehensweise der Studierenden wird durch die internen Prozesse und Funktionsweise der KI beeinflusst. Diese Beschreibung des Vertrauens ist aus der Literatur von Jacovi et al. (2021) als intrinsisches Vertrauen (vgl. ebd., S. 628) bekannt.

Thema/Kontext

Ein weiterer Einflussfaktor auf das Vertrauen in die KI ist das themen- oder kontextspezifische Vertrauen. Ein vergleichbares Konzept aus der Literatur stellt dabei der *Situational Trust* von Hoff und Bashir (2015) dar. Je nachdem in welchem Kontext und zu welchem Zweck die KI verwendet wird, haben die Studierenden mehr oder weniger Vertrauen. Bezuglich der Dimension des Vertrauens in den Output beschreibt ein Maschinenbaustudent sein Vertrauen:

“I would say just for many basic things I trust it a lot more than for, I don't know, detailed or more complex data. I mean, because sometimes I would say for daily tasks it's too good and I trust it so well. And maybe for other things, more complex, than it's worth it, my trust isn't that 100 Prozent.” (TS_086, 134-138)

Er beschreibt eine klare Unterscheidung des Vertrauens bei alltäglichen Anwendungen und komplexen Aufgaben. Für alltägliche, einfache Aufgaben hat der Student ein größeres Vertrauen in die KI als für komplexe Aufgaben. Im weiteren Interview führt er aus, dass das Vertrauen in die KI reflektiert werden muss und er sein Vertrauen besonders bei einfachen Aufgaben selbst kritisiert. Der Grund für diese kontextabhängige Unterscheidung ist zum einen, ob die Informationen der KI durch Vorwissen als richtig oder plausibel erkannt werden können und zum anderen, welche Auswirkungen die Informationen der KI haben. Letzteres beschreibt der Medizinstudent in seinem Interview auf die Frage, ob er den Output der KI nachprüft:

„Ja, sobald es halt für mich irgendwie eine Handlungsrelevanz hätte, vor allem halt im Arbeitskontext. Oder sobald ich irgendwie die Inhalte die die KI mir gibt in einem Austausch mit anderen verwenden würde, als meine Wissensgrundlage, dann prüfe ich das schon nach, indem ich halt die Referenzen einfach nachlese.“ (TS_045, 150-155)

Auf das Kontrollieren von Output der KI als Strategie, um Misstrauen in die KI auszugleichen, gehe ich im anschließenden Kapitel 5.2.3 ausführlicher ein. Das Zitat des Medizinstudenten macht deutlich, dass er diese Strategie nur in Kontexten anwendet, in denen er weniger Vertrauen hat. Dieses

wenige Vertrauen ist beeinflusst durch den Kontext, in dem er die Informationen verwenden möchte. Er nennt den „Arbeitskontext“ und den „Austausch mit anderen“ als Bereiche, in denen die Korrektheit des Outputs besonders wichtig sind und das Vertrauen in die KI nicht ausreicht, um den Output ohne Kontrolle zu übernehmen. Sobald die KI also als Wissensgrundlage im professionellen Kontext oder im Austausch von Informationen mit anderen Menschen genutzt wird, schildert der Student ein geringeres Vertrauen. Ähnlich begründet die Psychologiestudentin ihr kontextabhängiges Vertrauen. In ihrem Face-to-Face Interview wird sie danach gefragt, wie viel sie der KI zutraut. Darauf antwortet sie:

„Also zumindest im aktuellen Stand würde ich da jetzt nicht 100 Prozent drauf vertrauen. Vor allem in wichtigen Fragen nicht. So wie zum Beispiel bei der Uni würde ich jetzt nicht nur dem seine Lösung glauben, sondern beschäftige mich schon noch auch selber damit. Aber bei so kreativen Sachen oder sowas kann man das, glaube ich, schon gut nutzen.“ (TS_021, 249-254)

Sie hat ebenfalls weniger Vertrauen, wenn es um „wichtige Fragen“ geht, z.B. im universitären Kontext. Auch sie verfolgt die Strategie, die Informationen der KI nachzuprüfen und selbst einzuschätzen. Bei Themen, die die Freizeit betreffen und weniger faktenbasiert sind, wie z.B. kreative Inspiration, vertraut sie der KI eher. An dieser Stelle ist ein wirkliches Nachprüfen auch nicht möglich, da es um subjektive Einschätzungen oder Vorschläge geht und nicht um das nachvollziehbare Beantworten von wissenschaftlichen Fragen.

Neben den Auswirkungen der Informationen, begründet ein Sportmanagementstudent sein Vertrauen in den Output auch mit seinem Vorwissen und der eigenen Einschätzung, ob etwas korrekt und vertrauenswürdig ist. Kennt er sich mit einem Thema gut aus, kann er den Wahrheitsgehalt der Informationen der KI auch leichter einschätzen und vertraut ihr mehr als bei Themen, zu denen er kein Vorwissen hat. Auf die Frage, wie er mit Informationen der KI umgeht, antwortet er im Face-to-Face Interview:

„Also es kommt natürlich auch wieder stark auf das Thema an aber ich würde schon so weit gehen, dass ich sagen würde, wenn ich einer KI jetzt die Frage stellen würde, ob sie mir die eine Website einfach einmal zusammenfasst über dieses und solches Thema was jetzt mein Studium angeht, dass ich dem Ganzen schon so weit vertrauen würde, wie es für mich persönlich Sinn ergibt. Was natürlich auch abhängig davon ist, wie sehr ich mich mit dem Thema

auskenne, um überhaupt bewerten zu können, ob diese Antwort Sinn ergibt.“ (TS_085, 161-167)

Er macht das Vertrauen in die KI von seinem eigenen Wissen über das Thema abhängig. Geht es bspw. um ein Thema aus seinem Studium, mit dem er vertraut ist, kann er den Output der KI dahingehend leichter einordnen und für richtig oder falsch befinden, als bei einem Thema, zu dem er sich nicht gut auskennt. In dem Fall wäre sein Vertrauen geringer, da er nicht anhand von Vorwissen oder eigenen Erfahrungen zu dem Thema einschätzen kann, ob der Output richtig ist oder nicht.

Aus meiner Perspektive setzt sich der zweite Schritt des Vertrauensprozesses aus folgenden Einflüssen auf das Vertrauen zusammen: Die Erfahrungen mit der KI, das Verstehen und der Kontext der Anwendung. Diese drei Gruppen stellen dabei keinen Vollständigkeitsanspruch, sondern sind eine von mir ausgearbeitete Zusammenfassung der am häufigsten genannten Einflüsse aus den Face-to-Face Interviews. Außerdem sind die Einflussfaktoren selbst, aber auch wie sich diese genau auf das Vertrauen auswirken, individuell und können sich verändern, noch dazu ist es nicht möglich sie in ihrer Gänze zu erfassen. Zur weiteren Untersuchung sind wiederholte Datenerhebungen und Auswertungen notwendig.

5.2.3. Strategien zum Umgang mit eingeschränktem Vertrauen

Die dritte Phase des Vertrauensprozesses setzt sich zusammen aus Strategien, wie die Studierenden mit dem nur eingeschränkt möglichen Vertrauen in die KI umgehen.

Dazu möchte ich vier Strategien vorstellen, die ich aus den Daten herausgearbeitet habe. Sie beschreiben, wie die Studierenden die KI verantwortungsbewusst nutzen, trotz Skepsis und eingeschränktem Vertrauen. Die erste Strategie bezieht sich auf die Dimension des Für-wahr-haltens des Outputs. Die Studierenden berichten davon, die Informationen der KI zu kontrollieren, bzw. nachzuprüfen. Sie umgehen ihr eingeschränktes Vertrauen, indem sie die Informationen nicht übernehmen, sondern zuerst selbst auf den Wahrheitsgehalt prüfen. Die zweite Strategie betrifft die Dimension des Anvertrauens. Da die Studierenden auch ein eingeschränktes Vertrauen bezüglich des Datenschutzes haben, berichten viele von der Strategie, die Daten, die sie der KI anvertrauen, zu beschränken und zu anonymisieren. Die dritte Strategie betrifft die Grundhaltung gegenüber der KI. Manche Teilnehmende berichten davon, ihr Vertrauen in die KI nicht komplett selbst zu beurteilen, sondern richten sich nach Personen oder Institutionen, denen sie bereits vertrauen, und wie diese die KI nutzen. Die vierte Strategie bezieht sich auf die Eigenverantwortung der Studierenden. Dabei

beschreiben sie konkret die individuelle und selbstverantwortliche Bewertung und Aushandlung des Vertrauens in die KI.

Kontrolle

Geht es um das Vertrauen in den Output, berichten die meisten Studierenden davon, wie sie die Informationen der KI kontrollieren. Die Architekturstudentin erzählt davon, dass sie aufgrund der besseren Nachvollziehbarkeit und der einfacheren Kontrolle eine KI verwendet, die automatisch die Quellen des Outputs angibt (vgl. TS_046, 235-241). Auch andere Teilnehmende schildern ihr Vorgehen bei der Überprüfung der Informationen. Die dritte Soziologiestudentin sagt, dass sie die Antworten der KI nachrecherchiert, weil sie weiß, dass es nicht zuverlässig wahr ist (vgl. TS_010, 128-130). Und auch die erste Soziologiestudentin führt aus:

„Ich überprüfe es definitiv, also ich habe noch nie irgendwie komplett verwendet, was er genommen hat. Also was heißt noch nie, aber im Endeffekt muss man immer einen Double-Check machen.“ (TS_070, 149-151)

Die Studentin spricht hier von einem „Double-Check“, den sie macht, um die Antworten der KI zu prüfen. Aus meiner Perspektive meint sie damit einerseits die Plausibilität des Outputs anhand eigenen Vorwissens, Kontext und Intuition zu bewerten, andererseits aber nicht darauf zu vertrauen, dass diese Kontrolle ausreicht, sondern noch einen zweiten Check durchzuführen, die Informationen zu recherchieren und mit Quellen zu belegen. Für die Studentin ist klar, dass der Output nicht einfach übernommen werden kann, sondern immer überprüft und angepasst werden muss. Das Vertrauen ist hier außerdem an die Korrektheit des Outputs geknüpft und beschäftigt sich weniger mit dem Verstehen und Bewerten der internen Prozesse. Damit folgt diese Strategie eher dem extrinsischen Vertrauen von Jacovi et al. (2021), in dem die Erwartbarkeit der Antworten bewertet wird.

Anonymität

Die zweite Strategie bezieht sich auf das Anvertrauen von Informationen an die KI. Hierfür schildern die Studierenden, wie sie die Daten anpassen und anonymisieren. Damit nehmen sie den Datenschutz selbst in die Hand. Sensible Informationen geben sie also gar nicht erst an die KI weiter. Sie umgehen damit ein Risiko und müssen der KI in dieser Hinsicht gar nicht vertrauen.

Der Informatikstudent beschreibt in seinem Interview, wie er Chat GPT nutzt. Dabei schildert er, dass er bei der Eingabe von Informationen in die KI „Namen oder ähnliches“ (TS_065, 90) weglassen würde und darauf achtet, keine personenbezogenen Daten in die KI zu speisen. Genauso schildert auch die erste Soziologiestudentin das Beispiel, sich von der KI eine Mail schreiben zu lassen:

„Also zum Beispiel wenn ich mit ihm eine Mail schreibe, ich schreibe ungern noch unten meinen Namen oder so hin, weil ich mir denke, mein Name muss er jetzt nicht unbedingt wissen (...), oder an welche Person es geht Irgendwie denke ich mir manchmal, ja ich weiß ja nicht, wo die Info landet.“ (TS_070, 216-220)

Sie formuliert ihre Bedenken, dass die Daten zwischengespeichert oder veröffentlicht werden. Um diesem Misstrauen in den Datenschutz der KI entgegenzuwirken, verzichtet sie zum Beispiel darauf, persönliche Informationen, wie z.B. Namen an die KI weiterzugeben.

Institutionen/Expert*innen

Die dritte Strategie, die ich aus den Daten herausgelesen habe, besteht darin, bei eingeschränktem Vertrauen in die KI auf bereits bekannte Vertrauensobjekte und deren Einschätzung der KI zu setzen. Die Informatikstudentin spricht in ihrem Face-to-Face Interview über die Zukunft der KI und ihre Einschätzung potenzieller Risiken. Dabei führt sie aus, dass es zum einen schwierig ist, sich vollständig zu informieren und dass es viele unterschiedliche Meinungen zu KI gibt. Dazu schlussfolgert sie:

„Irgendwo muss man ja dann sozusagen auch den Expert*innen auf dem Gebiet vertrauen. Oder muss nicht, aber es ist ja immer gut sich damit auseinanderzusetzen, was sagen Leute, die wirklich da Ahnung haben.“ (TS_033, 275-277)

Sie beschäftigt sich mit unterschiedlichen Meinungen zu KI und vertraut schließlich Expert*innen, die sich mit der KI auskennen. Dadurch vertraut sie der KI nicht selbst, sondern dem, was andere informierte Menschen über die KI wissen und sagen.

Im Face-to-Face Interview mit der Public Social Managementstudentin geht es um das Vertrauen in Chat GPT verglichen mit Qualia, der KI, die im KI-Interview zuvor verwendet wurde. Auf die Frage, ob sie dem Output von Chat GPT vertraue, antwortet sie:

„Vielleicht da auch ein bisschen weniger, weil dieses Qualia, kannte ich vorher noch nicht. Und ich dachte mir auch, wenn das im Rahmen eurer Forschung da mitverwendet wird, dann wird das schon so einigermaßen okay sein in diesem Bereich mit Datenschutz und so. Bei

Chat GPT ist es halt schon so ein bisschen, dass man sich nicht sicher sein kann, was es jetzt alles speichert und was nicht.“ (TS_064, 67-71)

Sie vertraut Chat GPT weniger als Qualia. Die Begründung ist dabei, dass sie Chat GPT besser kennt, und dem eher skeptisch gegenübersteht, also eingeschränktes Vertrauen hat. Da sie Qualia nicht kannte und keine skeptische Grundhaltung gegenüber KI hat, misst sie ihr Vertrauen in Qualia an der Situation. Maßgeblich für die Situation, in der sie Qualia getestet hat ist, dass es im Rahmen des Forschungspraktikums stattgefunden hat. Sie vertraut der Universität in deren Rahmen die Forschung stattgefunden hat und uns als forschende Personen. Da wir Qualia verwenden und sie uns vertraut, ist ihr Vertrauen in Qualia größer, obwohl sie keine eigenen Erfahrungen oder andere Schritte durchlaufen ist, um ihre Vertrauen einzuschätzen. Sie evaluiert ihr Vertrauen also nicht selbst, sondern macht es stattdessen daran fest, wie vertrauenswürdige Personen oder Institutionen die KI verwenden.

Eigenes Einschätzen

Die vierte Strategie habe ich als das Verantwortungsbewusstsein bei der Nutzung der KI zusammengefasst. Die Beschreibung dieses Schrittes im Vertrauensprozess von den Teilnehmenden zeigt, dass sie ihren eigenen Einschätzungen und Fähigkeiten, die KI zu bewerten, eine große Bedeutung beimessen und sie in keinem Fall ein unüberlegtes Vertrauen haben. Das eingeschränkte Vertrauen in die KI wird durch Vertrauen in die eigenen Fähigkeiten, im Einzelfall zu bewerten, ausgeglichen. Die Statistikstudentin begründet damit ihren vertrauensvollen Umgang mit der KI. Auch wenn ihr Vertrauen in die KI begrenzt ist, vertraut sie auf ihre eigenen Fähigkeiten, verantwortungsvoll mit der KI umzugehen. Im Face-to-Face Interview sagt sie:

„Also ich vertraue immer darauf, dass das was ich von der KI kriege, dass ich das selber wieder einschätzen kann, wie sehr ich das dann benutze oder halt auf mich einwirken lasse.“ (TS_032, 236-238)

Sie traut sich zu, die KI trotz ihrer eingeschränkten Vertrauenswürdigkeit verantwortungsvoll zu nutzen, gestützt auf ihre Fähigkeit, deren Zuverlässigkeit einzuschätzen. Dieses Verantwortungsbewusstsein sieht auch der Maschinenbaustudent. In seiner Reflektion über die Verwendung von KI sagt er:

“it's a tool. It's also the way you use it. It's like a car. A car is made to move people from A to B, but it also can kill people. So it depends on the use, and it shouldn't be mishandled. So

yeah, (...) I mean, to be conscious that if you put your trust on the AI, if it does mistakes as well, then you are also guilty, because you already know that you made the decision of trusting it. I mean, it can also happen with all the other tasks, so yeah." (TS_086, 143-149)

Er vergleicht die KI mit anderen Maschinen oder Hilfsmitteln, wie einem Auto. Dabei zieht er die Schlussfolgerung, dass es im Verantwortungsbereich der Nutzenden liegt, die Maschine so einzusetzen und ihr so zu vertrauen, dass sie selbst die Verantwortung für etwaige Fehler tragen. Die KI wird also als Hilfsmittel gesehen, zu dessen Verwendung und Vertrauenswürdigkeit man sich als Nutzer*in selbst Gedanken machen muss und im Zuge dessen auch für Fehler einsteht.

6. Diskussion

Die vorgestellten Ergebnisse werde ich nun diskutieren und zentrale Erkenntnisse mit der eingangs vorgestellten Literatur verbinden. Dabei gehe ich auf die von mir entwickelten drei Vertrauensdimensionen (Grundvertrauen, Anvertrauen und Für-wahr-halten) und den Vertrauensprozess ein und stelle Verknüpfungen zu theoretischen und empirischen Konzepten aus Kapitel 2 her. Anschließend spreche ich Limitationen und Herausforderungen der Forschung an.

6.1. Diskussion zentraler Ergebnisse

Die Ergebnisse meiner Arbeit zeigen ein eingeschränktes Vertrauen in KI, das sich aus drei Dimensionen zusammensetzt und in einem dreischrittigen Prozess kontinuierlich ausgehandelt wird. Die Forschung zu Vertrauen in KI erweiterte ich durch den spezifischen Blick auf die studentische Perspektive in Zeiten neuer Möglichkeiten mit KI zu arbeiten. Und stelle die Ergebnisse anhand des Vertrauensprozesses, der die eigene Einschätzung und Bewertung des Vertrauens beschreibt, dar.

Die Ergebnisse zeigen, dass die Studierenden der KI skeptisch und mit einem eingeschränkten Vertrauen gegenüberstehen. Dies verdeutlicht, dass sie ihr Vertrauen bewusst und reflektiert gestalten, anstatt es unüberlegt oder verantwortungslos zu vergeben. Zugleich positionieren sie sich nicht ablehnend gegenüber der KI, sondern sehen darin Chancen und sind bereit, schrittweise Vertrauen in die KI zu entwickeln. Die Studierenden zeigen sich in ihren Perspektiven selbstreflektiert und wägen das Vertrauen für die Dimensionen des Anvertrauens und Für-wahr-haltens in Beziehung zueinander ab. Die Grundhaltung als vorangehende Dimension erfährt eine weniger aktive Einschätzung und entwickelt sich passiv über einen längeren Zeitraum anhand der Entwicklung der beiden anderen Dimensionen. Die Studierenden durchlaufen mit neuen Erfahrungen und Veränderungen immer wieder den Prozess und bewerten ihr Vertrauen neu. Die Ergebnisse zeigen auch, dass durch diese

prozesshafte Entwicklung immer wieder ein schwankendes Potenzial für mehr und auch weniger Vertrauen in die KI besteht und von den Studierenden bewertet wird. Es wird deutlich, dass keine eindeutige klare Trennung von Misstrauen und Vertrauen existiert, sondern immer eine Mischform von beiden Aspekten vorhanden ist. Die Zusammensetzung aus den Anteilen von Vertrauen und Misstrauen ist flexibel und individuell. Außerdem kann sich das Verhältnis für die Vertrauensdimensionen unterscheiden, bspw. ist ein großes Vertrauen in den Output möglich und eine gleichzeitig große Skepsis gegenüber den Datenschutzvereinbarungen.

Die Ergebnisse zeigen zudem, dass die Anwendung der KI auch mit eingeschränktem Vertrauen möglich ist. Ebenso leitet sich die flexible, prozesshafte Aushandlung von Vertrauen möglicherweise durch generationsbedingte Technikaffinität der Teilnehmenden ab. Dadurch könnten bereits vorhandene Strukturen und Erfahrungen mit Vertrauen in neue Technologien als Referenzrahmen für die Entwicklung von Vertrauen in KI dienen. Besonders die eigenverantwortliche Anwendung neu entwickelter Technologien, wozu auch das Wissen um potenzielle Gefahren und Risiken gehört, spiegelt sich jetzt im Umgang mit der KI.

Die Unterscheidung in KI spezifische Dimensionen des Vertrauens, die sich auf unterschiedliche Aspekte der KI (Datenschutz und Output) beziehen, wurde bisher in der Literatur noch nicht aufgegriffen. Die Dimension des Grundvertrauens ist bereits als *Dispositional trust* von Hoff und Bashir (2015) bekannt und findet sich auch in diesen Daten wieder. Hoff und Bashirs Analyse zu Vertrauen in automatische Systeme kann also in diesem Punkt durch die aktuelle KI Entwicklung ergänzt werden. Auch die Konzepte des *Situational trust* und *Learned trust* lassen sich mit dieser Analyse bestätigen. Mit dem hier entwickelten Modell verwende ich zwar nicht dieselben drei Dimensionen. Dass das Konzept Vertrauen aber situationsabhängig und Vertrauen durch Erfahrungen oder angeeignetes Wissen beeinflusst wird, findet sich im Vertrauensprozess wieder. Die eigene Einschätzung der KI, Erfahrungen, Wissen und die Nachvollziehbarkeit werden im zweiten Schritt des Vertrauensprozess herangezogen, um das Vertrauen neu zu bewerten. Dabei entsteht auch themen- und kontextspezifisches Vertrauen in die KI.

Die Bedingungen im zweiten Schritt des Vertrauensprozess', insbesondere der Wunsch nach Nachvollziehbarkeit und Verständnis der KI, sind mit dem von Choung et al. (2022) eingeführten *functionality trust* vergleichbar. Das Vertrauen der Studierenden orientiert sich an der Leistung und der Funktion der KI, es wird vor allem durch Lernen und Erfahrung geprägt. Auch der Aspekt des *human-*

like trust kann den Daten zugeordnet werden. Zwar spielt diese Dimension in den von mir analysierten Daten eine wesentlich geringere Rolle, als das *functionality trust*. Jedoch kann das Vertrauen, ausgelöst durch menschliche Vergleichspunkte, zur Grundhaltung und zum ersten Schritt des Vertrauensprozess, in dem die Grundhaltung wahrgenommen wird, zugeordnet werden. Dabei spielen Erfahrungen mit menschlichen Vertrauensobjekten eine Rolle. Diese können auf den Umgang mit KI übertragen werden: Umso ähnlicher die KI menschliches Verhalten simuliert, desto mehr könnte das Vertrauen in die KI mit den selben Kriterien wie für menschliches Vertrauen bewertet werden.

Die Einteilung in *warranted* und *unwarranted trust* von Jacovi et al. (2021) ist insofern überholt, als dass die KI, beispielsweise Chat GPT, explizit in ihrer Datenschutzrichtlinie die Erhebung und Weiterverwendung von Daten formuliert, und auf jeder Chatoberfläche schreibt, dass Fehler auftreten können und wichtige Informationen geprüft werden müssen. Der von Jacovi et al. bezeichnete Vertrag beschreibt eine Abmachung zwischen den Nutzenden und der KI. In diesem Fall könnte die Datenschutzbestimmung als ein solcher Vertrag verstanden werden. Darin schreiben die Betreiber*innen der KI allerdings keine eindeutigen vertrauensbasierten Leistungen, sondern vielmehr, dass die Daten eben weiterverwendet und gespeichert werden (vgl. Open AI, 2025b). Die Frage, die sich damit stellt, ist vielmehr, wie trotzdem Vertrauen hergestellt werden kann und in welchem Rahmen und unter welchen Bedingungen. Ebenfalls von Jacovi et al. (2021) vorgestellt, findet sich die Unterscheidung in intrinsisches und extrinsisches Vertrauen in meiner Arbeit wieder. Diese Differenzierung beeinflusst die Unterscheidung der Vertrauenswürdigkeit eines KI-Modells. Das intrinsische Vertrauen, basierend auf den Funktionen und internen Prozessen der KI, spielt für die Studierenden eine große Rolle, wie man im zweiten Schritt des Vertrauensprozess, bei der Einschätzung des Vertrauens, erkennen kann. Der dabei am häufigsten genannte Faktor ist die Nachvollziehbarkeit der internen Prozesse. Das extrinsische Vertrauen zeigt sich dann im dritten Schritt des Prozess', durch die Strategie der Kontrolle des Outputs. Nicht die Nachvollziehbarkeit wie die KI zu ihrem Output gekommen ist, beeinflusst das Vertrauen, sondern die Bewertung der Qualität und Korrektheit des Outputs selbst.

6.2. Diskussion der Limitationen

Wie bereits im Kapitel 4.4 der forschungsethischen Reflexion angesprochen, wurden die Daten nicht von vornherein mit dem Ziel erhoben, die in dieser Arbeit konkret gestellte Forschungsfrage zu beantworten. Vielmehr wurden die Interviews aus einem breiteren Interessenspektrum heraus geführt, aus dem sich meine Frage im Laufe des Forschungsprozess' entwickelt hat. Dementsprechend

ist die Konzeption gerade während der Datenerhebung nicht auf das Forschungsinteresse zugeschnitten. Dadurch, dass der Leitfaden der Interviews unterschiedliche Themen beinhaltet, darunter auch Erfahrungen mit einem Interviewbot, und nur manche der Interviews auch Fragen zu Vertrauen in KI gestellt haben, unterscheidet sich der Fokus und der Informationsgehalt für meine Forschungsfrage in den Interviews. Außerdem sind sie die Interviews durch unterschiedliche Interviewer*innen geführt wurden, die sich durch mehr oder weniger Erfahrung in der qualitativen Sozialforschung ausgezeichnet haben. Dadurch sind heterogene Interviews mit unterschiedlich ausführlichen Antworten und Nachfragen entstanden.

Eine weitere Limitation der Forschung ist die fehlende Reflexion der Teilnehmenden in ihrem soziokulturellen Umfeld. Zwar wurde darauf geachtet, Studierende aus unterschiedlichen Fachrichtungen zu interviewen, jedoch wurden neben dem Studienfach nur Alter und Geschlecht festgehalten. Über beispielsweise den kulturellen Hintergrund, Bildungsabschluss der Eltern oder Migrationserfahrung wurden keine Informationen erhoben. Durch das Fehlen dieser Informationen über die Teilnehmenden und der breitgefächerten Thematik in den Interviews können keine Aussagen über einen möglichen kulturellen Einfluss auf das Vertrauen in KI getroffen werden. Außerdem kann keine Perspektive auf Unterschiede von Studierenden aus Akademikerfamilien und Arbeiterfamilien geworfen werden. Und auch die Auswahl der Studierenden aus München kann als Limitation angesehen werden. Es ist möglich, dass Studierende aus anderen Städten und von anderen Universitäten und Hochschulen anderen oder ähnlichen Einflüssen ausgesetzt sind. Vorstellbar wären weitere Interviewteilnehmer*innen aus ländlichen Regionen oder anderen Bundesländern zu befragen.

Daran anschließend komme ich nun final noch auf die Limitationen dieser Studie zu sprechen, die aufgrund des Umfangs einer Masterarbeit eine Rolle spielen. Durch den Bearbeitungszeitraum ist nur eine begrenzte Erhebung und Durchführung der Forschung nach der RGTM möglich. Der iterative Prozess kann in diesem Fall nicht weiter ausgeführt werden, ebenso wie der Umfang der Reflexion einen kleineren Anteil einnimmt. An dieser Stelle wäre es angebracht, im Sinne der RGTM, mit den Ergebnissen erneut ins Feld zu gehen, neue Daten zu erheben und sich der theoretischen Sättigung anzunähern. Die hier entwickelten Vertrauensdimensionen und der Vertrauensprozess könnte anhand weiterer kontrastierender Daten geprüft und erweitert werden. Ebenfalls ist es für eine Masterarbeit nicht möglich, wie in der RGTM eigentlich vorhergesehen, umfänglich in Forschungsgruppen zu arbeiten, auch wenn der Austausch mit anderen natürlich trotzdem möglich ist und Interpretationsworkshops gebildet werden können. Dies erweist sich allerdings als nur eingeschränkt möglich, wenn im Team nicht mit denselben Daten gearbeitet wird und die Prüfungsleistung als

eigenständige Forschungsarbeit durchgeführt werden muss. Trotz der Herausforderungen und Limitationen ist es möglich eine gelungene Forschung im Stil der RGTM durchzuführen und interessante Ergebnisse zu präsentieren, an die in weiterer Forschung angeschlossen werden kann.

7. Fazit

Mit dieser Arbeit ordne ich das Vertrauen von Studierenden in KI in die aktuelle Forschung zu KI und den technischen Stand ein. Dabei liegt der Fokus auf der Nutzer*innen Perspektive, konkret der studentischen Perspektive. Damit leiste ich einen Forschungsbeitrag zum konkreten Umgang und zur Integration von KI in den Alltag. Die Ergebnisse zeigen, dass Studierende nicht hilflos und unüberlegt mit der neuen Technologie umgehen, sondern aktiv und verantwortungsbewusst Wege suchen, die KI zu nutzen und ihr eigenes Vertrauen in sie zu beurteilen. Die grundsätzliche Mischung aus Zuversicht und Skepsis verdeutlicht das Potenzial für eine konstruktive und reflektierte Integration im Studium und im Alltag, aber auch die Berücksichtigung und Einschätzung von potenziellen Gefahren.

Da sich die KI kontinuierlich weiterentwickelt, muss auch das an die Bedingungen der KI geknüpfte Vertrauen als dynamischer, anpassungsfähiger Prozess verstanden werden. Weil die KI eine flexible und neue Technologie ist, die sich in den nächsten Jahren verändern wird, wird sich auch die Nutzer*innenperspektive und deren Vertrauen verändern. Darin liegt vor allem die Chance die KI zuverlässiger und nachvollziehbarer zu gestalten, um dadurch das Vertrauenspotenzial der Nutzer*innen zu erhöhen.

Um einen Blick über die aktuelle Einordnung hinauszuerwerfen und zukünftige Entwicklung und das Potenzial von KI zu berücksichtigen, gilt es die Interaktion zwischen Nutzenden und KI auch weiterhin zu erforschen und dabei neue Entwicklungen der KI und soziale Umstände und Veränderungen für die Nutzenden zu berücksichtigen.

Literaturverzeichnis

- Amoozadeh, M., Daniels, D., Nam, D., Kumar, A., Chen, S., Hilton, M., Srinivasa Ragavan, S., & Ali-pour, M. A. (2024). Trust in Generative AI among Students: An exploratory study. *Proceedings of the 55th ACM Technical Symposium on Computer Science Education V. 1*, 67–73.
<https://doi.org/10.1145/3626252.3630842>
- Banh, L., & Strobel, G. (2023). Generative artificial intelligence. *Electronic Markets*, 33(1), Article 1.
- Breuer, F., Muckel, P., & Dieris, B. (2019). *Reflexive Grounded Theory: Eine Einführung für die Forschungspraxis* (4. Aufl.). Springer Fachmedien. <http://link.springer.com/10.1007/978-3-658-22219-2>
- Charmaz, K. (2014). *Constructing grounded theory* (2nd edition). SAGE.
- Choung, H., David, P., & Ross, A. (2022). Trust in AI and Its Role in the Acceptance of AI Technologies. *International Journal of Human–Computer Interaction*, 39, 1727–1739.
<https://doi.org/10.1080/10447318.2022.2050543>
- Der Spiegel. (2023, März 7). ChatGPT an Hochschulen: Gen Z sieht Einsatz von KI im Studium kritisch. *Der Spiegel (online)*. <https://www.spiegel.de/start/chatgpt-an-hochschulen-gen-z-sieht-einsatz-von-ki-im-studium-kritisch-a-e17cc400-4d74-4281-9b3b-e52819e1c941>
- Dröge, K. (2024). *noScribe* (Version 0.5) [Software]. <https://github.com/kaixxx/noScribe>
- Endreß, M. (2024). Vertrauen. In J. Kopp & A. Steinbach (Hrsg.), *Grundbegriffe der Soziologie* (13., S. 551–554). Springer VS. https://doi.org/10.1007/978-3-658-42676-7_94
- Felder, E., & Kückelhaus, M. (2025). Das definierende Sprachmodell (LLM): Anthropomorphisierung in der Mensch-Maschine-Interaktion. *Zeitschrift für Literaturwissenschaft und Linguistik*, 55(2), 431–448. <https://doi.org/10.1007/s41244-025-00380-7>
- Giddens, A. (1995). *Die Konsequenzen der Moderne*. Suhrkamp.
- Glaser, B. G., & Strauss, A. L. (1967). *The Discovery of Grounded Theory: Strategies for Qualitative Research*. Aldine Transaction.
- Hawley, K. (2014). Trust, Distrust and Commitment. *Noûs*, 48(1), 1–20.
<https://doi.org/10.1111/nous.12000>

- Hitzler, R. (2016). Zentrale Merkmale und periphere Irritationen interpretativer Sozialforschung. *Zeitschrift für Qualitative Forschung*, 17(1–2), 171–184. <https://doi.org/10.3224/zqf.v17i1-2.25549>
- Hockey, J. (1993). Research methods -- researching peers and familiar settings. *Research Papers in Education*, 8(2), 199–225. <https://doi.org/10.1080/0267152930080205>
- Hoff, K. A., & Bashir, M. (2015). Trust in Automation: Integrating Empirical Evidence on Factors That Influence Trust. *Human Factors*, 57(3), 407–434. <https://doi.org/10.1177/0018720814547570>
- Jacovi, A., Marasović, A., Miller, T., & Goldberg, Y. (2021). Formalizing Trust in Artificial Intelligence: Prerequisites, Causes and Goals of Human Trust in AI. *Proceedings of the 2021 ACM Conference on Fairness, Accountability, and Transparency*, 624–635. <https://doi.org/10.1145/3442188.3445923>
- Kuhn, A., & Hartmann, S. (2023). Das "Black-Box Phänomen" In der KI-Entwicklung. *Kompetenzzentrum Arbeitswelt.Plus*, 5–16. <https://doi.org/10.55594/SFAH4426>
- Lee, J. D., & See, K. A. (2004). Trust in Automation: Designing for Appropriate Reliance. *Human Factors*, 46(1), 50–80. https://doi.org/10.1518/hfes.46.1.50_30392
- Luhmann, N. (2014). *Vertrauen: Ein Mechanismus der Reduktion sozialer Komplexität* (5. Aufl.). utb.
- Lukyanenko, R., Maass, W., & Storey, V. C. (2022). Trust in artificial intelligence: From a Foundational Trust Framework to emerging research opportunities. *Electronic Markets*, 32(4), 1993–2020. <https://doi.org/10.1007/s12525-022-00605-4>
- Marsh, S., & Dibben, M. R. (2003). The role of trust in information science and technology. *Annual Review of Information Science and Technology*, 37(1), 465–498. <https://doi.org/10.1002aris.1440370111>
- Mayer, R. C., Davis, J. H., & Schoorman, F. D. (1995). An Integrative Model of Organizational Trust. *The Academy of Management Review*, 20(3), 709–734. <https://doi.org/10.2307/258792>
- Mcknight, D. H., Carter, M., Thatcher, J. B., & Clay, P. F. (2011). Trust in a specific technology: An investigation of its components and measures. *ACM Trans. Manage. Inf. Syst.*, 2(2), 12:1–

12:25. <https://doi.org/10.1145/1985347.1985353>

Murašov, E. (2023, Juli 11). Wie ChatGPT im Studium nutzen? HU-Studierende vermissen KI-Leitlinien. *Der Tagesspiegel Online*. <https://www.tagesspiegel.de/wissen/wie-chatgpt-im-studium-nutzen-hu-studierende-vermissen-ki-leitlinien-10126477.html>

Open AI. (2025a). *ChatGPT Deutsch*. chatopenai.de. <https://chatopenai.de/>

Open AI. (2025b, Juni 27). *Datenschutzerklärung*. openai.com. <https://openai.com/de-DE/policies/row-privacy-policy/>

Strübing, J. (2018). Grounded Theory. In *Qualitative Sozialforschung: Eine kompromierte Einführung* (2. Aufl., S. 121–144). De Gruyter.

von Unger, H. (2014). Forschungsethik in der qualitativen Forschung: Grundsätze, Debatten und offene Fragen. In *Forschungsethik in der qualitativen Forschung* (S. 15–39). Springer VS, Wiesbaden. https://doi.org/10.1007/978-3-658-04289-9_2

Von Unger, H. (2018). Archivierung und Nachnutzung qualitativer Daten aus forschungs- ethischer Perspektive. *RatSWD Working Paper*, 267, 91–100.

Witzel, A., & Reiter, H. (2022). *Das problemzentrierte Interview – eine praxisorientierte Einführung* (1. Aufl.). Beltz Juventa.

Anhang

Leitfäden

a) Für Face-to-Face Interviews anschließend an das KI-Interview

1. Wie war es? Wie fandest Du das Interview mit dem Qualiabot?
2. Was fandest Du gut?
3. Was fandest Du nicht so gut?
4. Hat Dich irgendwas überrascht?
5. Würdest Du Dich nochmal von einer KI interviewen lassen?
6. Anwendungen mit KI werden sehr schnell weiterentwickelt, wie stellst du dir die Zukunft mit KI vor?

Wir wollen jetzt noch einen Fokus setzen. (Individueller forschungsgruppenspezifischer Fragenteil)

7. ...
8. ...
9. Interview Abschluss: Hast Du noch irgendwelche Anmerkungen oder Fragen?

b) Für Face-to-Face Interviews ohne vorheriges KI-Interview

1. Du hast Dich gegen ein KI-gestütztes Interview entscheiden – was waren dabei die Haupt-Beweggründe?
2. Wie würdest Du Deine Einstellung oder Haltung zu KI allgemein beschreiben?
3. Hast Du praktische Erfahrungen in der Anwendung?
4. Welche Rolle spielt KI in deinem Studium?
5. Unter welchen Umständen würdest Du Dich evtl. doch mal von einer KI interviewen lassen?
6. Anwendungen mit KI werden sehr schnell weiterentwickelt, wie stellst du dir die Zukunft mit KI vor?

Wir wollen jetzt noch einen Fokus setzen. (Individueller forschungsgruppenspezifischer Fragenteil)

7. ...
8. ...
9. Interview Abschluss: Hast Du noch irgendwelche Anmerkungen oder Fragen?

Informierte Einwilligung



Einwilligung zum Interview

„Mit KI über KI qualitativ forschen“
Lehrforschungsprojekt 2024-2025

Am Institut für Soziologie der Ludwig-Maximilians-Universität (LMU) München führen Studierende im Rahmen ihrer methodischen Ausbildung Forschungsprojekte durch. Ein solches Lehrforschungsprojekt beschäftigt sich mit dem Thema Künstliche Intelligenz (KI). In Übungen, Seminaren und Forschungspraktika untersuchen Studierende Einstellungen zu KI. Gleichzeitig erproben sie die Grenzen und Möglichkeiten des Einsatzes von künstlicher Intelligenz in der qualitativen Forschung. Das Lehrforschungsprojekt „Mit KI über KI qualitativ forschen“ wird von Prof. Dr. Hella von Unger geleitet und gemeinsam mit den Studierenden und wissenschaftlichen Mitarbeiterinnen und Mitarbeitern des Instituts (u.a. PD Dr. Yves Jeanrenaud) umgesetzt.

Interviews

Dazu werden qualitative Interviews mit Studierenden durchgeführt. Die Interviews sind offene Gespräche zum Thema Künstliche Intelligenz. Um an der Studie teilzunehmen, ist es nicht erforderlich, praktische Erfahrungen mit KI (z.B. ChatGPT) zu haben. Von Interesse sind vor allem Einstellungen, Überzeugungen und Haltungen zum Thema.

Das Interview dauert ca. 30-35 Minuten. Es wird aufgezeichnet und anschließend verschriftlicht. **Ich kann selbst entscheiden, ob ich mit oder ohne KI-Unterstützung interviewt werden möchte.** Mit KI-Unterstützung bedeutet: Ich werde zunächst schriftlich von einem Interviewbot („Qualia“) befragt (ca. 15 min) und führe anschließend ein mündliches Gespräch (ca. 15 min) über diese Erfahrung mit einer menschlichen Interviewerin/einem menschlichen Interviewer. Ohne KI-Unterstützung bedeutet: Das Interview findet ausschließlich mit einer menschlichen Interviewerin/ einem menschlichen Interviewer statt.

- Ich möchte mit KI-Unterstützung interviewt werden. Der Datenverarbeitung (siehe „KI Tools“ unten) stimme ich zu.
- Ich möchte ohne KI-Unterstützung interviewt werden.

Vertraulichkeit und Datenschutz

Die Daten werden vertraulich behandelt und anonymisiert, d.h. Informationen zu meiner Person werden unkenntlich gemacht. Außerdem werden Fragen von vorn herein so gestellt, dass keine personenbezogenen Daten generiert werden. Die anonymen bzw. anonymisierten Daten werden auf Servern des Rechenzentrums der LMU (LRZ in Garching) sicher verwahrt. Sie werden datenschutzrechtlichen und forschungsethischen Grundsätzen entsprechend verarbeitet (siehe unten). Die Audioaufnahmen werden im August 2025 gelöscht. Alle anderen Forschungsdaten werden (in anonymer/anonymisierter Form) für 5 Jahre (bis August 2029) aufbewahrt und anschließend gelöscht.

Die Interviews werden von den Studierenden zu Lehr- und Forschungszwecken durchgeführt. Die Studierenden werten die Daten für Prüfungsleistungen aus und schreiben Hausarbeiten, ggf. auch Abschlussarbeiten, darüber. Außerdem sind wissenschaftliche Publikationen zu den o.g. und benachbarten Themen (studentische Einstellungen zu KI, Möglichkeiten und Grenzen der Nutzung von KI in der qualitativen Forschung und Lehre) geplant.

KI-Tools

Nicht nur in den Interviews, sondern auch bei der Aufbereitung und Auswertung der Daten werden KI-Tools erprobt. Ich kann dem zustimmen, oder dies ablehnen. Wenn ich der Verwendung von KI-Tools zustimme, bedeutet dies, dass die anonymen bzw. anonymisierten Daten auch auf Rechnersystemen außerhalb des europäischen Wirtschaftsraums unter Berücksichtigung entsprechender Auftragsdatenverarbeitungsverträge, z. B. mit OpenAI (OpenAI L.L.C., 3180 18th St, San Francisco, CA 94110, USA) und/oder Amazon AWS (Amazon Web Services EMEA SARL, 38 Avenue John F. Kennedy, L-1855, Luxemburg) gespeichert und verarbeitet werden. Die dort gespeicherten Daten werden nicht zu Trainingszwecken von KI-Systemen genutzt und nach spätestens 30 Tagen gelöscht. Die Zustimmung dazu ist optional.

- Ich stimme der Auswertung mit KI-Tools zu
- Ich stimme der Auswertung mit KI-Tools nicht zu

Teilnahme und Erklärung

Die Teilnahme an den Interviews ist freiwillig und ohne Gegenleistung. Ich habe zu jeder Zeit die Möglichkeit, das Interview abzubrechen, Fragen nicht zu beantworten oder meine Einwilligung zurückzuziehen, ohne dass mir dadurch ein Nachteil entsteht. Bei Nachfragen oder dem Widerruf meiner Einwilligung wende ich mich an die Interviewer/in oder an die Seminarleiterin (Prof. Dr. Hella von Unger, Institut für Soziologie der LMU München, Konradstraße 6, 80801 München; Tel. (089) 2180-6315; E-Mail: unger@lmu.de). Weiterführende Informationen zum Datenschutz und meinen Rechten als Studienteilnehmer/in finde ich untenstehend.

- Ich habe eine Kopie der Studieninformation und Einwilligungserklärung erhalten und bin damit einverstanden, an einem Interview teilzunehmen. Ich hatte ausreichend Gelegenheit, Rückfragen an den/die Interviewer/in zu stellen. Mir ist alles erklärt worden, um diese Einwilligung freiwillig und informiert abzugeben. **Ohne** diese Erklärung kann ich **nicht teilnehmen**.

..... / /

Datum

.....
Name des Teilnehmers/der Teilnehmerin

.....
Unterschrift

.....
Name des Interviewers/der Interviewerin

.....
Unterschrift

Weiterführende Informationen

Die erhobenen Daten werden gemäß § 4 Abs. 1 Bundesdatenschutzgesetz (BDSG) und mit Geltung der Datenschutzgrundverordnung (DSGVO), ab dem 25. Mai 2018, gemäß Art. 6 Abs. 1 lit. a DSGVO verarbeitet. Sofern besondere Kategorien personenbezogener Daten betroffen sind, werden diese gemäß § 4 Abs. 1 BDSG und mit Geltung der Datenschutzgrundverordnung (DSGVO) gemäß Art. 9 Abs. 2 lit. a DSGVO verarbeitet.

Die Einwilligungserklärungen werden für 10 Jahre (getrennt von den Daten) am Institut für Soziologie aufbewahrt und dann vernichtet. Für die Teilnehmer/innen dieser Studie wurde keine Versicherung (z.B. Wegeversicherung) abgeschlossen. Dies erfolgte aufgrund der Einschätzung, dass die Teilnahme mit sehr geringen Risiken verbunden ist.

Im Rahmen der gesetzlichen Vorgaben habe ich als Teilnehmer/in grundsätzlich Anspruch auf:

- Eine Bestätigung, ob mich betreffende personenbezogenen Daten verarbeitet werden,
- Auskunft über diese Daten und die Umstände der Verarbeitung,
- Berichtigung, soweit diese Daten unrichtig sind,
- Löschung, soweit für die Verarbeitung keine Rechtfertigung und keine Pflicht zur Aufbewahrung (mehr) besteht,
- Einschränkung der Verarbeitung in besonderen gesetzlich bestimmten Fällen und
- Übermittlung meiner personenbezogenen Daten – soweit ich diese bereitgestellt habe – an mich oder einen Dritten in einem strukturierten, gängigen und maschinenlesbaren Format.

Das bedeutet, ich habe das Recht, Auskunft über die mich betreffenden, personenbezogenen Daten zu erhalten sowie deren Berichtigung oder deren Löschung zu verlangen, so lange die Daten mit meiner Person in Verbindung gebracht werden können. Bei einer erfolgreichen Anonymisierung ist dies ggf. nicht mehr möglich.

**Behördlicher
Datenschutzbeauftragter der
Universität:**
Ludwig-Maximilians-Universität (LMU)
Behördlicher Datenschutzbeauftragter
Geschwister-Scholl-Platz 1
80539 München
E-Mail: datenschutz@lmu.de

Datenschutz-Aufsichtsbehörde:
Bayerischer Landesbeauftragter für den
Datenschutz (BayLfD)
Prof. Dr. Thomas Petri
Wagmüllerstr.1
80538 München
Tel.: (089) 212672-0

Verantwortliche Dozierende des Lehrforschungsprojekts:

Prof. Dr. Hella von Unger
Institut für Soziologie der LMU München
Konradstraße 6
80801 München
Tel. (089) 2180-6315
E-Mail: unger@lmu.de

PD Dr. habil. Yves Jeanrenaud
Institut für Soziologie der LMU München
Konradstraße 6
80801 München
Tel. (089) 2180-5606
E-Mail: yves.jeanrenaud@lmu.de