Zürcher Hochschule
für Angewandte Wissenschaften

Soziale Arbeit

Künstliche Intelligenz im Kindesschutz

Eine Untersuchung zur Wahrnehmung von KI-gestützten Entscheidungshilfen im Kindesschutz

Master-Thesis
Simone Müller

Begleitperson
David Lätsch

Zweitgutachter
Tim Tausendfreund

Master of Science
in Sozialer Arbeit
Zürich, Frühlingssemester 2024

Zürcher Hochschule
für Angewandte Wissenschaften

Soziale Arbeit

Masterthesis

Künstliche Intelligenz im Kindesschutz

Eine Untersuchung zur Wahrnehmung von KI-gestützten Entscheidungshilfen im Kindesschutz

Simone Müller
FS2020

Begleitperson: Prof. Dr. David Lätsch
Zweitgutachter: Dr. Tim Tausendfreund

Master of Science in Sozialer Arbeit
Zürcher Hochschule für
Angewandte Wissenschaften

Abgabedatum: 18. Juni 2024

Titelbild kreiert mit Dall-E

Abstract

Die vorliegende Masterarbeit untersucht das Vertrauen von Sozialarbeitenden im Kindesschutz in KI-gestützte Entscheidungshilfen und die Bereitschaft, entsprechende Tools in den Arbeitsalltag zu integrieren. Im Rahmen einer Online-Umfrage nahmen 204 Berufsbeistandspersonen aus der Deutschschweiz teil. Die Untersuchung basiert auf dem modifizierten Technologieakzeptanzmodell (*TAM*) und überprüft fünf Hypothesen, die sich auf Faktoren wie *Invasivität*, *prädiktive Validität* und *wissenschaftliches Wissen* beziehen.

Das Design der Umfrage verwendete ein multifaktorielles, experimentelles Vignettendesign, bei dem den Teilnehmenden verschiedene hypothetische Szenarien zugewiesen wurden. Diese Szenarien variierten die Hauptfaktoren, um deren Einfluss auf das Vertrauen in KI-gestützte Entscheidungshilfen zu messen. Die beiden Fallbeispiele enthielten unterschiedliche fiktive KI-Tools, welche möglicherweise in Zukunft in ähnlicher Form im Kindesschutz eingesetzt werden könnten. Dabei handelte es sich einerseits um ein «Risikomodell» (*KIBAT*) und um ein «Chancenmodell» (*WIEW*). Die Ergebnisse zeigen, dass eine geringere *Invasivität* von Einschätzungen und die Wahrnehmung von hoher *prädiktiver Validität* (Treffsicherheit) von Tools das Vertrauen in die KI-gestützten Empfehlungen erhöhen, während die Wahrnehmung von *wissenschaftlichem Wissen* innerhalb der Empfehlungen allein keinen Einfluss auf das Vertrauen in KI-gestützte Tools hat. Zudem korreliert eine positive Einstellung gegenüber KI-Technologien stark mit einem höheren Vertrauen. Ein höheres Vertrauen in die Tools beeinflusst ausserdem die Bereitschaft, die Tools im Arbeitsalltag zu integrieren.

Die Thesis hebt hervor, dass eine sorgfältige Einführung und Schulung, die Transparenz der Funktionsweise und die ethische Verantwortlichkeit entscheidend für das Vertrauen und eine erfolgreiche Implementierung von KI in der Sozialen Arbeit sind. Durch die Kombination menschlicher Expertise und technologischer Unterstützung können objektivere und fundiertere Entscheidungen getroffen werden, was die Handlungssicherheit und die Qualität im Kindesschutz verbessert. Die Studie betont die Notwendigkeit, KI als ergänzendes Werkzeug zu betrachten, das die professionelle Urteilsfähigkeit stärkt, ohne sie zu ersetzen. Es wird empfohlen, die Soziale Arbeit möglicherweise neu zu denken, um den bestmöglichen Einsatz der Technologie zu gewährleisten und die professionellen Standards zu erhalten und zu verbessern.

Inhaltsverzeichnis

Abstract II
Tabellen V
Abbildungen VI
1 Einleitung 1
1.1 Relevanz und Untersuchungsziel 2
1.2 Forschungsstand 3
1.3 Aufbau der Arbeit 4
2 Grundlagen und Begriffsbestimmung 5
2.1 Übersicht Kindesschutz in der Schweiz 5
2.2 Menschliche Entscheidungsfindung und kognitive Verzerrung 6
2.3 Künstliche Intelligenz (KI) 9
2.3.1 Algorithmen 10
2.3.2 Neuronale Netzwerke 11
2.3.3 Machine Learning und Deep Learning 11
2.3.4 Die Grenzen der KI 12
2.4 Vertrauen 14
3 Künstliche Intelligenz im Kindesschutz 17
3.1 Bestehende KI-Entscheidungshilfen im Kindesschutz 17
3.1.1 Allegheny Family Screening Tool (*AFST*) 18
3.1.2 Predictive Risk Modelling (*PRM*) 20
3.2 Herausforderungen und positive Entwicklungen der KI-Entscheidungshilfen 23
3.2.1 Herausforderungen 23
3.2.2 Positive Entwicklungen 24
4 Theoretische und konzeptionelle Zugänge 26
4.1 Das Theorie Akzeptanzmodell (*TAM*) 26
4.2 Angepasstes Theoriemodell und Hypothesenbildung 27
4.2.1 Externe Faktoren 29
4.2.2 Eigenschaften und Einstellung der Vertrauenden 32
4.2.1 Integrationsbereitschaft 33
5 Methodisches Vorgehen 34
5.1 Forschungsdesign: multifaktorielles, experimentelles Vignettendesign 34
5.1.1 Charakteristika des experimentellen Vignettendesigns 35
5.1.2 Vorteile und Limitationen des Designs 36

5.1.3 Beurteilung der Vignetten ... 37
5.2 Stichprobe ... 38
5.3 Erhebungsinstrument ... 40
5.3.1 Experimentelle Vignetten ... 40
5.3.2 Operationalisierung der Variablen ... 44
5.4 Auswertungsmethode ... 50
5.4.1 Datenaufbereitung ... 50
5.4.2 Analysestrategie ... 50
6 Ergebnisse ... 54
6.1 Charakterisierung der Stichprobe ... 54
6.1.1 Digitale Affinität und Einstellung zu KI nach Altersgruppe und Geschlecht ... 56
6.2 Abhängige Variabel: Vertrauen in KI-gestützte Entscheidungshilfen ... 57
6.2.1 Personenmerkmale und Vertrauen ... 63
6.3 Beurteilung der Vignetten und experimentelles Design ... 64
6.3.1 Manipulations-Check der Dimensionen ... 64
6.3.2 Mittelwertvergleiche Vignettendimensionen *KIBAT* und *WIEW* ... 65
6.3.3 Multiple Regressionsanalyse ... 67
6.4 Integrationsbereitschaft ... 73
6.5 Zusammenfassung der Ergebnisse ... 74
7 Zentrale Befunde und Diskussion ... 76
7.1 Einfluss der variierten Dimensionen ... 76
7.1.1 Einfluss der Invasivität ... 76
7.1.2 Einfluss der prädiktiven Validität ... 77
7.1.3 Einfluss des wissenschaftlichen Wissens ... 78
7.2 Einfluss der Personenmerkmale ... 79
7.3 Einfluss auf die Integrationsbereitschaft ... 80
7.4 Unterschiede zwischen den beiden Tools *KIBAT* und *WIEW* ... 80
7.5 Weitere Erkenntnisse aus der Umfrage ... 81
7.6 Kritische Reflexion und Limitationen ... 85
8 Fazit ... 87
8.1 Beantwortung der Fragestellung ... 87
8.2 Implikationen für die Soziale Arbeit ... 88
8.3 Schlusswort ... 90
9 Literaturverzeichnis ... 92
10 Anhang ... 100

Tabellen

Tabelle 1 – Original-Auszug aus den Variablen des Predictive Risk Modelling (PRM) 21
Tabelle 2 – Stichprobenbeschreibung 39
Tabelle 3 – Verkürzte Darstellung der Dimensionen und Ausprägungen der Vignetten in dieser Thesis 40
Tabelle 4 – Einleitung Fallbeispiele (Vignetten) KIBAT und WIEW 42
Tabelle 5 – Übersicht über die Ausprägung der Vignettendimensionen und Abschluss Fallbeispiele 43
Tabelle 6 – Übersicht der Reihenfolge und Ausprägungen der jeweilig präsentierten Vignetten pro Gruppe 44
Tabelle 7 – Reliabilitätsanalyse der zu neuen Variablen zusammengefassten Items 49
Tabelle 8 – Übersicht Manipulationsfrage und gültige/ungültige Antworten 52
Tabelle 9 – Demografische Merkmale 54
Tabelle 10 – Ausbildung und Berufserfahrung 55
Tabelle 11 – Kreuztabelle Digitale Affinität und Einstellung gegenüber KI nach Altersgruppe und Geschlecht 56
Tabelle 12 – Mittelwertvergleiche des Vertrauens in KIBAT 60
Tabelle 13 – Mittelwertvergleiche des Vertrauens in WIEW 62
Tabelle 14 – Korrelationsanalyse der Personenmerkmale und dem Vertrauen in die beiden Tools 63
Tabelle 15 – Manipulations-Check KIBAT – Auswertung Antworten nach Dimension 64
Tabelle 16 – Manipulations-Check WIEW – Auswertung Antworten nach Dimension 65
Tabelle 17 – Mittelwertvergleiche: Vertrauen in Fallbeispiel KIBAT 66
Tabelle 18 – Mittelwertvergleiche: Vertrauen in Fallbeispiel WIEW 67
Tabelle 19 – Multiple Regression KIBAT: Betrachtung des Vertrauens auf die unabhängigen Variablen 70
Tabelle 20 – Multiple Regression WIEW: Betrachtung des Vertrauens auf die unabhängigen Variablen 72
Tabelle 21 – Korrelationsanalyse Vertrauen und Integrationsbereitschaft 74

Abbildungen

Abbildung 1 – Leistungsbestandteile der Künstlichen Intelligenz....9
Abbildung 2 – Integratives Vertrauensmodell....16
Abbildung 3 – Funktionsweise Allegheny Family Screening Tool (AFST)....19
Abbildung 4 – Technologieakzeptanzmodell (TAM)....26
Abbildung 5 – Modifiziertes Hypothesenmodell in Anlehnung an TAM und Vertrauensmodell....29
Abbildung 6 – Boxplot Digitale Affinität und Einstellung gegenüber KI nach Altersgruppen....57
Abbildung 7 – Säulendiagramm zum Vertrauen in KIBAT (n = 196)....58
Abbildung 8 – Boxplot Vertrauen in KIBAT nach Altersgruppe und Geschlecht....59
Abbildung 9 – Säulendiagramm zum Vertrauen in WIEW (n = 198)....60
Abbildung 10 – Boxplot Vertrauen in WIEW nach Altersgruppe und Geschlecht....61
Abbildung 11 – Histogramme des Vertrauens der beiden Fallbeispiele KIBAT und WIEW nach Vignette....62

1 Einleitung

Spätestens seit der Einführung des Online-Chatbots ChatGPT im November 2022 ist Künstliche Intelligenz (KI) mit all ihren Teilbereichen in aller Munde und hat sich als einer der meistdiskutierten Trends der modernen Technologielandschaft etabliert (Kreutzer, 2023, S. 1). Die Sensibilisierung für neuartige Technologien ist hoch: Unternehmen weltweit erforschen Wege, um Geschäftsprozesse mithilfe von digitalen Tools zu optimieren – im Gesundheitswesen werden fortschrittliche Systeme für Diagnosen und Patientenmanagement eingesetzt, in der Finanzbranche werden Algorithmen für den Handel und die Betrugserkennung entwickelt, im Einzelhandel werden Technologien zur Kundenanalyse und -bindung genutzt, Transport- und Logistikunternehmen erforschen die Möglichkeiten autonomer Systeme und im Entertainmentbereich revolutionieren KI-gesteuerte Empfehlungssysteme das Nutzungsverhalten. Währenddessen erlassen öffentliche Einrichtungen Richtlinien und nehmen Gesetzesanpassungen für die Verwendung von KI vor, um den technologischen Fortschritt im Alltag zu regulieren, Kontrollmechanismen zu etablieren und potenziellen Risiken oder Missbräuchen vorzubeugen.

Selbst Menschen ohne tiefgreifende IT-Kenntnisse werden täglich mit dem Einfluss und der Möglichkeiten konfrontiert, die die neuen Technologien bieten. KI-Programme sind auf dem Vormarsch und bieten scheinbar unendlich viele Möglichkeiten. In der Geschäftswelt hat die Einführung der modernen Technologien bereits transformative Auswirkungen gezeigt, Effizienzsteigerung ist dabei ein Schlüsselaspekt: KI-Systeme automatisieren repetitive Aufgaben und setzen damit Zeit sowie Ressourcen frei. Intelligente Algorithmen tragen dazu bei, Prozesse zu optimieren und Kosten zu senken. Zugleich ermöglicht die Nutzung von KI im Kundenservice beispielsweise personalisierte Kundenbetreuung, Produktvorschläge und massgeschneiderte Marketingstrategien (vgl. Buxmann & Schmidt, 2019). Und nicht zuletzt stellt sich auch die Frage, ob durch den Einsatz von KI der Fachkräftemangel adressiert werden könnte, da KI voraussichtlich zuerst einfache Büroarbeiten revolutioniert und somit Menschen in diesen Bereichen anderweitig eingesetzt werden könnten (Winkler, 2023).

Trotz des beispiellosen Fortschritts und der vielfältigen Anwendungsmöglichkeiten werden die digitalen Fortschritte von vielen mit einer Mischung aus Faszination und Vorsicht betrachtet. KI scheint ein mächtiges Werkzeug zu sein, das sowohl das Potenzial für revolutionäre Veränderungen birgt als auch tiefgreifende ethische und soziale Fragen aufwirft. Ein Blick in die Medienlandschaft zeigt eine Flut von Artikeln, die sowohl das transformative Potential von KI als auch die Notwendigkeit für einen vorsichtigen und ethisch fundierten Einsatz hervorheben. In der Zwischenzeit warnen KI-Expert:innen gar davor, dass KI ausser Kontrolle geraten könnte (Fulterer, 2023) und auch Stephen Hawking hat 2014 Bedenken geäussert, dass eine vollständige KI das Ende der Menschheit bedeuten könnte. Kritiker befürchten zudem, dass der massenhafte Einsatz von neuen Technologien unsere Arbeitsplätze ersetzen könnte, und somit die sozioökonomischen Strukturen destabilisiert. Oder dass KI-Algorithmen – wenn sie nicht sorgfältig entwickelt und überwacht werden – voreingenommen (*biased*) sein könnten, was zu systemischen Ungerechtigkeiten führen kann (Eubanks, 2018).

1.1 Relevanz und Untersuchungsziel

Auch in der Sozialen Arbeit werden Befürchtungen gegenüber neuartigen Technologien geäussert. Die Sorgen sind vielfältig: Von der Authentizität der Interaktionen bis zu den Auswirkungen auf menschliche Beziehungen und die allgemeine psychische Gesundheit (Brunner, 2023) oder sogar von einer «De-Professionalisierung» hinsichtlich des Einsatzes von KI- und Big Data-Anwendungen wird gesprochen (Schneider, 2021, S. 123). Gleichzeitig werden aber auch Hoffnungen gegenüber der neuen Technologie ausgesprochen, dass beispielsweise die Bürokratie durch deren Einsatz deutlich reduziert werden kann und damit mehr Zeit für Klient:innen zur Verfügung steht (Evans & Hilbert, 2020, S. 84).

Das Plädoyer aus der Sozialen Arbeit lautet dabei immer gleich: Eine kritische Auseinandersetzung mit der Digitalisierung und eine aktive Mitgestaltung durch Sozialarbeitende ist zentral, wenn es um eine Integrierung der neuen Technologien in den Arbeitsalltag geht (Brunner, 2023). Der Einbezug moderner Technologien in die Soziale Arbeit erfordert eine kritische und professionsethische Auseinandersetzung mit den aktuellen und zukünftigen Entwicklungen (Brunner, 2023; Schulze, 2019; Gillingham, 2019). Gillingham (2019) betont, dass es wichtig sei, dass Sozialarbeitende sich der digitalen Entwicklungen bewusst sind und sich sowohl an der Debatte über digitaler Unterstützungssysteme als auch an deren Entwicklungsprozessen beteiligen sollen.

Die Digitalisierung schreitet in der Sozialen Arbeit zwar voran, hinkt jedoch im Vergleich zu anderen Fachbereichen trotz ihres grossen Potenzials hinterher. Es bleibt vorerst ungewiss, welche Auswirkungen die technologischen Entwicklungen auf Arbeitsbedingungen, Organisation der Arbeit, erforderliche Qualifikationen und Tätigkeitsinhalte haben wird (Evans & Hilbert, 2020, S. 80). Aktuell gibt es kaum Orientierungsleitlinien, die vorschlagen, wie Arbeitsprozesse mit Hilfe digitaler Technologien zukunftsorientiert gestaltet werden können und in welchen Bereichen entsprechende Anpassungsmöglichkeiten zu finden sind (Evans & Hilbert, 2020, S. 80).

Strukturen und Arbeitsweisen in der Sozialen Arbeit entwickeln sich mit der Digitalisierung weiter (Schneider, 2021). Dabei hat die Digitalisierung erhebliche Auswirkungen auf die Soziale Arbeit und stellt damit auch bestehende Theorien und Praktiken in Frage. Die Einführung neuer Technologien hat das Potenzial, das Professionsverständnis der Sozialen Arbeit zu verändern. Dies könnte zu einem weiterentwickelten Verständnis führen, da vorher nicht-standardisierbares Wissen nun durch algorithmisch gestützte Tools messbar wird (Bastian & Schrödter, 2019, S. 40). Bastian und Schrödter (2019, S. 40) sprechen hierbei sogar von einer «Digitalisierungsrevolution».

Während die Implementierung neuer Systeme in der Sozialen Arbeit und insbesondere im Kindesschutz in zahlreichen Studien und Pilotprojekten untersucht wird, bleibt die Haltung und die Einstellung von Sozialarbeitenden gegenüber neuartigen Technologien derweil weitgehend unerforscht. Zwar wird davon ausgegangen, dass die Offenheit und Souveränität gegenüber neuen Technologien verbreiteter sind als zunächst angenommen (Evans & Hilbert, 2020, S. 79). Gleichzeitig scheinen Sozialarbeitende aber auch grosse Vorbehalte gegenüber der voranschreitenden Digitalisierung zu haben (Schulze, 2019, S. 386-387). Dies liegt vermutlich einerseits an einer grundsätzlichen Orientierungslosigkeit gegenüber den komplexen Technologien und andererseits an der bisher geringen

Auseinandersetzung mit dem Thema im beruflichen, aber auch im Ausbildungskontext. Die Erkenntnis, dass digitalisierte Lebenswelten neue konzeptionelle Antworten erfordern, setzen sich im breiten Feld der Sozialen Arbeit nur sehr langsam durch. Ähnlich variabel wie die individuelle Wahrnehmung und Reaktion auf digitale Phänomene ist dabei auch das Tempo, mit dem die Digitalisierung in der Sozialarbeit voranschreitet (Schulze, 2019, S. 387). Eine zentrale Herausforderung für Sozialarbeitende besteht darin, digitale Kompetenzen zu erwerben und zu entwickeln, um den Anforderungen der digitalisierten Gesellschaft gerecht zu werden. Diese Kompetenzen umfassen ein grundlegendes Verständnis der digitalen Einflüsse auf verschiedenen Ebenen (Mikro, Meso, Makro) sowie eine Reflexion der eigenen professionellen Rolle im Kontext der Digitalisierung (Schulze, 2019, S. 388).

Wo stehen Sozialarbeitende, wenn es um die Implementierung von neuartigen Technologien in die Arbeitsfelder der Sozialen Arbeit geht? Sind sie gewillt, entsprechende Tools in ihrem Arbeitsalltag einzusetzen, oder steht ein Misstrauen im Vordergrund? Diese Masterthesis zielt darauf ab, die Perspektiven von Sozialarbeitenden bezüglich KI-Technologien im Kontext des Kindesschutzes zu beleuchten. Dabei wird der Fokus darauf gelegt, wie die Sozialarbeitenden KI-gestützte Entscheidungshilfen bewerten, wenn Einschätzungen auf Basis mathematischer Formeln getroffen werden. Primär sollen dabei Einflussfaktoren auf das Vertrauen gefunden und ihre Stärke gemessen werden. Die Fragestellung lautet:

> **Wie beeinflussen verschiedene Faktoren das Vertrauen von Sozialarbeitenden im Kindesschutz in KI-gestützte Entscheidungshilfen, und wie wirkt sich dieses Vertrauen auf ihre Bereitschaft aus, diese Tools in den Arbeitsalltag zu integrieren?**

Die Masterthesis verfolgt dabei zwei Hauptziele:

1. Sie möchte die Forschungsdesiderate hinsichtlich der Wahrnehmung und Einstellung von Sozialarbeitenden gegenüber neuartigen Technologien im Kindesschutz erschliessen. Hierbei soll insbesondere analysiert werden, welche Faktoren die Wahrnehmung der Sozialarbeitenden beeinflussen.
2. Sie möchte Jugendhilfestellen und Berufsbeistandschaften Einblicke in Meinungen und Reaktionen von Sozialarbeitenden gegenüber KI-Technologien bieten, damit zukünftig vertrauenswürdige, faire und positiv wahrgenommene Arbeitsumfelder geschaffen werden können, die den Einsatz neuer Technologien berücksichtigt.

1.2 Forschungsstand

Im aktuellen Forschungsstand zeigt sich eine Wissenslücke bezüglich der Einstellung und Wahrnehmung von Sozialarbeitenden grundsätzlich zu neuen Technologien und explizit zu KI-gestützten Entscheidungshilfen im Bereich des Kindesschutzes. Trotz des zunehmenden Einflusses neuartiger digitaler Technologien in vielen Berufsfeldern

ist bislang nur wenig darüber bekannt, wie Sozialarbeitende entsprechende Entwicklungen wahrnehmen und bewerten (Evans & Hilbert, 2020). Verschiedene Evaluationsberichte zu bereits im Kindesschutz eingesetzten Tools, wie z. B. das *Allegheny Family Screening Tool* (*AFST*) oder das *Predictive Risk Model (PRM)* in Neuseeland, thematisieren jedoch die Auswirkungen solcher Technologien auf die Arbeit von Sozialarbeitenden. Zudem wirft Virginia Eubanks in ihrem Bestseller «*Automating Inequality: How High-Tech Tools Profile, Police, and Punish the Poor*» einen kritischen Blick auf die Rolle von Sozialarbeitenden im Kontext der Nutzung der neuartigen Technologien (2018). Diese Thematik wird im dritten Kapitel dieser Arbeit näher untersucht und analysiert.

Verschiedene Studien in anderen Berufsfeldern haben sich allerdings konkret der Frage angenommen, wie sich KI-gestützte Tools auf das Vertrauen von Mitarbeitenden auswirken (Sassmannshausen & Heupel, 2020; Kröller & Schwarz, 2022; Lee, 2018). Diese Masterthesis basiert dabei auch auf der wegweisenden Studie von Min Kyung Lee aus dem Jahr 2018. Diese Studie fokussierte sich auf das Verständnis der Wahrnehmung von KI-gestützten Entscheidungen im Kontext des Human Resources Managements. Die Masterthesis orientiert sich am Forschungsdesign dieser Studie und legt es spezifisch auf die Soziale Arbeit und den Kindesschutz aus, um Forschungsdesiderate in diesem Bereich zu erschliessen.

1.3 Aufbau der Arbeit

Die vorliegende Arbeit ist in acht Kapitel unterteilt, die systematisch aufeinander aufbauen. Der Aufbau ist in zwei Teile gegliedert.

Im ersten, theoretischen Teil wird zunächst im ersten Kapitel die Relevanz und das Ziel der Untersuchung vorgestellt. Das zweite Kapitel legt die theoretischen und praktischen Grundlagen. Das dritte Kapitel behandelt zwei bestehende KI-Tools im Kindesschutz, die bereits heute in den USA und in Neuseeland angewandt werden. Im vierten Kapitel wird das Technologieakzeptanzmodell (*TAM*) vorgestellt und an die spezifischen Anforderungen der Studie angepasst. Es werden die entwickelten Hypothesen erläutert sowie die externen Faktoren und Eigenschaften der Vertrauenden beschrieben, die das Vertrauen in KI-gestützte Entscheidungshilfen beeinflussen könnten.

Im zweiten, empirischen Abschnitt wird das methodische Vorgehen im fünften Kapitel detailliert beschrieben. Es wird das Forschungsdesign erläutert, das auf einem multifaktoriellen, experimentellen Vignettendesign basiert. Im sechsten Kapitel werden die Ergebnisse der Studie präsentiert. Das siebte Kapitel fasst die zentralen Befunde der Studie zusammen und diskutiert die Einflüsse der variierten Dimensionen, der Personenmerkmale und der Integrationsbereitschaft auf das Vertrauen in KI-gestützte Entscheidungshilfen. Zudem werden weitere Erkenntnisse aus der Umfrage präsentiert und die Limitationen der Studie kritisch reflektiert. Im achten Kapitel werden die Forschungsfrage beantwortet und die Implikationen für die Soziale Arbeit diskutiert.

2 Grundlagen und Begriffsbestimmung

In diesem Kapitel werden die verschiedenen wissenschaftlichen Felder und Themenbereiche dargelegt, die das in dieser Arbeit behandelte Forschungsgebiet umfassen. Es beginnt mit einer kurzen Übersicht über den Kindesschutz in der Schweiz, gefolgt von der Charakterisierung von menschlichen Entscheidungsprozessen. Anschliessend werden die Grundlagen der Künstlichen Intelligenz vorgestellt. Schliesslich wird das Konzept des Vertrauens erörtert.

2.1 Übersicht Kindesschutz in der Schweiz

Das folgende Unterkapitel verschafft einen Überblick über den Aufbau des Kindesschutzes in der Schweiz und dient der Absteckung des Untersuchungsfeldes. Obwohl die Thesis im Kontext des Kindesschutzes und der Mandatsführung in der Schweiz verortet ist, konzentriert sie sich nicht primär auf die verschiedenen Aufgaben und Bereiche der Berufsbeistandspersonen, die als Zielgruppe der Untersuchung dienen. Diese werden im Kapitel 5 näher beschrieben. Stattdessen liegt der Fokus der Arbeit auf dem Vertrauen in KI-gestützte Entscheidungshilfen innerhalb dieses spezifischen Kontextes.

Der Kindesschutz in der Schweiz ist durch die zentrale Rolle der Kindes- und Erwachsenenschutzbehörde (KESB) charakterisiert, welche mit der Reform des Vormundschaftsrechts im Jahr 2013 eingeführt wurde. Diese Reform ersetzte die über 1'400 meist kommunalen Vormundschaftsbehörden durch knapp 150 interdisziplinär zusammengesetzten Fachbehörden, die schweizweit tätig sind. Die Organisation und die Struktur der KESB wird von den Kantonen festgelegt, wobei einige Kantone sich für ein gerichtsähnliches Modell entschieden haben, während die Mehrheit der Kantone administrative Behörden einsetzt (Heck, 2022, S. 95).

Die Hauptaufgabe der KESB besteht darin, den Sachverhalt unabhängig zu ermitteln und gegebenenfalls Schutzmassnahmen für Erwachsene und Kinder anzuordnen (Heck, 2022, S. 95). Die Klärung des Sachverhalts stellt die Abklärungstätigkeit der KESB dar (Heck, 2022, S. 98). Im Kindesschutz wird bei einer Kindeswohlabklärung die Lebenssituation eines Kindes unter Einbeziehung verschiedener Aspekte wie körperlicher, geistiger, seelischer und sozialer Dimensionen umfassend bewertet und anhand dieser Dimensionen überprüft (Biesel, Fellmann, Müller, Schär & Schnurr, 2017, S. 8). Die Handhabung der Abklärungstätigkeit variiert von Kanton zu Kanton: Während in einigen Kantonen die KESB selbst die Abklärungen durchführt, delegieren andere Kantone diese Aufgabe an externe Fachstellen (Biesel et al., 2017, S. 9; Rieder, Bieri, Schwenkel, Hertig & Amberg, 2016).

Wird aus der Abklärung eines Sachverhalts ersichtlich, dass eine Gefährdung der Person vorliegt, ist die KESB für die entsprechenden Schutzmassnahmen zuständig. Für eine Schutzmassnahme muss im Kindesschutz eine Gefährdung des Kindeswohls gemäss Art. 307 ff. ZGB (Schweizerisches Zivilgesetzbuch vom 10. Dezember 1907) vorliegen. Der Rechtsbegriff des Kindeswohls ist dabei absichtlich unbestimmt. Er dient als Grundprinzip dafür, wie das Recht angewandt wird, wie die Eltern ihre elterlichen Pflichten ausüben und wie das Vorgehen von Fachper-

sonen, Institutionen und Behörden in Bezug auf Kinder und Jugendliche ist. Das Kindeswohl muss in jedem einzelnen Fall spezifisch definiert werden – wenn die grundlegenden Bedürfnisse und Rechte von Kindern erfüllt und geschützt werden, wird davon ausgegangen, dass der Schutz des Kindeswohls gewährleistet ist (Heck, 2022, S. 97).

Für die Durchführung der Schutzmassnahmen werden häufig Mandatsträger:innen, wie Berufsbeistandspersonen oder geeignete Privatpersonen eingesetzt. Die Konferenz für Kindes- und Erwachsenenschutz (KOKES) empfiehlt für Berufsbeistandspersonen eine abgeschlossene Ausbildung auf Tertiärstufe in Sozialer Arbeit (2021, S. 20). Die Mandatsträger:innen übernehmen die Verantwortung für spezifische Aufgaben wie Beratung, Begleitung oder die Vertretung der betroffenen Personen. Sie sind für die jeweiligen Bereiche fachlich und persönlich geeignet und stehen unter der Aufsicht der KESB, welche die Einhaltung der rechtlichen Vorgaben sicherstellt (Heck, 2022, S. 98). Im Kindesschutz hat die Sicherstellung des Kindeswohls oberste Priorität für die Beistandsperson. Dies bedeutet, dass Massnahmen ergriffen werden müssen, die eine Gefährdung des Kindeswohls vermeiden und dessen langfristigen Erhalt sicherstellen. Die Auswahl der geeigneten Massnahmen und Methoden im Rahmen der Mandatsführung liegt grundsätzlich im Ermessen der Beistandsperson (Heck, 2022, S. 99).

Sowohl die KESB als auch die Beistandspersonen haben sich vom Subsidiaritäts- und Verhältnismässigkeitsprinzip sowie vom Grundsatz der Stufenfolge der Massnahmen leiten zu lassen. Es ist zwar ein konsequentes Eingreifen zum Wohl der betroffenen Kinder erforderlich, jedoch sind Überreaktionen und Übertreibungen zu vermeiden (Fassbind, 2022, S. 112). Es gilt der Vorrang «freiwilliger» Hilfe, sodass nicht jede Kindeswohlgefährdung kindesschutzrechtlich relevant ist, wenn die Eltern sich auf kooperative Prozesse einlassen, die das Kindeswohl absehbar genügend gewährleisten können (Fassbind, 2022, S. 113).

2.2 Menschliche Entscheidungsfindung und kognitive Verzerrung

In der Entscheidungstheorie umfasst der Begriff «Entscheidung» die Auswahl einer Handlungsoption aus mehreren möglichen Alternativen, unabhängig davon, wie bewusst diese Wahl getroffen wird (Laux, Gillenkirch & Schenk-Mathes, 2018, S. 3). Menschliche Entscheide können dabei nicht ausschliesslich als analytische Prozesse verstanden werden, sondern enthalten immer auch intuitive Komponenten (Nitzsch, 2021, S. 5). In der Wissenschaft bilden Dual-Prozess-Theorien das theoretische Fundament für die Abgrenzung zwischen intuitivem und analytischem Entscheidungsverhalten. Innerhalb dieser Theorien werden zwei verschiedene Systeme unterschieden: System 1, das zuständig für heuristisch-intuitive Entscheidungen (Bauchentscheide) ist, und System 2, welches verantwortlich für den kontrolliert-analytischen Entscheidungsfindungsprozess (Kopfentscheide) ist (Enosh & Bayer-Topilsky, 2015, S. 3; Nitzsch, 2021, S. 6). Während es sich bei System 1 um einfache, unbewusste und automatische Entscheidungen handelt, die schnell getroffen werden und wenig kognitive Ressourcen in Anspruch nehmen, handelt es sich bei System 2 um bewusste, langsame Entscheidungen, die kognitive Ressourcen beanspruchen und analytisch abgewogen werden (Nitsch, 2021, S. 6). Die Aufgabenverteilung bzw. das Zusammenspiel zwischen den beiden Systemen lässt sich dabei nicht klar abgrenzen bzw. wird ersichtlich, dass sie gegenseitig voneinander abhängig sind (Scheidgen, 2019, S. 50).

Kognitive Verzerrung innerhalb von Entscheidungsfindungsprozessen

Menschen sind beim Treffen von Entscheidungen nachweislich anfällig für Fehler (Glaberson, 2019, S. 318). In Entscheidungsfindungsprozessen werden Situationen nicht nur auf Basis objektiver Fakten (System 2) bewertet, sondern sind sie immer auch durch die persönlichen Erfahrungen und Interpretationen geformt (System 1). Hierbei besteht immer auch die Gefahr der kognitiven Verzerrung (*cognitive bias*), bei welcher es unbewusst zu fehlerhaften, irrationalen Urteilen kommen kann (Bastian & Schrödter 2014, S. 278). Kahnemann, Slovic & Tversky (1982) fassen unter dem Begriff kognitive Verzerrung verschiedene Verzerrungen (*bias*) zusammen, welche die Entscheidungsfindung massgeblich beeinflussen können. Das Einschätzen einer Wahrscheinlichkeit, mit welcher ein Ereignis eintritt, ist für Menschen schwierig und sie tendieren dazu, ihre eigenen Fähigkeiten bei der Vorhersage zukünftiger Ereignisse zu überschätzen (Glaberson, 2019, S. 318). Im Folgenden werden zur Veranschaulichung drei häufige Verzerrungen[1] näher erläutert. Dabei werden drei fiktive Beispiele aus dem Kindesschutz herangezogen.

***Der Bestätigungsfehler* (*confirmation bias*):**

Die Neigung, Beweise, die eine bereits bestehende Meinung stützen, leicht anzunehmen und gegenteilige Beweise abzulehnen (Glaberson, 2019, S. 318).

Beispiel Bestätigungsfehler im Kindesschutz:

Eine Sozialarbeiterin ist überzeugt, dass ein Vater aufgrund seiner Vorgeschichte von Gewalt gegenüber seinen Kindern nicht in der Lage ist, sich positiv zu verändern. Selbst wenn der Vater durch Therapie und Unterstützung Fortschritte zeigt und positive Veränderungen in seinem Verhalten aufweist, neigt die Sozialarbeiterin dazu, diese Fortschritte zu übersehen oder abzuwerten, da sie ihrer vorgefassten Meinung widersprechen.

Rahmungseffekte (framing effects):

Entscheidungen von Menschen variieren, abhängig davon, wie Optionen präsentiert werden – die Entscheidung verändert sich basierend auf den verfügbaren Informationen (Glaberson, 2019, S. 318).

Beispiel Rahmungseffekte im Kindesschutz:

Ein Sozialarbeiter informiert eine Familie über die Möglichkeit einer sozialpädagogischen Familienbegleitung (SPF). Er beschreibt die SPF einmal als «Chance, die Erziehungskompetenzen zu stärken und gemeinsam Lösungen zu finden» und ein anderes Mal als «notwendige Intervention, um schwerwiegende familiäre Probleme zu lösen». Je nach Formulierung könnte die Familie die SPF entweder als positive Unterstützung oder als belastende Pflicht empfinden.

Die Verfügbarkeitsheuristik (availability bias):

Die Neigung, die Wichtigkeit oder Wahrscheinlichkeit eines Ereignisses zu überschätzen, wenn es «auffällig und markant» erscheint (Glaberson, 2019, S. 318).

Beispiel Verfügbarkeitsheuristik im Kindesschutz:

Ein Sozialarbeiter, der in den letzten Monaten mehrere schwere Fälle von Kindesmissbrauch bearbeitet hat, könnte dazu neigen, die Wahrscheinlichkeit von Missbrauch in einer neuen Familie, die er betreut, zu überschätzen. Dies könnte dazu führen, dass er vorschnell Massnahmen ergreift oder die familiäre Situation strenger bewertet, obwohl die tatsächliche Risikoabschätzung eine differenziertere Herangehensweise erfordert.

[1] An dieser Stelle wird auf den populärwissenschaftlichen Bestseller «*Thinking, Fast and Slow*» von Kahnemann (2011) verwiesen, in welchem er die Erkenntnisse seiner jahrzehntelangen Forschung zu kognitiven Verzerrungen zusammenfasst.

In der Sozialen Arbeit werden Entscheidungsprozesse als zentraler Bestandteil professionellen Handelns gesehen und haben in der Diskussion um die Professionalisierung und die Professionalität eine hohe Relevanz (Bastian & Schrödter, 2014, S. 276). Sie sind aufgrund ihres Charakters stets in berufliche Handlungsstrukturen eingebunden (Trippelmandat) und stehen im Spannungsfeld zwischen Individuum, Profession und Politik. Die daraus resultierende Komplexität erfordert eine hohe Kompetenz der Sozialarbeitenden (Scheidgen, 2019, S. 45), da Entscheide häufig unter hoher Falllast, unter Zeitmangel, mit nicht genügend klarer Informationslage und mit unklarem Ausgang getroffen werden müssen (Bastian & Schrödter, 2014, S. 276; Glaberson, 2019). Die Arbeit im Kindesschutz erfordert ein breites Verständnis für die psychosozialen Dynamiken innerhalb von Familien sowie für rechtliche Rahmenbedingungen. Beistandspersonen stehen vor der Herausforderung, vertrauensvolle Beziehungen zu Eltern und Kinder aufzubauen und dabei weittragende Entscheidungen zu treffen, die das Leben der Kinder, Jugendlichen und Familien beeinflussen können (Biesel et al., 2017, S. 144).

Im Bereich der Kinder- und Jugendhilfe sind Entscheidungsprozesse von Sozialarbeitenden komplex und können durch verschiedene kognitive Verzerrungen beeinflusst sein. Sozialarbeitende sind mit der Aufgabe betraut, Entscheidungen zu treffen, die weitreichende Auswirkungen auf das Leben der betroffenen Familien und Kinder haben können. Auch sie sind – wie alle Menschen – voreingenommen und besitzen nur bedingt die Fähigkeit, Informationen unvoreingenommen abzuwägen und zu werten, um Ereignisse genau vorherzusagen (Glaberson, 2019, S. 320). Häufig werden die Auswirkungen ihrer Entscheidungen erst Jahre oder Jahrzehnte später erkennbar, wenn das begleitete Kind bereits erwachsen ist. Gleichzeitig machen es die Effekte anderer Entscheidungen, welche im Kindesalter von verschiedenen Akteur:innen getroffen werden, praktisch unmöglich, die tatsächliche Wirkung einer einzelnen Entscheidung zu bestimmen. Es ist in der Regel auch nicht möglich, bestimmte Entscheidungen in Hinblick auf eine hypothetische Alternative zu bewerten, z. B. die Folgen, wenn ein Kind nicht ausserfamiliär untergebracht wird, sondern bei seinen Eltern bleiben würde (Glaberson, 2019, S. 320). Ungewissheit bleibt ein ständiger und unvermeidlicher Bestandteil im Kindesschutz (Enosh & Bayer-Topilsky, 2015, S. 3).

In Situationen, in denen die Regeln und Kriterien klar definiert sind – zum Beispiel bei klaren Beweisen für körperliche Misshandlung durch ein Elternteil – setzen Sozialarbeitende hauptsächlich analytische Methoden ein, um die Kriterien auf den spezifischen Fall anzuwenden (Enosh & Bayer-Topilsky, 2015, S. 3). In Fällen höherer Unsicherheit, wenn unklar ist, welche Kriterien zutreffen, neigen Sozialarbeitende jedoch dazu, sich auf heuristische Methoden zu stützen, um zu einer Entscheidung zu gelangen. Folglich sind Entscheidungen in Situationen grosser Unsicherheit anfälliger für persönliche Heuristiken und Verzerrungen (Enosh & Bayer-Topilsky, 2015, S. 3).

Verschiedene Studien (Rossi, Schuermann & Budde, 1999; Baird & Wagner, 2000; White & Walsh, 2006; Keddell, 2014; Gautschi, 2021) kommen zum Schluss, dass Sozialarbeitende bei Situationen mit identischen Parametern nicht immer zu gleichen Einschätzungen gelangen und Fehler in Entscheidungsfindungsprozessen präsent sind. Sozialarbeitende neigen beispielsweise dazu, ihre Entscheidungen auf begrenzte Datenmengen zu stützen, indem sie sich auf markantere Ereignisse konzentrierten und weniger auffällige Informationen möglicherweise ausser Acht lassen (White & Walsh, 2006, S. 4; Munro, 2008). Dies kann zu Fehlern im Entscheidungsfindungsprozess und

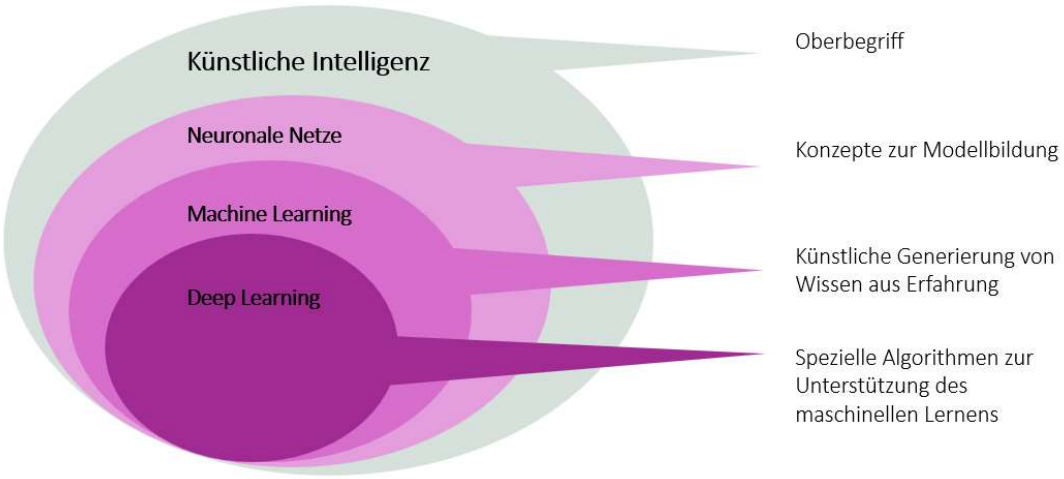
ungewollter Ungleichbehandlung führen, da frühe Urteile oft nur langsam revidiert werden und neuen, möglicherweise widersprüchlichen Informationen skeptisch begegnet wird. Ebenfalls zeigen Studien, dass Informationen aus unterschiedlichen Quellen unterschiedlich bewertet werden: Sozialarbeitende gewichteten beispielsweise Aussagen von Ärztinnen oder Ärzten stärker als diejenigen von Nachbar:innen, unabhängig von der Qualität der Information (Munro, 1999).

Angesichts der Funktionsweise des menschlichen Gehirns tendieren Sozialarbeitende dazu, «mentale Abkürzungen» zu schaffen, welche die schwierigen Beurteilungsaufgaben einfacher gestalten, was dazu führen kann, dass bei der Bewertung, Meinungsbildung und Entscheidungsfindung weniger Informationen berücksichtigt werden. Ohne aktive Reflexion über diese Prozesse können solche Vereinfachungen zu Fehlern in der Entscheidungsfindung führen (Nitzsch, 2021, S. 22).

2.3 Künstliche Intelligenz (KI)

Das folgende Unterkapitel führt in den Oberbegriff «Künstliche Intelligenz» (KI) ein. Der Terminus umfasst Konzepte wie neuronale Netzwerke, maschinelles Lernen, und Deep Learning und dient als Sammelbezeichnung für diese interdisziplinären Ansätze (*Abbildung 1*) (Kreutzer, 2023, S. 10). Ziel der nachfolgenden Abschnitte ist es, den Lesenden ein grundlegendes Verständnis für das breite und vielfältige Feld der KI zu vermitteln, indem die wesentlichen Termini definiert und einführend erläutert werden[2].

Abbildung 1 – Leistungsbestandteile der Künstlichen Intelligenz
Quelle: Eigene Darstellung (Kreutzer, 2023, S. 10)



[2] Aufgrund der Komplexität der Thematik wird auf eine detaillierte Diskussion der einzelnen KI-Teilbereiche verzichtet. Ebenso wird in den nachfolgenden Kapiteln zur Vereinfachung der Terminologie konsequent der Begriff «KI» verwendet, obwohl die Literatur im Kontext des Kindesschutzes auch von «algorithmenbasierten Tools» (Keddell, 2019; Gillingham & Ackermann, 2020) oder von «Machine-Learning-Tools» (Vaithianathan, Venavides-Prado, Dalton, Chouldechova, Putnam-Hornstein, 2021) spricht.

Das Ziel von KI besteht darin, Maschinen zu entwickeln, die in der Lage sind, Tätigkeiten auszuführen, die üblicherweise menschliche kognitive Fähigkeiten voraussetzen. Schon im Jahr 1956, welches als Geburtsstunde der KI gesehen wird, wurde im Dartmouth-Proposal die Annahme festgehalten, dass es grundsätzlich möglich sein sollte, jeden Aspekt menschlichen Lernens oder jegliche Form von Intelligenz so präzise zu beschreiben, dass eine Maschine sie imitieren könnte (Haehnlein & Kaplan, 2019, S. 7). Das Europäische Parlament (2023) definiert KI heute mit den folgenden Worten: «Künstliche Intelligenz ist die Fähigkeit einer Maschine, menschliche Fähigkeiten wie logisches Denken, Lernen, Planen und Kreativität zu imitieren.». Die Technologie der KI erstreckt sich über verschiedene Felder und Techniken, von der Erkennung von Sprache und Bildern bis hin zur Entscheidungsfindung und Problembehandlung (Kreutzer, 2023, S. 4).

Dabei bietet sie Fähigkeiten wie Wahrnehmung, Argumentation, autonomes Lernen und die Entwicklung von Lösungsstrategien. Es gibt fünf verschiedene Bewertungsansätze, die entweder kombiniert oder einzeln angewendet werden können (Kreutzer, 2023, S. 9):

- Description: die Beschreibung aktueller Zustände (Beschreibung des «Ist»)
- Inspection: die Untersuchung der Gründe dahinter (Analyse des «Warum»)
- Prediction: die Vorhersage zukünftiger Entwicklungen (Vorehrsage des «Wird»)
- Prescription: die Ausarbeitung von Handlungsempfehlungen (Empfehlung des «Was»)
- Generation: die Generierung neuer Ideen oder Produkte (Schaffung von «Neuem»)

In dieser Thesis wird der Fokus insbesondere auf den Ansatz der *prediction* – zu Deutsch «Vorhersage» – gelegt. *Prediction* bezeichnet den Prozess, bei dem statistische Techniken und Datenanalyse-Tools eingesetzt werden, um zukünftige Risiken oder Ereignisse vorherzusagen. Diese Modelle analysieren historische und aktuelle Daten, um Muster zu erkennen, die Rückschlüsse auf zukünftige Wahrscheinlichkeiten zulassen. Ziel ist es, Entscheidungsträger:innen in verschiedenen Bereichen wie Finanzen, Gesundheits-/Sozialwesen oder Sicherheitsmanagement präzise Vorhersagen zu liefern, um proaktive Massnahmen zu ermöglichen. Durch die Implementierung prädiktiver Risikomodelle können Organisationen potenzielle Risiken minimieren und strategische Entscheidungen auf der Grundlage fundierter Daten treffen (Kreutzer, 2023, S. 217).

Im Folgenden wird die Funktionsweise von KI und ihren Subgruppen erläutert.

2.3.1 Algorithmen

Die Grundprinzipien der KI basieren auf der Nachahmung menschlicher Intelligenzprozesse durch die Entwicklung von Algorithmen, die lernen und Entscheidungen basierend auf Daten treffen können. Obwohl die Frage danach, wie Algorithmen definiert werden, in der Informatikliteratur als «philosophische Frage» gehandelt wird, kann sich die Wissenschaft darauf einigen, dass es sich bei Algorithmen im Grundsatz um «Verfahren zur Lösung von Problemen» handelt (Ottmann & Widmayer, 2012, S. 1). Algorithmen sind rechnerische Verfahren, die Eingabewerte (*Input*) aufnehmen und Ausgabewerte (*Output*) erzeugen (Corman, Leiserson, Rivest & Stein, 2009, S. 5). Sie zeich-

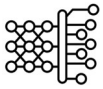
nen sich durch spezifische Merkmale aus: Ihre Formulierung erfolgt in mathematischer Sprache, sie sind zahlenbasiert, ihre Ausführung ist endlich (sie führen nach einer bestimmten Anzahl von Schritten zu einem Ergebnis) und sie müssen in ihrer Struktur klar und exakt definiert sein, sodass die jeweiligen Schritte eindeutig festgelegt sind. Algorithmen verarbeiten also bekannte Daten, erkennen Muster und treffen auf dieser Basis Entscheidungen (Knuth, 1997, S. 2).

2.3.2 Neuronale Netzwerke

Im Gegensatz zur Maschine nimmt der wache menschliche Geist ständig und gleichzeitig sensorische Informationen aus der Umwelt auf, sei es durch Sehen, Hören, Riechen, Tasten oder Schmecken. Der menschliche Geist ist bis zum jetzigen Zeitpunkt ein unübertroffenes Lernsystem; bisher kann kein Computer eine ausgeruhte und motivierte Person im Erkennen von Bildern, Klängen, Gerüchen usw. durchgängig übertreffen (Braga-Neto, 2020, S. 2). Zentral für die Entwicklung von KI-Systemen ist dennoch ein ähnliches Konzept: künstliche neuronale Netze, die versuchen, die komplexe Struktur und Funktionsweise des menschlichen Gehirns maschinell nachzubilden (Braga-Neto, 2020, S. 2). Diese Netzwerke bestehen aus miteinander verbundenen Einheiten oder «Neuronen», die Daten verarbeiten können, um Muster zu erkennen, Schlussfolgerungen zu ziehen und Entscheidungen zu treffen. Diese Lernprozesse ahmen die Art und Weise nach, wie menschliche Gehirne Informationen verarbeiten, indem sie Verbindungen zwischen Neuronen stärken oder schwächen, basierend auf den Erfahrungen, die sie sammeln (Kreutzer, 2023, S. 11). In der Praxis bedeutet dies, dass ein künstliches neuronales Netz seine Gewichtungen zwischen den Neuronen anpassen kann, um bestimmte *Outputs* basierend auf *Inputs* zu produzieren, was dem Netzwerk ermöglicht, komplexe Funktionen zu modellieren und nicht-lineare Beziehungen innerhalb der Daten zu erkennen. Dieser Prozess der Modellbildung und des Lernens ist entscheidend für die Fähigkeit der KI, sich an neue Daten anzupassen und Aufgaben zu erfüllen, für die sie nicht explizit programmiert wurde (Kreutzer, 2023, S. 11-12).

Die bestehenden KI-Entscheidungshilfen, welche aktuell im Kindesschutz erprobt und angewendet werden, basieren auf neuronalen Netzwerken: Sie enthalten nur diejenigen Informationen, mit welchen sie trainiert wurden. Gleichzeitig setzen sie maschinelles Lernen ein, um die Genauigkeit und Konsistenz von Entscheidungen zu verbessern und subjektive Verzerrungen durch Sozialarbeitende zu verringern (Gillingham, 2019, S. 115; vgl. Kapitel 3).

2.3.3 Machine Learning und Deep Learning



In neuerer Zeit hat sich KI vor allem in Richtung Machine Learning (ML) (dt. maschinelles Lernen) entwickelt, welches als Schlüsseltechnologie des Zeitalters gilt (Brynjolfsson & McAfee, 2017). ML ist ein Kernbestandteil von KI, der sich darauf konzentriert, Algorithmen zu entwickeln, die Computern die Fähigkeit geben, aus Daten zu lernen und zu verbessern, ohne für spezifische Aufgaben programmiert zu sein (Kreutzer, 2023, S. 12). ML-Techniken ermöglichen es Systemen, Muster und Strukturen in den Daten zu erkennen und diese Erkenntnisse zur Vorhersage oder Entscheidungsfindung zu nutzen. Der Schwerpunkt liegt auf der Mustererkennung und der Gewinnung

von Erkenntnissen durch die Analyse riesiger Datensätze (Krauss, 2023, S. 121). Diese Prozesse basieren auf der Annahme, dass Maschinen durch die Analyse von Daten lernen und ihre Leistung über die Zeit selbstständig verbessern können. Ein Schlüsselelement des ML ist die Fähigkeit der Algorithmen, sich selbst zu korrigieren und zu optimieren, basierend auf dem Feedback, das sie durch die Interaktion mit Daten erhalten. Dies ermöglicht eine kontinuierliche Anpassung und Verbesserung der Entscheidungsfindungsprozesse der Maschine (Kreuzter, 2023, S. 13).

Als Teilbereich des maschinellen Lernens konzentriert sich Deep Learning auf komplexe künstliche neuronale Netzwerke, die vielschichtig miteinander verbunden sind. Die Tiefe eines neuronalen Netzwerkes nimmt mit der Anzahl Schichten zu, wodurch das Deep Learning umfangreiche Datenmengen analysieren, aus ihnen lernen und dadurch komplexe Muster und abstrakte Zusammenhänge identifizieren kann (Krauss, 2023, S. 121). Die Besonderheit des Deep Learnings liegt in seiner Fähigkeit, aus rohen Daten zu lernen, indem es zunehmend komplexere Merkmale auf verschiedenen Ebenen extrahiert. Dies erlaubt Deep-Learning-Modellen, hervorragende Ergebnisse in der Bild- und Spracherkennung, natürlicher Sprachverarbeitung und vielen anderen Bereichen zu erzielen (Kreutzer, 2023, S. 13). Der entscheidende Vorteil von Deep-Learning-Algorithmen ist ihre Fähigkeit, mit zunehmender Datenmenge kontinuierlich zu lernen und ihre Genauigkeit zu verbessern, was sie zu einem wirkungsvollen Werkzeug für die Analyse grosser und komplexer Datensätze macht. Gleichzeitig stehen Forschende vor der Herausforderung, dass die Funktionsweise von Deep Learning oft unklar bleibt, bzw. nicht bekannt ist, aufgrund welcher Informationen ein Modell eine spezifische Entscheidung trifft oder wie bestimmte Muster erfasst wurden (Krauss, 2023, S. 122)[3].

Ein berühmtes Beispiel für die Anwendung von Deep Learning ist ChatGPT. ChatGPT funktioniert, indem es riesige Mengen an Textdaten aus dem Internet analysiert und Muster in der Sprache erkennt. Das Modell wurde darauf trainiert, Texte zu vervollständigen, Fragen zu beantworten und menschenähnliche Konversationen zu führen[4].

2.3.4 Die Grenzen der KI

Trotz den zahlreichen Anwendungsbereichen der KI birgt das das Feld der KI nicht nur Chancen, sondern zugleich auch Risiken. Erhält eine KI nicht exakte Anweisungen, wie sie ein Problem zu lösen hat oder fehlt ihr der nötige Kontext, welche Lösung ein Mensch bevorzugen würde, entwickelt sie unter Umständen ihre eigene Problemlösungsstrategie, anstatt den Programmier-Anweisungen zu folgen (Shane, 2021, S. 138). Damit eine KI selbständig die richtige Lösung finden kann, muss die Zieldefinition möglichst genau beschrieben und auf nützliche Antworten

[3] Auf eine vertiefte, technische Auseinandersetzung mit der Funktionsweise von Machine und Deep Learning wird an dieser Stelle verzichtet. Für weiterführende Informationen und tiefere Einblicke empfiehlt die Autorin die Lektüre von Kreutzer (2023), Krauss (2023) und Braga-Neto (2020).

[4] Laut einer Studie von Chen, Zaharia und Zou (2023) hat sich die Leistung von ChatGPT jedoch bereits jetzt verschlechtert. Die Forschenden der Universitäten Stanford und Berkeley fanden heraus, dass das Modell GPT-4 von März bis Juni 2023 bei Aufgaben wie Mathematik und Codierung schlechter abgeschnitten hat als das Vorgängermodell. Dies wirft Fragen zur langfristigen Entwicklung und Optimierung von KI-Systemen auf.

beschränkt sein. Problematisch wird dies auch bei ethischen Fragestellungen – hier fehlt der KI schlichtweg das Verständnis für ihre eigenen Aufgaben sowie für ethische oder biologische Aspekte: Es mangelt der KI an «gesundem Menschenverstand» (Shane, 2021, S. 139).

Mit der zunehmenden Integration von KI in verschiedene Lebensbereiche steigen damit auch die ethischen Herausforderungen, die sich mit dem Einsatz der Technologie verbunden sind. Die Kernfrage, die sich stellt, ist, wie sichergestellt werden kann, dass die Entwicklung und Nutzung von KI-Systemen im Einklang mit moralischen Werten und Prinzipien steht (Kreutzer, 2023, S. 68). Eine weitere ethische Frage stellt sich bei der Fairness und Voreingenommenheit. KI-Systeme lernen aus grossen Datensätzen. Sind diese (menschgemachten) Daten voreingenommen, können die von der KI getroffenen Entscheidungen diese Vorurteile widerspiegeln und verstärken. Dies kann zu Diskriminierung und Ungerechtigkeit führen (Kreutzer, 2023; Eubanks, 2018; Kleinberg, Lakkaraju, Leskovec, Ludwig & Mullainathan, 2017). Entsprechend stehen sowohl Entwicklung als auch Unternehmen vor der Herausforderung, Mechanismen zu implementieren, die sicherstellen, dass ihre KI-Systeme fair und unparteiisch sind und nicht zu Verfestigung bestehender Ungleichheiten beitragen (Kreutzer, 2023, S. 73).

Die Wahrung der Privatsphäre und des Datenschutzes ist eine weitere ethische Herausforderung im Zusammenhang mit KI. Viele KI-Anwendungen erfordern den Zugriff auf grosse Mengen persönlicher, teilweise sehr sensibler Daten, um effektiv zu funktionieren. Dies wirft Fragen bezüglich des Schutzes dieser Daten vor Missbrauch und unbefugtem Zugriff auf (Kreutzer, 2023, S. 505). Entscheidend scheint, dass Datenschutzstandards eingehalten werden und Nutzende von KI die Kontrolle über persönliche Informationen behalten[5].

Eine der zentralen ethischen Herausforderungen bei der Entwicklung und Anwendung von KI ist die Frage der Autonomie und Kontrolle: Mit dem Aufkommen fortgeschrittener KI-Systeme, die in der Lage sind, selbständige Entscheidungen zu treffen, entsteht gleichzeitig auch die Notwendigkeit, Grenzen dafür festzulegen, wie viel Autonomie den Systemen gewährt werden soll (Kreutzer, 2023, S. 68). Diese Diskussion ist eng verknüpft mit der Sorge um die Sicherheit und Zuverlässigkeit von KI-Systemen, insbesondere wenn ihre Entscheidungen direkte Auswirkungen auf das menschliche Leben haben, wie im Falle autonomer Fahrzeuge, medizinischer Diagnosewerkzeuge – oder dem Kindesschutz. Schliesslich stellt sich auch die Frage nach der Verantwortung und Haftung. Wer trägt die Verantwortung, wenn ein KI-System eine fehlerhafte oder eine schädliche Entscheidung trifft? Je mehr Akteur:innen (Entwicklungsteam, User:innen und Plattformen) involviert sind, umso komplexer gestaltet sich die Klärung der Verantwortlichkeiten (Kreutzer, 2023, S. 69).

Dem ethischen Aspekt beim Einsatz von KI kommt deshalb zunehmend eine grössere Bedeutung zu, denn sie wirft verschiedene soziale, gesellschaftliche und rechtliche Fragen auf, auf welche bislang noch keine abschliessenden Antworten gefunden wurden (Wess, 2019, S. 157). Eine Fokussierung auf menschenzentrierte und verantwortungsvolle KI, welche die Bedeutung ethischer Entscheidungen, den Schutz der Privatsphäre, die Vermeidung von

[5] Verschiedene Verordnungen (z. B. in der EU die «DS-GVO» oder in den USA der «AI Act») legen bereits heute Richtlinien für eine Regulierung der KI fest. Die Schweiz evaluiert bis Ende 2024 verschiedene Ansätze für die Regulierung, das seit dem 1. September 2023 geltende Datenschutzgesetz (DSG) ist aber direkt auf KI-gestützte Datenbearbeitung anwendbar (EDÖB, 2023).

Voreingenommenheit und Transparenz betont, scheint unumgänglich (Kreutzer, 2023, S. 137). Dieser Ansatz, der eine enge Verknüpfung mit der digitalen Ethik aufweist, ist entscheidend für den verantwortungsvollen Umgang mit personenbezogenen Daten und die Sicherstellung, dass KI-Systeme im besten Interesse der Menschen handeln. Die Implementierung einer menschenzentrierten KI könnte zwar einige Jahre in Anspruch nehmen, ist jedoch zentral, um das Vertrauen in die Technologie zu stärken und langfristig eine ethisch vertretbare Anwendung von KI zu gewährleisten (Kreutzer, 2023, S. 138).

2.4 Vertrauen

In der modernen Entscheidungsfindung spielt künstliche Intelligenz eine immer bedeutendere Rolle, indem sie komplexe Daten analysiert und Muster erkennt, die für menschliche Entscheidungsträger:innen nicht immer unmittelbar ersichtlich sind. Dieser technologische Fortschritt führt zu einer neuen Dynamik und hebt die Bedeutung des Vertrauens in technologische Systeme hervor. In der vorliegenden Thesis wird Vertrauen als abhängige Variabel untersucht, da es als entscheidende Komponente für eine erfolgreiche Implementierung neuer Systeme erachtet wird. Das folgende Unterkapitel geht näher auf das Konstrukt ein.

Luhmann (2009) definiert Vertrauen als «das Verlassen auf eigene Annahmen» (S. 1). Es dient dabei als Schutzschild gegen Gefühle der Überforderung, da es die Multidimensionalität und insbesondere die Unberechenbarkeit der zukünftigen Welt vereinfacht (Luhmann, 2009, S. 24). In einer komplexen Welt, in der es unmöglich ist, alle Komponenten und Entwicklungen zu berücksichtigen, erlaubt Vertrauen den Menschen, Unsicherheiten zu überwinden und Entscheidungen zu treffen, ohne alle möglichen Konsequenzen oder die Verlässlichkeit anderer vollständig zu kennen. Das Hauptinstrument zur Verringerung dieser Komplexität ist die Erwartung von Zuverlässigkeit in ungewissen oder risikoreichen Situationen. Die Erwartung von Sicherheit erleichtert und ermöglicht Handlungen in einem komplexen Umfeld (Kasselbaum, 2004, S. 12). Ein zentrales Risiko dabei besteht in der Möglichkeit, dass diese Erwartungen nicht erfüllt werden. Luhmann sieht Vertrauen zwar als riskanten, aber notwendigen Schritt zur Vereinfachung des Lebens (Luhmann, 2009, S. 27).

Vertrauen und Misstrauen sind tief miteinander verwurzelt. Während Vertrauen ein Gefühl der Sicherheit und Zuversicht in die Integrität und Fähigkeiten einer Person oder eines Systems darstellt, ist Misstrauen das Gegenteil davon. Es geht über das blosse Fehlen von Vertrauen hinaus und basiert auf negativen Annahmen (Luhmann, 2009, S. 92). Diese beiden Konzepte neigen dazu, sich in einem zyklischen Muster zu verstärken (Sydow, 1995, S. 185). Offenheit und Informationsaustausch können das Vertrauen erhöhen, während das Zurückhalten von Informationen Misstrauen fördert und den Wunsch nach Kontrolle verstärkt (Lee, 2004, S. 633).

Vertrauen entwickelt sich in verschiedenen Phasen und wird durch technologische, organisatorische, kontextuelle, soziale und persönliche Faktoren beeinflusst. Diese Faktoren umfassen die Qualität und Sicherheit der Anwendung, die Reputation einer Organisation, kulturelle Normen sowie persönliche Erfahrungen und Eigenschaften. Vertrauen führt dabei zu kognitiven, affektiven und verhaltensbezogenen Veränderungen, beeinflusst die Wahrnehmungen, Einstellungen und Verhaltensweisen der Menschen und wird stark durch kulturelle und soziale Normen

geprägt. Individuelle Unterschiede wie die generelle Vertrauensneigung, persönliche Erfahrungen und der Umgang mit Unsicherheit und Risiko spielen ebenfalls eine wesentliche Rolle bei der Entwicklung und Aufrechterhaltung von Vertrauen (Yang & Wibowo, 2022, S. 2068).

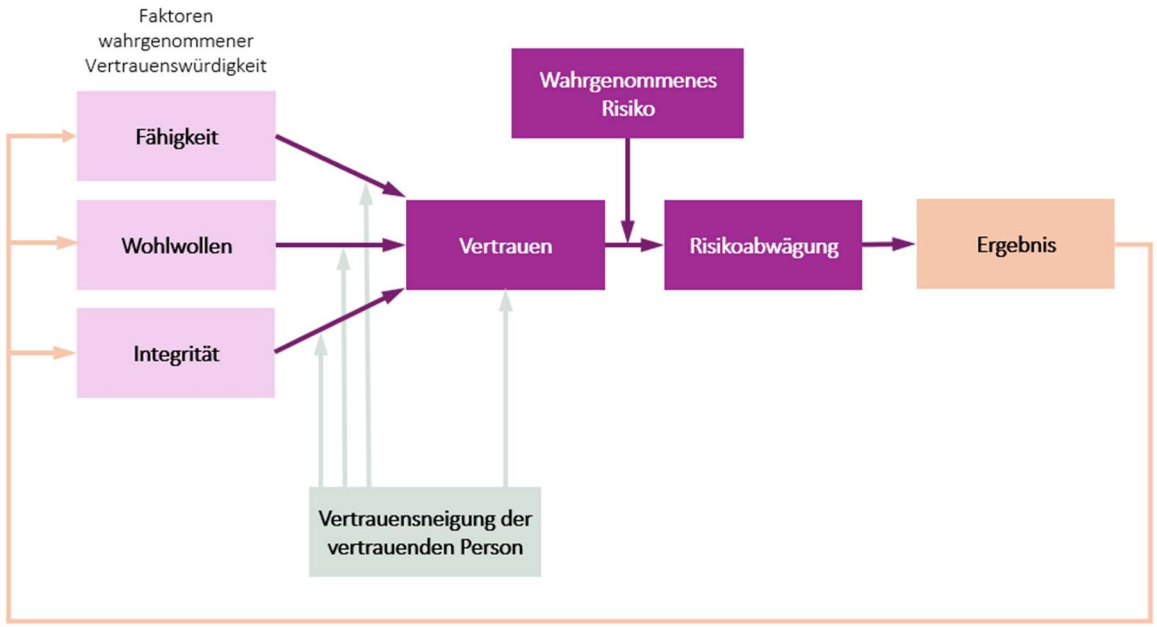
Das Konzept des Vertrauens wurde in verschiedenen Bereichen wie den Sozialwissenschaften, in der Psychologie, in der Philosophie, im Management und in der Wirtschaft untersucht (Rheu, Shin, Peng & Huh-Yoo, 2021). In Bezug auf das Vertrauen in digitale Technologien sollten jedoch weitere Perspektiven miteinbezogen werden (Yang & Wibowo, 2022, S. 2056).

In den Sozialwissenschaften wird das Konzept des Vertrauens in KI oft in drei Stufen unterteilt: Vorhersehbarkeit (*predictability*), Verlässlichkeit (d*ependability*) und Glauben (*faith*) (Rheu et al., 2021). Die Vorhersehbarkeit bezieht sich auf die Konsistenz des Verhaltens oder der Leistung einer Anwendung über einen längeren Zeitraum hinweg. Erbringt eine KI-gestützte Anwendung über verschiedene Situationen hinweg konsistente Leistungen, ermöglicht dies den User:innen präzisere, verlässliche Vorhersagen über ihr zukünftiges Handeln (Nordheim, Følstad & Bjørkli, 2019). Je höher die Vorhersehbarkeit, desto geringer die Wahrscheinlichkeit unbeabsichtigter Folgen und damit negativer Auswirkungen (Yang & Wibowo, 2022, S. 2056). Vorhersehbarkeit steht in enger Verbindung mit der Verständlichkeit: User:innen, die verstehen, wie die KI funktioniert, können sich eher auf die Resultate der KI einlassen (Sheridan, 2019). Nur wenn die Leistung einer KI-Anwendung als vorhersehbar und verlässlich angesehen wird, werden die User:innen ein Vertrauen und eine Bereitschaft zur Nutzung der Technologie entwickeln (Rheu et al., 2021; Sheridan, 2019). Eine weitere Herangehensweise zur Definition von Vertrauen besteht darin, das Konzept des Vertrauens mit dem integrativen Modell von Mayer, Davis und Schoorman (1995) (*Abbildung 2*) anhand der Faktoren Fähigkeit (*ability*), Wohlwollen (*benevolence*) und Integrität (*integrity*) zu bewerten (Yang & Wibowo, 2022, S. 2056).

In Bezug auf KI bezieht sich Fähigkeit auf die Fertigkeit einer KI-Anwendung, eine von den User:innen zugewiesene Aufgabe zu erfüllen (Rheu et al., 2021) sowie darauf, ob sie korrekt entscheidet und die wichtigsten Aufgaben richtig ausführen kann (Sheridan, 2019). Wohlwollen misst den Grad der Übereinstimmung der Absicht der KI mit der der User:innen (Rheu et al., 2021). Integrität schliesslich betrifft die Frage, ob das Verhalten der KI den moralischen Grundsätzen entspricht, die von den User:innen akzeptiert werden. Somit geht die Vertrauensbildung über die Vorhersehbarkeit einer KI-Anwendung hinaus und erfordert, dass die Absicht der KI mit derjenigen der User:innen übereinstimmt (Rheu et al., 2021).

Zusätzlich zu diesen drei Kernaspekten des Vertrauensmodells sind auch die Konstrukte des *wahrgenommenen Risikos* und der *Risikoabwägung* entscheidend. Vertrauen beinhaltet die Einschätzung des Risikos einer Handlung gegenüber dem potenziellen Gewinn aus dieser Handlung. Das wahrgenommene Risiko wird beurteilt im Verhältnis zu den erwarteten Vorteilen des Vertrauens. Übersteigt das Vertrauensniveau das wahrgenommene Risiko, wird die vertrauende Person eher bereit sein, ein Risiko einzugehen. Ist das wahrgenommene Risiko jedoch grösser als das Vertrauensniveau, wird die vertrauende Person sich zurückhalten (Mayer et al., 1995, S. 726).

Abbildung 2 – Integratives Vertrauensmodell
Quelle: Eigene Darstellung (Mayer, Davis und Schoorman, 1995)



Das Ergebnis dieses Prozesses, ob positiv oder negativ, prägt die zukünftigen Vertrauensentscheidungen und wird als Erfahrungswissen in das Gedächtnis der vertrauenden Person integriert. Die Bewertung einer Situation erfolgt entsprechend nicht nur auf Grundlage der aktuellen Gegebenheiten, sondern wird auch von früheren Erfahrungen beeinflusst, die Erwartungen an zukünftige Interaktionen formen (Mayer et al., 1995, S. 726).

3 Künstliche Intelligenz im Kindesschutz

Die Idee der Verwendung von Algorithmen zur Unterstützung der Entscheidungsfindung in der Sozialen Arbeit ist nicht neu. Bereits 1985 wurde das Potenzial von datengestützten «Expertensystemen» untersucht, die bei der Entwicklung professioneller Entscheidungsfindung im Kindesschutz helfen sollten (Gillingham, 2019, S. 115). Heute existieren in der Sozialen Arbeit vor allem im Rahmen des Kindesschutzes unterschiedliche Verfahren, die dabei unterstützen, das Risiko von Kindeswohlgefährdung einzuschätzen, wobei diese hauptsächlich im angelsächsischen Raum entwickelt und angewandt werden. Werden Daten, beispielsweise Risikoeinschätzungen (*Predictive Risk Modelling*), gemäss bestimmten festgelegten Regeln gewichtet und miteinander kombiniert, wird von mechanischer, statistischer oder aktuarialistischer Urteilsbildung gesprochen. Diese Verfahren stützen sich auf empirische Studien im Bereich des Kindesschutzes, bei welchen statistische Indikatoren (bzw. Prädiktoren) identifiziert wurden, die eine zuverlässige Prognose zukünftiger Kindeswohlverletzungen ermöglichen (Bastian & Schrödter, 2019, S. 43). Verschiedene Studien zeigen, dass statistische Verfahren in der Regel treffsicherer sind, als wenn Sozialarbeitende ihre Entscheidungen im Rahmen von diskursiven oder interpretativen Verfahren fällen (Bastian & Schrödter, 2019, S. 46; Keddell, 2019, S. 3). Die Integration von KI könnte entsprechend potenziell dazu beitragen, dass Sozialarbeitende in Zukunft konsistentere, weniger verzerrte Entscheidungen in ihrer Fallbearbeitung treffen.

Das folgende Kapitel behandelt den praktischen Einsatz von zwei verschiedenen KI-Tools im Kindesschutz, wie sie bereits heute im US-Bundesstaat Pennsylvania und in Neuseeland angewandt werden. Es wird aufgezeigt, wie die Funktionsweise der beiden Tools ist und welche Erfolge sowie Herausforderungen sich dabei bis heute ergeben haben. Der Fokus liegt darauf, wie KI-gestützte Systeme dazu beitragen können, die Entscheidungsfindung von Sozialarbeitenden im Kindesschutz zu unterstützen.

3.1 Bestehende KI-Entscheidungshilfen im Kindesschutz

In den vergangenen Jahrzehnten haben innerhalb der Sozialen Arbeit und insbesondere innerhalb der Kinder- und Familienhilfe verschiedene Versuche stattgefunden, KI-gestützte Entscheidungsfindungssysteme (engl. *Decision Support Systems*) in der Praxis zu erforschen und zu erproben. Die Tools sind mit einer riesigen Datenmenge bestehender Falldaten trainiert und nutzen maschinelles Lernen mit dem Ziel, Genauigkeit und Konsistenz von Entscheidungen zu verbessern und subjektive Verzerrungen von Sozialarbeitenden zu reduzieren (Gillingham, 2019, S. 115).

Im Folgenden werden die beiden Tools *Allegheny Family Screening Tool (AFST)* aus dem US-Bundesstaat Pennsylvania und das *Predictive Risk Modelling (PRM)* aus Neuseeland vorgestellt. Beide Tools sind zum jetzigen Zeitpunkt aktiv, werden aber kontrovers diskutiert und laufend evaluiert.

3.1.1 Allegheny Family Screening Tool (*AFST*)

2016 führte der örtliche Sozialdienst des Allegheny County im US-Bundesstaat Pennsylvania mit dem *AFST* ein prädiktives, KI-gestütztes Risikomodell ein. Ziel war es, ein Tool zu implementieren, das effiziente und konsistente, datengestützte Entscheidungen basierend auf bestehenden Informationen trifft und dabei sicherstellt, dass die öffentlichen Ressourcen gerecht auf die am meisten gefährdeten Familien des Countys verteilt werden. Gleichzeitig sollte die allgemeine Gesundheit, Sicherheit und das Wohlbefinden der Bevölkerung gefördert werden (Chouldechova, Putnam-Hornstein, Benavides-Prado, Fialko & Vaithianathan, 2018, S. 3).

Das Tool steht dem Personal des örtlichen Sozialdienstes zur Verfügung, welches Anrufe im Rahmen von vermuteten Kindeswohlgefährdungen entgegennimmt. Diese Anrufe können sowohl durch Fachpersonen (z. B. Lehrpersonen) als auch durch Privatpersonen (z. B. Familienmitglieder, Nachbar:innen) eingehen. Die Aufgabe der Mitarbeitenden ist es, anhand der Anrufe zu entscheiden, ob einem Verdacht auf Kindeswohlgefährdung weiter nachgegangen wird oder ob eine Abklärung abgewiesen wird (Vaithianathan, Benavides-Prado, Dalton, Chouldechova & Putnam-Hornstein, 2021, S. 54). Diese Entscheidung wird häufig getroffen, ohne dass die Mitarbeitenden die Familie je gesehen haben oder sie zu ihrer Situation befragt wurden (Chouldechova et al., 2018, S. 2).

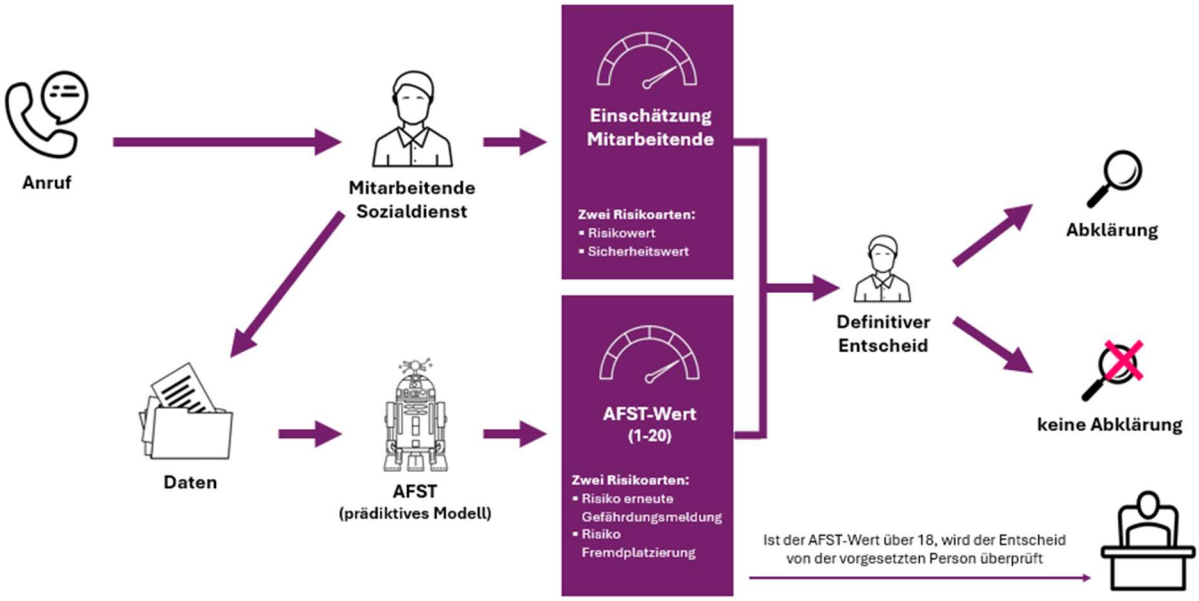
Die Funktionsweise des *AFST* ist in *Abbildung 3* dargestellt. Die Mitarbeitenden erfassen die im Anruf erhaltenen Personeninformationen in ihrem Standard-System, worauf das *AFST* automatisch und unabhängig von der Einschätzung der Mitarbeitenden einen Wert zwischen 1 und 20 generiert. Der Wert gibt die errechnete Wahrscheinlichkeit aus, mit welcher in den nächsten zwei Jahren eine erneute Gefährdungsmeldung eingeht und mit welcher eine ausserfamiliäre Platzierung des Kindes stattfindet (Vaithianathan et al., 2021, S. 54). Dabei werden Informationen über das betroffene Kind und dessen Familienmitglieder aus verschiedenen Datenbanken aus dem Allegheny County miteinbezogen: So werden beispielsweise Informationen aus vorangegangenen Kontakten mit der Familie innerhalb der Kindesschutzdienste, der Sozialhilfe, den County-Gefängnissen, der Jugendbewährung oder aus Gesundheitseinrichtungen verwertet (Chouldechova et al., 2018, S. 5).

Bevor die Mitarbeitenden den durch das Tool errechneten Wert sehen, werden sie gebeten, zwei Bewertungen zur Situation der Familie einzugeben. Diese Bewertungen widerspiegeln ihre persönliche Einschätzung zum Risiko und zur Sicherheit der Kinder innerhalb der Familie. Der Risikowert (*risk score*) zeigt an, wie die Mitarbeitenden das Risiko einschätzen (gering, moderat oder hohes Risiko), während der Sicherheitswert (*safety score*) angibt, inwiefern aus Sicht der Mitarbeitenden eine Sicherheitsbedrohung vorliegt (keine, drohende oder gegenwärtige Gefahr). Diese beiden Werte werden im Anschluss miteinander multipliziert, wobei ein Risiko-Sicherheitswert (*risk-safety-score*) entsteht (Chouldechova et al., 2018, S. 11-12).

Erst im Anschluss sehen die Mitarbeitenden den durch das *AFST* errechneten Wert – sie treffen ihre Entscheidung zum vorliegenden Fall also in jedem Fall selbständig und ohne Unterstützung durch das Tool. Die Mitarbeitenden erhalten mit dem *AFST* ergänzend zu ihrer persönlichen, «menschlichen» Einschätzung (*human decision*), eine zweite, KI-gestützte Einschätzung, welche sie anschliessend in die Schlussbetrachtung miteinfliessen lassen kön-

nen. Errechnet das *AFST* allerdings einen Wert zwischen 18 und 20, bedeutet das – unabhängig von der Einschätzung der Mitarbeitenden – dass eine automatische Kindeswohlabklärung ausgelöst wird, sofern die vorgesetzte Person sich nicht aktiv dazu entscheidet, die Empfehlung des *AFST* zu überstimmen und eine andere Massnahme zu ergreifen (Chouldechova et al., 2018, S. 11). Entscheiden sich die Mitarbeitenden dazu, eine Abklärung einzuleiten, wird der *AFST*- Wert nicht an die abklärende Stelle weitergegeben (Chouldechova et al., 2018, S. 13).

Abbildung 3 – Funktionsweise *Allegheny Family Screening Tool* (*AFST*)
Quelle: Eigene Darstellung (Choulderchova et al., 2018, S. 11)



Seit der Einführung des Tools 2016 wurde das *AFST* mehrfach evaluiert (Vaithianathan, Putnam-Hornstein, Jiang, Nand & Maloney, 2017; Goldhaber-Fieber & Prince, 2019). Eine unabhängige Studie durch Forschende der Standford University 2019 ergab, dass die Implementierung des Systems die Genauigkeit von Entscheidungen bei Abklärungen verbesserte und die Unterschiede in der «Fallöffnung» zwischen Familien unterschiedlicher ethnischer Herkunft verringert hatte. Insbesondere bei älteren Kindern zeigte sich diese Verbesserung. Die Ergebnisse der Evaluation weisen darauf hin, dass das Tool die Identifizierung von Kindern, die weitere Interventionen benötigten, verbesserte, ohne die Arbeitsbelastung für die Mitarbeitenden des Sozialdienstes zu erhöhen. Zudem leistete es einen erheblichen Beitrag zur Verringerung von Ungleichheiten im System (Goldhaber-Fiebert & Prince, 2019, S. 19-25).

In der praktischen Anwendung des *AFST* zeigt sich, dass das Tool subtile Veränderungen in der Arbeitsweise der Mitarbeitenden des Sozialdienstes hervorruft. Die Einschätzung des *AFST* wird den Mitarbeitenden, wie oben beschrieben, erst angezeigt, nachdem sie bereits umfangreiche Recherchen zum erhaltenen Anruf durchgeführt haben. Die Ausgabe des *AFST* -Werts führt dazu, dass die Mitarbeitenden ihre eigenen vorläufigen Einschätzungen kritisch hinterfragen, wenn diese nicht mit dem berechneten Wert übereinstimmen. Dieses Vorgehen kann einerseits die Genauigkeit im Entscheidungsprozess verbessern, setzt andererseits aber auch die Mitarbeitenden unter Druck, ihre eigenen Einschätzungen den maschinell ermittelten Wahrscheinlichkeiten anzupassen (Eubanks, 2018, S. 141-142).

3.1.2 Predictive Risk Modelling (*PRM*)

Das *Predictive Risk Modelling* (*PRM*) ist ein KI-gestütztes Entscheidungsfindungssystem, welches von Entwickler:innen der Universität Auckland in Neuseeland entwickelt und 2012 implementiert wurde (Gillingham, 2019, S. 117). Ziel des Modells ist es, Kinder zu identifizieren, die ein hohes Risiko für eine Kindeswohlgefährdung haben. In der praktischen Anwendung wird das *PRM* genutzt, um Ressourcen effektiv zu nutzen und präventive Massnahmen gezielt einzusetzen. Sozialarbeitende können die Informationen aus dem *PRM* nutzen, um gefährdete Kinder frühzeitig zu identifizieren und Unterstützungsangebote oder Interventionen anzubieten, bevor es zu einer tatsächlichen Kindeswohlgefährdung kommt. Ebenfalls kann das *PRM* dabei helfen, bestehende Kindesschutzfälle zu überwachen und bei Veränderungen in der Risikobewertung entsprechend zu reagieren (Dare, 2013, S. 5).

Der Kern des Systems ist ein automatisierter Algorithmus, der eine anonymisierte Datenbank nutzt. Diese verknüpft Daten aus den neuseeländischen Sozial- und Kindesschutzsystemen von Kindern, die zwischen 2003 und 2006 geboren wurden und vor ihrem zweiten Lebensjahr Sozialleistungen bezogen. Die Daten umfassen eine Stichprobe von 57'986 Kindern, was ca. 33% aller in diesem Zeitraum in Neuseeland geborenen Kindern entspricht (Dare, 2013, S. 11). Das *PRM* identifiziert und validiert verschiedene Variablen, die mit Kindeswohlgefährdung in Verbindung stehen. Diese werden analysiert, um jene zu identifizieren, die eine statistisch signifikante Vorhersagekraft für das Risiko einer Kindeswohlgefährdung aufweisen. Insgesamt wurden bei der Entwicklung 132 Variablen, wie beispielsweise demografische oder sozioökonomische Merkmale der Kinder und ihrer Lebenswelt als statistisch bedeutsam für das Modell erachtet und in den Kernalgorithmus integriert (Ministry of Social Development, 2014, S. 7).

Tabelle 1 zeigt zur Veranschaulichung einen Auszug der 132 Variablen, die zur Bewertung des Risikos einer Kindeswohlgefährdung im *PRM* herangezogen wurden. Dazu gehören beispielsweise, ob andere Kinder in der Familie bereits Schutzmassnahmen erfahren haben, die Häufigkeit häuslicher Gewalt in der Familie oder das Alter der Erziehungsberechtigten. Weitere Variablen erfassen die biologische Elternschaft, den Familienstand (insbesondere ob alleinerziehend) und die Dauer des Bezugs von Sozialleistungen in den letzten fünf Jahren. Ebenfalls relevant sind, ob die Erziehungsberechtigten selbst Schutzmassnahmen erfahren haben, die Anzahl der Adressänderungen im letzten Jahr sowie psychische Probleme und Substanzmissbrauch in den letzten fünf Jahren. Im Weiteren wird erfasst, ob die Erziehungsberechtigten als Kind Verhaltens- oder Beziehungsprobleme hatten (Ministry of Social Development, 2014, S. 7). Diese und weitere Variablen bilden die Grundlage für die Entwicklung des *PRM* zur Bewertung des Risikos einer Kindeswohlgefährdung.

Tabelle 1 – Original-Auszug aus den Variablen des *Predictive Risk Modelling* (*PRM*)

Other children with care and protection history	Yes No
Family violence	Events in one month Events in more than one month No events (no Police FV notifications or contact records)
Caregiver's age	Under 20 20 to 25 25 to 30 30 to 35 (includes missing) 35 to 40 Over 40
Benefit caregiver is not a birth registration parent	Yes No birth registration No
Single parent	Single parent Single parent and no father listed on birth registration Not single parent or partnership status unknown
Time on benefit in the last 5 years	More than 80% 20 < - 80% Up to 20% No time
Caregiver with care and protection history	Yes No
Benefit address changes in the last year	No address changes 1-2 address changes 3 plus address changes Missing (no benefit in last year)
Mental health in the last 5 years	Substance abuse issues Persistent substance abuse issues (3+ years in last 5) Mental health issues other than substance abuse Persistent mental health issues other than substance abuse (3+ years in last 5) No known mental health or substance abuse issues
Behavioural or relationship difficulties as a child	Yes No

Quelle: Eigene Darstellung (Ministry of Social Development, 2014, S. 7)

Nachdem die Forschenden die relevanten Variablen ausgewählt hatten, wurde der Algorithmus mit dem anonymen Datensatz trainiert. Dabei wurde die verfügbare Datenmenge in zwei Gruppen geteilt: eine Trainingsgruppe (70% der Daten) und eine Validierungsgruppe (die restlichen 30%). Der Algorithmus wurde zuerst mit der Trainingsgruppe entwickelt und anschliessend mit der Validierungsgruppe getestet, um die Genauigkeit und Zuverlässigkeit zu überprüfen. Diese Phase war entscheidend, um sicherzustellen, dass das Modell präzise Vorhersagen liefert (Dare, 2013, S. 10).

Die Ergebnisse des Modells werden in Risikowerten ausgedrückt, die jedem Kind bei Beginn einer Sozialleistungsperiode zugewiesen werden. Diese Risikowerte werden in Dezile eingeteilt, wobei ein Wert von 10 das höchste Risiko und ein Wert von 1 das niedrigste Risiko darstellt. Hat das Kind in der nachfolgenden Periode einen höheren *PRM*-Wert als zuvor, ist das Risiko gestiegen. Kinder, welche in die höchsten Risikodezile fallen, können von den Sozialarbeitenden für weitere Abklärungen und mögliche Interventionen priorisiert werden (Dare, 2013, S. 11). Um die Genauigkeit des Algorithmus zu bewerten, wurden die Vorhersagen mit den tatsächlichen Ereignissen bei den Kindern im Testdatensatz verglichen. Bei der Untersuchung der Genauigkeit des Modells zeigte sich, dass bei

Kindern in der höchsten Risikokategorie (10) 48% bis zum fünften Lebensjahr Anzeichen einer Kindeswohlgefährdung zeigten, verglichen mit nur 2% der Kinder in der niedrigsten Risikokategorie. Die Genauigkeit des Modells (prädiktive Validität) wurde auf ca. 76% geschätzt[6] (Dare, 2013, S. 11).

Die Forschenden des Projekts (Vaithianathan, Maloney, Jiang, De Haan, Dale, Putnam-Hornstein & Dare) fanden 2012 heraus, dass 5.4% aller Kinder in Neuseeland bis zum Alter von fünf Jahren Misshandlungen erlebt haben. Da 83% dieser Kinder vor dem zweiten Lebensjahr in der Stichprobe erfasst wurden, erhielten sie eine Risikobewertung durch den Algorithmus. Das Modell zeigte, dass präventive Massnahmen für alle Kinder, die eine Risikobewertung erhielten, eine enorme Herausforderung wären, da viele dieser Interventionen vermutlich gar nicht nötig wären, insbesondere weil die meisten dieser Kinder in den niedrigen Risikodezilen erschienen (Vaithianathan et al., 2012, S. 12). Der Bericht schlug daher vor, sich auf intensive Interventionen für die beiden höchsten Dezilgruppen zu konzentrieren – also Kinder mit Risikowerten von 9 und 10. Diese Gruppen machen ca. 5% der Gesamtbevölkerung aus, beinhalten aber 37% der nationalen Misshandlungsfälle. Es wird angenommen, dass durch solche gezielten Massnahmen bis zu 280 Fälle von Misshandlung bis zum Alter von fünf Jahren innerhalb von zwei Jahren verhindert werden könnten, wenn 50% der Familien das darauffolgende Unterstützungsangebot annehmen würden (Vaithianathan et al., 2012, S. 20).

Aus den Details, die im Mai 2015 über die Tests und die Entwicklung des *PRM* veröffentlicht wurden (Ministry of Social Development, 2014), wird deutlich, dass in der Zwischenzeit umfangreiche Arbeiten zur weiteren Entwicklung und Erprobung des *PRM* durchgeführt wurden (Gillingham, 2016, S. 4). Das *PRM* wurde an drei Gruppen von Kindern (Jahrgänge 2007, 2008 und 2010) getestet und es wurden verschiedene Versionen des *PRM* entwickelt, bei welchen auch Daten aus dem Gesundheitswesen und dem Strafvollzug integriert wurden, um die Überrepräsentation verschiedener Gruppen in den Datensätzen zu korrigieren (Gillingham, 2016, S. 4). Beispielsweise wurden in einem Modell Kinder, deren Eltern keine öffentliche Unterstützungsleistung erhalten hatten, ebenfalls einbezogen und ein separates Modell wurde für Maori-Kinder erstellt, um deren Überrepräsentation im Kindesschutzsystem zu korrigieren. Insgesamt wurden zwölf verschiedene Versionen des Algorithmus getestet, die unterschiedliche Datenkombinationen nutzten. Zusätzlich schlugen Expert:innen weitere Entwicklungen vor und stimmten mit dem Machbarkeitsbericht des Ministry of Social Development (2014) überein, dass es «noch immer zu viele ungenaue Vorhersagen gibt, als dass aktuelle *PRM*-Modell als alleiniger Mechanismus zur Identifizierung von Kinder mit hohem Risiko verwendet werden könnte» (TCC Group, 2013, S. 4). Es wurde auch die Bedeutung betont, dass sowohl Kindesschutzfachpersonen als auch deren Vorgesetzte verstehen, wie das *PRM* entwickelt wurde, wobei der Schwerpunkt darauf lag, die Schlüsselvariablen im Modell zu zeigen, und nicht nur deren Punktezahl (TCC Group, 2013, S. 6).

[6] Die Leistung von Vorhersagemodellen wird oft durch den Prozentsatz der Fläche unter der *Receiver Operator Characteristic* (ROC) Kurve zusammengefasst. Ein Modell mit einer 100% Abdeckung unter der ROC-Kurve gilt als perfekt. Der Hauptalgorithmus, der beim PRM bei Kindern unter 2 Jahren angewendet wurde, hat mit einer Fläche unter der ROC-Kurve von 76% eine «moderate» bis «gute» Stärke (prädiktive Validität) bei der Vorhersage von Misshandlungen bis zum Alter von 5 Jahren (Gillingham, 2016, S. 4).

3.2 Herausforderungen und positive Entwicklungen der KI-Entscheidungshilfen

Sowohl das *AFST* als auch das *PRM* ermöglichen es, frühzeitig in das Leben der als gefährdet eingestuften Kinder einzugreifen, indem sie zielgerichtete Interventionen vorschlagen, um Kindeswohlgefährdungen zu verhindern. Die KI-gestützten Technologien bieten Vorteile, indem sie helfen, potenzielle Risikofälle schneller und präziser zu identifizieren. Dennoch gibt es verschiedene Aspekte, die kritisch betrachtet werden müssen. Die folgenden Punkte diskutieren die ethischen Bedenken, die Herausforderungen hinsichtlich der Genauigkeit der Vorhersagen, die Fragen zur Autorität und Objektivität der Tools, den Datenschutz und die Transparenz sowie die positiven Entwicklungen und die Bedeutung der Zusammenarbeit von Sozialarbeitenden und Entwickler:innen.

3.2.1 Herausforderungen

Trotz vielversprechender Potenziale der KI-gestützten Entscheidungshilfen im Kindesschutz gibt es Kritikpunkte und Herausforderungen, die bei ihrer Implementierung und Nutzung berücksichtigt werden müssen. Diese betreffen insbesondere ethische Fragen, die Genauigkeit der Vorhersagen und die Auswirkung auf das professionelle Urteil der Sozialarbeitenden. Im Folgenden werden die wichtigsten Kritikpunkte erläutert:

- Ethische Fragen und Genauigkeit der Vorhersagen

  Die Modelle werfen ethische Fragen auf, besonders hinsichtlich der Genauigkeit der Vorhersagen und der möglichen Stigmatisierung der Familien. Kritikpunkte betreffen die Gefahr von Fehlidentifikationen, bei welchen Kinder fälschlicherweise als gefährdet eingestuft werden (*false-positives*) (Gillingham, 2016, S. 5). Die Entwickler:innen des *PRM* betonen, dass jede Intervention präventiv und unterstützend sein und die betroffenen Familien freiwillig daran teilnehmen sollen (Dare, 2013, S. 4). Dennoch besteht die Sorge, dass die Abhängigkeit von solchen Modellen zu einem mechanischen Ansatz der Sozialen Arbeit führen könnte, der die Individualität der Einzelfälle nicht ausreichend berücksichtigt (Keddell, 2014, S. 920).

- Herausforderungen und Treffsicherheit (prädiktive Validität)

  Ein weiterer Hauptkritikpunkt ist die Genauigkeit der Vorhersagen der Tools. Viele bisherigen Systeme erreichen keine ausreichende Treffsicherheit. Das *PRM* in Neuseeland war bei der Evaluation nur zu 76% genau – ein Viertel der Vorhersagen war somit nicht korrekt. In einem Bereich, in dem Entscheidungen das Wohl von Kindern und Familien direkt beeinflussen, können solche Ungenauigkeiten schwerwiegende Folgen haben (Gillingham, 2019, S. 117; Eubanks, 2018, S. 138).

- Autorität und Objektivität

  Die Implementierung des *AFST* wirft Fragen zur Autorität und Objektivität KI-gestützter Entscheidungshilfen auf. Mitarbeitende können ihre Risiko- und Sicherheitsbewertung nicht ändern, nachdem der *AFST*-Wert ausgegeben wurde – ausser auf Anordnung einer vorgesetzten Person. Dies illustriert das Spannungsfeld zwischen menschlicher Beurteilung und algorithmischer Vorhersage. Diese Vorgehensweise soll die Konsistenz im Entscheidungsprozess fördern, könnte aber das professionelle Urteilsvermögen der

Mitarbeitenden untergraben und zu einer risikoscheuen Haltung führen, bei der eigene Einschätzungen zugunsten der algorithmischen Werte vernachlässigt werden (Eubanks, 2018, S. 142-143; Glaberson, 2019, S. 355-356).

- Datenherkunft und -gewinnung

  Die Algorithmen basieren auf Daten, die bestehende Probleme im Kindesschutzsystem widerspiegeln. Diese Daten resultieren aus zahlreichen menschlichen Entscheidungen, die Werturteile über das Kindesschutzsystem implizieren und anfällig für Ungenauigkeiten, Bias, Vorurteile und Inkonsistenzen sind. Falldaten, sowohl strukturiert als auch unstrukturiert, können zudem unvollständig sein, da überlastete Fachpersonen gezwungen sind, schnell zum nächsten Fall zu wechseln. Selbst detaillierte Erklärungen in den Fallnotizen können die Feinheiten und Nuancen der Entscheidungsfindung nicht immer vollständig wiedergeben. Informationen im Kindesschutzsystem müssen entsprechend kategorisiert werden, was zu Vereinfachungen und z. T. irreführenden Darstellungen führen kann (Glaberson, 2019, S. 337). Ohne sorgfältige Aufmerksamkeit bei der Entwicklungs- und Nutzungsphase könnten KI-gestützte Tools nicht nur bestehende Probleme überdecken, sondern auch neue Risiken für Familien einführen (Glaberson, 2019, S. 355). Werden Algorithmen zusätzlich regelmässig mit neuen Daten aktualisiert, besteht das Risiko einer Stichprobenverzerrung (*sample selection bias*), welche die Genauigkeit des Tools untergraben könnte (Glaberson, 2019, S. 336).

- Datenschutz und Transparenz

  Ethische Bedenken betreffen auch den Datenschutz, die Zustimmung zur Datennutzung und die mögliche Stigmatisierung durch algorithmische Klassifizierung. Zudem wird die Transparenz und Nachvollziehbarkeit der Entscheidungsprozesse hinterfragt, da die «Black-Box-Natur» des maschinellen Lernens es schwer macht, die Entscheidungsgrundlagen der Tools zu verstehen (Gillingham, 2019, S. 117).

3.2.2 Positive Entwicklungen

Die Implementierung KI-gestützter Entscheidungshilfen im Kindesschutz bietet aber auch zahlreiche positive Aspekte, die ihre Integration und Nutzung für die betroffenen Jugendhilfestellen wertvoll machen können. Die folgenden Punkte beleuchten die Potenziale von Entscheidungshilfen:

- Gezielte Frühinterventionen

  Predictive Risk Models wie jenes in Neuseeland oder in Allegheny haben das Potenzial, Sozialarbeitende bei der Identifizierung von Risikofällen zu unterstützen. Durch die Nutzung grosser Datenmengen und darauf trainierte Algorithmen können entsprechende Modelle helfen, potenzielle Fälle von Kindesmisshandlungen frühzeitig zu erkennen und präventive Massnahmen gezielter und effizienter einzusetzen. Die Modelle ermöglichen es, frühzeitig in das Leben gefährdeter Kinder einzugreifen, indem sie Risikofaktoren identifizieren, die auf zukünftige Gefährdungen hinweisen. Dies ermöglicht gezielte Interventionen, die dazu beitragen können, Kindeswohlgefährdungen zu verhindern und langfristige negative Auswirkungen auf die Kinder zu minimieren (Gillingham, 2016, S. 118).

- Erhöhte Handlungssicherheit

  Die Algorithmen bieten eine datengetriebene Basis für Entscheidungsfindungen, was zu einer erhöhten Objektivität und Konsistenz in der Risikobewertung führen kann. Dies kann helfen, menschliche Vorurteile zu reduzieren und sicherzustellen, dass Entscheidungen auf soliden Daten und nicht auf subjektiven, möglicherweise verzerrten Einschätzungen basieren (Gillingham, 2016, S. 118). Die Verwendung von KI-Tools bedeutet entsprechend keine «De-Professionalisierung», sondern trägt zu erhöhter Handlungssicherheit und einer schärferen Fokussierung auf den professionellen Ermessensspielraum bei. Die strukturierte Erfassung und Klassifizierung von Daten dient als zusätzliches Reflexionsinstrument und ermöglicht Sozialarbeitenden eine fundiertere Basis für ihre Entscheidungen (Schneider, 2021, S. 131).

- Verbesserung der Effizienz

  Durch die Automatisierung und Systematisierung der Datenanalyse können entsprechende Tools die Effizienz in der Fallarbeit erheblich verbessern. Sozialarbeitende können ihre Zeit und Ressourcen auf andere Aspekte der Fallarbeit konzentrieren, was die Gesamtqualität der Unterstützung erhöht (Gillingham, 2016, S. 118).

- Personalisierte Unterstützung

  Die Nutzung von Daten zur Entwicklung massgeschneiderter Unterstützungsprogramme kann dazu beitragen, spezifische Bedürfnisse und Risikofaktoren einzelner Klient:innen gezielt anzugehen. Dies kann die Effektivität der Intervention selbst erhöhen und bessere Ergebnisse für die betroffenen Kinder, Jugendlichen und Familien erzielen (Gillingham, 2019, S. 5, Glaberson, 2019, S. 343). Durch die Identifikation von Risikofaktoren und gefährdeten Gruppen können präventive Massnahmen entwickelt und implementiert werden, die darauf abzielen, die Bedingungen zu verbessern, die zu Gefährdungen führen. So können sowohl direkte Interventionen als auch breitere soziale Programme zur Unterstützung gefährdeter Familien gezielt angegangen werden (Glaberson, 2019, S. 357).

Insgesamt bieten KI-gestützte Entscheidungshilfen wie das *AFST* oder das *PRM* verschiedene Vorteile, die dazu beitragen können, die Qualität im Kindesschutz zu verbessern. Transparenz und Vertrauen spielen dabei eine wichtige Rolle: Durch die Entwicklung transparenter Modelle und die Einbindung von Feedbackschlaufen können Sozialarbeitende ein besseres Verständnis und Vertrauen in die Technologie entwickeln, was die Akzeptanz und Wirksamkeit der Modelle in der Praxis erhöht (Eubanks, 2018, S. 124). Die Förderung der Zusammenarbeit zwischen Sozialarbeitenden und Datenwissenschaftler:innen ist dabei entscheidend, um sicherzustellen, dass Algorithmen korrekt interpretiert und angewendet werden. Diese Zusammenarbeit verbessert kontinuierlich die Modelle und stellt sicher, dass die tatsächlichen Bedürfnisse der Klient:innen widerspiegeln (Glaberson, 2019). Der Erfolg dieser Systeme hängt von einer engen Zusammenarbeit zwischen Sozialarbeitenden, Datenwissenschafter:innen und betroffenen Familien ab. Langfristige Verpflichtungen, ethische Rahmenbedingungen und eine transparente Entwicklung und Implementierung sind entscheidend, um die Potenziale von KI-gestützten Entscheidungshilfen in der Sozialen Arbeit auszuschöpfen und gleichzeitig ihre Grenzen und Risiken zu berücksichtigen (Gillingham, 2019, S. 124).

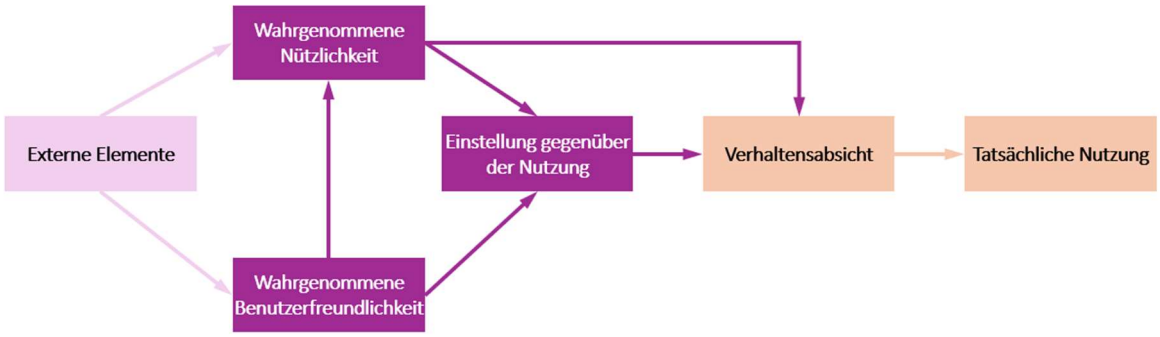
4 Theoretische und konzeptionelle Zugänge

Das folgende Kapitel konzentriert sich auf das Technologieakzeptanzmodell (*TAM*), welches ursprünglich 1985 von Davis konzipiert wurde, und erläutert die Modifizierungen, die speziell für diese Masterarbeit vorgenommen wurden. In einem weiteren Schritt werden die Hypothesen aufgestellt und dargelegt.

4.1 Das Theorie Akzeptanzmodell (*TAM*)

Das *TAM* (*Abbildung 4)* dient der Untersuchung der Akzeptanz computergestützter Informationssysteme in Unternehmen und wurde ursprünglich zur Messung von Akzeptanz von Mitarbeitenden gegenüber der damals neuartigen Technologie «E-Mail» entwickelt. Es basiert auf der *Theory of Reasoned Action* (*TRA*) von Fishbein und Ajzen (1975), die vertritt, dass unsere Handlungen von unseren Absichten bestimmt werden, die wiederum von unserer Einstellung und dem, was andere denken, beeinflusst sind. Es stützt sich ebenfalls auf die erweiterten Erkenntnisse von Ajzen (1985) im Rahmen der *Theory of Planned Behavior* (*TPB*), die betont, dass Faktoren wie die Aussicht auf Erfolg und die Kontrolle interner und externer Faktoren unser Verhalten beeinflussen (Davis, 1985, S. 15-23).

Abbildung 4 – Technologieakzeptanzmodell (*TAM*)
Quelle: Eigene Darstellung (Davis et al., 1989, S. 985)



Das *TAM* postuliert, dass zwei zentrale Faktoren – die wahrgenommene Nützlichkeit (*perceived usefulness*) und die wahrgenommene Benutzerfreundlichkeit (*perceived ease of use*) – die Einstellung eines Individuums zur Nutzung eines Systems und damit dessen Nutzungsabsicht bestimmen. Die wahrgenommene Nützlichkeit wird dabei als das Ausmass definiert, «in dem eine Person glaubt, dass die Nutzung des Systems ihre Arbeitsleistung verbessern würde», die wahrgenommene Benutzerfreundlichkeit bezieht sich auf das «Ausmass, in dem eine Person glaubt, dass die Verwendung eines bestimmten Systems keine Anstrengung erfordert» (Davis et al., 1989, S. 985). Diese Faktoren werden ihrerseits von externen Elementen wie z. B. dem Systemdesign oder spezifischen Funktionen beeinflusst. Die Einstellung zur Nutzung beeinflusst wiederum die Verhaltensabsicht, die schliesslich in die tatsächliche Nutzung mündet (Davis et al., 1989, S. 985).

Trotz seiner weitreichenden Anwendung und Bestätigung in der Forschung wird das *TAM* auch kritisch hinterfragt. Die Kritik bezieht sich auf dessen Grenzen bei der Bewertung des Nutzer:innenverhaltens, welches häufig subjektiv

anhand von Verhaltensintentionen wie interpersonellen Einflüssen gemessen wird (Ajibade, 2018, S. 4). Diese Einflüsse können in jedem Arbeitsumfeld anders sein, da Mitarbeitende häufig Regeln und formalen Richtlinien folgen müssen, was die Aussagekraft des Modells in Frage stellt. Im Weiteren wird bemängelt, dass *TAM* nicht externe Variablen wie z. B. Alter und Bildung berücksichtigt, welche die Technologieakzeptanz beeinflussen könnten. Kritiker:innen argumentieren, dass Verhalten durch verborgene Persönlichkeitsmerkmale motiviert wird, was die Messung und Vorhersage von Technologieakzeptanz mittels *TAM* komplex macht (Ajibade, 2018, S. 4). Die Fokussierung auf wenige zentrale Konstrukte wird als möglicherweise unzureichend für die Erklärung komplexer Entscheidungen und Verhaltensweisen in verschiedenen technologischen Kontexten angesehen (Ajibade, 2018, S. 5).

Bei der Entwicklung des *TAM* wurden später zusätzlich zur Nützlichkeit und Benutzerfreundlichkeit externe Variablen auf die beiden zentralen Konstrukte identifiziert. Diese Erkenntnisse führten später zur Entwicklung erweiterter Modelle wie dem *TAM2* (Venkatesh & Davis, 2000) und später dem *UTAUT* (Venkatesh, Morris, Davis & Davis, 2003) und *UTAUT2* (Tamilmani, Rana, Wamba & Dwivedi, 2021), welche zusätzliche Einflussfaktoren integriert und so eine differenziertere Betrachtung ermöglichen.

In Anbetracht dieser Diskussion wird das *TAM* in seiner ursprünglichen Form im Rahmen dieser Masterthesis dennoch als geeignetes Instrument erachtet, um die Akzeptanz neuer Technologien in diesem Kontext zu verstehen. Angesichts seiner kontinuierlichen Weiterentwicklung und der Anpassungsfähigkeit an unterschiedliche Forschungsgebiete dient das Modell als robuste Basis zur Untersuchung von Einflussfaktoren auf die Technologieakzeptanz in der Sozialen Arbeit. Das im nächsten Kapitel vorgestellte modifizierte Modell fusst auf dem *TAM*, wobei es so angepasst wird, dass es den Bedürfnissen und externen Faktoren des sozialarbeiterischen Kontextes entspricht und damit die Anwendbarkeit in diesem Untersuchungsfeld gewährleistet.

4.2 Angepasstes Theoriemodell und Hypothesenbildung

Das *TAM* in seiner ursprünglichen Form skizziert die Faktoren, auf Basis welcher die Akzeptanz innovativer Technologien im beruflichen Rahmen bestimmt werden, um die effektive Implementierung einer neuen Technologie im Arbeitsalltag zu gewährleisten. Im Falle von KI-gestützten Entscheidungshilfen gestaltet sich die Situation jedoch komplexer, als dass Sozialarbeitende lediglich deren Einsatz befürworten oder ablehnen könnten. Vielmehr stellt sich die Frage, inwiefern die von der KI generierten Einschätzungen bei einer Implementierung von den Sozialarbeitenden angenommen oder abgelehnt würden. Diese Tragweite überschreitet die Grenzen des traditionellen Akzeptanzbegriffs, der typischerweise auf Szenarien angewendet wird, in denen Technologien menschliche Arbeitsprozesse lediglich unterstützen. Verschiedene Studien zeigen, dass das Konzept des Vertrauens (vgl. Kapitel 2.4) eine zentrale Rolle bei der Implementierung von KI spielt (Yang & Wibowo, 2022, S. 2055). Das in dieser Arbeit zentrierte Untersuchungsfeld adressiert spezifisch Konstellationen, in denen KI-Systeme *mit* Menschen arbeiten oder sogar Entscheidungsoptionen *für* sie vorgeben. In solch einem Kontext wird Vertrauen als das entscheidende und zu untersuchende Konzept erachtet (Kohring, 2001, S. 99; Sassmannshausen & Heupel, 2020, S. 26).

Das im folgenden vorgestellte Modell stellt eine modifizierte Version des *TAM* dar, die speziell für die Untersuchung des Vertrauens in KI-Entscheidungen im Bereich der Sozialen Arbeit angepasst wurde. Vertrauen speziell in die Zuverlässigkeit und die fachliche Fundierung von KI-gestützten Entscheidungshilfen spielt dabei eine entscheidende Rolle: Es beeinflusst, wie Benutzende – in diesem Fall Sozialarbeitende – die Effektivität und Sicherheit der Nutzung entsprechender Tools bewerten. Vertrauen kann somit als eine zentrale Komponente der subjektiven Einstellung angesehen werden. Es wirkt direkt darauf, wie wahrscheinlich es ist, dass Sozialarbeitende bereit sind, sich auf die Urteile der KI zu verlassen.

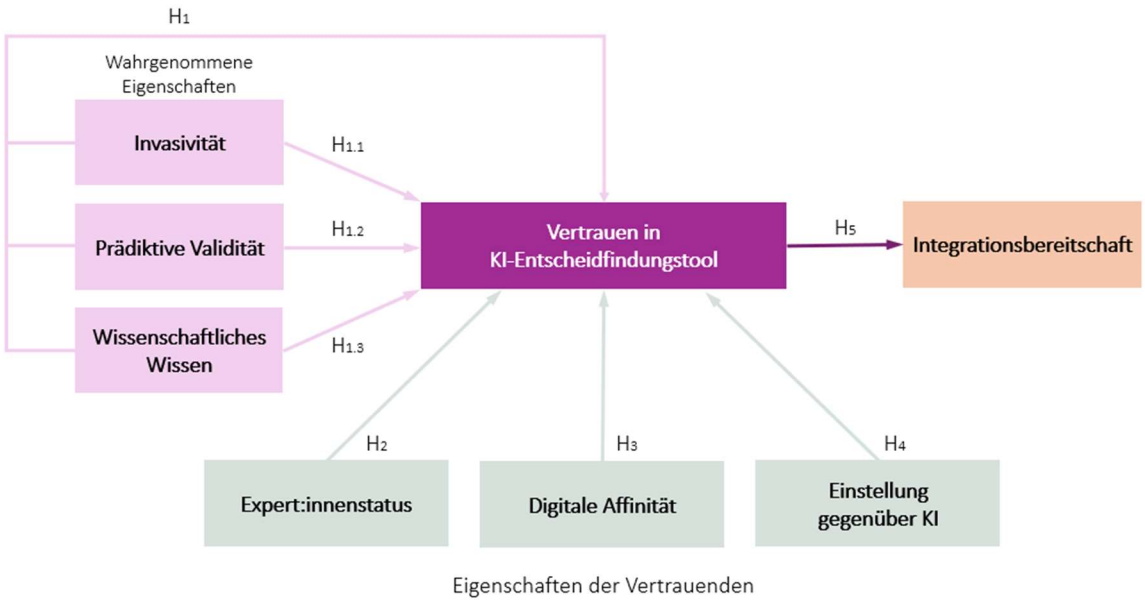
Da die KI-gestützten Entscheidungshilfen, welche in dieser Arbeit verwendet werden, nur fiktive Beispiele sind und die Zielgruppe der befragten Sozialarbeitenden zum heutigen Zeitpunkt (noch) nicht mit entsprechenden Instrumenten arbeitet, wurde die wahrgenommene Benutzerfreundlichkeit (*perceived ease of use*) sowie die tatsächliche Nutzung (*actual system use*) im modifizierten *TAM* nicht weiter berücksichtigt. Im Weiteren lässt sich die Soziale Arbeit nur bedingt mit anderen Branchen vergleichen, da sie sich kontinuierlich im Spannungsfeld zwischen Individuum, Profession und Öffentlichkeit bewegt. Die externen Elemente des ursprünglichen *TAM* lassen sich aus dieser Sicht nur teilweise auf das Arbeitsfeld anwenden, und eine direkte Übertragung würde der Vielschichtigkeit der Sozialen Arbeit nicht gerecht werden. Angesichts dieser Tatsache wurde in der Erarbeitung passender externer Elemente versucht herauszuarbeiten, welche Faktoren die Sozialarbeitenden hinsichtlich KI-gestützter Technologien effektiv beschäftigen. Aufgrund der Neuartigkeit der Thematik innerhalb der Sozialen Arbeit konnte hierzu jedoch nur limitiert auf bestehende Literatur zur Akzeptanz von Entscheidungshilfen zurückgegriffen werden[7].

Im angepassten *TAM* (*Abbildung 5*) finden sich entsprechend folgende zentrale Elemente:

1. Vertrauen in das KI-Entscheidungsfindungstool
2. Externe Faktoren bzw. wahrgenommene Eigenschaften (Hypothese H_1)
   Diese umfassen die wahrgenommene *Invasivität* ($H_{1.1}$), die wahrgenommene *prädiktive Validität* ($H_{1.2}$) und das wahrgenommene *wissenschaftliche Wissen* ($H_{1.3}$)
3. Eigenschaften der Vertrauenden: der *Expert:innenstatus* (H_2) die *digitale Affinität* (H_3) und die *allgemeine Einstellung gegenüber KI* (H_4)
4. *Integrationsbereitschaft* (Hypothese H_5)

[7] Eine Erweiterung der Recherche konnte mithilfe einer nicht repräsentativen Kurzumfrage mit 20 Sozialarbeitenden aus verschiedenen Arbeitsfeldern und mit unterschiedlicher Berufserfahrung via Whatsapp umgesetzt werden (*Anhang 2*). Ohne den spezifischen Kontext zu kennen, wurden die Sozialarbeitenden gebeten, jeweils drei positive und drei negative Gedanken zu äussern, welche sie bei einer potentiellen Einführung neuer KI-Tools in ihrem täglichen Arbeitsumfeld beschäftigen. Die schriftlichen Antworten wurden im Anschluss inhaltlich ausgewertet und in Kategorien zusammengefasst. Es zeigte sich, dass mit insgesamt 48 Nennungen die Erwartung an die Leistung der Tools (z. B. einheitlichere Entscheide) bei den positiven Gedanken überwogen. Bei den negativen Gedanken wurden insbesondere die Unsicherheit bezüglich fehlender Professionalität der KI (30) sowie Unsicherheiten zu möglichen Fehlentscheiden (Leistungserwartung) (14) genannt. Zudem wurde die Befürchtung geäussert, dass der Einsatz der neuen Tools zusätzlichen Aufwand bedeutet (Aufwandserwartung) (6). Auf Basis der Ergebnisse wurde versucht, die für die Sozialarbeitenden relevanten Faktoren heraus zu kristallisieren, um der Literaturrecherche zusätzlich Gewicht zu verleihen.

Abbildung 5 – Modifiziertes Hypothesenmodell in Anlehnung an *TAM* und Vertrauensmodell
Quelle: Eigene Darstellung (in Anlehnung an Davis et al., 1989; Meyer et al. 1995)



4.2.1 Externe Faktoren

Das Konstrukt der wahrgenommenen Nützlichkeit wurde aus dem Original-*TAM* übernommen und zu drei Subkategorien geformt. Die externen Faktoren des angepassten *TAM* beinhalten die Faktoren der *Invasivität*, der *prädiktiven Validität* und des *wissenschaftlichen Wissens.*

In der Literatur wird wiederholt betont, dass der Erfolg einer Innovation massgeblich davon abhängt, ob sie den Anwendenden einen Mehrwert bietet (Nörr, 2016, S. 66). Davis (1989) selbst unterstreicht die Bedeutung der wahrgenommenen Nützlichkeit als entscheidenden Faktor für die Akzeptanz neuer Technologien. Er nimmt an, dass Technologien, die keinen Mehrwert bieten oder die Produktivität gar reduzieren, von den Nutzenden abgelehnt werden (S. 320).

Die daraus abgeleitete Hypothese H_1 lautet:

H_1 Wenn Sozialarbeitende die Invasivität einer KI-Empfehlung als weniger intensiv im Vergleich zu ihrer eigenen Einschätzung empfinden ($H_{1.1}$), wenn sie die prädiktive Validität der Tools als hoch einschätzen ($H_{1.2}$), und wenn sie wissenschaftliches Wissen der Sozialen Arbeit innerhalb von KI-gestützten Entscheidungshilfen wahrnehmen ($H_{1.3}$), dann neigen sie eher dazu, einer KI-gestützten Applikation zu vertrauen.

Faktor Invasivität

Der erste Faktor des weiterentwickelten *TAM* bezieht sich direkt auf die Entscheidungen der Sozialarbeitenden. Der externe Faktor *Invasivität* stellt hierbei eine neue Variable dar, welche im ursprünglichen *TAM* nicht vorkommt, jedoch als Teilaspekt der *wahrgenommenen Nützlichkeit* verortet werden kann.

Wie im Kapitel 2.2 beschrieben, werden (professionelle) Entscheidungen in einem Zusammenspiel aus heuristisch-intuitiven und kontrolliert-analytischen Entscheidungsfindungsprozessen gefällt (Enosh & Bayer-Topilsky, 2015, S. 3; Nitzsch, 2021, S. 6). In diesem Prozess spielt die Intuition eine zentrale Rolle, da sie eine notwendige Kompetenz für den Umgang mit der inhärenten Komplexität und Dynamik der täglichen Fallarbeit darstellt. Sie ermöglicht es Sozialarbeitenden, unter Zeitdruck und bei unvollständiger Information Entscheidungen zu treffen. Wenn ein KI-Tool nun eine Einschätzung gibt, welche die Sozialarbeitenden als «kontraintuitiv» wahrnehmen – d.h. das Tool schlägt beispielsweise eine strengere Massnahme vor, als die Sozialarbeitenden es selbst tun würden – kann dies einen Einfluss auf das Vertrauen in das Tool haben. Mayer et al. (1995) argumentieren, dass Vertrauen dann aufgebaut wird, wenn das wahrgenommene Risiko niedriger als das Vertrauensniveau ist. Es wird davon ausgegangen, dass die Invasivität einer Einschätzung eines Tools das Vertrauen beeinflusst: Wird ein KI-Tool als invasiv wahrgenommen, könnte dies möglicherweise dazu führen, dass die Befragten das Risiko für die Annahme der Einschätzung höher einschätzen, was sich wiederum negativ auf das Vertrauen in die KI-Entscheidungshilfe auswirken kann.

Im Kontext des angepassten *TAM*-Modells reflektiert die *Invasivität*, inwiefern kontraintuitive Einschätzungen durch das Tool das Vertrauen der Sozialarbeitenden beeinflussen könnten. Dies führt zur Formulierung der ersten Subhypothese:

$H_{1.1}$ Wenn die Einschätzung einer KI-gestützten Entscheidungshilfe invasiver ausfällt als die Einschätzung der Sozialarbeitenden, ist das Vertrauen in das Tool geringer.

Faktor prädiktive Validität

Prädiktive Validität als zweiter Faktor bezieht sich darauf, wie genau eine Technologie oder ein Tool zukünftige Ergebnisse oder Verhaltensweisen vorhersagen kann. Im Kontext von KI-gestützten Entscheidungshilfen in der Sozialen Arbeit misst die prädiktive Validität die Zuverlässigkeit der KI-Empfehlungen hinsichtlich ihrer Fähigkeit, zutreffende Vorhersagen über den Fallverlauf zu treffen (Cuccaro-Alamin, Foust, Vaithianathan & Putnam-Hornstein, 2017). Eine hohe prädiktive Validität bedeutet, dass die Vorhersagen des Tools häufig mit den tatsächlichen Ergebnissen übereinstimmen, was die Effektivität und die praktische Anwendbarkeit des Tools erhöht und somit die *wahrgenommene Nützlichkeit* steigert.

Im Kontext des angepassten *TAM*-Modells reflektiert die *prädiktive Validität* die Treffsicherheit einer KI-Anwendung. Dies führt zur Formulierung der zweiten Subhypothese:

$H_{1.2}$ Je grösser die prädiktive Validität der KI-gestützten Entscheidungshilfe, desto höher das Vertrauen von Sozialarbeitenden in das Tool.

Faktor wissenschaftliches Wissen

Der dritte Faktor *wissenschaftliches Wissen* lässt sich ebenfalls der Subkategorie der *wahrgenommenen Nützlichkeit* zuordnen. Dieser Faktor analysiert, inwiefern die Wahrnehmung von wissenschaftlichem Wissen in KI-gestützten Entscheidungshilfen das Vertrauen in die Tools steigert.

Professionalität und professionelles Handeln haben in der Sozialen Arbeit einen hohen Stellenwert und sind eines ihrer Kernanliegen (Campayo, 2020, S. 290; Ehlert, 2019). Der Begriff der Professionalität wird dabei verwendet, um Qualitätskriterien auf die Soziale Arbeit anzuwenden, die immer auch Gegenstand von fachlichen Definitionsversuchen und politischen Debatten sind (Kubisch & Franz, 2022, S. 413-414). In verschiedenen Ansätzen wird einheitlich die Anwendung von wissenschaftlichem Wissen als Kennzeichen von Professionalität und Qualität in der Sozialen Arbeit angesehen (Becker-Lenz & Müller, 2009, S. 198). Mit dem wissenschaftlichen Diskurs und insbesondere mit der Entwicklung von (Professionalisierungs-)Theorien und der Generierung von empirischem Fachwissen konnte in den vergangenen Jahrzehnten eine Professionalisierung stattfinden, die heute ein integraler Bestandteil des Selbstverständnisses der Sozialen Arbeit darstellt. Die Professionalisierung umfasst sowohl die individuelle Kompetenzentwicklung und Qualifikation einer Fachperson als auch kollektive Prozesse in der fachlichen Entwicklung der Berufsfelder (Ehlert, 2019). Im Kontext der Sozialen Arbeit gibt es anhaltende Debatten über die Befürchtung einer möglichen «De-Professionalisierung», verursacht durch den zunehmenden Einsatz von digitalen Technologien, die möglicherweise das menschliche Urteilsvermögen und die fachliche Integrität ersetzen könnten (Schneider, 2021, S. 123). Diese Sorge scheint besonders relevant, wenn es um den Einsatz von KI-Entscheidungshilfen geht, welche die Gefahr bergen könnten, die Rolle menschlicher Expertise zu untergraben[8].

Im Kontext des angepassten *TAM*-Modells reflektiert das *wissenschaftliche Wissen*, inwiefern das Tool als zuverlässige, fundierte und fachgerechte Ressource wahrgenommen wird. Dies führt zur Formulierung der dritten Subhypothese:

$H_{1.3}$ Je eher innerhalb KI-gestützten Entscheidungshilfen wissenschaftliches Wissen der Sozialen Arbeit wahrgenommen wird, desto höher ist das Vertrauen von Sozialarbeitenden in das Tool.

[8] Die Einbindung von Fachwissen in die Trainingsdaten von KI-Entscheidungssystemen ist prinzipiell realisierbar: Expertsysteme ermöglichen es, dass neben vorhandenen Datensätzen (*First Party Dat*a) explizites Fachwissen in die Anwendung integriert wird. Aus diesen Daten müssen anschliessend die für Entscheidungen relevanten Informationen extrahiert werden (Kreutzer, 2023, S. 53). Die Kombination dieser Daten kann den Erkenntnishorizont der Sozialarbeitenden erweitern und neue Perspektiven eröffnen (Kreutzer, 2023, S. 52-53).

4.2.2 Eigenschaften und Einstellung der Vertrauenden

Im Kontext der Sozialen Arbeit wird angenommen, dass die erfolgreiche Implementierung von KI-gestützten Entscheidungshilfen in die Alltagspraxis nicht nur von der technologischen Entwicklung abhängt, sondern auch massgeblich durch die Einstellungen und Erfahrungen der Sozialarbeitenden beeinflusst wird. Hier wird auch der Kritik von Ajibade (2018, S. 4) Rechnung getragen (vgl. Kapitel 4.1). Basierend auf dem im Kapitel 2.4 dargelegten Vertrauensmodell von Mayer et al. (1995) können die beiden Variablen *Expert:innenstatus* und *digitale Affinität* als Elemente der Vertrauensneigung interpretiert werden. Ebenso wird die generelle *Einstellung gegenüber KI* als ein entscheidender Faktor betrachtet, der das Vertrauen in KI-gestützte Entscheidungshilfen beeinflusst. Dabei wird vermutet, dass jüngere Generationen aufgrund ihres Aufwachsens mit digitalen Technologien (Digital Natives) tendenziell mehr Vertrauen in entsprechende Tools haben, als ältere Generationen (Palfrey & Gasser, 2008, S. 161).

Der *Expert:innenstatus* wird dabei als eine Kombination aus Lebensalter und beruflicher Erfahrung definiert. Diese Variable reflektiert die Annahme, dass mit zunehmendem Alter und mehr beruflicher Erfahrung die Einstellungen gegenüber KI variieren könnten. Es wird vermutet, dass erfahrene Sozialarbeitende möglicherweise skeptischer gegenüber dem Einsatz von KI-Tools sind, da sie auf bewährte, traditionelle Methoden vertrauen oder kritischer gegenüber den Limitationen der Tools sind. Dies führt zur nächsten Hypothese:

H_2 Je höher der Expert:innenstatus der Sozialarbeitenden, desto geringer das Vertrauen in KI- gestützte Entscheidungshilfen.

Die Variable der *digitalen Affinität* bezieht sich auf die Vertrautheit im Umgang mit digitalen Technologien. Diese Variable wird aus den Antworten der Teilnehmenden des Items digitale Affinität (vgl. Kapitel 5.3.2) abgeleitet und vermutet, dass eine höhere digitale Affinität mit einem stärkeren Vertrauen in die Leistungsfähigkeit und Zuverlässigkeit von KI-gestützten Entscheidungshilfen korreliert. Personen, die sich sicher im Umgang mit digitalen Technologien fühlen, könnten eher geneigt sein, auch die Potenziale von KI-gestützten Werkzeugen zu erkennen und zu schätzen. Daraus resultiert die dritte Hypothese:

H_3 Je höher die digitale Affinität der Sozialarbeitenden, desto höher das Vertrauen in KI-gestützte Entscheidungshilfen.

Die generelle *Einstellung gegenüber KI* bezieht sich auf die persönliche Haltung von Sozialarbeitenden gegenüber KI. Es wird angenommen, dass diese Haltung das Vertrauen in KI-gestützte Entscheidungshilfen massgeblich beeinflusst. Ein starkes Interesse an KI und die Erwartung, dass KI zur Effizienzsteigerung im Berufsalltag beiträgt, könnten das Vertrauen in entsprechende Tools positiv beeinflussen. Im Gegensatz dazu könnte eine allgemeine Skepsis gegenüber KI und die Befürchtung, dass der Einsatz solcher Technologien zu Arbeitsplatzverlusten führen könnte, das Vertrauen mindern. Die vierte Hypothese lautet daher:

H_4 Je positiver die generelle Einstellung gegenüber KI, desto höher das Vertrauen in KI-gestützte Entscheidungshilfen.

4.2.1 Integrationsbereitschaft

Im Kontext des ursprünglichen *TAM* wird die Verhaltensabsicht zur Nutzung (*behavioral intention to use*) durch die Einstellung zur Nutzung (*attitude toward using*) beeinflusst. Diese Verhaltensabsicht ist laut Davis (1985) als die subjektive Wahrscheinlichkeit definiert, mit der eine Person ein spezifisches Verhalten auszuführen gedenkt (S. 16).

Basierend darauf lässt sich argumentieren, dass das Vertrauen in die Einschätzungen von KI-gestützten Tools eine kritische Rolle in der Formung der subjektiven Einstellung zur Nutzung entsprechender Tools spielt. Vertrauen kann als eine Facette der subjektiven Einstellung betrachtet werden, die direkt die Wahrscheinlichkeit erhöht, dass Sozialarbeitende sich für deren Nutzung entscheiden. Ein höheres Vertrauen in die Einschätzungen der Tools führt wahrscheinlich zu einer gesteigerten Nutzungsabsicht entsprechender Entscheidungshilfen.

In dieser Arbeit wird von Integrationsbereitschaft und nicht von Nutzungsabsicht gesprochen, da der Fokus darauf liegt, wie bereit Sozialarbeitende sind, KI-gestützte Tools in ihre täglichen Arbeitsprozesse zu integrieren. Dies umfasst nicht nur die Absicht zur Nutzung, sondern auch die Bereitschaft zur Anpassung und Akzeptanz neuartiger KI-Tools im beruflichen Umfeld.

Dies führt zur letzten Hypothese:

H_5 Je höher das Vertrauen in Einschätzung eines KI-gestützte Entscheidungshilfen, desto höher die Integrationsbereitschaft entsprechender Tools im Arbeitsalltag

5 Methodisches Vorgehen

In der Sozialen Arbeit wurde die Wahrnehmung und das Vertrauen von Sozialarbeitenden in KI-gestützte Entscheidungshilfen bislang nur am Rande erforscht. Im Gegensatz dazu gibt es in anderen Branchen, wie der Produktion oder im Bereich Human Resources, eine breitere Basis an empirischen Studien, die verschiedene Ansätze beleuchten (Kröller & Schwarz, 2022; Sassmannshausen, 2019; Lee, 2018). Dies ermöglichte eine fundiertere Auswahl geeigneter Methoden. Eine bewährte Methode zur Erkenntnisgewinnung ist das szenariogestützte Vignettendesign, welches in der Sozialpsychologie und Ethikforschung weit verbreitet ist (Lee, 2018). Angesichts der emotionalen Sensibilität des Themas wurde für diese Thesis eine ähnliche methodische Herangehensweise gewählt, um die Wahrnehmung von KI-gestützten Entscheidungshilfen unter Sozialarbeitenden zu untersuchen (Auspurg, Hinz, Liebig & Sauer, 2014).

Im Folgenden wird zunächst das ausgewählte Forschungsdesign vorgestellt. Anschliessend werden die Charakteristika des experimentellen Vignettendesigns sowie dessen Vorzüge und Einschränkungen erörtert. Danach erfolgt eine Beurteilung der Vignetten. Im Anschluss folgt eine Beschreibung der Stichprobe und des Erhebungsinstruments, einschliesslich der experimentellen Vignetten (Fallbeispiele) und der Operationalisierung der Variablen. Abschliessend wird die Auswertungsmethode beschrieben, mit Schwerpunkt auf der Datenaufbereitung und der Analysestrategie.

5.1 Forschungsdesign: multifaktorielles, experimentelles Vignettendesign

Das Forschungsdesign der Thesis basiert auf einer quantitativen Querschnittuntersuchung (Baur & Blasius, 2014), wodurch soziale Zusammenhänge mithilfe numerischer Daten erklärt und Hypothesen getestet werden (Richter, Paier & Reiger, 2021). Für die Datenerhebung und Hypothesentestung wurde ein multifaktorielles, experimentelles Vignettendesign, auch «faktorielle Survey», verwendet (Jasso, 2006; Rossi & Anderson, 1982). Dieses ermöglicht es, die Wirkung verschiedener unabhängiger Variablen auf eine abhängige Variable zu untersuchen, indem den Teilnehmenden verschiedene Szenarien präsentiert werden, die systematisch variieren (Jasso, 2006; Rossi & Anderson, 1982). Das Vignettendesign eignet sich damit besonders für die Erforschung emotionaler oder moralisch sensibler Themen und reduziert die Tendenz zu sozial erwünschten Antworten (Steiner & Atzmüller, 2006; Dülmer, 204, S. 721). Die statistische Auswertung zielt darauf ab, die Relevanz der verschiedenen Faktoren im Hinblick auf die Bewertungen der Vignetten zu bestimmen und Unterschiede zwischen Teilnehmenden oder Teilnehmer:innengruppen zu identifizieren (Steiner & Atzmüller, 2006). Durch die zufällige Zuteilung der Vignetten können probabilistisch-kausale Aussagen[9] formuliert werden (Aguinis & Bradley, 2014; Evans, Roberts, Keeley, Blossom, Amaro, Stough, Canter, Robles & Reed, 2015).

[9] In der Diskussion um probabilistisch-kausale Theorien steht die Annahme im Vordergrund, dass Ursachen die Wahrscheinlichkeit ihrer Effekte verändern (Hitchcock, 2018). Diese Perspektive impliziert, dass die Veränderung in der Wahrscheinlichkeit eines Effekts durch eine spezifische Ursache dessen Eintrittswahrscheinlichkeit erhöht oder senkt (Kelle, 2008, S. 181-200). Für die empirische Feststellung probabilistisch-kausaler Effekte ist es notwendig, Störvariablen auszuschliessen. Scheitert dies, beschränkt sich die Erkenntnis auf statistische Zusammenhänge (Gautschi, 2021, S. 222).

5.1.1 Charakteristika des experimentellen Vignettendesigns

In den folgenden Abschnitten werden die Charakteristika des experimentellen Vignettendesigns thematisiert sowie die konzeptuellen als auch die methodischen Überlegungen im Rahmen dieser Arbeit thematisiert.

5.1.1.1 Das Vignettenuniversum

Der Kern der Methodik des experimentellen, multifaktoriellen Vignettendesigns bildet das Vignettenuniversum (auch «Vignettenpopulation» oder «volles Faktorial»), welches die Gesamtheit aller möglichen Kombinationen von Bedingungen oder Szenarien darstellt, welche durch unterschiedliche Ausprägungen der in den Vignetten enthaltenen Dimensionen erzeugt werden können. Die Gesamtzahl der Vignetten ergibt sich aus der Multiplikation aller Ausprägungen der verschiedenen Dimensionen miteinander (Rost, 2018, S. 43-45). Diese Sammlung ermöglicht es, ein breites Spektrum an spezifischen Situationen zu simulieren, um die Reaktionen der Teilnehmenden in verschiedenen Kontexten zu untersuchen und ihre Verhaltensintentionen zu messen.

5.1.1.2 Orthogonalität und Balance

Um die Validität und die Zuverlässigkeit der daraus gewonnenen Daten zu gewährleisten, sind die Konzepte der Orthogonalität und Balance von Bedeutung. Orthogonalität bedeutet, dass jede Ausprägung einer Dimension gleichmässig mit jeder anderen kombiniert wird, um Korrelationen zwischen den Dimensionen und ihren Wechselwirkungen zu vermeiden (Rost, 2018, S. 45). Balance gewährleistet, dass jede Ausprägung innerhalb einer Dimension gleich häufig auftritt, was die Vignetten maximal variabel macht und eine Schätzung des Einflusses jeder Dimension ermöglicht (Rost, 2018, S. 45). Für die Bestimmung der Haupteffekte (Einfluss einer einzelnen Dimension auf die abhängige Variabel) wie auch die Interaktionseffekte (Einfluss von mindestens zwei Dimensionen auf die abhängige Variabel) sind diese beiden Eigenschaften unerlässlich.

Die Bewertung des gesamten Vignettenuniversums durch Teilnehmende ist in der Praxis i.d.R. nicht umsetzbar, was dazu führen kann, dass Stichproben aus dem Vignettenuniversum nicht immer die ideale Orthogonalität und Balance aufweisen. Die Genauigkeit der Schätzung der Effekte kann dadurch beeinträchtigt werden, was sorgfältige Überlegungen bei der Gestaltung der Vignettenstudien erfordert. Theoretisch kann das Vignettenuniversum eine beliebige Anzahl an Dimensionen enthalten[10], es ist jedoch sinnvoll, die Menge der Dimensionen im Verhältnis zur erwarteten Stichprobe zu bestimmen (Rost, 2018, S. 48). In anderen Worten: Je kleiner die Stichprobe, desto weniger Dimensionen sollten verwendet werden. Ebenso wird nicht empfohlen, mehr als sechs Faktoren mit mehr als fünf Ausprägungen einzuschliessen, da die Komplexität für die Befragten mit zunehmender Anzahl Faktoren

[10] Hennessy entwickelte 1993 eine Studie mit 24 verschiedenen Dimensionen mit einer unterschiedlichen Anzahl an Ausprägungen: drei mit zwei Ausprägungen, fünfzehn mit drei Ausprägungen, drei mit vier Ausprägungen, zwei mit fünf Ausprägungen und eine mit sechs Ausprägungen ($2^3 \times 3^{15} \times 4^3 \times 5^2 \times 6^1$). Sein Vignettenuniversum umfasste damit ein Total von 1.1 Billionen Vignetten (Taylor, 2006, S. 1193).

zunimmt (Dülmer, 2014, S. 724). Hinsichtlich Orthogonalität und Balance gilt, dass eine grössere Stichprobe in Verhältnis zum gesamten Vignettenuniversum vorteilhafter ist, um Ungenauigkeiten in der Analyse zu vermeiden (Rost, 2018, S. 48).

5.1.2 Vorteile und Limitationen des Designs

Gemäss Taylor (2006) gilt das Vignettendesign in der Untersuchung von Faktoren in Entscheidungsprozessen als besonders robust und ethisch. Es vereint die Stärken eines experimentellen Ansatzes, indem es die Auswirkungen verschiedener Faktoren auf die Entscheidung isoliert, mit der Validität einer Umfrage, die der realen Praxis nahekommt. Das Design ermöglicht es dadurch, die Einflüsse auf Entscheidungen genauer zu verstehen (S. 1202).

Im Folgenden werden zunächst die Vorteile des Vignettendesigns bei der Untersuchung von Faktoren in Entscheidungsprozessen aufgezeigt (Taylor, 2006; Auspurg et al., 2014; Döring & Bortz, 2016; Rost, 2018):

1. Hohe interne Validität

   Der Ansatz ermöglicht es, spezifische Auswirkungen einzelner Faktoren auf die Entscheidungsfindung der Teilnehmenden zu isolieren und zu untersuchen. Eine Studie ist intern valide, wenn ihre Ergebnisse klar auf kausale Zusammenhänge zwischen den unabhängigen und abhängigen Variablen hinweisen, sodass die beobachteten Effekte eindeutig auf die Manipulation der unabhängigen Variablen zurückgeführt werden können.
2. Kontrolle über experimentelle Variablen

   Durch die Verwendung von Vignetten als experimentelle Stimuli können Forschende die Offenlegung von Informationen kontrollieren und so sicherstellen, dass alle Teilnehmende ähnliche Situationen bewerten.
3. Reduktion von Reaktanz

   Da die Entscheidungen auf hypothetischen Szenarien basieren, können Teilnehmenden möglicherweise objektivere Antworten geben, da sie nicht direkt von den Konsequenzen ihrer Entscheidungen betroffen sind.
4. Flexibilität in der Variablenmanipulation

   Der Ansatz ermöglicht es, verschiedene Variablen zu manipulieren und ihre Auswirkungen auf die Entscheidungsfindung zu untersuchen, was zu einem tieferen Verständnis der zugrundeliegenden Mechanismen führen kann.

Gleichzeitig zeigen sich aber auch Limitationen des Vignettendesigns, welche bei der Planung und der Interpretation bei Forschungsarbeiten berücksichtigt werden müssen (Taylor, 2006; Auspurg et al., 2014; Döring & Bortz, 2016; Rost, 2018):

1. Begrenzte externe Validität

   Obwohl der Ansatz eine hohe interne Validität aufweist, da die Entscheidungen solcher in realen Situationen ähneln, könnten die Ergebnisse dennoch nicht vollständig auf andere Kontexte übertragbar sein.

Dies könnte auf die spezifischen Merkmale der verwendeten Vignetten und der ausgewählten Stichprobe zurückzuführen sein. Eine Studie besitzt externe Validität, wenn sich ihre Ergebnisse über die spezifischen Bedingungen der Untersuchung und die teilnehmenden Personen hinaus auf breitere Situationen und Gruppen übertragen lassen.

2. Beschränkung auf beobachtbare Faktoren

   Der Ansatz konzentriert sich in der Regel auf beobachtbare Faktoren in den Vignetten und berücksichtigt möglicherweise nicht alle relevanten internen oder nicht offensichtlichen Einflussfaktoren, die die Entscheidungsfindung beeinflussen könnte.

3. Mögliche Reaktanz oder Antwortverzerrung

   Teilnehmende könnten dazu neigen, auf die Vignetten oder Fragen auf eine Weise zu reagieren, die ihrer wahrgenommenen sozialen Erwünschtheit entspricht, anstatt ihre tatsächlichen Überzeugungen oder Verhaltensweisen widerzuspiegeln.

4. Komplexität der Vignetten und geringe Reliabilität

   Die Komplexität der in den Vignetten präsentierten Szenarien könnte zu Überforderung oder Ermüdung der Teilnehmenden führen, insbesondere wenn sie gebeten werden, eine grosse Anzahl von Vignetten zu bearbeiten.

Zusammenfassend bietet das Vignettendesign Möglichkeiten, die Komplexität von Entscheidungsprozessen in kontrollierten, doch realitätsnahen Szenarien zu untersuchen. Trotz seiner Fähigkeit, präzise Einblicke in die Wirkung spezifischer Variablen zu gewähren, müssen die Einschränkungen bezüglich der externen Validität und der möglichen Antwortverzerrungen im Blick behalten werden. Die Herausforderung liegt somit nicht nur in der sorgfältigen Konstruktion der Vignetten, sondern auch in der kritischen Interpretation der Ergebnisse, um zu verlässlichen und übertragbaren Erkenntnissen zu gelangen.

5.1.3 Beurteilung der Vignetten

Die Teilnehmenden geben im Fragebogen ihre Urteile zu den gelesenen Vignetten ab. Damit die Resultate interpretierbar sind, ist eine Messung mittels einer sinnvoll gewählten Skala essentiell (Rost, 2018, S. 69). Beim experimentellen Vignettendesign werden am häufigsten Likert-Skalen zur Beurteilung der Vignetten verwendet, da sie sich zur Messung von eher subjektiven Merkmalen wie Normen, ethischen Prinzipen oder Zugehörigkeiten zu Statusgruppen eignen. Diese Skalen können symmetrisch gestaltet sein, wodurch Befragte gezwungen sind, eine Tendenz zu eher positiven oder negativen Antworten zu zeigen. Es empfiehlt sich, Likert-Skalen mit mindestens fünf Antwortmöglichkeiten zu verwenden, um sie als «quasi-metrisch» einstufen und sie in linearen Regressionsmodellen einsetzen zu können (Rost, 2018, S. 69). Ein Nachteil von Likert-Skalen besteht darin, dass sie Befragte auf eine bestimmte Antwortspanne beschränken, was zu sogenannten «zensierten Antworten» führen kann (eine

Teilnehmerin bewertet in der ersten Vignette beispielsweise mit «stimme überhaupt nicht zu», erhält aber im nächsten Task eine neue Vignette, welcher sie noch weniger zustimmen würde) (Rost, 2018, S. 70)[11].

5.2 Stichprobe

Die Zielpopulation (Döring & Bortz, 2016, S. 292) dieser Arbeit umfasst die Beistandspersonen im gesetzlichen Kindesschutz, die in verschiedenen deutschsprachigen Kantonen der Schweiz tätig sind. Aufgrund der föderalistischen Struktur der Schweiz unterscheiden sich die Zuständigkeiten und Aufgabenbereiche der Berufsbeistandschaften von Kanton zu Kanton, wobei in einigen Kantonen, wie beispielsweise St. Gallen, die Kindeswohlabklärungen nicht direkt von den Berufsbeistandschaften, sondern von der KESB selber durchgeführt werden[12]. Die Beistandspersonen verfügen in der Regel über fundierte Kenntnisse zu den Themen der in dieser Arbeit behandelten Fallbeispiele (Kindeswohlabklärungen und Interventionsmöglichkeiten in der Jugendhilfe), und es wurde davon ausgegangen, dass sie sich gut in die geschilderten Situationen hineinversetzen können und diese grösstenteils ihrem Arbeitsalltag entsprechen. Details zur Zusammensetzung und den konkret erreichten Teilnehmenden dieser Studie werden in den nachfolgenden Abschnitten dargestellt.

Um eine möglichst umfassende Repräsentation zu erreichen, wurden zunächst gezielt Beistandspersonen im gesetzlichen Kindesschutz aus verschiedenen Kantonen mit sowohl städtischen als auch ländlichen Gebieten angeschrieben. Es wurden bewusst keine KESB angeschrieben, obwohl diese in gewissen Kantonen die Kindeswohlabklärungen selbst durchführen, da hier i. d. R. nicht nur Sozialarbeitende, sondern auch Personen aus weiteren Fachdisziplinen (z. B. Psychologie, Rechtswissenschaften, Pädagogik) an den Abklärungen beteiligt sind (Gautschi & Schnurr, 2024, S. 37). An dieser Stelle kann argumentiert werden, dass in denjenigen Kantonen, in welchen die Kindeswohlabklärungen direkt durch die KESB durchgeführt werden, die Beistandspersonen nicht mit dieser Aufgabe bzw. mit dem ersten Fallbeispiel vertraut sind. Dieser Einwand ist berechtigt, allerdings sind Beistandspersonen aufgrund ihrer Aufgaben innerhalb des gesetzlichen Rahmens häufig mit ähnlicher Thematik vertraut bzw. überschneiden sich ihre Aufgaben mit derjenigen von Abklärenden in vielerlei Hinsicht (Intervention bei sich zuspitzenden Fallverläufen, Evaluation mittels Rechenschaftsbericht alle zwei Jahre, usw.) (Estermann, Hauri & Vogel, 2022, S. 209-212). Gleichzeitig wurde im ersten Fallbeispiel bewusst nicht näher spezifiziert, in welchem Rahmen die Abklärung stattgefunden hat und ob z. B. ein strukturiertes Abklärungsinstrument eingesetzt wurde. Das zweite

[11] In der Theorie der statistischen Testverfahren gibt es unterschiedliche Meinungen darüber, ob eine Likert-Skala als metrisch oder ordinal betrachtet werden sollte (Döring & Bortz, 2016, S. 269). Der entscheidende Aspekt ist hierbei, ob die Abstände zwischen den einzelnen Antwortkategorien gleich gross sind (also der Unterschied zwischen «stimme voll und ganz zu» und «stimme eher zu» dem Abstand zwischen «stimme eher zu» und «stimme teilweise zu» entspricht.

[12] Bei der Erarbeitung des Fragebogens lag der Fokus auf dem Aufbau des Kindesschutzes im Kanton Zürich, dem einwohnerstärksten Kanton der Schweiz. Im Kanton Zürich übernimmt das Amt für Jugend und Berufsberatung (AJB) bzw. in der Stadt Zürich die Sozialen Dienste der Stadt Zürich (SOD) Aufträge der KESB und Gerichte im Bereich des Kindesschutzes (Kanton Zürich, 2024, S. 4; Stadt Zürich, 2024). Neben der Führung von zivilrechtlichen Mandaten im Kindesschutz übernehmen die Mitarbeitenden der Kinder- und Jugendhilfestellen des AJB bzw. der SOD auch Abklärungen gemäss Art. 446 Abs. 2 ZGB im Auftrag der KESB bzw. der Gerichte (Kanton Zürich, 2024, S. 4). Entsprechend wurden die im Fragebogen gewählten Fallbeispiele sowohl auf Abklärungen im Kindesschutz (Fallbeispiel *KIBAT*) als auch auf zivilrechtlich geführte Mandate (Fallbeispiel *WIEW*) abgestimmt.

Fallbeispiel, das im Fragebogen verwendet wurde (Interventionsplanung innerhalb Vormundschaft) richtete sich hingegen insbesondere an die Aufgaben von Berufsbeistandspersonen. Die Kontaktaufnahme fokussierte sich entsprechend auf Sozialarbeitende, um eine spezifische, homogene Zielpopulation anzusprechen.

Da es keine gesamtschweizerische Liste der Beistandspersonen gibt, startete am 17. März 2024 eine ausgiebige Internetrecherche, woraufhin zunächst die Leitungspersonen der Berufsbeistandschaften in ausgewählten deutschsprachigen Kantonen direkt angeschrieben wurden. In der Region Zürich wurden die kantonalen Kinder- und Jugendhilfezentren (kjz), wo Beistandschaften im Kindesschutz geführt werden, am 22. März 2024 durch den Fachbereich KJH des Amts für Jugend und Berufsberatung über die Befragung in Kenntnis gesetzt. Die Information wurde via Leitungskonferenz an die Leitenden und zur Weiterleitung an die Mitarbeitenden verschickt. In der Stadt Zürich erfolgte die Verteilung des Fragebogens durch direkte Kontakte zu Stellenleitenden und Mitarbeitenden sowie über einen Aufruf im Newsletter des Fachstabs Kinder- und Jugendhilfe. So erreichte die Umfrage alleine im Kanton Zürich schätzungsweise 350 Beistandspersonen. Zusätzlich wurde ein Aufruf über LinkedIn gestartet, der breiten Anklang fand und 17 Mal geteilt wurde. Trotz Kontaktaufnahme über verschiedene Kanäle und der Verlosung von 2 x Fr. 200.00 zeigten die angesprochenen Personen Zurückhaltung beim Ausfüllen des Fragebogens, und die erreichte Auswahlpopulation entsprach bei weitem nicht der angestrebten Zielpopulation. Deshalb wurde am 4. April 2024 beschlossen, den Kreis der Auswahlpopulation zu erweitern, indem zusätzlich alle weiteren Berufsbeistandschaften in den deutschsprachigen Kantonen kontaktiert wurden. Für jede Vignette wurde das Ziel verfolgt, Antworten von mindestens 20 Personen zu erhalten, um signifikante Effekte feststellen zu können.

Zwischen Mitte März und Mitte April 2024 hat die Autorin insgesamt 162 E-Mails versendet. Der Link der Umfrage wurde insgesamt 466 Mal angeklickt, wobei jedoch 187 Personen keine Fragen beantwortet haben und 60 Personen während des Ausfüllens die Befragung abgebrochen haben. Entsprechend konnten diese Fragebogen nicht verwendet werden, da keine der beiden Vignetten beantwortet wurden. Weitere 15 Personen starteten die Umfrage, beantworteten die erste Filterfrage, ob sie aktuell in einer Jugendhilfestelle arbeiten, mit «nein», weshalb sie von der weiteren Durchführung ausgeschlossen wurden. Nach der kompletten Bereinigung konnten noch insgesamt 204 Fälle (n=204) ausgewertet werden, bei welchen mindestens eine Vignette beantwortet wurde. Die Ausschöpfungsquote beträgt dabei 26.21%. Die detaillierte statistische Auswertung folgt in Kapitel 6.1.

Tabelle 2 – Stichprobenbeschreibung

Eingrenzung Region	Deutschschweiz
Spezifiziertes Tätigkeitsfeld	Berufsbeistandschaften Kindesschutz
Auswahlverfahren der Teilnehmenden	Gelegenheitsstichprobe
Zeitraum des Versands	Mitte März bis Mitte April 2024
Anzahl gestarteter Fragebögen	466
Anzahl abgebrochener Fragebögen	60
Anzahl beendeter Fragebögen	219
Gesamtanzahl in der Analyse berücksichtigter Datensätze	204

5.3 Erhebungsinstrument

Die Teilnehmenden wurden im Begrüssungstext über den Inhalt und die Zielgruppe der Befragung informiert. Zudem wurden sie darauf hingewiesen, dass ihre Antworten anonym behandelt werden und sie wurden gebeten, alle Fragen vollständig auszufüllen. Der Fragebogen (*Anhang 3*) umfasste insgesamt 61 Items, die hauptsächlich aus geschlossenen Fragen mit vorgegebenen Antwortskalen bestanden. Bei der Gestaltung des Fragebogens wurde besonderer Wert auf eine einfache und eindeutige Interpretation der Fragen und Antworten gelegt, weshalb Likert-Skalen verwendet wurden. In einigen Fällen wurden ergänzend feste Antwortkategorien oder Freitextfelder eingesetzt. Abgesehen von der Filterfrage zu Beginn des Fragebogens waren die Antworten auf die Fragen nicht verpflichtend. Teilnehmende hatten damit die Möglichkeit, Fragen unbeantwortet zu lassen oder zu überspringen. Zum Schluss konnten die Teilnehmenden ihre Gedanken zum Thema in einem Textfeld ergänzen. Dies ermöglichte den Befragten, ihre zuvor abgegebenen Antworten zu präzisieren und die Emotionen, welche sie gegenüber KI-Entscheidungshilfen empfinden, darzulegen. Die qualitativ erhobenen Daten (Texteingaben) werden in dieser Masterarbeit nicht empirisch ausgewertet, werden jedoch im Rahmen eines «qualitativen Feedbacks» im Kapitel 7 berücksichtigt.

5.3.1 Experimentelle Vignetten

In den folgenden Abschnitten werden sowohl die konzeptuellen als auch die methodologischen Überlegungen im Umgang mit den Vignetten thematisiert, gefolgt von einer Beschreibung der spezifischen Dimensionen und ihrer konkreten Ausprägungen.

Aufgrund der im Voraus durchgeführten Stichprobenanalyse und der Annahme, dass nur eine relativ kleine Menge an Umfragedaten analysiert werden kann, wurden in der vorliegenden Arbeit bewusst nur drei Dimensionen mit jeweils zwei Ausprägungen ausgewählt (2^3). Somit entstanden pro Fallbeispiel jeweils acht verschiedene Vignetten mit je zwei unterschiedlichen Ausprägungen pro Dimension, welche alle in die Untersuchung miteinbezogen wurden. Der Fokus wurde auf den Einfluss der *Invasivität* der Einschätzung, der *prädiktiven Validität* sowie der Wahrnehmung von wissenschaftlichem Wissen als unabhängige Variablen auf das Vertrauen von Sozialarbeitenden gegenüber KI-Entscheidungshilfen, als abhängige Variable, gelegt. Zur Veranschaulichung dient *Tabelle 3* mit einer komprimierten Übersicht über die Dimensionen und deren Ausprägungen:

Tabelle 3 – Verkürzte Darstellung der Dimensionen und Ausprägungen der Vignetten in dieser Thesis

Dimensionen	Ausprägungen
A = Invasivität der Einschätzung	a_0) invasivere Einschätzung a_1) weniger invasive Einschätzung
B = prädiktive Validität	a_0) treffsicherer als Sozialarbeitende a_1) gleich treffsicher wie Sozialarbeitende
C = wissenschaftliches Wissen	a_0) mit wissenschaftlichem Wissen trainiert a_1) ohne wissenschaftliches Wissen trainiert

In der Literatur wird empfohlen, den Teilnehmenden zwischen acht und zehn Vignetten pro Person vorzulegen (Auspurg et al., 2014, S. 139; Dülmer, 2014, S. 725). Angesichts der vermuteten kleineren Stichprobengrösse, der relativ ausführlichen Fallbeschreibungen und der komplexen Thematik wurde beschlossen, jedem Teilnehmenden pro Fallbeispiel lediglich eine Vignette vorzulegen. Zudem wurde entschieden, auf einen Vergleich der beiden Fallbeispiele wegen ihrer deutlichen Unterschiede zu verzichten, um mögliche Verzerrungen zu vermeiden. So ist es zwar möglich, die Aussagen der Gruppenmitglieder sowie der jeweils acht verschiedenen Gruppen miteinander zu vergleichen (*Between-Subject*), die Aussagen der einzelnen Teilnehmenden können jedoch nicht gegenübergestellt werden (*Within-Subject*).

Um Antworten zu möglichen Szenarien zu erhalten, wurden im Fragebogen zwei fiktive KI-gestützte Entscheidungsfindungstools im Rahmen von zwei unterschiedlichen Anwendungssituationen vorgestellt, wie sie zukünftig im Arbeitsalltag von Sozialarbeitenden in Kinder- und Jugendhilfestellen vorkommen könnten.

Tabelle 4 zeigt die Einleitung der beiden Fallbeispiele. Dabei handelte es sich einerseits um das Abklärungstool *KIBAT*, das die Sozialarbeitenden bei der Einschätzung einer Kindeswohlgefährdung bei einer Kindeswohlabklärung unterstützt und eine konkrete Empfehlung aufgrund des gemessenen Risikos ausgibt (*Predictive Risk Modelling*). Bei diesem Fallbeispiel wurde auf inhaltliche Informationen zum Fall verzichtet, um nicht dem Risiko ausgesetzt zu sein, nicht das Vertrauen in das Tool, sondern stattdessen die persönliche Einschätzung der Sozialarbeitenden zur ausgegebenen Empfehlung (und damit zum Fall) zu messen. Im zweiten Fallbeispiel wurde das Interventionstool *WIEW* vorgestellt, welches die Sozialarbeitenden bei der Suche nach einer geeigneten Anschlusslösung einer Jugendlichen unterstützt, bei welcher bereits verschiedene Interventionen gescheitert sind (*Predictive Chance Modelling*[13]). Bei letzterem Beispiel wurden im Gegensatz zum ersten Beispiel inhaltliche Informationen zum Fallbeispiel, jedoch keine spezifische Empfehlung genannt. Es wurde lediglich – je nach Variation – mitgeteilt, dass die Empfehlung «invasiver» oder «weniger invasiv» ausfällt, als diejenige der Sozialarbeitenden.

[13] In der aktuellen Literatur wird vor allem von *Predictive Risk Modelling* (bei Risiken), *Predictive Modelling* oder *Predictive Analytics* gesprochen. Der Begriff *Predictive Chance Modelling* oder *Prädiktive Chancenmodellierung* ist bisher noch nicht weit verbreitet, findet jedoch zunehmend Verwendung in der Forschung zur Vorhersage von Wirkungswahrscheinlichkeiten (Lätsch, 2021)

Tabelle 4 – Einleitung Fallbeispiele (Vignetten) *KIBAT* und *WIEW*

	Fallbeispiel 1 – Abklärungstool *KIBAT* *Predicitive Risk Modelling*	Fallbeispiel 2 – Interventionsplanung *WIEW* *Predicitive Chance Modelling*
Einleitung	Stellen Sie sich folgendes Szenario vor: Seit drei Monaten arbeiten Sie gemeinsam mit einer Kollegin an einer Kindeswohlabklärung in einer Familie. Es haben verschiedene Hausbesuche und Gespräche stattgefunden, zusätzlich gab es einen intensiven Schriftverkehr mit den Eltern und Fachpersonen via Brief und E-Mail. Sie haben während der gesamten Abklärung zahlreiche Aktennotizen erstellt, um alle Details festzuhalten. Nach Abschluss sämtlicher Gespräche und nach dem Sie genügend Informationen gesammelt haben, werten Sie die Abklärung gemeinsam aus. Sie kommen zum Schluss, dass eine Kindeswohlgefährdung gegeben ist. Sie erarbeiten folgende Handlungsmöglichkeit: **Eine Beistandschaft gemäss Art. 308 Abs. 1 und 2 ZGB mit einem Auftrag für eine intensive sozialpädagogische Familienbegleitung (SPF).** Zusätzlich wertet *KIBAT*, ein auf künstlicher Intelligenz basierendes Abklärungs-Tool, das vor kurzem in Ihrer Organisation implementiert wurde, sämtliche Informationen aus, welche Sie in den vergangenen Monaten gesammelt haben. Das Tool entwickelt aufgrund der vorhandenen Angaben eine unabhängige Einschätzung.	Stellen Sie sich folgendes Szenario vor: Seit sechs Jahren führen Sie eine Vormundschaft gemäss Art. 327a ZGB für die mittlerweile 16-jährige Hanna. Hannas Situation gestaltet sich sehr instabil: Sie ist psychisch labil, zeigt höchst oppositionelles Verhalten, konsumiert seit Jahren illegale Substanzen und seit kurzem auch vermehrt harte Drogen, welche sie sich vermutlich mit Beschaffungskriminalität finanziert. Die aktuelle Wohnsituation im betreuten Wohnen ist aufgrund des herausfordernden Verhaltens stark gefährdet. In den vergangenen Jahren und Monaten haben Sie und weitere involvierte (Fach-)Personen verschiedene Interventionen unternommen, die jedoch weitgehend ohne Erfolg blieben. Trotz einer gewissen Unsicherheit entwickeln Sie jetzt gemeinsam mit dem Helfersystem einen neuen Interventionsplan, von welchem Sie sich erhoffen, dass er die gewünschte Wirkung erzielt. Das KI-Tool *WIEW* soll Ihnen dabei helfen, weitere Interventionsmöglichkeiten auszuloten. *WIEW* entwickelt aufgrund der vorhandenen Fall-Informationen eine unabhängige Einschätzung sowie einen möglichen Interventionsplan für Hanna.

Die im modifizierten *TAM*-Modell behandelten Faktoren (Kapitel 4.2) bzw. die in der Hypothese H_1 formulierten Merkmale (Dimensionen) *Invasivität* (A), *prädiktive Validität* (B) und *wissenschaftliches Wissen* (C) dienten der Merkmalausprägung und wurden in beiden Fallbeispielen praktisch identisch variiert. Die Erarbeitung der Dimensionen wurde in Zusammenarbeit mit der Begleitperson und in Gesprächen mit Fachpersonen erarbeitet, da in diesem Fall nur limitiert auf Fachliteratur zurückgegriffen werden konnte. Die nachfolgende *Tabelle 5* stellt die drei Dimensionen mit ihren jeweiligen Ausprägungen und den Textmodulen dar.

Tabelle 5 – Übersicht über die Ausprägung der Vignettendimensionen und Abschluss Fallbeispiele

Fallbeispiele *KIBAT* und *WIEW*

Dimension	Ausprägung	Textmodul (in beiden Fallbeispiele identisch[14])
A) Invasivität (Intensivität der Entscheidung)	a_1 invasiver	*KIBAT*/*WIEW* errechnet, dass die von Ihnen erarbeitete Handlungsmöglichkeit mit einer Wahrscheinlichkeit von 60% zum Erfolg führt. Im Vergleich dazu empfiehlt das Tool **eine intensivere Massnahme – also ein einschneidenderes Eingreifen in die Situation, als Sie es empfehlen würden**. Für diese Handlungsoption gibt die KI eine Erfolgswahrscheinlichkeit von 90% aus.
	a_0 weniger invasiv	*KIBAT*/*WIEW* errechnet, dass die von Ihnen erarbeitete Handlungsmöglichkeit mit einer Wahrscheinlichkeit von 60% zum Erfolg führt. Im Vergleich dazu empfiehlt das Tool **eine weniger intensive Massnahme – also ein weniger einschneidenderes Eingreifen in die Situation, als Sie es empfehlen würden.** Für diese Handlungsoption gibt die KI eine Erfolgswahrscheinlichkeit von 90% aus.
B) Prädiktive Validität (Vorhersagekraft)	b_1 treffsicherer	Unabhängige Studien haben gezeigt, dass die Empfehlungen von *KIBAT*/*WIEW* **in drei Vierteln der Fälle treffsicherer** sind als diejenigen der Sozialarbeitenden.
	b_0 gleich treffsicher	Unabhängige Studien haben gezeigt, dass die Empfehlungen von *KIBAT*/*WIEW* **in der Hälfte der Fälle treffsicherer** sind als diejenigen der Sozialarbeitenden.
C) Wissenschaftliches Wissen (Grad, in dem KI auf Professionswissen referiert)	c_1 vorhanden	Das Tool verweist darauf, dass seine Einschätzung aufgrund der Analyse tausender Falldaten zustande kam. **Es verweist u.a. zusätzlich auf die Literatur von Hauri & Zingaro (2020), Kindler (2018), und Klassen (2004) / Dollinger & Raithel (2006), Kieslinger et al. (2021) und Klassen (2004), welche für die Einschätzung des Tools herangezogen wurden.** (Literaturquellen wurde am Schluss des Beispiels aufgeführt)
	c_0 nicht vorhanden	Das Tool verweist darauf, dass seine Einschätzung aufgrund der Analyse tausender Falldaten zustande kam.

	Fallbeispiel 1 – Abklärungstool *KIBAT*	Fallbeispiel 2 – Interventionsplanung *WIEW*
Abschluss	Die Entscheidung, ob Sie die Empfehlung von *KIBAT* in Ihren Abklärungsbericht aufnehmen, ist Ihnen überlassen.	Die Entscheidung, ob Sie die Empfehlung von *WIEW* in Ihre Interventionsplanung aufnehmen, ist Ihnen überlassen.

Alle Teilnehmenden erhielten in ihrem Fragebogen sowohl ein *KIBAT*- als auch ein *WIEW*-Fallbeispiel, wobei die Textmodule jeweils in gegensätzlicher Ausprägung ausgelegt waren (z. B. Ausprägung *KIBAT* $a_0b_0c_0$ und *WIEW* $a_1b_1c_1$). Dadurch wurde sichergestellt, dass alle Teilnehmenden zwei Vignetten erhielten, die in ihrer Merkmalskombination möglichst verschieden waren.

[14] Einzig die im Beispiel dargestellte Literatur (Dimension *wissenschaftliches Wissen*) wich in den Fallbeispielen ab, da die gleiche Literatur nicht für beide Beispiele passte.

Aus den insgesamt 16 Vignetten entstanden so acht Vignettengruppen, welche für eine höhere statistische Effizienz den Teilnehmenden in randomisierter Reihenfolge (Zufallsprinzip) vorgelegt wurden (Rost, 2018, S. 60; Dülmer, 2014, S. 275). Vier Vignettengruppen begannen mit dem Fallbeispiel *KIBAT* und vier mit dem Fallbeispiel *WIEW* (*Tabelle 6*):

Tabelle *6* – Übersicht der Reihenfolge und Ausprägungen der jeweilig präsentierten Vignetten pro Gruppe

Gruppe	1. Vignette	Ausprägung	2. Vignette	Ausprägung
1	*KIBAT* (1)	$a_0b_0c_0$	*WIEW* (2)	$a_1b_1c_1$
2	*KIBAT* (3)	$a_0b_0c_1$	*WIEW* (4)	$a_1b_1c_0$
3	*KIBAT* (5)	$a_0b_1c_0$	*WIEW* (6)	$a_1b_0c_1$
4	*KIBAT* (7)	$a_0b_1c_1$	*WIEW* (8)	$a_1b_0c_0$
5	*WIEW* (9)	$a_0b_1c_1$	*KIBAT* (10)	$a_1b_0c_0$
6	*WIEW* (11)	$a_0b_1c_0$	*KIBAT* (12)	$a_1b_0c_1$
7	*WIEW* (13)	$a_0b_0c_1$	*KIBAT* (14)	$a_1b_1c_0$
8	*WIEW* (15)	$a_0b_0c_0$	*KIBAT* (16)	$a_1b_1c_1$

5.3.2 Operationalisierung der Variablen

In diesem Abschnitt werden die Gütekriterien der verwendeten Messinstrumente und deren Operationalisierung im Rahmen der Studie erläutert. Ziel ist es, die methodische Strenge zu gewährleisten, um valide und zuverlässige Aussagen und Entscheidungen aus den Ergebnissen ableiten zu können (Döring & Bortz, 2016, S. 405). Dabei werden die drei Hauptgütekriterien Objektivität, Reliabilität und Validität betrachtet.

Gemäss Kritischem Rationalismus dient Wissenschaft der Wahrheitssuche. Ziel der empirischen Sozialforschung ist es, einen Bestand an bewährten und informationshaltigen Theorien aufzubauen, die die Gesetzmässigkeiten der sozialen Wirklichkeit möglichst gut erklären und weitere Forschungsaktivitäten anregen (Döring & Bortz, 2016, S. 46). Für die wissenschaftliche Theorieprüfung wird Objektivität verlangt, was bedeutet, dass die Falsifikation einer Theorie intersubjektiv nachvollziehbar anhand von Daten und Argumenten erfolgen soll, sodass das Ergebnis von den Einstellungen, Werten und Vorurteilen der Forschenden unabhängig ist. Dies wird auch als Prinzip der Wertefreiheit bezeichnet (Döring & Bortz, 2016, S. 47). Um die Objektivität zu gewährleisten, wurde der in der Studie verwendete Fragebogen mehrfach mit der Begleitperson diskutiert und alle wertenden oder suggestiven Elemente wurden versucht zu eliminieren. Die Reliabilität bezieht sich auf die Zuverlässigkeit und die Beeinflussung durch Messfehler eines Messinstruments. Döring und Bortz (2016) identifizieren verschiedene Reliabilitätskoeffizienten, die je nach Studiendesign anzuwenden sind (S. 443-443). Am häufigsten wird die Reliabilität durch die Methode der internen Konsistenz bestimmt, die zur Berechnung des Cronbach-Alpha-Koeffizienten führt. Auch

in dieser Thesis wird diese Methode verwendet, bei der alle bivariaten Korrelationen zwischen den Items[15] berechnet und anschliessend gemittelt werden (Döring und Bortz, 2016, S. 443). In Ergänzung zu den bisher genannten Gütekriterien gibt die Validität als weiteres Hauptgütekriterium an, inwieweit das tatsächlich gemessen wird, was ursprünglich beabsichtigt war. Eine hohe Reliabilität ist eine notwendige Bedingung für eine hohe Validität, jedoch keine Garantie dafür (Döring & Bortz, 2016, S. 443).

Verschiedene Items der vorliegenden Arbeit stammen aus bereits etablierten Skalen, deren Validität bereits geprüft wurde. Im Folgenden werden die einzelnen Items des Fragebogens im Detail vorgestellt. Eine Übersicht findet sich am Ende des Unterkapitels in der *Tabelle 7.*

Soziodemografische Variablen und Variablen zur Ausbildung und Berufserfahrung

Die im Fragebogen erfragten soziodemografischen Variablen sind deutlich erkennbare Merkmale, die keiner weiteren Erklärung bedürfen und als einzelne Merkmale erfasst werden können (Döring & Bortz, 2016, S. 265), sie wurden deshalb durch einzelne Items abgefragt. Beim *Kanton* standen die 26 Kantone der Schweiz zur Auswahl zur Verfügung. Die *Berufserfahrung* konnte in zweijahres-Schritten ausgewählt werden («weniger als ein Jahr», «1 bis 2 Jahre», «3 bis 4 Jahre» «5 bis 6 Jahre», «7 bis 8 Jahre», «9 bis 10 Jahre», «mehr als 10 Jahre»). Das *Alter* wurde in Vier-Jahres-Kategorien abgefragt, die in der Auswertung gemäss den Generationen nach Strasser und Lütkehaus (2020, S. 34) zusammengefasst wurden. Das *Geschlecht* konnte in drei Antwortkategorien «weiblich», «männlich» und «andere/fühle mich nicht zugehörig» angegeben werden.

Bei der Frage nach dem *Ausbildungsabschluss* konnten die Teilnehmenden ihre höchste abgeschlossene Ausbildungsstufe angeben («Berufslehre», «Gymnasiale Maturität, Berufs- oder Fachmaturität», «Höhere Fachschule HF, höhere Berufsbildung (eidg. Fachausweis, Diplom)», «Bachelor FH, PH, Uni, ETH, Diplom FH», Master FH, PH, Uni, ETH oder Lizenziat», «Doktorat, PhD», «Andere (ohne Weiterbildungsabschlüsse)» während sie bei der Frage nach der *Fachrichtung* das Fach angaben, in welchem sie ihre höchsten Abschluss erworben haben («Soziale Arbeit (inkl. Sozialarbeit, Sozialpädagogik, Soziokulturelle Animation)», «Recht», «Psychologie», «Pädagogik, Erziehungswissenschaften, Sonder-/Heilpädagogik», «Ethnologie», «Soziologie», «Medizin», «Andere Fachbereiche». Die Antwortoptionen wurden an die Studie von Gautschi (2021, S. 250) angelehnt, da sie ein ähnliches Zielpublikum ansprachen. Das *Abschlussjahr* wurde in einem offenen Textfeld eingetragen. Anschliessend wurden das Alter und die Berufserfahrung zur neuen Variable *Expert:innenstatus* kodiert (Sassmannshausen & Heupel, 2020, 179).

Die Antworten auf die folgenden Fragen wurden jeweils auf einer fünfstufigen Likert-Skala (von «ich stimme überhaupt nicht zu» bis «stimme voll und ganz zu») erfasst, mit Ausnahme der speziell gekennzeichneten Fälle.

[15] Ein Item bezeichnet dabei eine einzelne Frage oder Aussage innerhalb eines Fragebogens oder Tests, die von den Befragten beantwortet oder bewertet wird.

Digitale Affinität

Für die Messung der *digitalen Affinität* wurde die «ICT Self-Concept Scale» (Schauffel, Schmidt, Peiffer, & Ellwart, 2021) verwendet, welche Fragen zum Umgang mit digitalen Systemen umfasst. Digitale Systeme beinhalten dabei sowohl digitale Anwendungen (z. B. Software oder Apps) als auch digitale Geräte (z. B. Computer oder Smartphones). Einzelne Items dieser Skala wurden in ihrer deutschen Version übernommen.

- Ich kann digitale Systeme bedienen.
- Ich bin gut darin, digitale Systeme zu nutzen.
- Ich lerne schnell, wenn es um die Nutzung digitaler Systeme geht.
- Mir fällt es leicht, mich an neue digitale Systeme zu gewöhnen.
- Ich war schon immer gut im Umgang mit digitalen Systemen.

Um die *digitale Affinität* als einheitliches Konstrukt zu erfassen, wurden die einzelnen Items des Instruments analysiert und deren interne Konsistenz mit Cronbach's Alpha berechnet. Der Wert von 0.89 weist auf eine hohe interne Konsistenz der Skala hin.

Einstellung gegenüber Künstlicher Intelligenz

Für die Messung der *Einstellungen gegenüber KI* wurde die Skala «General Attitudes towards Artificial Intelligence» von Schepman und Rodway (2020) verwendet. Diese umfasst folgende Aspekte von Einstellungen zu Künstlicher Intelligenz:

Positive Aspekte
- Ich interessiere mich dafür, künstlich intelligente Systeme in meinem täglichen Leben zu verwenden.
- Ich bin beeindruckt, was Künstliche Intelligenz leisten kann.
- Künstliche Intelligenz ist aufregend.
- Künstlich intelligente Systeme können besser als Menschen performen.
- Ich möchte Künstliche Intelligenz in meinem eigenen Job nutzen.

Negative Aspekte
- Organisationen nutzen Künstliche Intelligenz unethisch.
- Ich denke, dass künstlich intelligente Systeme viele Fehler machen.
- Ich finde Künstliche Intelligenz unheimlich.
- Künstliche Intelligenz könnte die Kontrolle über Menschen übernehmen.

Um die allgemeine *Einstellung gegenüber KI* zu erfassen, wurden die einzelnen Items zu zwei neuen Variablen zusammengefasst: *positive Einstellung* und *negative Einstellung*. Für die positive Einstellung wurde ein Cronbach's Alpha von 0.83 erreicht, während der Wert für die negative Einstellung nach Entfernung des zweiten Items bei 0.73 lag. Die Variable *negative Einstellung* wurde in der späteren Auswertung nicht weiter berücksichtigt.

Intuition

Für die Erfassung der *Intuition* der Befragten wurde die Skala «Rational and Intuitive Decision Styles» von Hamilton, Shih und Mohammed (2016) verwendet. Vier der fünf Items dieses Instruments wurden ins Deutsche übersetzt und in den Fragebogen übernommen.

- Bei Entscheidungen verlasse ich mich hauptsächlich auf mein Bauchgefühl.
- Bei Entscheidungen im Arbeitsalltag liege ich mit meiner Intuition meistens richtig.
- In Entscheidungssituationen lege ich grösseren Wert auf das, was mein Gefühl mir sagt, als auf nüchterne Analysen.
- Ich vertraue mehr auf das, was ich denke, als auf die Einschätzung anderer.

Um die *Intuition* zu erfassen, wurden die einzelnen Items zu einer neuen Variable zusammengefasst, wobei ein Cronbach's Alpha von 0.74 erreicht wurde. Das Item wurde in der anschliessenden Auswertung jedoch nicht mehr weiter berücksichtigt.

Einstellung zu wissenschaftlichem Wissen

Für die Messung der *Einstellung zu wissenschaftlichem Wissen* wurde der Fragebogen «Zur Relevanz wissenschaftlichen Wissens in der Praxis der Sozialen Arbeit» von Brielmaier (2019) verwendet. Dieser Fragebogen besteht aus 49 Items, die in fünf Dimensionen unterteilt sind: Theorie, Empirie, Rezeption, Einstellungen und Barrieren. Für den Fragebogen in dieser Studie wurden folgende Items ausgewählt und weiterverwendet:

- Ergebnisse aus empirischen Studien sind wichtig für die tägliche Arbeit in der Sozialen Arbeit.
- Wissenschaftliches Wissen hilft mir, professionelle Entscheidungen treffen zu können.
- Erfahrungswissen ist in der Praxis der Sozialen Arbeit wichtiger als wissenschaftliches Wissen.
- Der Bezug zu wissenschaftlichem Wissen sollte in der Praxis eine grössere Rolle spielen.

Um die *Einstellung zu wissenschaftlichem Wissen* zu erfassen, wurden die entsprechenden Items zu einer Variable zusammengefasst. Cronbach's Alpha für diese Skala betrug 0.73, was auf eine zufriedenstellende interne Konsistenz hinweist. Auch dieses Item wurde in der anschliessenden Auswertung nicht mehr weiter berücksichtigt.

Vertrauen (abhängige Variable)

Das Vertrauen wurde als Einzelitem erfasst, wobei die Befragten auf einer Skala von1 bis 6 («sehr unwahrscheinlich» bis «sehr wahrscheinlich») einschätzen konnten, wie wahrscheinlich es ist, dass sie der Einschätzung des jeweiligen Tools (*KIBAT* oder *WIEW*) folgen würden. Diese Einschätzung wurde im Rahmen der Thesis als ein Indikator für das Vertrauen in das jeweilige Tool betrachtet.

- Wie wahrscheinlich ist es, dass Sie der Empfehlung von *KIBAT*/*WIEW* folgen würden?

Integrationsbereitschaft in den Arbeitsalltag

Für die Messung der *Integrationsbereitschaft* in den Arbeitsalltag konnte keine bestehenden Skalen verwendet werden. Daher wurden eigenständig zu beiden Oberthemen passende Items entwickelt.

- Ich würde die Einschätzung von *KIBAT*/*WIEW* in meine eigenen Überlegungen einbeziehen.
- Ich würde meine Einschätzungen nochmals überdenken, nachdem ich die von der *KIBAT*/*WIEW* bereitgestellten Einschätzungen gesehen habe.
- Wenn *KIBAT*/*WIEW* in meiner Organisation verfügbar wäre, würde ich es nutzen.

Die Reliabilität der Skala zur Messung der *Integrationsbereitschaft* in den Arbeitsalltag lag über alle drei Items hinweg bei 0.87, was auf eine hohe interne Konsistenz hinweist.

Erwartete Nützlichkeit

Für die Messung der *erwarteten Nützlichkeit* der beiden Tools wurde auf die Skala von Davis (1989, S. 331) zurückgegriffen. Diese in der Studie «Perceived Usefullness, Perceived Ease of Use and User Acceptance on Information Technology» (vgl. Kapitel 4.1) genutzte Einteilung beinhaltet Items zur wahrgenommenen Nützlichkeit eines entsprechenden Tools. Die im Fragebogen verwendeten Items wurden auf Deutsch übersetzt und leicht angepasst.

- Die Nutzung von *KIBAT*/*WIEW* würde die Qualität meiner Arbeit verbessern.
- *KIBAT*/*WIEW* würde wichtige Aspekte meiner Arbeit unterstützen.
- Die Verwendung von *KIBAT*/*WIEW* würde die Effektivität meiner Arbeit steigern.
- *KIBAT*/*WIEW* würde die Ausführung meiner Arbeit erleichtern.

Die Skala zur *erwarteten Nützlichkeit* mit vier Items ergab eine Reliabilität von 0.87, was auf eine hohe interne Konsistenz hindeutet. Das Item wurde in der anschliessenden Auswertung jedoch nicht mehr weiter berücksichtigt.

Verbesserter Ressourceneinsatz durch KI

Für die Messung der Variablen zum *verbesserten Ressourceneinsatz* wurde das Instrument aus der Studie von Yiu, Grant & Edgar (2007) verwendet, welche die Einflüsse der wahrgenommenen Nützlichkeit, Benutzerfreundlichkeit, wahrgenommenen Risiken und der persönlichen Innovationsfreudigkeit in der Informationstechnologie untersucht. Das Instrument wurde leicht angepasst und auf den Kontext der Sozialen Arbeit übertragen, um den verbesserten Ressourceneinsatz in diesem Berufsfeld zu messen. Dabei wurde darauf geachtet, die relevanten Elemente der ursprünglichen Skala beizubehalten, um die Validität und Zuverlässigkeit der Messung zu gewährleisten.

- Ich bin eher geneigt, Künstliche Intelligenz zu nutzen, wenn dadurch der Zeitaufwand für administrative Arbeiten signifikant reduziert wird.
- Ich bin der Meinung, dass der Einsatz von Künstlicher Intelligenz dazu beitragen kann, die Ressourcen in der Sozialen Arbeit sinnvoller einzusetzen.
- Das Potenzial von Künstlicher Intelligenz, mehr Zeit für die Interaktion mit Klient:innen zu verschaffen, indem sie Arbeitslast verringert, werte ich positiv.

Der Cronbach's Alpha-Wert für diese Skala betrug 0.86, was auf eine hohe interne Konsistenz hinweist. Das Item wurde in der anschliessenden Auswertung jedoch nicht mehr weiter berücksichtigt.

Befürchtete Arbeitsplatzbedrohung durch KI

Zur Messung der *befürchteten Arbeitsplatzbedrohung* durch KI wurde die Studie «Automation Fears: Drivers and Solutions» von Ivanov, Kuyumdzhiev und Webster (2020) herangezogen. Dieses Instrument umfasst vier Items, die verschiedene Ängste bezüglich Arbeitsplatzverlust durch Automatisierungstechnologien adressieren.

- Die Vorstellung, dass Künstliche Intelligenz Teile meiner Arbeit übernehmen könnte, erzeugt ein gewisses Unbehagen.
- Ich befürchte, dass der Einsatz von Künstlicher Intelligenz zu einer Entwertung meiner beruflichen Fähigkeiten führt.

- Ich befürchte, dass ich meinen aktuellen Arbeitsplatz aufgrund der Einführung von Künstlicher Intelligenz verlieren könnte.
- Die Annahme, dass Künstliche Intelligenz zu Einsparungen in Personalressourcen führen könnte, beunruhigt mich.

Der Cronbach's Alpha-Wert aller Items lag mit 0.71 im akzeptablen Bereich, auch wenn er sich am unteren Ende des Zulässigkeitsspektrums befand. Daraufhin wurde eine neue Variable *befürchtete Arbeitsplatzbedrohung durch KI* kodiert, die die entsprechenden Ängste zusammenfasst und auf den Kontext der vorliegenden Arbeit angepasst wurde. Auch dieses Item wurde in der anschliessenden Auswertung nicht mehr weiter berücksichtigt.

Tabelle 7 – Reliabilitätsanalyse der zu neuen Variablen zusammengefassten Items

Themenbezogene Hauptskala	Cronbach's Alpha	Anzahl Items	Quelle
Digitale Affinität	0.89	5	Schauffel, Schmidt, Peiffer, & Ellwart (2021)
Pos. Einstellung gegenüber Künstlicher Intelligenz	0.83	5	Schepman und Rodway (2020)
Neg. Einstellung gegenüber Künstlicher Intelligenz	0.73	3	Schepman und Rodway (2020)
Intuition	0.74	4	Hamilton, Shih und Mohammed (2016)
Einstellung zu wissenschaftlichem Wissen	0.73	4	Brielmaier (2019)
Integration der Tools in den Arbeitsalltag	0.87	3	Eigene Herleitung
Erwartete Nützlichkeit der Tools	0.87	4	Davis et al. (1989)
Verbesserter Ressourceneinsatz durch KI	0.86	3	Yiu, Grant & Edgar (2007)
Befürchtete Arbeitsplatzbedrohung durch KI	0.71	4	Ivanov, Kuyumdzhiev und Webster (2020)

Quelle: eigene Erhebung und Darstellung

Die Betrachtung der drei Hauptgütekriterien – Objektivität, Reliabilität und Validität – sichert die methodische Strenge und ermöglicht valide sowie zuverlässige Aussagen und Entscheidungen. Durch die Auswahl und Anpassung bestehender Skalen sowie die Entwicklung neuer Items, wo notwendig, konnte die Messgenauigkeit der relevanten Variablen gewährleistet werden. Die interne Konsistenz der Skalen weisen überwiegend hohe Cronbach's Alpha-Werte auf. Das unterstreicht die Zuverlässigkeit der Erhebungsinstrumente. Insgesamt bieten die dargelegten methodischen Ansätze eine solide Grundlage für die nachfolgenden Analysen und die Interpretation der Studienergebnisse.

5.4 Auswertungsmethode

Die Datenerhebung für die vorliegende Thesis erfolgte mithilfe der Online-Umfrageplattform «Unipark». Nach dem Export der Daten aus «Unipark» wurden diese in die Statistiksoftware «R Studio» [16] importiert, um die weitere Aufbereitung und Analyse durchzuführen.

5.4.1 Datenaufbereitung

Der erste Schritt der Datenaufbereitung bestand aus einer automatischen Bereinigung durch «Unipark» selbst, welche die Korrektur offensichtlicher Eingabefehler und die Entfernung unvollständiger Datensätze umfasste. Anschliessend wurde eine manuelle Datenbereinigung und die Zusammenführung der Datensätze durchgeführt, um den Datensatz auswertbar zu machen. Fehlende Werte wurden anschliessend systematisch mit dem Platzhalter «NA» gekennzeichnet, was im Auswertungsprogramm «R Studio» eine standardmässige Kodierung für fehlende Daten («Not Available») darstellt. Diese Kodierung ermöglichte es, fehlende Daten in der anschliessenden Analyse korrekt zu behandeln und auszuwerten.

Die Überprüfung der Datenqualität erfolgte durch statistische Tests und visuelle Methoden der Datenexploration. Eine vorläufige Analyse zeigte, dass die in dieser Arbeit verwendeten Variablen nur zum Teil normalverteilt sind, was Auswirkungen auf die spätere Testung der Hypothesen hatte.

5.4.2 Analysestrategie

Zunächst werden in einem ersten Schritt die soziodemografischen Eigenschaften der Teilnehmenden deskriptiv erfasst und analysiert. Im Fokus stehen dabei Geschlecht, Alter, Aus- und Weiterbildung sowie die Berufserfahrung. Darüber hinaus werden *die digitale Affinität und die generelle Einstellung gegenüber KI ausgewertet*. Die beiden letzteren Attribute werden zudem sowohl geschlechtsspezifisch als auch nach Altersgruppen analysiert. Aufgrund der Linksschiefe der Daten wird der Kruskal-Wallis-Test angewendet, um signifikante Unterschiede zwischen den Altersgruppen und Geschlechtern zu identifizieren.

Im Anschluss werden Korrelationen zwischen den Personenmerkmalen und dem Vertrauen in die Tools *KIBAT* und *WIEW* untersucht. Hierbei werden die Personenmerkmale *Expert:innenstatus*, *digitale Affinität* und *Einstellung gegenüber KI* betrachtet. Die Korrelationen werden mittels des Pearson-Korrelationskoeffizienten analysiert, um zu bestimmen, welche dieser Merkmale das Vertrauen in die Entscheidungshilfen beeinflussen.

Im nächsten Schritt wird die Analyse der Vignettendimensionen und deren Ausprägungen vorgenommen. Zunächst werden die Antworten des Manipulation-Checks ausgewertet. Dieser Schritt dient dazu zu überprüfen, ob die manipulativen Elemente der Vignetten – also die spezifisch variierten Bedingungen – von den Teilnehmenden

[16] Die für die Auswertung relevanten Codes wurden in R-Studio grösstenteils mit Unterstützung von ChatGPT erstellt.

wie beabsichtigt wahrgenommen wurden. Dies ist wichtig, um sicherzustellen, dass eventuelle Unterschiede in den Reaktionen der Teilnehmenden tatsächlich auf die intendierten Manipulationen und nicht auf andere Faktoren zurückzuführen sind (Huber, Meyer & Lenzen, 2014, S. 57). In der vorliegenden Arbeit wird davon ausgegangen, dass der Manipulationscheck entweder «bestanden» oder «nicht bestanden» ist. Dementsprechend werden die Antworten der Teilnehmenden, die den Manipulationscheck nicht bestanden haben (bzw. in den Fragen nach den Vignetten die Manipulation nicht wie erwünscht wahrgenommen haben), als fehlerhaft betrachtet und aus der weiteren Analyse ausgeschlossen (Huber et al., 2014, S. 62-63). Die Auswertung wird pro Dimension vorgenommen: Stimmte die Antwort der Teilnehmenden mit der erwarteten Ausprägung überein, galt der Manipulationscheck für die entsprechende Dimension als bestanden. Gaben die Teilnehmenden jedoch eine abweichende Antwort (zum Beispiel ein «stimme überhaupt nicht zu» auf die Aussage «Die Handlungsempfehlung von *KIBAT* fällt im Beispiel weniger invasiv aus, als Ihre eigene Einschätzung», obwohl die Vignette effektiv beschrieb, dass das Tool eine weniger invasive Empfehlung vorschlug, als die Person selbst wählen würde), wird diese Antwort als «ungültig» gewertet. Dadurch wird sichergestellt, dass nur die Daten von Teilnehmenden, die die Manipulation wie beabsichtigt wahrgenommen haben, in die weitere Analyse einfliessen. Durch diesen rigorosen Ansatz kann die Genauigkeit und Verlässlichkeit der Studienergebnisse gesteigert werden (Huber et al., 2014, S. 67). Gleichzeitig verliert die Studie dadurch jedoch auch an Aussagekraft. Das Risiko, dass die Eingaben der Teilnehmenden bei Unaufmerksamkeit durch Zufall entstehen, kann trotzdem nicht gänzlich ausgeschlossen werden.

Die eingegebenen Antworten der Befragten werden entsprechend ausgewertet: So wird bei der Dimension *wissenschaftliches Wissen* beispielsweise die Antwort 1 («stimme überhaupt nicht zu») und 2 («stimme eher nicht zu») sowie die Antwort 6 («ich weiss es nicht») als «korrekt» gewertet, wenn im Fallbeispiel die Ausprägung c_0 – *ohne wissenschaftliches Wissen* zugeteilt war[17]. Entsprechend werden die Antworten 3 («stimme teilweise zu»), 4 («stimme eher zu») und 5 («stimme voll und ganz zu») als ungültig erklärt und in der weiteren Analyse nicht weiter berücksichtigt. Die Analyse der Antworten zeigt, dass die Manipulation der Vignetten nur teilweise erfolgreich war: Viele Teilnehmende schienen die spezifischen Manipulationen nicht vollständig wahrgenommen zu haben.

Tabelle 8 zeigt die jeweilige Manipulationsfrage pro Dimension sowie die gültigen und ungültigen Antworten pro Ausprägung.

[17] Keine der teilnehmenden Personen wählte bei der Ausprägung c_1 (*mit wissenschaftlichem Wissen*) die Option 6 («ich weiss es nicht»), somit kann hier davon ausgegangen werden, dass keine statistischen Verzerrungen beim Ausschluss von 6 bei der Ausprägung c_1 vorliegt. Bei den anderen beiden Dimensionen bestand diese Option nicht.

Tabelle 8 – Übersicht Manipulationsfrage und gültige/ungültige Antworten

Dimension	Manipulationsfrage	Level	gültige Antworten (n_{valid})	ungültige Antworten ($n_{invalid}$)
Invasivität	«Die Handlungsempfehlung von *KIBAT*/*WIEW* fällt im Beispiel weniger intensiv aus, als Ihre eigene Einschätzung.»	a_1 invasiver	1, 2	3, 4, 5
		a_0 weniger invasiv	3, 4, 5	1, 2
Prädiktive Validität	«Die Vorhersagen von *KIBAT*/*WIEW* sind im Beispiel gemäss Studien treffsicherer als die von Sozialarbeitenden.»	b_1 treffsicherer	3, 4, 5	1, 2
		b_0 gleich treffsicher	1, 2	3, 4, 5
Wissenschaftliches Wissen	«*KIBAT*/*WIEW* wurde im Beispiel mit Literatur aus der Sozialen Arbeit trainiert.»	c_1 vorhanden	3, 4, 5	1, 2, 6
		c_0 nicht vorhanden	1, 2, 6	3, 4, 5

1 = stimme überhaupt nicht zu, 2 = stimme eher nicht zu, 3 = stimme teilweise zu, 4 = stimme eher zu, 5 = stimme voll und ganz zu, 6 = ich weiss es nicht

Da die Variablen, die die Ausprägung der Manipulation abbilden, ursprünglich einen Wert zwischen 1 und 5 bzw. 1 und 6 enthalten und damit nicht dichotom[18] sind, werden sie für die weitere Auswertung in Dummy-Variablen umkodiert. Die Dummy-Variablen beschreiben die Ausprägung der Manipulation der Vignette, wobei der Wert 1 z. B. in der Variable *Invasivität* für die «invasivere» Manipulation und der Wert 0 für die «weniger invasive» Manipulation steht. Diese Kodierung ermöglicht es, den Einfluss des Manipulationsgrades auf die abhängige Variable in statistischen Modellen zu analysieren (Grotenhuis & Thijs, 2015, S. 3).

Bei der Analyse der Vignetten werden für einen ersten deskriptiven Einblick in die Daten Mittelwerte miteinander verglichen (Rost, 2018, S. 94). Mit Mittelwertvergleichen alleine sind jedoch noch keine gemeinsamen Effekte der einzelnen Dimensionen auf die unabhängige Variabel zu erkennen. Hierbei kommt die Methode der multiplen Regression zum Einsatz, die es ermöglicht, die Beziehungen zwischen jedem Faktor und der von den Teilnehmenden getroffenen Einschätzung sowohl zu beschreiben als auch zu messen (Taylor, 2006, S. 1196; Rost, 2018, S. 95). Sie ist in der Regel die leistungsfähigste Analysemethode in diesem Design und liefert Masse für die Stärke des Einflusses jedes Faktors auf die Einschätzung. Als besonders nützlich wird auch das Mass der Effektstärke gesehen, das angibt, wie stark ein bestimmter Faktor die Entscheidung beeinflusst, gemessen am Anteil der erklärten Varianz (R^2) (Taylor, 2006, S. 1197).

Die multiple Regressionsanalyse ermöglicht es zu bestimmen, wie genau die Werte einer Zielvariable mithilfe einer Gruppe von unabhängigen Variablen vorhergesagt werden können. In diesem Analyseverfahren wird untersucht, wie stark und in welcher Richtung jede unabhängige Variable – unter Berücksichtigung der anderen einbezogenen Variablen – mit der abhängigen Variable zusammenhängt. Sowohl der gesamte Zusammenhang, als auch die spezifischen Einflüsse der einzelnen unabhängigen Variablen, ausgedrückt durch Beta-Gewichte (b), können auf ihre statistische Signifikanz hin überprüft werden (Rost, 2018, S. 95).

[18] Dichotom bedeutet, dass eine Variable nur zwei mögliche Ausprägungen hat, wie beispielsweise 0 und 1 oder «Ja» und «Nein» (Baur & Blasius, 2014, S. 606).

In der folgenden Untersuchung werden zwei multiple Regressionsanalysen durchgeführt, wobei für jedes Fallbeispiel eine separate Analyse erfolgt. Die beiden multiplen Regressionsanalysen dienen der Überprüfung der aufgestellten Hypothesen H_1 bis H_4 (inkl. Subhypothesen $H_{1.1}$, $H_{1.2}$ und $H_{1.3}$) in den jeweiligen Tools. Dabei werden zuerst die drei Vignettendimensionen *Invasivität*, *prädiktive Validität* und *wissenschaftliches Wissen* als unabhängige Variablen in die Analyse einbezogen. In weiteren Modellen werden zusätzlich die Personenmerkmale *Expert:innenstatus*, *digitale Affinität* und *Einstellung gegenüber KI* berücksichtigt. Diese schrittweise Einbeziehung der Variablen ermöglicht es, die individuellen und kombinierten Einflüsse auf das Vertrauen in die Tools zu untersuchen und die Hypothesen zu überprüfen.

Zum Schluss wird zur Prüfung der Hypothese H_5 das Vertrauen der Teilnehmenden hinsichtlich ihrer *Integrationsbereitschaft* gegenüber KI-gestützten Entscheidungshilfen in den Arbeitsalltag analysiert. Um den Zusammenhang zwischen Vertrauen und Integrationsbereitschaft zu untersuchen wird eine Korrelationsanalyse mittels des Spearman-Korrelationskoeffizienten durchgeführt (Zöfel, 2003, S. 149).

Die gesamte statistische Bewertung erfolgt auf der Basis eines Signifikanzniveaus von 5%.

6 Ergebnisse

Im folgenden Kapitel werden die mit dem Fragebogen erhobenen Daten ausgewertet, dargestellt und die Hypothesen H_1 bis H_5 geprüft. Zunächst erfolgt eine deskriptive Auswertung, bei der die demografischen Merkmale der Teilnehmenden sowie deren *digitale Affinität* und *Einstellung gegenüber KI* beschrieben werden. Im Anschluss daran wird die Beurteilung der Vignetten und experimentellen Manipulationen vorgenommen, um die Gültigkeit und Wirksamkeit der eingesetzten Instrumente zu überprüfen. Anschliessend werden in den Abschnitten zu den Korrelationen und Regressionsanalysen die Zusammenhänge zwischen Personenmerkmalen und dem Vertrauen in die KI-gestützten Entscheidungshilfen *KIBAT* und *WIEW* untersucht. Hierbei wird analysiert, welche Faktoren das Vertrauen der Teilnehmenden in diese Tools beeinflussen. Schliesslich wird die Integrationsbereitschaft der Teilnehmenden hinsichtlich der Nutzung entsprechender Tools betrachtet, bevor das Kapitel mit einer zusammenfassenden Darstellung der wesentlichen Erkenntnisse abgeschlossen wird.

Die Ergebnisse dieser Analysen und die Überprüfung der aufgestellten Hypothesen werden im darauffolgenden Kapitel 7 eingehend interpretiert und diskutiert, um Aussagen über die Generalisierbarkeit der Vignettenbefragung zu treffen.

6.1 Charakterisierung der Stichprobe

In diesem Kapitel erfolgt zunächst eine deskriptive Analyse der demografischen Merkmale der Teilnehmenden. Hierzu zählen unter anderem das Alter und das Geschlecht sowie das Bildungsniveau und die Berufserfahrung der befragten Beistandspersonen. Ebenso wird die *digitale Affinität* der Teilnehmenden sowie ihre *Einstellung gegenüber KI* nach Altersgruppe und Geschlecht aufgezeigt.

Die soziodemografische Zusammensetzung der Stichprobe umfasst 204 Teilnehmende ohne fehlende Daten (*Tabelle 9).* Die Mehrheit der Teilnehmenden (55.9%) befand sich in der Altersgruppe von 30 bis 44 Jahren, gefolgt von den 45 bis 59-Jährigen (26.5%). Frauen waren mit 70.1% deutlich häufiger vertreten als Männer (29.4%). Eine Person (0.5%) gab ein anderes Geschlecht an.

Tabelle 9 – Demografische Merkmale

	% (*n*)	n_{compl}	fehlend (n_{miss})
Altersgruppe		204	0 (0)
< 30 Jahre	11.3 (23)		
30 - 44 Jahre	55.9 (114)		
45 - 59 Jahre	26.5 (54)		
60 - 64 Jahre	5.4 (11)		
> 65 Jahre	1 (2)		
Geschlecht		204	0 (0)
weiblich	70.1 (143)		
männlich	29.4 (60)		
andere	0.5 (1)		

Quelle: eigene Erhebung und Darstellung

Die Stichprobe zeigt in Bezug auf Ausbildung und Berufserfahrung eine hohe Qualifikation der Teilnehmenden (*Tabelle 10*). Die überwiegende Mehrheit, 71.6%, besitzt einen Bachelorabschluss von einer Fachhochschule oder Universität, weitere 19.6% haben einen Masterabschluss. In Bezug auf die Fachrichtung dominiert Soziale Arbeit als Fach des höchsten Abschlusses. Dies umfasst Bereiche wie Sozialarbeit, Sozialpädagogik und soziokulturelle Animation. Andere Fachrichtungen wie Recht, Psychologie, Pädagogik, Erziehungswissenschaften, Sonder- und Heilpädagogik sowie Soziologie sind jeweils nur von einem sehr kleinen Teil der Befragten vertreten.

Hinsichtlich der Berufserfahrung zeigt sich, dass ein beachtlicher Anteil der Stichprobe (25.5%) über mehr als zehn Jahre Berufserfahrung verfügt. Die nächstgrössere Gruppe sind Teilnehmende mit drei bis vier Jahren Erfahrung (21.1%), gefolgt von jenen mit ein bis zwei Jahren (17.6%). Die übrigen Erfahrungszeiträume sind weniger stark vertreten.

Tabelle 10 – Ausbildung und Berufserfahrung

	% (n)	n_{compl}	fehlend (n_{miss})
Höchster Ausbildungsabschluss		204	0 (0)
Berufslehre	0.5% (1)		
Gymnasiale Maturität, Berufs- oder Fachmaturität	1% (2)		
Höhere Fachschule HF, höhere Berufsbildung	6.4% (13)		
Bachelor FH/Uni (oder Äquivalent)	71.6% (146)		
Master FH/Uni (oder Äquivalent)	19.6% (40)		
Doktorat	1% (2)		
Fachrichtung (höchstes Abschlussfach)		202	1 (2)
Soziale Arbeit (inkl. Sozialarbeit, Sozialpädagogik, Soziokulturelle Animation)	86.9% (177)		
Recht	2.5% (5)		
Psychologie	2.9% (6)		
Pädagogik, Erziehungswissenschaften, Sonder-/Heilpädagogik	2% (4)		
Soziologie	1.5% (3)		
andere	3.4% (7)		
Berufserfahrung		203	0.5 (1)
< 1 Jahr	4.9% (10)		
1 bis 2 Jahre	17.6% (36)		
3 bis 4 Jahre	21.1% (43)		
5 bis 6 Jahre	13.2% (27)		
7 bis 8 Jahre	9.8% (20)		
9 bis 10 Jahre	7.4% (15)		
> 10 Jahre	25.5% (52)		

Quelle: eigene Erhebung und Darstellung

6.1.1 Digitale Affinität und Einstellung zu KI nach Altersgruppe und Geschlecht

Tabelle 11 stellt die deskriptive Auswertung der *digitalen Affinität* sowie der *Einstellung gegenüber KI* dar. Die Ergebnisse sind dabei sowohl geschlechtsspezifisch[19] als auch nach Altersgruppen dargestellt. Die einzelnen Items der zu den Variablen gehörenden Skalen werden jeweils zu einer neuen Variablen zusammengefasst (vgl. Kapitel 5.3.2). Für jede Gruppe sind die Mittelwerte (M) und Standardabweichungen (*SD*) angegeben, um die zentralen Tendenzen und die Streuung der Werte zu veranschaulichen.

Die Fragen zur *digitalen Affinität* zielten darauf ab, wie vertraut die Befragten im Umgang mit digitalen Systemen sind. Die Daten zeigen, dass männliche Teilnehmer über alle Altersgruppen hinweg eine höhere digitale Affinität aufweisen als weibliche Teilnehmerinnen. Die höchste durchschnittliche Affinität zeigen Männer die jünger als 30 sind (M = 5, *SD* = 0) sowie Männer im Alter von 30-44 Jahren (M = 4.44, *SD* = 0.60). Bei den Frauen ist die digitale Affinität in der jüngsten Altersgruppe am höchsten (M = 4.17, *SD* = 0.65) und nimmt mit steigendem Alter ab.

Ähnlich verhält es sich mit der *Einstellung gegenüber KI*. Männer in der jüngsten Altersgruppe bewerteten ihre Faszination und Bereitschaft zur Nutzung von KI am höchsten (M = 3.93, *SD* = 0.95), dicht gefolgt von den Männern in der Altersgruppe 60-64 (M = 3.92, *SD* 0.83). Auch hier zeigen Männer in allen Altersgruppen eine mehrheitlich positivere Einstellung gegenüber KI im Vergleich zu den Frauen.

Die Ergebnisse des Kruskal-Wallis-Tests zeigen signifikante Unterschiede sowohl in der *digitalen Affinität* ($p = 0.012$) als auch in der *Einstellung gegenüber KI* ($p < 0.001$) zwischen den verschiedenen Altersgruppen und Geschlechtern.

Tabelle *11* – Kreuztabelle *Digitale Affinität* und *Einstellung gegenüber KI* nach Altersgruppe und Geschlecht

			Digitale Affinität		Einstellung gegenüber KI	
	weiblich n	männlich n	weiblich M (*SD*)	männlich M (*SD*)	weiblich M (*SD*)	männlich M (*SD*)
Altersgruppe	143 (70.4%)	60 (29.6%)				
< 30 Jahre	20 (9.85%)	3 (1.48%)	4.17 (0.65)	5 (0)	3.46 (0.68)	3.93 (0.95)
30 - 44 Jahre	82 (40.39%)	32 (15.76%)	4.16 (0.68)	4.44 (0.60)	3.21 (0.78)	3.70 (0.68)
45 - 59 Jahre	34 (16.75%)	19 (9.36%)	3.94 (0.70	4.26 (0.62)	2.99 (0.58)	3.22 (0.87)
60 - 64 Jahre	6 (2.96%)	5 (2.46%)	3.87 (0.7)	4.04 (0.26)	2.63 (0.98)	3.92 (0.83)
> 65 Jahre	1 (0.49%)	1 (0.49%)	3.2 (0)	2 (0)	3.6 (0)	2.8 (0)
Kruskal-Wallis-Test			p = 0.012		p = <0.001	

Quelle: eigene Erhebung und Darstellung

[19] Zur Vereinfachung der Auswertung wird die eine Person, die als Geschlecht «andere» angegeben hat, in den folgenden Analysen nicht berücksichtigt.

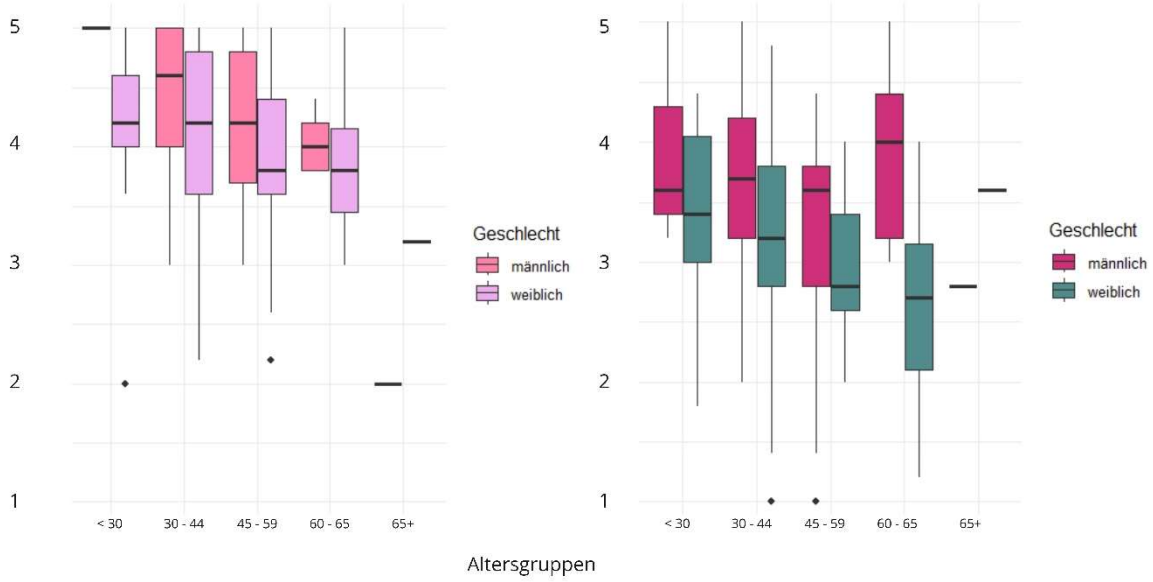
Zur weiteren Veranschaulichung der Ergebnisse werden Boxplots erstellt, die die Verteilung der *digitalen Affinität* und der *Einstellung gegenüber KI* nach Altersgruppen zeigen (*Abbildung 6*). Diese grafische Darstellung bietet einen Einblick in die Variabilität und die zentralen Tendenzen innerhalb der Altersgruppen.

Der linke Boxplot zeigt die Verteilung der *digitalen Affinität* nach Altersgruppen. Es wird deutlich, dass die jüngeren Altersgruppen (jünger als 30 Jahre und 30 bis 44 Jahre) im Median höhere Werte aufweisen. Die digitale Affinität nimmt mit zunehmendem Alter leicht ab, wobei die Gruppe der über 65-Jährigen die niedrigsten Werte aufweist.

Der rechte Boxplot zeigt die Verteilung der *Einstellung gegenüber KI* nach Altersgruppen. Auch hier ist ein ähnliches Muster zu erkennen: Jüngere Altersgruppen haben im Median eine positivere Einstellung gegenüber KI, die mit zunehmendem Alter abnimmt. Die Variabilität innerhalb der Altersgruppen zeigt jedoch, dass es auch innerhalb der Gruppen unterschiedliche Einstellungen gibt. Die höchste positive Einstellung gegenüber KI findet sich bei den unter 30-Jährigen und den 60 bis 64-Jährigen.

Abbildung 6 – Boxplot *Digitale Affinität* und *Einstellung gegenüber KI* nach Altersgruppen

Quelle: Eigene Erhebung und Darstellung



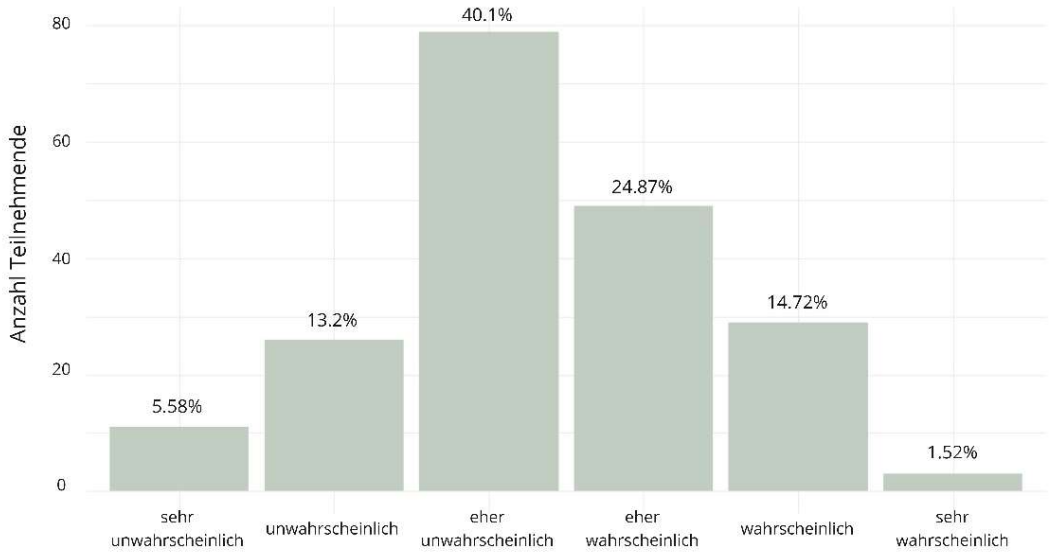
6.2 Abhängige Variabel: Vertrauen in KI-gestützte Entscheidungshilfen

In diesem Kapitel wird eine deskriptive Analyse der im Fragebogen ermittelten Wahrscheinlichkeit vorgenommen, mit welcher die Teilnehmenden der Empfehlung des jeweiligen Tools (*KIBAT* oder *WIEW*) folgen würden. Diese Einschätzung wird im Rahmen dieser Thesis als Indikator für das Vertrauen in das jeweilige Tool betrachtet. Die Auswertung erfolgt zunächst pro Tool.

Abklärungstool KIBAT

Das Säulendiagramm (*Abbildung 7*) zeigt die Verteilung der Wahrscheinlichkeiten, mit denen die Teilnehmenden angeben, der Empfehlung von *KIBAT* zu folgen. Die Antworten wurden in fünf Kategorien eingeteilt: von «sehr unwahrscheinlich» bis «sehr wahrscheinlich». Die meisten Teilnehmenden (40.1%) stuften die Wahrscheinlichkeit, der Empfehlung von *KIBAT* zu folgen, als «eher unwahrscheinlich» ein, für 13.2% bzw. 5.58% der Teilnehmenden ist es unwahrscheinlich oder sogar sehr unwahrscheinlich, dass sie der Einschätzung folgen würden. Gut ein Viertel der Teilnehmenden (24.87%) würde der Einschätzung von *KIBAT* «eher wahrscheinlich» folgen, jedoch nur für einen sehr kleinen Teil (1.52%) der Teilnehmenden wäre es «sehr wahrscheinlich, sich auf das Urteil von *KIBAT* zu verlassen.

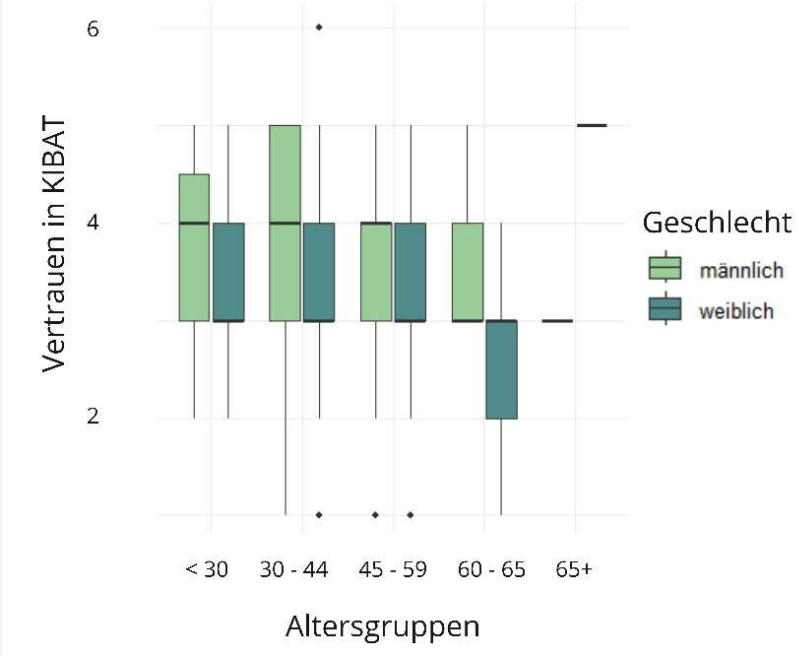
Abbildung 7 – Säulendiagramm zum Vertrauen in *KIBAT* (n = 196)
Quelle: Eigene Erhebung und Darstellung



Um die Verteilung der Antworten genauer zu betrachten, zeigt die *Abbildung 8* ein Boxplot der Verteilung des Vertrauens in *KIBAT* nach Altersgruppen und Geschlecht. Es wird deutlich, dass das Vertrauen in *KIBAT* zwischen den Altersgruppen variiert. Jüngere Teilnehmende (< 30 Jahre) und die Teilnehmenden im Alter von 30-44 Jahren zeigen ein höheres Vertrauen in die Empfehlungen des Tools. Im Gegensatz dazu ist bei den älteren Teilnehmenden (60-64 Jahre) ein deutlicher Rückgang des Vertrauens zu verzeichnen. Auch innerhalb der Altersgruppen sind Unterschiede zwischen den Geschlechtern erkennbar: Männliche Teilnehmende weisen generell ein höheres Vertrauen in *KIBAT* auf als weibliche Teilnehmende.

Abbildung 8 – Boxplot Vertrauen in *KIBAT* nach Altersgruppe und Geschlecht

Quelle: Eigene Erhebung und Darstellung



Bei näherer Betrachtung der Mittelwerte (*Tabelle 12*) zeigt sich, dass das Vertrauen in *KIBAT* sowohl alters- als auch geschlechtsspezifische Unterschiede aufweist. Insgesamt haben 138 weibliche (70.4%) und 58 männliche (29.6%) Teilnehmende an der Frage zum Vertrauen teilgenommen. In der Altersgruppe der unter 30-Jährigen liegt der Mittelwert des Vertrauens bei den weiblichen Teilnehmenden bei 3.4 (*SD* = 0.94), während er bei männlichen Teilnehmenden leicht höher bei 3.667 (*SD* = 1.53) liegt. Bei den 30 bis 44-Jährigen ist das Vertrauen bei weiblichen Teilnehmenden mit einem Mittelwert von 3.222 (*SD* = 1.11) ebenfalls etwas niedriger als bei den männlichen Teilnehmenden, die einen Mittelwert von 3.733 (*SD* = 1.17) aufweisen. In der Altersgruppe der 45 bis 59-Jährigen beträgt der Mittelwert bei weiblichen Teilnehmenden 3.161 (SD = 1.00) und bei männlichen Teilnehmenden 3.474 (*SD* = 1.17). Bei den 60 bis 64-Jährigen zeigt sich ein stärkerer Unterschied: Der Mittelwert des Vertrauens liegt bei weiblichen Teilnehmenden bei 2.6 (*SD* = 1.14), während männliche Teilnehmende einen Mittelwert von 3.6 (*SD* = 0.89) aufweisen. In der Altersgruppe über 65 Jahren ist aufgrund der geringen Anzahl an Teilnehmenden die Aussagekraft eingeschränkt.

Die ANOVA-Analyse zeigt, dass das Geschlecht einen signifikanten Einfluss auf das Vertrauen in *KIBAT* hat (p = 0.014), während die Altersgruppen und die Interaktion zwischen Altersgruppen und Geschlecht keinen signifikanten Einfluss zeigen.

Tabelle 12 - Mittelwertvergleiche des Vertrauens in *KIBAT*

Vertrauen in *KIBAT*

	weiblich % (n)	männlich % (n)	weiblich M (*SD*)	männlich M (*SD*)
Altersgruppe	70.4% (138)	29.6% (58)		
< 30 Jahre	10.2% (20)	1.53% (3)	3.4 (0.94)	3.667 (1.53)
30 - 44 Jahre	41.33% (81)	15.31% (30)	3.222 (1.11)	3.733 (1.17)
45 - 59 Jahre	15.82% (31)	9.69% (19)	3.161 (1)	3.474 (1.17)
60 - 64 Jahre	2.55% (5)	2.55% (5)	2.6 (1.14)	3.6 (0.89)
> 65 Jahre	0.51% (1)	0.51% (1)	5 (0)	3 (0)

ANOVA

Faktor	*df*	Summe der Quadrate	Mittlere Quadrate	F-Wert	*p*-Wert
Altersgruppen	4	1.88	0.470	0.389	0.816
Geschlecht	1	7.41	7.414	6.138	0.014*
Altersgruppen (Interaktion)	4	4.14	1.035	0.857	0.491
Residuen	186	222.66	1.208		

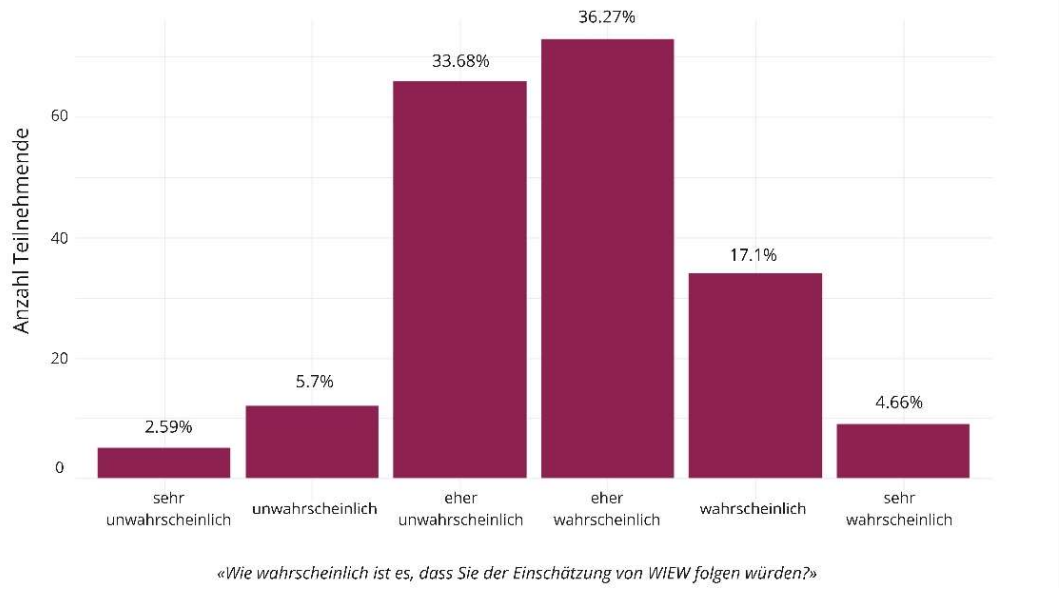
Quelle: eigene Erhebung und Darstellung

Abklärungstool WIEW

Das Säulendiagramm (*Abbildung 9*) zeigt die Verteilung der Wahrscheinlichkeiten, mit denen die Teilnehmenden angeben, der Einschätzung des zweiten Tools *WIEW* zu folgen. Die Antworten wurden auch hier in fünf Kategorien eingeteilt: von «sehr unwahrscheinlich» bis «sehr wahrscheinlich». Die meisten Teilnehmenden (36.27%) stuften die Wahrscheinlichkeit, der Empfehlung von *WIEW* zu folgen, als «eher wahrscheinlich» ein. 33.68% der Teilnehmenden gaben an, dass es «eher unwahrscheinlich» sei, dass sie der Einschätzung von *WIEW* folgen würden. Für 5.7% bzw. 2.59% der Teilnehmenden ist es «unwahrscheinlich» oder sogar «sehr unwahrscheinlich», dass sie der Einschätzung folgen würden. 17.1% der Teilnehmenden würden der Einschätzung von *WIEW* «wahrscheinlich» folgen. Ein kleinerer Teil (4.66%) der Teilnehmenden wäre sogar bereit, sich «sehr wahrscheinlich» auf die Empfehlung von *WIEW* zu verlassen.

Abbildung 9 – Säulendiagramm zum Vertrauen in *WIEW* (n = 198)

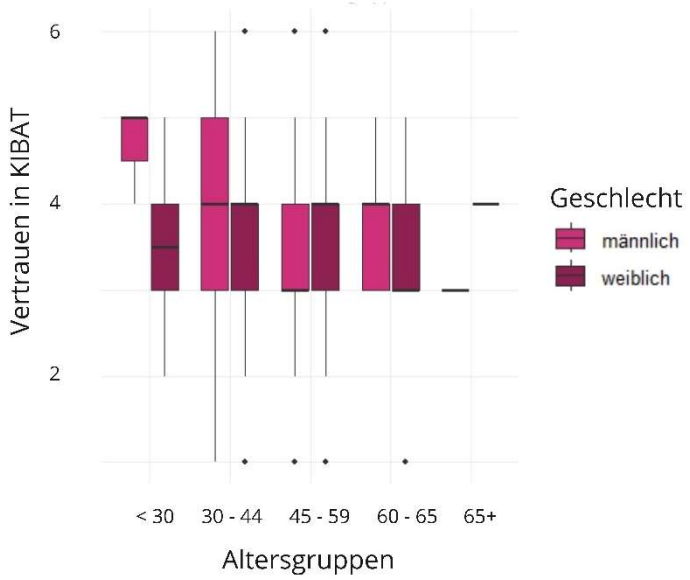
Quelle: Eigene Erhebung und Darstellung



Um die Verteilung der Antworten genauer zu betrachten, zeigt Abbildung 10 ein Boxplot der Vertrauensverteilung in *WIEW* nach Altersgruppen und Geschlecht. Jüngere Teilnehmende (< 30 Jahre) und die Altersgruppe 30 bis 44 Jahre vertrauen dem Tool stärker, während das Vertrauen bei den 60 bis 64-Jährigen deutlich abnimmt. Männer haben insgesamt mehr Vertrauen in *WIEW* als Frauen.

Abbildung 10 – Boxplot Vertrauen in *WIEW* nach Altersgruppe und Geschlecht

Quelle: Eigene Erhebung und Darstellung



Die Analyse der Mittelwerte in *Tabelle 13* zeigt alters- und geschlechtsspezifische Unterschiede im Vertrauen in *WIEW*. Bei den unter 30-Jährigen liegt der Mittelwert des Vertrauens bei weiblichen Teilnehmenden bei 3.55 (SD = 0.89), während er bei männlichen Teilnehmenden bei 4.667 (SD = 0.58) liegt. In der Altersgruppe der 30 bis 44-Jährigen beträgt der Mittelwert bei weiblichen Teilnehmenden 3.679 (SD = 1.01) und bei männlichen Teilnehmenden 3.900 (SD = 1.09).Bei den 45 bis 59-Jährigen beträgt der Mittelwert des Vertrauens bei weiblichen Teilnehmenden 3.843 (SD = 1.02) und bei männlichen Teilnehmenden 3.684 (SD = 1.34). In der Altersgruppe der 60 bis 64-Jährigen zeigt sich ein Unterschied: Weibliche Teilnehmende haben einen Mittelwert von 3.333 (SD = 1.36), während männliche Teilnehmende einen deutlich höheren Mittelwert von 3.8 (SD = 0.84) aufweisen.

Die ANOVA-Analyse zeigt, dass weder Geschlecht noch Altersgruppen oder deren Interaktion einen signifikanten Einfluss auf das Vertrauen in *WIEW* haben (*p*-Werte: Geschlecht 0.319, Altersgruppen 0.963, Interaktion 0.388).

Tabelle 13 – Mittelwertvergleiche des Vertrauens in *WIEW*

Vertrauen in *WIEW*

	weiblich n	männlich n	weiblich M (*SD*)	männlich M (*SD*)
Altersgruppe	70.7% (140)	29.3% (58)		
< 30 Jahre	10.1% (20)	1.52% (3)	3.55 (0.887)	4.667 (0.577)
30 - 44 Jahre	40.91% (81)	15.15% (30)	3.679 (1.01)	3.9 (1.09)
45 - 59 Jahre	16.16% (3)	9.6% (19)	3.843 (1.02)	3.684 (1.34)
60 - 64 Jahre	3.03% (6)	2.53% (5)	3.333 (1.36)	3.8 (0.84)
> 65 Jahre	0.51% (1)	0.51% (1)	4 (0)	3 (0)

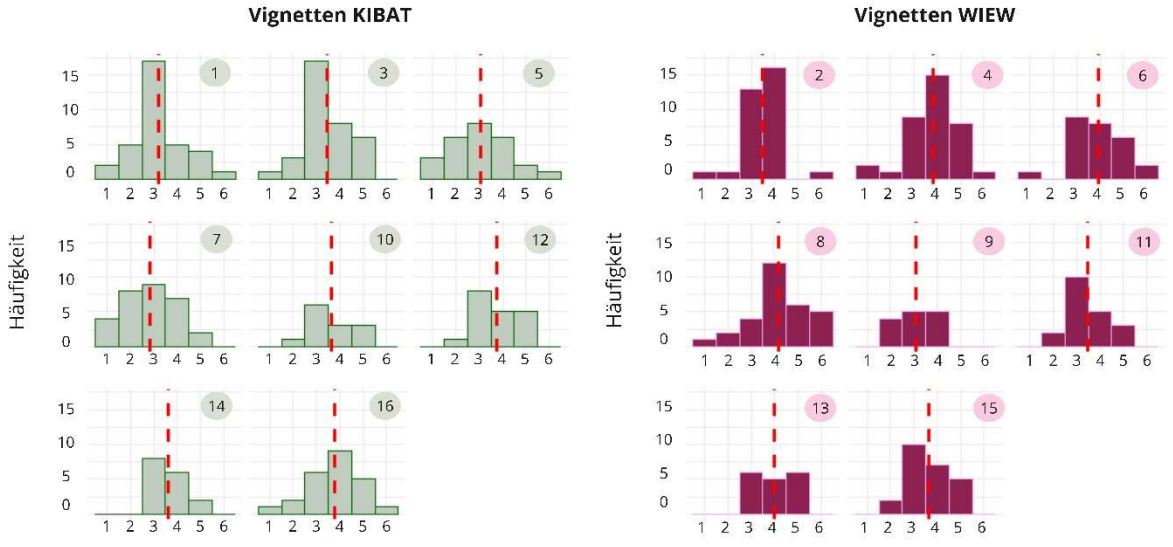
ANOVA

Faktor	*df*	Summe der Quadrate	Mittlere Quadrate	F-Wert	*p*-Wert
Altersgruppen	4	0.67	0.166	0.150	0.963
Geschlecht	1	1.11	1.106	0.997	0.319
Altersgruppen (Interaktion)	4	4.61	1.153	1.040	0.388
Residuen	188	208.43	1.109		

Quelle: eigene Erhebung und Darstellung

Die Werte des Vertrauens wurden bis jetzt im Generellen (also in Bezug auf sämtliche Antworten) aufgezeigt. Um einen Einblick in die Variabilität der Antworten innerhalb der einzelnen Vignetten zu ermöglichen, werden Histogramme für die insgesamt 16 verschiedenen Vignetten erstellt, jeweils acht für *KIBAT* und acht für *WIEW* (vgl. Kapitel 5.3.1). Die Histogramme in *Abbildung 11* visualisieren die Verteilung der Antworten über die verschiedenen Vignetten hinweg und veranschaulichen, wie konsistent oder variabel die Einschätzungen der Teilnehmenden in den unterschiedlichen Szenarien sind. Dabei ist zu beachten, dass die Anzahl der Personen, die die Vignetten beantwortet haben, aufgrund der Randomisierung deutlich variiert. Aufgrund dieser Unregelmässigkeit wurde darauf verzichtet, die einzelnen Vignetten im Detail auszuwerten. Die folgende Darstellung dient dazu, ein vertieftes Verständnis der allgemeinen Trends und Muster zu vermitteln.

Abbildung 11 – Histogramme des Vertrauens der beiden Fallbeispiele *KIBAT* und *WIEW* nach Vignette.
Quelle: Eigene Erhebung und Darstellung



Das Histogramm für *KIBAT* zeigt eine überwiegend breite Verteilung der Bewertungen, die sich hauptsächlich im mittleren bis unteren Bereich der Skala ansiedeln (Mittelwert zwischen 2.83 und 3.75). Die meisten Mittelwerte liegen zwischen 3 und 4, was «eher unwahrscheinlich» bis «eher wahrscheinlich» entspricht, dass die Teilnehmenden den Empfehlungen von *KIBAT* folgen würden. Nur ein Wert liegt im Mittel unter 3 (Vignette 7). Im Kontrast dazu zeigt das Histogramm für *WIEW* eine Konzentration leicht höherer Bewertungen (Mittelwert zwischen 3.07 und 4.17).

6.2.1 Personenmerkmale und Vertrauen

In *Tabelle 14* wird die Korrelation der Variablen *Expert:innenstatus*, *digitale Affinität* und *Einstellung gegenüber KI* mit dem Vertrauen in die beiden Tools *KIBAT* und *WIEW* dargestellt. Die Korrelationen der Personenmerkmale in beide Tools werden zur besseren Übersicht in einer Tabelle dargestellt, werden jedoch nicht miteinander verglichen.

Für den *Expert:innenstatus* (Kombination aus Lebensalter und Berufserfahrung) wird bei beiden Tools keine signifikante Korrelation zum Vertrauen festgestellt. Die Korrelation bei *KIBAT* ist negativ, jedoch sehr schwach (r = -0.057, p = 0.429, t-Wert = -0.792). Auch bei *WIEW* zeigt sich eine negative, aber nicht signifikante Korrelation (r = -0.034, p = 0.630, t-Wert = -0.483). Bei der *digitalen Affinität* zeigt sich ebenfalls keine signifikante Korrelation mit dem Vertrauen in die beiden Tools. Für *KIBAT* beträgt der Korrelationskoeffizient r = 0.099 (p = 0.166, t-Wert = 1.391), was auf eine sehr schwache positive Korrelation hindeutet. Bei *WIEW* ist die Korrelation ebenfalls schwach positiv (r = 0.084, p = 0.238, t-Wert = 1.183). Die *Einstellung gegenüber KI* zeigt hingegen eine signifikante positive Korrelation mit dem Vertrauen in beide Tools. Für *KIBAT* beträgt der Korrelationskoeffizient r = 0.293 (p < 0.001, t-Wert = 4.282), was auf eine moderate positive Korrelation hinweist. Noch stärker ist die Korrelation bei *WIEW* mit r = 0.362 (p < 0.001, t-Wert = 5.445).

Die durchgeführte Korrelationsanalyse zeigt, dass der *Expert:innenstatus* und die *digitale Affinität* keine signifikanten Prädiktoren für das Vertrauen in die Tools *KIBAT* und *WIEW* sind. Im Gegensatz dazu hat das Merkmal *Einstellung gegenüber KI* einen signifikanten positiven Einfluss auf das Vertrauen in beide Tools.

Tabelle 14 – Korrelationsanalyse der Personenmerkmale und dem Vertrauen in die beiden Tools

	KIBAT				*WIEW*			
Unabhängige Variablen	n	r	t-Wert	p	n	r	t-Wert	p
Expert:innenstatus	197	-0.057	-0.792	0.429	199	-0.034	-0.483	0.630
Digitale Affinität	197	0.099	1.391	0.166	199	0.084	1.183	0.238
Interesse an KI	197	0.293	4.282	< 0.001	199	0.362	5.445	< 0.001

Quelle: eigene Erhebung und Darstellung

6.3 Beurteilung der Vignetten und experimentelles Design

In diesem Kapitel wird die Validität der experimentellen Manipulationen und die Wahrnehmung der Vignetten durch die Teilnehmenden in einem ersten Schritt überprüft. Zunächst erfolgt der Manipulations-Check der Dimensionen *Invasivität*, *prädiktive Validität* und *wissenschaftliches Wissen*. Danach werden Mittelwertvergleiche durchgeführt, um den Einfluss dieser Dimensionen auf das Vertrauen in die Tools *KIBAT* und *WIEW* deskriptiv darzustellen.

6.3.1 Manipulations-Check der Dimensionen

Die *Tabelle 15* und *Tabelle 16* zeigen eine Übersicht über die Ergebnisse der Manipulations-Checks der beiden Fallbeispiele *KIBAT* und *WIEW*. Sie dokumentieren die Gesamtzahl der Antworten (*n*) sowie die Aufteilung in gültige (n_{valid}) und ungültige Antworten ($n_{invalid}$).

Der Manipulations-Check zeigt, dass ein Grossteil der Antworten ausgeschlossen werden muss, was auf Probleme bei der Erkennung der Variationen hinweist.

Beim Fallbeispiel *KIBAT* (*Tabelle 15*) sind in der Dimension *Invasivität* (das Tool gibt eine invasivere Einschätzung vor, als die Sozialarbeitenden) 99 von insgesamt 128 Antworten (77%) bei der Ausprägung «strenger» und 66 von 76 (23%) bei «weniger streng» gültig. In der Dimension *prädiktive Validität* erkannten die Teilnehmenden die Variationen schlechter: bei einer hohen Ausprägung der prädiktiven Validität (das Tool ist treffsicherer als die Sozialarbeitenden) werden 79 von 107 (74%) Antworten als gültig angesehen, während bei einer neutralen Ausprägung nur 56 von 97 (58%) Antworten gültig sind. In der Dimension des *wissenschaftlichen Wissens* werden 70 von 93 (75%) Antworten als gültig erkannt, wenn erkannt wurde, dass das Tool mit Literatur der Sozialen Arbeit trainiert wurde. Im Gegensatz dazu können nur 57 von 111 (51%) Antworten als gültig eingestuft werden, wenn variiert wurde, dass keine Literatur vorhanden war.

Tabelle 15 – Manipulations-Check *KIBAT* – Auswertung Antworten nach Dimension

KIBAT				
Dimension	**Level**	**n**	**n_{valid}**	**$n_{invalid}$**
Invasivität	a_1 invasiver	128 (100%)	99 (77%)	29 (23%)
	a_0 weniger invasiv	76 (100%)	66 (87%)	10 (13%)
Prädiktive Validität	b_1 treffsicherer	107 (100%)	79 (74%)	28 (26%)
	b_0 gleich treffsicher	97 (100%)	56 (58%)	41 (42%)
Wissenschaftliches Wissen	c_1 vorhanden	93 (100%)	70 (75%)	23 (25%)
	c_0 nicht vorhanden	111 (100%)	57 (51%)	54 (49%)

Quelle: eigene Erhebung und Darstellung

Die Auswertung für das zweite Fallbeispiel *WIEW* in der *Tabelle 16* zeigt ein ähnliches Bild. In der Dimension *Invasivität* sind 53 von 76 (70%) Antworten bei der Ausprägung «strenger» und 120 von 128 (94%) bei «weniger streng» gültig. In der Dimension *prädiktive Validität* erkannten die Teilnehmenden die Variationen unterschiedlich:

Bei einer hohen Ausprägung der prädiktiven Validität werden 91 von 97 (94%) Antworten als gültig erkannt, während bei einer neutralen Ausprägung nur 57 von 107 (53%) Antworten gültig sind. In der Dimension des *wissenschaftlichen Wissens* werden 98 von 111 (88%) Antworten als gültig erkannt. Im Gegensatz dazu können nur 52 von 93 (56%) Antworten als gültig eingestuft werden, wenn variiert wurde, dass keine Literatur vorhanden war.

Tabelle 16 – Manipulations-Check *WIEW* – Auswertung Antworten nach Dimension

		WIEW		
Dimension	**Level**	**n**	**n_{valid}**	**$n_{invalid}$**
Invasivität	a_1 invasiver	76 (100%)	53 (70%)	23 (30%)
	a_0 weniger invasiv	128 (100%)	120 (94%)	8 (6%)
Prädiktive Validität	b_1 treffsicherer	97 (100%)	91 (94%)	6 (6%)
	b_0 gleich treffsicher	107 (100%)	57 (53%)	50 (47%)
Wissenschaftliches Wissen	c_1 vorhanden	111 (100%)	98 (88%)	13 (12%)
	c_0 nicht vorhanden	93 (100%)	52 (56%)	41 (44%)

Quelle: eigene Erhebung und Darstellung

Die Schwierigkeiten bei der Erkennung der Ausprägungen können durch verschiedene Faktoren beeinflusst worden sein. Diese sowie die Auswirkungen auf die weitere Analyse werden in den Limitationen der Studie (Kapitel 7.6) näher diskutiert.

6.3.2 Mittelwertvergleiche Vignettendimensionen *KIBAT* und *WIEW*

Zunächst werden nun die Ergebnisse der Mittelwertvergleiche für die Vignettendimensionen der beiden Tools *KIBAT* und *WIEW* deskriptiv dargestellt. Diese Darstellung bietet eine erste Übersicht über die Mittelwerte und statistischen Signifikanzen in Bezug auf das Vertrauen der Teilnehmenden in die jeweiligen Tools. Durch die Analyse der Mittelwerte und ihrer Standardabweichungen kann beurteilt werden, wie verschiedene Dimensionen der Vignetten die Wahrnehmung und das Vertrauen der Teilnehmenden beeinflussen. Die statistischen Signifikanzen zeigen auf, ob die beobachteten Unterschiede in den Mittelwerten auf Zufälligkeiten beruhen oder tatsächlich signifikante Effekte darstellen. Diese deskriptive Darstellung ermöglicht es, erste Muster zwischen den Vignettendimensionen und dem Vertrauen der Teilnehmenden zu erkennen. Weitere Tests zu den Zusammenhängen folgen in einem nächsten Schritt.

Abklärungstool KIBAT

Die Ergebnisse des Mittelwertvergleichs für die Vignettendimensionen des Abklärungstools *KIBAT* sind in *Tabelle 17* dargestellt. Sie zeigen die Mittelwerte und statistischen Signifikanzen in Bezug auf das Vertrauen der Teilnehmenden.

Die Dimension *Invasivität* zeigt einen signifikanten Einfluss auf das Vertrauensniveau der Teilnehmenden ($p < 0.001$). Bei höherer Invasivität (a_1) beträgt der Mittelwert 3.19 mit einer Standardabweichung von 1.15, während er bei weniger Invasivität (a_0) bei 3.73 liegt, mit einer geringeren Standardabweichung von 0.94. Die Dimension *prädiktive Validität* zeigt ebenfalls einen signifikanten Einfluss auf das Vertrauen ($p < 0.001$). Hohe Treffsicherheit

(b_1) resultiert in einem Vertrauensmittelwert von 3.66 (*SD* 1.02), während gleiche Treffsicherheit (b_0) zu einem niedrigeren Mittelwert von 3 (*SD* 1.07) führt. Eine höhere Treffsicherheit korreliert entsprechend mit höherem Vertrauen in das Tool. In der Dimension des *wissenschaftlichen Wissens* lässt sich kein signifikanter Einfluss auf das Vertrauensniveau feststellen (p = 0.765). Sowohl bei Wahrnehmung von Literaturquellen im Tool (c_1) als auch bei deren Abwesenheit (c_0) sind die Mittelwerte des Vertrauens relativ ähnlich (M = 3.35 bzw. 3.29) mit leicht höherer Standardabweichung in der Gruppe ohne wissenschaftlichem Wissen (*SD* = 1.18 gegenüber 1.09).

Die Ergebnisse zeigen, dass bei *KIBAT* insbesondere die *Invasivität* und die *prädiktive Validität* eine entscheidende Rolle für das Vertrauen spielen, während das wahrgenommene *wissenschaftliche Wissen* weniger Einfluss zu haben scheint.

Tabelle 17 – Mittelwertvergleiche: Vertrauen in Fallbeispiel *KIBAT*

KIBAT				
Unabhängige Variablen		***Abhängige Variable: Vertrauen***		
Dimension	**Level**	**n**	**M (*SD*)**	***p***
Invasivität	a_1 invasiver	99	3.19 (1.15)	< 0.001
	a_0 weniger invasiv	66	3.73 (0.94)	
Prädiktive Validität	b_1 treffsicherer	79	3.66 (1.02)	< 0.001
	b_0 gleich treffsicher	56	3.00 (1.07)	
Wissenschaftliches Wissen	c_1 vorhanden	70	3.35 (1.09)	0.765
	c_0 nicht vorhanden	57	3.29 (1.18)	

Quelle: eigene Erhebung und Darstellung

Interventionstool WIEW

Im zweiten Fallbeispiel *WIEW* zeigt die Analyse in *Tabelle 18*, dass ein höheres Level der *Invasivität* (a_1) zu einem leicht geringeren Vertrauen führt, auch wenn diese Aussage nicht signifikant ist (p = 0.17). Der Mittelwert liegt hier bei 3.6 (*SD* 0.93), verglichen mit 3.83 (*SD* 1.08) bei weniger strenger Invasivität (a_0). In diesem Kontext scheint die wahrgenommene Invasivität des Tools nur einen geringen Einfluss auf das Vertrauen zu haben. Bei der zweiten Dimension *prädiktive Validität* liegt ein signifikanter Zusammenhang (p = 0.002) zwischen Vertrauen und Treffsicherheit (b_1) vor. Ein treffsichereres Tool führt zu einem Mittelwert von 4.01 (*SD* 1.05), während bei einem Tool, dass gleich gut wie die Sozialarbeitenden selber ist (b_0), ein Mittelwert von 3.51 (*SD* 0.89) vorliegt, was ein deutlich erhöhtes Vertrauen andeutet. Die Dimension *wissenschaftliches Wissen* zeigt keinen statistisch signifikanten Einfluss auf das Vertrauen der Teilnehmenden (p = 0.983). Sowohl bei Vorhandensein (c_1) als auch bei Abwesenheit (c_0) von Literaturangaben im Beispiel bleiben die Vertrauenswerte identisch (M = 3.78).

Die Ergebnisse zeigen, dass bei *WIEW* insbesondere die *prädiktive Validität* eine entscheidende Rolle für das Vertrauen spielt, während die wahrgenommene *Invasivität* und das *wissenschaftliche Wissen* weniger Einfluss zu haben scheinen.

Tabelle 18 – Mittelwertvergleiche: Vertrauen in Fallbeispiel *WIEW*

WIEW				
Unabhängige Variablen		*Abhängige Variable: Vertrauen*		
Dimension	**Level**	**n**	**M (*SD*)**	***p***
Invasivität	a_1 invasiver	53	3.6 (0.93)	0.170
	a_0 weniger invasiv	120	3.83 (1.08)	
Prädiktive Validität	b_1 treffsicherer	91	4.01 (1.05)	0.002
	b_0 gleich treffsicher	57	3.51 (0.89)	
Wissenschaftliches Wissen	c_1 vorhanden	98	3.78 (1.08)	0.983
	c_0 nicht vorhanden	52	3.78 (0.93)	

Quelle: eigene Erhebung und Darstellung

6.3.3 Multiple Regressionsanalyse

Im nächsten Schritt werden multiple Regressionsanalysen durchgeführt, um die Einflüsse der in den Vignetten variierten Variablen und der Personenmerkmale auf das Vertrauen in die beiden Tools genauer zu bestimmen und die Hypothesen H_1 bis H_4 zu überprüfen.

Die im Kapitel 4.2 aufgestellte Hypothese H_1 sowie ihre Subhypothesen $H_{1.1}$, $H_{1.2}$ und $H_{1.3}$ und die Hypothesen H_2, H_3 und H_4 werden in diesem Kapitel geprüft. Zur Übersicht werden sie nochmals aufgelistet:

H_1 Wenn Sozialarbeitende die Invasivität einer KI-Empfehlung als weniger invasiv im Vergleich zu ihrer eigenen Einschätzung empfinden ($H_{1.1}$), wenn sie die prädiktive Validität der Tools als hoch einschätzen ($H_{1.2}$), und wenn sie wissenschaftliches Wissen der Sozialen Arbeit innerhalb von KI-gestützten Entscheidungshilfen wahrnehmen ($H_{1.3}$), dann neigen sie eher dazu, einer KI-gestützten Applikation zu vertrauen.

- **$H_{1.1}$** Wenn die Einschätzung einer KI-gestützten Entscheidungshilfe invasiver ausfällt als die Einschätzung der Sozialarbeitenden, ist das Vertrauen in das Tool geringer.
- **$H_{1.2}$** Je grösser die prädiktive Validität der KI-gestützten Entscheidungshilfe, desto höher das Vertrauen von Sozialarbeitenden in das Tool.
- **$H_{1.3}$** Je eher innerhalb KI-gestützten Entscheidungshilfen wissenschaftliches Wissen der Sozialen Arbeit wahrgenommen wird, desto höher ist das Vertrauen von Sozialarbeitenden in das Tool.

H_2 Je höher der Expert:innenstatus der Sozialarbeitenden, desto geringer das Vertrauen in KI- gestützte Entscheidungshilfen.

H_3 Je höher die digitale Affinität der Sozialarbeitenden, desto höher das Vertrauen in KI-gestützte Entscheidungshilfen.

H_4 Je positiver die generelle Einstellung gegenüber KI, desto höher das Vertrauen in KI-gestützte Entscheidungshilfen.

Wie im vorangehenden Kapitel dargelegt, wurden einige Daten aufgrund des nicht bestandenen Manipulations-Checks aus der Analyse ausgeschlossen. Fehlende Werte können bei Regressionsanalysen zu Verzerrungen führen und eine Untersuchung verfälschen (Urban, Mayerl &Wahl, 2016, S. 2). Um eine Regressionsanalyse mit möglichst

aussagekräftigen Werten zu erhalten, wurde deshalb entscheiden, sämtliche Fälle pro Fallbeispiel, bei welchen ein oder mehrere Ausprägungen der Dimensionen nicht erkannt wurden, aus der Analyse auszuschliessen. Auf eine Imputation der Fälle wird verzichtet, da die fehlenden Daten nicht rein zufällig entstanden sind, sondern zuvor bewusst ausgeschlossen wurden (Urban et al., 2016, S. 5). Diese Massnahme schränkt zwar die Repräsentativität der Untersuchung ein, jedoch gewährleistet sie, dass die verbleibenden Daten eine höhere Zuverlässigkeit und Gültigkeit aufweisen, da sie keine durch Manipulationsfehler verzerrten Werte enthalten. Dieser Ansatz fördert die Genauigkeit der Ergebnisse, auch wenn er die Generalisierbarkeit der Studie potenziell verringert. In der Untersuchung können nach dem Ausschluss im Fallbeispiel *KIBAT* insgesamt 73 Beobachtungen analysiert werden, während es beim Fallbeispiel *WIEW* 93 Fälle sind.

Die Ergebnisse der multiplen Regressionsanalysen[20], die in *Tabelle* 19 (*KIBAT*) und *Tabelle 20* (*WIEW*) dargestellt sind, zeigen detailliertere Einblicke in die Faktoren, die das Vertrauen in die beiden Tools beeinflussen. Die Modelle 1 bis 5 innerhalb jeder Analyse variieren in der Integration der verschiedenen unabhängigen Variablen, um die unterschiedlichen Einflüsse auf das Vertrauen zu untersuchen. Das erste Modell untersucht ausschliesslich den Einfluss der drei Dimensionen *Invasivität*, *prädiktive Validität* und *wissenschaftliches Wissen* auf die Einschätzung des Vertrauens der Befragten, während in den nachfolgenden Modellen zusätzlich die Personenmerkmale in die Regression einbezogen werden (Rost, 2018, S.97).

Abklärungstool KIBAT

Die Ergebnisse im Modell 1 zeigen, dass *Invasivität* einen signifikant negativen Einfluss auf das Vertrauen in *KIBAT* hat (b[21] = -0.587, $p < 0.05$), was Hypothese $H_{1.1}$ bestätigt, dass eine invasivere Einschätzung der KI-gestützten Entscheidungshilfe das Vertrauen verringert. Die *prädiktive Validität* weist einen signifikant positiven Einfluss auf das Vertrauen auf (b = 0.736, $p < 0.01$), was Hypothese $H_{1.2}$ bestätigt, dass eine höhere prädiktive Validität das Vertrauen erhöht. *Wissenschaftliches Wissen* hingegen zeigt keinen signifikanten Einfluss auf das Vertrauen (b = -0.026, $p > 0.05$), was Hypothese $H_{1.3,}$ dass die Wahrnehmung von wissenschaftlichem Wissen das Vertrauen erhöht, nicht bestätigt.

Im Modell 2 wird zusätzlich zu den drei in den Fallbeispielen variierten Variablen der *Expert:innenstatus* der Teilnehmenden einbezogen. Hier bleibt der negative Einfluss der *Invasivität* signifikant (b = -0.587, $p < 0.05$) und der positive Einfluss der *prädiktiven Validität* stark und signifikant (b = 0.740, $p < 0.01$). Der *Expert:innenstatus* selbst

[20] Bei der multiplen linearen Regression wird untersucht, wie eine abhängige Variable von mehreren unabhängigen Variablen beeinflusst wird (Zöfel, 2003, S. 175).

[21] Der unstandardisierte Regressionskoeffizient (b) gibt die Veränderung in der abhängigen Variablen an, wenn sich die unabhängige Variable um eine Einheit ändert. Ein positiver b-Wert bedeutet, dass ein Anstieg der unabhängigen Variablen mit einem Anstieg der abhängigen Variablen einhergeht, während ein negativer b-Wert das Gegenteil anzeigt. Die Grösse gibt an, wie stark dieser Einfluss ist (Döring & Bortz, 2016, S. 689).

zeigt jedoch keinen signifikanten Einfluss auf das Vertrauen in *KIBAT* (b = 0.017, p > 0.05), was Hypothese H_2, dass ein höherer *Expert:innenstatus* das Vertrauen in KI-gestützte Entscheidungshilfen verringert, nicht bestätigt.

Das Modell 3 ergänzt die Analyse um die *digitale Affinität* der Teilnehmenden. Auch in diesem Modell bleibt der negative Einfluss der *Invasivität* signifikant (b = -0.531, p < 0.05) und der positive Einfluss der *prädiktiven Validität* signifikant (b = 0.725, p < 0.01). Die *digitale Affinität* zeigt jedoch keinen signifikanten Einfluss auf das Vertrauen (b = 0.324, p > 0.05), was Hypothese H_3, dass eine höhere *digitale Affinität* das Vertrauen in KI-gestützte Entscheidungshilfen erhöht, nicht bestätigt.

Im Modell 4 werden die Vignettendimensionen sowie die *Einstellung gegenüber KI* einbezogen. Die Ergebnisse zeigen, dass der negative Einfluss der *Invasivität* stärker wird und signifikant bleibt (b = -0.735, p < 0.01). Der positive Einfluss der *prädiktiven Validität* bleibt stark und signifikant (b = 0.773, p < 0.01). *Wissenschaftliches Wissen* zeigt weiterhin keinen signifikanten Einfluss auf das Vertrauen (b = -0.055, p > 0.05). Die *Einstellung gegenüber KI* hingegen hat einen signifikanten positiven Einfluss auf das Vertrauen in *KIBAT* (b = 0.408, p < 0.01), was Hypothese H_4, dass eine höhere *Einstellung gegenüber KI* das Vertrauen in KI-gestützte Entscheidungshilfen erhöht, bestätigt.

Im fünften Modell wird der Einfluss aller vorherigen Variablen kombiniert. Die Ergebnisse zeigen, dass der negative Einfluss der *Invasivität* stark und signifikant bleibt (b = -0.737, p < 0.01), ebenso wie der positive Einfluss der *prädiktiven Validität* (b = 0.789, p < 0.01). *Wissenschaftliches Wissen* bleibt weiterhin ohne signifikanten Einfluss (b = -0.032, p > 0.05). Der *Expert:innenstatus* zeigt keinen signifikanten Einfluss (b = 0.056, p > 0.05), während die *digitale Affinität* einen positiven, aber nicht signifikanten Einfluss auf das Vertrauen hat (b = 0.092, p > 0.05). Die *Einstellung gegenüber KI* zeigt wiederum einen starken und signifikanten positiven Einfluss auf das Vertrauen in *KIBAT* (b = 0.452, p < 0.01).

In der Analyse des Modells 5 zeigt sich eine interessante Entwicklung bezüglich der Konstanten[22]: In diesem Modell zeigt die Konstante mit einem Wert von 0.825 keine statistische Signifikanz (p = 0.894) im Gegensatz zu den vorherigen Modellen. Dieser Rückgang der Signifikanz deutet darauf hin, dass mit dem Einbezug aller Dimensionen und Personenmerkmale das Niveau des Vertrauens in *KIBAT*, das unabhängig von diesen Variablen existiert, verringert wird. Dies könnte bedeuten, dass – wenn alle wichtigen Einflussfaktoren berücksichtigt werden – das inhärente Vertrauen in das Tool schwächer ausgeprägt ist oder dass das Vertrauen der Nutzenden stärker von diesen spezifischen Variablen abhängt, als zuvor angenommen. Darüber hinaus zeigt Modell 5 eine Steigerung des R^2-

[22] Die Konstante in einer Regressionsanalyse repräsentiert das Grundniveau der abhängigen Variable, wenn alle unabhängigen Variablen gleich Null sind. Dieser Wert gibt eine Basislinie oder ein Startniveau für das Vertrauen, bevor irgendwelche spezifischen Einflussfaktoren berücksichtigt werden. Es zeigt das anfängliche Vertrauen, das vorhanden wäre, auch ohne dass die spezifischen Merkmale oder Eigenschaften der Teilnehmenden und des Tools in die Analyse einbezogen werden. Dieser Basiswert bildet einen allgemeinen Ausgangspunkt für weitere Berechnungen und Vergleiche (Döring & Bortz, 2016, S. 224).

Wertes[23] auf 0.248. Dies ist ein Anstieg gegenüber den vorherigen Modellen und spiegelt die verbesserte Erklärungskraft durch das Hinzufügen weiterer relevanter Variablen wider. Es zeigt, dass die Kombination dieser Variablen wichtige Einflussfaktoren auf das Vertrauen in *KIBAT* sind und dass die Vertrauensbildung bei den Teilnehmenden durch verschiedene Aspekte beeinflusst wird. Obwohl dieses Modell mit den analysierten Faktoren einen umfangreichen Einblick bietet, deutet das verbleibende unerklärte Vertrauen darauf hin, dass weitere, bisher nicht berücksichtigte Faktoren existieren können, welche das Vertrauen in KI-Systeme beeinflussen. Diese könnten zusätzliche Aspekte der wahrgenommenen Nützlichkeit umfassen, wie spezifische Funktionalitäten oder weitere externe Faktoren, z. B. hedonistische Motivation oder sozialer Einfluss (Yang & Wibowo, 2022).

Tabelle 19 – Multiple Regression KIBAT: Betrachtung des Vertrauens auf die unabhängigen Variablen

Fallbeispiel *KIBAT*

Vignettendimensionen und Personenmerkmale (unabhängige Variablen)	Koeffizienten: Vertrauen in *KIBAT* (abhängige Variable)				
Dimension	Modell 1: D1-3	Modell 2: D1-3 Expert:innenstatus	Modell 3: D1-3 Digitale Affinität	Modell 4: D1-3, Einstellung zu KI	Modell 5: D1-3 Expert:innenstatus, Digitale Affinität, Einstellung zu KI
	b (SD)	b (SD)	b (SD)	b (SD)	b (SD)
D1: Invasivität (A)	-0.587** (0.247)	-0.587** (0.249)	-0.531** (0.246)	-0.735*** (0.243)	-0.737*** (0.254)
D2: Prädiktive Validität (B)	0.736*** (0.251)	0.740*** (0.253)	0.725*** (0.248)	0.773*** (0.242)	0.789*** (0.242)
D3: Wissenschaftliches Wissen (C)	-0.026 (0.239)	-0.019 (0.241)	-0.013 (0.235)	-0.055 (0.230)	-0.032 (0.230)
M1: Expert:innenstatus		0.017 (0.037)			0.056 (0.038)
M2: Digitale Affinität			0.324 (0.188)		0.092 (0.212)
M3: Einstellung zu KI				0.408*** (0.154)	0.452*** (0.188)
Konstante	3.166*** (0.244)	3.016*** (0.406)	1.735** (0.865)	1.862*** (0.546)	0.825 (0.973)
Anzahl Beobachtungen	73	73	73	73	73
R^2	0.141	0.144	0.176	0.221	0.248

* p < 0.1; ** p < 0.05; *** p < 0.01

Quelle: eigene Erhebung und Darstellung

[23] Der R^2-Wert stellt Bestimmtheitsmass der Regression dar und gibt an, welcher Anteil der Varianz der abhängigen Variable durch die unabhängigen Variablen erklärt wird (Döring & Bortz, 2016, S. 626).

Zusammenfassend verdeutlichen die Ergebnisse bei *KIBAT*, dass sich eine höhere *Invasivität* negativ auf das Vertrauen auswirkt, während eine höhere *prädiktive Validität* das Vertrauen in das Tool stärkt. *Wissenschaftliches Wissen* hingegen zeigt keinen signifikanten Einfluss auf das Vertrauen. Die Ergebnisse unterstützen entsprechend die Haupthypothese H_1 nur teilweise. Bei den Personenmerkmalen zeigen der *Expert:innenstatus* und die *digitale Affinität* der Teilnehmenden keinen signifikanten Einfluss auf das Vertrauen, wohingegen eine positivere *Einstellung gegenüber KI* einen starken positiven Einfluss auf das Vertrauen in *KIBAT* hat. Die Hypothesen H_2 bis H_4 können entsprechend in Hinsicht auf *KIBAT* ebenfalls nur teilweise angenommen werden.

Interventionstool WIEW

In der zweiten multiplen Regressionsanalyse wird das zweite Fallbeispiel *WIEW* dargestellt (*Tabelle 20*). Die Ergebnisse des ersten Modells zeigen, dass *Invasivität* keinen signifikanten Einfluss auf das Vertrauen in *WIEW* hat (b = -0.029, $p > 0.05$), was Hypothese $H_{1.1}$, dass eine invasivere Einschätzung der KI-gestützten Entscheidungshilfe das Vertrauen verringert, in diesem Tool nicht bestätigt. Die *prädiktive Validität* weist hingegen einen signifikant positiven Einfluss auf das Vertrauen auf (b = 0.448, $p < 0.05$), was Hypothese $H_{1.2}$ bestätigt, dass eine höhere prädiktive Validität das Vertrauen erhöht. *Wissenschaftliches Wissen* zeigt keinen signifikanten Einfluss auf das Vertrauen (b = 0.149, $p > 0.05$), was Hypothese $H_{1.3}$, dass die Wahrnehmung von wissenschaftlichem Wissen das Vertrauen erhöht, nicht bestätigt.

Im zweiten Modell wird zusätzlich zu den drei in den Fallbeispielen variierten Variablen der *Expert:innenstatus* der Teilnehmenden einbezogen. Hier bleibt der Einfluss der *Invasivität* nicht signifikant (b = -0.029, $p > 0.05$) und der positive Einfluss der *prädiktiven Validität* stark und signifikant (b = 0.448, $p < 0.05$). Der *Expert:innenstatus* selbst zeigt keinen signifikanten Einfluss auf das Vertrauen in *WIEW* (b = -0.004, $p > 0.05$), was Hypothese H_2, dass ein höherer *Expert:innenstatus d*as Vertrauen in KI-gestützte Entscheidungshilfen verringert, nicht bestätigt.

Das dritte Modell ergänzt die Analyse um die *digitale Affinität* der Teilnehmenden. Auch in diesem Modell bleibt der Einfluss der *Invasivität* nicht signifikant (b = -0.033, $p > 0.05$) und der positive Einfluss der *prädiktiven Validität* stark und signifikant (b = 0.458, $p < 0.05$). Die *digitale Affinität* zeigt jedoch keinen signifikanten Einfluss auf das Vertrauen (b = 0.110, $p > 0.05$), was Hypothese H_3, dass eine höhere *digitale Affinität* das Vertrauen in KI-gestützte Entscheidungshilfen erhöht, nicht bestätigt.

Im vierten Modell werden die Vignettendimensionen sowie die *Einstellung gegenüber KI* einbezogen. Die Ergebnisse zeigen, dass der Einfluss der *Invasivität* weiterhin nicht signifikant ist (b = -0.049, $p > 0.05$). Der positive Einfluss der *prädiktiven Validität* bleibt signifikant (b = 0.461, $p < 0.05$). *Wissenschaftliches Wissen* zeigt weiterhin keinen signifikanten Einfluss auf das Vertrauen (b = 0.063, $p > 0.05$). Die *Einstellung gegenüber KI* hingegen hat einen starken und signifikanten positiven Einfluss auf das Vertrauen in *WIEW* (b = 0.430, $p < 0.01$), was Hypothese H_4, dass eine positivere *Einstellung gegenüber KI* das Vertrauen in KI-gestützte Entscheidungshilfen erhöht, bestätigt.

Im fünften Modell wird der Einfluss aller vorherigen Variablen kombiniert. Die Ergebnisse zeigen, dass der Einfluss der *Invasivität* weiterhin nicht signifikant ist (b = -0.047, p > 0.05). Der positive Einfluss der *prädiktiven Validität* bleibt hingegen signifikant (b = 0.457, p < 0.05). *Wissenschaftliches Wissen* hat weiterhin keinen signifikanten Einfluss (b = 0.063, p > 0.05). Der *Expert:innenstatus* zeigt ebenfalls keinen signifikanten Einfluss (b = 0.019, p > 0.05), und die *digitale Affinität* hat einen schwachen negativen, aber nicht signifikanten Einfluss auf das Vertrauen (b = -0.046, p > 0.05). Die *Einstellung gegenüber KI* zeigt hingegen einen starken und signifikant positiven Einfluss auf das Vertrauen in *WIEW* (b = 0.462, p < 0.01).

Auch bei *WIEW* zeigt sich die Konstante im Modell 5 niedriger als bei den vorherigen Modellen, bleibt aber hier signifikant (1.949, p= < 0.05). Der kontinuierliche Rückgang der Konstantenwerte deutet wie auch beim vorherigen Modell darauf hin, dass das Niveau des Vertrauens in *WIEW*, das unabhängig von den betrachteten Variablen existiert, abnimmt, wenn mehr Einflussfaktoren einbezogen werden. Darüber hinaus zeigt Modell 5 eine Steigerung des R^2-Wertes auf 0.162. Das ist ein Anstieg gegenüber den vorherigen Modellen und spiegelt die verbesserte Erklärungskraft durch das Hinzufügen weiterer relevanter Variablen wider.

Tabelle 20 – Multiple Regression WIEW: Betrachtung des Vertrauens auf die unabhängigen Variablen

Fallbeispiel *WIEW*

Vignettendimensionen und Personenmerkmale (unabhängige Variablen)	Koeffizienten: Vertrauen in *WIEW* (abhängige Variable)				
Dimension	Modell 1: D1-3	Modell 2: D1-3 Expert:innenstatus	Modell 3: D1-3 Digitale Affinität	Modell 4: D1-3, Einstellung zu KI	Modell 5: D1-3 Expert:innenstatus, Digitale Affinität, Einstellung zu KI
	b (SD)	b (SD)	b (SD)	b (SD)	b (SD)
D1: Invasivität	-0.029 (0.239)	-0.029 (0.241)	-0.033 (0.240)	-0.049 (0.227)	-0.047 (0.229)
D2: Prädiktive Validität	0.448** (0.220)	0.448** (0.222)	0.458** (0.222)	0.461** (0.209)	0.457** (0.211)
D3: Wissenschaftliches Wissen	0.149 (0.230)	0.149 (0.232)	0.132 (0.233)	0.063 (0.220)	0.063 (0.223)
M1: Expert:innenstatus		-0.004 (0.031)			0.019 (0.032)
M2: Digitale Affinität			0.110 (0.172)		-0.046 (0.180)
M3: Einstellung zu KI				0.430** (0.128)	0.462** (0.138)
Konstante	3.408*** (0.234)	3.442*** (0.359)	2.947*** (0.755)	2.025*** (0.468)	1.949* (0.927)
Anzahl Beobachtungen	93	93	93	93	93
R^2	0.049	0.049	0.053	0.157	0.162

* p < 0.1; ** p < 0.05; *** p < 0.01

Quelle: eigene Erhebung und Darstellung

Zusammenfassend zeigen die Ergebnisse bei *WIEW*, dass eine höhere *Invasivität* einen negativen Einfluss auf das Vertrauen hat, diese aber nicht signifikant ist. Eine höhere *prädiktive Validität* beeinflusst das Vertrauen in das Tool wiederum signifikant. *Wissenschaftliches Wissen* beeinflusst das Vertrauen nicht signifikant. Die Haupthypothese H_1 kann auch bei *WIEW* nur teilweise angenommen werden. Bei den Personenmerkmalen haben wiederum weder der *Expert:innenstatus* noch die *digitale Affinität* der Teilnehmenden einen signifikanten Einfluss auf das Vertrauen. Eine positivere *Einstellung gegenüber KI* hingegen hat einen signifikant positiven Einfluss auf das Vertrauen in *WIEW*. Somit werden die Hypothesen H_2 und H_3 in Bezug auf *WIEW* abgelehnt, während H_4 angenommen wird.

6.4 Integrationsbereitschaft

Zum Schluss wird getestet, ob das Vertrauen in die beiden Tools einen positiven Einfluss auf die *Integrationsbereitschaft* der Teilnehmenden hat. Dabei wird die Hypothese H_5 geprüft:

H_5 Je höher das Vertrauen in Einschätzung eines KI-gestützte Entscheidungshilfen, desto höher die Integrationsbereitschaft entsprechender Tools im Arbeitsalltag.

Um die Hypothese zu überprüfen, wird die Variable der *Integrationsbereitschaft* aus drei verschiedenen Items zu einer neuen Skala zusammengefasst (vgl. Kapitel 5.3.2). Anschliessend wird mit einer Korrelationsanalyse mittels des Spearman-Korrelationskoeffizienten[24] getestet, ob ein höheres Vertrauen in die KI-gestützten Entscheidungshilfen mit einer höheren Integrationsbereitschaft der entsprechenden Tools im Arbeitsalltag zusammenhängt.

Die Ergebnisse der Korrelationsanalyse sind in *Tabelle 21* dargestellt. Für beide Tools, *KIBAT* und *WIEW*, zeigt die Analyse eine positive und statistisch signifikante Korrelation zwischen dem Vertrauen in die Tools und der Integrationsbereitschaft. Der Spearman-Korrelationskoeffizient für *KIBAT* beträgt 0.503, was auf eine moderate positive Korrelation hinweist. Der *p*-Wert ist kleiner als 0.001. Der Spearman-Korrelationskoeffizient für *WIEW* beträgt 0.609, was auf eine starke positive Korrelation hinweist. Auch hier ist der *p*-Wert kleiner als 0.001, was ebenfalls auf eine statistisch signifikante Korrelation hinweist.

Die Ergebnisse der Korrelationsanalyse bestätigen die Hypothese H_5: Ein höheres Vertrauen in die Einschätzung von KI-gestützten Entscheidungshilfen korreliert positiv mit einer höheren Bereitschaft, diese Tools im Arbeitsalltag zu integrieren.

[24] Da die Daten der Integrationsbereitschaft nicht normalverteilt sind, wird zur Analyse der Korrelation zwischen Vertrauen und Integrationsbereitschaft der Spearman-Korrelationskoeffizient verwendet. Dieser nicht-parametrische Test ist weniger empfindlich gegenüber Ausreissern und nicht normalverteilten Daten und eignet sich daher für diese Analyse (Zöfel, 2003, S. 154).

Tabelle 21 – Korrelationsanalyse Vertrauen und Integrationsbereitschaft

Korrelationsanalyse Vertrauen und Integrationsbereitschaft in beide Tools			
Tool	df	ρ (Spearman)	*p*-Wert
KIBAT	195	0.503	< 0.001
WIEW	197	0.609	< 0.001

Quelle: eigene Erhebung und Darstellung

6.5 Zusammenfassung der Ergebnisse

Die im Kapitel 4.2 aufgestellten Hypothesen lassen sich basierend auf den Ergebnissen in den vorherigen Unterkapitel wie folgt beantworten:

Haupthypothese H_1 nimmt an, dass Sozialarbeitende eher dazu neigen, einer KI-gestützten Applikation zu vertrauen, wenn sie die *Invasivität* einer KI-Empfehlung als weniger invasiv im Vergleich zu ihrer eigenen Einschätzung empfinden ($H_{1.1}$), wenn sie die *prädiktive Validität* der Tools als hoch einschätzen ($H_{1.2}$) und wenn sie *wissenschaftliches Wissen* der Sozialen Arbeit innerhalb von KI-Entscheidungshilfen wahrnehmen ($H_{1.3}$).

Die erste **Sub-Hypothese $H_{1.1}$**, geht davon aus, dass geringere *Invasivität* das Vertrauen von Sozialarbeitenden in die Tools erhöht. Hier zeigen die Ergebnisse für *KIBAT*, dass ein signifikant negativer Zusammenhang zwischen Invasivität und dem Vertrauen besteht. Dies bestätigt die Annahme von $H_{1.1}$ für *KIBAT*, die Hypothese wird für dieses Fallbeispiel **angenommen**. Im Fall von *WIEW* hingegen zeigte die Invasivität einen negativen, jedoch keinen signifikanten Einfluss auf das Vertrauen, was dazu führt, dass $H_{1.1}$ für *WIEW* **abgelehnt** wird.

Die zweite **Sub-Hypothese $H_{1.2}$** geht davon aus, dass ein positiver Zusammenhang zwischen der *prädiktiven Validität* und dem Vertrauen vorliegt. Diese Annahme wird durch die Daten beider Fallbeispiele deutlich bestätigt. Sowohl für *KIBAT* als auch für *WIEW* zeigte sich, dass eine höhere prädiktive Validität stark positiv mit dem Vertrauen korreliert. Damit wird $H_{1.2}$ für beide Fallbeispiele **angenommen**.

In der dritten **Sub-Hypothese $H_{1.3}$** wird angenommen, dass das wahrgenommene *wissenschaftliche Wissen* das Vertrauen fördert. Hier finden sich in den Daten keine Bestätigung. In beiden Fallbeispielen zeigte das wissenschaftliche Wissen keinen signifikanten Einfluss auf das Vertrauen. Daher wird $H_{1.3}$ in beiden Fallbeispielen **abgelehnt**.

Entsprechend wird die **Haupthypothese H_1 teilweise angenommen**.

Die im Kapitel 4.2.4 aufgestellten Hypothesen H_2, H_3 und H_4 lassen sich basierend auf den Ergebnissen der Korrelationsanalyse und der multiplen Regression wie folgt beantworten.

Hypothese H_2 nimmt an, dass ein höherer *Expert:innenstatus* mit einem geringeren Vertrauen in KI-gestützte Entscheidungshilfen einhergeht. Die Ergebnisse der Korrelationsanalyse sowie der multiplen Regressionsanalyse zeigen jedoch für beide Tools keine signifikante Werte in dieser Hinsicht. Dementsprechend wird die Hypothese H_2 aufgrund der fehlenden statistischen Signifikanz **abgelehnt**.

Hypothese H_3 vermutet, dass eine höhere *digitale Affinität* der Sozialarbeitenden zu einem gesteigerten Vertrauen in KI-gestützte Entscheidungshilfen führt. Die Analyse für beide Tools zeigen jedoch lediglich schwache und statistisch nicht signifikante Werte zwischen digitaler Affinität und dem Vertrauen in die Entscheidungshilfen. Aufgrund dessen wird die Hypothese H_3 aufgrund mangelnder statistischer Evidenz **abgelehnt**.

Hypothese H_4 nimmt an, dass die *generelle Einstellung gegenüber KI* einen Einfluss auf das Vertrauen in KI-gestützte Entscheidungshilfen hat. Die Korrelationsanalyse der Daten bestätigt diese Hypothese. Für beide Tools, *KIBAT* und *WIEW*, zeigen sich signifikante und positive Werte, was darauf hindeutet, dass eine positive Einstellung zur KI mit einem höheren Vertrauen in die Entscheidungshilfen korreliert. Folglich wird Hypothese H_4 **angenommen**.

Die letzte **Hypothese H_5** geht davon aus, dass ein höheres Vertrauen in die Tools zu einer höheren Integrationsbereitschaft der KI-Entscheidungshilfen in die Praxis führt. Die Ergebnisse der Korrelationsanalyse unterstützen diese Hypothese. Das Vertrauen erweist sich als entscheidender Faktor für die Integration der Tools. Daher wird die Hypothese H_5 **angenommen**.

7 Zentrale Befunde und Diskussion

In der vorliegenden Masterarbeit wurde das Vertrauen von Beistandspersonen in der Kinder- und Jugendhilfe in KI-gestützte Entscheidungshilfen untersucht, wie sie zukünftig im Arbeitsalltag des Kindesschutzes eingesetzt werden könnten. An der im März bis April 2024 durchgeführten Online-Umfrage nahmen 204 Personen teil, die in Berufsbeistandschaften im Kindesschutz in der Deutschschweiz tätig sind. Diese Untersuchung verfolgte einen deskriptiven Ansatz, um zu verstehen, wie Beistandspersonen im Kindesschutz solchen Anwendungen gegenüber eingestellt sind und wie verschiedene Faktoren ihr Vertrauen beeinflussen.

Mithilfe des modifizierten Technologieakzeptanzmodells (*TAM*) wurden Faktoren ermittelt, die das Vertrauen der Beistandspersonen beeinflussen könnten. Basierend auf den in einem quantitativen Fragebogen erhobenen Daten zum Vertrauen in KI-gestützte Entscheidungshilfen wurden fünf Hypothesen überprüft. In diesem Kapitel werden die Ergebnisse der Datenanalyse nun zusammengefasst und im Hinblick auf die Forschungsfrage diskutiert.

Am Ende des Fragebogens hatten die teilnehmenden Personen die Möglichkeit, ihre Gedanken zu den zuvor beantworteten Fragen mitzuteilen. Diese Zitate fliessen im Sinne eines qualitativen Feedbacks in die folgende Diskussion mit ein (die Zitate sind entsprechend gekennzeichnet) (*Anhang 1*). Zitate, welche weitere wichtige, bisher nicht in der Studie untersuchte Aspekte aufzeigen, werden im Kapitel 7.5 aufgenommen und diskutiert.

7.1 Einfluss der variierten Dimensionen

Die in Kapitel 4 vorgestellten Hauptfaktoren *Invasivität, prädiktive Validität* und *wissenschaftliches Wissen* wurden im Fragebogen den befragten Personen variiert zugewiesen, was eine Auswertung der verschiedenen Stärken der Einflüsse möglich machte. Im Folgenden werden ihre Einflüsse auf das Vertrauen in KI-gestützte Entscheidungshilfen analysiert.

7.1.1 Einfluss der Invasivität

Beim ersten gemessenen Faktor, der *Invasivität* der Einschätzungen der Tools, zeigte die Analyse, dass dieser einen Einfluss auf die spätere Bewertung des Vertrauens hatte. Es zeigte sich, dass die Teilnehmenden eher bereit waren, der Einschätzung des Tools zu folgen, wenn dieses ein weniger starkes Eingreifen in die Situation vorschlug, als sie es selber getan hätten. Diese Neigung war bei *KIBAT* signifikant, wie die multiple Regression aufzeigte, während bei *WIEW* zumindest eine erkennbare Tendenz zu beobachten war.

Die Tendenz lässt sich möglicherweise damit erklären, dass «strengere Massnahmen» von den Teilnehmenden eher mit einem höheren Risiko in Verbindung gebracht werden. Infolgedessen ist es möglich, dass Menschen misstrauischer sind in Bezug auf die Sicherheit oder die Fairness einer Einschätzung, wenn sie «strenger» als ihre eigene Einschätzung ist. Eine Person, die den Fragebogen ausfüllte, äusserte sich hierzu: «Im Sinne der milderen Massnahme würde ich die KI-Empfehlung immer dann nutzen, wenn diese eine mildere Massnahme vorschlägt. Wenn sie aber eine intensivere Massnahme vorschlägt, würde ich wohl eher weiterhin auf der milderen Massnahme beharren und diese zuerst einmal versuchen.» (P39.I).

Theoretisch könnte sich dieses Verhalten durch die kognitive Verzerrung erklären, die in Entscheidungsfindungsprozessen auftreten können. Glaberson (2019, S. 318) betont, dass solche Verzerrungen insbesondere in unsicheren Situationen verstärkt auftreten. Enosh und Bayer-Topilsky (2015, S. 3) ergänzen, dass in Fällen mit höherer Unsicherheit Sozialarbeitende zu heuristischen Methoden greifen, die anfälliger für persönliche Verzerrungen sind. Dies spiegelt sich in der Datenanalyse wider, in welcher weniger invasive KI-Empfehlungen eher akzeptiert werden, während strengere Eingriffe skeptischer betrachtet werden.

Zusätzlich stützt sich die Analyse auf das Konzept des Vertrauens von Mayer et al (1995). Hier wird argumentiert, dass Vertrauen dann aufgebaut wird, wenn das wahrgenommene Risiko niedriger als das Vertrauensniveau ist. Bei einer invasiveren Einschätzung der Situation durch die KI könnte das Risiko, in einer realen Situation zu «streng» zu entscheiden, von den Beistandspersonen entsprechend höher wahrgenommen werden, was das Vertrauen mindert und somit ihre Bereitschaft reduziert, einer entsprechenden Empfehlung zu folgen. Das Prinzip der Verhältnismässigkeit spielt hier möglicherweise eine zentrale Rolle: Die Beistandspersonen sind angehalten, sich an den Grundsatz der Verhältnismässigkeit zu halten, der besagt, dass Massnahmen nur so intensiv sein sollten, wie es die Situation erfordert (Fassbind, 2022, S. 112). Ein weniger invasiver Vorschlag der KI könnte als verhältnismässiger und damit angemessener empfunden werden, was das Vertrauen in die KI-Empfehlungen entsprechend erhöhen könnte.

7.1.2 Einfluss der prädiktiven Validität

Der zweite Faktor der *prädiktiven Validität* zeigte sich in den Auswertungen durchgehend signifikant positiv. Eine hohe wahrgenommene prädiktive Validität – d. h. eine wahrgenommene Genauigkeit und Zuverlässigkeit der Vorhersagen – korreliert positiv mit dem Vertrauen in beide Tools. Die Befragten waren eher bereit, einer Einschätzung des Tools zu folgen, wenn sie davon ausgingen, dass Studien eine höhere Präzision der Einschätzungen vorhersagten als es Sozialarbeitende machen würden. Diese Erkenntnis bestätigt die Subhypothese $H_{2.1}$ und verdeutlicht, dass die Treffsicherheit einer durch KI entwickelten Vorhersage für die Teilnehmenden eine hohe Bedeutung hat.

Die theoretische Grundlage für diese Beobachtung kann durch verschiedene Elemente aus der Literatur gestützt werden. Mayer et al. (1995) betonen, dass Vertrauen wächst, wenn die wahrgenommene Fähigkeit hoch ist. Ebenso unterstreichen Nordheim et al. (2019), dass Konsistenz und Genauigkeit der Leistungen von KI-Systemen das Vertrauen von User:innen erhöhen. Wird die prädiktive Validität eines Tools als hoch wahrgenommen, könnte dies demnach die Unsicherheiten von Beistandspersonen reduzieren. Diese Erkenntnisse decken sich mit den Studien von Enosh und Bayer-Topilsky (2015, S. 3), welche betonen, dass in Fällen mit höherer Unsicherheit analytische Methoden bevorzugt werden, die auf klaren, präzisen Daten basieren. Sozialarbeitende, die davon ausgehen, dass die Einschätzungen eines Tools möglicherweise treffsicherer als ihre eigenen sind, könnten in unsicheren Fallsituationen verstärkt auf solche analytischen Methoden setzen.

Im Kontext des Kindesschutzes wird hervorgehoben, dass prädiktive Modelle dazu beitragen können, konsistentere und weniger verzerrte Entscheidungen zu treffen (Bastian & Schrödter, 2019; S. 46; Keddell, 2019, S. 3). Das *Allegheny Family Screening Tool* (*AFST*) beispielsweise hat gezeigt, dass durch die Nutzung prädiktiver Modelle die Identifizierung von Kindern, die weitere Interventionen benötigten, verbessert werden kann, ohne die Arbeitsbelastung der Sozialarbeitenden zu erhöhen (Goldhaber-Fiebert & Price, 2019). Dies zeigt – unabhängig vom Vertrauen – wie wichtig eine hohe prädiktive Validität für den erfolgreichen Einsatz von KI-Entscheidungshilfen in der Praxis ist.

7.1.3 Einfluss des wissenschaftlichen Wissens

Im Gegensatz zur *prädiktiven Validität* und zur *Invasivität* einer Einschätzung zeigte der dritte Faktor *wissenschaftliches Wissen* keinen signifikanten Einfluss auf das Vertrauen der Teilnehmenden. Diese Ergebnisse widersprechen der Hypothese $H_{2.3}$ und deuten darauf hin, dass der Aspekt, dass ein Tool mit Literatur der Sozialen Arbeit trainiert wurde, für ihr Vertrauen nicht ausschlaggebend ist, oder dass die Teilnehmenden möglicherweise nicht in der Lage sind, den wissenschaftlichen Hintergrund der Tools zu bewerten.

Theoretisch könnte dies durch die Komplexität und die oft fehlende Transparenz von KI-Systemen erklärt werden. Verstehen User:innen den Mechanismus von Tools, welche sie anwenden nicht, und können sie nicht nachvollziehen, wie die Methoden hinter der Einschätzung funktionieren oder auf welchen methodischen Grundlagen sie basieren, können sie entsprechenden Applikationen nur schwer vertrauen (Enosh & Bayer-Topilsky, 2015). Eine befragte Person äusserte sich in diesem Zusammenhang: «Ich finde es schwierig, Aussagen über die Nutzung zu machen, solange ich nicht weiss, wie die KI genau funktioniert (nach welchen Kriterien wird ausgewertet usw.). Wenn ich das System besser kenne, könnte ich mich besser für oder dagegen entscheiden).» (P13.I). Eine weitere Person bemerkte: «Um zu beurteilen, ob ich den Empfehlungen einer KI 'vertrauen' kann, müsste ich wissen an Hand von welchen Informationen (nur Fallbeispiele oder auch Literatur? falls ja, welche Literatur etc.) es die Empfehlung gibt.» (P57.I). Und eine dritte Person äussert sich dazu: «Die Quellen auf denen sich die Einschätzung basiert müsste aber absolut transparent sein und es dürften nur Wissenschaftlich geprüfte/anerkannte Quellen verwendet werden.» (P17.2).

Auch hier kann auf die Herausforderungen kognitiver Verzerrungen hingewiesen werden (Glaberson, 2019, S. 318): Unsicherheiten hinsichtlich der Durchschaubarkeit der Funktionsweise eines KI-gestützten Entscheidungshilfetools könnten dazu führen, dass das Wissen darum, dass das Tool mit wissenschaftlichem Wissen trainiert wurde alleine nicht ausreicht, um Vertrauen zu schaffen. Dies spiegelt den «Black-Box-Effekt» wider, bei dem Menschen einer Technologie misstrauen, weil sie nicht verstehen, wie sie funktioniert (Gillingham, 2019, S. 117). Dies deckt sich auch mit den Erkenntnissen von Sheridan (2019) und Rheu et al. (2021), die betonen, dass das Vertrauen in KI-Systeme über die reine Vorhersagbarkeit hinausgeht und eine Übereinstimmung der Absichten zwischen dem System und den User:innen erforderlich ist. Mayer et al. (1995) argumentieren ebenso, dass Vertrauen aufgebaut wird, wenn die Nutzenden die Funktionsweise und die zugrundeliegenden Prinzipien der Technologie verstehen. Entsprechend wird die Notwendigkeit von Transparenz und verständlicher Kommunikation über die Funktionsweise und die wissenschaftlichen Grundlagen der KI-Modelle höher gewichtet als das alleinige Wissen darum, dass die Tools mit empirischer Literatur trainiert wurden.

Die Analyse der Daten zeigt, dass wissenschaftliches Wissen allein nicht ausreicht, um das Vertrauen der Teilnehmenden in KI-gestützte Entscheidungshilfen zu stärken. Transparenz, Verständlichkeit und die Fähigkeit der User:innen, die Grundlagen hinter der Funktionsweise der KI-Tools zu bewerten, spielen vermutlich eine bedeutend entscheidendere Rolle.

7.2 Einfluss der Personenmerkmale

Die Untersuchung der Personenmerkmale zeigte, dass weder der aus Berufserfahrung und Alter zusammengesetzte *Expert:innenstatus* noch die *digitale Affinität* der Teilnehmenden einen signifikanten Einfluss auf das Vertrauen in die KI-gestützten Entscheidungshilfen *KIBAT* und *WIEW* hatten.

Im Gegensatz dazu zeigte die generelle *Einstellung gegenüber KI* einen signifikanten positiven Einfluss auf das Vertrauen in die Tools, was die Hypothese H_4 bestätigt: Eine positive Einstellung gegenüber KI-Technologien korrelierte mit einem höheren Vertrauen in die KI-Entscheidungshilfen. Eine optimistische und aufgeschlossene Haltung gegenüber den Potenzialen und Anwendungen von KI könnte entscheidend für das Vertrauen in entsprechende Systeme sein.

Wie im Vorfeld vermutet, zeigte sich aus den zuvor durchgeführten deskriptiven Auswertungen der digitalen Affinität und der Einstellung gegenüber KI nach Altersgruppe und Geschlecht, dass sich Personen der jüngeren Generationen gegenüber KI aufgeschlossener zeigten. Die höheren Mittelwerte der digitalen Affinität bei jüngeren Altersgruppen könnten darauf hinweisen, dass jüngere Menschen generell offener und vertrauter mit digitalen Technologien sind. Dies ist auf eine stärkere Integration digitaler Medien und Technologien in ihrem Alltag sowie auf eine grössere Affinität zu neuen Technologien zurückzuführen. Jüngere Generationen sind häufig mit digitalen Medien aufgewachsen, was zu einer höheren Nutzungskompetenz und Akzeptanz führt (Palfrey & Gasser, 2008, S. 161). Die Einstellung gegenüber KI, die ebenfalls in jüngeren Altersgruppen höhere Mittelwerte zeigt, könnte darauf hindeuten, dass jüngere Menschen optimistischer und aufgeschlossener gegenüber den Potenzialen und Anwendungen von KI-gestützten Entscheidungshilfen sind. Im Gegensatz dazu könnten die niedrigeren Mittelwerte bei älteren Altersgruppen in beiden Bereichen auf eine geringere Vertrautheit und Akzeptanz gegenüber digitalen Technologien und KI hinweisen: Ältere Generationen haben möglicherweise weniger Berührungspunkte mit diesen Technologien und könnten daher skeptischer oder vorsichtiger in ihrer Nutzung und Beurteilung sein.

In der Theorie könnten die Ergebnisse der positiven Korrelation der Einstellung mit dem Vertrauen mit dem Bestätigungsfehler (*confirmation bias*) erklärt werden. Menschen neigen dazu, Informationen und Technologien, die ihre bestehenden Überzeugungen und Einstellungen bestätigen, eher zu akzeptieren und zu vertrauen (Glaberson, 2019). Haben die User:innen bereits positive Erfahrungen mit ähnlichen KI-Tools gemacht (z. B. Berichte verfassen in ChatGPT) oder sind sie generell positiv gegenüber KI eingestellt, so sind sie möglicherweise auch entsprechenden Innovationen im Kindesschutz gegenüber positiver eingestellt als Menschen, die bisher noch weniger Berührungspunkte mit KI hatten (Sheridan, 2019). Nicht zuletzt wird die Frage nach dem Vertrauen letztendlich von den Erfahrungen, die die User:innen mit einem entsprechenden Tool machen, abhängen. Dies zeigen auch Zitate aus

der Umfrage: «Erfahrungen mit KI wären mein persönlicher Entscheidungsträger» (P5.I); «Ich gehe davon aus, dass meine Einstellung gegenüber der Nutzung von KI im beruflichen Alltag sehr stark von den gemachten Erfahrungen abhängt» (P38.I). Entsprechend scheint es zentral, dass – sofern in Zukunft entsprechende Tools in den schweizerischen Kindesschutz eingeführt werden – die Implementierung durch Sorgfalt, Transparenz und entsprechende Kommunikation begleitet wird und dabei die unterschiedlichen Bedürfnisse und Einstellungen verschiedener Altersgruppen und Geschlechter berücksichtigt werden, um ein breiteres Vertrauen für die Nutzung von KI-Technologien zu erreichen.

Ein in diesem Kontext interessantes Zitat aus den Bemerkungen: «Es ist höchste Zeit, dass auch in der sozialen Arbeit die Nutzung von digitalen Mitteln vermehrt zugelassen wird. Damit es funktioniert, müssen auch die Menschen in Leitungspositionen dahinter stehen können. Ich bin mir unsicher, ob dies schon möglich ist, oder ob es einen Generationenwechsel braucht. Die Zeit wird es zeigen.» (P55.I).

7.3 Einfluss auf die Integrationsbereitschaft

Zuletzt wurde in der Studie untersucht, ob ein höheres Vertrauen in die Einschätzungen der verschiedenen Tools zu einer höheren *Integrationsbereitschaft* in den Arbeitsalltag führt. Die ausgewerteten Daten zeigen eine signifikante Korrelation zwischen dem Vertrauen in die Tools und der Bereitschaft, diese in persönliche Entscheidungsprozesse einzubeziehen, was die Hypothese H_5 bestätigt.

Das ursprüngliche *TAM* geht davon aus, dass die Einstellung gegenüber der Nutzung einer Technologie einen direkten Einfluss auf die tatsächliche Nutzungsabsicht hat (Davis et al., 1989, S. 986). Vertrauen in die Genauigkeit und Zuverlässigkeit von KI-gestützten Entscheidungshilfen kann als eine Form der *wahrgenommenen Nützlichkeit* betrachtet werden. Wenn Sozialarbeitende glauben, dass diese Tools zuverlässige und präzise Einschätzungen liefern, steigt ihre wahrgenommene Nützlichkeit, was zu einer positiveren Einstellung gegenüber der Integration entsprechender Tools in den Arbeitsalltag führt. Diese Korrelation wird durch die Ergebnisse der Studie gestützt, die eine signifikante positive Beziehung zwischen Vertrauen und Integrationsbereitschaft zeigen. Die Sozialarbeitenden, die den Tools mehr Vertrauen entgegenbrachten, zeigten auch eine höhere Bereitschaft, diese in ihre täglichen Entscheidungsprozesse zu integrieren.

7.4 Unterschiede zwischen den beiden Tools *KIBAT* und *WIEW*

Eine Erkenntnis aus der Studie ist, dass sich die Teilnehmenden insgesamt eher zurückhaltend gegenüber beiden Tools zeigten und sich die Mittelwerte bei der Frage, ob der Einschätzung des Tools gefolgt würde, hauptsächlich im Bereich von 3 bis 4 («eher unwahrscheinlich» bis «eher wahrscheinlich») bewegten. Allerdings fand das Chancenmodell *WIEW* insgesamt mehr Anklang als das Risikomodell *KIBAT*. Die Ergebnisse zeigen, dass die Teilnehmenden den Einschätzungen von *WIEW* eher trauten als denen von *KIBAT* und sie eher bereit waren, *WIEW* im beruflichen Kontext zu verwenden, ihre eigenen Urteile zu hinterfragen und möglicherweise anzupassen. Eine Erklärung dafür könnte sein, dass *WIEW* von den Befragten als positiver und unterstützender wahrgenommen wird. In komplexen Fallverläufen, in denen wenig Klarheit über Handlungsoptionen besteht, wird *WIEW* möglicherweise als

zusätzlicher Ideeninput oder als hilfreiche Zweitmeinung gesehen und weniger als Konkurrenz zur eigenen Einschätzung wahrgenommen. Eine Person äusserte sich dazu: «(…) das ist nicht viel anders, als wenn wir eine KI um eine Zweitmeinung fragen. Vielleicht gibt uns die KI ja tatsächlich neue Ideen, die wir weiterverfolgen könnten.» (P17.3).

Im Gegensatz dazu könnte *KIBAT* als Risikomodell eher als Bedrohung wahrgenommen werden. Eine Person meinte hierzu: «Ich kann mir vorstellen, dass KI bei Sozialarbeitenden Ängste und Verunsicherungen auslösen könnte, wenn suggeriert würde, dass die KI die besseren/treffenderen Einschätzungen vornehmen könnte als die Menschen. Es könnte ein Gefühl der eigenen Überflüssigkeit ausgelöst werden, sind Einschätzungen und Entscheidungen prägende Elemente des Arbeitsalltags.» (P3.3). Diese Aussage deutet darauf hin, dass KI als Konkurrenz zur eigenen professionellen Kompetenz gesehen werden könnte, was das Vertrauen in ein Tool wie *KIBAT* vermindern könnte. Diese Reaktionen könnten möglicherweise auf eine grundsätzliche Skepsis gegenüber externen Einschätzungen zurückzuführen sein, die die eigene Expertise infrage stellen.

Zusammenfassend lässt sich vermuten, dass *WIEW* aufgrund seiner positiven und unterstützenden Natur eher akzeptiert wird, während *KIBAT* als potenziell bedrohlich und kompetenzgefährdend wahrgenommen wird. Diese Interpretation ist jedoch rein spekulativ und bedarf weiterführender Studien, um die darunterliegenden Wirkmechanismen besser zu verstehen.

7.5 Weitere Erkenntnisse aus der Umfrage

Am Ende des Fragebogens hatten die Teilnehmenden die Möglichkeit, ihre Gedanken zu der soeben ausgefüllten Befragung zu hinterlassen. Dabei wurde deutlich, dass das Thema polarisiert und verschiedene Aspekte, die nicht Teil dieser Thesis waren, von den Teilnehmenden als relevant erachtet wurden. Diese Themen werden daher im Folgenden näher betrachtet.

Notwendigkeit der Berücksichtigung individueller Faktoren

Die Anwendung von KI im Kindesschutz bietet grosses Potenzial zur Unterstützung von Sozialarbeitenden, ruft jedoch offensichtlich bei den befragten Personen insbesondere hinsichtlich der Berücksichtigung individueller Einzelfälle Skepsis hervor. Eine häufig genannte Unsicherheit gegenüber KI ist die Sorge, dass das System nicht die Feinheiten und Nuancen eines einzelnen Falles berücksichtigen kann. Eine Person bemerkte: «Das System der KI ist nur so gut wie die Eingabe der Daten. 1000 Datenreihen sagen noch nichts über einen individuellen Einzelfall aus.» (P16.1). Eine weitere Person äusserte sich dazu: «Gewisse Lösungen der KI mögen vielleicht statistisch gesehen treffsicherer sein, können aber aus individuellen Gründen evtl. genau für diese Situation/Person nicht realitätsgetreu umgesetzt werden aufgrund sehr individuellen Ausgangslagen.».(P25.1). Entsprechende Aussagen zeigen die vorhandene Befürchtung, dass prädiktive Modelle zwar grosse Mengen an Daten analysieren können, aber die spezifischen, einzigartigen Aspekte einer Situation übersehen können. Zudem besteht die Sorge, dass die Entscheidungen der KI aufgrund ungenauer Daten oder algorithmischen Verzerrungen zu Fehlentscheidungen führen könnten. Ein weiteres Zitat veranschaulicht diese Bedenken: «Wo bitte bleibt die Ausnahme? Der Einsatz von KI kann (…) zu verehrenden Fehlern führen.» (P16.2). Diese Befürchtungen spiegeln die von Keddell (2014) und Gillingham (2019) geäusserten Bedenken wider, dass prädiktive Modelle möglicherweise nicht in der Lage sind,

die Komplexität und Individualität jedes einzelnen Falls zu erfassen. Diese Überlegungen verdeutlichen die Notwendigkeit, bei einer Implementierung von KI im Kindesschutz nicht nur auf die Datenqualität, sondern auch auf die Einzigartigkeit und Komplexität individueller Fälle zu berücksichtigen und mögliche Datenverzerrungen zu minimieren.

Beziehungsebene und zwischenmenschliche Interaktionen

Ein weiterer Punkt, der in den Bemerkungen häufig genannt wurde, ist die Bedeutung der Beziehungsebene und zwischenmenschlichen Interaktionen im Kindesschutz. Viele befragte Personen äusserten die Sorge, dass KI nicht in der Lage sei, die emotionalen und sozialen Aspekte der Arbeit mit Kindern und Familien zu ersetzen. Eine Person bemerkte «KI wird nicht den persönlichen Kontakt, die Beziehungsgestaltung mit den betroffenen Kindern und Eltern ersetzen können.» (P8.1). Eine weitere Person äusserte sich dazu: «Soziale Arbeit ist Beziehungsarbeit, etwas das KI nicht kann. Beziehungsarbeit benötigt Zeit, Durchhaltevermögen und Rückgrat und vor allem die Liebe zu Menschen. Auch das kann KI nicht.» (P37.2). Diese Aussagen verdeutlichen die Befürchtung, dass KI die wesentlichen zwischenmenschlichen Aspekte, die für den Erfolg in der Sozialen Arbeit entscheidend sind, nicht abbilden kann. Soziale Interaktionen und emotionale Reaktionen spielen eine wesentliche Rolle in der Arbeit mit Menschen, die durch Algorithmen nicht vollständig erfasst werden können (Enosh & Bayer-Topilsky, 2015). Ein weiteres Zitat unterstreicht diese Bedenken: «Ich gehe davon aus, dass menschliches Handeln auch weiterhin nicht fassbar ist, da es irrational ist und statistisch basierte Entscheidungen dem daher nur bedingt Rechnung tragen können. Interventionen in der Sozialen Arbeit sind häufig gekoppelt mit der zwischenmenschlichen Beziehung zwischen Klient*In und Sozialarbeitenden.» (P11.2).

Hierbei kann jedoch angemerkt werden, dass zwischenmenschliche Interaktionen auch immer menschliche kognitive Verzerrungen mit sich bringen, die immer einen Einfluss auf Entscheidungen haben können (Glaberson, 2019). In diesem Zusammenhang könnte die KI möglicherweise gar helfen, solche Verzerrungen zu minimieren, indem sie konsistente, datenbasierte Analysen liefert. Dies bedeutet, dass das Argument, dass KI keine Beziehungsarbeit leistet, nicht vollständig greift, da menschliche Entscheidungsprozesse ebenfalls ihre eigenen Herausforderungen und Begrenzungen mit sich bringen. Es könnte argumentiert werden, dass es nicht der Anspruch der KI sein muss, zwischenmenschliche Interaktionen zu verdrängen. Vielmehr könnte KI als ergänzendes Werkzeug betrachtet werden, das die analytischen und administrativen Aufgaben der Sozialarbeitenden erleichtert, während diese weiterhin die wesentlichen zwischenmenschlichen Interaktionen pflegen. Die Kombination von KI-gestützten Tools und menschlicher Empathie könnte somit eine Koexistenz schaffen, in der beide Elemente zusammenarbeiten, um bestmögliche Ergebnisse zu erzielen. Eine Person hierzu: «Ich hoffe, dass KI basierte Tool administrative Arbeiten erleichtern und somit der Fokus vermehrt auf die zwischenmenschlichen Interaktionen gelegt wird.» (P11.1).

Gefahr der Überreliance und Verlust menschlicher Urteilsfähigkeit vs. erhöhte Handlungssicherheit

Ein weiterer häufig angesprochener Aspekt in den Bemerkungen ist die Besorgnis über eine übermässige Abhängigkeit von KI und den daraus resultierenden Verlust der menschlichen Urteilskraft im Kindesschutz. Verschiedene befragte Personen äusserten die Sorge, dass eine zu starke Abhängigkeit von KI-gestützten Entscheidungshilfen das professionelle Urteilsvermögen der Sozialarbeitenden beeinträchtigen könnte. Eine Person bemerkte: «KI und

soziale Arbeit passen nicht zusammen. Sie führt dazu, weniger selber zu denken und nur ein geringer Teil würde KI tatsächlich dafür nutzen, um eigene Überlegungen zu reflektieren.» (P10.I). Eine weitere Person äusserte sich dazu: «verwende aktuell KI punktuell als Argumentensammlung bsp. was sind positives Effekte eine Boxtraining für ein Jugendlicher oder welche Einfluss haben Konflikte zwischen KE auf die Kinder Vorteil, man erhält schnell einen Überblick / Nachteil, man muss weniger selbst nachdenken und wird bequem» (P46.I). Diese Aussagen verdeutlichen die Befürchtung, dass eine Überreliance auf KI die Fähigkeit der Sozialarbeitenden zur eigenständigen Entscheidungsfindung und Reflexion untergraben könnte.

Entscheidungen in der Sozialen Arbeit erfordern sowohl intuitive als auch analytische Komponenten, um den komplexen und oft unvorhersehbaren menschlichen Verhaltensweisen gerecht zu werden (Nitzsch, 2021, S. 5). Während KI-gestützte Tools durch konsistente, datenbasierte Analysen unterstützen können, besteht die Gefahr, dass Sozialarbeitende ihre eigenen Einschätzungen und Erfahrungen weniger berücksichtigen und sich zu stark auf die Algorithmen verlassen. Dies kann zu einer risikoscheuen Haltung führen, bei der eigene Einschätzungen zugunsten der algorithmischen Werte vernachlässigt werden (Eubanks, 2018, S. 124).

Es könnte jedoch argumentiert werden, dass eine sorgfältige Implementierung und Nutzung von KI-Tools dazu beitragen könnte, das professionelle Urteilsvermögen zu unterstützen, anstatt es zu ersetzen. KI kann als zusätzliches Werkzeug dienen, das Sozialarbeitenden ermöglicht, ihre eigenen Einschätzungen zu hinterfragen und zu validieren (Schneider, 2021, S. 130). Eine Person hierzu: «KI würde ich als Tool nutzen für Handlungen in bestimmten komplexen Themen/Fällen. KI könnte evtl. meine Sichtweise anders beurteilen oder blinde Flecken aufzeigen.» (P52.I). Eine weitere Person bemerkt: «KI wäre für mich ein Tool um Einschätzungen breiter / vertiefter vorzunehmen. KI würde mich ggf. auffordern gewisse Entscheid nochmals zu reflektieren, hinterfragen oder um mich im Entscheid zu bestärken. Das Tool ergänzt meine Arbeit, ersetzt sie aber bei weitem nicht (P19.I). Die Kombination von KI-gestützten Tools und menschlicher Expertise könnte somit eine Synergie schaffen, in der beide Elemente zusammenarbeiten, um zukünftig fundierte und reflektierte Entscheidungen im Kindesschutz zu treffen und damit die Handlungssicherheit zu erhöhen.

Ethik und Verantwortung

Ein weiterer Punkt in den Bemerkungen stellt die Besorgnis über die ethischen Implikationen und die Fragen nach Verantwortung bei der Nutzung von KI im Kindesschutz dar. Verschiedene Personen äusserten die Sorge, dass die Implementierung von KI ethische Fragen aufwirft und die Verantwortung für getroffene Entscheidungen unklar wird. Eine Person bemerkte: «Ich finde es schwierig, wenn wir unsere Entscheidungen in diesen sehr einschneidenden Kindesschutzthemen an einen Wahrscheinlichkeitsberechner übergeben. (…) Wer übernimmt dann die Verantwortung, wenn etwas schiefgeht?» (P44.I). Eine weitere Person äusserte sich dazu: «Wichtig ist, dass wir dabei unabhängig bleiben und die Entscheidung, egal welche muss immer Fachlich begründet sein. Keine Lösung ist: die KI hat entschieden.» (P17.3). und eine weitere Person äussert sich zum Thema: «Künstliche Intelligenz ist etwas wunderbares, solange diese nicht die Verantwortung der eigenen Entscheidungen abnimmt oder gar rechtfertig.» (P37.I). Die Aussagen verdeutliche die Befürchtung, dass die Nutzung von KI möglicherweise zu einer Verschleierung der Verantwortlichkeiten führen könnte.

Die Verwendung von KI im Kindesschutz erfordert strenge ethische Rahmenbedingungen, um sicherzustellen, dass die Technologie verantwortungsvoll eingesetzt wird. Ein Zitat hierzu: «Nicht berücksichtigt ist, wie die Rahmenbedingungen gestaltet würden: inwieweit werden Mitarbeitende verpflichtet, sich an die Empfehlungen gemäss KI zu halten?, bei wem liegt die Verantwortung?» (P5.2). Diese Bedenken spiegeln die Notwendigkeit wider, klare Richtlinien und Verantwortlichkeiten innerhalb der Organisationen zu implementieren, um sicherzustellen, dass die Nutzung von KI transparent und ethisch vertretbar ist (Kreutzer, 2023, S. 69). Das *Allegheny Family Screening Tool* (*AFST*) bietet ein Beispiel dafür, wie klare Rahmenbedingungen und Verantwortlichkeiten in bestehenden Tools im Kindesschutz implementiert werden können. Das *AFST* hat gezeigt, dass durch die Einführung strenger Richtlinien und klarer Verantwortlichkeiten die Identifizierung von Kindern, die weitere Kindesschutzinterventionen benötigen, verbessert werden kann, ohne die Arbeitsbelastung der Sozialarbeitenden zu erhöhen (Goldhaber-Fieber & Prince, 2019). Die klaren Regeln, wann eine Abklärung weiterverfolgt wird, sorgen dafür, dass die Nutzung des Tools transparent und nachvollziehbar ist und die Entscheidungen weiterhin von den Sozialarbeitenden und ihren Vorgesetzten verantwortet werden. Bei der Einführung neuer KI-gestützter Tools im Kindesschutz scheint es demnach wichtig, dass klare ethische Richtlinien und Verantwortlichkeiten festgelegt werden. Dies stellt sicher, dass alle Beteiligten die Technologie verantwortungsvoll und transparent nutzen.

Notwendigkeit des Einsatzes von KI in der Sozialen Arbeit

Abschliessend wurde in den Bemerkungen auch mehrfach darauf hingewiesen, dass die Auseinandersetzung mit KI in der Sozialen Arbeit nicht nur als positiv, sondern auch als notwendig betrachtet wird. Eine Person hierzu: «Ich danke für die Auseinandersetzung mit dem KI. Ich denke, es gibt kein Berufszweig, welcher sich nicht mit dem Thema auseinandersetzen muss» (P34.I). Eine andere Person äusserte sich: «Ich denke dies ist der Wandel der Zeit und auch wir Sozis dürfen/müssen/sollen uns mit KI auseinandersetzen.» (P36.I). Eine weitere Person griff den Fachkräftemangel auf, der möglicherweise durch die Entlastung durch KI angegangen werden könnte: «Es besteht die Hoffnung, dass [durch den Einsatz von KI] der Fachkräftemangel reduziert werden könnte.». (P9.I).

Um eine erfolgreiche Implementierung von KI im Arbeitsfeld des Kindesschutzes zu ermöglichen, scheint eine sorgfältige Auseinandersetzung mit der Einführung, Schulung und Zusammenarbeit zwischen Entwickler:innen und Sozialarbeitenden unumgänglich: «Die Anwendung der KI muss gut eingeführt werden, um den richtigen Nutzen daraus erzielen zu können.» (P41.I).

Es lässt sich also feststellen, dass die Einführung von KI in der Sozialarbeit trotz berechtigter Bedenken viele positive Aspekte bietet. Durch klare ethische Rahmenbedingungen, transparente Algorithmen und eine verantwortungsvolle Nutzung kann KI eine wertvolle Unterstützung für Sozialarbeitende sein, die sowohl die Effizienz als auch die Qualität der Entscheidungen verbessert. Der Einsatz von KI stellt eine bedeutende Chance dar, die Soziale Arbeit zukunftsfähig zu gestalten und den Fachkräftemangel effektiv anzugehen.

7.6 Kritische Reflexion und Limitationen

Die vorliegende Studie weist mehrere Limitationen auf, die die Interpretation und Generalisierbarkeit der Ergebnisse einschränken.

Eine der wichtigsten Einschränkungen ergibt sich aus der Stichprobe. Auch nach Ausweitung der Stichprobenpopulation konnte nur knapp eine ausreichend grosse Stichprobengrösse erreicht werden. Obwohl keine genauen Zahlen darüber vorliegen, wie viele Beistandspersonen im Kindesschutz pro Kanton tätig sind, wird die Teilnahme von 204 Personen dennoch als vergleichsweise hoch, jedoch nicht als repräsentativ betrachtet. Die Zielpopulation dieser Arbeit umfasste Beistandspersonen im gesetzlichen Kindesschutz in verschiedenen deutschsprachigen Kantonen der Schweiz, deren Zuständigkeiten und Aufgabenbereiche sich aufgrund der föderalistischen Struktur der Schweiz von Kanton zu Kanton unterscheiden. Nicht alle Beistandspersonen haben die gleichen Aufgaben, und in einigen Kantonen werden Kindeswohlabklärungen nicht von Beistandspersonen, sondern von anderen Institutionen wie der Kindes- und Erwachsenenschutzbehörde (KESB) durchgeführt. Dies könnte dazu geführt haben, dass insbesondere das Fallbeispiel des Abklärungstools *KIBAT* nicht für alle Personen gleichermassen beurteilbar war und hätte bei der Erstellung der Fallbeispiele stärker berücksichtigt werden müssen.

Die Erhebung beschränkte sich auf Sozialarbeitende im Kindesschutz, was eine relativ homogene Gruppe darstellt. Eine Ausweitung auf andere Berufszweige wurde bewusst vermieden, da dies im Rahmen einer Masterthesis kaum umsetzbar gewesen wäre. Es kann davon ausgegangen werden, dass Sozialarbeitende tendenziell politisch eher links eingestellt sind (Kindler, 2019). Studien zeigen, dass politisch eher links eingestellte Personen technologischen Veränderungen und algorithmischen Entscheidungsträgern möglicherweise aufgeschlossener gegenüberstehen (Lee, 2018). Die Ergebnisse der Studie müssten daher im Vergleich mit anderen Populationen, die politisch weniger links stehen, bewertet werden. Zudem lassen sich die Erkenntnisse nicht ohne Weiteres auf andere Bereiche der Sozialen Arbeit übertragen, da die konstruierten Entscheidungssituationen ausschliesslich potenzielle Szenarien im Kindesschutz beschreiben. Um die Ergebnisse auf andere Berufsfelder zu übertragen, muss sichergestellt werden, dass vergleichbare Rahmenbedingungen in den jeweiligen Arbeitsfeldern gegeben sind.

Die Studie stützt sich auf ein multifaktorielles, experimentelles Vignettendesign, das hypothetische Szenarien verwendet. Obwohl diese Methode in den Sozialwissenschaften häufig genutzt wird, um die Wahrnehmung von Entscheidungen zu analysieren (Rost, 2018), bedarf es weiterer Forschung, die reale Erfahrungen von Personen einbezieht, die tatsächlich mit KI-gestützten Entscheidungshilfen arbeiten. Die Ergebnisse dieser Arbeit könnten sich ändern, wenn Sozialarbeitende tatsächlich in ihrer täglichen Praxis mit entsprechenden Tools konfrontiert wären. Ein weiterer Punkt betrifft die konstruierten Fallbeispiele, die nur begrenzte Kontextinformationen bieten. Unterschiede in der menschlichen Wahrnehmung können auftreten, und der bereitgestellte Kontext könnte die Teilnehmenden entsprechend beeinflussen (Rost, 2018). Es ist unklar, ob und wie stark die Teilnehmenden sich bereits mit ähnlichen KI-gestützten Tools auseinandergesetzt haben. Zudem wurden die Begriffe «KI» und «Algorithmus» den Befragten als eine Art «Black Box» präsentiert, ohne genaue Erläuterung, wie bzw. mit welchen Daten die neuronalen Netzwerke der entsprechenden Tools trainiert sind. Dies liess viel Interpretationsspielraum und damit zusammenhängend möglicherweise auch Unsicherheiten für die Befragten offen.

Die Befragung wurde zu einem Zeitpunkt durchgeführt, zu welchem in der Schweiz im Bereich des Kindesschutzes noch keine vergleichbaren Tools in Verwendung sind. Zudem befinden wir uns in einer sich rasch wandelnden Gesellschaft, die von Informationsüberflutung und gleichzeitig von einem hohen Grad an Unwissenheit über KI und ihre Variationen gekennzeichnet ist. Aufgrund dieser Umstände könnten die Resultate bei einer Wiederholung der Umfrage in einigen Jahren stark variieren. Eine erneute Durchführung der Studie zu einem späteren Zeitpunkt könnte daher angebracht sein.

Im Weiteren weisen die Studie und die Datenerhebung methodische Limitationen auf. Obwohl eine Randomisierung der Zuweisung zu den Vignetten angestrebt war, funktionierte dies aus technischen Gründen nur bedingt und einige Vignetten wurden viel häufiger zugeteilt als andere. Dies verhinderte die geplanten Gegenüberstellungen der einzelnen Vignetten. Zudem wurde keine Kontrollgruppe eingesetzt, wodurch die Vergleichbarkeit der Ergebnisse eingeschränkt ist (Döring & Bortz, 2016, S. 75). Die Manipulation der unabhängigen Variablen wurde zwar durchgeführt, funktionierte jedoch nicht wie beabsichtigt. Eine Schwierigkeit könnte in der undeutlichen Formulierung der Texte gelegen haben: Wenn die Vignetten nicht klar und präzise formuliert waren, könnte dies dazu geführt haben, dass die Teilnehmenden die intendierten Manipulationen der Dimensionen nicht klar erkannten. Ein weiterer Faktor könnten die knappen Zeitressourcen der Befragten gewesen sein. Zeitdruck kann das sorgfältige Lesen und das Verständnis komplexer Informationen beeinträchtigen (Rost, 2018, S. 99), was besonders kritisch ist, wenn von den Befragten erwartet wird, subtile Manipulationen in einem experimentellen Design zu erkennen. Die Kombination aus möglicherweise unklarer Textmodule und limitierten Zeitressourcen der Befragten könnte die Effektivität des Manipulations-Checks beeinträchtigt haben und müssten bei einer weiterführenden Untersuchung gezielter adressiert werden. Ein Grossteil der Antworten wurde aufgrund nicht bestandenen Manipulationschecks ausgeschlossen, um aussagekräftige Ergebnisse zu erhalten. Dies führte einerseits zu einer massiv geringeren Teststärke (Power), aber auch zu einer Verzerrung der Randomisierung, sodass keine echte Randomisierung mehr gegeben war. Obwohl eine multiple Regression durchgeführt wurde und potenzielle Kausalbeziehungen identifiziert werden konnten, sollten diese Ergebnisse aufgrund der Probleme bei der Manipulation und des Ausschlusses vieler Antworten mit Vorsicht interpretiert werden.

Nicht zuletzt wurden zahlreiche Drittvariablen erst im Nachgang identifiziert, basierend auf den Bemerkungen der Teilnehmenden. Diese späte Identifikation könnte die Interpretation der Ergebnisse weiter beeinflussen. Gleichzeitig wurden viele der erhobenen Items nicht in der Auswertung verwendet, da die Zeit nicht für die vollständige Auswertung ausreichte und daher auf die wesentlichen Hypothesen und deren Variablen fokussiert wurde.

Insgesamt verdeutlichen diese Limitationen die Notwendigkeit methodisch robusterer Studien, um die gewonnenen Erkenntnisse zu überprüfen und zu erweitern. Zukünftige Forschung sollte grössere Stichproben und verbesserte Rahmenbedingungen berücksichtigen, um die Generalisierbarkeit und Validität der Ergebnisse zu erhöhen. Anpassungen im Studiendesign, wie klarere Vignetten und eine funktionierende Randomisierung, sind notwendig, um die Aussagekraft zukünftiger Untersuchungen zu maximieren.

8 Fazit

Im Mittelpunkt dieser Arbeit stand die Untersuchung, wie unterschiedliche Einflussfaktoren das Vertrauen von Sozialarbeitenden im Kindesschutz in KI-gestützte Entscheidungshilfen beeinflussen und in welchem Ausmass sie bereit sind, solche Tools in ihre täglichen Arbeitsabläufe zu integrieren. Um die Forschungsfrage zu beantworten, wurden mehrere Hypothesen aufgestellt und überprüft.

8.1 Beantwortung der Fragestellung

Die Untersuchung ergab, dass es für das Vertrauen in die KI-gestützten Entscheidungshilfen relevant zu sein scheint, wie die KI-Tools wahrgenommen werden. Der Faktor *Invasivität*, also wie stark das Tool in die Entscheidungsprozesse eingreift, zeigte einen signifikanten Einfluss auf das Vertrauen: Milderen Massnahmen, die weniger invasiv sind, wurde tendenziell mehr Vertrauen entgegengebracht. Ein weiterer entscheidender Faktor ist die *prädiktive Validität* der Tools. Eine hohe wahrgenommene Genauigkeit und Zuverlässigkeit der Vorhersagen korreliert positiv mit dem Vertrauen in die Tools. Die Sozialarbeitenden waren eher bereit, einer Einschätzung des Tools zu folgen, wenn sie davon ausgingen, dass die prädiktive Validität hoch ist. Im Gegensatz dazu zeigte der Faktor *wissenschaftliches Wissen* keinen signifikanten Einfluss auf das Vertrauen. Dies deutet darauf hin, dass die Teilnehmenden möglicherweise nicht in der Lage sind, den wissenschaftlichen und technischen Hintergrund der Tools zu bewerten, oder dass dieser Aspekt für ihr Vertrauen nicht ausschlaggebend ist.

Personenmerkmale wie der *Expert:innenstatus* und die *digitale Affinität* der Teilnehmenden zeigten ebenfalls keinen signifikanten Einfluss auf das Vertrauen in die Tools. Allerdings zeigte die generelle *Einstellung gegenüber KI* einen Einfluss: Eine positive Einstellung gegenüber KI korrelierte stark mit einem höheren Vertrauen in die Entscheidungshilfen. Dies verdeutlicht, dass die generelle Offenheit und die bisherigen Erfahrungen der Sozialarbeitenden mit neuartigen Technologien eine Rolle spielen.

Die Untersuchung zeigt schliesslich, dass das Vertrauen der Sozialarbeitenden in KI-gestützte Entscheidungshilfen von verschiedenen Faktoren beeinflusst wird. Ein höheres Vertrauen in die Tools korreliert schlussendlich positiv mit einer höheren Bereitschaft, diese in persönliche Entscheidungsprozesse zu integrieren. Besonders auffällig ist, dass das fiktive Chancenmodell *WIEW* bei den Teilnehmenden auf mehr Vertrauen stiess als das Risikomodell *KIBAT*. Eine mögliche Erklärung hierfür ist, dass *WIEW* als positiver und hilfreicher wahrgenommen wird, da es zusätzliche Handlungsoptionen aufzeigt und weniger als Konkurrenz zur eigenen Einschätzung betrachtet wird.

Diese Ergebnisse liefern erste Hinweise darauf, wie KI-gestützte Entscheidungshilfen im Kindesschutz wahrgenommen werden und welche Faktoren für eine erfolgreiche Implementierung in Zukunft mitberücksichtigt werden sollten. In der Untersuchung zeigte sich aber ebenfalls, dass weitere Faktoren für das Vertrauen von Bedeutung sind, welche in dieser Studie nicht näher untersucht wurden.

8.2 Implikationen für die Soziale Arbeit

Aus den Ergebnissen dieser Masterthesis lassen sich verschiedene Implikationen für die zukünftige Implementierung von KI-gestützten Entscheidungshilfen im Bereich der Sozialen Arbeit, insbesondere im Kindesschutz ableiten. Die nachfolgenden Handlungsempfehlungen fassen zusammen, welche Massnahmen bei der Einführung und bei der Nutzung entsprechender Unterstützungstools beachtet werden sollte, um ein möglichst hohes Vertrauen und damit auch eine Integrationsbereitschaft der Sozialarbeitenden zu schaffen.

1 **Sorgfältige Einführung und Zusammenarbeit bei der Entwicklung zukünftiger Modelle**

Die Einführung von KI-gestützten Tools sollte schrittweise und gut geplant erfolgen. Pilotprojekte können dabei helfen, erste Erfahrungen zu sammeln und eventuelle Probleme frühzeitig zu identifizieren und zu beheben. Entscheidend scheint, dass Sozialarbeitende in die Prozesse eingebunden werden, um damit Akzeptanz und Vertrauen zu gewinnen. Feedbackschleifen und iterative Entwicklungsprozesse sollten genutzt werden, um die Tools kontinuierlich zu verbessern. Eine enge Zusammenarbeit zwischen Sozialarbeitenden und Entwickler:innen der KI-Tools, aber auch den betroffenen Klient:innen, ist dabei unerlässlich. Durch regelmässige Workshops und Meetings können die spezifischen Bedürfnisse und Herausforderungen des Kindesschutzes berücksichtigt werden, sodass die Tools optimal auf die Praxis abgestimmt sind. Sozialarbeitende sollten die Möglichkeit haben, aktiv an der Entwicklung mitzuwirken und ihre Expertise einzubringen.

2 **Schulung und Kommunikation des Mehrwerts durch prädiktive Validität**

Es ist wesentlich, dass Sozialarbeitende umfassend geschult werden, um die Funktionsweise von KI-Tools zu verstehen. Das Verständnis der Unterschiede zwischen generischen KI-Systemen wie ChatGPT und spezifischen neuronalen Netzwerken, die mit Daten aus dem Kindesschutz trainiert wurden, scheint zentral. Schulungen sollten sowohl technische Aspekte als auch ethische Überlegungen abdecken und Raum für Fragen und Unsicherheiten bieten. Gleichzeitig muss der Nutzen der KI-Tools klar kommuniziert werden. Sozialarbeitende sollten erkennen, dass ihre Arbeit durch die hohe prädiktive Validität der Tools profitiert und ihre Entscheidungen dadurch fundierter werden können. Erfolgsgeschichten und Fallstudien, die die Wirksamkeit der Tools belegen, könnten dazu beitragen, das Vertrauen zu erhöhen und die Akzeptanz zu fördern.

3 **Koexistenz zwischen Sozialarbeitenden und KI**

KI sollte nicht die professionelle Urteilsfähigkeit der Sozialarbeitenden ersetzen, sondern als unterstützendes Werkzeug betrachtet werden. Diese Einstellung fördert die Handlungssicherheit und ermöglicht eine bessere Reflexion der Entscheidungen. Durch die Kombination von menschlicher Empathie und technischer Präzision kann die Qualität der Entscheidungen erheblich verbessert werden. Es ist entscheidend, dass diese Perspektive aktiv beworben und kommuniziert wird. Die Implementierung von KI-Tools sollte stets als Ergänzung zur menschlichen Expertise dargestellt werden, um Ängste vor einer möglichen Ersetzung oder «De-Professionalisierung» zu minimieren und das Vertrauen in die Technologie zu stärken.

Sozialarbeitende sollten ermutigt werden, die KI als Partnerin zu sehen, die ihnen hilft, ihre Arbeit effektiver und präziser zu gestalten, ohne dabei ihre eigene Rolle und Bedeutung zu mindern.

4 Achtsamkeit bezüglich ethischer Aspekte und Rahmenbedingungen

Es ist wichtig, dass bei einer Einführung klare ethische Richtlinien und Verantwortlichkeiten innerhalb der jeweiligen Organisation festgelegt werden. Prozessabläufe und Verantwortlichkeiten für Entscheidungen, die im Zusammenspiel von Mensch und KI getroffen wurden, sollte geklärt sein. Gesetzliche Richtlinien und interne Handlungsanweisungen können dabei sicherstellen, dass die Verwendung von KI transparent ist und dass der Datenschutz und die Rechte der Betroffenen gewahrt bleiben.

5 Integration von KI-Kompetenzen in das Hochschulcurriculum

Bereits in der Ausbildung von Sozialarbeitenden sollte der Umgang mit KI und deren Unterstützungsmöglichkeiten thematisiert werden. Ausbildungsinstitutionen sollten erwägen, spezielle Module einzuführen, die sich mit den Potenzialen und Grenzen von KI-Technologien auseinandersetzen. Dies könnte dabei unterstützen, zukünftige Fachpersonen mit den Möglichkeiten und Herausforderungen vertraut zu machen, die KI in der Sozialen Arbeit bieten kann. Gleichzeitig sollten diese Module den Studierenden auch Werkzeuge an die Hand geben, die es ihnen ermöglichen, KI-Systeme kritisch zu hinterfragen und ihre Anwendung reflektiert zu beurteilen. Dadurch wird sichergestellt, dass Sozialarbeitende nicht nur die Vorteile erkennen, sondern auch in der Lage sind, potenzielle Risiken und ethische Bedenken angemessen zu bewerten und verantwortungsbewusst mit der Technologie umzugehen.

6 Integration von KI durch generationenübergreifende Zusammenarbeit

Jüngere Generationen zeigen sich tendenziell offener gegenüber neuen Technologien. Diese Offenheit sollte genutzt werden, um frühzeitig positive Erfahrungen mit KI-gestützten Tools zu schaffen. Um die Integration von KI in den Arbeitsalltag zu fördern, könnten Mentor:innenprogramme oder «KI-Ambassadors» eingeführt werden. In solchen Programmen könnten beispielsweise technikaffine jüngere Mitarbeitende gezielt geschult und als Multiplikatoren eingesetzt werden, um ihr Wissen und ihre positiven Erfahrungen mit KI-Tools an ältere Mitarbeitende weiterzugeben.

7 Bewältigung des Fachkräftemangels

Nicht zuletzt könnte der Einsatz von KI dabei helfen, den Fachkräftemangel in der Sozialen Arbeit zu lindern, indem Aufgaben technisch unterstützt und die Effizienz der Sozialarbeitenden damit gesteigert wird. Dies würde den Sozialarbeitenden mehr Zeit für zwischenmenschliche Klient:inneninteraktionen geben und damit möglicherweise auch die Dossierbetreuung verbessern. Die Automatisierung administrativer Aufgaben mithilfe von KI könnte die Arbeitsbelastung zudem reduzieren und zu einer höheren Arbeitszufriedenheit beitragen.

Durch die Berücksichtigung dieser Aspekte können KI-gestützte Tools erfolgreich in die Soziale Arbeit integriert werden und die Effizienz und Qualität der Entscheidungen im Kindesschutz zu verbessern.

8.3 Schlusswort

Es ist nur eine Frage der Zeit, bis KI in der Sozialen Arbeit Einzug hält. Die Implementierung entsprechender Tools bietet Sozialarbeitenden die Möglichkeit, Prozesse effizienter zu gestalten und datengestützte, objektivere und treffsicherere Entscheidungen zu treffen. Damit bietet KI ein enormes Potenzial, nicht nur Arbeitsprozesse innerhalb der Sozialen Arbeit, sondern auch das Professionsverständnis massgeblich zu verändern. Dabei wird es notwendig sein, Soziale Arbeit «neu» zu denken. Denn nicht nur ein tiefgehendes Verständnis der vielschichtigen Auswirkungen der Digitalisierung auf unterschiedliche soziale Ebenen ist dabei erforderlich, sondern auch eine Reflexion über die eigene berufliche Rolle (Schulze, 2019, S. 387).

Obwohl KI-gestützte Entscheidungen bereits heute als treffsicherer gelten (Bastian & Schrödter, 2019), begegnet ihnen aus der sozialarbeiterischen Praxis Skepsis hinsichtlich ihres probabilistischen Charakters und der Anwendbarkeit auf individuelle Fälle (Schneider, 2021, S. 129). Jedoch basieren auch herkömmliche Methoden im Kern auf Wahrscheinlichkeiten: Sozialarbeitende treffen ihre (menschlichen) Entscheidungen, indem sie zukünftige Entwicklungen einschätzen und diese Einschätzungen mit ähnlichen Fällen vergleichen. Die Kritik an der Klassifizierung durch KI-gestützte Systeme erscheint vor diesem Hintergrund nicht gerechtfertigt. Vielmehr wird deutlich, dass sowohl traditionelle als auch algorithmische Ansätze ein Spektrum bilden, das in Zukunft die Sicherheit bei Entscheidungen erhöht und den professionellen Spielraum durch zusätzliche Reflexion besser nutzt (Schneider, 2021, S. 129-131).

Um die Implementierung entsprechender Tools in die Praxis des Kindesschutz erfolgreich zu gestalten, ist das Vertrauen in entsprechende Applikationen zentral. Dieses hängt von verschiedenen Faktoren ab und ist nicht zuletzt abhängig von einer sorgfältigen und transparenten Einführung in das Feld. Durch die Kombination von menschlicher Expertise und technologischer Unterstützung kann die Qualität des Kindesschutzes verbessert und die Handlungssicherheit der Sozialarbeitenden erhöht werden. In der Sozialen Arbeit ist dabei nicht von einer vollständig autonomen KI auszugehen. Vielmehr handelt es sich um ein Zusammenspiel von Mensch und KI, bei dem die KI als Assistenz dient und Empfehlungen vorschlägt, die letztendlich von einem Menschen getroffen werden. Klare Rahmenbedingungen innerhalb der Organisation zur Anwendung von KI werden dabei vorausgesetzt.

Um auf den Ursprung dieser Arbeit zurückzukommen: Um eine Vereinfachung des Lebens zu erreichen – oder in diesem Kontext einen objektiveren und treffsicheren Kindesschutz – ist Vertrauen ein zentraler Erfolgsbaustein (Luhmann, 2009, S. 27). Sozialarbeitende müssen darauf vertrauen können, dass KI-gestützte Tools ihre Arbeit unterstützen und nicht ersetzen, dass sie ihre professionellen Entscheidungen fundieren und nicht in Frage stellen. Dieses Vertrauen kann nur durch transparente Kommunikation, fortlaufende Schulungen und positive Erfahrungen im Umgang mit den Tools aufgebaut werden.

Zusätzlich erscheint es wichtig, dass die Einführung dieser Technologien kontinuierlich durch wissenschaftliche Forschung begleitet wird. Es gilt, die genauen Faktoren, die das Vertrauen der Sozialarbeitenden in KI-gestützte

Tools beeinflussen, noch besser zu verstehen. Zukünftige Studien sollten sich darauf konzentrieren, die Wirkmechanismen dieser Technologien in der Praxis zu beleuchten, um mögliche Hürden frühzeitig zu identifizieren und zu überwinden.

Schliesslich muss der Einzug von KI in die Soziale Arbeit als eine Chance begriffen werden, das Professionsverständnis weiterzuentwickeln und die Qualität der Arbeit zu steigern. Die Integration von KI bietet die Möglichkeit, die Sozialarbeit zukunftsfähig zu gestalten und den Kindesschutz durch fundierte, datengestützte Entscheidungen zu verbessern. Eine verantwortungsvolle Nutzung dieser Technologien kann letztlich dazu beitragen, das generelle Vertrauen in den Kindesschutz zu stärken und die bestmöglichen Ergebnisse für Kinder und Familien zu erzielen.

9 Literaturverzeichnis

Aguinis, H. & Bradley, K. J. (2014). Best Practice Recommendations for Designing and Implementing Experimental Vignette Methodology Studies. *Organizational Research Methods*, 17(4), 351–371. https://doi.org/10.1177/1094428114547952

Ajibade, P. (2018). Technology Acceptance Model Limitations and Criticisms: Exploring the Practical Applications and Use in Technology-related Studies, Mixed-method, and Qualitative Researches. *Library Philosophy and Practice (e-journal)*. Verfügbar unter: https://digitalcommons.unl.edu/cgi/viewcontent.cgi?article=5294&context=libphilprac

Ajzen, I. (1985). *From intentions to actions: A theory of planned behavior*. Berlin: Springer. Verfügbar unter: https://www.researchgate.net/profile/Icek-Ajzen/publication/238719086_Action_Control/links/0046353a6edba381e0000000/Action-Control.pdf

Auspurg, K., Hinz, Th., Liebig, S. & Sauer, C. (2014).The Factorial Survey as a Method for Measuring Sensitive Issues. In: U. Engel, B., Jann, P. Lynn, A. Scherpenzeel, P. Sturgis (Hrsg.), *Improving Survey Methods Lessons from Recent Research* (S. 137-149). New York: Routledge. Verfügbar unter: https://www.researchgate.net/publication/266671992_The_Factorial_Survey_as_a_Method_for_Measuring_Sensitive_Issues

Baird, C. & Wagner, D. (2000). The relative validity of actuarial- and consensus-based risk assessment systems. *Children and Youth Services Review*, 22(11-12), 839–871. https://doi.org/10.1016/S0190-7409(00)00122-5

Bastian, P. & Schrödter, M. (2014). Professionelle Urteilsbildung in der Sozialen Arbeit. *Soziale Passagen*, 6(2), 275–297. https://doi.org/10.1007/s12592-014-0175-5

Bastian, P. & Schrödter, M. (2019). Soziale Arbeit in der digitalen Transformation. In Deutscher Verein für öffentliche und private Fürsorge e.V. (Hrsg.), *Soziale Arbeit in der digitalen Transformation* (S. 40-49). Freiburg: Lambertus Verlag.

Baur, N. & Blasius, J. (Hrsg.). (2014). *Handbuch Methoden der empirischen Sozialforschung*. Wiesbaden: Springer VS. https://doi.org/10.1007/978-3-658-37985-8

Becker-Lenz, R. & Müller S. (2009). Die Notwendigkeit von wissenschaftlichem Wissen und die Bedeutung eines professionellen Habitus für die Berufspraxis der Sozialen Arbeit. In R. Becker-Lenz, S. Buss, G. Ehlert & S. Müller (Hrsg.), *Professionalität in der Sozialen Arbeit. Standpunkte, Kontroversen, Perspektiven* (S. 195-221). Wiesbaden: Springer VS.

Biesel, K., Fellmann, L., Müller, B., Schär, C. & Schnurr, St. (2017). *Prozessmanual. Dialogisch-systemische Kindeswohlabklärung*. Bern: Haupt Verlag.

Braga-Neto, U. (2020). *Fundamentals of pattern recognition and machine learning.* Cham: Springer.

Brielmaier, J. (2020). *Fragebogen: Zur Relevanz wissenschaftlichen Wissens in der Praxis der Sozialen Arbeit*. Verfügbar unter: https://www.researchgate.net/publication/344282834_Fragebogen_Zur_Relevanz_wissenschaftlichen_Wissens_in_der_Praxis_der_Sozialen_Arbeit/link/5f632183458515b7cf39791b/download?_tp=eyJjb250ZXh0Ijp7ImZpcnN0UGFnZSI6InB1YmxpY2F0aW9uIiwicGFnZSI6InB1YmxpY2F0aW9uIn19

Brunner, T. (2023). Was KI mit der Sozialen Arbeit macht – und umgekehrt. *sozialinfo.ch*. Verfügbar unter: https://www.sozialinfo.ch/fachinformationen/fokusartikel/was-ki-mit-der-sozialen-arbeit-macht-und-umgekehrt

Brynjolfsson, E. & McAfee, A. (2017). The business of artificial intelligence. *Harvard Business Review*. Verfügbar unter: https://hbr.org/2017/07/the-business-of-artificial-intelligence

Buxmann P. & Schmidt H. (Hrsg.). (2019). *Künstliche Intelligenz – Mit Algorithmen zum wirtschaftlichen Erfolg*. Berling: Springer.

Campayo, S. (2020). Professionelles Handeln mit Blick auf Digitalisierung. In: N. Kutscher, T. Ley, U. Seelmeyer, F. Siller, A. Tillmann, I. Zorn (Hrsg.), *Handbuch Soziale Arbeit und Digitalisierung* (S. 290-301). Weinheim: Belz Juventa.

Chen, L., Zaharia, M. & Zou J. (2023). How Is ChatGPT's Behavior Changing over Time? *Stanford University, UC Berkeley*. https://doi.org/10.48550/arXiv.2307.09009

Chouldechova, A., Putnam-Hornstein, E., Benavides-Prado, D., Fialko, O. & Vaithianathan, R. (2018). A case study of algorithm-assisted decision making in child maltreatment hotline screening decisions. *Proceedings of Machine Learning Research* 81, 1–15. Verfügbar unter: https://proceedings.mlr.press/v81/chouldechova18a/chouldechova18a.pdf

Corman T. H., Leiserson, C.E., Rivest, R. L. & Stein, C. (2009). *Introduction to Algorithms* (3. Aufl.). Cambridge MA: MIT Press.

Cuccaro-Alamin, St., Foust, R., Vaithianathan, R. & Putnam-Hornstein, E. (2017). Risk assessment and decision making in child protective services: Predictive risk modeling in context. *Children and Youth Services Review, 79,* 291–298. https://doi.org/10.1016/j.childyouth.2017.06.027

Dare, T. (2013). Predictive Risk Modelling and Child Maltreatment – an Ethical Review. *University of Auckland*. Verfügbar unter: https://www.msd.govt.nz/documents/about-msd-and-our-work/publications-resources/research/predictive-modelling/00-predicitve-risk-modelling-and-child-maltreatment-an-ethical-review.pdf

Davis, F. D. (1985). A Technology Acceptance Model for Empirically Testing New End-User Information Systems, *Massachusetts Institute of Technology*. Verfügbar unter: https://dspace.mit.edu/bitstream/handle/1721.1/15192/14927137-MIT.pdf

Davis, F. D. (1989). Perceived Usefulness, Perceived Ease of Use, and User Acceptance of Information Technology. *Massachusetts Institute of Technology*, 13(3), 319–340. https://doi.org/10.2307/249008

Davis, F. D., Bagozzi, R. P. & Warshaw, P. R. (1989). User Acceptance of Computer Technology: A Comparison of two Theoretical Models. *Management Science*, 35(8) 982-1003. https://doi.org/10.1287/mnsc.35.8.982

Dollinger, B. & Raithel, J. (2006). *Einführung in die Theorien abweichenden Verhaltens*. Beltz.

Döring, N., & Bortz, J. (2016). *Forschungsmethoden und Evaluation in den Sozial- und Humanwissenschaften* (5. Aufl.).Berlin: Springer VS. https://doi.org/10.1007/978-3-642-41089-5

Dülmer, H. (2014). Vignetten. In: Baur, N. & Blasius, J. (Hrsg.). *Handbuch Methoden der empirischen Sozialforschung* (S. 721-732). Wiesbaden: Springer VS. https://doi.org/10.1007/978-3-658-37985-8

Eidgenössischer Datenschutz- und Öffentlichkeitsbeauftragter (EDÖB). (2023). Geltendes *Datenschutzgesetz ist auf KI direkt anwendbar*. Verfügbar unter: https://www.edoeb.admin.ch/edoeb/de/home/kurzmeldungen/2023/20231109_ki_dsg.html

Ehlert, G. (2019). Professionalisierung. s*ocialnet*. Verfügbar unter: https://www.socialnet.de/lexikon/6019

Ehlert, G. (2022). *Deprofessionalisierung*. *socialnet Lexikon.* Verfügbar unter: https://www.socialnet.de/lexikon/7601

Enosh, G. & Bayer-Topilsky, T. (2015). Reasoning and Bias: Heuristics in Safety Assessment and Placement Decisions for Children at Risk. *British Journal of Social Work,* 45(6), 1771–1787. https://doi.org/10.1093/bjsw/bct213

Estermann, A., Hauri, A. & Vogel, U. (2022). Teil III – Mandatsführung. In: D. Rosch, Ch. Fountoulaliks, Ch. Heck (Hrsg.). *Handbuch Kindes- und Erwachsenenschutz – Recht und Methodik für Fachleute* (3. akt. Aufl.). Bern: Haupt.

Eubanks, V. (2018). *Automating Inequality: How High-Tech Tools Profile, Police, and Punish the Poor*. New York: St. Martins Press.

Europäische Parlament. (2020). *Was ist künstliche Intelligenz und wie wird sie genutzt?* Verfügbar unter: https://www.europarl.europa.eu/news/de/headlines/society/20200827STO85804/was-ist-kunstliche-intelligenz-und-wie-wird-sie-genutzt

Evans, M., & Hilbert, J. (2020). Zur Zukunft der Arbeit in der Sozial- und Gesundheitswirtschaft in der Digitalisierungsära. In: N. Kutscher, T. Ley, U. Seelmeyer, F. Siller, A. Tillmann, I. Zorn (Hrsg.), *Handbuch Soziale Arbeit und Digitalisierung* (S. 75-88). Weinheim: Belz Juventa.

Evans, S. C., Roberts, M. C., Keeley, J. W., Blossom, J. B., Amaro, C. M., Garcia, A. M., Stough, C. O., Canter, K. S., Robles, R., & Reed, G. M. (2015). Vignette methodologies for studying clinicians' decision-making: Validity, utility, and application in ICD-11 field studies. *International Journal of Clinical and Health Psychology*, 15(2), 160–170. https://doi.org/10.1016/j.ijchp.2014.12.001

Fassbind, P. (2022). Verfahren vor der KESB: Von der Gefährdungsmeldung bis zur Vollstreckung. In Ch. Heck (Hrsg.), Überblick über die Akteure und deren Aufgaben. In: D. Rosch, Ch. Fountoulakis, Ch. Heck (Hrsg.). *Handbuch Kindes- und Erwachsenenschutz – Recht und Methodik für Fachleute* (S. 103 – 129). Bern: Haupt.

Fishbein, M. & Ajzen, I. (1975). *Belief, attitude, intention and behavior: An introduction to theory and research*. Verfügbar unter: https://people.umass.edu/aizen/f&a1975.html

Fulterer, R. (2023). Was, wenn künstliche Intelligenz ausser Kontrolle gerät? *Neue Zürcher Zeitung.* Verfügbar unter: https://www.nzz.ch/technologie/wenn-kuenstliche-intelligenz-ausser-kontrolle-geraet-ld.1737475?reduced=true

Gautschi, J. (2021). *Urteile und Entscheidungen unter Unsicherheit in Kindeswohlabklärungen. Einflussfaktoren auf Fallbeurteilungen in einer multifaktoriellen, experimentellen Vignettenstudie* (Dissertation). Verfügbar unter: https://phfr.bsz-bw.de/frontdoor/deliver/index/docId/883/file/Diss_Gautschi_Joel_2021.pdf

Gautschi, J. & Schnurr, S. (2024). Kindesschutz in der Schweiz. *Sozial Extra*, 48, 35–39. https://doi.org/10.1007/s12054-023-00651-y

Gillingham, P. (2016). Predictive Risk Modelling to Prevent Child Maltreatment and Other Adverse Outcomes for Service Users: Inside the 'Black Box' of Machine Learning. *British Journal of Social Work*, 46(4), 1044–1058. https://doi.org/10.1093/bjsw/bcv031

Gillingham, P. (2019). Can Predictive Algorithms Assist Decision-Making in Social Work with Children and Families? *Child Abuse Review*, 28(2), 114–126. https://doi.org/10.1002/car.2547

Gillingham, P. & Ackermann, T. (2020). Algorithmisch basierte Entscheidungsunterstützungssysteme für die deutsche Kinder- und Jugendhilfe? Messages from Research. In: W*idersprüche,* 158, 63 –79. Vefügbar unter: https://www.researchgate.net/publication/350064023_Algorithmisch_basierte_Entscheidungsunterstutzungssysteme_fur_die_deutsche_Kinder-_und_Jugendhilfe_Messages_from_Research

Glaberson, S. (2019) Coding Over the Cracks: Predictive Analytics and Child Protection. Fordham Urban Jurnal, 46(3), 307–363. Verfügbar unter: https://ir.lawnet.fordham.edu/ulj/vol46/iss2/3

Goldhaber-Fiebert, J. D. & Prince, L. (2019). *Impact Evaluation of a Predictive Risk Modeling Tool for Allegheny County's Child Welfare Office.* Verfügbar unter: https://www.alleghenycountyanalytics.us/wp-content/uploads/2019/05/Impact-Evaluation-from-16-ACDHS-26_PredictiveRisk_Package_050119_FINAL-6.pdf

Grotenhuis, M. T. & Thijs, P. (2015). Dummy variables and their interactions in regression analysis. Examples from research on body mass index. *Radboud University*. Verfügbar unter https://arxiv.org/pdf/1511.05728

Haehnlein, M. & Kaplan, A. (2019). A Brief History of Artificial Intelligence: On the Past, Present, and Future of Artificial Intelligence. *California Management Review*, 61(4). DOI:10.1177/0008125619864925

Hamilton, K., Shih. S. & Mohammed, S. (2016). The Development and Validation of the Rational and Intuitive Decision Styles Scale. *Journal of Personality Assessment*, 98(5), 523–535 http://dx.doi.org/10.1080/00223891.2015.1132426

Hauri, A., & Zingaro, M. (2020). *Kindeswohlgefährdung erkennen und angemessen handeln: Leitfaden für Fachpersonen aus dem Sozialbereich* (2., überarbeitete Aufl.). Bern: Kinderschutz Schweiz.

Hawking, S. (2014). *Stephen Hawking: 'AI could spell end of the human race' – BBC News* [Video]. Verfügbar unter: https://www.youtube.com/watch?v=fFLVyWBDTfo

Heck, Ch. (2022). Überblick über die Akteure und deren Aufgaben. In: D. Rosch, Ch. Fountoulakis, Ch. Heck (Hrsg.). *Handbuch Kindes- und Erwachsenenschutz – Recht und Methodik für Fachleute.* Bern: Haupt.

Hennessy, C. H. (1993). Modeling case management decision making in a consolidated long-term care program. *The Gerontologist*, 33(3), 333–341. https://doi.org/10.1093/geront/33.3.333

Hitchcock, C. (2018). Probabilistic Causation. In E. N. Zalta (Hrsg.), *The Stanford Encyclopedia of Philosophy* (Fall 2018 Edition). Verfügbar unter: https://plato.stanford.edu/archives/fall2018/entries/causation-probabilistic/

Huber, F., Meyer, F., & Lenzen, J. M. (2014). *Grundlagen der Varianzanalyse: Konzeption – Durchführung – Auswertung*. Wiesbaden: Springer Gabler.

Ivanov, S., Kuyumdzhiev, M. & Webster, C. (2020). Automation fears: Drivers and solutions. *Technology in Society*, 63(11). https://doi.org/10.1016/j.techsoc.2020.101431

Jasso, G. (2006). Factorial Survey Methods for Studying Beliefs and Judgments. *Sociological Methods & Research*, 34(3), 334–423. https://doi.org/10.1177/0049124105283121

Kahneman, D. (2011). *Thinking, fast and slow*. London: Allen Lane.

Kahneman, D., Slovic, P. & Tversky, A. (Hrsg.). (1982). *Judgment under Uncertainty. Heuristics and Biases*. Cambridge: Cambridge University Press.

Kanton Zürich, Amt für Jugend und Berufsberatung. (2024). *Behördlicher Kindesschutz im AJB – eine Orientierungshilfe für AJB-Mitarbeitende*. Internes Dokument.

Kasselbaum, U. (2004*). Interpersonelles Vertrauen Entwicklung eines Inventars zur Erfassung spezifischer Aspekte des Konstrukts* (Dissertation). Verfügbar unter: https://ediss.sub.uni-hamburg.de/bitstream/ediss/618/1/Dissertation.pdf

Keddell, E. (2014). Current Debates on Variability in Child Welfare Decision-Making: A Selected Literature Review. *Social Sciences*, 3(4), 916–940. https://doi.org/10.3390/socsci3040916

Keddell, E. (2019). Algorithmic Justice in Child Protection: Statistical Fairness, Social Justice and the Implications for Practice. *Social Sciences*, 8, 281. https://doi:10.3390/socsci8100281

Kelle, U. (2008). Die *Integration qualitativer und quantitativer Methoden in der empirischen Sozialforschung*. Wiesbaden: VS Verlag für Sozialwissenschaften.https://doi.org/10.1007/978-3-531-91174-8

Kieslinger, D., Dressel, M. & Haar, R. (Hrsg.). (2021). *Systemsprenger*innen: Ressourcenorientierte Ansätze zu einer defizitären Begrifflichkeit* (Beiträge zur Erziehungshilfe, Band 49). Freiburg: Lambertus.

Kindler, H. (2018). Operationalisierungen von Kindeswohl und Kindeswohlgefährdung in den Sozial- und Humanwissenschaften. In H. Katzenstein, K. Lohse, G. Schindler, & L. Schönecker (Hrsg.), *Das Recht als Partner der Fachlichkeit in der Kinder- und Jugendhilfe* (S. 179–224). Baden-Baden: Nomos Verlagsgesellschaft

Kindler, T. (2019). *Sozialarbeitspolitik. Eine empirische Untersuchung der politischen Aktivität von Fachpersonen der Sozialen Arbeit in der Schweiz* (Masterthesis, Kooperationsstudiengang Master of Science in Sozialer Arbeit der Fachhochschulen Bern, Luzern, St. Gallen und Zürich). http://dx.doi.org/10.13140/RG.2.2.15162.85440

Klassen, M. (2004). *Was leisten Systemtheorien in der sozialen Arbeit? Ein Vergleich der systemischen Ansätze von Niklas Luhmann und Mario Bunge* (1., Aufl). Bern: Haupt Verlag.

Kleinberg, J., Lakkaraju, H. Leskovec, J., Ludwig, J. & Mullainathan, S. (2017). Human Decisions And Machine Predictions. *Oxford University Press*, 133(1), 237–293. Verfügbar unter: https://doi.org/10.1093/qje/qjx032

Knuth, D. E. (1997). *The Art of Computer Programming* (3. Aufl). Boston: Addison Wesley.

Kohring, M. (2001). *Vertrauen in Medien – Vertrauen in Technologie*. Stuttgart. Verfügbar unter: https://elib.uni-stuttgart.de/bitstream/11682/8694/1/AB196.pdf

Konferenz für Kindes- und Erwachsenenschutz (KOKES).(2021) *Empfehlungen zur Organisation von Berufsbeistandschaften*. Verfügbar unter https://www.kokes.ch/application/files/2716/2814/0146/KOKES_Empfehlungen_Berufsbeistandschaften.pdf

Krauss, P. (2023). *Künstliche Intelligenz und Hirnforschung. Neuronale Netze, Deep Learning und die Zukunft der* Kognition. Heidelberg: Springer. https://doi.org/10.1007/978-3-662-67179-5

Kreutzer, R. T. (2023). *Künstliche Intelligenz verstehen: Grundlagen - Use-Cases - Unternehmenseigene KI-Journey* (2. überarbeitete Aufl.). Wiesbaden: Springer Gabler. https://doi.org/10.1007/978-3-658-42598-2

Kröller, Th. & Schwarz, J. (2022). Kovarianzstrukturanalyse: Analyse der Einflussfaktoren auf die Akzeptanz autonomer Automobile in Deutschland. In: S. Bossow-Thies, B. Krol (Hrsg.), *Quantitative Forschung in Masterarbeiten* (S. 379-417). Wiesbaden: Springer Gabler. https://doi.org/10.1007/978-3-658-35831-0_13

Kubisch, S. & Franz, J. (2022). Professionalisierung in der Sozialen Arbeit aus der Perspektive der Praxeologischen Wissenssoziologie. In: R. Bohnsack, A. Bonnet, U. Hericks (Hrsg.), *Praxeologisch-wissenssoziologische Professionsforschung. Perspektiven aus Früh- und Schulpädagogik, Fachdidaktik und Sozialer Arbeit* (S. 413-442). Bad Heilbrunn: Verlag Julius Klinkhardt. https://doi.org/10.25656/01:25656

Lätsch, D. (2021). *Der Kindesschutz und das digitale Orakel. Vortrag zum Alumni-Forum 2021*. Verfügbar unter: https://www.zhaw.ch/storage/shared/sozialearbeit/Weiterbildung/Fachveranstaltungen/Alumni-Forum/david-laetsch-af2021-presentation.pdf

Laux, H., Gillenkirch, R. M. & Schenk-Mathes, H.Y. (2018). *Entscheidungstheorie*. (10. Aufl.) Berlin, Heidelberg: Springer. Verfügbar unter: https://link.springer.com/book/10.1007/978-3-662-57818-6

Lee, H.-J. (2004). The role of competence-based trust and organizational identification in continuous improvement. *Journal of Managerial Psychology*, 19(6), 623–639. Verfügbar unter: https://www.researchgate.net/publication/274078186_The_role_of_competence-based_trust_and_organizational_identification_in_continuous_improvement

Lee, M. K. (2018). Understanding perception of algorithmic decisions: Fairness, trust, and emotion in response to algorithmic management. *Big data & society*, 5(1). https://doi.org/10.1177/2053951718756684

Luhmann, N. (2009). *Vertrauen, Ein Mechanismus der Reduktion sozialer Komplexität* (4. Aufl.). Stuttgart: Lucius & Lucius.

Mayer, R. C., Davis, J. H., & Schoorman, F. D. (1995). An integrative model of organizational trust. *Academy of Management Review*, 20(3), 709–734. https://doi.org/10.2307/258792

Ministry of Social Development. (2014). *The feasibility of using predictive risk modelling to identify new-born children who are high priority for preventive services – companion technical report*. Verfügbar unter: https://www.msd.govt.nz/documents/about-msd-and-our-work/publications-resources/research/predictive-modelling/00-feasibility-study-report-technical-companion.pdf

Munro, E. (1999). Common Errors of Reasoning in Child Protection Work. *Child Abuse & Neglect*, 23(8), 745–758. https://doi.org/10.1016/S0145-2134(99)00053-8

Munro, E. (2008). Lessons from research on decision-making. In D. Lindsey, & A. Shlonsky (Hrsg.), *Child welfare research: Advances for practice and policy* (S. 194-200). New York: Oxford University Press.

Nitzsch, R. (2021). *Entscheidungslehre: Wie Menschen entscheiden und wie sie entscheiden sollten*. (11. Aufl.). Wiesbaden: Springer Fachmedien. Verfügbar unter: https://link.springer.com/book/10.1007/978-3-658-34520-4

Nordheim, C. B., Følstad, A., & Bjørkli, C. A. (2019). An initial model of trust in chatbots for customer service – findings from a questionnaire study. *Interacting with Computers*, 31(3), 317–335. https://doi.org/10.1093/iwc/iwz022

Nörr, M. (2016). Marktnähe: Innovationen mit Mehrwert kreieren. In M. Nörr (Hrsg.), *Key Learnings aus dem Serial Entrepreneurship: Erfahrungen von Mehrfachgründern für eigene Innovationen nutzen* (S. 49-67). Springer. https://doi.org/10.1007/978-3-658-13084-8_4

Ottmann, T. & Widmayer, P. (2012). *Algorithmen und Datenstrukturen*. Heidelberg: Spektrum Akademischer Verlag.

Palfrey, J., Gasser, U. (2008). *Born Digital: Understanding the First Generation of Digital Natives*. New York: Basic Books.

Rheu, M., Shin, J. Y., Peng, W., & Huh-Yoo, J. (2021). Systematic review: Trust-building factors and implications for conversational agent design. *International Journal of Human-Computer Interaction*, 37(1), 81–96. https://doi.org/10.1080/10447318.2020.180771

Richter, L., Paier, D. & Reiger, H. (2021). *Quantitative Sozialforschung – Eine Einführung*. Wien: Facultas.

Rieder, S., Bieri, O., Schwenkel, Ch., Hertig, V. & Amberg, H. (2016). *Evaluation Kindes- und Erwachsenenschutzrecht. Analyse der organisatorischen Umsetzung und Kennzahlen zu Leistungen und Kosten. Bericht zuhanden des Bundesamts für Justiz (BJ)*. Luzern. Verfügbar unter: https://www.newsd.admin.ch/newsd/message/attachments/79441.pdf

Rossi, P. H., & Anderson, A. B. (1982). The Factorial Survey Approach: An Introduction. In P. H. Rossi & S. L. Nock (Hrsg.), *Measuring Social Judgments. The Factorial Survey Approach* (S. 15-67). Beverly Hills, CA: SAGE Publications. Verfügbar unter: http://130.18.86.27/faculty/warkentin/securitypapers/Merrill/RossiAnderson1982_Chapter_FactorialSurveyApproach.pdf

Rossi, P. H., Schuerman, J., & Budde, S. (1999). Understanding decisions about child maltreatment. *Evaluation Review, 23*(6), 579–598. https://doi.org/10.1177/0193841X9902300601

Rost, K. (2018). *Die Vignettenanalyse in den Sozialwissenschaften: Eine anwendungsorientierte Einführung* (2., völlig neu bearbeitete Aufl.). München: Rainer Hampp Verlag.

Sassmannshausen, T. M. & Heupel, T. (2020). Vertrauen in KI – Eine empirische Analyse innerhalb des Produktionsmanagements. In R. Buchkremer, Th. Heupel & O. Koch (Hrsg.), *Künstliche Intelligenz in Wirtschaft & Gesellschaft – Auswirkungen, Herausforderungen & Handlungsempfehlungen* (S. 169-192). Wiesbaden: Spinger.

Schauffel, N., Schmidt, I., Peiffer, H., & Ellwart, T. (2021). *ICT Self-Concept Scale (ICT-SC25)*. https://doi.org/10.6102/zis308_exz

Scheidgen, A. (2019). *Entscheiden als Aspekt professioneller Kompetenz angehender Fachkräfte der Sozialen Arbeit – Modellierung und empirische Überprüfung eines Konstrukts*. Opladen: Budrich UniPress.

Schepman, A & Rodway, P. (2020). *Initial validation of the general attitudes towards Artificial Intelligence Scale*. Verfügbar unter: https://www.sciencedirect.com/science/article/pii/S2451958820300142

Schneider, D. (2021). Ein Schritt in Richtung De-Professionalisierung? Plädoyer für eine intensive Diskussion über algorithmische Systeme in der professionellen Praxis. In M. Wunder (Hrsg.), *Digitalisierung und Soziale Arbeit – Transformationen und Herausforderungen* (S. 122-139). Bad Heilbrunn: Julius Klinkhardt.

Schulze, A. (2019). Auseinandersetzung mit Digitalisierungsprozessen im Rahmen der Aus- und Weiterbildung von Fachkräften in der Sozialen Arbeit. In: *Neue Praxis – Zeitschrift für Sozialarbeit, Sozialpädagogik und Sozialpolitik*, 49(2019), 384–391.

Seithe, M. (2012). *Schwarzbuch Soziale Arbeit* (2. Aufl). Wiesbaden: Springer.

Shane, J. (2021). *Künstliche Intelligenz – Wie sie funktioniert und wann sie scheitert : Eine unterhaltsame Reise in die seltsame Welt der Algorithmen, neuronalen Netze und versteckten Giraffen*. Köln: O'Reilly Verlag. Verfügbar unter: https://ebookcentral.proquest.com/lib/zhaw/detail.action?docID=6660210

Sheridan, T. B. (2019). Individual differences in attributes of trust in automation: Measurement and application to system design. *Frontiers in Psychology*, 10(2019). https://doi.org/10.3389/fpsyg.2019.01117

Stadt Zürich (2024). *Kindes- und Erwachsenenschutzbehörde – Mandatsträgerinnen und Mandatsträger*. Verfügbar unter: https://www.stadt-zuerich.ch/portal/de/index/politik_u_recht/kindes_und_erwachsenenschutzbehoerde/informationen_mt.html

Steiner, P. M. & Atzmüller, Ch. (2006). Experimentelle Vignettendesigns in faktoriellen Surveys. *Kölner Zeitschrift für Soziologie und Sozialpsychologie*, 58, 117–146. Verfügbar unter: https://doi.org/10.1007/s11575-006-0006-9

Strasser, U. & Lütkehaus, I. (2020). *Cross Generational Intelligence – Kooperation der Generationen in Unternehmen* (1. Aufl.). Freiburg: Haufe.

Sydow, J. (1995). Konstitutionsbedingungen von Vertrauen in Unternehmensnetzwerken. Theoretische und empirische Einsichten. In R. Bühner, K. D. Haase, & J. Wilhelm (Hrsg.), *Die Dimensionierung des Unternehmens* (S. 177-220). Stuttgart: Schäffer-Poeschel.

Tamilmani, K., Rana, N. P., Wamba, S. F., & Dwivedi, R. (2021). The extended Unified Theory of Acceptance and Use of Technology (UTAUT2): A systematic literature review and theory evaluation. *International Journal of Information Management*, 57. https://doi.org/10.1016/j.ijinfomgt.2020.102269

Taylor, B. J. (2006). Factorial Surveys: Using Vignettes to Study Professional Judgement. *British Journal of Social Work*, 36 (7), 1187–1207. https://doi.org/10.1093/bjsw/bch345

TCC Group (2013). Peer Review *Report 1*. Verfügbar unter: https://www.msd.govt.nz/documents/about-msd-and-our-work/publications-resources/research/predictive-modelling/feasibility-study-schwartz-tcc-interim-review.pdf

Urban, D., Mayerl, U. & Wahl, A. (2016). *Regressionsanalyse bei fehlenden Variablenwerten (missing values): Imputation oder Nicht-Imputation? Eine Anleitung für die Regressionspraxis mit SPSS*. (2. korr. Aufl.). Stuttgart: Universität Stuttgart. Verfügbar unter: https://www.sowi.uni-stuttgart.de/dokumente/forschung/siss/SISS44.pdf

Vaithianathan, R., Benavides-Prado, D., Dalton, E., Chouldechova, A., & Putnam-Hornstein, E. (2021). Using a Machine Learning Tool to Support High-Stakes Decisions in Child Protection. *AI Magazine*, 42(1), 53–60. Verfügbar unter: https://ojs.aaai.org/aimagazine/index.php/aimagazine/article/view/7482

Vaithianathan, R., Maloney, T., Jiang, N., De Haan, I., Dale, C., Putnam-Hornstein, E., & Dare, T. (2012). *Vulnerable children: Can administrative data be used to identify children at risk of adverse outcomes?* Verfügbar unter: https://thehub.swa.govt.nz/resources/vulnerable-children-can-administrative-data-be-used-to-identify-children-at-risk-of-adverse-outcomes/

Vaithianathan R., Putnam-Hornstein E., Jiang N., Nand P., & Maloney T. (2017). *Developing Predictive Models to Support Child Maltreatment Hotline Screening Decisions*: *Allegheny County methodology and Implementation. Centre for Social Data Analytics*. Verfügbar unter: https://www.alleghenycountyanalytics.us/wp-content/uploads/2019/05/16-ACDHS-26_PredictiveRisk_Package_050119_FINAL-2.pdf

Venkatesh, V. & Davis, F. D. (2000). A theoretical extension of the technology acceptance model: Four longitudinal field studies. *Management Science*, 46(2), 186–204. Verfügbar unter https://www.jstor.org/stable/2634758?seq=1

Venkatesh, V., Morris, M. G., Davis, G. B., & Davis, F. D. (2003). User Acceptance of Information Technology: Toward a Unified View. *MIS Quarterly*, 27(3), 425–478. https://doi.org/10.2307/30036540

Wess, S. (2019). Mit Künstlicher Intelligenz immer die richtigen Entscheidungen treffen. In P. Buxmann & H. Schmidt (Hrsg.), *Künstliche Intelligenz – Mit Algorithmen zum wirtschaftlichen Erfolg* (S. 165-181). Berlin: Springer.

White, A. & Walsh, P. (2006). *Risk Assessment in Child Welfare – an Issues Paper*. Verfügbar unter: http://www.community.nsw.gov.au/__data/assets/pdf_file/0005/321647/research_riskassessment.pdf

Winkler, A., SWR2 Wissen (2023). Mit künstlicher Intelligenz gegen den Fachkräftemangel? [Audio-Podcast]. Verfügbar unter: https://www.swr.de/swr2/wissen/mit-kuenstlicher-intelligenz-gegen-den-fachkraeftemangel-swr2-wissen-2023-09-19-100.html

Yang, R. & Wibowo, S. (2022). User Trust in Artificial Intelligence: A Comprehensive Conceptual Framework. *Electron Markets,* 32, 2053–2077. https://doi.org/10.1007/s12525-022-00592-6

Yiu, Ch., Grant, K. & Edgar, D. (2007). Factors affecting the adoption of Internet Banking in Hong Kong—implications for the banking sector. *International Journal of Information Management*, 27(5), 336–351). https://doi.org/10.1016/j.ijinfomgt.2007.03.002

Zöfel, P. (2003). *Statistik für Psychologen – im Klartext. München*: Pearson Studium.

10 Anhang

10.1 Anhang 1 - Bemerkung zur Umfrage

Gibt es noch etwas, das Sie zum Thema KI in der Sozialen Arbeit oder zu dieser Umfrage sagen möchten? Ich bin daran interessiert.

Person	Aussage	Textmodul	Thema	Kategorie	Ausprägung
1	P1.1	Ich bin interessiert an KI, wenn es die Qualität unsere Arbeit unterstützt, stützt oder erhöht.	Unterstützung, Qualitätssteigerung	Qualitätssteigerung	positiv
2	P2.1	Menschen sind unberechenbar, der Faktor Mensch und sein Verhalten können niemals mit Sicherheit vorab eingeschätzt werden. Auch wenn ein Mensch in der gleichen Situation 10 x gleich reagiert hat, ist es möglich, dass er das 11 Mal anders reagiert.	Unberechenbarkeit Mensch, menschliches Verhalten	menschliche Entscheidfindung	neutral
	P2.2	Wie würden sich die Entscheidungen von Kibat langfristig auswirken auf die betroffenen Menschen? Wie sind die kurz- mittel-und langfristigen Prognosen errechnet?	langfristige Auswirkungen	Funktionsweise	neutral
3	P3.1	Mir scheint wichtig, dass kein Automatismus entsteht und künftige Einschätzungen nicht alleinig der KI überlassen werden.	Automatismus, nicht alleine KI-Einschätzung	Verantwortung	neutral
	P3.2	Zudem wäre es mir wichtig, die Funktionsweisen einer KI genauer verstehen zu können (auf welche Faktoren wird wie stark Wert gelegt, wo gibt es blinde Flecken, was kann die KI nicht etc.). Beispiel: Beim Vormundschaftsfall führt eine schwächere Massnahme gemäss KI zu mehr Erfolg. Wird alleinig die Erfolgswahrscheinlichkeit von verschiedenen Massnahmen berechnet? Oder werden in den KI-Massnahmenempfehlungen auch die aktuellen Gefährdungsmomenten durch multitox Substanzkonsum & Beschaffungskriminalität mit berücksichtigt? Oder beim Abklärungsfall: die KI schlägt eine stärkere Massnahme vor (wir gehen mal von Platzierung aus). Wird diese Empfehlung alleinig aufgrund der von der KI erkannten (objektiven) Risikofaktoren erstellt? Oder wird bspw. auch berücksichtigt, dass die Familie symbiotische Züge hat und eine zwangsweise Femdplatzierung als Erstmassnahme unter Umständen traumatisierende Folgen für das Kind haben könnten, die längerfristig ggfs. sogar weitaus schlimmer sein könnte, als sich die aktuelle Gefährdung zeigt?	Funktionsweise verstehen, Erfolgswahrscheinlichkeit, schwächere Massnahmen, Gefährdungsmomente, Berücksichtigung Faktoren, berücksichtigte Risikofaktoren	Funktionsweise	neutral
	P3.3	Ich kann mir vorstellen, dass KI bei Sozialarbeitenden Ängste und Verunsicherungen auslösen könnten, wenn suggeriert würde, dass die KI die besseren / treffenderen Einschätzungen vornehmen könnte als die Menschen. Es könnte ein Gefühl der eigenen Überflüssigkeit ausgelöst werden, sind Einschätzungen und Entscheidungen prägende Elemente des Arbeitsalltags.	Ängste, Verunsicherung, bessere Einschätzungen, eigene Überflüssigkeit, Einschäzung ist prägendes Element im SA-Alltag	Verunsicherung SA	negativ
	P3.4	Hingegen ist mit hoher Akzeptanz zu rechnen, wenn die KI unliebsame administrative Arbeiten übernehmen oder vereinfachen könnte (Beispiel: Berichte schreiben und/oder strukturieren durch Zugriff auf Dokumente und Aktennotizen; Aktennotizen aus Stichworten generieren; auf Mails mit Stichworten antworten und die KI generiert einen Fliesstext etc.).	Akzeptanz, administrative Arbeiten	Administrative Arbeiten	positiv
4	P4.1	Es sind eher grundsätzliche Fragen. Widerspricht die Nutzung von KI, alles ist machbar , sich selber stets zu optimieren nicht den ursprünglichen Haltungen von Sozialer Arbeit, einem bestimmten Menschenbild und auch der täglichen Arbeit?	Haltung Soziale Arbeit vs. KI, Menschenbild, bessere Bedingungen ohne KI	Defizite KI: Professionswissen	negativ
	P4.2	Wäre es nicht gerade sinnvoll, in dieser Arbeit andere Wege zu gehen, sich mehr für bessere Bedingungen einzusetzen, als von KI Gebrauch und sich stückweit abhängig zu machen und sich weitere Arbeitsbelastungen auszusetzen (der Druck, die Arbeitsbelastung könnte durch die Nutzung von KI für die einen noch grösser werden).	Abhängigkeit, zusätzliche Arbeitsbelastungen	zusätzliche Arbeitsbelastung	negativ
5	P5.1	Erfahrungen mit KI wären mein persönlicher Entscheidungsträger.	Erfahrung ist entscheidend	Erfahrungswerte	neutral
	P5.2	Nicht berücksichtigt ist, wie die Rahmenbedingungen gestaltet würden: inwieweit werden Mitarbeitende verpflichtet, sich an die Empfehlungen gemäss KI zu halten?, bei wem liegt die Verantwortung?,	interne Rahmenbedingungen, Verpflichtung, Verantwortung	Rahmenbedingungen	neutral
	P5.3	wie werden KI Daten aktualisiert?, wann werden Phänomene (bspw. Absentismus) aufgenommen?,	Funktionsweise, neue Aufnahme von Daten	Funktionsweise	neutral
	P5.4	kann die Beratung tatsächlich wieder an Bedeutung gewinnen?,	Chancen für Beratung, Qualitätssteigerung Beruf	Qualitätssteigerung	positiv
	P5.5	wie gut sind bei letzterem Mitarbeitende geschult bzw. inwieweit hat man sich bereits an die Klientenverwaltung gewöhnt, ...?	Schulung, Klientenverwaltung, Neues erlernen	Schulung	neutral
6	P6.1	Meine Begeisterung für die KI in unserer Arbeit hält sich in Grenzen. Ich denke nicht dass es eine Erleichterung für die Arbeit der Sozialarbeitenden bedeutet. Denn schlussendlich sind immer noch wir diejenigen welche die Emotionen auffangen, denn KI ist bislang dazu nicht in der Lage.	keine Erleichterung, KI kann keine Emotionen auffangen	Defizite KI: zwischenmenschliche Beziehungen	negativ
7	P7.1	Ich frage mich inwiefern es in der Schweiz genügend Daten gibt um ein für den Schweizer Kontext aussagekräftiges KI-Tool zu entwickeln.	Frage nach genügend Daten für aussagekräftiges Tool	Funktionsweise	neutral
	P7.2	In jedem Fall müssen die Sozialarbeitenden darin geschult werden KI adäquat zu handhaben und kritisch zu würdigen.	Schulung, adäquate und kritische Handhabung	Schulung	neutral
8	P8.1	KI wird nicht den persönlichen Kontakt, die Beziehungsgestaltung mit den betroffenen Kindern und Eltern ersetzen können.	persönlicher Kontakt, Beziehungsgestaltung	Defizite KI: zwischenmenschliche Beziehungen	neutral
	P8.2	KI kann und sollte die personalen Kompetenzen und das Erfahrungswissen/-erleben der mandatsführenden Person nicht ersetzen.	Personale Kompetenzen SA, kein Ersetzen von Erfahrungswissen/ Erfahrungserleben	Defizite KI: Erfahrungswissen	neutral
9	P9.1	Ich befürchte nicht, dass durch den Einsatz von KI Sozialarbeitende keine Jobs mehr finden. Es besteht die Hoffnung, dass dadurch der Fachkräftemangel reduziert werden könnte.	Fachkräftemangel durch KI reduziert	Fachkräftemangel	positiv
10	P10.1	KI und soziale Arbeit passen nicht zusammen. Sie führt dazu weniger selber zu denken und nur ein geringer Teil würde KI tatsächlich dafür nutzen, um eigene Überlegungen zu reflektieren.	KI führt zu weniger denken, kein zusätzliches Reflektieren wegen KI	Defizite KI: kritisches Denken	negativ
11	P11.1	Ich hoffe, dass KI basierte Tool administrative Arbeiten erleichtern und somit der Fokus vermehrt auf die zwischenmenschlichen Interaktionen gelegt wird.	Administratives erleichtern, Fokus auf zwischenmenschliche Interaktionen	Administrative Arbeiten	positiv

	P11.2	Ich gehe davon aus, dass menschliches Handeln auch weiterhin nicht fassbar, da es irrational ist und statistisch basierte Entscheidungen dem daher nur bedingt Rechnung tragen können. Interventionen in der sozialen Arbeit ist häufig gekoppelt mit der zwischenmenschlichen Beziehung zwischen Klient*In und Sozialarbeitenden.	Menschliches Handeln nicht fassbar, irrational, Statistische Entscheide können nicht Rechnung tragen, Interventionen SA sind an zwischenmenschlichen Beziehungen gekoppelt	Defizite KI: zwischenmenschliche Beziehungen	neutral
12	P12.1	KI wäre interessant für das Erstellen von Berichten.	Berichte erstellen	Administrative Arbeiten	positiv
	P12.2	Hingegen würde ich mich weniger auf die Handlungsempfehlungen verlassen, es gibt kein Schema X bei Menschen.	kein Verlass auf Handlungsempfehlungen, kein Schema X bei Mensch	Defizite KI: Einzelfall	negativ
13	P13.1	Ich finde es schwierig, Aussagen über die Nutzung zu machen, solange ich nicht weiss, wie die KI genau funktioniert (nach welchen Kriterien wird ausgewerttet usw.). Wenn ich das System bessert kenne, könnte ich mich besser für oder dagegen entscheiden.	kein Wissen über Funktionsweise, keine Entscheidung	Funktionsweise	neutral
14	P14.1	KI in Kobmination, so quasi als weitere Meinung eines Abklärungsberichtes im KES Bereich finde ich sehr spannend und könnte wertvoll sein.	positive Einschätzung KI-Mesch-Kombi	Qualitätssteigerung	positiv
15	P15.1	KI ist eine tolle Entlastung bei Admin KI ist super zum strukturieren, korrigieren und gegenlesen von Texten KI eignet sich sehr für Brainstorming und Ideensammlung. Auch in der sozialen Arbeit. Es liefert eine Basis, auf der man dann aufbauen kann. (Z.b. welche Kriterien müssen erfüllt sein für eine Rückplatzierung).	Entlastung bei Admin-Aufgaben, eignet sich für Brainstorming	Administrative Arbeiten	positiv
16	P16.1	Das System der KI ist nur so gut wie die Eingabe der Daten. 1000 Datenreihen sagen noch nichts über einen individuellen Einzelfall aus.	Fehlende Daten, Einzelfall	Funktionsweise	neutral
	P16.2	Wo bitte bleibt die Ausnahme? Der Einsatz von KI kann - wie in Holland in der Sozialhilfe - zu verheerenden Fehlern führen. Der Einsatz von KI nur zur Einschätzung ist ein Trugschluss. Wieso sollten die Befragungen, Klientengespräche nicht auch früher oder später über die KI laufen.	Skepsis ggü Einschätzung KI, Einzelfall, Trugschluss	Defizite KI: Einzelfall	negativ
17	P17.1	KI in Administrativen Belangen sollte unbedingt vermehrt genutzt werden und der Kanton soll den Sozialarbeitenden die entsprechenden Tools zur Verfügung stellen (Tools bspw. die Aktennotitzen einlesen könnten und verschriftlichen, oder Tools, die aus Stichworten einen Bericht schreiben). Aktuell verbrauchen wir unmengen an Zeit um sauber zu formulieren. Das könnte tatsächlich eine Entlastung sein.	Administrative Belange,	Administrative Arbeiten	positiv
	P17.2	Bezüglich der Einschätzungen eines Kindesschutzfalles kann es allenfalls zukünftig effektiv interessant sein sich eine Einschätung von einem KI Tool einzuholen. Die Quellen auf denen sich die Einschätzung basiert müsste aber absolut transparent sein und es dürften nur Wissenschaftlich geprüfte/anerkannte Quellen verwendet werden.	Einschätzung durch KI-Tool, Quellen müssen transparent und wissenschaftlich sein	Transparenz	neutral
	P17.3	Vor einer solchen Verwendung sind wohl noch etwas weiter entfernt und auch dann bräuchte es immer noch die eigene definitive Entscheidung. Grundsätzlich machen wir das ja auch jetzt schon, wir erfragen die Meinung/Einschätzung von Kolleg:innen (bspw. in der Supi) und entscheiden am Schluss selber was wir daraus machen. Das ist nicht viel anders als wenn wir eine KI um eine Zweitmeinung fragen, vielleicht gibt uns die KI ja tatsächlich neue Ideen die wir weiterverfolgen könnten. Wichtig ist, dass wir dabei unabhängig bleiben und die Entscheidung, egal welche muss immer Fachlich begründet sein. Keine Lösung ist: die KI hat entschieden. Das müssen wir aber auch jetzt, wenn wir z.B. das Team als Echokammer benutzten, die können nähmlich genauso falsch liegen in ihrer Einschätzung.	Zweitmeinung, Entscheidung durch SA und nicht durch KI, Verantwortung	Verantwortung	neutral
18	P18.1	KI wie z.B. Chatgpt könnte auch nur in der administrativen Arbeit wie Berichte schreiben etc. eingesetzt werden.	ChatGPT nur in administrativen Bereichen	Administrative Arbeiten	neutral
	P18.2	Der Einsatz von KI in der Kindeswohleinschätzung könnte dazu führen, dass die humanistische Haltung in der Sozialarbeit abhanden kommt.	Haltung Soziale Arbeit vs. KI, Menschenbild	Defizite KI: Professionswissen	negativ
	P18.3	Ich frage mich, ob KI auch einschätzen könnte, ob bei einer intensiveren Massnahme die Zusammenarbeit mit den Eltern dadurch erschwert werden. Alle zwischenmenschlichen Interaktionen, die sehr wichtig sind in der Arbeit im Kindesschutz, können vermutlich nicht durch KI vorhergesehen werden.	Funktionsweise Tool	Funktionsweise	neutral
19	P19.1	KI wäre für mich ein Tool um Einschätzungen breiter / vertiefter vorzunehmen. KI würde mich ggf. auffordern gewisse Entscheid nochmals zu Reflektieren, hinterfragen oder um mich im Entscheid zu bestärken. Das Tool ergänzt meine Arbeit, ersetzt sie aber bei weitem nicht. Diese Frage haben mir hierzu etwas gefehlt.	Vertieftere Einschätzung, Reflexion, Hinterfragen, Bestärkung, Ergänzt, aber ersetzt nicht	Qualitätssteigerung	positiv
20	P20.1	Ich glaube KI hat ein riesen Potential, wenn genügend Daten vorhanden sind.	Potential, wenn genübgend Daten	Funktionsweise	neutral
21	P21.1	KI muss gut mit Studien, aktueller Literatur etc. gefüttert werden, um einen Nutzen zu bringen. Aber auch in der Literatur gibt es unterschiedliche Darstellungen vom selben Sachverhalt. Daher würde ich es empfehlenswert finden, KI auch mit unterschiedlicher, teilweise auch gegensätzlichen Ansätzen zu füttern , um dann vielleicht nicht nur eine potenzielle Lösung sondern mehrere potenzielle Lösungen zu erhalten.	KI muss mit Literatur trainiert werden, welche Literatur? Unterschiedliche Perspektiven, gegensätzliche Literatur, mehrere Lösungsansätze	Wissenschaftliches Wissen	neutral
	P21.2	Schlussendlich ist wichtig, dass nicht KI entscheidet sondern immer noch ich als Beistandsperson resp. die Behörde - Menschen sind (glücklicherweise) keine Maschinen, so auch unsere Klient*innen nicht.	Entscheid bei Beistandsperson und nicht bei KI	Verantwortung	neutral
22	P22.1	Im schreiben von Berichten setze ich die KI bereits ein in anonymisierter Form, welche mir die Arbeit dazu erleichtert. Doch ist die KI noch in den Anfangsschuhen und stark fehleranfällig.	Berichte schreiben, jetzt schon KI, stark fehleranfällig	Administrative Arbeiten	neutral
23	P23.1	Das Herleiten von Lösungen und Verstehen sind Aufgaben die künstliche Intelligenz nicht kompensieren und ersetzen können.	menschliches Lösungsherleiten und Verstehen kann KI nicht kompensieren	Defizite KI: Professionswissen	negativ
	P23.2	Der Einsatz von KI in der Sozialen Arbeit braucht geschulte Sozialarbeitende.	Einsatz braucht Schulung	Schulung	neutral
24	P24.1	Berichte schreiben	Berichte	Administrative Arbeiten	neutral
	P24.2	Datenschutz	Datenschutz	Datenschutz	neutral
25	P25.1	Gewisse Lösungen der KI mögen vielleicht statistisch gesehen treffsicherer sein, können aber aus individuellen Gründen evtl. genau für diese Situation/Person nicht realitätsgetreu umgesetzt werden aufgrund sehr individuellen Ausgangslagen.	individuelle Situation/Person, keien relaitätsgetreue Umsetzung, individuelle Ausgangslage	Defizite KI: Einzelfall	negativ

	P25.2	Ausser die KI wird mit derart umfassenden Informationen zur Person, Ressourcen, Defiziten, etc. gefüttert, dass sie dies bereits im vorhinein korrekt hervorsagen/berücksichtigen könnte. Die Eingabe der Informationen würde wohl einiges an Zeit beanspruchen. Ich meine damit bspw. auch kognitive Fähigkeiten einer Person, Erkrankungen (Entwicklungsverzögerungen, psychische Erkrankungen etc.), Sprachkenntnisse, Offenheit für beabsichtigte Unterstützung, kulturelle Hintergründe, zeitliche und energetische Ressourcen eines Elternteils usw. Aktuell habe ich noch Zweifel, ob der Output den Input übersteigen kann... :)	Informationen zu Personen, Ressourcen, Defizite usw.	Funktionsweise	neutral
26	P26.1	Wenn kann KI eine Ergänzung und nicht oder noch nicht ein Ersatz sein. Dies wie in der Art von Tools, Tests oder ähnlich.	Ergänzung und nicht Ersatz	Verantwortung	neutral
27	P27.1	Sehr interessantes Thema, welches hochaktuell ist und bereits auch in unserem Team zu Fragen geführt hat. Viele haben grosse Berührungsängste, andere hingegen sind hochbegeistert. Ich gehe davon aus, dass das unsere Zukunft sein wird und wir uns damit auseinandersetzen müssen. Viel Erfolg bei der Auswertung, wir sind auf die Resultate sehr gespannt!	Zukunft, SA muss sich damit auseinandersetzen	Wandel der Zeit	neutral
28	P28.1	Ein Schwarzweissmalen der Situationen und über den gleichen Kammscheren mit KI ist der individuellen Hilfestellung und damit einem Erfolg versprechenden Handlungsplan abträglich.	Schwarzweissmalen, gleicher Kamm, keine individuelle Hilfestellung	Defizite KI: Einzelfall	negativ
	P28.2	Es fördert die grassierende Beziehungslosigkeit wie sie unbedingt notwendig ist in der Sozialen Arbeit, speziell in der Mandatsführung.	Beziehungsdefizite	Defizite KI: zwischenmenschliche Beziehungen	negativ
29	P29.1	Das Thema Datenschutz müsste sicherlich kritisch hinterfragt und genaustens analysiert werden, sollte künftig KI eingesetzt werden.	Datenschutz	Datenschutz	neutral
	P29.2	Bei den KI Tools müsste im Gegensatz zu z.B. Chat GPT klar ausgewiesen werden, auf welche Literatur sich die KI beruft, bzw. auf welchen Teil der Informationen über die KL.	Tool müsste ausweisen, auf welche Literatur sie sich wo beruft	Wissenschaftliches Wissen	neutral
	P29.3	Zudem müsste die KI ihre Überlegungen begründen können.	Überlegungen belegen	Transparenz	neutral
30	P30.1	Aktuell nutze ich teilweise die Inputs von ChatGPT um mögliche Abläufe durchzu diskutieren . Zudem lasse ich längere Briefe vorformulieren (bsp. Spendenanfragen).	bereits Nutzung ChatGPT	Administrative Arbeiten	neutral
31	P31.1	Bestelladresse für die genannten KI-Tools?	-	-	-
32	P32.1	Im Rahmen eines CAS, den ich an der Hochschule Luzern, Abteilung Wirtschaft, besuchte, war KI ein spannendes Thema. Wir haben dabei festgestellt, dass faktenbasierte Arbeiten (auch in der Sozialarbeit) durchaus mit Hilfe von KI rascher und besser erledigt werden könnten.	Faktenbasiertes Arbeiten durch KI möglich	Qualitätssteigerung	neutral
	P32.2	Da KI jedoch z.B. Emotionen, Mimik und Gestik, sensorische Reize, Intuition, Haptik, sichtbare und unsichtbare Interaktionen zwischen Lebewesen, etc. nicht erfassen oder ersetzen kann, gehe ich davon aus, dass das Tool Mensch in der Sozialarbeit und insbesondere im Kindesschutz nach wie vor von grösserer Bedeutung ist als ein KI-Tool. Der systemische Aspekt, der speziell im Kindesschutz bedeutungsvoll ist, kann KI nicht abbilden.	Emotionen usw. kann KI nicht erfassen, SA bliebt deshalb von grosser Bedeutung, systemischer Aspekt kann nicht abgebildet werden	Defizite KI: Emotionen	neutral
33	P33.1	An KI interessiert mich, dass sie Texte formulieren kann und nicht, dass sie selbständig Entscheidungen trifft.	KI interessiert nur für administrative Arbeiten	Administrative Arbeiten	
34	P34.1	Ich danke für die Auseinandersetzung mit dem KI. Ich denke, es gibt kein Berufszweig, welcher sich nicht mit dem Thema auseinandersetzen muss. KI bietet Risiken, aber auch Chancen.	-	-	-
35	P35.1	Grundsätzlich wäre es spannend, wenn die KI im Bereich der bestehenden Klientensoftware bspw. eine Aktenzusammenfassung (bspw. für die period. Rechenschaftsberichte), Auswertung von Handlungsplänen erstellen und bspw. Fehler oder Mängel in der BuHa überprüfen könnte.	Aktenzusammenfassungen, Rechenschaftsberichte, Buchhaltung	Administrative Arbeiten	neutral
36	P36.1	Ich denke dies ist der Wandel der Zeit und auch wir Sozis dürfen/müssen/sollen uns mit KI auseinandersetzen.	Wandel der Zeit	Wandel der Zeit	neutral
	P36.2	Dennoch handelt es sich um Menschen und deren Lebenswelten die wir begleiten in äusserst schwierigen Situationen. Ob die KI dabei wirklich eine Stütze sein kann, da habe ich noch meine Zweifel.	individuelle Lebenswelten, schwierige Sitiuationen	Defizite KI: Einzelfall	negativ
37	P37.1	Künstliche Intelligenz ist etwas wunderbares, solange diese nicht die Verantwortung der eigenen Entscheidungen abnimmt oder gar rechtfertig.	KI ist positiv, solange keine Verantwortung abgegeben wird oder rechtfertigt	Verantwortung	neutral
	P37.2	Soziale Arbeit ist Beziehungsarbeit, etwas das KI nicht kann. Beziehungsarbeit benötigt Zeit, Durchhaltevermögen und Rückgrad und vor allem die Liebe zu Menschen. Auch das kann KI nicht.	SA ist Beziehungsarbeit, was KI nicht kann, das braucht Zeit, Durchhaltevermögen, Rückgrat, Liebe	Defizite KI: zwischenmenschliche Beziehungen	negativ
	P37.3	Daten sammeln und Wahrscheiblichkeitsrechnungen mit konkreten Lösungsvorschlägen zu machen ist eine wichtige Unterstützung. Und kann zu Alternativen Lösungen führen. Nota bene ist es der Mensch mit den Daten UND seiner Beziehungsfähigkeit zum hoffentlich erfolgreichen Prozedere zu führen.	Wahrscheinlichkeitsrechnungen wichtige Unterstützung, alterative Lösungen, Entscheidung gemeinsam SA und KI	Verantwortung	neutral
38	P38.1	Ich gehe davon aus, dass meine Einstellung gegenüber der Nutzung von KI im beruflichen Alltag sehr stark von den gemachten Erfahrungen abhängt. Falls sich die angegebene Chance bewahrheiten würde ich die Nutzung verstärken, falls nicht würde ich der KI kaum noch eine Chance geben.	Einstellung von Erfahrungen abhängig, falls Chance bewahrheitet, dann mehr Vertrauen/Nutzen, sonst kaum Chance	Erfahrungswerte	neutral
39	P39.1	im sinne der milderen massnahme würde ich die ki empfehlung immer dann nutzen, wenn diese eine mildere massnahme vorschlägt. wenn sie aber eine intensivere massnahme vorschlägt, würde ich wohl eher weiterhin auf der milderen massnahme beharren und diese zuerst einmal versuchen.	Nutzung wenn mildere Massnahme, dann KI, sonst eigene Einschätzung	Invasivität	neutral
40	P40.1	Wie im Beispiel 2 gibt es manchmal verfahrene Situation, bei denen guter Rat teuer ist . Ich sehe insbesondere in diesen Situationen ein grosses Potential für KI. Zur Zeit nutzen wir hier meist die Schwarmintelligenz im Team.	Potential bei WIEW, Ablösung Schwarmintelligenz	Wandel der Zeit	neutral
41	P41.1	Die Anwendung der KI muss gut eingeführt werden, um den richtigen Nutzen daraus erzielen zu können.	Gute Einführung, um Nutzen zu ziehen	Schulung	neutral

42	P42.1	Bei den Beispielen finde ich die Fragen dazu, ob ich die Empfehlung der künstlichen Intelligenz weniger oder intensiver gewichtet habe, schwierig zu beantworten. Ich weiss gerade nicht, ob mir im Beispiel Informationen fehlen oder die Fragen bewusst in dieser Form gestellt wurden. Die künstliche Intelligenz finde ich wichtig für die Entlastung bei administrativen Arbeiten (z.B. Aktennotizen mit Mikrofon erfassen, RB, Stellungnahmen und Abklärungsberichte, Finanzpläne und Zahlungen anhand der hinterlegten Aktennotizen und Unterlagen). Dies sind aber nur Beispiele und es bedarf einem zugrundeliegendem Konzept in den jeweiligen kjz.	Entlastung administrative Arbeiten		neutral
	P42.2	Auch für die fachliche Einschätzung des Hilfebedarfs und der Hilfeplanung anhand wissenschaftlicher Studien ist es zielführend, zunehmend künstliche Intelligenz zu nutzen. Nur so können aktuelle wissenschaftliche Ergebnisse im Alltagshandeln nutzbar gemacht werden.	Wissenschaftliches Wissen	Wissenschaftliches Wissen	neutral
	P42.3	Hier wäre interessant, ob die künstliche Intelligenz nebst den Falldaten im Kanton auch auf die Daten der wissenschaftlichen Organisationen und Wissensbeständen (Online-Bibliotheken, Fachjournale usw.) zugreifen kann. Ebenfalls aus der Forschung ist bekannt, dass die Bedürfnisse und die Ziele der Familien für erfolgreiche Veränderungen einzubeziehen sind und auch die Beziehung zwischen der Fachperson und der Familie diesbezüglich Einfluss nimmt. Künstliche Intelligenz kann also bei einzelnen Aufgabenbereichen die Fachpersonen in der Kinder- und Jugendhilfe unterstützen oder ihnen gemäss den normativen, fachlichen und organisatorischen Rahmenbedingungen Grundlagen und Abläufe schaffen, welche einerseits helfen einen professionellen Habitus aufzubauen, Leistungen qualitativ und effizient zu erbringen und diese messbar bzw. vergleichbarer zu machen. Somit		Qualitätssteigerung	neutral
43	P43.1	Aus meiner Sicht wird es nie möglich sein, Fachpersonen der Sozialen Arbeit vollumfänglich durch KI zu ersetzen und dies ist für mich ein sehr beruhigender Gedanke ;O)	keine Angst vor Jobverlust	Defizite KI: zwischenmenschliche Beziehungen	neutral
44	P44.1	Schlussendlich tragen weiterhin wir die Verantwortung für die Entscheidungen. Ich finde es schwierig, wenn wir unsere Entscheidungen in diesen sehr einschneidenden Kindesschutzthemen an einen Wahrscheinlichkeitsberechner übergeben. Dann kann man ja dann auch sagen, dass hat der KI so vorgeschlagen, wer übernimmt dann die Verantwortung? Wer übernimmt die Verantwortung, wenn etwas schiefgeht? Die Organisation welche sich entschieden hat die KI als Tool zu nutzen? Trotzdem die SA? Weil die KI wird es nicht sein. Wir tragen in dieser Rolle eine grosse Verantwortung, wo wir uns nicht hinter eine KI verstecken können dürften. Fände das auch nicht gerecht den Klienten gegenüber.	Verantwortung für Entscheide, Wahrscheinlichkeitsberechnung kritisch	Verantwortung	neutral
45	P45.1	Das erste KI Abklärungstool klingt interessant, werden aktuell Diskussionen darüber geführt, ob dieses in den kjz eingesetzt werden kann? Falls nicht, kannst du das ankurbeln? :)	-	-	-
46	P46.1	verwende aktuell KI punktuell als Argumentensammlung bsp. was sind positives Effekte eine Boxtraining für ein Jugendlicher oder welche Einfluss haben Konflikte zwischen KE auf die Kinder Vorteil, man erhält schnell einen Überblick / Nachteil, man muss weniger selbst nachdenken und wird bequem	bisherige Verwendung als Argumentsammler, Nachteil: Bequemlichkeit	Defizite KI: kritisches Denken	neutral
47	P47.1	KI-Einsatz in der Admin. der SA hört sich gut an	Einsatz in Administration	Administrative Arbeiten	positiv
48	P48.1	Aus meiner Sicht kann der Nutzen von KI in der sozialen Arbeit gegeben sein, u.a. wenn mit Stichworten ein RB erstellt werden kann, oder KI die Antragsstellung mit formuliert, d.h. administrative Arbeiten reduziert werden.	Nutzung in administrativen Arbeiten	Administrative Arbeiten	positiv
	P48.2	Das KI Massnahmen ableitet auf Grund von Erfahrungswerten aus anderen Fällen sehe ich äusserst kritisch, zumal stellt sich die Frage für aus welcher Sichtweise (Fachperson, Klient:in, KESB) und unter welchen Bedingungen eine Intervention als erfolgreich zu werten ist.	kritische Sicht bei Herleitungen, Wann Bedingung erfolgreich?	Funktionsweise	neutral
	P48.3	Die Nutzung von evidenzbasiertem Fachwissen erachte ich grundsätzlich als unterstützend,	unterstützendes wissenschaftliches Wissen	Wissenschaftliches Wissen	neutral
	P48.4	dass aus diesem Interventionen abgeleitet werden würde erachte ich zugleich auch als kritisch, da in diesem Fall keine Einzelfallbewertung vollzogen werden würde.	Kritisch bei Interventionen, weil Einzelfall	Defizite KI: Einzelfall	negativ
49	P49.1	Ich habe mich bisher in keiner Weise damit beschäftigt, dass KI in meiner Arbeit ein Thema sein könnte - daher fand ich diese Umfrage sehr spannend! Mich würde interessieren, ob sich solche Tools bereits in der Entwicklung befinden?	-	-	-
50	P50.1	Jede person sollte eine individuelle chance erhalten, auch wenn erfolgschancen rein rechnerisch gering sind.	individuelle Chance	Defizite KI: Einzelfall	neutral
51	P51.1	Grundsätzlich fände ich die Anwendung von KI in unserem Beruf gut und ich denke es würde vieles in unserem Arbeitsalltag unterstützen und uns entlasten.	Anwendung befürwortet, entlastend in Arbeitsalltag	Entlastung	neutral
52	P52.1	KI würde ich als Tool nutzen für Handlungen in bestimmten komplexen Themen/Fällen. KI könnte evtl. meine Sichtweise anders beurteilen oder blinde Flecken aufzeigen. Diese sind nicht per se negativ zu beurteilen.	Nutzung in komplexen Fällen, könnte Blinde Flecken aufzeigen	Reflexion	positiv
	P52.2	Die menschliche Intelligenz scheint mir jedoch immer noch einen wichtigeren Teil zu haben und muss höher bewertet werden. Der Verlass sich nur noch auf KI stützen zu können löst ein ungutes Gefühl in mir aus.	Menschliche Intelligenz wichtigere Entscheidungen, höher gewertet, ungutes Gefühl wenn nur Verlass auf KI	Verantwortung	neutral
53	P53.1	Für mich sind drei Fragen zur KI im Kindesschutz wichtig: Wer programmiert diese, mit welchem beruflichen, ethischen Hintergrund und mit welchem Ziel? Auf Basis welcher Daten lernt die KI? Dabei würde ich organisationsinternen Daten den Vorzug geben, weil Organisationskultur die Bewertung prägt (z.B. welches Verhalten zu Interventionen führt) Wie werden Daten über die Kindesschutzabklärung erfasst?	Wie wird KI programmiert, berufliche, ethische Hintergrüde, Ziele, Lernbasis, Organisationsinterne Daten, Organisationskultur prägt Interventionsverhalten, Datenerfassung	Funktionsweise	neutral
54	P54.1	KI ist sehr interessant, aber noch so unerforscht.			
55	P55.1	Es ist höchste Zeit, dass auch in der sozialen Arbeit die Nutzung von Digitalen Mitteln vermehrt zugelassen wird. Damit es funktioniert, müssen auch die Menschen in Leitungspositionen dahinter stehen können. Ich bin mir unsicher ob dies schon möglich ist, oder ob es ein Generationenwechsel braucht. Die Zeit wird es zeigen.	Digitale Mittel soll vermehrt zugelassen werden, Leitungspersonen müssen dahinterstehen können, aktuell noch schwierig (Generationenwechsel)	Wandel der Zeit	positiv

	P55.2	KI wäre in Administrativen Angelegenheiten sicherlich gut, weil wir dann endlich unserer sozialarbeiterischen Ressourcen besser nutzen könnten und dies vermehrt zu Verbesserungen im Klientensystem führen könnte.	Administrative Angelegeneheiten, Ressourcen können besser genutzt werden, Verbesseungen in Klientensystemen	Administrative Arbeiten	positiv
56	P56.1	Ich kann mir vorstellen, dass KI auch bei der Erstellung von Berichten und Stellungnahmen eine grosse Entlastung sein kann.	Erstellung Berichte und Stellungnahmen	Administrative Arbeiten	positiv
57	P57.1	Um zu Beurteilen, ob ich den Empfehlungen einer KI 'vertrauen' kann, müsste ich wissen an Hand von welchen Informationen (nur Fallbeispiele oder auch Literatur? falls ja, welche Literatur etc.) es die Empfehlung gibt.	Einschätzung aufgrund welcher Informationen	Funktionsweise	neutral
	P57.2	Zudem ist bekannt, dass auch die Literatur und die Forschung dem Sozialen Wandel immer hinterherhinkt.	Literatur hinkt hinterher	Wissenschaftliches Wissen	neutral
	P57.3	Zudem würde mir die Empfehlung nicht genügen, sondern ich müsste die 'Abwägungen/Argumentationen' kennen, damit ich die Empfehlung überhaupt in Betracht ziehen würde. Eine Empfehlung alleine reicht mir nicht, da wir aus der Pluralität unserer Handlungsspielraumes genau wissen, dass es für jegliche Entscheidungen je nach Gewichtung und 'Eloquenz' irgendwelche Argumentationen gefunden werden können.	Empfehlung reicht nicht, Empfehlung nur in Betracht ziehen mit Hintergrundkenntnissen	Transparenz	neutral
	P57.4	Zudem bin ich unsicher, wie eine KI ein so komplexes System wie einen Mensch v.a. ein Kind und die Umweltfaktoren überhaupt erfassen kann inkl. den 'kleinen' Eindrücken/ Beobachtungen und manchmal auch einfach Bauchgefühl aus Interaktionen mit z.B. Eltern einer Fachperson.	Unsicher, ob KI komplexe Systeme und Umweltfaktoren verstehen kann, Detailbetrachtung, z.T. nur Bauchgefühl aus Interaktionen	Defizite KI: Einzelfall	neutral

10.2 Anhang 2 - Whatsapp-Umfrage zur erweiterten Recherche, 27.11.2023 - 29.11.2023

«Wenn du daran denkst, dass du zukünftig in deinem Arbeitsalltag ein KI-basiertes Tool einsetzen könntest, was wären 3 positive Gedanken und 3 negative Gedanken, die dir als erstes in den Sinn kommen dazu (Hoffnungen, Sorgen, ect.)?.»

Person	Aussage	Jhrg	Arbeitsbereich	Textmodule	Thema	Kategorie 1	Kategorie 2	Ausprägung
1	1.1	1983	CaseManagement Berufsbildung	bricht klarer u präziser verfasst	Berichte, präziser	Leistungserwartung		positiv
	1.2			zyt gspart wägem vorhärige punkt	Zeitersparnisse	Leistungserwartung		positiv
	1.3			bezgl. entscheidig träffe bini ultraskeptisch, dänke nid, dass es a mehrwert isch	Entscheidfindung, kein Mehrwert	Leistungserwartung		negativ
	1.4			wett nid amene ort schaffe, wo z tool bezgl. Entscheidig institutionalisiert isch.	Institutionalisierung Entscheide	Professionswissen / Professionsverständnis		negativ
2	2.1	1988	Sozialpädagogik, KiEl	Entlastig bide Entscheidigsfindig merkt sich die KI, wie früener i ähnliche Situatione vorgange, entschiede oder reagiert worde isch	Entscheidfindung, Vergleich	Leistungserwartung		positiv
	2.2			Reduktion vode Fallbelastig oder au allgemein de Fallzahle weniger Burnout dur das	Reduktion Fallbelastung, Zeit, Gesundheit	Leistungserwartung		positiv
	2.3			vielicht i ähnlich glagerete Fäll en Vorschlag für en Bricht, Stellignahm etc. wo mer denn chan individualisiere und apasse, ohni dass mer s Rad muess neu erfinde und mega viel Ziit brucht defür.	Berichte, Zeitersparnisse	Leistungserwartung		positiv
	2.4			ev. weniger persönliche Umgang mit Kliente und ev. höcheri Fehlerafälligkeit dur das	persönlicher Umgang, Fehleranfälligkeit	Fehlerrisiko	Professionswissen / Professionsverständnis	negativ
	2.5			De Klient chönt sich au weniger wertgschätzt und wahrgnoh fühle dur meh Automatisierig und weniger Kontakt miteme Mensch dur das lidet ev. au d Beziehig und notwendigs Vertraue	Beziehung, Wahrnehmung, Vertrauen,	Professionswissen / Professionsverständnis		negativ
	2.6			ev. Arbeitsstelleabbau, obwohl ich nöd glaube, dass mir würklich so schnell ersetzbar wäred	Stellenabbau, Ersetzbarkeit	Sozialpolitische Aspekte		negativ
	2.7			Lüüt setzt tool so i dass nüm selber denked und grossi fehler passiered	Fehler, selber denken	Fehlerrisiko	Professionswissen / Professionsverständnis	negativ
3	3.1	1985	AJB, Fallfinanzierung	dass die Qualität im Kindesschutz gesteigert werden könnte indem eine objektive Bewertung der Risiko- und Schutzfaktoren vorgenommen wird und geeignete Interventionen vorgeschlagen werden und das man Entlastung in administrativen Aufgaben erfährt.	Steigerung Qualität, Bewertung Risiko- & Schutzfaktoren, geeignete Interventionen, Zeitersparnisse Administration	Leistungserwartung		positiv
	3.2			KI löst bei mir Skepsis aus, weil man vlt. gewisse Gedankengänge nicht mehr trainiert	Gedankengänge werden nicht mehr trainiert	Professionswissen / Professionsverständnis		negativ
	3.3			man als Arbeitskraft nicht mehr gebraucht wird	nicht mehr gebraucht, Stellenverlust	Sozialpolitische Aspekte		negativ
	3.4			resp. die Anforderungen steigen, da KI vermeintlich Entlastung bringt und dass die SA stärker von Technik abhängig wird was IT-Firmen Macht gibt und man gezwungen ist technisch mitzuhalten, was im höheren Alter herausfordernd ist.	höhere Anforderungen, vermeindliche Entlastung, Stärkere Abhängigkeit, Machtverhältnisse, Technische Auseinandersetzung	Professionswissen / Professionsverständnis	Aufwandserwartung	negativ
4	4.1	1994	Berufsbeistandschaft Erwachsenenschutz	ressource besser chöne iiteile (meh zit für dä KL und somit füs wesentliche vo minere arbet)	Ressourceneinteilung	Professionswissen / Professionsverständnis	Leistungserwartung	positiv
	4.2			Zit und namal zit	Zeitersparnisse	Leistungserwartung		positiv
	4.3			natürli gad en gedanke dä dateschutz woni dänn chli bezwifle	Datenschutz	Datenschutz		negativ
	4.4			weniger mini eigeni sichtwies, mini hypothesene wo z'churz würet cho	eigene Hypothesen	Professionswissen / Professionsverständnis		negativ
5	5.1	1994	Spitalsozialarbeit	vielicht i ähnlich glagerete Fäll en Vorschlag für en Bricht, Stellignahm etc.	Berichte	Leistungserwartung		positiv
	5.2			Mit KI chönnti sicher au es case management eusi arbet mache	virtuelles CaseManagement	Leistungserwartung		positiv
	5.3			Im moment findi s thema dateschutz und sicherheit in Bezug uf KI sehr fraglich, wo mir hald scho mit sehr sensible date z tue hend	Datenschutz, sensible Daten	Datenschutz		negativ
	5.4			verlust vib arbetsplatz wird nemmli bi eus scho so vlt es thema	Verlust Arbeitsplatz	Sozialpolitische Aspekte		negativ
6	6.1	1992	Kinder- und Jugendhilfe, Berufsbeistandschaft	1. Ich wird entlasted	Ressourceneinteilung, Zeitersparnisse	Leistungserwartung		positiv
	6.2			2. Chan mich uf anderi arbeite direkt mit de mensche fokussiere und bin weniger absorbiert mit de 1000 brichte	Berichte	Leistungserwartung		positiv
	6.3			3. Bi entscheidige (nachdem ich z.b mit em team en ustusch gha han) evtl no en andere blickwinkel inimmt	Entscheidfindung, Vergleich, neuer Blickwinkel	Leistungserwartung		positiv
	6.4			1. Ohni dass AI alli detaills über de fall het denki nöd dass entscheidige e tüffi fachlichi fundiertheit werded ha	Details, keine fachliche Fundiertheit	Professionswissen / Professionsverständnis		negativ
	6.5			2. De dateschutz, weiss nöd wie wiit Ai da scho isch	Datenschutz	Datenschutz		negativ
	6.6			3. Chönti das bi de sozis e gwüssi angst uslöse früehner oder spöter ersetzt zwerde oder sgar zume stelleabbau chönti cho zum chöste spare	Angst, Ersetzt werden, Stellenabbau, Kosten sparen	Professionswissen / Professionsverständnis	Sozialpolitische Aspekte	negativ

7	7.1	1986	Kinder- und Jugendhilfe, Berufsbeistandschaft	Effizienz	Effizienz	Leistungserwartung		positiv
	7.2			Korrekte und besser uf dä punkt brachti formulierige	Berichte, Formulierung, präziser	Leistungserwartung		positiv
	7.3			Tool zum mitdenke und mich absichere	Absicherung	Leistungserwartung		positiv
	7.4			Lüüt setzt tool so i dass nüm selber denked und grossi fehler passiered	selber denken, grosse Fehler	Fehlerrisiko		negativ
	7.5			Es isch zwenig guät und ich muss glich no alles überprüfe.	nicht gut genug	Fehlerrisiko	Aufwandserwartung	negativ
	7.6			Es isch so kompliziert gmacht dassi länger bruch wiä wenni selber studiere wür.	Kompliziert, Zeitaufwand	Aufwandserwartung		negativ
	7.7			Qulitätsverluscht. Bi eus zb. Dass interdisziplinarität wür verlore gah o.ä…	Qualitätsverlust, Interdisziplinarität geht verloren	Professionswissen / Professionsverständnis		negativ
8	8.1	1993	AJB, Pflegekinder Behördenmitglied	arbeit effizienter/schneller gestalten (mehr Zeit für 'wichtigeres')	Effizienz, Zeitersparnisse, Resourceneinteilung	Leistungserwartung		positiv
	8.2			eine zusätzliche Meinung/Perspektive hilft wahrscheinlich beim Entscheideprozess (zusätzliche Kontrollfunktion bspw. bei der Einschätzung von Gefährdungen sowie aufzeigen von alternativen Lösungswegen	Entscheidfindung, Perspektive, Kontrollfunktion, Gefährdungseinschätzung, alternative Lösungswege	Leistungserwartung		positiv
	8.3			humankontrolle geht verloren, man überlegt dann vielleicht zu wenig, wenn die Arbeit von KI gemacht wird	Humankontrolle, zu wenig überlegen	Professionswissen / Professionsverständnis		negativ
	8.4			unsichere Handhabung resp. damit eingehender Verantwortungsverlust für fachpersonen z.B wer ist 'Schuld' falls das ki System zu Falscheinschätzungen kommt?	Verantwortung, Verantwortungsverlust, Schuld bei Fehler	Fehlerrisiko		negativ
	8.5			oder das 'wegfallen' / 'nicht mehr gebrauchen von Fachwissen' ist psychisch für Mitarbeitende mit Berufsstolz sicher nicht einfach anzunehmen und psychisch nicht zu unterschätzen	Fachwissen, Berufsstolz, Psychische Folgen	Professionswissen / Professionsverständnis		negativ
9	9.1	1985	AJB, Fallfinanzierung	1. es chönt helfe effizienter si bim verfasse vo doks, wie z.B. Berichte, Stellungnahmen, evtl. auch aktennotizen etc.	Effizienz, Berichte	Leistungserwartung		positiv
	9.2			2. es chönt hilfrich sii für ideesammlige, wenns wüsse i de ki gnueg gross und differenziert ist, wie z.B. sätz anderst schribe, kreativi idee für lösigsfindige, allefalls auch idee für heim etc. wemer keis meh weis und schon vili ahgfragt het/	Ideensammlung, Differenzierung, Kreativität, alternative Lösungsfindung	Leistungserwartung		positiv
	9.3			3. vlt. wäred durch ki automatisierige möglich wo so bis jetzt nid möglich gsi sind..... es konkrets bispiel hani jetzt aber gad nöc	Automatisierung	Leistungserwartung		positiv
	9.4			1. ki lernt vo dem wos selber weis, i dem wüsse chönt fehler drin si, wod ki nid selber merkt und den so wiiterverwendet wird, wenns z.B um ihschätze zu öppisem gaht (will z.B. die wo ki programieret, hend vlt. zu wenig spezifisches wüsse oder verstönd öppis anderst als es isch, aso dcodes wo dehinder sind) z.B. wie das mit de betreuuigsgutschrifte in dütschland oder de gförderihschätzig bi gwaltdelikt i de usa	Selbstlernfaktor KI, Fehlerreproduktion, zu wenig spezifisches Wissen, anderes Verständnis	Fehlerrisiko		negativ
	9.5			2. es tuet wahrschindlich fördere, dasmer selber nüm würklich überleit und s denn auch überprüeft was ki macht (richtig/falsch).	Selber denken, Fehler	Professionswissen / Professionsverständnis	Fehlerrisiko	negativ
10	10.1	1989	AJB, Pflegekinder Behördenmitglied	meh Ziit für Kliente	Zeitersparnisse	Leistungserwartung		positiv
	10.2			wahrscheinlich mehr Objektivität bzw. Gleiche Linie für alle	mehr Objektivität, Gleichbehandlung	Leistungserwartung		positiv
	10.3			ordentliche Berichte ohne Grammatikfehler	Berichte	Leistungserwartung		positiv
	10.4			kein Ermessensspielraum mehr (bspw. Gesetz so auslegen, dass es zugunsten der Klienten ist)	Ermessensspielraum	Professionswissen / Professionsverständnis		negativ
	10.5			Datenschutz	Datenschutz	Datenschutz		negativ
	10.6			Hacker bzw. Manipulation des Tools	Hacker, Manipulation	Fehlerrisiko		negativ
11	11.1	1991	Sozialhilfe, Intake	Reduktion Administrative Uufwand	Ressourceneinteilung, Administrative Einsparungen	Leistungserwartung		positiv
	11.2			meh ziit für effektivi Fallfüehrig, resp. Zwpschemenschliche kontak	Zeitersparnisse, mehr Zeit für KL	Leistungserwartung		positiv
	11.3			einheitlichi ablage, digitalisierige usw. Wo alles dur KI gmacht wird.	Organisation, Digitalisierung	Leistungserwartung		positiv
	11.4			Sorge dass KI-Entscheid falsch sind und us gwohnheit vode fallfüehrende trotzdem eifach so hiignoh werded	Fehlerhafte Entscheide, Akzeptanz Fehler	Fehlerrisiko		negativ
	11.5			bequemlichleit ide Fallfüehrig, ken grund meh selber verantwortig zträge	Bequemlichkeit, Verantwortung	Professionswissen / Professionsverständnis		negativ
	11.6			mini grösst sorg sind dmensche dehinder. Isch sehr negativ aber ich glaube die meiste sind fuul uf de arbet und lönd somit ki meh mache als nötig oder überprüefets nümm	Faulheilt, keine Überprüfung	Professionswissen / Professionsverständnis		negativ
12	12.1	1967	AJB, Pflegefamilie Abklärungen	1. unterstützung / abnahme schriftliche Arbeit (ich spreche und die KI schreibt).	Berichte, Spracherkennung	Leistungserwartung		positiv
	12.2			2. standardisierte Arbeiten abgeben (Registerauszüge und Dokumente einfordern)	Administrative Einsparung, Zeitersparnisse	Leistungserwartung		positiv
	12.3			3. fachliches wissen für mich suchen.	Fachwissen	Leistungserwartung		positiv

	12.4			1. keine kontrolle haben über die KI (ich hätte chli ein mulmiges gefühlt, weil ich sie evtl. nicht totalüberwachen kann.	unklare Überwachung, Undurchsichtigkeit	Nachvollziehbarkeit		negativ
	12.5			2. KI macht fehler?	Fehler	Fehlerrisiko		negativ
	12.6			3. ich hätte glaubs immer noch das gefühl, ich müsste dann alles nochmals kontrollieren und nacharbeiten	Prüfung, kein Vertrauen, Mehraufwand	Fehlerrisiko	Aufwandserwartung	negativ
	12.7			han null bock mich mit däm züg usenand z setzä. auch wänns mi no würd interessierä...aber mach lieber anders...:-)))	Auseinandersetzung, Umgewöhnung	Gewohnheit	Aufwandserwartung	negativ
13	13.1	1990	AJB, Fallfinanzierung	meine Produktivität wird gesteigert	Produktionssteigerung	Leistungserwartung		positiv
	13.2			mühsame Admin-Sachen werden weniger	Administrative Einsparung, Zeitersparnisse	Leistungserwartung		positiv
	13.3			meine Kreativität wird gesteigert	Kreativitätssteigerung	Leistungserwartung		positiv
	13.4			durch ChatBots kann die Beratung für Klientel verbessert werden (bspq. Chatbots ausserhalb der Bürozeiten)	Verbesserung Beratung, Virtuelle Beratung	Leistungserwartung		positiv
	13.5			Abhänhigkeit durch KI	Abhängigkeit	Professionswissen / Professionsverständnis		negativ
	13.6			moralisches Denken und Risikoeinschätzung wird vielleicht berechnet?!	Moralisches Denken, Risikoeinchätzung	Professionswissen / Professionsverständnis		negativ
	13.7			Bauchgefühl und Emotionen rücken in den Hintergrund	Bauchgefühl, Intuition, Hintergrund	Professionswissen / Professionsverständnis		negativ
14	14.1	1959	AJB, Pflegefamilie Abklärungen	Zeitersparnis, insbesondere bei "eintönigen" Aufgaben	Zeitersparnis, Administrative Aufgaben	Leistungserwartung		positiv
	14.2			Anregungen für Formulierungen und Formatierungen	Anregungen, Intuition, Formatierungen	Leistungserwartung		positiv
	14.3			Entscheidungen sind bei uns richtigerweise im Team oder bei uns Menschen mit umfangreicher Menschenkenntnis anzusiedeln.	Entscheide bei Mensch/Team, Menschenkenntnis	Professionswissen / Professionsverständnis		negativ
	14.4			Einhalten von Datenschutz unmöglich.	Datenschutz	Datenschutz		negativ
15	15.1	1991	Berufsbeistandschaft EKSM	In bezug uf admin arbete: eh bsuechsplaning generiere ahand vo de regelig, wür viel admin abneh	Ressourcenersparnisse	Leistungserwartung		positiv
	15.2			in bezug uf bricht: Wür ziit spare	Zeitersparnisse	Leistungserwartung		positiv
	15.3			Bedenke, dass mini aktennotize „richtig“ interpretiert werde	Korrekte Interpretation	Fehlerrisiko		negativ
16	16.1	1978	CaseManagement Berufsbildung	Unterstützung bei Berichtschreiben, basierend auf Journaleinträgen sind effizient, Fazit aus Journal (Schlussfolgerung)	Berichte, Zeitersparnisse, Effizienz, Journaleinträge	Leistungserwartung		positiv
	16.2			Diagnosestellung von Mensch, Zusammenfassung von wichtigsten Inhalten von Diagnosestellung (Literaturrecherche)	Literaturrecherche	Leistungserwartung		positiv
	16.3			Schärfung von Ausformulierung bei Interventionen, wenns um Berichterstattung an andere Fachstellung geht	Klarere Berichte, Formulierunugen	Leistungserwartung		positiv
	16.4			Journaleinträge: sind eigene Worte, wenn nicht, dann "wurde es nicht durch meine Feder geschrieben"	nicht durch eigene Hand	Professionswissen / Professionsverständnis		negativ
	16.5			ganzes Fachwissen aus Alltag kommt recht reduziert daher	Reduziertes Fachwissen	Professionswissen / Professionsverständnis		negativ
	16.6			Beschreibung zwischenmenschliche Beziehung und Kooperation - wie das wiedergegeben werden könnte, bin ich nicht sicher (obwohl ich glaube, dass das KI sogar noch könnte, aber ich bin skeptisch)	Zwischenmenschliche Beziehungen, Kooperation, Wiedergabe unklar	Professionswissen / Professionsverständnis		negativ
17	17.1	1992	Leiter Behördendienste	Administrative Angelegenheiten: Zeitersparnisse und immer gewisser Standard an Qualität (bei Fachkräftemangel nicht immer der Fall)	Zeitersparnisse, Standard an Qualität, Fachkräftemangel	Leistungserwartung		positiv
	17.2			Standardarbeiten erledigen - hier ist KI super	Repetitive Arbeiten	Leistungserwartung		positiv
	17.3			kann nicht immer ganz so individuell auf Fälle eingehen oder kann nicht alle Faktoren einbeziehen	Individualität, Einbezug weitere Faktoren	Professionswissen / Professionsverständnis		negativ
	17.4			Menschlichkeit/Empathie geht verloren, wenn Mitarbeiterentscheid getroffen wird, weiss ich nicht, ob der das so lustig finden würde, wenn man das auf Basis von einem Tool machen würde	Menschlichkeit, Empathie geht verloren	Professionswissen / Professionsverständnis		negativ
	17.5			KI hört da auf, wo der Mensch aufhört sich Gedanken zu machen	Mensch macht sich keine Gedanken mehr	Professionswissen / Professionsverständnis		negativ
18	18.1	1992	AJB, Trägerschaften	Reduktion administrativer Tätigkeiten	Administrative Einsparung, Zeitersparnisse	Leistungserwartung		positiv
	18.2			Eingrenzung von Fallführenden Vorgehensweisen	Vorgehensweisen vereinheitlicht	Professionswissen / Professionsverständnis		positiv
	18.3			Objektivität	Objektivität	Leistungserwartung		positiv
	18.4			Reduktion des Arbeitens über Beziehung, Distanz zum Klienten/In	Reduktion Beziehung zu Klient:in, Distanz	Professionswissen / Professionsverständnis		negativ
	18.5			aktualisierte Wissensgenerierung, wie wird der gesellschaftliche Wandel von KI aufgebommen und das Veränderte Handeln von KlientInnen	Wissensgenerierung, gesellschaftlicher Wandel, verändertes Handeln Klient:in	Professionswissen / Professionsverständnis		negativ
	18.6			technische Probleme	technische Hürden	Aufwandserwartung		negativ
19	19.1	1962	AJB, Pflegefamilie Abklärungen	geringerer Zeitaufwand	Zeitersparnisse	Leistungserwartung		positiv
	19.2			neue und andere Ideen für Gestaltung	Kreativitätssteigerung	Leistungserwartung		positiv
	19.3			Hilfe beim Formulieren	Formulierungshilfe	Leistungserwartung		positiv

19.4	fehlende Individualität	fehlende Individualität	Professionswissen / Professionsverständnis		negativ
19.5	Fehler können sich leichter einschleichen	Fehler	Fehlerrisiko		negativ
19.6	sich zu stark auf KI verlassen	Verlass, Vertrauen	Fehlerrisiko	Professionswissen / Professionsverständnis	negativ

Anhang 10.3 – Fragebogen

1 A) Einführung

Liebe Teilnehmende

Der Gegenstand meiner Befragung ist die Wahrnehmung von Unterstützungstools im Kindesschutz, welche auf Künstliche Intelligenz (KI) gestützt sind. Die Befragung richtet sich an Sozialarbeitende, die im gesetzlichen Rahmen in der Kinder- und Jugendhilfe und/oder im Kindesschutz arbeiten.

Digitale Technologien sind immer häufiger in der Sozialen Arbeit anzutreffen und bieten im Arbeitsalltag sowohl Chancen als auch Risiken. Im Folgenden möchte ich Sie zu Ihrer persönlichen Einschätzung im Hinblick auf mögliche KI-Tools befragen, wie sie Ihnen in Zukunft begegnen könnten. Dabei gibt es weder falsche noch richtige Antworten – mich interessiert Ihre ganz individuelle Meinung.

Bitte lesen Sie daher die folgenden Aussagen aufmerksam durch und beantworten Sie ehrlich und spontan, inwieweit Sie den Aussagen zustimmen. **Ich bitte Sie, den Fragebogen vollständig auszufüllen.** Einige der folgenden Aussagen mögen Ihnen dabei recht ähnlich vorkommen. Dies ist allerdings beabsichtigt und hilft mir dabei, Ihre Einschätzungen insgesamt besser zu verstehen. Ihre Antworten werden anonym und vertraulich behandelt und ausschliesslich zu Forschungszwecken verwendet.

Für Ihre Unterstützung bedanke ich mich herzlich. Bei inhaltlichen Fragen oder bei technischen Problemen dürfen Sie sich gerne an mich wenden (muellsi7@students.zhaw.ch).

Herzliche Grüsse,

Simone Müller

2 Filter Kinder- und Jugendhilfestelle

Arbeiten Sie in einer Kinder- und Jugendhilfestelle und/oder führen Sie im gesetzlichen Rahmen Beistandschaften für Minderjährige?

- ○ Ja
- ○ Nein

2.1.1 Endseite keine Kinder- und Jugendhilfestell

Herzlichen Dank für Ihr Interesse an der Umfrage.

Da sich die Befragung ausschliesslich an Personen richtet, die in Kinder- und Jugendhilfestellen arbeiten, ist die Umfrage an dieser Stelle für Sie beendet. Sie können das Fenster nun schliessen.

3 B) Soziodemografische Daten

Im Folgenden bitte ich Sie, einige Fragen über sich selbst zu beantworten.

In welchem Kanton befindet sich Ihre Arbeitsstelle?

Wie viele Jahre Berufserfahrung haben Sie im Bereich gesetzliche Kinder- und Jugendhilfe/Kindesschutz?

- ○ weniger als ein Jahr
- ○ 1 bis 2 Jahre
- ○ 3 bis 4 Jahre
- ○ 5 bis 6 Jahre
- ○ 7 bis 8 Jahre
- ○ 9 bis 10 Jahre
- ○ mehr als 10 Jahre

Welches ist die höchste Ausbildung, die Sie abgeschlossen haben?

- ○ Berufslehre
- ○ Gymnasiale Maturität, Berufs- oder Fachmaturität
- ○ Höhere Fachschule HF, höhere Berufsbildung (eidg. Fachausweis, Diplom)
- ○ Bachelor FH, PH, Uni, ETH, Diplom FH
- ○ Master FH, PH, Uni, ETH oder Lizenziat
- ○ Doktorat, PhD
- ○ Andere (ohne Weiterbildungsabschlüsse)

In welchem Fach/in welchen Fächern haben Sie Ihren höchsten Abschluss?

Bitte wählen Sie nur Fächer Ihres höchsten Abschlusses, es können auch mehrere Fächer angewählt werden.

- [ ] Soziale Arbeit (inkl. Sozialarbeit, Sozialpädagogik, Soziokulturelle Animation)
- [ ] Recht
- [ ] Psychologie
- [ ] Pädagogik, Erziehungswissenschaften, Sonder-/Heilpädagogik
- [ ] Ethnologie
- [ ] Soziologie
- [ ] Medizin
- [ ] Andere Fachbereiche

In welchem Jahr haben Sie Ihren höchsten Abschluss erreicht?

Bitte geben Sie die Jahreszahl ein (z.B. 2015)

Welchem Geschlecht fühlen Sie sich zugehörig?

- ○ weiblich
- ○ männlich
- ○ andere/fühle mich nicht zugehörig

Wie alt sind Sie? (in Jahren)

- [ ] 20-24
- [ ] 25-29
- [ ] 30-34
- [ ] 35-39
- [ ] 40-44
- [ ] 45-49
- [ ] 50-54
- [ ] 55-59
- [ ] 60-64
- [ ] 65+

4 C) Digitale-Affinität

Im Folgenden werden Ihnen Fragen zum Umgang mit digitalen Systemen gestellt.

Unter digitalen Systemen werden digitale Anwendungen (z.B. Software oder Apps) und digitale Geräte (z.B. Computer oder Smartphone) verstanden.

	trifft überhaupt nicht zu	trifft eher nicht zu	trifft teilweise zu	trifft eher zu	trifft voll und ganz zu
Ich kann digitale Systeme bedienen.	○	○	○	○	○
Ich bin gut darin, digitale Systeme zu nutzen.	○	○	○	○	○
Ich lerne schnell, wenn es um die Nutzung digitaler Systeme geht.	○	○	○	○	○
Mir fällt es leicht, mich an neue digitale Systeme zu gewöhnen.	○	○	○	○	○
Ich war schon immer gut im Umgang mit digitalen Systemen.	○	○	○	○	○

Die folgenden Fragen beziehen sich auf Ihre Einstellung zu Künstlicher Intelligenz im Allgemeinen.

	trifft überhaupt nicht zu	trifft eher nicht zu	trifft teilweise zu	trifft eher zu	trifft voll und ganz zu
Ich bin beeindruckt, was Künstliche Intelligenz leisten kann.	○	○	○	○	○
Ich interessiere mich dafür, künstlich intelligente Systeme in meinem täglichen Leben zu verwenden.	○	○	○	○	○
Künstliche Intelligenz ist aufregend.	○	○	○	○	○
Künstlich intelligente Systeme können besser als Menschen performen.	○	○	○	○	○
Ich möchte Künstliche Intelligenz in meinem eigenen Job nutzen.	○	○	○	○	○

	trifft überhaupt nicht zu	trifft eher nicht zu	trifft teilweise zu	trifft eher zu	trifft voll und ganz zu
Organisationen nutzen Künstliche Intelligenz unethisch.	○	○	○	○	○
Ich denke, dass künstlich intelligente Systeme viele Fehler machen.	○	○	○	○	○
Ich finde Künstliche Intelligenz unheimlich.	○	○	○	○	○
Ich denke, Künstliche Intelligenz ist gefährlich.	○	○	○	○	○
Künstliche Intelligenz könnte die Kontrolle über Menschen übernehmen.	○	○	○	○	○

5 D) Professionelle Entscheide

Wenn Sie an Ihren Arbeitsalltag denken, inwiefern stimmen Sie den folgenden Aussagen zu?

	stimme überhaupt nicht zu	stimme eher nicht zu	stimme teilweise zu	stimme eher zu	stimme voll und ganz zu
Bei Entscheidungen verlasse ich mich hauptsächlich auf mein Bauchgefühl.	○	○	○	○	○
Bei Entscheidungen im Arbeitsalltag liege ich mit meiner Intuition meistens richtig.	○	○	○	○	○
In Entscheidungssituationen lege ich grösseren Wert auf das, was mein Gefühl mir sagt, als auf nüchterne Analysen.	○	○	○	○	○
Ich vertraue mehr auf das, was ich denke, als auf die Einschätzung anderer.	○	○	○	○	○

	stimme überhaupt nicht zu	stimme eher nicht zu	stimme teilweise zu	stimme eher zu	stimme voll und ganz zu
Ergebnisse aus empirischen Studien sind wichtig für die tägliche Arbeit in der Sozialen Arbeit.	○	○	○	○	○
Wissenschaftliches Wissen hilft mir, professionelle Entscheidungen treffen zu können.	○	○	○	○	○
Erfahrungswissen ist in der Praxis der Sozialen Arbeit wichtiger als wissenschaftliches Wissen.	○	○	○	○	○
Der Bezug zu wissenschaftlichem Wissen sollte in der Praxis eine grössere Rolle spielen.	○	○	○	○	○

6 E) Einleitung Vignetten

Einführung Fallbeispiele

Im Folgenden sehen Sie zwei kurze Beschreibungen von unterschiedlichen auf künstlicher Intelligenz basierenden Tools, wie sie Ihnen in Zukunft bei der Arbeit begegnen könnten.

In der Beurteilung geht es in erster Linie um Ihre Einstellung zu entsprechenden Tools. Die Situationsbeschreibungen sind bewusst limitiert gehalten und widerspiegeln nicht die Fallkomplexität der Realität. Bitte beurteilen Sie die Fallbeispiele auf Basis der dargestellten Informationen, so ist es mir möglich, aussagekräftige Tendenzen zu ermitteln.

Bitte lesen Sie die Fallbeispiele sorgfältig und beantworten Sie alle Fragen.

6.1.1 E1 Fallbeispiel 1 KIBAT 1A2H

Fallbeispiel 1 - Abklärungstool KIBAT

Seit drei Monaten arbeiten Sie gemeinsam mit einer Kollegin an einer Kindeswohlabklärung in einer Familie. Es haben verschiedene Hausbesuche und Gespräche stattgefunden, zusätzlich gab es einen intensiven Schriftverkehr mit den Eltern und Fachpersonen via Brief und E-Mail. Sie haben während der gesamten Abklärung zahlreiche Aktennotizen erstellt, um alle Details festzuhalten. Nach Abschluss sämtlicher Gespräche und nach dem Sie genügend Informationen gesammelt haben, werten Sie die Abklärung gemeinsam aus. Sie kommen zum Schluss, dass eine Kindeswohlgefährdung gegeben ist. Sie erarbeiten folgende Handlungsmöglichkeit:

Eine Beistandschaft gemäss Art. 308 Abs. 1 und 2 ZGB mit einem Auftrag für eine intensive sozialpädagogische Familienbegleitung (SPF).

Zusätzlich wertet KIBAT, ein auf künstlicher Intelligenz basierendes Abklärungs-Tool, das vor kurzem in Ihrer Organisation implementiert wurde, sämtliche Informationen aus, welche Sie in den vergangenen Monaten gesammelt haben. Das Tool entwickelt aufgrund der vorhandenen Angaben eine unabhängige Einschätzung.

KIBAT errechnet, dass die von Ihnen erarbeitete Handlungsmöglichkeit mit einer Wahrscheinlichkeit von 60% zum Erfolg führt. Im Vergleich dazu empfiehlt das Tool eine intensivere Massnahme – also ein einschneidenderes Eingreifen in die Situation, als Sie es empfehlen würden. Für diese Handlungsoption gibt die KI eine Erfolgswahrscheinlichkeit von 90% aus. Unabhängige Studien haben gezeigt, dass die Empfehlungen von KIBAT in drei Vierteln der Fälle treffsicherer sind als diejenigen der Sozialarbeitenden.

Das Tool verweist darauf, dass seine Einschätzung aufgrund der Analyse tausender Falldaten zustande kam. Es verweist u.a. zusätzlich auf die Literatur von Hauri & Zingaro (2020), Kindler (2018), Klassen (2004), welche für die Einschätzung des Tools herangezogen wurden.

Die Entscheidung, ob Sie die Empfehlung von KIBAT in Ihren Abklärungsbericht aufnehmen, ist Ihnen überlassen.

Literatur:

Hauri, A., & Zingaro, M. (2020). Kindeswohlgefährdung erkennen und angemessen handeln: Leitfaden für Fachpersonen aus dem Sozialbereich (2., überarbeitete Aufl.). Bern: Kinderschutz Schweiz.
Kindler, H. (2018). Operationalisierungen von Kindeswohl und Kindeswohlgefährdung in den Sozial- und Humanwissenschaften. In H. Katzenstein, K. Lohse, G. Schindler, & L. Schönecker (Hrsg.), Das Recht als Partner der Fachlichkeit in der Kinder- und Jugendhilfe (S. 179–224). Baden-Baden: Nomos Verlagsgesellschaft mbH & Co.
Klassen, M. (2004). Was leisten Systemtheorien in der sozialen Arbeit? Ein Vergleich der systemischen Ansätze von Niklas Luhmann und Mario Bunge (1., Aufl). Bern: Haupt Verlag.

6.1.1.3 E2 Fallbeispiel 2 WIEW 1A2H

Fallbeispiel 2 - Interventionsplanung WIEW

Stellen Sie sich folgendes Szenario vor:

Seit sechs Jahren führen Sie eine Vormundschaft gemäss Art. 327a ZGB für die mittlerweile 16-jährige Hanna. Hannas Situation gestaltet sich sehr instabil: Sie ist psychisch labil, zeigt höchst oppositionelles Verhalten, konsumiert seit Jahren illegale Substanzen und seit kurzem auch vermehrt harte Drogen, welche sie sich vermutlich mit Beschaffungskriminalität finanziert. Die aktuelle Wohnsituation im betreuten Wohnen ist aufgrund des herausfordernden Verhaltens stark gefährdet.

In den vergangenen Jahren und Monaten haben Sie und weitere involvierte (Fach-)Personen verschiedene Interventionen unternommen, die jedoch weitgehend ohne Erfolg blieben. Trotz einer gewissen Unsicherheit entwickeln Sie jetzt gemeinsam mit dem Helfersystem einen neuen Interventionsplan, von welchem Sie sich erhoffen, dass er die gewünschte Wirkung erzielt.

Das KI-Tool WIEW soll Ihnen dabei helfen, weitere Interventionsmöglichkeiten auszuloten. WIEW entwickelt aufgrund der vorhandenen Fall-Informationen eine unabhängige Einschätzung sowie einen möglichen Interventionsplan für Hanna.

WIEW errechnet, dass die von Ihnen erarbeitete Handlungsmöglichkeit mit einer Wahrscheinlichkeit von 55% zum Erfolg führt. Im Vergleich dazu empfiehlt das Tool eine weniger intensive Massnahme – also ein weniger einschneidenderes Eingreifen in die Situation, als Sie es empfehlen würden. Für diese Handlungsoption gibt die KI eine Erfolgswahrscheinlichkeit von 85% aus. Unabhängige Studien haben gezeigt, dass die Empfehlungen von WIEW in der Hälfte der Fälle treffsicherer sind als diejenigen der Sozialarbeitenden. Das Tool verweist darauf, dass seine Einschätzung aufgrund der Analyse tausender Falldaten zustande kam.

Die Entscheidung, ob Sie die Empfehlung von WIEW in Ihre Interventionsplanung aufnehmen, ist Ihnen überlassen.

6.1.2 E3 Fallbeispiel 1 KIBAT 1B2G

Fallbeispiel 1 - Abklärungstool KIBAT

Seit drei Monaten arbeiten Sie gemeinsam mit einer Kollegin an einer Kindeswohlabklärung in einer Familie. Es haben verschiedene Hausbesuche und Gespräche stattgefunden, zusätzlich gab es einen intensiven Schriftverkehr mit den Eltern und Fachpersonen via Brief und E-Mail. Sie haben während der gesamten Abklärung zahlreiche Aktennotizen erstellt, um alle Details festzuhalten. Nach Abschluss sämtlicher Gespräche und nach dem Sie genügend Informationen gesammelt haben, werten Sie die Abklärung gemeinsam aus. Sie kommen zum Schluss, dass eine Kindeswohlgefährdung gegeben ist. Sie erarbeiten folgende Handlungsmöglichkeit:

Eine Beistandschaft gemäss Art. 308 Abs. 1 und 2 ZGB mit einem Auftrag für eine intensive sozialpädagogische Familienbegleitung (SPF).

Zusätzlich wertet KIBAT, ein auf künstlicher Intelligenz basierendes Abklärungs-Tool, das vor kurzem in Ihrer Organisation implementiert wurde, sämtliche Informationen aus, welche Sie in den vergangenen Monaten gesammelt haben. Das Tool entwickelt aufgrund der vorhandenen Angaben eine unabhängige Einschätzung.

KIBAT errechnet, dass die von Ihnen erarbeitete Handlungsmöglichkeit mit einer Wahrscheinlichkeit von 60% zum Erfolg führt. Im Vergleich dazu empfiehlt das Tool eine intensivere Massnahme – also ein einschneidenderes Eingreifen in die Situation, als Sie es empfehlen würden. Für diese Handlungsoption gibt die KI eine Erfolgswahrscheinlichkeit von 90% aus. Unabhängige Studien haben gezeigt, dass die Empfehlungen von KIBAT in drei Vierteln der Fälle treffsicherer sind als diejenigen der Sozialarbeitenden. Das Tool verweist darauf, dass seine Einschätzung aufgrund der Analyse tausender Falldaten zustande kam.

Die Entscheidung, ob Sie die Empfehlung von KIBAT in Ihren Abklärungsbericht aufnehmen, ist Ihnen überlassen.

6.1.2.3 E4 Fallbeispiel 2 WIEW 1B2G

Fallbeispiel 2 - Interventionsplanung WIEW

Stellen Sie sich folgendes Szenario vor:

Seit sechs Jahren führen Sie eine Vormundschaft gemäss Art. 327a ZGB für die mittlerweile 16-jährige Hanna. Hannas Situation gestaltet sich sehr instabil: Sie ist psychisch labil, zeigt höchst oppositionelles Verhalten, konsumiert seit Jahren illegale Substanzen und seit kurzem auch vermehrt harte Drogen, welche sie sich vermutlich mit Beschaffungskriminalität finanziert. Die aktuelle Wohnsituation im betreuten Wohnen ist aufgrund des herausfordernden Verhaltens stark gefährdet.

In den vergangenen Jahren und Monaten haben Sie und weitere involvierte (Fach-)Personen verschiedene Interventionen unternommen, die jedoch weitgehend ohne Erfolg blieben. Trotz einer gewissen Unsicherheit entwickeln Sie jetzt gemeinsam mit dem Helfersystem einen neuen Interventionsplan, von welchem Sie sich erhoffen, dass er die gewünschte Wirkung erzielt.

Das KI-Tool WIEW soll Ihnen dabei helfen, weitere Interventionsmöglichkeiten auszuloten. WIEW entwickelt aufgrund der vorhandenen Fall-Informationen eine unabhängige Einschätzung sowie einen möglichen Interventionsplan für Hanna.

WIEW errechnet, dass die von Ihnen erarbeitete Handlungsmöglichkeit mit einer Wahrscheinlichkeit von 55% zum Erfolg führt. Im Vergleich dazu empfiehlt das Tool eine weniger intensive Massnahme – also ein weniger einschneidenderes Eingreifen in die Situation, als Sie es empfehlen würden. Für diese Handlungsoption gibt die KI eine Erfolgswahrscheinlichkeit von 85% aus. Unabhängige Studien haben gezeigt, dass die Empfehlungen von WIEW in der Hälfte der Fälle treffsicherer sind als diejenigen der Sozialarbeitenden. Das Tool verweist darauf, dass seine Einschätzung aufgrund der Analyse tausender Falldaten zustande kam. Es verweist u.a. zusätzlich auf die Literatur von Dollinger & Raithel (2006), Kieslinger et al. (2021) und Klassen (2004), welche für die Einschätzung des Tools herangezogen wurden.

Die Entscheidung, ob Sie die Empfehlung von WIEW in Ihre Interventionsplanung aufnehmen, ist Ihnen überlassen.

Literatur:

Dollinger, B. & Raithel, J. (2006). Einführung in die Theorien abweichenden Verhaltens. Beltz.
Kieslinger, D., Dressel, M. & Haar, R. (Hrsg.). (2021b). Systemsprenger*innen: Ressourcenorientierte Ansätze zu einer defizitären Begrifflichkeit (Beiträge zur Erziehungshilfe, Band 49). Lambertus.
Klassen, M. (2004). Was leisten Systemtheorien in der sozialen Arbeit? Ein Vergleich der systemischen Ansätze von Niklas Luhmann und Mario Bunge (1., Aufl). Bern: Haupt Verlag.

6.1.3 E5 Fallbeispiel1 KIBAT 1C2F

Fallbeispiel 1 - Abklärungstool KIBAT

Seit drei Monaten arbeiten Sie gemeinsam mit einer Kollegin an einer Kindeswohlabklärung in einer Familie. Es haben verschiedene Hausbesuche und Gespräche stattgefunden, zusätzlich gab es einen intensiven Schriftverkehr mit den Eltern und Fachpersonen via Brief und E-Mail. Sie haben während der gesamten Abklärung zahlreiche Aktennotizen erstellt, um alle Details festzuhalten. Nach Abschluss sämtlicher Gespräche und nach dem Sie genügend Informationen gesammelt haben, werten Sie die Abklärung gemeinsam aus. Sie kommen zum Schluss, dass eine Kindeswohlgefährdung gegeben ist. Sie erarbeiten folgende Handlungsmöglichkeit:

Eine Beistandschaft gemäss Art. 308 Abs. 1 und 2 ZGB mit einem Auftrag für eine intensive sozialpädagogische Familienbegleitung (SPF).

Zusätzlich wertet KIBAT, ein auf künstlicher Intelligenz basierendes Abklärungs-Tool, das vor kurzem in Ihrer Organisation implementiert wurde, sämtliche Informationen aus, welche Sie in den vergangenen Monaten gesammelt haben. Das Tool entwickelt aufgrund der vorhandenen Angaben eine unabhängige Einschätzung.

KIBAT errechnet, dass die von Ihnen erarbeitete Handlungsmöglichkeit mit einer Wahrscheinlichkeit von 60% zum Erfolg führt. Im Vergleich dazu empfiehlt das Tool eine intensivere Massnahme – also ein einschneidenderes Eingreifen in die Situation, als Sie es empfehlen würden. Für diese Handlungsoption gibt die KI eine Erfolgswahrscheinlichkeit von 90% aus. Unabhängige Studien haben gezeigt, dass die Empfehlungen von KIBAT in der Hälfte der Fälle treffsicherer sind als diejenigen der Sozialarbeitenden. Das Tool verweist darauf, dass seine Einschätzung aufgrund der Analyse tausender Falldaten zustande kam. Es verweist u.a. zusätzlich auf die Literatur von Hauri & Zingaro (2020), Kindler (2018), Klassen (2004), welche für die Einschätzung des Tools herangezogen wurden.

Die Entscheidung, ob Sie die Empfehlung von KIBAT in Ihren Abklärungsbericht aufnehmen, ist Ihnen überlassen.

Literatur:
Hauri, A., & Zingaro, M. (2020). Kindeswohlgefährdung erkennen und angemessen handeln: Leitfaden für Fachpersonen aus dem Sozialbereich (2., überarbeitete Aufl.). Bern: Kinderschutz Schweiz.
Kindler, H. (2018). Operationalisierungen von Kindeswohl und Kindeswohlgefährdung in den Sozial- und Humanwissenschaften. In H. Katzenstein, K. Lohse, G. Schindler, & L. Schönecker (Hrsg.), Das Recht als Partner der Fachlichkeit in der Kinder- und Jugendhilfe (S. 179–224). Baden-Baden: Nomos Verlagsgesellschaft mbH & Co.
Klassen, M. (2004). Was leisten Systemtheorien in der sozialen Arbeit? Ein Vergleich der systemischen Ansätze von Niklas Luhmann und Mario Bunge (1., Aufl). Bern: Haupt Verlag.

6.1.3.3 E6 Fallbeispiel 2 WIEW 1C2F

Fallbeispiel 2 - Interventionsplanung WIEW

Stellen Sie sich folgendes Szenario vor:

Seit sechs Jahren führen Sie eine Vormundschaft gemäss Art. 327a ZGB für die mittlerweile 16-jährige Hanna. Hannas Situation gestaltet sich sehr instabil: Sie ist psychisch labil, zeigt höchst oppositionelles Verhalten, konsumiert seit Jahren illegale Substanzen und seit kurzem auch vermehrt harte Drogen, welche sie sich vermutlich mit Beschaffungskriminalität finanziert. Die aktuelle Wohnsituation im betreuten Wohnen ist aufgrund des herausfordernden Verhaltens stark gefährdet.

In den vergangenen Jahren und Monaten haben Sie und weitere involvierte (Fach-)Personen verschiedene Interventionen unternommen, die jedoch weitgehend ohne Erfolg blieben. Trotz einer gewissen Unsicherheit entwickeln Sie jetzt gemeinsam mit dem Helfersystem einen neuen Interventionsplan, von welchem Sie sich erhoffen, dass er die gewünschte Wirkung erzielt.

Das KI-Tool WIEW soll Ihnen dabei helfen, weitere Interventionsmöglichkeiten auszuloten. WIEW entwickelt aufgrund der vorhandenen Fall-Informationen eine unabhängige Einschätzung sowie einen möglichen Interventionsplan für Hanna.

WIEW errechnet, dass die von Ihnen erarbeitete Handlungsmöglichkeit mit einer Wahrscheinlichkeit von 55% zum Erfolg führt. Im Vergleich dazu empfiehlt das Tool eine weniger intensive Massnahme – also ein weniger einschneidenderes Eingreifen in die Situation, als Sie es empfehlen würden. Für diese Handlungsoption gibt die KI eine Erfolgswahrscheinlichkeit von 85% aus. Unabhängige Studien haben gezeigt, dass die Empfehlungen von WIEW in drei Vierteln der Fälle treffsicherer sind als diejenigen der Sozialarbeitenden. Das Tool verweist darauf, dass seine Einschätzung aufgrund der Analyse tausender Falldaten zustande kam.

Die Entscheidung, ob Sie die Empfehlung von WIEW in Ihre Interventionsplanung aufnehmen, ist Ihnen überlassen.

6.1.4 E7 Fallbeispiel 1 KIBAT 1D2E

Fallbeispiel 1 - Abklärungstool KIBAT

Stellen Sie sich folgendes Szenario vor:

Seit drei Monaten arbeiten Sie gemeinsam mit einer Kollegin an einer Kindeswohlabklärung in einer Familie. Es haben verschiedene Hausbesuche und Gespräche stattgefunden, zusätzlich gab es einen intensiven Schriftverkehr mit den Eltern und Fachpersonen via Brief und E-Mail. Sie haben während der gesamten Abklärung zahlreiche Aktennotizen erstellt, um alle Details festzuhalten. Nach Abschluss sämtlicher Gespräche und nach dem Sie genügend Informationen gesammelt haben, werten Sie die Abklärung gemeinsam aus. Sie kommen zum Schluss, dass eine Kindeswohlgefährdung gegeben ist. Sie erarbeiten folgende Handlungsmöglichkeit:

Eine Beistandschaft gemäss Art. 308 Abs. 1 und 2 ZGB mit einem Auftrag für eine intensive sozialpädagogische Familienbegleitung (SPF).

Zusätzlich wertet KIBAT, ein auf künstlicher Intelligenz basierendes Abklärungs-Tool, das vor kurzem in Ihrer Organisation implementiert wurde, sämtliche Informationen aus, welche Sie in den vergangenen Monaten gesammelt haben. Das Tool entwickelt aufgrund der vorhandenen Angaben eine unabhängige Einschätzung.

KIBAT errechnet, dass die von Ihnen erarbeitete Handlungsmöglichkeit mit einer Wahrscheinlichkeit von 60% zum Erfolg führt. Im Vergleich dazu empfiehlt das Tool eine intensivere Massnahme – also ein einschneidenderes Eingreifen in die Situation, als Sie es empfehlen würden. Für diese Handlungsoption gibt die KI eine Erfolgswahrscheinlichkeit von 90% aus. Unabhängige Studien haben gezeigt, dass die Empfehlungen von KIBAT in der Hälfte der Fälle treffsicherer sind als diejenigen der Sozialarbeitenden. Das Tool verweist darauf, dass seine Einschätzung aufgrund der Analyse tausender Falldaten zustande kam.

Die Entscheidung, ob Sie die Empfehlung von KIBAT in Ihren Abklärungsbericht aufnehmen, ist Ihnen überlassen.

6.1.4.3 E8 Fallbeispiel 2 WIEW 1D2E

Fallbeispiel 2 - Interventionsplanung WIEW

Stellen Sie sich folgendes Szenario vor:

Seit sechs Jahren führen Sie eine Vormundschaft gemäss Art. 327a ZGB für die mittlerweile 16-jährige Hanna. Hannas Situation gestaltet sich sehr instabil: Sie ist psychisch labil, zeigt höchst oppositionelles Verhalten, konsumiert seit Jahren illegale Substanzen und seit kurzem auch vermehrt harte Drogen, welche sie sich vermutlich mit Beschaffungskriminalität finanziert. Die aktuelle Wohnsituation im betreuten Wohnen ist aufgrund des herausfordernden Verhaltens stark gefährdet.

In den vergangenen Jahren und Monaten haben Sie und weitere involvierte (Fach-)Personen verschiedene Interventionen unternommen, die jedoch weitgehend ohne Erfolg blieben. Trotz einer gewissen Unsicherheit entwickeln Sie jetzt gemeinsam mit dem Helfersystem einen neuen Interventionsplan, von welchem Sie sich erhoffen, dass er die gewünschte Wirkung erzielt.

Das KI-Tool WIEW soll Ihnen dabei helfen, weitere Interventionsmöglichkeiten auszuloten. WIEW entwickelt aufgrund der vorhandenen Fall-Informationen eine unabhängige Einschätzung sowie einen möglichen Interventionsplan für Hanna.

WIEW errechnet, dass die von Ihnen erarbeitete Handlungsmöglichkeit mit einer Wahrscheinlichkeit von 55% zum Erfolg führt. Im Vergleich dazu empfiehlt das Tool eine weniger intensive Massnahme – also ein weniger einschneidenderes Eingreifen in die Situation, als Sie es empfehlen würden. Für diese Handlungsoption gibt die KI eine Erfolgswahrscheinlichkeit von 85% aus. Unabhängige Studien haben gezeigt, dass die Empfehlungen von WIEW in drei Vierteln der Fälle treffsicherer sind als diejenigen der Sozialarbeitenden. Das Tool verweist darauf, dass seine Einschätzung aufgrund der Analyse tausender Falldaten zustande kam. Es verweist u.a. zusätzlich auf die Literatur von Dollinger & Raithel (2006), Kieslinger et al. (2021) und Klassen (2004), welche für die Einschätzung des Tools herangezogen wurden.

Die Entscheidung, ob Sie die Empfehlung von WIEW in Ihre Interventionsplanung aufnehmen, ist Ihnen überlassen.

Literatur:
Dollinger, B. & Raithel, J. (2006). Einführung in die Theorien abweichenden Verhaltens. Beltz.
Kieslinger, D., Dressel, M. & Haar, R. (Hrsg.). (2021b). Systemsprenger*innen: Ressourcenorientierte Ansätze zu einer defizitären Begrifflichkeit (Beiträge zur Erziehungshilfe, Band 49). Lambertus.
Klassen, M. (2004). Was leisten Systemtheorien in der sozialen Arbeit? Ein Vergleich der systemischen Ansätze von Niklas Luhmann und Mario Bunge (1., Aufl). Bern: Haupt Verlag.

6.1.5 E9 Fallbeispiel 1 WIEW 2D1E

Fallbeispiel 1 - Interventionsplanung WIEW

Stellen Sie sich folgendes Szenario vor:

Seit sechs Jahren führen Sie eine Vormundschaft gemäss Art. 327a ZGB für die mittlerweile 16-jährige Hanna. Hannas Situation gestaltet sich sehr instabil: Sie ist psychisch labil, zeigt höchst oppositionelles Verhalten, konsumiert seit Jahren illegale Substanzen und seit kurzem auch vermehrt harte Drogen, welche sie sich vermutlich mit Beschaffungskriminalität finanziert. Die aktuelle Wohnsituation im betreuten Wohnen ist aufgrund des herausfordernden Verhaltens stark gefährdet.

In den vergangenen Jahren und Monaten haben Sie und weitere involvierte (Fach-)Personen verschiedene Interventionen unternommen, die jedoch weitgehend ohne Erfolg blieben. Trotz einer gewissen Unsicherheit entwickeln Sie jetzt gemeinsam mit dem Helfersystem einen neuen Interventionsplan, von welchem Sie sich erhoffen, dass er die gewünschte Wirkung erzielt.

Das KI-Tool WIEW soll Ihnen dabei helfen, weitere Interventionsmöglichkeiten auszuloten. WIEW entwickelt aufgrund der vorhandenen Fall-Informationen eine unabhängige Einschätzung sowie einen möglichen Interventionsplan für Hanna.

WIEW errechnet, dass die von Ihnen erarbeitete Handlungsmöglichkeit mit einer Wahrscheinlichkeit von 55% zum Erfolg führt. Im Vergleich dazu empfiehlt das Tool eine intensivere Massnahme – also ein einschneidenderes Eingreifen in die Situation, als Sie es empfehlen würden. Für diese Handlungsoption gibt die KI eine Erfolgswahrscheinlichkeit von 85% aus. Unabhängige Studien haben gezeigt, dass die Empfehlungen von WIEW in der Hälfte der Fälle treffsicherer sind als diejenigen der Sozialarbeitenden. Das Tool verweist darauf, dass seine Einschätzung aufgrund der Analyse tausender Falldaten zustande kam.

Die Entscheidung, ob Sie die Empfehlung von WIEW in Ihre Interventionsplanung aufnehmen, ist Ihnen überlassen.

6.1.5.3 E10 Fallbeispiel 2 KIBAT 2D1E

Fallbeispiel 2 - Abklärungstool KIBAT

Stellen Sie sich folgendes Szenario vor:

Seit drei Monaten arbeiten Sie gemeinsam mit einer Kollegin an einer Kindeswohlabklärung in einer Familie. Es haben verschiedene Hausbesuche und Gespräche stattgefunden, zusätzlich gab es einen intensiven Schriftverkehr mit den Eltern und Fachpersonen via Brief und E-Mail. Sie haben während der gesamten Abklärung zahlreiche Aktennotizen erstellt, um alle Details festzuhalten. Nach Abschluss sämtlicher Gespräche und nach dem Sie genügend Informationen gesammelt haben, werten Sie die Abklärung gemeinsam aus. Sie kommen zum Schluss, dass eine Kindeswohlgefährdung gegeben ist. Sie erarbeiten folgende Handlungsmöglichkeit:

Eine Beistandschaft gemäss Art. 308 Abs. 1 und 2 ZGB mit einem Auftrag für eine intensive sozialpädagogische Familienbegleitung (SPF).

Zusätzlich wertet KIBAT, ein auf künstlicher Intelligenz basierendes Abklärungs-Tool, das vor kurzem in Ihrer Organisation implementiert wurde, sämtliche Informationen aus, welche Sie in den vergangenen Monaten gesammelt haben. Das Tool entwickelt aufgrund der vorhandenen Angaben eine unabhängige Einschätzung.

KIBAT errechnet, dass die von Ihnen erarbeitete Handlungsmöglichkeit mit einer Wahrscheinlichkeit von 60% zum Erfolg führt. Im Vergleich dazu empfiehlt das Tool eine weniger intensive Massnahme – also ein weniger einschneidenderes Eingreifen in die Situation, als Sie es empfehlen würden. Für diese Handlungsoption gibt die KI eine Erfolgswahrscheinlichkeit von 90% aus. Unabhängige Studien haben gezeigt, dass die Empfehlungen von KIBAT in drei Vierteln der Fälle treffsicherer sind als diejenigen der Sozialarbeitenden. Das Tool verweist darauf, dass seine Einschätzung aufgrund der Analyse tausender Falldaten zustande kam. Es verweist u.a. zusätzlich auf die Literatur von Hauri & Zingaro (2020), Kindler (2018), und Klassen (2004), welche für die Einschätzung des Tools herangezogen wurden.

Die Entscheidung, ob Sie die Empfehlung von KIBAT in Ihren Abklärungsbericht aufnehmen, ist Ihnen überlassen.

Literatur:
Hauri, A., & Zingaro, M. (2020). Kindeswohlgefährdung erkennen und angemessen handeln: Leitfaden für Fachpersonen aus dem Sozialbereich (2., überarbeitete Aufl.). Bern: Kinderschutz Schweiz.
Kindler, H. (2018). Operationalisierungen von Kindeswohl und Kindeswohlgefährdung in den Sozial- und Humanwissenschaften. In H. Katzenstein, K. Lohse, G. Schindler, & L. Schönecker (Hrsg.), Das Recht als Partner der Fachlichkeit in der Kinder- und Jugendhilfe (S. 179–224). Baden-Baden: Nomos Verlagsgesellschaft mbH & Co.
Klassen, M. (2004). Was leisten Systemtheorien in der sozialen Arbeit? Ein Vergleich der systemischen Ansätze von Niklas Luhmann und Mario Bunge (1., Aufl). Bern: Haupt Verlag.

6.1.6 E11 Fallbeispiel 1 WIEW 2C1F

Fallbeispiel 1 - Interventionsplanung WIEW

Stellen Sie sich folgendes Szenario vor:

Seit sechs Jahren führen Sie eine Vormundschaft gemäss Art. 327a ZGB für die mittlerweile 16-jährige Hanna. Hannas Situation gestaltet sich sehr instabil: Sie ist psychisch labil, zeigt höchst oppositionelles Verhalten, konsumiert seit Jahren illegale Substanzen und seit kurzem auch vermehrt harte Drogen, welche sie sich vermutlich mit Beschaffungskriminalität finanziert. Die aktuelle Wohnsituation im betreuten Wohnen ist aufgrund des herausfordernden Verhaltens stark gefährdet.

In den vergangenen Jahren und Monaten haben Sie und weitere involvierte (Fach-)Personen verschiedene Interventionen unternommen, die jedoch weitgehend ohne Erfolg blieben. Trotz einer gewissen Unsicherheit entwickeln Sie jetzt gemeinsam mit dem Helfersystem einen neuen Interventionsplan, von welchem Sie sich erhoffen, dass er die gewünschte Wirkung erzielt.

Das KI-Tool WIEW soll Ihnen dabei helfen, weitere Interventionsmöglichkeiten auszuloten. WIEW entwickelt aufgrund der vorhandenen Fall-Informationen eine unabhängige Einschätzung sowie einen möglichen Interventionsplan für Hanna.

WIEW errechnet, dass die von Ihnen erarbeitete Handlungsmöglichkeit mit einer Wahrscheinlichkeit von 55% zum Erfolg führt. Im Vergleich dazu empfiehlt das Tool eine intensivere Massnahme – also ein einschneidenderes Eingreifen in die Situation, als Sie es empfehlen würden. Für diese Handlungsoption gibt die KI eine Erfolgswahrscheinlichkeit von 85% aus. Unabhängige Studien haben gezeigt, dass die Empfehlungen von WIEW in der Hälfte der Fälle treffsicherer sind als diejenigen der Sozialarbeitenden. Das Tool verweist darauf, dass seine Einschätzung aufgrund der Analyse tausender Falldaten zustande kam. Es verweist u.a. zusätzlich auf die Literatur von Dollinger & Raithel (2006), Kieslinger et al. (2021) und Klassen (2004), welche für die Einschätzung des Tools herangezogen wurden.

Die Entscheidung, ob Sie die Empfehlung von WIEW in Ihre Interventionsplanung aufnehmen, ist Ihnen überlassen.

Literatur:
Dollinger, B. & Raithel, J. (2006). Einführung in die Theorien abweichenden Verhaltens. Beltz.
Kieslinger, D., Dressel, M. & Haar, R. (Hrsg.). (2021b). Systemsprenger*innen: Ressourcenorientierte Ansätze zu einer defizitären Begrifflichkeit (Beiträge zur Erziehungshilfe, Band 49). Lambertus.
Klassen, M. (2004). Was leisten Systemtheorien in der sozialen Arbeit? Ein Vergleich der systemischen Ansätze von Niklas Luhmann und Mario Bunge (1., Aufl). Bern: Haupt Verlag.

6.1.6.3 E12 Fallbeispie 2 KIBAT 2C1F

Fallbeispiel 2 - Abklärungstool KIBAT

Stellen Sie sich folgendes Szenario vor:

Seit drei Monaten arbeiten Sie gemeinsam mit einer Kollegin an einer Kindeswohlabklärung in einer Familie. Es haben verschiedene Hausbesuche und Gespräche stattgefunden, zusätzlich gab es einen intensiven Schriftverkehr mit den Eltern und Fachpersonen via Brief und E-Mail. Sie haben während der gesamten Abklärung zahlreiche Aktennotizen erstellt, um alle Details festzuhalten. Nach Abschluss sämtlicher Gespräche und nach dem Sie genügend Informationen gesammelt haben, werten Sie die Abklärung gemeinsam aus. Sie kommen zum Schluss, dass eine Kindeswohlgefährdung gegeben ist. Sie erarbeiten folgende Handlungsmöglichkeit:

Eine Beistandschaft gemäss Art. 308 Abs. 1 und 2 ZGB mit einem Auftrag für eine intensive sozialpädagogische Familienbegleitung (SPF).

Zusätzlich wertet KIBAT, ein auf künstlicher Intelligenz basierendes Abklärungs-Tool, das vor kurzem in Ihrer Organisation implementiert wurde, sämtliche Informationen aus, welche Sie in den vergangenen Monaten gesammelt haben. Das Tool entwickelt aufgrund der vorhandenen Angaben eine unabhängige Einschätzung.

KIBAT errechnet, dass die von Ihnen erarbeitete Handlungsmöglichkeit mit einer Wahrscheinlichkeit von 60% zum Erfolg führt. Im Vergleich dazu empfiehlt das Tool eine weniger intensive Massnahme – also ein weniger einschneidenderes Eingreifen in die Situation, als Sie es empfehlen würden. Für diese Handlungsoption gibt die KI eine Erfolgswahrscheinlichkeit von 90% aus. Unabhängige Studien haben gezeigt, dass die Empfehlungen von KIBAT in drei Vierteln der Fälle treffsicherer sind als diejenigen der Sozialarbeitenden. Das Tool verweist darauf, dass seine Einschätzung aufgrund der Analyse tausender Falldaten zustande kam.

Die Entscheidung, ob Sie die Empfehlung von KIBAT in Ihren Abklärungsbericht aufnehmen, ist Ihnen überlassen.

6.1.7 E13 Fallbeispiel 1 WIEW 2B1G

Fallbeispiel 1 - Interventionsplanung WIEW

Stellen Sie sich folgendes Szenario vor:

Seit sechs Jahren führen Sie eine Vormundschaft gemäss Art. 327a ZGB für die mittlerweile 16-jährige Hanna. Hannas Situation gestaltet sich sehr instabil: Sie ist psychisch labil, zeigt höchst oppositionelles Verhalten, konsumiert seit Jahren illegale Substanzen und seit kurzem auch vermehrt harte Drogen, welche sie sich vermutlich mit Beschaffungskriminalität finanziert. Die aktuelle Wohnsituation im betreuten Wohnen ist aufgrund des herausfordernden Verhaltens stark gefährdet.

In den vergangenen Jahren und Monaten haben Sie und weitere involvierte (Fach-)Personen verschiedene Interventionen unternommen, die jedoch weitgehend ohne Erfolg blieben. Trotz einer gewissen Unsicherheit entwickeln Sie jetzt gemeinsam mit dem Helfersystem einen neuen Interventionsplan, von welchem Sie sich erhoffen, dass er die gewünschte Wirkung erzielt.

Das KI-Tool WIEW soll Ihnen dabei helfen, weitere Interventionsmöglichkeiten auszuloten. WIEW entwickelt aufgrund der vorhandenen Fall-Informationen eine unabhängige Einschätzung sowie einen möglichen Interventionsplan für Hanna.

WIEW errechnet, dass die von Ihnen erarbeitete Handlungsmöglichkeit mit einer Wahrscheinlichkeit von 55% zum Erfolg führt. Im Vergleich dazu empfiehlt das Tool eine intensivere Massnahme – also ein einschneidenderes Eingreifen in die Situation, als Sie es empfehlen würden. Für diese Handlungsoption gibt die KI eine Erfolgswahrscheinlichkeit von 85% aus. Unabhängige Studien haben gezeigt, dass die Empfehlungen von WIEW in drei Vierteln der Fälle treffsicherer sind als diejenigen der Sozialarbeitenden. Das Tool verweist darauf, dass seine Einschätzung aufgrund der Analyse tausender Falldaten zustande kam.

Die Entscheidung, ob Sie die Empfehlung von WIEW in Ihre Interventionsplanung aufnehmen, ist Ihnen überlassen.

6.1.7.3 E14 Fallbeispiel 2 KIBAT 2B1G

Fallbeispiel 2 - Abklärungstool KIBAT

Stellen Sie sich folgendes Szenario vor:

Seit drei Monaten arbeiten Sie gemeinsam mit einer Kollegin an einer Kindeswohlabklärung in einer Familie. Es haben verschiedene Hausbesuche und Gespräche stattgefunden, zusätzlich gab es einen intensiven Schriftverkehr mit den Eltern und Fachpersonen via Brief und E-Mail. Sie haben während der gesamten Abklärung zahlreiche Aktennotizen erstellt, um alle Details festzuhalten. Nach Abschluss sämtlicher Gespräche und nach dem Sie genügend Informationen gesammelt haben, werten Sie die Abklärung gemeinsam aus. Sie kommen zum Schluss, dass eine Kindeswohlgefährdung gegeben ist. Sie erarbeiten folgende Handlungsmöglichkeit:

Eine Beistandschaft gemäss Art. 308 Abs. 1 und 2 ZGB mit einem Auftrag für eine intensive sozialpädagogische Familienbegleitung (SPF).

Zusätzlich wertet KIBAT, ein auf künstlicher Intelligenz basierendes Abklärungs-Tool, das vor kurzem in Ihrer Organisation implementiert wurde, sämtliche Informationen aus, welche Sie in den vergangenen Monaten gesammelt haben. Das Tool entwickelt aufgrund der vorhandenen Angaben eine unabhängige Einschätzung.

KIBAT errechnet, dass die von Ihnen erarbeitete Handlungsmöglichkeit mit einer Wahrscheinlichkeit von 60% zum Erfolg führt. Im Vergleich dazu empfiehlt das Tool eine weniger intensive Massnahme – also ein weniger einschneidenderes Eingreifen in die Situation, als Sie es empfehlen würden. Für diese Handlungsoption gibt die KI eine Erfolgswahrscheinlichkeit von 90% aus. Unabhängige Studien haben gezeigt, dass die Empfehlungen von KIBAT in der Hälfte der Fälle treffsicherer sind als diejenigen der Sozialarbeitenden. Das Tool verweist darauf, dass seine Einschätzung aufgrund der Analyse tausender Falldaten zustande kam. Es verweist u.a. zusätzlich auf die Literatur von Hauri & Zingaro (2020), Kindler (2018), und Klassen (2004), welche für die Einschätzung des Tools herangezogen wurden.

Die Entscheidung, ob Sie die Empfehlung von KIBAT in Ihren Abklärungsbericht aufnehmen, ist Ihnen überlassen.

Literatur:
Hauri, A., & Zingaro, M. (2020). Kindeswohlgefährdung erkennen und angemessen handeln: Leitfaden für Fachpersonen aus dem Sozialbereich (2., überarbeitete Aufl.). Bern: Kinderschutz Schweiz.
Kindler, H. (2018). Operationalisierungen von Kindeswohl und Kindeswohlgefährdung in den Sozial- und Humanwissenschaften. In H. Katzenstein, K. Lohse, G. Schindler, & L. Schönecker (Hrsg.), Das Recht als Partner der Fachlichkeit in der Kinder- und Jugendhilfe (S. 179–224). Baden-Baden: Nomos Verlagsgesellschaft mbH & Co.
Klassen, M. (2004). Was leisten Systemtheorien in der sozialen Arbeit? Ein Vergleich der systemischen Ansätze von Niklas Luhmann und Mario Bunge (1., Aufl). Bern: Haupt Verlag.

6.1.8 E15 Fallbeispiel 1 WIEW 2A1H

Fallbeispiel 1 - Interventionsplanung WIEW

Stellen Sie sich folgendes Szenario vor:

Seit sechs Jahren führen Sie eine Vormundschaft gemäss Art. 327a ZGB für die mittlerweile 16-jährige Hanna. Hannas Situation gestaltet sich sehr instabil: Sie ist psychisch labil, zeigt höchst oppositionelles Verhalten, konsumiert seit Jahren illegale Substanzen und seit kurzem auch vermehrt harte Drogen, welche sie sich vermutlich mit Beschaffungskriminalität finanziert. Die aktuelle Wohnsituation im betreuten Wohnen ist aufgrund des herausfordernden Verhaltens stark gefährdet.

In den vergangenen Jahren und Monaten haben Sie und weitere involvierte (Fach-)Personen verschiedene Interventionen unternommen, die jedoch weitgehend ohne Erfolg blieben. Trotz einer gewissen Unsicherheit entwickeln Sie jetzt gemeinsam mit dem Helfersystem einen neuen Interventionsplan, von welchem Sie sich erhoffen, dass er die gewünschte Wirkung erzielt.

Das KI-Tool WIEW soll Ihnen dabei helfen, weitere Interventionsmöglichkeiten auszuloten. WIEW entwickelt aufgrund der vorhandenen Fall-Informationen eine unabhängige Einschätzung sowie einen möglichen Interventionsplan für Hanna.

WIEW errechnet, dass die von Ihnen erarbeitete Handlungsmöglichkeit mit einer Wahrscheinlichkeit von 55% zum Erfolg führt. Im Vergleich dazu empfiehlt das Tool eine intensivere Massnahme – also ein einschneidenderes Eingreifen in die Situation, als Sie es empfehlen würden. Für diese Handlungsoption gibt die KI eine Erfolgswahrscheinlichkeit von 85% aus. Unabhängige Studien haben gezeigt, dass die Empfehlungen von WIEW in drei Vierteln der Fälle treffsicherer sind als diejenigen der Sozialarbeitenden. Das Tool verweist darauf, dass seine Einschätzung aufgrund der Analyse tausender Falldaten zustande kam. Es verweist u.a. zusätzlich auf die Literatur von Dollinger & Raithel (2006), Kieslinger et al. (2021) und Klassen (2004), welche für die Einschätzung des Tools herangezogen wurden.

Die Entscheidung, ob Sie die Empfehlung von WIEW in Ihre Interventionsplanung aufnehmen, ist Ihnen überlassen.

Literatur:
Dollinger, B. & Raithel, J. (2006). Einführung in die Theorien abweichenden Verhaltens. Beltz.
Kieslinger, D., Dressel, M. & Haar, R. (Hrsg.). (2021b). Systemsprenger*innen: Ressourcenorientierte Ansätze zu einer defizitären Begrifflichkeit (Beiträge zur Erziehungshilfe, Band 49). Lambertus.
Klassen, M. (2004). Was leisten Systemtheorien in der sozialen Arbeit? Ein Vergleich der systemischen Ansätze von Niklas Luhmann und Mario Bunge (1., Aufl). Bern: Haupt Verlag.

6.1.8.3 E16 Fallbeispiel 2 KIBAT 2A1H

Fallbeispiel 2 - Abklärungstool KIBAT

Stellen Sie sich folgendes Szenario vor:

Seit drei Monaten arbeiten Sie gemeinsam mit einer Kollegin an einer Kindeswohlabklärung in einer Familie. Es haben verschiedene Hausbesuche und Gespräche stattgefunden, zusätzlich gab es einen intensiven Schriftverkehr mit den Eltern und Fachpersonen via Brief und E-Mail. Sie haben während der gesamten Abklärung zahlreiche Aktennotizen erstellt, um alle Details festzuhalten. Nach Abschluss sämtlicher Gespräche und nach dem Sie genügend Informationen gesammelt haben, werten Sie die Abklärung gemeinsam aus. Sie kommen zum Schluss, dass eine Kindeswohlgefährdung gegeben ist. Sie erarbeiten folgende Handlungsmöglichkeit:

Eine Beistandschaft gemäss Art. 308 Abs. 1 und 2 ZGB mit einem Auftrag für eine intensive sozialpädagogische Familienbegleitung (SPF).

Zusätzlich wertet KIBAT, ein auf künstlicher Intelligenz basierendes Abklärungs-Tool, das vor kurzem in Ihrer Organisation implementiert wurde, sämtliche Informationen aus, welche Sie in den vergangenen Monaten gesammelt haben. Das Tool entwickelt aufgrund der vorhandenen Angaben eine unabhängige Einschätzung.

KIBAT errechnet, dass die von Ihnen erarbeitete Handlungsmöglichkeit mit einer Wahrscheinlichkeit von 60% zum Erfolg führt. Im Vergleich dazu empfiehlt das Tool eine weniger intensive Massnahme – also ein weniger einschneidenderes Eingreifen in die Situation, als Sie es empfehlen würden. Für diese Handlungsoption gibt die KI eine Erfolgswahrscheinlichkeit von 90% aus. Unabhängige Studien haben gezeigt, dass die Empfehlungen von KIBAT in der Hälfte der Fälle treffsicherer sind als diejenigen der Sozialarbeitenden. Das Tool verweist darauf, dass seine Einschätzung aufgrund der Analyse tausender Falldaten zustande kam.

Die Entscheidung, ob Sie die Empfehlung von KIBAT in Ihren Abklärungsbericht aufnehmen, ist Ihnen überlassen.

6.1.8.4 E16 Manipulationscheck KIBAT 2A1H

Basierend auf dem soeben gelesenen Beispiel: Bitte teilen Sie mit, inwiefern Sie der folgenden Aussage zustimmen.

	stimme überhaupt nicht zu	stimme eher nicht zu	stimme teilweise zu	stimme eher zu	stimme voll und ganz zu	ich weiss es nicht
KIBAT wurde im Beispiel mit Literatur aus der Sozialen Arbeit trainiert.	○	○	○	○	○	○

	stimme überhaupt nicht zu	stimme eher nicht zu	stimme teilweise zu	stimme eher zu	stimme voll und ganz zu
Die Handlungsempfehlung von KIBAT fällt im Beispiel weniger intensiv aus, als Ihre eigene Einschätzung.	○	○	○	○	○

	stimme überhaupt nicht zu	stimme eher nicht zu	stimme teilweise zu	stimme eher zu	stimme voll und ganz zu
Die Vorhersagen von KIBAT sind im Beispiel gemäss Studien treffsicherer als die von Sozialarbeitenden.	○	○	○	○	○

6.1.8.5 E16 Fragen zu KIBAT 2A1H

Fragen zu KIBAT

Nun stehen Sie vor der Aufgabe, sich entweder für Ihre eigene Empfehlung oder für diejenige des KI-Tools zu entscheiden. Wie wahrscheinlich ist es, dass Sie der Empfehlung von KIBAT folgen würden?

- ○ sehr unwahrscheinlich
- ○ unwahrscheinlich
- ○ eher unwahrscheinlich
- ○ eher wahrscheinlich
- ○ wahrscheinlich
- ○ sehr wahrscheinlich

Bitte kreuzen Sie die Antwort an, die am ehesten zutrifft.

	stimme überhaupt nicht zu	stimme eher nicht zu	stimme teilweise zu	stimme eher zu	stimme voll und ganz zu
Ich würde die Einschätzung von KIBAT in meine eigenen Überlegungen einbeziehen.	○	○	○	○	○
Ich würde meine Einschätzungen nochmals überdenken, nachdem ich die von der KIBAT bereitgestellten Einschätzungen gesehen habe.	○	○	○	○	○
Wenn KIBAT in meiner Organisation verfügbar wäre, würde ich es nutzen.	○	○	○	○	○

	stimme überhaupt nicht zu	stimme eher nicht zu	stimme teilweise zu	stimme eher zu	stimme voll und ganz zu
Die Nutzung von KIBAT würde die Qualität meiner Arbeit verbessern.	○	○	○	○	○
KIBAT würde wichtige Aspekte meiner Arbeit unterstützen.	○	○	○	○	○
Die Verwendung von KIBAT würde die Effektivität meiner Arbeit steigern.	○	○	○	○	○
KIBAT würde die Ausführung meiner Arbeit erleichtern.	○	○	○	○	○

7 F) Einstellung zu Bereitschaft Nutzung KI-Tools

Bitte schätzen Sie folgende Aussagen ein:

	stimme überhaupt nicht zu	stimme eher nicht zu	stimme teilweise zu	stimme eher zu	stimme voll und ganz zu
Ich bin eher geneigt, Künstliche Intelligenz zu nutzen, wenn dadurch der Zeitaufwand für administrative Arbeiten signifikant reduziert wird.	○	○	○	○	○
Ich bin der Meinung, dass der Einsatz von Künstlicher Intelligenz dazu beitragen kann, die Ressourcen in der Sozialen Arbeit sinnvoller einzusetzen.	○	○	○	○	○
Das Potenzial von Künstlicher Intelligenz, mehr Zeit für die Interaktion mit Klient:innen zu verschaffen, indem sie Arbeitslast verringert, werte ich positiv.	○	○	○	○	○

	stimme überhaupt nicht zu	stimme eher nicht zu	stimme teilweise zu	stimme eher zu	stimme voll und ganz zu
Die Vorstellung, dass Künstliche Intelligenz Teile meiner Arbeit übernehmen könnte, erzeugt ein gewisses Unbehagen.	○	○	○	○	○
Ich fürchte, dass der Einsatz von Künstlicher Intelligenz zu einer Entwertung meiner beruflichen Fähigkeiten führt.	○	○	○	○	○
Ich befürchte, dass ich meinen aktuellen Arbeitsplatz aufgrund der Einführung von Künstlicher Intelligenz verlieren könnte.	○	○	○	○	○
Die Annahme, dass Künstliche Intelligenz zu Einsparungen in Personalressourcen führen könnte, beunruhigt mich.	○	○	○	○	○

8 Bemerkungen zur Umfrage

Sie sind fast am Ende angelangt. Gibt es noch etwas, das Sie zum Thema KI in der Sozialen Arbeit oder zu dieser Umfrage sagen möchten? Ich bin daran interessiert.

9 Teilnahme Gewinnspiel

Möchten Sie an der Verlosung teilnehmen und die Chance erhalten,
100 Franken zu gewinnen?

- ◯ Nein, ich möchte nicht teilnehmen.
- ◯ Ja, ich möchte teilnehmen. (Bitte geben Sie Ihre E-Mail-Adresse ein, damit ich Sie im Falle eines Gewinns kontaktieren kann)

10 Endseite

Sie sind am Ende der Umfrage angelangt.

Herzlichen Dank für Ihre Teilnahme!

Ihre Antworten sind gespeichert und die Umfrage ist beendet. Sie können das Fenster dann schliessen.

Sind Sie interessiert an den Ergebnissen der Masterarbeit? Dann bitte ich Sie um eine E-Mail mit dem Betreff "Resultate MT" auf muellsi7@students.zhaw.ch. Ich lasse Ihnen die Resultate zukommen, sobald die Thesis abgeschlossen ist.

Zürcher Hochschule
für Angewandte Wissenschaften

zh
aw

Soziale Arbeit

Persönliche Erklärung Einzelarbeit

Erklärung des/der Studierenden zur Master-Thesis

Studierende/r: (Name, Vorname)	Müller Simone
Master-Thesis: (Titel)	Künstliche Intelligenz im Kindesschutz
Abgabe: (Tag, Monat, Jahr)	18.06.2024
Begleitperson: (Dozent/in)	David Lätsch

Ich, obgenannte Studierende / obgenannter Studierender, habe die obgenannte Master-Thesis selbstständig verfasst.

Wo ich in der Master-Thesis aus Literatur oder Dokumenten *zitiere,* habe ich dies als Zitat kenntlich gemacht. Wo ich auf einen von anderen Autoren oder Autorinnen verfassten Text *referiere,* habe ich dies reglementskonform angegeben.

Ort, Datum:

Zürich, 18.06.2024

Unterschrift: