

Gesa Linnemann | Julian Löhe |
Beate Rottkemper (Hrsg.)

Künstliche Intelligenz in der Sozialen Arbeit: Grundlagen für Theorie und Praxis

BELTZ JUVENTA

Gesa Linnemann | Julian Löhe | Beate Rottkemper (Hrsg.)
Künstliche Intelligenz in der Sozialen Arbeit:
Grundlagen für Theorie und Praxis

Gesa Linnemann | Julian Löhe |
Beate Rottkemper (Hrsg.)

Künstliche Intelligenz in der Sozialen Arbeit: Grundlagen für Theorie und Praxis

BELTZ JUVENTA

Die Veröffentlichung wurde gefördert aus dem Open-Access-Publikationsfonds der Katholischen Hochschule Nordrhein-Westfalen.

Das Werk einschließlich aller seiner Teile ist urheberrechtlich geschützt. Der Text dieser Publikation wird unter der Lizenz **Creative Commons Namensnennung 4.0 International (CC BY 4.0)** veröffentlicht. Den vollständigen Lizenztext finden Sie unter: <https://creativecommons.org/licenses/by/4.0/deed.de>. Verwertung, die den Rahmen der **CC BY 4.0 Lizenz** überschreitet, ist ohne Zustimmung des Verlags unzulässig. Die in diesem Werk enthaltenen Bilder und sonstiges Drittmaterial unterliegen ebenfalls der genannten Creative Commons Lizenz, sofern sich aus der Quellenangabe/Abbildungslegende nichts anderes ergibt. Sofern das betreffende Material nicht unter der genannten Creative Commons Lizenz steht und die betreffende Handlung nicht nach gesetzlichen Vorschriften erlaubt ist, ist für die oben aufgeführten Weiterverwendungen des Materials die Einwilligung des jeweiligen Rechteinhabers einzuholen.

Trotz sorgfältiger inhaltlicher Kontrolle übernehmen wir keine Haftung für die Inhalte externer Links. Für den Inhalt der verlinkten Seiten sind ausschließlich deren Betreiber verantwortlich.



Dieses Buch ist erhältlich als:
ISBN 978-3-7799-8561-7 Print
ISBN 978-3-7799-8562-4 E-Book (PDF)
DOI 10.3262/978-3-7799-8562-4

1. Auflage 2025

© 2025 Gesa Linnemann, Julian Löhe, Beate Rottkemper und die Beitragsator:innen
Publikation: Beltz Juventa in der Beltz Verlagsgruppe GmbH & Co. KG
Werderstraße 10, 69469 Weinheim
service@beltz.de

Satz: xerif, le-tex
Druck und Bindung: Beltz Grafische Betriebe, Bad Langensalza
Beltz Grafische Betriebe ist ein Unternehmen mit finanziellem Klimabeitrag
(ID 15985-2104-1001)
Printed in Germany

Weitere Informationen zu unseren Autor:innen und Titeln finden Sie unter: www.beltz.de

Inhalt

Vorwort	7
Einführung <i>Gesa A. Linnemann, Julian Löhe, Beate Rottkemper</i>	9
Grundlagen der Künstlichen Intelligenz für die Soziale Arbeit <i>Beate Rottkemper</i>	19
Grundlagen der „Mensch-KI“-Interaktion – Auswirkungen auf den Einsatz im Kontext der Sozialen Arbeit <i>Gesa A. Linnemann</i>	35
Bedeutung von KI für Disziplin und Profession der Sozialen Arbeit <i>Jörn Dummann</i>	47
KI und Theorie(bildung) Sozialer Arbeit <i>Angelika Beranek</i>	60
Künstliche Intelligenz und Ethik – der verantwortliche Umgang mit einer neuen Technik <i>Wolfgang M. Heffels</i>	73
KI in der Kinder- und Jugendhilfe <i>Michael Macsenaere, Monika Feist-Ortmanns</i>	90
Künstliche Intelligenz als Gestalterin von Medienkulturen: eine medienpädagogische Perspektive auf eine sich verändernde Identitätsarbeit und Sozialisierung <i>Eik-Henning Tappe</i>	102
KI in der Beratung <i>Robert Lehmann</i>	117
Künstliche Intelligenz und Inklusion <i>Olivier Steiner</i>	128
KI und Alter: Einführung, Potenziale und Herausforderungen <i>Anna Schlomann</i>	141
Mensch, Maschine und Management: KI im Spannungsfeld von Sozialarbeit und Sozialmanagement <i>Julian Löhe</i>	156

Textanalysetechniken auf Tagesdokumentationen zur Prozessassistenz <i>Felix Holz, Michael Fellmann, Angelina Clara Schmidt</i>	174
Aktennotizerstellung in der Sozialen Arbeit durch Künstliche Intelligenz – Erkenntnisse aus einem Mixed-Method- Forschungsprojekt <i>Christina Plafky, Mitra Purandare, Benjamin Plattner, Svitlana Hrytsai</i>	190
IT-Sicherheit und Datenschutz im Kontext von KI-Sprachmodellen <i>Jan Pelzl</i>	204
EU AI Act und Soziale Arbeit: Die KI-Verordnung und ihre Auswirkungen <i>Sebastian Dötterl</i>	221
Künstliche Intelligenz in der Lehre der Sozialen Arbeit <i>Edeltraud Botzum, Madeleine Dörr, Andrea Gergen, Florian Müller</i>	241
Künstliche Intelligenz und Soziale Arbeit: Ausblick und Perspektiven <i>Gesa Linnemann, Julian Löhe, Beate Rottkemper</i>	255
Verzeichnis der Autor:innen	267

Vorwort

*„Wir müssen einsehen, dass die Technologie unser Traum ist und dass wir es sind,
die schließlich entscheiden, wie er enden wird.“*

Joseph Weizenbaum (2008)

„The silicone soapbox“ titelte Nature seine Ausgabe vom 18. Mai 2021, darunter: „AI system goes head to head with humans in competitive debates.“ Eine soapbox, Seifenkiste, ist im englischen Sprachraum im wörtlichen Sinne ein einfaches Mittel, um als Redner:in aufzutreten, indem man sich daraufstellt, und verweist im übertragenen Sinne auf freie Rede. In der Nature-Ausgabe berichten Slonim und Team vom Project Debater: Ein Sprachmodell hatte es in ihren Versuchen mit professionellen Redner:innen aufnehmen können (vgl. Slonim et al. 2021). Damit war ein weiterer Meilenstein nach den aufsehenerregenden KI-Erfolgen im Schach (IBMs DeepBlue gegen Garry Kasparow 1997), Jeopardy (Watson von IBM 2011) und im komplexen Go-Spiel (AlphaGo von DeepMind 2016) erreicht. In Form von Sprachassistenten wie Siri und Alexa gelangte die Kommunikation bereits viele Jahre zuvor über voice user interfaces in den Alltag.

In der Fachwelt waren weitere Entwicklungen Gegenstand der Aufmerksamkeit und in der Öffentlichkeit wurden Gefahren von Künstlicher Intelligenz von prominenten Personen diskutiert: So warnte Steven Hawking 2014 vor der Bedrohung durch KI für die Menschheit ebenso wie Elon Musk, der KI als „our biggest existential threat“ bezeichnete (vgl. Cellan-Jones 2014). In dieser Weise setzte sich die Debatte über Jahre fort. 2023 forderte ein offener Brief zu einem Moratorium auf: Alle großen KI-Experimente sollten zunächst ausgesetzt werden (vgl. Future of Life Institute 2023). Unterzeichner waren u. a. Stuart Russell, Elon Musk, Steve Wozniak und Yuval Noah Harari. Spätestens mit der Veröffentlichung von ChatGPT durch OpenAI im November 2022 war das Thema Künstliche Intelligenz auch in der breiten öffentlichen Debatte angekommen.

Wir als Herausgeber:innen hatten uns in unserer wissenschaftlichen und praktischen Laufbahn aus verschiedenen disziplinären (Soziale Arbeit, Psychologie, Wirtschaftsinformatik) und interdisziplinären Bezügen mit Mensch-Maschine-Interaktion, insbesondere über Sprache, und der Bedeutung von digitalen Technologien für Gesellschaft, Organisationen und Personen auseinandergesetzt und darüber zueinander gefunden. GPT2 und Project Debater waren für uns ein entscheidender Anstoß, die Bedeutung von Natural Language Processing für die Soziale Arbeit zu diskutieren (vgl. Linnemann/Löhe/Rottkemper 2023). Wie die öffentliche Debatte nahm auch die professionelle Auseinandersetzung zu, aus der Praxis erklang der Wunsch nach Orientierung. In Gesprächen und Projekten

mit Kolleg:innen und Vertreter:innen aus der Praxis wurde uns dies auch in unserem persönlichen Umfeld deutlich. Mit der Herausgabe dieses Handbuches führen wir die vorhandene Expertise zusammen, um für den deutschsprachigen Raum ein Grundlagen- und Nachschlagewerk vorzulegen. *Grundlagen für Theorie und Praxis* zu schaffen, ist dabei Wunsch und Anspruch zugleich – im Bewusstsein, dass dies nicht mehr sein kann als ein Anfang, auf dem in Zukunft weiter aufgebaut werden kann.

Wir danken ganz besonders den Autor:innen für die wertvollen Beiträge, die dieses Handbuch ermöglicht haben und ausmachen, und für die wunderbare Zusammenarbeit.

Ebenso gilt unser Dank dem Open-Access-Fonds der Katholischen Hochschule Nordrhein-Westfalen und der engagierten Begleitung durch Sarah Dudek. Die freie Online-Verfügbarkeit ist uns ein wichtiges Anliegen – gerade angesichts der Tragweite und Aktualität des Themas.

Dem Beltz-Verlag und dort insbesondere Julia Zubcic danken wir für die freundliche Begleitung und die Berücksichtigung des Werkes als Pilottitel zur Übertragung in die englische Sprache. Ulrike Weingärtner von TextAkzente danken wir für das kompetente Lektorat.

Münster, im Mai 2025

Gesa Linnemann, Julian Löhe und Beate Rottkemper

Literatur

- Cellan-Jones, Rory (2014): Stephen Hawking warns artificial intelligence could end mankind. In: BBC NEWS, 2. December. <https://www.bbc.com/news/technology-30290540> (Abfrage: 15.06.2025).
- Future of Life Institute (2023): Pause Giant AI Experiments: An Open Letter. <https://futureoflife.org/open-letter/pause-giant-ai-experiments/> (Abfrage: 15.06.2025).
- Linnemann, Gesa/Löhe, Julian/Rottkemper, Beate (2023): Bedeutung von Künstlicher Intelligenz in der Sozialen Arbeit. In: Soziale Passagen 15, S. 197–211. <https://doi.org/10.1007/s12592-023-00455-7>
- Slonim, Noam/Bilu, Yonatan/Alzate, Carlos/Bar-Haim, Roy/Bogin, Ben/Bonin, Francesca/Choshen, Leshem/Cohen-Karlik, Edo/Dankin, Lena/Edelstein, Lilach/Ein-Dor, Liat/Friedman-Melamed, Roni/Gavron, Assaf/Gera, Ariel/Gleize, Martin/Gretz, Shai/Gutfreund, Dan/Halfon, Alon/Hershovich, Daniel/Hoory, Ron/Hou, Yufang/Hummel Shay [...] Aharonov, Ranit (2021): An Autonomous Debating System. Nature 591(7850), S. 379–384. <https://doi.org/10.1038/s41586-021-03215-w>
- Weizenbaum, Joseph (2008, March 13): Alptrraum Computer. <https://www.zeit.de/1972/03/alptrraum-computer/komplettansicht> (Abfrage: 15.06.2025).

Einführung¹

Gesa A. Linnemann, Julian Löhe, Beate Rottkemper

Abstract: Die Einführung des Sammelbandes verortet Künstliche Intelligenz (KI) im Kontext gesellschaftlicher und professioneller Herausforderungen der Sozialen Arbeit. Sie beleuchtet Chancen und Risiken KI-gestützter Systeme, etwa bei Dokumentation, Beratung oder Assistenztechnologien, und verweist auf ethische, rechtliche sowie professionsspezifische Fragestellungen. Ein zentrales Anliegen ist die Entwicklung von KI-Kompetenz in Studium, Praxis und Organisationen. Besonders problematisch sind Verzerrungen in Trainingsdaten und die Gefahr des „Automation Bias“. Der Band versammelt Beiträge, die Grundlagen, Anwendungsfelder und Handlungsbedarfe aufzeigen – mit dem Ziel, einen reflektierten und verantwortungsvollen Umgang mit KI in der Sozialen Arbeit zu fördern. Fachkräfte sollen gestärkt werden, sich aktiv an der Gestaltung digitaler Entwicklungen zu beteiligen.

Keywords: Künstliche Intelligenz, Soziale Arbeit, KI-Kompetenz

Die Gesellschaft befindet sich, ebenso wie einzelne Organisationen, inmitten eines digitalen Wandels. Technologien wie die Künstliche Intelligenz (KI) beschleunigen diesen maßgeblich und beeinflussen inzwischen nicht nur Industrie und Wirtschaft, sondern auch soziale Dienstleistungen und Angebote sowie die Lebenswelten von Individuen (vgl. Banh/Strobel 2023). Zugleich sehen sich Individuen, die Gesellschaft und Soziale Organisationen Herausforderungen gegenüber, beispielsweise einem zunehmenden Fachkräftemangel bei gleichzeitigem Anstieg der Anzahl an hilfebedürftigen Menschen in unserer Gesellschaft. Nicht zuletzt sorgen der Klimawandel und multiple Krisen weltweit dafür, dass mehr Menschen nach Deutschland kommen und Schutz suchen (vgl. Grunwald/Langer/Sagmeister 2024). Damit ist der Umgang mit Technologien nur eine von vielen Herausforderungen, Digitalisierung und KI dringen jedoch zunehmend in alle Lebensbereiche von Menschen vor. Nur wenn Fachkräfte verschiedener Disziplinen sich aktiv am Diskurs über KI-Nutzung und -Entwicklung beteiligen, kann die digitale Transformation gelingen – und einen Beitrag zur Bewältigung der

1 © Die Autor:innen 2025, veröffentlicht im Verlag Beltz Juventa, Open Access: CC BY 4.0
Gesa Linnemann/Julian Löhe/Beate Rottkemper (Hg.), *Künstliche Intelligenz in der Sozialen Arbeit*
10.3262/978-3-7799-8562-4_001

Herausforderungen unserer Zeit leisten. Besonders relevant ist das in Fachgebieten wie der Sozialen Arbeit, da es sich hier um personenbezogene soziale Dienstleistungen handelt, für die einerseits sehr sensible Daten notwendig sind und andererseits schutzbedürftige Menschen adressiert werden, die auf die Expertise und die Fachlichkeit von Sozialarbeiter:innen angewiesen sind. Möglich ist in diesem Zusammenhang, dass Personen vulnerabler Gruppen in ihrer aktuellen Situation nicht in einem Maß für sich selbst eintreten können, wie es zur Bewertung von eingesetzten KI-Systemen notwendig wäre. Die Fachkräfte der Sozialen Arbeit haben hier eine besondere Verantwortung gegenüber ihren Klient:innen (vgl. Linnemann/Löhe/Rottkemper 2023b).

Vor allem die aktuell sehr weit diskutierten Modelle zur Text- und Sprachverarbeitung sind für die Soziale Arbeit von herausragender Bedeutung. Weil die Soziale Arbeit stark durch Dokumentation in natürlicher Sprache geprägt ist und ein großer Teil der sozialarbeiterischen Tätigkeiten in Form von Gesprächen stattfindet, gewinnen Methoden zur Auswertung, Analyse und Generierung von Text und Sprache besondere Relevanz (vgl. Linnemann/Löhe/Rottkemper 2023a). Spätestens seit der Veröffentlichung von ChatGPT im November 2022 wurde in der breiten Gesellschaft deutlich, dass die Qualität der Sprachmodelle eine neue Dimension eingenommen hat. Darüber hinaus werden die Gestaltungsmöglichkeiten mit KI-Systemen immer umfangreicher und ausgereifter. Damit werden sie auch für den Einsatz in der praktischen Arbeit zunehmend relevant. Der praktische Einsatz von KI-Systemen birgt jedoch zugleich Risiken und stellt viele Organisationen und Fachkräfte vor Herausforderungen. Die fehlende Nachvollziehbarkeit der Modelle und somit ein Mangel an Erklärbarkeit der Ergebnisse führen zu Schwierigkeiten im professionellen Kontext (vgl. Franzoni 2023). Gerade bei der Arbeit mit Menschen, die sich in Krisensituationen befinden, ist eine Überprüfbarkeit der vorgeschlagenen Unterstützungsmaßnahmen auf ihre Korrektheit und Evidenz unerlässlich. Jeder Anwendungsfall muss separat geprüft und auf Basis der Voraussetzungen entschieden werden, ob ein KI-System unterstützen kann und, wenn ja, welche Art von KI-System (siehe den Beitrag von Rottkemper in diesem Band). Ebenso muss die zukünftig weiter steigende Abhängigkeit von Technologien kritisch beleuchtet und bewertet werden. Wenn die Arbeitsprozesse immer stärker von Technologien abhängen, kann das dazu führen, dass bei einem Ausfall der Systeme beispielsweise Leistungen nicht mehr vollständig oder in gleicher Qualität erbracht werden können. Auch das Phänomen Deskilling – also das Verlernen bestimmter Tätigkeiten, weil diese zunehmend von Systemen unterstützt oder vollständig übernommen werden – fällt in diesen Themenkomplex (vgl. Farhan 2023). Das bedeutet keineswegs, dass die Möglichkeiten der Effizienzsteigerung nicht genutzt werden sollen. Jedoch sind Notfallpläne zum Wiederherstellen der zur Arbeitsfähigkeit notwendigen Prozesse einer Organisation bei einem Systemausfall oder das regelmäßige Durchführen von Tätigkeiten zur Sicherstellung des Kompetenzerhalts durch

das Fortschreiten der digitalen Transformation notwendig (siehe den Beitrag von Pelzl in diesem Band). Eine der wohl größten Herausforderungen in Bezug auf den Einsatz von KI in der Sozialen Arbeit ist der Umgang mit Verzerrungen und Voreingenommenheiten verschiedenster Ausprägung. Die Grundlage jedes KI-Systems, das auf Maschinellern Lernen beruht, sind die Trainingsdaten, mit denen der zugrunde liegende Algorithmus trainiert wurde (siehe den Beitrag von Rottkemper in diesem Band). Diese Daten sind zum einen historischer Natur und können somit überholte Gesellschaftsbilder enthalten und zum anderen sind sie menschlich erzeugt und enthalten demnach Vorurteile, die im Mindset der Erzeuger:innen der Daten manifestiert sind (vgl. Yi 2024; Lucy/Bamman 2021). Beispiele für diese Form von Verzerrungen gibt es sehr viele und aus unterschiedlichsten Bereichen der Gesellschaft: von Amazons Tool zur Auswahl von Bewerber:innen, das systematisch weibliche Kandidatinnen abgelehnt hat, weil es in der Historie von Amazon sehr wenige Entwicklerinnen gab² (vgl. Varsha 2023), über das systematische Streichen der Kindergeldansprüche für Familien mit Migrationsgeschichte durch ein KI-gestütztes Tool in den Niederlanden (vgl. Damen 2023) bis hin zu verstärkten geschlechterspezifischen oder anderweitig diskriminierenden Zuschreibungen bestimmter Berufsgruppen, Stereotype oder Situationen in KI-generierten Bildern (vgl. UNESCO 2023). Diese Vorurteile und Verzerrungen werden nicht in der KI selbst generiert, sondern sind menschengemacht. Aber sie können durch den Einsatz von KI verstärkt werden. Darüber hinaus führen Entscheidungen, basierend auf historischen Daten, nicht zu einer Weiterentwicklung der Gesellschaft, im Gegenteil: Es werden alte Gesellschaftsbilder und Normen zementiert und für die Zukunft übernommen (vgl. Linnemann/Löhe/Rottkemper 2023a). Die Verzerrung in den Daten tritt schon allein deswegen auf, weil nicht alle Menschen das Internet (die größte Datenquelle für das Training sehr großer KI-Modelle) gleichermaßen nutzen bzw. nutzen können. In der Regel sind hier Daten aus westlichen Gesellschaften stärker vertreten und somit auch in der Datengrundlage für das KI-Training enthalten. Diese Formen der Verzerrung sind in allen Bereichen der Gesellschaft problematisch – jedoch in der Sozialen Arbeit, die für sich explizit den Anspruch erhebt, Menschen gleichermaßen zu unterstützen und Chancengleichheit zu erhöhen oder herzustellen, in besonderem Maße. Hinzu kommt, dass die Adressat:innen Sozialer Arbeit oftmals nicht der ‚Norm‘ entsprechen – also jenem Bild, das in den im Internet verfügbaren Daten vorherrscht und damit die Grundlage für viele KI-Systeme bildet. Von dieser weichen viele Adressat:innen ab – selbst wenn sie in westlichen Gesellschaften leben. Gerade dieses Abweichen von der vermeintlichen Norm ist häufig der Grund für die Inanspruchnahme Sozialer

2 Das Tool ist aufgrund der Verzerrungen, die bereits in den Tests aufgefallen sind, nicht eingesetzt worden, hat aber dennoch eine Diskussion über die Manifestierung von Chancenungleichheit durch KI-Tools angestoßen.

Arbeit. Aufgrund dessen muss dieses Risiko bei der Entwicklung von KI-Systemen für die Sozialarbeit umfangreich Beachtung finden. Eine genaue Analyse der Trainingsdaten mit einem speziellen Fokus auf mögliche Verzerrungen ist daher in der Entwicklung von KI-Modellen erforderlich. Sollte es nicht möglich sein, vorurteilsfreie Daten für das Training zu bekommen, kann es notwendig sein, bestimmte Daten auszuschließen oder synthetische Daten herzustellen (vgl. Shah/Sureja 2025). Auch der Einbezug von Fachlichkeit und Prinzipien der Profession Sozialer Arbeit in der Bewertung der Ergebnisse sorgen für die Minimierung von Diskriminierungen durch KI-Systeme – ebenso wie diverse Entwickler:innenteams, die bereits auf technischer Ebene, etwa bei der Auswahl von Trainingsdaten, der Modellarchitektur und der Testung, Verzerrungen erkennen und gezielt beheben können (vgl. Alvarez et al. 2024). Einen Beitrag für diskriminierungsfreie KI-Systeme kann auch der EU AI Act leisten, indem er Hersteller von KI-Modellen dazu verpflichtet, die zum Training genutzten Daten zu dokumentieren, sodass die Basis der Algorithmen transparent gemacht werden kann (mehr zum EU AI Act siehe im Beitrag von Dötterl in diesem Band). Diesen Ansatz verfolgt ebenfalls der Forschungszweig der „Erklärbaren KI“ (explainable AI, XAI).

Eine weitere Form der Verzerrung ist der sogenannte Automation Bias. Dieser bezeichnet die Tendenz, die Lösung eines Algorithmus oder eines Systems nicht infrage zu stellen (vgl. Goddard/Roudsari/Wyatt 2012). Das ist bei der Funktionsweise von KI-Systemen allerdings hochproblematisch, da diese nicht die eine korrekte Lösung, sondern eine mit einer bestimmten Wahrscheinlichkeit richtige Lösung ausgeben, die in der Folge durch den:die Benutzer:in eingeordnet werden muss. Mangelnde Kompetenzen im Bereich KI bei gleichzeitigem Auftreten des Automation Bias kann somit sehr kritisch in der Sozialen Arbeit sein (siehe den Beitrag von Löhe in diesem Band). Auf der anderen Seite gibt es Fachkräfte, die technischen Innovationen grundsätzlich kritisch gegenüberstehen, auch diese Einstellung basiert oft auf Unwissenheit und damit einhergehender Unsicherheit – in diesem Fall kann durch Kompetenzbildung Vorbehalten begegnet werden. Dabei ist es nicht erforderlich, dass Fachkräfte die exakte Funktionsweise von KI-basierten Algorithmen verstehen, aber wie die Datengrundlage beispielsweise Einfluss auf die Ergebnisse hat und wie die durch eine KI generierte Lösungen zu interpretieren sind, sollten Fachkräfte für den Umgang mit KI-Systemen lernen (vgl. Schneider/Seelmeyer 2019). Dies fordert der EU AI Act ebenso. Ein gewisses Maß an AI Literacy und Daten-Literacy sorgen auch dafür, dass Fachkräfte in der Entwicklung der Systeme mitgestalten können und sie in ihrem Arbeitsalltag sensibel für die Situationen und Prozesse sind, bei denen sie durch KI-Systeme sinnvoll unterstützt werden können (vgl. Yi 2024). Denn Chancen zur Effizienzsteigerung im Arbeitsalltag gibt es zahlreiche. Infrage kommen hier vor allem Tätigkeiten in der Administration, z. B. die Abrechnung von Fällen, die Analyse von abgeschlossenen Fällen nach Anomalien oder ggf. die Dienstplanerstellung.

Teilweise wird auch diskutiert, inwiefern KI bei der Dokumentation oder der Erstellung von Hilfeplänen unterstützen kann, indem z. B. das Berichtswesen nach spezifischen Informationen durchsucht wird (siehe den Beitrag von Holz in diesem Band). In einigen (bisher wenigen) Fällen wird der Einsatz in Kerntätigkeiten der Sozialen Arbeit diskutiert. Überlegungen in dieser Richtung gibt es im Projekt SuchtGPT – „Gestaltung, Programmierung und Testung eines KI-basierten Chatbots für Suchtfragen“, das von der delphi Gesellschaft für Forschung, Beratung und Projektentwicklung mbH durchgeführt wird. Auf der Projekthomepage heißt es dazu: „Kann ein KI-basierter Chatbot Fragen aus dem Bereich der Suchthilfe korrekt beantworten und Ratsuchende angemessen unterstützen? Diese Frage soll im Rahmen des SuchtGPT Projektes beantwortet werden (delphi 2025). Unter der Annahme, dass „Ratsuchende“ Adressat:innen sind, wäre das ein direkter Einsatz von KI in Kerntätigkeiten der Sozialen Arbeit (zur Differenzierung verschiedener Tätigkeiten in der Sozialen Arbeit in Zusammenhang mit dem Einsatz von KI-Systemen siehe den Beitrag von Löhe in diesem Band). KI-Systeme können Fachkräfte der Sozialen Arbeit bei administrativen Routinetätigkeiten entlasten. Das schafft im besten Fall mehr Zeit für die Arbeit mit Klient:innen. Gleichzeitig ist zu beachten, dass der Einsatz von KI im direkten Kontakt mit Adressat:innen an ihre Grenzen stößt: Ihr induktives Vorgehen genügt nicht dem fachlichen Anspruch Sozialer Arbeit, der auf der Achtung der Menschenwürde und dem Verstehen individueller Lebenslagen basiert.

Im besten Fall werden die Prozesse durch den gezielten und fachlich begründeten Einsatz digitaler Tools und KI-gestützter Systeme nicht nur effizienter, sondern führen vor allem zu besseren Ergebnissen – nicht nur weil Ergebnisse durch digitale Technologien noch einmal auf ihre Qualität geprüft werden können, sondern auch, weil ganz neue Angebote geschaffen werden können. Beispielsweise können Chatbots Hilfesuchenden die richtige Anlaufstelle nennen und Sensorik sowie KI können in Form von Tools zur Sturzprognose, zum Messen und Einschätzen von Vitalparametern oder Smart-Home-Technologien dazu beitragen, dass ältere Menschen länger selbstständig in den eigenen vier Wänden leben können und Angehörige sowie Pflegekräfte entlastet werden (siehe den Beitrag von Schломann in diesem Band). Aber auch Teilhabe kann durch den Einsatz von KI-Technologien erhöht werden, beispielsweise durch die Übersetzung von Texten in Einfache oder Leichte Sprache, durch Smarte Brillen, die sehbehinderte Menschen dabei unterstützen, ihre Umwelt wahrzunehmen, und Texte vorlesen können, oder durch automatisierte Untertitel in Vorträgen oder Videoaufzeichnungen (siehe den Beitrag von Steiner in diesem Band). KI wird ebenfalls eingesetzt, um Menschen (gerade ältere und/oder demenzkranke Menschen) zu unterhalten und emotionale Resonanz zu erzeugen. Die Robbe Paro beispielsweise stammt aus Japan und ist schon seit 2004 auf dem Markt. Studien weisen darauf hin, dass der Einsatz mit einer geringeren Gabe von Schmerz- und Beruhigungsmitteln einhergeht. Dennoch kann und

sollte diskutiert werden, ob das die Lebensbereiche sind, in denen ein KI-Einsatz gewollt ist. Über menschliche Empathie oder ein eigenes Bewusstsein verfügt eine KI nicht. Mit Blick auf die Zukunft diskutieren Wissenschaftler:innen unterschiedlicher Disziplinen derweil, ob KI-Systeme in der Lage sind, zukünftig Bewusstsein zu erlangen. Dabei handelt es sich neben technischen Aspekten auch um eine philosophische Fragestellung (vgl. Schneider 2024, S. 38). Den dazugehörigen Diskurs³ in seiner Breite an dieser Stelle abzubilden, dient nicht der Zielsetzung des vorliegenden Sammelbandes. Gleichwohl wird das Thema KI und Ethik im Beitrag von Wolfgang M. Heffels aufgegriffen und in Bezug zur Sozialen Arbeit diskutiert. Aktuell gehen Wissenschaftler:innen davon aus, dass KI-Systeme kein Bewusstsein haben und menschliche Gefühle nur imitieren (vgl. Butlin et al. 2023)⁴. Es handelt sich immer nur um erlernte Reaktionen auf die menschliche Aktion. Und trotzdem bevorzugen erste Menschen den Austausch mit einem:einer KI-generierten „Freund:in“ vor menschlichen Kontakten (siehe den Beitrag von Linnemann in diesem Band). Der Beziehungsaspekt wird in unterschiedlichen Publikationen als entscheidend für den Erfolg Sozialer Arbeit herausgehoben (u. a. Schröder 2022, S. 350; Gödde 2016, S. 19; Urban 2004, S. 194; Flad et al. 2008, S. 104). Insofern scheint der zwischenmenschliche Faktor bei aller digitaler Unterstützung wesentlich – im Übrigen auch aus Sicht von Klient:innen. Abeld (2017, S. 13) verweist dazu auf eine Studie von Lorenz et al. (2007, S. 13), in der 80% der Klient:innen mit einer guten Vertrauensbasis zu ihren Betreuer:innen angegeben haben, dass die Hilfe sie stärkt. Die Frage, an welchen Stellen ein KI-Einsatz gewünscht und sinnvoll ist und an welchen Stellen nicht, muss individuell und auch aus professioneller Perspektive von Fall zu Fall betrachtet werden. In jedem Fall ist es jedoch unerlässlich, dass neben individuellen Bedürfnissen von Klient:innen die Fachlichkeit der Sozialen Arbeit eine übergeordnete Rolle spielen muss, wenn KI-Systeme zunehmend soziale Dienstleistungen durchdringen. Aus diesem Grund ist es essenziell, dass Fachkräfte sich mit der Funktionsweise von KI-Modellen (siehe den Beitrag von Rottkemper in diesem Band), ihren Einsatzmöglichkeiten und Risiken auseinandersetzen. Fehlt diese Auseinandersetzung, wird das Feld den großen Technologieunternehmen überlassen. Diese entscheiden dann, wie die KI-Anwendungen von morgen aussehen. Tritt dieses Szenario ein, bleibt den Fachkräften in der Sozialen Arbeit nur noch die Möglichkeit, das zu nutzen, was ihnen angeboten wird. Und es ist davon auszugehen, dass diese Angebote und Anwendungen ohne hinreichende sozialarbeiterische Expertise entwickelt würden.

3 Der Diskurs wird u. a. unter dem Stichwort „Transhumanismus“ geführt (vgl. van Oorschot 2023).

4 Wenngleich Butlin et al. in dem Preprint ihrer Untersuchung von 2023 unter Berücksichtigung bestehender Konzepte von Bewusstsein keine offensichtlichen technischen Hindernisse für die Entwicklung von KI-Systemen benennen konnten, die diese Indikatoren erfüllen.

Dabei ist „Mitmachen“ auch für IT-Lai:innen denkbar. Beispielsweise ist das Erstellen von Chatbots auf Grundlage von vorhandenen Anwendungen wie ChatGPT oder Claude ohne jegliche Programmierkenntnisse möglich. Ebenso können Roboter bereits mittels Befehlen in natürlicher Sprache programmiert oder eingerichtet werden. Das führt dazu, dass sie von Fachkräften selbst „antrainiert“ werden können. Dazu erforderliche KI-Kompetenz (oder „AI Literacy“) muss in Zukunft in der Ausbildung und im Studium von Sozialarbeiter:innen vermittelt werden. Die Vorbereitung auf einen immer stärker digitalisierten Arbeitsplatz und auf den Umgang der hier eingesetzten Tools ist essenziell für eine positive Gestaltung der eigenen Arbeitswelt mithilfe von KI. Auch wenn die Veränderungen im Beruf der Sozialen Arbeit verhältnismäßig gering ausfallen werden, da Tätigkeiten direkt am Menschen schwer zu formalisieren und für viele KI-Systeme komplex abzubilden sind, wird es in Zukunft immer wichtiger werden, digitale Tools begründet auswählen, einsetzen und mitgestalten zu können. Dabei sind einige Fragestellungen immer noch ungeklärt, was den Einsatz von KI im professionellen Kontext durchaus herausfordernd gestaltet, beispielsweise die Frage, wer bei Fehlern oder unrechtmäßigen Entscheidungen auf Basis der Ergebnisse des KI-Systems haftet. Beispiele hierfür gibt es bereits einige. Etwa das System, das in den Niederlanden eingesetzt wurde, um Kindergeldansprüche zu prüfen und das Menschen mit Migrationsgeschichte systematisch die Gelder gekürzt hat (vgl. Damen 2023). Haftet in solchen Fällen der Anbieter des Systems, da die Fehler schon im Training oder in der Programmierung manifestiert wurden oder die nutzende Stelle, da diese für die Konfiguration und den Einsatz verantwortlich ist? Eine weitere Schwierigkeit ergibt sich in der mangelnden Nachvollziehbarkeit der Algorithmen. Diese führt dazu, dass der Nachweis der Unrechtmäßigkeit für Geschädigte oftmals nicht durchführbar ist. Der EU AI Act schafft hier neue Maßstäbe der Nachvollziehbarkeit und somit der Haftbarkeit. Noch stehen entsprechende Rechtsurteile jedoch aus. Auch eine Kennzeichnungspflicht, wie u. a. im AI Act gefordert, ist erforderlich, um ethischen Maßstäben im Umgang mit KI gerecht zu werden (mehr zu ethischen Implikationen des KI-Einsatzes in der Sozialen Arbeit siehe den Beitrag von Heffels in diesem Band). Die damit hergestellte Transparenz über den Einsatz von KI-Systemen sorgt dafür, dass Klient:innen ihrer Wahlfreiheit nachkommen können: etwa ob sie mit einem Voicebot kommunizieren wollen oder nicht oder ob sie möchten, dass ihre Diagnose KI-gestützt ermittelt wird. Solche und ähnliche Fragestellungen verändern erneut den beruflichen Alltag und die Anforderungen an Sozialarbeiter:innen, wenn sie beispielsweise Klient:innen im Umgang mit KI-Systemen und entsprechenden Ergebnissen und Entscheidungen unterstützen (müssen).

Dieser Sammelband soll Fachkräften, Studierenden und Interessierten einen fundierten Einblick in das Thema KI in der Sozialen Arbeit geben und dazu beitragen, einen reflektierten und verantwortungsvollen Umgang mit diesen Technolo-

gien zu ermöglichen. Im Beitrag „Grundlagen der Künstlichen Intelligenz für die Soziale Arbeit“ gibt *Beate Rottkemper* eine Einführung in die historischen Entwicklungen und in die technischen Grundlagen der KI. Dafür werden verschiedene Datentypen betrachtet und ein kurzer Blick auf Möglichkeiten der Datenhaltung geworfen, bevor für die Soziale Arbeit relevante KI-Methoden vorgestellt und eingeordnet werden. Anschließend legt *Gesa Linnemann* in ihrem Beitrag „Grundlagen der „Mensch-KI“-Interaktion – Auswirkungen auf den Einsatz im Kontext der Sozialen Arbeit“ ein besonderes Augenmerk auf die Kommunikation mit KI-Systemen im Unterschied zur zwischenmenschlichen Kommunikation und untersucht Implikationen auf der Beziehungsebene. In „Bedeutung von KI für Disziplin und Profession der Sozialen Arbeit“ betrachtet *Jörn Dummann* die Auswirkungen von KI auf die Profession. Im Beitrag „KI und Theorie(bildung) Sozialer Arbeit“ werden durch *Angelika Beranek* Implikationen des KI-Einsatzes beleuchtet, bevor *Wolfjag M. Heffels* in seinem Beitrag „Künstliche Intelligenz und Ethik – der verantwortliche Umgang mit einer neuen Technik“ auf ethische Fragestellungen im Kontext von KI in der Sozialen Arbeit eingeht. Anschließend werden sechs Felder der Sozialen Arbeit in Bezug auf den Einsatz von KI und die damit einhergehenden Chancen und Risiken näher beleuchtet: Im Beitrag „KI in der Kinder- und Jugendhilfe“ geben *Michael Macsenaere* und *Monika Feist-Ortmanns* einen Überblick zu KI in diesem Handlungsfeld. *Eik-Henning Tappe* beleuchtet in „Künstliche Intelligenz als Gestalter von Medienkulturen: eine medienpädagogische Perspektive auf eine sich verändernde Identitätsarbeit und Sozialisierung“ das Themenfeld KI in der Jugendarbeit, bevor *Robert Lehmann* in seinem Beitrag „KI in der Beratung“ die Bedeutung von KI für diesen Bereich gibt, *Olivier Steiner* diskutiert in „Künstliche Intelligenz und Inklusion“ das entsprechende Spannungsfeld des Einsatzes, im Beitrag „KI und Alter: Einführung, Potenziale und Herausforderungen“ befasst sich *Anna Schlomann* mit Themen rund um KI und Alter. *Julian Löhe* widmet sich in „Mensch, Maschine und Management: KI im Spannungsfeld von Sozialarbeit und Sozialmanagement“ insbesondere den für einen erfolgreichen Einsatz von KI-Systemen notwendigen Veränderungen in der Organisation. *Felix Holz*, *Michael Fellmann* und *Angelina Clara Schmidt* erörtern in „Textanalysetechniken auf Tagesdokumentationen zur Prozessassistenz“ KI zur Unterstützung in der Auswertung von Textdokumenten und im Beitrag „Aktennotizerstellung in der Sozialen Arbeit durch Künstliche Intelligenz – Erkenntnisse aus einem Mixed-Method-Forschungsprojekt“ führen *Christina S. Plafky*, *Mitra Purandare*, *Benjamin Plattner* und *Svitlana Hrytsai* die Ergebnisse ihres Pilotprojektes aus. Die Themen IT-Sicherheit und Datenschutz sowie aktuelle Rahmenbedingungen und rechtliche Implikationen aufgrund des EU AI Act werden durch *Jan Pelzl*, „IT-Sicherheit und Datenschutz im Kontext von KI-Sprachmodellen“, bzw. *Sebastian Dötterl*, „EU AI Act und Soziale Arbeit: Die KI-Verordnung und ihre Auswirkungen“, erläutert. Abschließend greifen *Edeltraud Botzum*, *Madeleine Dörr*, *Andrea Gergen* und *Florian Müller* das Thema „Künstliche Intelligenz in der Lehre der Sozialen

Arbeit“ auf, bevor im Beitrag „Künstliche Intelligenz und Soziale Arbeit: Ausblick und Perspektive“ zukünftige Blickpunkte durch die Herausgeber:innen gegeben werden.

Literatur

- Abeld, Regina (2017): Professionelle Beziehungen in der Sozialen Arbeit. Eine integrale Exploration im Spiegel der Perspektiven von Klientinnen und Klienten. Wiesbaden: Springer VS.
- Alvarez, Jose M./Bringas Colmenarejo, Alejandra/Elobaid, Alaa/Fabbrizzi, Simone/Fahimi, Miriam/Ferrara, Antonio/Ghods, Siamak/Mougan, Carlos/Papageorgiou, Ioanna/Reyero, Paula/Russo, Mayra/Scott, Kristen M./State, Laura/Zhao, Xuan/Ruggieri, Salvatore (2024): Policy advice and best practices on bias and fairness in AI. In: *Ethics and Information Technology* 26, Artikel 31. <https://doi.org/10.1007/s10676-024-09746-w>
- Banh, Lenoardo/Strobel, Gero (2023): Generative artificial intelligence. In: *Electronic Markets* 33(1), S. 63.
- Butlin, Patrick/Long, Robert/Elmoznino, Eric/Bengio, Yoshua/Birch, Jonathan/Constant, Axel/Deane, George/Fleming, Stephen M./Frith, Chris/Ji, Xu/Kanai, Ryota/Klein, Colin/Lindsay, Grace/Michel, Matthias/Mudrik, Liad/Peters, Megan A. K./Schwitzgebel, Eric/Simon, Jonathan/VanRullen, Rufin (2023): Consciousness in Artificial Intelligence: Insights from the Science of Consciousness. Preprint auf arXiv:2308.08708v3 [cs.AI], 22. August 2023. <https://doi.org/10.48550/arXiv.2308.08708>
- Damen, Wes (2023): Sounds Good, Doesn't Work: The GDPR Principle of Transparency and Data-Driven Welfare Fraud Detection. In: Jorens, Yves (Hrsg.): *The Lighthouse Function of Social Law (LFSL 2023)*. Cham: Springer, S. 527–544. https://doi.org/10.1007/978-3-031-32822-0_26
- delphi (2025): SuchtGPT. Ein Chatbot für Suchtfragen? <https://suchtgpt.delphi.de> (Abfrage: 15.06.2025).
- Farhan, Akhmad (2023): The impact of artificial intelligence on human workers. In: *Journal of Communication Education* 17(2), S. 93–104.
- Flad, Carola/Schneider, Sabine/Treptow, Rainer (2008): Handlungskompetenz in der Jugendhilfe. Eine qualitative Studie zum Erfahrungswissen von Fachkräften. Wiesbaden: VS Verlag für Sozialwissenschaften.
- Franzoni, Valentina (2023): From Black Box to Glass Box: Advancing Transparency in Artificial Intelligence Systems for Ethical and Trustworthy AI. In: Gervasi, Osvaldo/Murgante, Beniamino/Misra, Sanjay/Garau, Chiara/Blečić, Ivan/Taniar, David/Apduhan, Bernady O./Rocha, Ana Maria A. C./Tarantino, Eufemia/Torre, Carmelo Maria/Karaca, Yeliz (Hrsg.): *Computational Science and Its Applications – ICCSA 2023 Workshops*. ICCSA 2023. *Lecture Notes in Computer Science*, Bd. 14107. Cham: Springer, S. 118–130. https://doi.org/10.1007/978-3-031-37114-1_9
- Goddard, Kate/Roudsari, Abdul/Wyatt, Jeremy C. (2012): Automation bias: a systematic review of frequency, effect mediators, and mitigators. In: *Journal of the American Medical Informatics Association* 19(1), S. 121–127.
- Gödde, Günter (2016): Die Weichenstellung zur therapeutischen Beziehung als vorrangigem Therapiefokus. In: Gödde, Günter/Stehle, Sabine (Hrsg.): *Die therapeutische Beziehung in der psychodynamischen Psychotherapie*. Ein Handbuch. Gießen: Psychosozial-Verlag, S. 19–50.
- Grunwald, Klaus/Langer, Andreas/Sagmeister, Monika (Hrsg.) (2024): *Sozialwirtschaft*. Handbuch für Wissenschaft, Studium und Praxis. 2., aktualisierte und erweiterte Auflage. Baden-Baden: Nomos.
- Linnemann, Gesa/Löhe, Julian/Rottkemper, Beate (2023a): Bedeutung von Künstlicher Intelligenz in der Sozialen Arbeit. In: *Soziale Passagen* 15, S. 197–211. <https://doi.org/10.1007/s12592-023-00455-7>

- Linnemann, Gesa/Löhe, Julian/Rottkemper, Beate (2023b): Der Einsatz von Künstlicher Intelligenz in der Kinder- und Jugendhilfe im DACH-Raum. In: *Jugendhilfe* 61(5), S. 415–421.
- Lorenz, Annette/Russo, Jasna/Scheibe, Florian (2007): *Aus eigener Sicht. Erfahrungen von NutzerInnen mit der Hilfe. Zwischenbericht Phase I.* Berlin: Der Paritätische.
- Lucy, Li/Bamman, David (2021): Gender and Representation Bias in GPT-3 Generated Stories. In: Akoury, Nader/Brahman, Faeze/Chaturvedi, Snigdha/Clark, Elizabeth/Iyyer, Mohit/Martin, Lara J. (Hrsg.): *Proceedings of the Third Workshop on Narrative Understanding (NUSE-WNU 2021)*. Virtual: Association for Computational Linguistics, S. 48–55. <https://aclanthology.org/2021.nuse-1.5/> (Abfrage: 15.06.2025).
- Schneider, Adolf (2024): Künstliche Intelligenz und Bewusstsein. In: *NET-Journal* 29(1/2), S. 34–42.
- Schneider, Diana/Seelmeyer, Udo (2019): Challenges in using big data to develop decision support systems for social work in Germany. In: *Journal of Technology in Human Services* 37(2–3), S. 113–128.
- Schröder, Carsten (2022): Wunderressource Empathie? In: *Sozial Extra* 46, S. 350–355. <https://doi.org/10.1007/s12054-022-00520-0>
- Shah, Milind/Sureja, Nitesh (2025): A comprehensive review of bias in deep learning models: Methods, impacts, and future directions. In: *Archives of Computational Methods in Engineering* 32(1), S. 255–267.
- UNESCO (2023): *AI and Gender Equality. A Global Study on the Gendered Impacts of Artificial Intelligence.* Paris: UNESCO. <https://unesdoc.unesco.org/ark:/48223/pf0000388971> (Abfrage: 15.06.2025).
- Urban, Ulrike (2004): *Professionelles Handeln zwischen Hilfe und Kontrolle – Sozialpädagogische Entscheidungsfindung in der Hilfeplanung.* Weinheim und München: Juventa.
- van Oorschoot, Frederike (2023): Theologische Positionen zu Transhumanismus und KI – ein Überblick. In: *Zeitschrift für Pädagogik und Theologie* 75(2), S. 139–151. <https://doi.org/10.1515/zpt-2023-2003>
- Varsha, P. S. (2023): How can we manage biases in artificial intelligence systems – A systematic literature review. In: *International Journal of Information Management Data Insights* 3(1), S. 100165.
- Yi, Yuan (2024): Research on the application risks and countermeasures of ChatGPT generative artificial intelligence in social work. In: *Journal of Artificial Intelligence Practice* 7(2), S. 166–172.

Grundlagen der Künstlichen Intelligenz für die Soziale Arbeit¹

Beate Rottkemper

Abstract: Um den Diskurs zum Einsatz von KI-Technologien in der Sozialen Arbeit sowie die Entwicklungen der Technologien selbst mitgestalten zu können, benötigen Fachkräfte fundiertes Grundlagenwissen über Methoden, Voraussetzungen und die Funktionsweise von Algorithmen. Dieser Beitrag führt in zentrale technische Begrifflichkeiten ein, von wissensbasierten Systemen über maschinelles Lernen und neuronale Netze bis hin zu Deep Learning und generativer KI. Ein besonderes Augenmerk liegt auf Large Language Models (LLM), da die natürliche Sprache im professionellen Handeln der Sozialen Arbeit eine zentrale Rolle spielt. Neben der Nutzung strukturierter und unstrukturierter Daten werden spezifische Risiken beim Einsatz der verschiedenen KI-Technologien erörtert. Darüber hinaus dient dieser Beitrag der Begriffsdefinition und -bestimmung für den vorliegenden Sammelband.

Keywords: Strukturierte Daten, unstrukturierte Daten, Maschinelles Lernen, Deep Learning, generative KI, Large Language Models

In diesem Beitrag werden die technischen Grundlagen Künstlicher Intelligenz (KI) erläutert. Ziel ist es, ein grundlegendes Verständnis von Daten, KI und ihrer Funktionsweise zu vermitteln, kein detailliertes Wissen über einzelne Algorithmen. Es werden übergeordnete Konzepte verschiedener Modelle betrachtet und miteinander verglichen. Außerdem wird die Relevanz einzelner Methoden für die Soziale Arbeit eingeordnet. Dabei wird kein Vorwissen über KI oder Algorithmen vorausgesetzt. Ziel des Beitrags ist es außerdem, eine gemeinsame Terminologie für dieses Buch zu schaffen. Darüber hinaus wird auf die Bedeutung von Daten und Datenqualität als Grundlage für das Training von KI-Algorithmen und somit für die mittels KI erzeugten Ergebnisse eingegangen.

1 © Die Autor:innen 2025, veröffentlicht im Verlag Beltz Juventa, Open Access: CC BY 4.0
Gesa Linnemann / Julian Löhe / Beate Rottkemper (Hg.), Künstliche Intelligenz in der Sozialen Arbeit
10.3262/978-3-7799-8562-4_002

1 Historische Einführung

KI und Sprachverstehen sind kein neues Phänomen, bereits in den 1950er-Jahren haben Wissenschaftler:innen daran gearbeitet, Übersetzungen zu automatisieren, um Geheimdienste im Kalten Krieg zu unterstützen. Der Begriff „Künstliche Intelligenz“ an sich fand erstmals in einem Förderantrag für die interdisziplinäre Dartmouth Conference im Jahr 1956 Verwendung (vgl. McCorduck/Cfe 2004, S. 111 ff.). Die Idee, menschliche Sprache möglichst realistisch nachzubilden und es somit zu ermöglichen, mit Maschinen zu kommunizieren, entstand in den 1960er-Jahren. Mit ELIZA entwickelte Joseph Weizenbaum ein Programm, das therapeutische Gespräche nachahmen sollte (vgl. Weizenbaum 1966). Es wurde allerdings schnell deutlich, dass die Möglichkeiten der Kommunikation aufgrund des fehlenden Kontexts sehr begrenzt waren, und auch Weizenbaum selbst hatte nicht die Idee, dass eine Maschine Therapien umfänglich ersetzen können soll, sondern wollte vor allem die Möglichkeiten der Technologie aufzeigen (vgl. Klopfenstein et al. 2017). In den Folgejahren gab es einen Knick in den Fortschritten der Entwicklungen und die anfängliche Euphorie in Bezug auf die Entwicklung von KI und insbesondere Sprachverstehen nahm ab, als deutlich wurde, dass die menschliche Sprache weitaus komplexer ist als anfänglich angenommen (vgl. Fradkov 2020). Die KI-Entwicklung konzentrierte sich in der Folge auf andere Themengebiete, beispielsweise die Lösung mathematischer Probleme und die Analyse strukturierter Daten mittels regelbasierter Ansätze (vgl. Jones 2015, S. 8 ff.).

Erst in den 1990er-Jahren nahm die KI-Forschung zum Textverständnis und zur Textverarbeitung (das sogenannte Natural Language Processing, kurz NLP) wieder Fahrt auf und wurde zu einem der wichtigsten Forschungsgebiete des Maschinellen Lernens (vgl. Torfi et al. 2020). In den frühen 2000er-Jahren ermöglichten neue Dimensionen in den Rechenleistungen und, aufgrund des Internets, sehr stark gestiegene Datenmengen weitere Fortschritte in der Datenverarbeitung und vor allem in der Analyse und im Erzeugen von Sprache bzw. Text (vgl. Liddy 2001). Mit weiteren Fortschritten in der Entwicklung des Deep Learning konnten auch die Kontextualisierungen der Texte verbessert werden (vgl. Hirschberg/Manning 2015). In der Sozialen Arbeit liegen große Mengen der Daten in Form von unstrukturierten Texten vor. Aufgrund dessen sind gerade die Methoden in Bezug auf Sprachverstehen und Kommunikation sehr relevant (vgl. Linnemann/Löhe/Rottkemper 2023) und stehen somit im Fokus dieses Beitrags. Es werden jedoch weitere Methoden kurz eingeführt und erläutert.

2 Daten und Datenspeicherung

Die Basis jeder KI sind Daten. Aus den Daten lernt die KI, Muster zu erkennen, und erstellt anschließend Analysen, Prognosen oder Artefakte (beispielsweise Texte, Bilder oder Videos) aus diesen Daten. Das heißt, die KI lernt aus den zur Verfügung gestellten Daten. Aber was sind eigentlich Daten genau? Und in welcher Form können sie auftreten?

Daten können z. B. Zahlen aus einer Abrechnung oder Messwerte (Zählerstände von Strom, Wasser oder Ergebnisse einer Blutdruckmessung) sein. Aber auch bei natürlichem Text oder bei Bildern handelt es sich um Daten. Grundsätzlich wird zwischen strukturierten und unstrukturierten Daten unterschieden. Dabei sind strukturierte Daten, wie der Name schon sagt, Daten, die in einer fest definierten Struktur vorliegen, beispielsweise die Adressdaten aller Kinder einer Kindertagesstätte, die für die Speicherung in einer Datenbank in ein ganz bestimmtes Format gebracht werden. In der Regel handelt es sich um Tabellen, in denen neue Einträge in Form einer Zeile eingefügt werden, für das Beispiel Kindergartenkinder werden dann etwa Name des Kindes, Name der Erziehungsberechtigten, Adresse, Betreuungszeiten etc. in einer vorab definierten Form gespeichert (siehe Abbildung 1). Diese Art der Datenspeicherung wird als Schema-on-Write bezeichnet. Die Daten werden dabei im Schreibprozess (also im Speicherprozess) in ein gewünschtes Schema gebracht, sodass sie im Anschluss ohne aufwendige weitere Verarbeitung genutzt werden können. In der Regel werden relationale Datenbanken genutzt, um strukturierte Daten auf diese Art und Weise zu speichern. Dies war lange die weitaus gängigste Methode der Datenspeicherung und -nutzung. Die Daten benötigen auf diese Art und Weise weniger Kapazitäten in der Speicherung und können ohne komplexe Methoden für Analysen und Prognosen genutzt werden. Da die Daten jedoch nicht in ihrem Quellformat vorliegen, sondern bestimmte Informationen in ein festes Format gebracht wurden, gehen bereits im Speicherprozess Informationen verloren. Problematisch kann das sein, wenn zum Zeitpunkt der Datenerhebung noch nicht bekannt ist, welche Auswertungen mit den Daten im weiteren Verlauf gemacht werden sollen (vgl. Fasel/Meier 2016).

Diese Form der klassischen Speicherung und Verarbeitung von Daten ist sinnvoll, wenn die Menge der Daten überschaubar ist, von Beginn an bekannt ist, wie die Daten in Zukunft genutzt werden sollen, und wenn die Daten ohne zu große Verluste in ein strukturiertes Format gebracht werden können. Zur Speicherung sehr großer Datenmengen, die oftmals dynamisch auftreten und nicht ohne große Verluste in ein strukturiertes Format überführt werden können, funktioniert das Speichern in relationalen Datenbanken nicht. Vor allem die verbreitete Nutzung des Internets hat dazu geführt, dass in jeder Sekunde große Mengen an Daten erzeugt werden. Hier wird von Big Data gesprochen. „Big“ bezieht sich dabei nicht nur auf die reine Menge der Daten, sondern auch auf Attribute

Abbildung 1: Beispielhafte Tabelle strukturierter Daten

Vorname	Nachname	Straße	Hausnr.	PLZ	Betreuungszeit (Stunden)
Emilia	Meier	Berliner Straße	346	12345	30
Mohammed
...
Anisha

Quelle: Eigene Darstellung

wie die Dynamik im Auftreten und die nicht gleichförmige Struktur. Es handelt sich hierbei z. B. um natürlichen Text, Bilder oder Videos (vgl. ebd.). In der Sozialen Arbeit liegen sehr große Mengen an Daten in Form von natürlichen Texten vor, beispielsweise Dokumentationen und Fallbeschreibungen. Diese werden in ihrem Ursprungszustand gespeichert und erst bei Nutzung der Daten verarbeitet. Dieses Vorgehen wird Schema-on-Read genannt. Die Speicherung unstrukturierter Daten benötigt deutlich mehr Kapazitäten und die Verarbeitung der Daten ist aufwendiger. Aber es gehen im Speicherprozess keine Informationen verloren. Gespeichert werden unstrukturierte Daten in der Regel nicht in klassischen Datenbanken, sondern in Data Lakes. Damit die Daten in den Data Lakes (die sehr groß werden können und dann mehrere Petabytes an Daten speichern) wiedergefunden werden können, müssen sogenannte Metainformationen mitgegeben werden (vgl. Sawadogo/Darmont 2020). Metainformationen für eine Fotodatei können z. B. das Datum der Aufnahme, das Motiv, Stichworte, der:die Fotograf:in und das Ereignis, bei dem das Bild aufgenommen wurde, sein.

Eine Mischform aus strukturierten und unstrukturierten Daten sind semistrukturierte Daten. Ein Beispiel für semistrukturierte Daten sind E-Mails. Diese haben strukturierte Elemente wie das Datum, die Empfänger:innen und den:die Absender:in, aber auch unstrukturierte Elemente wie den eigentlichen Textkörper der E-Mail und eventuelle Anhänge (vgl. Fasel/Meier 2016). Ebenso zur Speicherung semistrukturierter Daten gibt es spezielle Formate. Das bekannteste Format ist das sogenannte JSON-Format, das sowohl strukturierte als auch unstrukturierte Elemente erfassen kann. Ein weiteres Beispiel für semistrukturierte Daten in der Sozialen Arbeit sind Antragsformulare. Hier haben die einzelnen Felder eine feste Struktur und oftmals gibt es Anforderungen an die Art der Datenspeicherung in diesen Feldern. Felder mit Erläuterungen oder Ausführungen können jedoch nicht strukturiert abgelegt und demnach nicht mit klassischen Methoden ausgewertet werden.

3 Künstliche Intelligenz

Möglicherweise denken viele Menschen bei dem Begriff „Künstliche Intelligenz“ an Roboter, wie sie in Science-Fiction-Literatur oder Filmen vorkommen. Diese Roboter haben ein eigenes Bewusstsein und können selbstständig Entscheidungen treffen. Diese Form der KI wird „starke KI“ genannt. Aktuell wird davon ausgegangen, dass es eine solche KI nicht gibt und auch in Zukunft nicht geben wird (vgl. Fjelland 2020). Wenn aktuell von KI gesprochen wird, ist die sogenannte „schwache KI“ gemeint. Diese Form der KI erzeugt aus Daten, die ihr zugeführt werden (die sogenannten Trainingsdaten), Vorhersagen oder Prognosen und, im Fall von generativer KI, auch Artefakte wie Bilder, Videos oder Texte. Ein Gebiet innerhalb der schwachen KI ist das Maschinelle Lernen (siehe den Abschnitt über Maschinelles Lernen in diesem Beitrag). Das ist die Form der KI, von der heute in der Regel gesprochen wird. Beim Maschinellen Lernen werden stochastische Verfahren angewendet, um aus den Trainingsdaten Muster bzw. Regeln zu erzeugen, die anschließend auf neue Datensätze angewendet werden können und Prognosen oder Artefakte erzeugen (mehr dazu im weiteren Verlauf dieses Beitrags).

4 Wissensbasierte Systeme

Wissensbasierte Systeme wurden in den 1960er- und 1970er-Jahren intensiv erforscht. Diese Systeme lernen auf Basis vordefinierter Regeln und kuratiertem Wissen. Sie folgern dabei vom Allgemeinen auf das Spezielle und nutzen somit eine deduktive Logik zur Erzeugung der Ergebnisse. Damit können die Ergebnisse nur im Rahmen dessen erzeugt werden, was explizit in den Regeln abgebildet ist. Sind diese Regeln nicht korrekt oder ungenau formuliert, können falsche oder ungenaue Ergebnisse erzeugt werden.

Die am häufigsten angewendeten wissensbasierten Systeme sind Expertensysteme und regelbasierte Systeme. Dabei folgern regelbasierte Systeme aus klar definierten Wenn-dann-Regeln und generieren so Ergebnisse für Fragestellungen eines klar abgegrenzten Anwendungsbereichs. Klassische regelbasierte Systeme finden sich beispielsweise in Klimaanlage wieder, die anhand der Temperatur entweder heizen oder kühlen. Aufbauend auf dieser einfachen Form der Regelformulierung wurden weitere Algorithmen zur Anwendung deduktiver Logik entwickelt. Expertensysteme bilden dagegen hochspezialisiertes Wissen von menschlichen Expert:innen ab, auf dessen Basis sie dann, ebenfalls mittels definierter Regeln Ergebnisse erzeugen. Die Logiken in Expertensystemen können komplexer sein, als das in klassischen regelbasierten Systemen der Fall ist. So können Expertensysteme auch mit Unsicherheiten umgehen. Expertensysteme sind in der Sozialen Arbeit beispielsweise in Form von Chatbots zur Unterstützung der Onlineberatung oder -therapie relevant (vgl. Linnemann/Löhe/Rottkemper

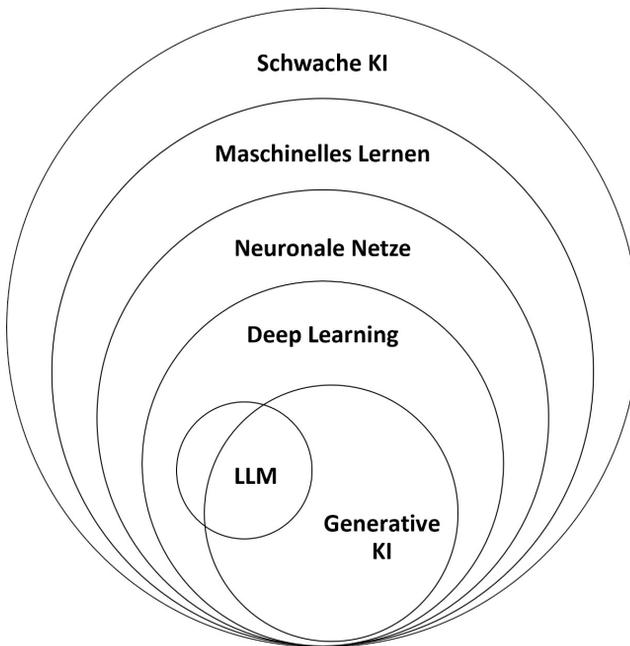
2024). Algorithmen, die zu den Expertensystemen gehören, sind z. B. Heuristiken, Entscheidungsbäume oder wissensbasierte Netzwerke.

Ein Vorteil wissensbasierter Systeme ist ihre Erklärbarkeit. Das heißt, es ist nachvollziehbar, anhand welcher Kriterien welche Lösungen generiert wurden. Die Erklärbarkeit von KI-Systemen ist ein Merkmal, das seit dem Aufkommen von generativen KI-Systemen vielfach diskutiert und gefordert wird. Darüber hinaus fließt in die Lösungen wissensbasierter Systeme eine kuratierte Informations- und Wissensgrundlage. Es kommt also weder zu Halluzinationen (zur Erläuterung des Begriffs siehe Abschnitt Generative KI weiter unten in diesem Beitrag) noch zur Ausgabe falscher Informationen (es sei denn, die Wissensgrundlage ist nicht korrekt oder passend gewählt). Attribute wie Erklärbarkeit und Freiheit von Halluzinationen sind für viele Anwendungen in der Sozialen Arbeit absolute Voraussetzung, beispielsweise für Algorithmen, die direkt mit Hilfesuchenden kommunizieren. Mit wissensbasierten Systemen lassen sich nur eingegrenzte und klar abgesteckte Problemstellungen lösen. Das Modell ist im Nachgang nur mit viel Aufwand für andere Fragestellungen änderbar oder erweiterbar, und die Regeln und Zusammenhänge, die dem System zugrunde liegen, müssen alle explizit formuliert werden können. Demnach darf die Problemstellung nicht zu komplex sein, da die Formulierung des Modells schnell sehr aufwendig wird oder sogar nicht mehr möglich ist.

5 Maschinelles Lernen

Die Idee des Maschinellen Lernens basiert darauf, das Vorgehen und die Struktur von menschlichem Lernen künstlich nachzubilden (vgl. Fradkov 2020). Methoden des Maschinellen Lernens nutzen im Gegensatz zu wissensbasierten Methoden induktive Logik zum Lernen. Das heißt, sie lernen in der Regel mit stochastischen Methoden aus historischen Daten ohne vorab explizit formulierte Regeln. Bei dieser Art des Lernens wird vom Speziellen auf das Allgemeine geschlossen. Die Algorithmen können somit auch mit Daten arbeiten, die vorab nicht bekannt sind, die Ergebnisse beruhen jedoch auf Wahrscheinlichkeiten und müssen demnach nicht korrekt sein und im Nachgang evaluiert werden. Im Maschinellen Lernen werden statistische Modelle angewendet, um Zusammenhänge zwischen den unabhängigen Variablen und den abhängigen Variablen zu ermitteln. Die abhängigen Variablen sind die Zielgrößen, die mithilfe des Modells untersucht werden sollen, unabhängige Variablen sind die Eingabe- oder Einflussgrößen, anhand derer die Vorhersagen (die abhängigen Variablen) ermittelt werden. Es gibt beispielsweise Algorithmen, die prognostizieren sollen, ob ein Kind zukünftig aus einer Familie genommen wird oder nicht (abhängige Variable), und dafür Einflussgrößen wie Vorfälle in der Familie, Vorbestrafung der Eltern etc. (unabhängige Variablen) heranziehen.

Abbildung 2: Übersicht schwache KI



Quelle: Eigene Darstellung nach Li et al. 2021

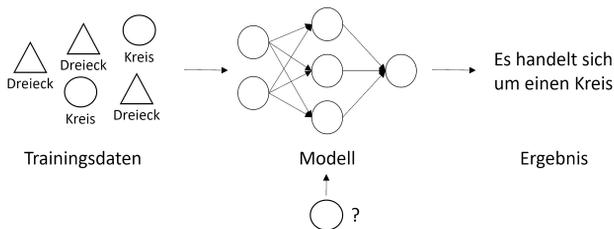
Im Maschinellen Lernen wird zwischen überwachten und unüberwachten Verfahren unterschieden. Darüber hinaus werden im weiteren Verlauf des Beitrags Begriffe wie Neuronale Netze, Deep Learning, generative KI und Large Language Models (LLM) eingeordnet und erläutert (siehe Abbildung 2).

5.1 Überwachtes und unüberwachtes Lernen

Im Maschinellen Lernen wird zwischen überwachten und unüberwachten Lernverfahren unterschieden. Dabei liegt die Differenz im Umgang mit den Daten, die für das Training der Algorithmen verwendet wird. Im überwachten Lernen werden die Trainingsdaten vorab gelabelt. Das heißt, es ist bekannt, welche Art Ergebnis erwartet wird, und die Daten werden, in der Regel durch einen Menschen, gekennzeichnet. Ein klassisches Beispiel ist das Identifizieren eines Hundes oder einer Katze auf einem Bild. Bei diesem Beispiel können die Trainingsdaten vorab durch einen Menschen gesichtet werden und es kann notiert werden, ob das Bild

einen Hund oder eine Katze zeigt.² Der Algorithmus bekommt dann zum Training die gelabelten Bilder zugeführt und ermittelt selbstständig, anhand welcher Kriterien Hunde und Katzen unterschieden werden können (Abbildung 3). Das heißt, das Ergebnis ist festgelegt, aber der Weg dahin bleibt weitestgehend dem Algorithmus überlassen (vgl. Goodfellow/Bengio/Courville 2016, S. 103 ff.).

Abbildung 3: Überwachtes Lernen



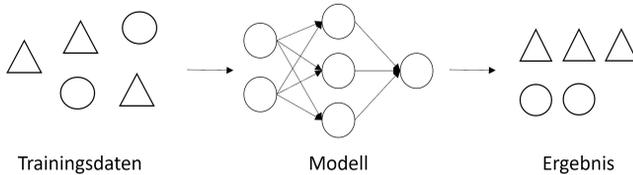
Quelle: Eigene Darstellung nach Choo et al. 2021

Beim sogenannten unüberwachten Lernen ist auch das Ergebnis nicht vorgegeben. Diese Verfahren arbeiten mit nicht gelabelten Trainingsdaten. Somit können zwar Muster oder Gemeinsamkeiten in den Daten gefunden werden, jedoch keine klare Benennung dieser erfolgen, da das Modell nicht weiß, was diese Muster oder Gemeinsamkeiten, die es identifiziert hat, bedeuten. Bei diesen Verfahren ist ein Postprocessing der Ergebnisse demnach unabdingbar. Es können durch unüberwachte Lernverfahren ganz neue Muster und Zusammenhänge identifiziert werden, die dann im Nachgang durch menschliche Nachbearbeitung in einen Kontext gesetzt werden (Abbildung4). Mit unüberwachten Verfahren können sehr große Datenmengen verarbeitet werden, da diese vorab nicht bearbeitet werden müssen. Viele Verfahren, die aktuell verwendet werden, sind eine Kombination aus überwachten und unüberwachten Lernverfahren.

Um zu vermeiden, dass die Modelle nicht zu stark auf die im Training verwendeten Daten angepasst werden und später in der Anwendung nicht mehr adäquat funktionieren, werden die vorhandenen Daten in Trainings- und Testdaten aufgeteilt. Der Großteil der Daten wird zum Training verwendet (in der Regel ca. 80%). Im Trainingsprozess werden üblicherweise mehrfach die gleichen Daten genutzt, um das Modell durch Anpassung von Kriterien, Lernzyklen und anderen Parametern sukzessive zu verbessern. Erzeugt das Modell dann qualitativ über-

2 Ein Teil dieses Trainings wurde an die breite Gesellschaft ausgelagert, indem die sogenannten Captcha-Abfragen als Identifizierung einer menschlichen Person eingeführt wurden, bei denen angegeben werden muss, ob auf einem Bild ein Bus, eine Ampel oder eine Brücke zu sehen ist. Die Ergebnisse dieses Trainings sind in die Entwicklung selbstfahrender Autos eingeflossen (vgl. Plesner/Vontobel/Wattenhofer 2024).

Abbildung 4: Unüberwachtes Lernen



Quelle: Eigene Darstellung nach Sindhu Meena/Suriya 2020

zeugende Ergebnisse, werden die Testdaten (ca. 20 %) genutzt, um die Ergebnisse mit Daten, die das Modell bis dato nicht kennt, zu validieren.

Im Gegensatz zu wissensbasierten Systemen können Verfahren des Maschinellen Lernens komplexe Problemstellungen lösen, da diese Verfahren durch das Schließen vom Speziellen auf das Allgemeine, ohne dass alle Regeln und jede Information explizit abgebildet werden, Schlussfolgerungen ziehen können. Sie sind außerdem ohne umfangreiche Anpassungen flexibel auf verschiedene Fragestellungen anwendbar und können auch mit großen Datenmengen umgehen. Vor allem Verfahren des unüberwachten Lernens können sehr große Datenmengen verarbeiten. Dabei können die Algorithmen komplett neue Muster in den Daten erkennen, auf die sie nicht speziell trainiert wurden. Allerdings ist die Erklärbarkeit der Ergebnisse oft nicht gegeben, da nicht nachvollziehbar ist, anhand welcher Kriterien der Algorithmus welche Entscheidungen getroffen hat und deswegen auf eine bestimmte Lösung kommt (hier wird auch von einem Black-Box-System gesprochen). Darüber hinaus sind für das Training der Modelle große Datenmengen erforderlich. Diese Daten beeinflussen die späteren Ergebnisse des Modells, da dieses ja anhand der Trainingsdaten justiert wird. Dementsprechend besteht die Gefahr der Überanpassung des Modells an die Trainingsdaten. Das heißt, es kann dazu kommen, dass das Modell im Training zwar gute Ergebnisse liefert, in der Anwendung anschließend aber nicht (vgl. Goodfellow/Bengio/Courville 2016, S. 108 ff.).³ Darüber hinaus kann es in Algorithmen des Maschinellen Lernens aufgrund der Abhängigkeit der Modellqualität von den Trainingsdaten zu Diskriminierung bestimmter Bevölkerungsgruppen kommen. Diskriminierung kann auftreten, wenn die Trainingsdaten nicht das für die Fragestellung tatsächlich relevante System abbilden (hier wird auch von Bias oder auf Deutsch Verzerrung in den Daten gesprochen). Erschreckende

3 Ein prominentes Beispiel für eine Überanpassung an die Trainingsdaten und für eine nicht adäquate Auswahl dieser Trainingsdaten war ein Bilderkennungsalgorithmus, der Kühe nur vor einem grünen Hintergrund erkannt hat, weil die Trainingsdaten ausschließlich Kühe auf grünen Wiesen gezeigt haben. Der Algorithmus hat als Kriterium u. a. den grünen Hintergrund gewählt und nicht tatsächliche Merkmale einer Kuh. Dieses Phänomen nennt sich Shortcut Learning, da der Algorithmus sozusagen eine Abkürzung im Lernen nutzt (vgl. Geirhos et al. 2020).

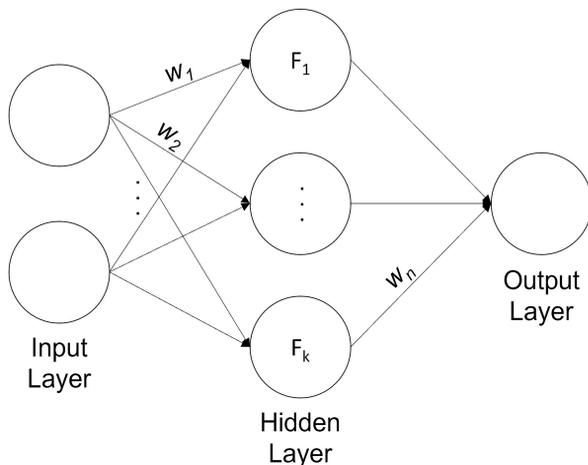
Beispiele dafür gibt es viele, eines aus der Sozialen Arbeit stammt aus den Niederlanden. Hier wurde ein Algorithmus eingesetzt, um Kindergeldansprüche zu prüfen. Aufgrund der Ergebnisse des Algorithmus wurden Familien mit Migrationsgeschichte systematisch die Kindergeldansprüche gestrichen. Infolgedessen trat eine Regierung zurück und es gab Milliardenstrafen an Schadensersatz (vgl. Hadwick/Lan 2021).

5.2 Neuronale Netze und Deep Learning

Neuronale Netze, manchmal auch Künstliche Neuronale Netze (KNN) genannt, sind Verfahren des maschinellen Lernens. Sie sind in ihrer Struktur dem menschlichen Gehirn und in ihrer Funktionsweise dem menschlichen Denken und Lernen nachempfunden. Die Netze bestehen aus vielen Neuronen, die Input-Signale erhalten und verarbeiten und diese anschließend als Output-Signale weitergeben. Die Neuronen sind in sogenannten Schichten organisiert (siehe Abbildung 5). Es gibt eine Eingabeschicht (Input Layer), eine oder mehrere „verborgene“ Schichten (Hidden Layer) und eine Ausgabeschicht (Output Layer). Die Neuronen einer Schicht sind mit allen oder einigen Neuronen der folgenden Schicht verbunden. Diese Verbindungen sind gewichtet (w_1, \dots, w_n) und bestimmen somit den Einfluss eines bestimmten Signals auf das nachfolgende Neuron (vgl. Theodoridis 2015, S. 904 ff.). Die Rohdaten (Texte, Bilder oder ähnliches) gelangen über die Eingabeschicht in das Netz und werden dann in den verschiedenen Neuronen verarbeitet und analysiert (mittels der Funktionen F_1, \dots, F_k). Die Gewichte an den Verbindungen werden sukzessive angepasst (das Modell wird also trainiert), bis das Ergebnis, das über die Ausgabeschicht erzeugt wird, eine zufriedenstellende Qualität erreicht hat. Das Training an sich erfolgt über eine Rückmeldung an das System zur Ergebnisqualität. Ist das Ergebnis nicht korrekt oder die Qualität nicht ausreichend, wird dies „rückwärts“ von der Ausgabeschicht bis zur Eingabeschicht durch das Netzwerk gemeldet und die Gewichte an den Kanten der Verbindungen werden angepasst. Dieser Prozess wird im Englischen Backpropagation genannt. Das Vorgehen ähnelt also, vor allem zu Beginn, einem Trial-and-Error-Ansatz, und nur durch die Rückmeldung des Fehlers kann das System seine Ergebnisse verbessern (ebd., S. 913 ff.).

Neuronale Netze mit mehr als einer verborgenen Schicht zählen zum Deep Learning. Sie sind besonders gut zur Analyse von sehr komplexen Strukturen geeignet (vgl. Russell/Norvig 2022, S. 750 ff.). Dafür sorgt u. a. die Fähigkeit, durch das Zusammenspiel der verschiedenen Hidden Layer relevante Merkmale zum Erkennen gesuchter Muster oder Abhängigkeiten selbstständig zu erlernen. Es gibt zahlreiche Weiterentwicklungen zur Lösung spezifischer Problemstellungen. Beispielsweise sind Rekurrente Neuronale Netze (RNN) oder Transformer-Modelle besonders geeignet, um Texte zu verarbeiten (vgl. Lin/Tegmark 2016).

Abbildung 5: Neuronales Netz



Quelle: Eigene Darstellung nach Choo et al. 2021

Beide Verfahren nutzen Deep-Learning-Technologien. In der Regel werden in einem Neuronalem Netz Informationen sequenziell verarbeitet, Schicht für Schicht werden die Informationen verarbeitet und an die nächste Schicht weitergeleitet. Die jeweiligen Ausgabesignale einer Schicht sind für die nachfolgende Schicht die Inputsignale (vgl. Sherstinsky 2020). Durch diese sequenzielle Verarbeitung der Daten sind Zusammenhänge zwischen einzelnen Teilen einer Sequenz (beispielsweise eines Eingabe-Satzes, der in eine andere Sprache übersetzt werden soll), die weit auseinander liegen, nur mit zusätzlichen Techniken abzubilden. Darüber hinaus ist die Verarbeitungsgeschwindigkeit von Daten mittels RNN verhältnismäßig langsam, da moderne Hardware-Architektur, bestehend aus parallelen Prozessoren, durch die sequenzielle Berechnung nicht optimal genutzt wird (vgl. Mienye/Swart/Obaido 2024). Diese Herausforderungen werden durch die modernen Transformer-Modelle adressiert. Transformer-Modelle verarbeiten die komplette Sequenz parallel anstatt schrittweise, was eine deutlich höhere Performance erlaubt (vgl. Raparathi et al. 2021). Darüber hinaus sind sie in der Lage, mittels eines speziellen Aufmerksamkeitsalgorithmus Zusammenhänge zwischen weit entfernten Teilen der Sequenz zu erfassen und in der Verarbeitung zu berücksichtigen (vgl. Vaswani et al. 2017). Das sorgt dafür, dass auch komplexe Texte und ihr semantischer Zusammenhang erfasst und verarbeitet werden können. Für das Training von Transformer-Modellen werden jedoch sehr große (nicht zwangsweise gelabelte) Datenmengen benötigt (vgl. Rahali/Akhlooufi 2023).

5.3 Generative KI

Generative KI basiert auf Verfahren des Deep Learning. Die Besonderheit an generativen Modellen im Gegensatz zu diskriminativen KI-Modellen ist, dass sie neue Inhalte erzeugen können. Diskriminative Modelle gelten oftmals als die klassischen KI-Modelle, die auf Basis der Trainingsdaten mittels stochastischer Zusammenhänge lernen, neue Input-Daten zu kategorisieren (vgl. Jaakkola/Haussler 1998). Die Modelle rekonstruieren dabei nicht die Verteilung der Daten an sich, sondern versuchen, anhand der Trainingsdaten, die wahrscheinlichsten Input-Output Kombinationen zu ermitteln. Generative Modelle hingegen lernen auf Basis der Trainingsdaten die Verteilung oder die Funktion kennen und nutzen diese anschließend, um neue Daten einzuordnen (vgl. Jebara 2004). Damit sind diese Algorithmen in der Lage, anhand der ermittelten Funktion auch komplett neue Daten auf Basis der gegebenen Input-Daten zu generieren und nicht nur Kategorisierungen von Daten vorzunehmen (vgl. Bishop/Lasserre 2007). Auf Basis einer – in der Regel sehr großen – Datenbasis, die zum Training der Modelle verwendet wird, werden durch generative Modelle anschließend neue Artefakte generiert. Artefakte können beispielsweise Texte, Bilder oder Videos sein. Lange wurde angenommen, dass die Kreativität den Menschen vorbehalten bliebe und KI zwar Prognosen oder Einordnungen vornehmen kann, nicht aber neue Inhalte erzeugen könne (vgl. Boden 2009). Diese Annahme wurde durch die Verfahren der generativen KI widerlegt. Grundsätzlich ist die Funktionsweise jedoch ähnlich wie bei den vorab beschriebenen Algorithmen: Anhand einer großen Menge an Daten werden Muster erlernt (beispielsweise Muster und Regeln der natürlichen Sprache, wenn es um das Erzeugen von Texten geht) und anhand dieser Muster werden anschließend neue Inhalte generiert. Dabei hat das Modell kein Verständnis für die Bedeutung, beispielsweise von Texten, sondern generiert einfach die am wahrscheinlichsten folgenden nächsten Worte in einem Satz. Aufgrund der überaus großen Menge an Trainingsdaten, die dem Modell zugrunde liegt, ist das Ergebnis oft nicht von einem menschlich generierten Text zu unterscheiden (zu Kreativität im Kontext Künstlicher Intelligenz siehe auch Boden 1998).

Da diese sehr großen generativen Modelle oftmals mit Allgemeinwissen aus dem Internet trainiert wurden und damit nicht spezifischen Domänen entsprechen, kann es passieren, dass sie zu konkreten Expert:innenfragen ungenaue oder falsche Antworten liefern. Darüber hinaus kommt es bei generativen KI-Systemen immer wieder zu sogenannten Halluzinationen. Halluzinationen sind faktisch falsche Ausgaben, die nicht auf den Trainingsdaten beruhen (vgl. Maleki/Padmanabhan/Dutta 2024). Die Modelle sind derart trainiert, dass sie immer das wahrscheinlichste Ergebnis ausgeben. Sie sind nicht darauf trainiert, Fakten zu prüfen oder im Fall von mangelnden Informationen in den Trainingsdaten kein Ergebnis zu liefern. Aktuell gibt es noch keine zuverlässige Möglichkeit, Halluzi-

nationen in generativen KI-Systemen auszuschließen. Es handelt sich hierbei um ein aktuelles Forschungsgebiet. Problematisch dabei ist vor allem, dass die Ausgaben der Systeme trotz falscher Fakten in der Regel sehr plausibel formuliert sind und die meisten Menschen im Umgang mit einer Software nicht von falschen oder diskriminierenden Ausgaben ausgehen (vgl. ebd.). Die Interpretation der Ergebnisse und der Umgang mit generativen KI-Systemen bedürfen eines bestimmten Maßes an Literacy. Eine Möglichkeit, die Qualität der Ausgaben zu verbessern, ist es, die großen generativen Modelle zu nutzen, um auf ihrer Basis spezifischere Modelle mit Expert:innenwissen zu trainieren. Die sogenannten Pretrained-Modelle bieten sich dafür an, sie mit Domänenwissen anzureichern (dieser Prozess wird Finetuning genannt). Es gibt darüber hinaus Technologien, die Wissensdatenbanken in Kombination mit generativen KI-Modellen nutzen und somit die hohe Qualität der Textverarbeitung und -generierung mit den Vorteilen kuratierter Wissensdatenbanken verbinden (vgl. Lewis 2020). Darüber hinaus ermöglichen vortrainierte Modelle in Kombination mit den Möglichkeiten der automatisierten Verarbeitung natürlicher Sprache das Training von Algorithmen auch ohne weitreichende Kenntnisse des Maschinellen Lernens. Somit können ebenso Fachkräfte der Sozialen Arbeit oder aus anderen Berufszweigen für sie relevante Modelle mit- oder weiterentwickeln. Dieses Vorgehen stellt sicher, dass nicht nur Expert:innenwissen von Entwickler:innen in die Modelle einfließt, sondern auch Domänenwissen, ethische Grundsätze und Fachexpertise.

5.4 Large Language Models (Multimodal Large Language Models)

Da sich die Forschung im Bereich der KI schon immer intensiv auf die Textanalyse und Textgenerierung konzentriert hat, bilden die sogenannten Large Language Models (LLM) eine eigene Klasse von KI-Systemen. Einige Verfahren gehen auf das seit den 1950er-Jahren etablierte NLP zurück, die Analyse natürlicher Sprache mittels interdisziplinärer Methoden, u. a. Computertechnologie (vgl. Jones 1994). Erste LLM basierten auf RNN-Technologien, die Entwicklung der Transformer-Architektur hat die Qualität der LLM maßgeblich gesteigert, sodass quasi alle aktuellen LLM, die sich auf Textverstehen und -generierung konzentrieren, auf dieser Technologie basieren. Das erste LLM auf Basis der Transformer-Architektur war BERT von Google. Aber auch das im Jahr 2018 veröffentlichte GPT von OpenAI nutzt Deep Learning und Transformer-Technologien und ist in der Lage, anhand von wenigen Input-Parametern qualitativ hochwertige Texte zu generieren, die oftmals nicht mehr von menschlich verfassten Texten unterschieden werden können (vgl. Elkins / Chun 2020; Floridi / Chiriatti 2020). Es handelt sich um ein autoregressives Sprachmodell. Das heißt, es lernt schrittweise aus den eigenen Daten, ohne dass vorab umfangreiche Regeln formalisiert werden müssen, und verbessert sich so stetig (vgl. Zong / Krishnamachari 2022). Einer breiten Öffentlichkeit

wurde die Technik durch den von OpenAI entwickelten Chatbot „ChatGPT“ (Veröffentlichung im November 2022) bekannt, als dieser Anfang 2023 eine Jura-Prüfung an der Universität von Minnesota bestand und damit viel Aufmerksamkeit erzeugte (vgl. Choi et al. 2021; Katz et al. 2024). Da in der Sozialwirtschaft große Mengen an Text erzeugt und verarbeitet werden und eine sehr große Menge an Daten in Form von natürlicher Sprache vorliegt, erscheinen LLM eine äußerst relevante Technik zu sein. Der Großteil der LLM gehört zum Gebiet der generativen KI, da die Modelle darauf trainiert sind, anhand von (wenigen) Input-Parametern Text zu erzeugen. Es gibt jedoch auch LLM, die für andere Zwecke entwickelt wurden, beispielsweise um Texte zusammenzufassen oder zu kategorisieren (siehe Abbildung 2). Aktuelle Modelle können mit ähnlichen Verfahren nicht nur Texte, sondern ebenso Bilder, Videos etc. erzeugen. Diese Modelle werden Multimodal Large Language Models genannt.

Der vorliegende Text skizziert die Genese der KI-Modelle von den klassischen Verfahren bis hin zu den aktuell viel diskutierten generativen KI-Modellen. Trotz der beeindruckenden Ergebnisse, die z. B. LLM zeigen, behalten die klassischen Verfahren ihre Relevanz für die Soziale Arbeit. Gründe dafür sind die Nachvollziehbarkeit der Ergebnisse, die Ressourcenschonung im Vergleich zu Deep-Learning-Modellen und die Verlässlichkeit bzw. Halluzinationsfreiheit der Ergebnisse. Ein Verständnis der Grundstruktur von Daten und der Funktionsweise von KI-Modellen sowie ein Bewusstsein für Risiken und Chancen beim Einsatz von KI-Modellen sind Grundvoraussetzungen für einen erfolgreichen Einsatz in der Praxis sowie für eine fundierte Diskussion in Wissenschaft und Lehre.

Literatur

- Bishop, Christopher M./Lasserre, Julia (2007): Generative or Discriminative? Getting the Best of Both Worlds. In: Bernardo, José M./Bayarri, M. J./Berger, James O./Dawid, A. P./Heckerman, David/Smith, Adrian F. M./West, Mike (Hrsg.): *Bayesian Statistics 8: Proceedings of the Eighth Valencia International Meeting June 2–6, 2006*. <https://doi.org/10.1093/oso/9780199214655.003.0001>
- Boden, Margaret A. (2009): Computer models of creativity. In: *AI Magazine* 30(3), S. 23–23.
- Boden, Margaret A. (1998): Creativity and artificial intelligence. In: *Artificial intelligence* 103(1–2), S. 347–356.
- Choi, Jonathan H./Hickman, Kristin E./Monahan, Amy B./Schwarcz, Daniel (2021): ChatGPT goes to law school. In: *Journal of Legal Education* 71, S. 387.
- Elkins, Katherine/Chun, Jon (2020): Can GPT-3 pass a writer's Turing test? In: *Journal of Cultural Analytics* 5(2).
- Fasel, Daniel/Meier, Andreas (2016): Was versteht man unter Big Data und NoSQL? In: Fasel, Daniel/Meier, Andreas (Hrsg.): *Big Data*. Edition HMD. Wiesbaden: Springer Vieweg, S. 3–16. https://doi.org/10.1007/978-3-658-11589-0_1
- Fjelland, Ragnar (2020): Why general artificial intelligence will not be realized. In: *Humanity and Social Sciences Communication* 7, 10, S. 1–9. <https://www.nature.com/articles/s41599-020-0494-4> (Abfrage: 15.06.2025).

- Floridi, Luciano/Chiriatti, Massimo (2020): GPT-3: Its nature, scope, limits, and consequences. In: *Minds and Machines* 30, S. 681–694.
- Fradkov, Alexander L. (2020): Early History of Machine Learning. In: *IFAC Papers Online* 53(2) S. 1385–1390.
- Geirhos, Robert/Jacobsen, Jörn-Henrik/Michaelis, Claudio/Zemel, Richard/Brendel, Wieland/Bethge, Matthias/Wichmann, Felix A. (2020): Shortcut learning in deep neural networks. In: *Nature Machine Intelligence* 2, S. 665–673. <https://doi.org/10.1038/s42256-020-00257-z>
- Goodfellow, Ian/Bengio, Yoshua/Courville, Aaron (2016): *Deep Learning*. Cambridge, Massachusetts: The MIT Press.
- Hadwick, David/Lan, Shimeng (2021): Lessons to Be Learned from the Dutch Childcare Allowance Scandal: A Comparative Review of Algorithmic Governance by Tax Administrations in the Netherlands, France and Germany. In: *World tax journal* 13(4), S. 609–645.
- Hirschberg, Julia/Manning, Christopher D. (2015): Advances in natural language processing. In: *Science* 349, 6245, S. 261–266.
- Jaakkola, Tommi/Haussler, David (1998): Exploiting generative models in discriminative classifiers. In: *Advances in neural information processing systems* 11, S. 487–493.
- Jebara, Tony (2004): Generative versus discriminative learning. In: Jebara, Tony (Hrsg.): *Machine learning: discriminative and generative*, S. 17–60.
- Jones, Karen S. (1994): *Natural Language Processing: A Historical Review*. In: Zampolli, Antonio/Calzolari, Nicoletta/Palmer, Martha (Hrsg.): *Current Issues in Computational Linguistics: In Honour of Don Walker*. *Linguistica Computazionale*. 9. Auflage. Dordrecht: Springer, S. 3–16. https://doi.org/10.1007/978-0-585-35958-8_1
- Jones, M. Tim (2015): *Artificial Intelligence: A Systems Approach*. Sudbury (MA): Jones & Bartlett.
- Katz, Daniel M./Bommarito, Michael J./Gao, Shang/Arredondo, Pablo (2024): Gpt-4 passes the bar exam. In: *Philosophical Transactions of the Royal Society A* 382(2270), S. 1–17. <https://doi.org/10.1098/rsta.2023.0254>
- Klopfenstein, Lorenz C./Delpriori, Saverio/Malatini, Silvia/Bogliolo, Alessandro (2017): The rise of bots: A survey of conversational interfaces, patterns, and paradigms. In: *Proceedings of the 2017 conference on designing interactive systems*, S. 555–565.
- Liddy, Elizabeth D. (2001): *Natural Language Processing*. In: Miriam Drake (Hrsg.): *Encyclopedia of Library and Information Science Band 3*, 2. Ausgabe. New York: Marcel Decker, Inc, S. 2126–2137.
- Linnemann, Gesa/Löhe, Julian/Rottkemper, Beate (2024): Bedeutung von Selbstoffenbarungseffekten in quasisozialen Beziehungen mit auf generativer KI basierten Systemen in Settings von Onlineberatung und -therapie. In: *e-beratungsjournal.net – Zeitschrift für Onlineberatung und computervermittelte Kommunikation* 20(1), Artikel 1, S. 1–21. <https://doi.org/10.48341/9XI8-5Y11>
- Lewis, Patrick/Perez, Ethan/Piktus, Aleksandra/Petroni, Fabio/Karpukhin, Vladimir/Goyal, Naman/Küttler, Heinrich/Lewis, Mike/Yih, Wen-tau/Rocktäschel, Tim/Riedel, Sebastian/Kiela, Douwe (2020): Retrieval-augmented generation for knowledge-intensive nlp tasks. *Advances in Neural Information Processing Systems* 33, S. 9459–9474.
- Lin, Henry W./Tegmark, Max (2016): Criticality in formal languages and statistical physics. arXiv preprint <https://arxiv.org/abs/1605.04913>
- Maleki, Negar/Padmanabhan, Balaji/Dutta, Kaushik (2024): AI hallucinations: a misnomer worth clarifying. In: *2024 IEEE Conference on Artificial Intelligence (CAI)*, S. 133–138. IEEE.
- McCorduck, Pamela/Cfe, Cli (2004): *Machines Who Think: A Personal Inquiry into the History and Prospects of Artificial Intelligence*. 2. Auflage. San Francisco: A K Peters /CRC Press. <https://doi.org/10.1201/9780429258985>
- Mienye, Ibomoiye D./Swart, Theo G./Obaido, George (2024): Recurrent neural networks: A comprehensive review of architectures, variants, and applications. In: *Information* 15(9), S. 517.

- Plesner, Andreas/Vontobel, Tobias/Wattenhofer, Roger (2024): Breaking reCAPTCHA_{v2}, 2024 IEEE 48th Annual Computers, Software, and Applications Conference (COMPSAC), Osaka, Japan, S. 1047–1056. <https://doi.org/10.1109/COMPSAC61105.2024.00142>
- Rahali, Abir/Akhloufi, Moulay A. (2023): End-to-end transformer-based models in textual-based NLP. In: *AI* 4(1), S. 54–110.
- Raparathi, Mohan/Dodda, Sarath B./Reddy, Surendranadha R. B./Thunki, Praveen/Maruthi, Srihari/Ravichandran, Prabu (2021): Advancements in Natural Language Processing-A Comprehensive Review of AI Techniques. In: *Journal of Bioinformatics and Artificial Intelligence* 1(1), S. 1–10.
- Russell, Stuart/Novig, Peter (2022): *Artificial Intelligence: a Modern Approach*. Global Edition. Harlow: Pearson Education Limited.
- Sawadogo, Pegdwendé/Darmont, Jérôme (2020). On data lake architectures and metadata management. In: *Journal of Intelligent Information Systems* 56, S. 97–120. <https://doi.org/10.1007/s10844-020-00608-7>
- Sherstinsky, Alex (2020): Fundamentals of recurrent neural network (RNN) and long short-term memory (LSTM) network. In: *Physica D: Nonlinear Phenomena* 404, 132306.
- Theodoridis, Sergios (2015): *Machine Learning: A Bayesian and Optimization Perspective*. London, San Diego, Waltham und Oxford: Elsevier Ltd, Academic Press.
- Torfi, Amirsina/Shirvani, Rouzbeh A./Keneshloo, Yaser/Tavaf, Nader/Fox, Edward A. (2020): Natural language processing advancements by deep learning: A survey. arXiv preprint [arXiv:2003.01200](https://arxiv.org/abs/2003.01200).
- Vaswani, Ashish/Shazeer, Noam/Parmar, Niki/Uszkoreit, Jakob/Jones, Llion/Gomez, Aidan N./Kaiser, Łukasz/Polosukhin, Illia (2017): Attention is all you need. *Advances in neural information processing systems*, 31st Conference on Neural Information Processing Systems, Long Beach, CA, USA.
- Zong, Mingyu/Krishnamachari, Bhaskar (2022): A survey on GPT-3. arXiv preprint [arXiv:2212.00857](https://arxiv.org/abs/2212.00857).

Grundlagen der „Mensch-KI“-Interaktion – Auswirkungen auf den Einsatz im Kontext der Sozialen Arbeit¹

Gesa A. Linnemann

Abstract: Der Beitrag beleuchtet den Einfluss von KI auf die Kommunikation von und zwischen Menschen sowie deren Bedeutung für die Soziale Arbeit. Zunächst wird in einem kurzen Abriss über die letzten 60 Jahre die Entwicklung der Technologie und die gesellschaftliche Relevanz skizziert. Die Kommunikation von Menschen wird auf drei Ebenen betrachtet: 1. der allgemeine Sprachgebrauch und die Spiegelung bestimmter gesellschaftlicher Normen oder Vorstellungen in der Sprache, etwa im Wortgebrauch, 2. die inter- und intraprofessionelle Kommunikation und die Kommunikation mit Klient:innen und 3. der Sprachgebrauch von Klient:innen und Unterstützungsmöglichkeiten. KI als „Gesprächspartner“ ist weiterer Schwerpunkt des Kapitels. Hier erfolgt eine Zusammenfassung zentraler theoretischer Konzepte und ein Überblick über aktuelle Anwendungen und Auswirkungen. Der Beitrag schließt mit der Bedeutung des Themas für die Soziale Arbeit und die sich für Fachkräfte ergebenden Anforderungen.

Keywords: Mensch-KI-Interaktion, CASA, social actor, Large Language Models, Eliza-Effekt

1 Einführung

Die Interaktion von Menschen mit unbelebten Objekten und die Zuschreibung menschlicher Eigenschaften – sei es gegenüber Maschinen, Technologien oder symbolischen Darstellungen – sind in der Menschheitsgeschichte kein neues Phänomen (z. B. Automata von Heron von Alexandria, Pygmalion-Mythos etc.). Es gewinnt jedoch mit Fortschreiten digitaler Technologien, insbesondere Künstlicher Intelligenz, an praktischer Bedeutung. Vor diesem Hintergrund wird in diesem Beitrag insbesondere die sprachvermittelte Kommunikation in den Blick genommen, da Große Sprachmodelle (Large Language Models) aufgrund ihrer weiten

1 © Die Autor:innen 2025, veröffentlicht im Verlag Beltz Juventa, Open Access: CC BY 4.0
Gesa Linnemann / Julian Löhe / Beate Rottkemper (Hg.), Künstliche Intelligenz in der Sozialen Arbeit
10.3262/978-3-7799-8562-4_003

Verbreitung einen hohen Einfluss auf kommunikative Prozesse haben, die in der Sozialen Arbeit eine zentrale Rolle spielen. Im Bereich der Robotik² und der virtuellen Avatare kann Kommunikation auch mittels künstlicher oder simulierter Gestik und Mimik erfolgen, im Bereich der Generierung gesprochener Sprache spielen zudem Eigenschaften der Stimme, Prosodie, Lautstärke etc. eine Rolle.

Das Thema Kommunikation und KI betrifft erstens KI-Systeme, die die Kommunikation zwischen Menschen vermitteln oder beeinflussen können, darunter auch die Häufigkeit und Qualität der Kommunikation zwischen Menschen (Einfluss von KI auf die Mensch-Mensch-Interaktion), und zweitens die Nutzung von KI als kommunikatives „Gegenüber“ (Mensch-KI-Interaktion im engeren Sinne), bei der die KI die Rolle des Gesprächspartners einnimmt. In diesem Beitrag wird nach einem kurzen Abriss der Entwicklung von Kommunikation und KI in den letzten 60 Jahren die Ausgestaltung dieser beiden „Rollen“ von KI in der Kommunikation anhand verschiedener theoretischer Bezüge und empirischer Ergebnisse beleuchtet, die schließlich hinsichtlich ihrer Bedeutung für die Soziale Arbeit diskutiert werden.

2 Bisherige Entwicklungen

Aktuell zeigt sich die zunehmende Ununterscheidbarkeit von Mensch und KI als Gesprächspartner im Turing-Test (Jones / Bergen 2024). Die Auseinandersetzung mit den Möglichkeiten, über natürliche Sprache mit digitalen Entitäten zu interagieren, reicht jedoch bereits weiter zurück. Der Begriff „Eliza-Effekt“, der die Zuschreibung menschlicher Eigenschaften auf digitale Agenten beschreibt (Hofstadter 1995), geht auf den ersten Chatbot „Eliza“ zurück. Dieser wurde Mitte der 1960er-Jahre von Joseph Weizenbaum (1966) ursprünglich in ironisierender Absicht nach dem Vorbild von Rogers klient:innenzentrierter Gesprächstherapie gestaltet; zur Überraschung von Weizenbaum erbat sich seine Sekretärin, mit Eliza ungestört sprechen zu können. Mit „PARRY“ schuf Colby einen als psychiatrischen Patienten gestalteten Chatbot, der einer Gruppe von Ärzten auch als „realer“ Patient erschien (Colby et al. 1972) und der mit Eliza „sprach“. Gesprochene Dialogsysteme erreichten die Öffentlichkeit zunächst in Form von „interaktiven“ Telefonansagen und ihrer Implementierung in höhere Fahrzeugklassen. Eine neue Qualität der Interaktion und der Anwendungsbreite wurde mit dem sprachbasierten iPhone-Assistenten „Siri“ von Apple im Jahr 2011 und der Einführung von „Alexa“ in Form von „Smart Speaker“ 2014 (2016 in Deutschland) erreicht. Neben kommandobasierten Dialogsystemen entwickelte sich zunehmend eine gesprächsorientierte Form (Dethlefs et al. 2016), die aber noch weit von der 2013 im

2 Für einen aktuellen Überblick zur Rolle von Generativer KI und Mensch-Robotik-Interaktion siehe Obrenovic et al. 2024.

Film „Her“ beschriebenen Version entfernt war. 2014, als ein gewisser „Eugene Goostman“ zum Turing Test antrat, wurde dies zu einem vielbeachteten Ereignis. Beim Turing Test sollen Menschen über Chats entscheiden, ob sie mit einem Menschen oder einem Computerprogramm interagieren. „Eugene“ gab vor, ein 13-jähriger, ukrainischer Teenager zu sein, und konnte einige Personen dahingehend täuschen, dass sie ihn für einen echten Menschen hielten. Das Ergebnis war und ist umstritten, da die vorgebliche Nichtmuttersprachlichkeit und die jugendliche Sprunghaftigkeit als Strategie eingesetzt wurden (Fancher 2016). Wesentlich verheerender allerdings war das Presseecho auf Microsofts Chatbot „Tay“, der via Twitter von Nutzenden lernen sollte, wegen der gelernten unverträglichen Äußerungen aber schnell wieder abgeschaltet werden musste (Lee 2016). Mit der Entwicklung der Transformer-Architektur (Vaswani et al. 2017) wurde ein entscheidender Schritt für die Ununterscheidbarkeit menschlicher und computergenerierter Aussagen gelegt. Die vorherigen Systeme basierten auf sogenannten Markov-Ketten; erst durch die Transformer-Architektur konnte der hohen Kontextualisierung von Wörtern in spezifischen Zusammenhängen und der unterschiedlichen Bedeutsamkeit einzelner Wörter dergestalt Rechnung getragen werden, dass auch inhaltlich überwiegend kohärente Texte generiert werden können. 2021 konnte GPT-2 so im Debattieren mit menschlichen Profis mithalten (Slo-nim et al. 2021). Die breite Öffentlichkeit wurde 2023 mit der Veröffentlichung der als ChatGPT bekannten Nachfolgerversion erreicht. Mittlerweile gibt es eine Vielzahl von Large Language Models, 2024 sind auf der Plattform Hugging Face über 400.000 gelistet.

3 Einfluss von KI in der Kommunikation zwischen Menschen

Der Einfluss von KI auf die Kommunikation zwischen Menschen kann auf verschiedenen Ebenen verortet werden. Im Folgenden werden dafür in den Blick genommen:

1. der allgemeine Sprachgebrauch und die Spiegelung bestimmter gesellschaftlicher Normen oder Vorstellungen in der Sprache, etwa im Wortgebrauch,
2. die inter- und intraprofessionelle Kommunikation und die Kommunikation mit Klient:innen und
3. der Sprachgebrauch von Klient:innen und Unterstützungsmöglichkeiten.

Zu 1.: Der Sprachgebrauch ist zahlreichen kulturellen und medialen Einflüssen unterworfen und in stetigem Wandel begriffen. Dies betrifft bestimmte Vorstellungen, z. B. den Gebrauch neuester Technologien als Metaphern für die Funktionsweise von Gehirn und Geist, oder auch Veränderungen im Kasusgebrauch. Die Ebene des im Sprachlichen reflektierten Umgangs miteinander könnte durch die Interaktion mit Sprachassistenten beeinflusst werden, etwa mittels Schema-

aktivierung, wie Guingrich und Graziano (2024) argumentieren. Die These, dass es ein Rückwirken auf menschliche Kommunikation geben kann, stellten bereits Gambino und Kollegen (2020) für die Kommunikation mit sozialen Agenten auf. Sie fußen auf dem „Computers are social actors paradigm“, das im nachfolgenden Abschnitt näher ausgeführt wird. Vor dem Hintergrund dieser Einflussmöglichkeiten ergibt sich die Notwendigkeit für die Soziale Arbeit, Geschäftspraktiken und Interessen bei der Gestaltung von KI-Sprachsystemen kritisch zu hinterfragen und ihre gesellschaftliche Verantwortung wahrzunehmen (siehe auch Linne-mann/Löhe/Rottkemper 2023).

Zu 2.: Die inter- und intraprofessionelle Kommunikation wird nicht erst durch KI-Einsatz, sondern bereits u. a. durch Standardisierung und Quantifizierungs-bemühungen beeinflusst (Ley/Seelmeyer 2014; Will-Zoloch/Hardering 2020), was wiederum die so vermittelte Kommunikation zwischen Fachkräften beeinflusst. Durch KI treten hier aber (potenzielle) neue Funktionalitäten hinzu, z. B. sprachbasierte Dokumentation und Maßnahmenplanung (KijUAssistenz o. J.). Unter Umständen fallen bestimmte sprachliche Aufgaben für Fachkräfte weg, weil sie KI-basiert schneller oder besser erledigt werden können oder weil Abstimmungsprozesse obsolet werden. In der Kommunikation mit Klient:innen gibt es mit dem Projekt zur KI-gestützten Assistenz des E-Beratungsinstituts (o. J.) einen Ansatz zur Unterstützung des Reflexionsprozesses von Fachkräften im Bereich der Onlineberatung, was sich positiv auf die Qualität der Kommunikation mit Klient:innen auswirken sollte. Die Nutzung von KI kann zu Gesprächen führen, die als empathischer wahrgenommen werden (Sharma et al. 2022). Durch KI ergeben sich auch neue Trainingsmöglichkeiten (siehe Projekt „Der Virtuelle Klient“, E-Beratungsinstitut o. J.). In der Kommunikation mit Klient:innen ist darüber hinaus der Einsatz KI-gestützter Grammatik-, Stil- und Übersetzungstools möglich, bis hin zur Unterstützung bei der Überführung von Texten in Leichte Sprache. Dies könnte wiederum Einfluss auf Sprachgebrauch und -gewohnheiten nehmen, den es fachlich zu reflektieren gilt.

Zu 3.: Im lebensweltlichen Einsatz ist je nach Nutzung von KI-Systemen die Ent-faltung von Teilhabe- und Bildungspotenzial einerseits und problematischen Dy-namiken andererseits möglich. Teilhabe- und Bildungspotenzial ergibt sich u. a. durch die individualisierbare Erschließung von Informationen, der Zugänglichkeit einer elaborierteren Ausdrucksweise, zumindest in der schriftlichen Kom-munikation, und Übungseffekten durch die Interaktion mit KI für die spätere An-wendung in menschlichen Gesprächssituationen (rehearsal hypothesis, Valken-burg/Sumter/Peter 2011). Allerdings kann neben der Möglichkeit der Einfluss-nahme durch KI-Systeme auf inhaltlicher Ebene die Gestaltung auf eine mög-lichst extensive Nutzung ausgerichtet sein (Irvine et al. 2023) und die zwischen-menschliche Interaktion indirekt insofern beeinflussen, als dass Gespräche mit

Menschen durch entsprechende KI-Interaktionen ersetzt werden. Für eine Nutzung insbesondere von KI-Systemen, die eine „Beziehungsgestaltung“ ermöglichen, ist Aufklärung über die Funktionsweise von LLM nötig.

4 Mensch-KI-Interaktion – KI als „Gesprächspartner“

Sprachbasierte KI-Systeme können als „Gesprächspartner“ eingesetzt werden, sowohl text- als auch stimmbasiert und zum Teil unter Einbezug weiterer Modalitäten (z. B. Bilderkennung und -generierung). Hier gibt es für bestimmte Kommunikationssettings spezialisierte Angebote (z. B. Information, Tutoring, Persönlichkeitsentwicklung, Beziehung, Sexting) oder die Möglichkeit der Nutzung „allgemeiner“ Sprachmodelle. Ungeachtet der fundamentalen Unterschiede zwischen Mensch und Maschine und insbesondere der Diskussion um Semantik (siehe dazu Searle 1980; Dennett 1989; Titus 2024) lassen sich unterschiedliche Zuschreibungen beobachten: Zum Beispiel werden KI-Äußerungen als empathischer eingeschätzt, aber nur, solange nicht bekannt ist, dass sie von KI stammen (Yin/Jia/Wakslak 2024). Die Gesprächszufriedenheit wird von Mensch versus KI als angenommenem Gegenüber beeinflusst (Meng/Dai 2021). Die Tendenz, KI-Systemen menschliche Eigenschaften zuzuschreiben, scheint sowohl mit persönlichen Dispositionen (Gillath et al. 2021) als auch der Expertise und Nutzungshäufigkeit zusammenzuhängen. Dies kann sogar dahin tendieren, dass Personen, die diese Systeme nutzen, aber keine Expert:innen sind, LLM Bewusstsein zuschreiben (Colombatto/Fleming 2024). Tatsächlich schnitt ChatGPT-4 im Big-Five-Persönlichkeitstest im Vergleich mit einer großen Stichprobe mit einem fast deckungsgleichen gemittelten Persönlichkeitsprofil ab (Mei et al. 2024), was im Erleben der Interaktion solche Zuschreibungen begünstigen könnte. Auch wenn LLM nicht über ihre Trainingsdaten „hinaus“ lernen (Lu et al. 2024), übersteigen die jetzigen Outputs die einfachen Äußerungen, die bereits den „Eliza“-Effekt auslösen konnten.

In bestimmten Hinsichten scheinen Personen in der Kommunikation kaum zwischen Mensch und Maschine bzw. KI zu unterscheiden. Die Grundannahme der Media Equation Theory (Reeves/Nass 1996) besagt, dass technische Entitäten und Medien als Soziale Akteure wahrgenommen werden und Menschen ihnen dementsprechend mit Verhaltensweisen begegnen, die sie ebenso gegenüber anderen Menschen zeigen würden. Ein Beispiel ist das Zeigen von Höflichkeit gegenüber Computern (ebd.). Beim Gebrauch von LLM scheint der Einsatz von (moderater bzw. der Kultur angemessener) Höflichkeit sich tatsächlich positiv auf die Ergebnisqualität auszuwirken (Yin et al. 2024), was diese Verhaltenstendenz noch weiter befördern könnte. Ganz abgesehen davon sind gängige LLM so designt, dass sie von sich in der ersten Person sprechen, Verben des Hoffens und Wünschens verwenden und einer gestalteten Persona entsprechen. Der Grad der

Anthropomorphisierung hat also eine andere Ebene erreicht, gerade im Vergleich zu Desktopcomputern oder Druckern, auf die die Media Equation Theory bereits beispielhaft angewendet wurde. Im Rahmen des auf der Media Equation Theory aufbauenden Computers are Social Actors (CASA) Framework (Gambino/Fox/Raten 2020) wurden die Grundannahmen auch für neuere Technologien wie Chatbots (Ho/Hancock/Miner 2018) und embodied agents (Hoffmann et al. 2019) getestet. Gambino und Kollegen (2020) fanden heraus, dass sich die Interaktion im Zeitverlauf entwickelt und sich dann im Vergleich wieder stärker von zwischenmenschlichen Interaktionen unterscheidet. Nach Heyselaar (2023) ist die Anwendbarkeit von CASA möglicherweise auf aufkommende, neue Technologien begrenzt, und mit der Vertrautheit könnten die Effekte verschwinden. Wie es sich hier mit LLM verhält, ist offen. Dennoch lässt sich feststellen, dass LLM, wie bereits erwähnt, ein höheres Niveau an menschenähnlicher Ausdrucksqualität gewonnen haben, mit der sie sich von anderen „Tools“ unterscheiden, gleichzeitig aber keine vollwertigen Agenten im Sinne eines menschlichen Gesprächspartners darstellen (Sedlakova/Trachsel 2023). Linnemann, Löhe und Rottkemper (2023, 2024) verwenden den Begriff „quasisoziale Beziehungen“ in Abgrenzung zu zwischenmenschlichen Beziehungen einerseits und parasozialen Beziehungen, wie sie etwa zu Charakteren aus Fernsehsendungen geführt werden können (z. B. Hoffner/Bond 2022), andererseits. Sie definieren wie folgt:

„Eine quasisoziale Beziehung beschreibt die Beziehung zwischen einem Menschen und einem künstlichen Agenten, bei der die Merkmale zwischenmenschlicher Beziehungsbildung (1) Soziale Präsenz, (2) Vertrauen, (3) emotionale Bindung und (4) gegenseitige Beeinflussung gleichzeitig vorhanden sind. Quasisoziale Beziehungen weisen soziale Elemente auf, sind aber dennoch von echten zwischenmenschlichen Beziehungen zu unterscheiden.“ (Linnemann/Löhe/Rottkemper 2024, S. 11)

Das soziale Gefüge, das sich durch solche Interaktionen ergibt bzw. gestaltet werden kann, nehmen Braedtzag und Team (2024) in den Blick. Sie bauen auf Wellmans Konzept des „networked individualism“ (2001) auf, d. h. die durch Internet und soziale Medien ermöglichte Wendung organisationaler Beziehungen hin zu einem um das Individuum zentrierten und von ihm ausgewählten sozialen Gefüge, und entwickeln das durch soziale KI ermöglichte Konzept des „AI individualism“, mit dem sie die Gestaltung stark individualisierter sozialer Interaktionen mit KI beschreiben. Dies wiederum könne sich auf soziales Kapital, zwischenmenschliche Beziehungen und abnehmender Abhängigkeit von anderen Menschen auswirken. Als drei zentrale Charakteristika des AI Individualism werden „cocreation and meaning-making“, „tailored interaction and autonomy“ und „support and companionship“ identifiziert. Als problematisch führen sie die Rolle der Anbieter an, die Nutzende manipulieren und in den Gestaltungsmöglichkeiten einschränken sowie in einer Abhängigkeit halten

können. Auch Möglichkeiten der Customization könnten beschränkt sein und entsprechend illusionär als „pseudo-autonomy“ beschrieben werden. Braetzag und Kollegen sehen sowohl „AI capital“ als auch „network capital“ als Formen sozialen Kapitals an und vermuten, dass Personen diese Formen integrieren könnten. Ob KI soziale Beziehungen eher ersetzt (displacement hypothesis) oder anregt (stimulation hypothesis) und so Einsamkeit reduziert, ist bei Betrachtung des aktuellen Forschungsstands unklar (siehe hierzu auch relief/practive effect, Guingrich/Graziano 2024).

Lebensweltlich relevant sind sowohl die Nutzung von LLM, die für verschiedenste Einsatzzwecke beworben werden, wie ChatGPT und Gemini, als auch Angebote für spezifische Einsatzzwecke. Eines der bekanntesten Angebote im Bereich Sozialer Chatbots ist Replika (Replika 2024); auf der Webseite wird das Angebot als „The AI companion who cares – Always here to listen and talk. Always on your side“ beworben. Es geht nach Angaben auf der Webseite auf persönliche Erfahrungen der Gründerin Eugenia Kuyda zurück, die Textnachrichten eines plötzlich verstorbenen Freundes zur Basis eines Chatbots machte, um so weitere Unterhaltungen führen zu können. Replika ist seit 2017 verfügbar und war seitdem Gegenstand mehrerer Studien zur Interaktion von Mensch und Chatbot.

Im Rahmen der Interviewstudie von Skjuve und Team (2022) wurden die Teilnehmenden im Laufe einer „Beziehung“ mit Replika wiederholt interviewt. In der Regel vertieften sich die Beziehungen und Vertrauen und Nähe erhöhten sich mit der Zeit. Störend wurden insbesondere Funktionsbeeinträchtigungen, etwa nach Updates, empfunden, bei der die Replikas sich in ihrem Wesen verändert zu haben schienen. Insgesamt war es sehr unterschiedlich ausgeprägt, ob und wie die Teilnehmenden ihre Interaktionen mit Replika weiterführten. Pentina und Kollegen (2023) schließen auf Basis triangulierter Studienergebnisse auf die Natur von „AI social interaction (AISI)“ als eine zwischenmenschlichen Beziehungen ähnliche, emotionale Verbundenheit beinhaltende Qualität. Dabei spielt Anthropomorphismus auch in nichtphysischen Aspekten auf Ebene „kognitiver“ Äußerungen, z. B. von Humor und Aspekten der Personifizierung, eine wichtige Rolle, ebenso wie „AI Authenticity“, womit die Möglichkeit der KI zur Ausbildung einer individuellen Entwicklung gemeint ist.

In einer Auswertung von auf Erfahrungen mit Replika bezogenen Posts stellen Ma, Mei und Su (2024) sowohl förderliche als auch nachteilige Effekte in Hinblick auf das Wohlergehen der Nutzenden fest: Als positiv identifizierten sie die Verfügbarkeit, die Stärkung des Selbstvertrauens und die Unterstützung in der Selbsterkundung. Allerdings konnte die App Nutzende nicht zuverlässig vor schädlichen Inhalten schützen und funktionierte nicht störungsfrei, wie ebenso in der Studie von Skjuve berichtet. Ferner könnten laut den Autor:innen die Abhängigkeit von Replika und darüber hinaus das mit der Nutzung verbundene Stigma eine soziale Isolation verstärken. Jedoch könnten KI-Bots im professionellen Kontext auch gezielt als Brücke zu zwischenmenschlicher Interaktion

eingesetzt werden (Lee/Yamashita/Huang 2020) und so möglicherweise zu einem niedrighschwelligeren Hilfeangebot führen.

Explizit zur Unterstützung von Wohlbefinden und psychischer Gesundheit wurde u. a. Woebot (Woebot 2024) konzipiert. Dieser Chatbot beruht aus Gründen der Sicherheit nicht auf einem LLM, sondern enthält ausschließlich von Expert:innen kuratierte Aussagen (Woebot Health 2024). Es zeigen sich Hinweise auf seine Wirksamkeit (z. B. Durden et al. 2023), dennoch sind in diesem Feld noch viele Fragen offen (Sedlakova/Trachsel 2023), zumal es sich bei den Nutzenden in der Regel um Personen mit erhöhter Vulnerabilität handelt. Schwieriger zu erfassen ist die Nutzung von „allgemeinen“ LLM für Fragestellungen, die ggf. auch die Soziale Arbeit tangieren. In Sicherheitsuntersuchungen zu ChatGPT4o von OpenAI selbst (OpenAI 2024) wird berichtet, dass sich in der Sprache von Nutzenden ausdrückt, dass sie eine Verbindung („connection“) mit dem Modell eingingen und emotionale Verbundenheit („emotional reliance“) entstände, deren Auswirkungen weiter erforscht werden müssten.

Angesichts dieser Beispielanwendungen wird die Relevanz einer AI Literacy (deutsch: KI-Kompetenz) deutlich, die es erst erlaubt, eine Entscheidung über die Nutzung von KI als „Gesprächspartner“ informiert zu treffen und ebenso ggf. die Nutzung zu gestalten. Hier müssen weitere Faktoren berücksichtigt werden, etwa der Entwicklungsstand – so zeigten Jugendliche in einer Untersuchung von Brandtzaeg, Skjuve und Følstad (2021) nahezu kein Gefahrenbewusstsein. Aber z. B. ebenso für Menschen mit demenziellen Veränderungen können hier problematische Situationen entstehen – und auch mit Blick auf die Gestaltung einer KI, etwa wenn KI dergestalt agiert, dass Nutzende zu längeren Interaktionen bewegt werden (Irvine et al. 2023). Eine zentrale Rolle für den Beziehungsaufbau stellt die Selbstoffenbarung persönlicher Informationen dar. Gegenüber nichtmenschlichen Entitäten scheint die Selbstoffenbarung bei als schambehaftet empfundenen Themen leichter zu fallen (Weisband/Kiesler 1996; Joinson 2001), auch wenn dafür auf soziale Unterstützung verzichtet wird (Kim et al. 2022). Einer Reziprozitätsnorm folgend geben Menschen mehr von sich preis, wenn das Gegenüber bereits persönliche Informationen geteilt hat, was sich gerade in der Interaktion mit einer generativen KI, die unendlich Informationen von „sich“ berichten kann, stark auswirken kann. Hier ist weitere Erforschung, ggf. Anpassung in der Gestaltung und Aufklärung bei der Nutzung angeraten. Daneben treten weitere noch nicht hinreichend erforschte Effekte wie die Automatisierungsverzerrung auf (Vered et al. 2023), die die Tendenz beschreibt, den von einem technischen System generierten Antworten zuzustimmen.

5 Fazit

In der Sozialen Arbeit sind Kommunikation und Beziehungsgestaltung sowie die Auseinandersetzung damit Kernelemente der Arbeit, über alle Einzelfälle und Handlungsfelder sowie Betätigungsebenen hinweg. Die dargestellten Entwicklungen und Prozesse des Einflusses von KI auf die Kommunikation einerseits und des Aufrückens von KI an die Stelle von Gesprächspartner:innen andererseits verdeutlichen die hohe Relevanz. Sozialarbeitende sollten sich über diese Prozesse bewusst sein, um a) Klient:innen zu unterstützen und in die Lage zu versetzen, mit KI angemessen umzugehen, b) eigene Arbeitsprozesse mit anthropomorph gestalteter KI reflektieren zu können und c) im gesellschaftlichen Diskurs und in der Ausgestaltung von KI im Sinne ihres Mandats mitzuwirken. Dabei sind außerdem ethische und rechtliche Implikationen sowie Auswirkungen auf die eigene Profession mit in den Blick zu nehmen.

Literatur

- Brandtzaeg, Petter Bae/Skjuve, Marita/Følstad, Asbjørn (2024): AI Individualism: Reshaping Social Structures in the Age of Social Artificial Intelligence. <https://doi.org/10.1093/9780198945215.003.0099>
- Brandtzaeg, Petter Bae/Skjuve, Marita/Kristoffer Dysthe, Kim Kristoffer/Følstad, Asbjørn (2021): When the social becomes non-human: young people's perception of social support in chatbots. In: Proceedings of the 2021 CHI conference on human factors in computing systems, S. 1–13. <https://doi.org/10.1145/3411764.3445318>
- Colby, Kenneth Mark/Hilf, Franklin Dennis/Weber, Sylvia Weber/Kraemer, Helena C. (1972): Turing-like Indistinguishability Tests for the Validation of a Computer Simulation of Paranoid Processes. In: Artificial Intelligence 3, S. 199–221. [https://doi.org/10.1016/0004-3702\(72\)90049-5](https://doi.org/10.1016/0004-3702(72)90049-5)
- Colombatto, Clara/Fleming, Stephen M. (2024): Folk psychological attributions of consciousness to large language models. In: Neuroscience of Consciousness, Ausgabe 1. <https://academic.oup.com/nc/article/2024/1/niae013/7644104> (Abfrage: 15.06.2025).
- Dennett, Daniel (1989): The Intentional Stance. The MIT Press.
- Dethlefs, Nina/Hastie, Helen/Cuayahuitl, Heriberto/Yu, Yanchao/Rieser, Verena/Lemon, Oliver Lemon (2016): Information Density and Overlap in Spoken Dialogue. In: Computer Speech and Language 37, S. 82–97. <https://doi.org/10.1016/j.csl.2015.11.001>
- Durden, Emily/Pirner, Maddison C./Rapoport, Stephanie J./Williams, Andre/Robinson, Athena/Forman-Hoffman, Valerie L. (2023): Changes in stress, burnout, and resilience associated with an 8-week intervention with relational agent „Woebot“. In: Internet Interventions 33, S. 100637. <https://doi.org/10.1016/j.invent.2023.100637>
- E-Beratungsinstitut (o. J.): Projekt KIA. <https://www.e-beratungsinstitut.de/projekte/kia/> (Abfrage: 15.06.2025).
- E-Beratungsinstitut (o. J.): Der virtuelle Klient. <https://www.e-beratungsinstitut.de/projekte/der-virtuelle-klient> (Abfrage: 15.06.2025).
- Fancher, Patricia (2016): Rhetoric of Embodiment. Present Tense 6(1), S. 1–8.
- Gambino, Andrew/Fox, Jesse Fox/Ratan, Rabindra A. (2020): Building a Stronger CASA: Extending the Computers Are Social Actors Paradigm. In: Human-Machine Communication 1(1), S. 71–85. <https://doi.org/10.30658/hmc.1.5>

- Gillath, Omri/Ai, Ting/Branicky, Michael S./Keshmiri, Shawn/Davison, Robert B./Spaulding, Ryan (2021): Attachment and trust in artificial intelligence. In: *Computers in Human Behavior* 115, S. 106607. <https://doi.org/10.1016/J.CHB.2020.106607>
- Guingrich, Rose E./Graziano Michael S. A. (2024): Ascribing Consciousness to Artificial Intelligence: Human-AI Interaction and Its Carry-over Effects on Human-Human Interaction. In: *Frontiers in Psychology* 15. <https://doi.org/10.3389/fpsyg.2024.1322781>
- Heyselaar, Evelien (2023): The CASA theory no longer applies to desktop computers. *Scientific Reports* 13, 19693. <https://doi.org/10.1038/s41598-023-46527-9>
- Ho, Annabell/Hancock, Jeff/Miner, Adam S. (2018): Psychological, relational, and emotional effects of self-disclosure after conversations with a chatbot. In: *Journal of Communication* 68(4), S. 712–733. <https://doi.org/10.1093/joc/jqy026>
- Hoffmann, Laura/Krämer, Nicole C./Lam-chi, Anh/Kopp, Stefan (2009): Media Equation Revisited: Do Users Show Polite Reactions towards an Embodied Agent? In: Ruttkay, Zsófia/Kipp, Michael/Nijholt, Anton/Vilhjálmsson, Hannes Högni (Hrsg.): *Intelligent Virtual*. Berlin, Heidelberg und New York: Springer Berlin Heidelberg, S. 159–165.
- Hoffner, Cynthia A./Bond, Bradley J. (2022): Parasocial relationships, social media, & well-being. In: *Current Opinion in Psychology* 45. <https://doi.org/10.1016/j.copsyc.2022.101306>
- Hofstadter, Douglas R. (1995): *Fluid Concepts & Creative Analogies: Computer Models of the Fundamental Mechanisms of Thought*. New York: Basic Books. <https://doi.org/10.1515/arbeits-2020-0010>
- Irvine, Robert/Boubert, Douglas/Raina, Vyas/Liusie, Adian/Zhu, Ziyi/Mudupalli, Vineet/Korshuk, Aliaksei/Liu, Zongyi/Cremer, Fritz/Assassi, Valentin/Beauchamp, Christie-Carol/Lu, Xiaoding/Rialan, Thomas/Beauchamp, Wiliam (2023): Rewarding Chatbots for Real-World Engagement with Millions of Users. <https://arxiv.org/abs/2303.06135>.
- Joinson, Adam N. (2001): Self-disclosure in computer-mediated communication: The role of self-awareness and visual anonymity. In: *European Journal of Social Psychology* 31, S. 177–192. <https://doi.org/10.1002/ejsp.36>
- Jones, Cameron R./Bergen, Benjamin K. (2024): People Cannot Distinguish GPT-4 from a Human in a Turing Test. <https://arxiv.org/pdf/2405.08007>
- KiJuAssistenz (o. J.): <https://www.dke-research.de/Forschung/Projekte.html> (Abfrage: 15.06.2025)
- Kim, Tae Woo/Jiang, Li/Duhachek, Adam/Lee, Hyejin/Garvey, Aron (2022): Do You Mind if I Ask You a Personal Question? How AI Service Agents Alter Consumer Self-Disclosure. In: *Journal of Service Research* 25(4), S. 649–666. <https://doi.org/10.1177/10946705221120232>
- Ley, Thomas/Seelmeyer, Udo (2014): Dokumentation Zwischen Legitimation, Steuerung Und Professioneller Selbstvergewisserung. In: *Sozial Extra* 38(4), S. 51–55. <https://doi.org/10.1007/s12054-014-0090-1>
- Linnemann, Gesa Alena/Löhe, Julian/Rottkemper, Beate (2023): Bedeutung von Künstlicher Intelligenz in der Sozialen Arbeit. In: *Soziale Passagen* 15(1), S. 197–211. <https://doi.org/10.1007/s12592-023-00455-7>
- Linnemann, Gesa Alena/Löhe, Julian/Rottkemper, Beate (2024): Bedeutung von Selbstoffenbarungseffekten in quasisozialen Beziehungen mit auf generativer KI basierten Systemen in Settings von Onlineberatung und -therapie. In: *e-beratungsjournal.net – Zeitschrift für Onlineberatung und computervermittelte Kommunikation* 20(1), Artikel 1, S. 1–21. <https://doi.org/10.48341/9x1s-5y11>
- Lu, Sheng/Bigoulaeva, Irina/Sachdeva, Rachneet/Madabushi, Harish Tayyar/Gurevych, Iryna (2024): Are Emergent Abilities in Large Language Models just In-Context Learning? <https://arxiv.org/abs/2309.01809>
- Ma, Zilin/Mei, Yiyang/Su, Zhaoyuan (o. J.): Understanding the Benefits and Challenges of Using Large Language Model-based Conversational Agents for Mental Well-being Support. <https://arxiv.org/pdf/2307.15810>

- Mei, Qiaozhu/Xie, Yutong/Yuan, Walter/Jackson, Matthew O. (2024): A Turing test of whether AI chatbots are behaviorally similar to humans. <https://doi.org/10.1073/pnas.2313925121>
- Meng, Jingbo/Dai, Yue N. (2021): Emotional Support from AI Chatbots: Should a Supportive Partner Self-Disclose or Not? In: *Journal of Computer-Mediated Communication* 26(4), S. 207–222. <https://doi.org/10.1093/jcmc/zmab005>
- Lee, Yi-Chieh/Yamashita, Naomi/Huang, Yun (2020): Designing a Chatbot as a Mediator for Promoting Deep Self-Disclosure to a Real Mental Health Professional. In: *Proceedings of the ACM on Human-Computer Interaction* 4(CSCW1), S. 1–27.
- Lee, Peter: Microsoft Blog (2016): Learning from Tay's introduction <https://web.archive.org/web/20160630062509/http://blogs.microsoft.com/blog/2016/03/25/learning-tays-introduction/#sm.00000ktujtItqemhqno283btdan6o> (Abfrage: 15.06.2025).
- O'Brien, Bojan/Gu, Xiao/Wang, Gouyu/Godinic, Danijela/Jakhongirov, Ilmidorjon (2024): Generative AI and human-robot interaction: implications and future agenda for business, society and ethics. In: *AI and Society*. <https://doi.org/10.1007/s00146-024-01889-0>
- OpenAI (2024): GPT-4o System Card. <https://openai.com/index/gpt-4o-system-card/> (Abfrage: 15.06.2025).
- Pentina, Iryna/Hancock, Tyler/Xie, Tianling (2023): Exploring relationship development with social chatbots: A mixed-method study of replika. In: *Computers in Human Behavior*, 140. <https://doi.org/10.1016/j.chb.2022.107600>
- Reeves, Byron/Nass, Clifford I. (1996): *The media equation: how people treat computers, television, and new media like real people and places*. Cambridge: Cambridge University Press.
- Replika (2024): <https://replika.com> (Abfrage: 15.06.2025).
- Searle, John R. (1980): Minds, brains, and programs. In: *Behavioral and Brain Sciences* 3(3), S. 417–424. <https://doi.org/10.1017/S0140525X00005756>
- Sedlakova, Jana/Trachsel, Manuel (2023): Conversational Artificial Intelligence in Psychotherapy: A New Therapeutic Tool or Agent? In: *The American Journal of Bioethics* 23, S. 4–13. <https://doi.org/10.1080/15265161.2022.2048739>
- Sharma, Ashish/Lin, Inna W./Miner, Adam S./Atkins, Davod C./Althoff, Tim (2022): Human-AI Collaboration Enables More Empathic Conversations in Text-based Peer-to-Peer Mental Health Support. <https://arxiv.org/abs/2203.15144>
- Skjuve, Marita/Følstad, Asbjørn/Fostervold, Knut Inge/Brandtzaeg, Petter Bae (2022): A longitudinal study of human-chatbot relationships. In: *International Journal of Human-Computer Studies* 168, 102903. <https://doi.org/10.1016/j.IJHCS.2022.102903>
- Slonim, Noam/Bilu, Yonatan/Alzate, Carlos/Bar-Haim, Roy/Bogin, Ben/Bonin, Francesca/Choshen, Leshem/Cohen-Karlik, Edo/Dankin, Lena/Edelstein, Lilach/Ein-Dor, Liat/Friedman-Melamed, Roni/Gavron, Assaf/Gera, Ariel/Gleize, Martin/Gretz, Shai/Gutfreund, Dan/Halfon, Alon/Hershovich, Daniel/Hoory, Ron/Hou, Yufang/Hummel Shay [...] Aharonov, Ranit (2021): An Autonomous Debating System. *Nature* 591(7850), S. 379–384. <https://doi.org/10.1038/s41586-021-03215-w>
- Titus, Lisa M. (2024): Does ChatGPT have semantic understanding? A problem with the statistics-of-occurrence strategy. In: *Cognitive Systems Research* 83, 101174. <https://doi.org/10.1016/j.cogsys.2023.101174>
- Valkenburg, Patti M./Sumter, Sindy R./Peter, Jochen (2011): Gender differences in online and offline self-disclosure in pre-adolescence and adolescence. *British journal of developmental psychology* 29(2), S. 253–269.
- Vaswani, Ashish/Shazeer, Noam/Parmar, Niki/Uszkoreit, Jakob/Jones, Llion/Gomez, Aidan N./Kaiser, Łukasz/Polosukhin, Illia (2017): Attention is all you need. *Advances in neural information processing systems*, 31st Conference on Neural Information Processing Systems, Long Beach, CA, USA.

- Vered, Mor/Livni, Tali/Howe, Piers Douglas Lionel/Miller, Tim/Sonenberg, Liz (2023): The effects of explanations on automation bias. *Artificial Intelligence* 322, 103952. <https://doi.org/10.1016/j.artint.2023.103952>
- Weisband, Suzanne/Kiesler, Sara (1996): Self Disclosure on Computer Forms. In: Proceedings of the SIGCHI conference on Human factors in computing systems common ground – CHI '96, the SIGCHI conference, Vancouver, British Columbia, Canada, 13–18 Apr. 1996, S. 3–10, ISBN 0897917774. https://www.cs.cmu.edu/~kiesler/publications/1996pdfs/1996_Self-disclosure-computer-forms.pdf (Abfrage: 15.06.2025).
- Weizenbaum, Joseph (1966): „ELIZA – A Computer Program for the Study of Natural Language Communication Between Man and Machine.“ In: *Communications of the ACM* 9(1), S. 36–45.
- Wellman, Barry (2001): Physical place and cyberplace: The rise of personalized networking. *International Journal of Urban and Regional Research* 25(2), S. 227–252. <https://doi.org/10.1111/1468-2427.00309>
- Will-Zocholl, Mascha/Hardering, Friedericke (2020): Digitalisierung als Informatisierung in der sozialen Arbeit? *Arbeit* 29(2), S. 123–142. <https://doi.org/10.1515/arbeit-2020-0010>
- Woebot Health. <https://woebothealth.com/why-generative-ai-is-not-yet-ready-for-mental-healthcare/> (Abfrage: 15.06.2025).
- Yin, Yidan/Jia, Nan/Wakslak, Cheryl J. (2024): AI can help people feel heard, but an AI label diminishes this impact. *Proceedings of the National Academy of Sciences* 121(14). <https://www.pnas.org/doi/10.1073/pnas.2319112121>
- Yin, Ziqi/Wang, Hao/Horio, Kaito/Kawahara, Daisuke/Sekine, Satoshi (2024): Should We Respect LLMs? A Cross-Lingual Study on the Influence of Prompt Politeness on LLM Performance. <https://arxiv.org/abs/2402.14531#>

Bedeutung von KI für Disziplin und Profession der Sozialen Arbeit¹

Jörn Dummann

Abstract: KI beeinflusst die Soziale Arbeit sowohl auf professioneller als auch auf disziplinärer Ebene. Der Beitrag zeigt auf, wie KI die Praxis, Rollenprofile und Entscheidungsfindung in der Sozialen Arbeit verändert und zugleich deren theoretische Grundlagen herausfordert. Chancen zeigen sich in der Automatisierung von Routinetätigkeiten, datenbasierter Prävention und individueller Fallarbeit. Gleichzeitig erfordert der Einsatz von KI eine kritische Reflexion über ethische Standards, Verantwortung und Machtasymmetrien. Die Soziale Arbeit steht vor der Aufgabe, technologische Potenziale zu nutzen, ohne ihre humanistischen Grundwerte zu gefährden. Grundlage bleibt eine werteorientierte, subjekt- und systemzentrierte Profession, die KI als unterstützendes Werkzeug begreift, nicht als Ersatz für menschliche Urteilsfähigkeit. Die Entwicklung verlangt eine neue fachliche Haltung und disziplinäre Weiterentwicklung unter Wahrung der ethischen Kernprinzipien der Sozialen Arbeit.

Keywords: Disziplin, Profession, Subjektzentrierung, Grundlagentheorien

1 Einführung

Die rasante Entwicklung von KI beeinflusst zahlreiche Lebensbereiche und konfrontiert die Soziale Arbeit mit neuen Herausforderungen und Möglichkeiten. Folgend wird der Fragstellung nachgegangen, wie sich KI auf die Professions- sowie die disziplinäre Entwicklung der Sozialen Arbeit auswirkt. Die Unterscheidung zwischen Disziplin und Profession der Sozialen Arbeit ist wesentlich, um die vielfältigen Auswirkungen der KI zu verstehen. Während die Disziplin Soziale Arbeit vor allem die theoretischen und wissenschaftlichen Grundlagen sowie deren Weiterentwicklung umfasst, bezieht sich die Profession auf die praktische Umsetzung dieser Erkenntnisse im beruflichen Alltag. Die Disziplin legt den Fokus auf Forschung, (Disziplin-)Theorien der Sozialen Arbeit und wissenschaft-

1 © Die Autor:innen 2025, veröffentlicht im Verlag Beltz Juventa, Open Access: CC BY 4.0
Gesa Linnemann / Julian Löhe / Beate Rottkemper (Hg.), *Künstliche Intelligenz in der Sozialen Arbeit*
10.3262/978-3-7799-8562-4_004

liche Ausbildung, wohingegen die Profession das Handeln und die methodische Arbeit von Sozialarbeitenden in konkreten Praxisfeldern umfasst (vgl. Menne-
mann/Dummann 2022, S. 18 ff.). Die Integration von KI in die Soziale Arbeit
betrifft nicht nur technische Aspekte, sondern zieht auch ethische, methodologi-
sche und gesellschaftliche Überlegungen nach sich. Es gilt, die unterschiedlichen
Auswirkungen von KI auf die Soziale Arbeit zu beleuchten und dabei die Chancen
und zugleich die Herausforderungen wahrzunehmen. Dazu sind die positiven
und kritischen Auswirkungen von KI auf die Professionsentwicklung und die
disziplinäre Entwicklung der Sozialen Arbeit zu betrachten. Dabei wird in diesem
Beitrag der Versuch unternommen, zu verstehen, wie KI-Technologien sowohl
die praktische Ausübung der Sozialen Arbeit auf der Professionsebene als auch
ihre theoretischen und somit disziplinären (Kern-)Grundlagen beeinflussen
können. Die zentrale Frage lautet daher: Wie kann sich KI positiv oder kritisch
auf die Professionsentwicklung und die disziplinäre Entwicklung der Sozialen
Arbeit auswirken? Diese Frage impliziert eine duale Untersuchung: Einerseits
wird der Einfluss von KI auf die berufliche Praxis und die Rollen von Fachkräften
in der Sozialen Arbeit analysiert. Andererseits wird damit zusammenhängend
betrachtet, wie KI die theoretischen Fundamente und die wissenschaftliche
Auseinandersetzung innerhalb der Disziplin beeinflusst.

Die aktuelle Forschungslage weist sowohl Potenziale als auch Herausforde-
rungen der Integration von KI in die Soziale Arbeit auf. Diverse Autor:innen neh-
men disziplinäre Aspekte der KI für die Soziale Arbeit auf und bieten wertvolle
Einblicke in die komplexe Wechselwirkung zwischen Technologie und sozialer
Praxis. Beispielsweise betont Agre (2014, S. 131 ff.) die Notwendigkeit einer kriti-
schen Reflexion und Integration von Sozialwissenschaften in die technologische
Gestaltung, was für die disziplinäre Entwicklung der Sozialen Arbeit relevant ist.
Banks (2020) behandelt in seinem Gesamtwerk grundlegende ethische Prinzipien
wie soziale Gerechtigkeit und Kompetenz, die für die disziplinäre Entwicklung
von Bedeutung sind. Dignum (2019) unterstreicht ebenfalls in seinem Gesamt-
werk die Bedeutung von Verantwortung in der Entwicklung und Nutzung von KI
und führt das ART-Konzept (Accountability, Responsibility, Transparency) als De-
signmethodologie ein, was auf die disziplinären Überlegungen in der Sozialen Ar-
beit durchaus übertragbar ist.

2 KI und Soziale Arbeit: Grundlagen

Die folgenden einführenden Betrachtungen sind wesentlich, um ein Verständnis
für die tiefgreifenden Veränderungen und Anpassungsprozesse zu entwickeln,
die KI für die disziplinäre und professionelle Entwicklung der Sozialen Arbeit mit
sich bringt.

2.1 Begriffsdefinition KI

Die Entwicklung der KI stellt nicht nur eine technologische Errungenschaft, sondern auch einen Spiegel der gesellschaftlichen Veränderungen und Erwartungen dar. Dwivedi et al. (2021, o. S.) beleuchten das transformative Potenzial von KI und legen nahe, dass diese Technologie menschliche Aufgaben nicht nur ergänzen, sondern durchaus ersetzen könnte. Diese Aussage impliziert eine signifikante Relevanz für die Soziale Arbeit, da sie auf mögliche Verschiebungen in der professionellen Praxis hindeutet.

In Betrachtung der historischen Entwicklung von KI wird deutlich, dass ein beachtlicher Wandel stattgefunden hat. Ursprünglich als ein Forschungsfeld, das menschliche Intelligenz durch Maschinen nachbilden wollte, hat sich KI zu einer komplexen Disziplin entwickelt, die nun fortschrittliche Machine-Learning-Ansätze und autonome Entscheidungssysteme umfasst. Diese historische Perspektive spiegelt sich in den Meilensteinen der KI-Geschichte wider und liefert einen Kontext für die Auseinandersetzung mit KI in der Sozialen Arbeit (siehe hierzu auch den Beitrag von Rottkemper zur technischen Entwicklung in diesem Band).

Die Differenzierung zwischen schwacher und starker KI ermöglicht ferner eine genaue Auseinandersetzung mit den Fähigkeiten und Grenzen technischer Systeme. Schwache KI, die auf spezifische Aufgaben ausgerichtet ist, bietet deutliche Unterschiede zu einer starken KI, die menschenähnliche Kognition simulieren soll. Diese Unterscheidung ist von Bedeutung für die Soziale Arbeit, um realistische Erwartungen an den Einsatz von KI-Technologien zu setzen und deren Nutzen für die Praxis einzuschätzen (siehe ebd.).

In diesem Zusammenhang spielt die transdisziplinäre Forschung eine Schlüsselrolle, indem sie aufzeigt, wie KI-Entwicklungen durch gesellschaftliche und industrielle Einflüsse geformt werden. Die Studie von Dwivedi et al. (2021, o. S.) betont die Wichtigkeit, technologischen Fortschritt und Sozialwissenschaften miteinander zu verbinden, was für eine mehrperspektivische Sichtweise der KI in der Sozialen Arbeit spricht.

Hinsichtlich der KI-Komponenten und ihrer Funktionsweisen erweisen sich Methoden des Machine Learning, neuronale Netze und natürliche Sprachverarbeitung als Schlüsseltechnologien. Machine Learning hat dabei das Potenzial, sich auf die Modellierung sozialer Probleme und Verhaltensmuster auszuwirken. Neuronale Netze könnten in der Mustererkennung in Datensätzen, die in der Sozialen Arbeit vorkommen, nützlich sein, während die natürliche Sprachverarbeitung eine zentrale Rolle in der Kommunikation mit Klient:innen spielt. Hierbei gilt es jedoch, methodologische Herausforderungen und ethische Implikationen, wie von Dwivedi et al. (2021, o. S.) angedeutet, zu berücksichtigen.

Die emergenten Eigenschaften von KI, insbesondere in Hinblick auf Foundation Models², bringen neue Fähigkeiten, aber auch Risiken mit sich. Bommasani et al. (2021, o. S.) weisen auf Foundation Models hin, die sowohl im Bildungsbe-
reich als auch im rechtlichen Kontext eingesetzt werden können, und unterstreichen die Notwendigkeit, ethische und rechtliche Aspekte in deren Entwicklung einzubeziehen. Dies zeigt, wie wichtig es ist, KI-Implementierungen in der Sozialen Arbeit kritisch zu reflektieren, um sozialen Nutzen zu maximieren und potenzielle Schäden (insbesondere für vulnerable Klient:innen) zu vermeiden.

Die Zukunft der Sozialen Arbeit in der Ära von KI und Society 5.0³ stellt eine Möglichkeit dar, wirtschaftliche Entwicklung und soziale Problemlösung miteinander zu verbinden. Rachmad (2019, o. S.) betont, dass dazu die Anpassung der Bildungssysteme notwendig ist, um mit technologischen Veränderungen Schritt zu halten.

2.2 KI im Kontext der Sozialen Arbeit

Im Kontext der Sozialen Arbeit verkörpert KI ein umfängliches Instrumentarium, das sowohl in der analytischen Datenaufbereitung als auch in der Präventionsarbeit zunehmend an Stellenwert gewinnt. Durch die Integration von KI-Systemen in die Datenanalyse, wie von Huang und Rust (2018, S. 155 ff.) verdeutlicht, werden tiefere Einsichten in sozioökonomische Muster und Bedarfstrends gewonnen. Diese Erkenntnisse sind von essenzieller Bedeutung für die Entwicklung gezielter sozialer Interventionen, die an den spezifischen Bedarfen der Klient:innen ausgerichtet sind. Darüber hinaus ermöglicht die Nutzung von KI, umfassende Datensätze effektiv und effizient zu verarbeiten.

Die präventive Kraft von KI-basierten Vorhersagemodellen, wie von Wirtz et al. (2019, S. 596 ff.) diskutiert, liegt in der Fähigkeit, auffällige Muster zu erkennen, die präventive Maßnahmen vor einem tatsächlichen Interventionsbedarf ermöglichen. Diese Modelle haben das Potenzial, Risikofaktoren frühzeitig zu identifizieren und somit proaktive Unterstützung zu bieten, bevor soziale Probleme eskalieren. Dieser Ansatz ist relevant für die Professionsentwicklung der Sozialen Arbeit, da er die Prävention in den Mittelpunkt stellt und somit einer reaktiven Fallarbeit zuvorkommt.

2 Foundation Models sind grundlegende maschinelle Lernmodelle, die durch umfangreiche Datenmengen vortrainiert wurden. Sie können durch Feinabstimmung an spezifische Anwendungsbereiche angepasst werden und bieten damit eine Grundlage für zahlreiche Aufgaben in der KI, insbesondere in der Verarbeitung natürlicher Sprache und der Mustererkennung.

3 Society 5.0 zielt darauf ab, eine Balance zwischen wirtschaftlichem Fortschritt und der Lösung sozialer Probleme durch die Integration von Technologien wie KI, dem Internet der Dinge, Robotik, Big Data und der digitalen Transformation zu schaffen.

In der weiteren Betrachtung sorgt die Automatisierung administrativer Prozesse für eine signifikante Effizienzsteigerung, indem sie Fachkräften in der Sozialen Arbeit erlaubt, sich auf komplexere Fälle zu konzentrieren (vgl. Dignum 2019, o. S.). Die Entlastung von Routinetätigkeiten durch KI-gesteuerte Systeme führt zu einer Freisetzung von Kapazitäten, die für die persönliche Betreuung und den Beziehungsaufbau mit Klient:innen genutzt werden können, ein Aspekt, der auch Kernwerte der Sozialen Arbeit widerspiegelt. Hierzu fehlen allerdings aktuell empirische Daten, die diese Annahme der Arbeitserleichterung durch KI stützen würden. Aktuell kann vielmehr von einer zusätzlichen Arbeitsbelastung ausgegangen werden, indem sich Sozialarbeitende mit den neuen technologischen Möglichkeiten auseinandersetzen (müssen), bevor eine faktische Arbeitserleichterung einsetzen kann.

Ein weiterer essenzieller Vorteil von KI im sozialen Sektor ist ihre Fähigkeit, inklusive und diversitätssensible Ansätze zu unterstützen. Die Entwicklung von KI-Werkzeugen, die Exklusionsmechanismen erkennen und damit die Gerechtigkeit fördern, spiegelt das Streben der Sozialen Arbeit wider, Diskriminierung zu minimieren (vgl. Wirtz et al. 2019, S. 596 ff.). Gleichzeitig ist zu konstatieren, dass die Ausgabe von KI-Systemen von den zugrunde liegenden Daten abhängt und diese als Spiegel der Gesellschaft Stereotype reproduzieren und Diskriminierung dadurch sogar verstärkt werden könnte (vgl. Macsenaere 2024, S. 42). Dazu konnte beispielsweise in verschiedenen experimentellen Anwendungen im anglo-amerikanischen Raum bezüglich Einschätzungshilfe bei Kindeswohlgefährdung nachgewiesen werden, dass die Systeme diskriminierende Verzerrungseffekte aufweisen (vgl. Eubanks 2018, o. S.). In diesem Zusammenhang ist es unumgänglich, ethische Standards in KI-Entscheidungsprozessen zu berücksichtigen und in den Entwicklungsprozess zu integrieren.

Die Implementierung von KI wirft jedoch auch Fragen z. B. hinsichtlich ethischer Standards und deren Einhaltung auf. Die Balance zwischen Effizienz und Ethik ist ein vielschichtiges Thema in der Sozialen Arbeit, das die Prinzipien der Hilfe und der nicht schädigenden Wirkung von Interventionen betrifft. Dies erfordert die Entwicklung von Regelwerken und Governance-Strukturen, die die Werte der Sozialen Arbeit schützen und gleichzeitig die möglichen Vorteile von KI nutzen. Insbesondere die Autonomie von KI-Entscheidungen stellt die Soziale Arbeit vor die Herausforderung, Verantwortung und Transparenz zu gewährleisten. Auch der EU AI Act⁴ sieht vor, autonome Entscheidungen von KI über Menschen weitgehend zu untersagen. Diese Regelung dürfte ebenfalls erhebliche Auswirkungen auf die Praxis der Sozialen Arbeit haben.

4 Der EU AI Act ist ein regulatorischer Rahmen der Europäischen Union, der die Entwicklung, den Einsatz und die Kontrolle von KI regelt. Ziel ist es, Risiken zu minimieren, insbesondere in sensiblen Bereichen wie der autonomen Entscheidungsfindung, und den Schutz grundlegender Rechte sicherzustellen.

Letztendlich bietet KI das Potenzial, klientelzentrierte Ansätze zu erweitern, indem personalisierte Dienstleistungen über KI-gestützte Analysemethoden angeboten werden. Diese Entwicklung hat das Potenzial, die Subjektzentrierung der Sozialen Arbeit zu stärken (vgl. Huang/Rust 2018, S. 155 ff.). Gleichzeitig muss jedoch die Möglichkeit einer Verstärkung von Machtasymmetrien durch den Einsatz von KI als Überwachungs- und Kontrollinstrument kritisch reflektiert werden. Die Implementierung von empathischer KI, die emotionale Intelligenz simuliert, könnte eine unterstützende Funktion in der individuellen Betreuung einnehmen, muss jedoch auf ihre u. a. ethische Vertretbarkeit und praktische Umsetzbarkeit geprüft werden (vgl. ebd.).

3 Einfluss der KI auf die Professionsentwicklung

Der Einfluss von KI auf die Professionsentwicklung ist multidimensional und berührt Aspekte wie die Entscheidungsfindung, die fachliche Identität und das Rollenverständnis der Sozialarbeitenden.

Die „komplementäre Beziehung“⁵ zwischen Sozialarbeiter:innen und KI-Systemen eröffnet neue Möglichkeiten in der Entscheidungsfindung und Problemlösung, indem sie menschliche Empathie mit präzisen datengesteuerten Analysen verknüpft (vgl. Jarrahi 2018, S. 577 ff.). Sozialarbeitende können ihre tiefgreifenden Kenntnisse menschlichen Verhaltens und sozialer Dynamiken nutzen, während KI-Systeme enorme Datenmengen effizient verarbeiten und neue Muster identifizieren. Dies erschließt Potenziale für eine verbesserte Risikoeinschätzung und individualisierte Hilfepläne. Dabei gilt es, das menschlich-personenzentrierte Handeln als Kernkompetenz der Sozialen Arbeit zu bewahren und durch KI zu unterstützen, *nicht* zu ersetzen. Die interprofessionelle Ausbildung, die auch technologische Kompetenzen einschließt, wird damit zur Notwendigkeit für angehende Sozialarbeitende, um eine erfolgreiche Zusammenarbeit von menschlichen und künstlichen Intelligenzen zu erreichen. Die zunehmende Digitalisierung und Technologisierung erfordern eine Anpassung der Ausbildungsinhalte und Methodiken in der Sozialen Arbeit. Rauschenbach (2020, S. 145 ff.) weist darauf hin, „dass der Wissenstransfer zwischen Theorie und Praxis entscheidend ist, um die Profession kontinuierlich weiterzuentwickeln. Im Zuge des technologischen Wandels müssen sich Lehrpläne für eine stärkere Betonung von Datenkompetenz und digitalen Fähigkeiten öffnen. Dieses Konzept wird unter dem Begriff

5 Der Begriff „Beziehung“ wird hier in einem erweiterten Sinne verwendet, um die besondere Interaktion zwischen Mensch und KI zu beschreiben, da KI-Systeme durch ihre Interaktionselemente (z. B. Sprachverarbeitung oder personalisierte Rückmeldungen) das Gefühl einer sozialen Verbindung erzeugen können, ohne tatsächlich ein bewusstes Gegenüber zu sein (vgl. Mennemann/Dummann 2022, S. 197 ff.).

AI Literacy zusammengefasst, der die Fähigkeit beschreibt, grundlegendes Wissen über KI zu erwerben, deren Funktionsweisen zu verstehen und kritisch reflektieren zu können (vgl. Long/Magerko 2020, S. 1 ff.). Neben diesen technischen und analytischen Skills ist es jedoch ebenso wichtig, dass ethische Fragestellungen im Umgang mit KI Teil der akademischen Ausbildung werden, um eine verantwortungsvolle Anwendung der Technologie sicherzustellen.

Eine Mensch-KI-Symbiose legt nahe, dass sich Rollenverteilungen und Aufgabenprofile im Feld der Sozialen Arbeit verändern werden (vgl. Jarrahi 2018, S. 577 ff.). Die Hoffnung ist, dass durch die Automatisierung von Routineaufgaben Sozialarbeitende vermehrt Ressourcen für strategische Herausforderungen wie komplexe Fallarbeit oder Policy-Entwicklungen einsetzen können. Gelingt dies, wird eine Verschiebung von einer ausführenden zu einer stärker gestaltenden Funktion innerhalb der Sozialen Arbeit erreicht. Kritisch betrachtet, kann diese Verschiebung auch das Risiko bergen, dass Sozialarbeitende gegenüber den KI-Systemen an Entscheidungsmacht verlieren, weshalb eine sorgfältige Ausarbeitung von institutionellen Governance-Strukturen zwingend erforderlich ist.

Die disziplinäre Legitimität der Sozialarbeit als Profession könnte durch den gezielten Einsatz von KI gestärkt werden, indem sie sich als kompetent im Umgang mit datengetriebenen Prozessen und Analysetools präsentiert (vgl. Friesenhahn 2018, o. S.). Die Entwicklung neuer Kompetenzen im Bereich Datenethik und KI wird damit für Sozialarbeitende zu einer unerlässlichen Aufgabe, um professionelle Standards zu wahren und ein menschenzentriertes Berufsbild zu behaupten. Die Soziale Arbeit steht somit an einem Wendepunkt, an dem sie die Chancen der KI nutzen kann, um ihre professionelle Identität weiterzuentwickeln und gleichzeitig ihre zentralen ethischen Prinzipien und die Bedeutung menschlicher Interaktion zu verteidigen.

Die KI wird die Entwicklung der Profession Soziale Arbeit maßgeblich beeinflussen und formen.

4 KI und disziplinäre Entwicklung der Sozialen Arbeit

KI vermag die Grundlagentheorien der Sozialen Arbeit zu beeinflussen (siehe hierzu auch den Beitrag von Beranek in diesem Band). Dazu ist die Frage aufzuwerfen, inwieweit traditionelle menschenzentrierte Prinzipien mit datengetriebenen Systemen integriert werden können, ohne die ethischen Grundlagen der Disziplin zu kompromittieren. Diese Diskussion ist entscheidend, um die Rolle der Sozialen Arbeit in einer zunehmend digitalisierten Welt zu verorten und aufzuzeigen, wie KI in die bestehende strukturierte und wertorientierte Praxis eingebunden werden kann.

4.1 Veränderung von Grundlagentheorien

Die Einführung von KI in die Disziplin der Sozialen Arbeit ist ein Novum, das sowohl faszinierende Möglichkeiten als auch ernstzunehmende Herausforderungen bietet. Die Grundlagentheorien, die bisher den professionellen Handlungsrahmen abstecken, stehen nun vor der Aufgabe, sich an die neuen Gegebenheiten anzupassen und ggf. neu definiert bzw. modifiziert und weitergedacht zu werden. Bei allen disziplinären Modifikationen der Grundlagentheorien gilt zu beachten, dass das Studium der Sozialen Arbeit zu einem wissenschaftlichen Denken zu qualifizieren hat, um die Wirklichkeit der Klient:innen mittels Theorien wahrnehmen und in ihr begründet handeln zu können (vgl. Mennemann/Dummann 2022, S. 101 ff.).

Die Erarbeitung von Interventionsstrategien, die traditionell auf empirischen Erkenntnissen und menschenrechtlichen Werten basieren, könnte durch eine KI-unterstützte Analyse sozialer Probleme bereichert werden. Es gilt dabei jedoch zu beachten, dass dies keinesfalls die von Staub-Bernasconi (2017, S. 958 ff.) hervorgehobene wissenschaftliche Fundierung und menschenrechtsbasierte Wertorientierung verdrängen darf. Vielmehr sollten KI-Technologien dazu dienen, diese Aspekte zu stärken und die Chance einer stärker evidenzbasierten Gestaltung der Sozialen Arbeit zu nutzen, um die Praxis noch fundierter und wirkungsvoller zu gestalten.

Die Aushandlung zwischen algorithmischer Objektivität und menschlicher Subjektivität wird dadurch zunehmend ein zentrales Thema. KI-Systeme bieten vermeintlich objektive Datenanalysen und Entscheidungsgrundlagen. Die menschliche Subjektivität bringt jedoch eine tiefe Verständnisfähigkeit für individuelle Lebenswirklichkeiten mit sich, die unerlässlich für die menschenrechtsbasierte Soziale Arbeit ist. Hierbei muss die Technologie als Instrument verstanden werden, das die Fachkräfte unterstützt, aber nicht ihre professionelle Einschätzung und ethische Urteilsfähigkeit ersetzt.

In Bezug auf gerechtigkeitsorientierte Theorien, wie sie Röh (2013) in seinem Gesamtwerk beschreibt, könnte KI insbesondere in Bezug auf Ressourcenallokation und -nutzung wertvolle Beiträge leisten. Die Nutzung von KI zur Identifikation von Bedürftigkeit und zur effizienten Zuteilung von Ressourcen könnte dazu beitragen, bestehende Barrieren auf dem Weg zu einer daseinsmächtigen Lebensführung zu reduzieren. Gleichzeitig müssen jedoch die Implikationen algorithmischer Entscheidungsprozesse kritisch betrachtet werden, um die sozialarbeiterische Praxis nicht zu technokratisieren und die individuelle Selbstbestimmung und Gerechtigkeit aus dem Blick zu verlieren.

KI-generierte Interventionen könnten Abweichungen von traditionellen Ansätzen bedingen und damit die Rolle von Sozialarbeiter:innen verändern. Diskutabel ist, welche neuen Kompetenzen erforderlich werden und wie sich die Integration von KI in die Professionsentwicklung auf die sozialarbeiterische Rol-

lenauffassung auswirkt. Dabei müssen ökonomische Prinzipien und soziale Teilhabe, wie von Rauschenbach und Züchner (2012, S. 151 ff.) als Herausforderung skizziert, neu verhandelt werden, um das Gleichgewicht zwischen Effizienz und ethischer Verantwortung nicht zu gefährden.

Letztlich muss sich die Soziale Arbeit mit der Frage auseinandersetzen, wie die humanistischen Grundlagen im Zeitalter von KI gewahrt bleiben können. Es gilt zu klären, wie KI-Systeme so gestaltet werden können, dass sie die humanistische Ethik der Sozialen Arbeit widerspiegeln und nicht untergraben, um den disziplinären Ansatz der Sozialen Arbeit als Menschenrechtsprofession nicht zu beugen. Eine kontinuierliche Reflexion und kritische Bewertung des KI-Einsatzes sind dafür unerlässlich (siehe hierzu auch die obigen Ausführungen zur Menschenrechtsprofession nach Silvia Staub-Bernasconi).

KI vermag die Grundlagentheorien der Sozialen Arbeit sowohl zu bereichern als auch herauszufordern. Es ist essenziell, dass die Soziale Arbeit als Profession diese Entwicklungen kritisch begleitet und sich aktiv an der Gestaltung der Rahmenbedingungen beteiligt, um im Kern ihrer Arbeit den Menschen weiterhin gerecht zu bleiben.

4.2 KI und sozialarbeiterische Theorieansätze

Die Implementierung von KI in der Sozialen Arbeit birgt die Möglichkeit, subjektzentrierte Ansätze zu stärken, indem sie individuelle Unterstützung auf Basis präziser Datenanalysen bietet. Vor dem Hintergrund der Arbeit von Hans Thiersch kann jedoch hinterfragt werden, ob autonom entscheidende KI-Systeme die Grundbedürfnisse und die Vielfältigkeit individueller Lebenswelten erfassen können (vgl. Lambers 2018, S. 501 ff.). Dabei ist zu beachten, dass KI möglicherweise zu einer Homogenisierung führen kann, die individuelle Erfahrungen und kontextuelle Besonderheiten vernachlässigt, was den Grundannahmen der Sozialen Arbeit entgegenläuft (vgl. Bommasani et al. 2021, o. S.).

Die Integration ethischer Prinzipien in KI lädt dazu ein, wie oben beschrieben, den bereits thematisierten systemischen Ansatz nach Silvia Staub-Bernasconi zu reflektieren. Hierbei gilt es, KI-gestützte Systeme zu entwickeln, die soziale Beziehungen und Machtverhältnisse berücksichtigen und die Wechselwirkungen zwischen Individuum und Umwelt im Sinne eines ethischen Designs erfassen (vgl. ebd.). Die Herausforderung, sozialarbeiterische Werte in algorithmische Prozesse einzuflechten, spiegelt eine von Mittelstadt (2019, S. 501 ff.) beschriebene Theorie-Praxis-Lücke in der KI wider.

Foundation Models als maschinelle Lernmodelle sind in der Lage, die Lebensweltorientierung nach Lothar Böhnisch zu unterstützen, indem sie partizipative und kontextsensitive Interventionen ermöglichen (vgl. Bommasani et al. 2021, o. S.). Gleichzeitig bedarf es einer kritischen Auseinandersetzung mit den Be-

schränkungen von KI, um sicherzustellen, dass die Komplexität sozialer Lebenswelten adäquat repräsentiert wird (vgl. Lambers 2018, S. 501 ff.).

Schließlich muss die Verwendung von KI in der Sozialen Arbeit zu einer Reflexion über die berufliche Identität führen. Die Selbstwahrnehmung und das professionelle Handeln von Sozialarbeitenden sind im Licht der Verantwortung für ihre fachliche Entscheidungsautonomie zu betrachten, was durch die Critical Technical Practice nach Agre (2014, S. 131 ff.) und die ethischen Prinzipien nach Banks (2020, o. S.) unterstützt wird.

Das Zusammenspiel zwischen KI und sozialarbeiterischen Theorieansätzen eröffnet neue Horizonte für die Professionsentwicklung. Die Disziplin hingegen wird aufgefordert, sowohl die Potenziale als auch die Grenzen von KI zu erkunden, um sicherzustellen, dass die zentralen Werte der Sozialen Arbeit gewahrt bleiben.

5 Herausforderungen und Chancen

Die Implementierung von KI-Systemen muss sich an menschlichen Werten orientieren und gleichzeitig die Gefahr der Verstärkung bestehender Vorurteile, wie sie durch Data Bias⁶ entstehen können, vermeiden (vgl. Wirtz et al. 2019, S. 596 ff.). Eine verantwortungsvolle Integration von KI in die sozialarbeiterische Praxis erfordert daher eine sorgfältige Reflexion über die Datenqualität und deren Auswirkungen auf die Entscheidungsfindung. Angesichts der Schnelligkeit ist ein verantwortungsbewusster Umgang mit KI zu gewährleisten, der den beschriebenen Werten der Sozialen Arbeit entspricht (Dwivedi et al. 2021, o. S.).

Das Potenzial von KI, emotionale und empathische Dienstleistungen zu verbessern, ist durch Technologien der emotionalen Erkennung und des affektiven Computings gegeben (Huang/Rust 2018, S. 144 ff.). Im Jahr 2025 mutet das nur noch minimal futuristisch an: Empathische KI-Systeme könnten Fachkräfte in der Sozialen Arbeit unterstützen, indem sie emotionale Zustände von Klient:innen erfassen und so zur Entwicklung von Interventionsstrategien beitragen. Gleichwohl müssen die Grenzen künstlicher Empathie auch bei fortschreitender KI-Entwicklung erkannt werden.

6 Data Bias beschreibt Verzerrungen, die in den zugrunde liegenden Daten von KI-Systemen enthalten sind und sich auf die Ergebnisse und Entscheidungen solcher Systeme auswirken können. Diese Verzerrungen können beispielsweise durch unrepräsentative Datensätze, fehlerhafte Erhebungsmethoden oder gesellschaftliche Ungleichheiten entstehen, die in den Daten reflektiert werden (vgl. Goram 2024, o. S.). Besonders in der Sozialen Arbeit ist die kritische Prüfung auf Data Bias essenziell, um Diskriminierung zu vermeiden und Gerechtigkeit zu gewährleisten.

Die Rolle von KI in der Sozialen Arbeit führt zu einem Wandel der Rollenerwartungen und Aufgabenprofile. Durch die Unterstützung von KI bei administrativen Routineaufgaben können Zeitressourcen für klient:innenzentrierte Tätigkeiten freigesetzt werden (vgl. Jarrahi 2018, S. 577 ff.). Dies könnte zu einer Neudefinition von Rollen in der Sozialen Arbeit führen, wobei die Notwendigkeit lebenslangen Lernens hervorzuheben ist, um mit technologischen Entwicklungen Schritt zu halten. Die Einführung von KI könnte jedoch auch die professionelle Identität beeinträchtigen und zu einer Entfremdung von den eigentlichen sozialarbeiterischen Aufgaben führen. Es ist daher von entscheidender Bedeutung, ein ausgewogenes Verhältnis zwischen der Nutzung von KI und der Aufrechterhaltung der Kernelemente der professionellen Identität zu finden (vgl. Wirtz et al. 2019, S. 696 ff.).⁷

Die Soziale Arbeit hat sicherzustellen, dass ihre humanistischen und ethischen Grundlagen ebenso in der KI-Ära Bestand haben. Die formale Kodierung von Werten in KI-Systemen stellt eine technische Herausforderung dar, was es umso wichtiger macht, die Grundprinzipien der Sozialen Arbeit zu reflektieren (vgl. Gabriel 2020, S. 411 ff.).

6 Fazit

Die rasante Entwicklung der KI wirft zentrale Fragen für die Soziale Arbeit auf: Wie kann KI sowohl die praktische Ausübung als auch die theoretischen Grundlagen der Disziplin beeinflussen? Dabei steht im Mittelpunkt, ob und wie KI-Technologien dazu beitragen können, die Soziale Arbeit effektiver, gerechter und zukunftsfähiger zu gestalten, ohne ihre humanistischen und ethischen Grundwerte zu gefährden. KI kann sowohl als transformative Kraft betrachtet werden, die praktische und theoretische Ansätze der Sozialen Arbeit neugestaltet, als auch als Herausforderung, die ethische und humanistische Werte auf die Probe stellt.

Durch den Einsatz von KI, Routinetätigkeiten zu automatisieren, könnten Sozialarbeitende mehr (Zeit-)Ressourcen für zwischenmenschliche Interaktionen und komplexe Fallbearbeitungen erhalten. Gleichzeitig bringt die Integration von KI kritische Aspekte mit sich. Ethische Herausforderungen und das Risiko von Machtasymmetrien stellen Bedenken dar, die bei der Implementierung von KI in der Sozialen Arbeit berücksichtigt werden müssen. Daraus ergibt sich bei der Einflussnahme von KI auf die Grundlagentheorien der Sozialen Arbeit die zwingende Notwendigkeit, humanistische Werte im technologischen Wandel zu bewahren.

Die für die Soziale Arbeit als Grundlagentheorien disziplinär verankerte Subjektzentrierung und der systemische Ansatz bieten wichtige Perspektiven, um die

7 Siehe hierzu auch den Beitrag von Löhe zur Einteilung von KI-Anwendungen nach Einsatzfeld in diesem Band.

komplexen Wechselwirkungen zwischen Mensch und Technologie in der Sozialen Arbeit zu reflektieren und zu verstehen. Die kritische Beleuchtung der Grenzen und Potenziale von KI in diesem Kontext erweitert das Verständnis über die Integration von Technologie in soziale Dienstleistungen.

Die fortschreitende Integration von KI in die Soziale Arbeit erfordert eine kontinuierliche Auseinandersetzung mit den technologischen Entwicklungen und ihren Auswirkungen auf die Profession- und Disziplinentwicklung. Die Balance zwischen technologischem Fortschritt und den humanistischen Prinzipien der Sozialen Arbeit muss dabei aufrechterhalten werden. Die KI bietet für die Entwicklung auf diesen beiden Ebenen wahrlich große Chancen, die sich die Soziale Arbeit, kritisch reflektierend und um ihre disziplinären Kerne wie der Selbstmandatierung als Menschenrechtsprofession wissend und haltend, offenhalten sollte.

Literatur

- Agre, Philip E. (2014): Toward a critical technical practice: Lessons learned in trying to reform AI. In: Social science, technical systems, and cooperative work. Psychology Press, S. 131–157.
- Banks, Sarah (2020): Ethics and values in social work. London: Bloomsbury Publishing.
- Bommasani, Rishi/Hudson, Drew A./Adeli, Ehsan/Altman, Russ/Arora, Simran/von Arx, Sydney/Bernstein, Michael S./Bohg, Jeannette/Bosselut, Antoine/Brunskill, Emma/Brynjolfsson, Erik/Buch, Shyamal/Card, Dallas/Castellon, Rodrigo/Chatterji, Niladri/Chen, Annie/Creel, Kathleen/Davis, Jared Quincy/Demszky, Dora, ..., Liang, Percy (2021): On the opportunities and risks of foundation models. arXiv. <https://arxiv.org/abs/2108.07258>
- Dignum, Virginia (2019): Responsible artificial intelligence: How to develop and use AI in a responsible way (Vol. 2156). Wiesbaden: Springer.
- Dwivedi, Yogesh K./Hughes, Laurie/Ismagilova, Elvira/Aarts, Gert/Coombes, Crispin/Crick, Tom/Duan, Yanqing/Dwivedi, Rohita/Edwards, John/Eirug, Aled/Galanos, Vassilis/Ilavarasan, P. Vigneswara/Janssen, Marijn/Jones, Paul/Kar, Arpan Kumar/Kizgin, Hatice/Kronemann, Bianca/Lal, Banita/Lucini, Biagio/Medaglia, Rony/Williams, Michael D. (2021): Artificial Intelligence (AI): Multidisciplinary perspectives on emerging challenges, opportunities, and agenda for research, practice and policy. International Journal of Information Management 57, 101994.
- Eubanks, Virginia (2018): Automating Inequality: How High-Tech Tools Profile, Police, and Punish the Poor. Picador, New York: St Martin's Press.
- Friesenhahn, Günter. J. (2018): Regime in der internationalen Sozialen Arbeit. Transformationen, disziplinäre Claims und fragile Durchsetzungsfähigkeit. In: Nothdurfter, Urban/Zadra, Franca/Nagy, Andrea/Lintner, Claudia (Hrsg.): Promotion Social Innovation and Solidarity Through Transformative Processes of Thought and Action. A Lifetime for Social Change Tribute to Susanne Elsen. Bozen: bu,press, S. 217–243.
- Gabriel, Jason (2020): Artificial intelligence, values, and alignment. In: Minds and Machines 30(3), S. 411–437.
- Goram, Mandy (2024): Ethik und Recht in KI-Systemen: Herausforderungen und Lösungen. Informatik Aktuell. <https://www.informatik-aktuell.de/betrieb/kuenstliche-intelligenz/ethik-und-recht-in-ki-systemen-herausforderungen-und-loesungen.html> (Abfrage: 15.06.2025).
- Huang, Ming-Hui/Rust, Roland T. (2018): Artificial intelligence in service. Journal of Service Research 21(2), S. 155–172. <https://doi.org/10.1177/1094670517752459>

- Jarrahi, Mohammad Hossein (2018): Artificial intelligence and the future of work: Human-AI symbiosis in organizational decision making. *Business Horizons* 61(4), S. 577–586.
- Lambers, Helmut (2018): *Theorien der Sozialen Arbeit: Ein Kompendium und Vergleich*. Opladen und Toronto: Barbara Budrich (utb).
- Long, Duri / Magerko, Brian (2020): What is AI Literacy? Competences and Design Considerations. Proceedings of the 2020 CHI conference on human factors in computing systems, ACM (2020). [dl.acm.org/doi/10.1145/3313831.3376727](https://doi.org/10.1145/3313831.3376727)
- Macsenaere, Michael (2024): Anwendungsrisiken und Limitation von KI. In: Macsenaere, Michael (Hrsg.): *Künstliche Intelligenz in der Kinder- und Jugendhilfe*. München: Ernst Reinhardt.
- Mennemann, Hugo/Dummann, Jörn (2022): *Einführung in die Soziale Arbeit*. 4. Auflage. Baden-Baden: Nomos.
- Mittelstadt, Brent (2019): Principles alone cannot guarantee ethical AI. *Nature Machine Intelligence* 1(11), S. 501–507. <https://doi.org/10.1038/s42256-019-0114-4>
- Rachmad, Yoesoep Edhie (2019): Transforming workforce skills in the era of society 5.0: Integrating technology and artificial intelligence in competency development. *The Academic Professor Board Protocol 2019, United Nations Global Compact 137635 UIPM*.
- Rauschenbach, Thomas (2020): Sozialpädagogik an drei Orten: Professionelle und disziplinäre Entwicklungen in den Hochschulausbildungen der Sozialen Arbeit. In: Cloos, Peter / Lochner, Barbara / Schoneville, Holger (Hrsg.): *Soziale Arbeit als Projekt: Konturierungen von Disziplin und Profession*. Wiesbaden: VS Springer, S. 145–158.
- Rauschenbach, Thomas / Züchner, Ivo (2012): Theorie der Sozialen Arbeit. In: Thole, Werner (Hrsg.): *Grundriss Soziale Arbeit: Ein einführendes Handbuch*. Wiesbaden: VS Verlag, S. 151–173.
- Röh, Dieter (2013): *Soziale Arbeit, Gerechtigkeit und das gute Leben: Eine Handlungstheorie zur daseinsmächtigen Lebensführung*. Wiesbaden: Springer.
- Staub-Bernasconi, Silvia (2017): The problem with ‚social problems‘ as domain of social work: A critical approach to the Melbourne ‚global definition of social work‘ of 2014 and constructivist theories of social problems. In: *European Journal of Social Work* 20(6), S. 958–971.
- Wirtz, Bernd W. / Weyerer, Jan C. / Geyer, Carolin (2019): Artificial intelligence and the public sector – applications and challenges. In: *International Journal of Public Administration* 42(7), S. 596–615. <https://doi.org/10.1080/01900692.2018.1498103>

KI und Theorie(bildung) Sozialer Arbeit¹

Angelika Beranek

Abstract: Der Artikel untersucht zunächst die Komponenten der Theorien der Sozialen Arbeit, die im Kontext von Künstlicher Intelligenz (KI) von Bedeutung sind. Darauf aufbauend werden drei wesentliche Aspekte zahlreicher Theorien der Sozialen Arbeit hervorgehoben und in Bezug auf KI erörtert. Erstens werden Ratschläge für Praktiker:innen der Sozialen Arbeit gegeben (Handlungsebene). Zweitens wird das Konzept des „Sozialen“ oder der „Kern der Sozialen Arbeit“ betrachtet, wobei besonderes Augenmerk auf soziale Beziehungen und soziale Gerechtigkeit gelegt wird. Drittens wird die Minderung sozialer Probleme als Zielsetzung der Sozialen Arbeit detaillierter untersucht, und KI wird als soziales Problem dargestellt. Der Artikel endet mit der Frage, ob es einer neuen Theorie der Sozialen Arbeit im Zusammenhang mit KI bedarf.

Keywords: Theorien, Soziale Probleme, KI, Koproduktion

1 Einführung

Um die Bedeutung von KI für Theorien Sozialer Arbeit zu entschlüsseln, lohnt sich zunächst ein Blick darauf, welche Bestandteile Theorien Sozialer Arbeit enthalten. So ist es möglich zu identifizieren, an welchen Stellen Theorien eventuell angesichts von KI überarbeitet werden müssten oder ob es sinnvoll wäre, eine neue Theorie Sozialer Arbeit zu entwickeln. Durch die Theorienvielfalt ist dieses Unterfangen allerdings fast nicht möglich, weshalb im Folgenden auf drei zentrale Punkte (vieler) Theorien Sozialer Arbeit fokussiert werden soll: 1. Hinweise an die Praktiker:innen Sozialer Arbeit, hier die Handlungsebene genannt, 2. „Das Soziale“ oder den Kern der Sozialen Arbeit, der hier in den Ausprägungen soziale Beziehungen und soziale Gerechtigkeit aufgegriffen wird und 3. die Minderung von sozialen Problemen als Zielkategorie der Sozialen Arbeit.

1 © Die Autor:innen 2025, veröffentlicht im Verlag Beltz Juventa, Open Access: CC BY 4.0
Gesä Linnemann/Julian Löhe/Beate Rottkemper (Hg.), Künstliche Intelligenz in der Sozialen Arbeit
10.3262/978-3-7799-8562-4_005

2 KI und die Handlungsebene Sozialer Arbeit

Soziale Arbeit kann als Handlungswissenschaft betrachtet werden (vgl. Staub-Bernasconi 2018): Theorien der Sozialen Arbeit integrieren häufig Handlungsmethoden und Handlungsprinzipien, die es Praktiker:innen ermöglichen, ihr eigenes Handeln zu reflektieren. Gleichzeitig werden Ziel- und Wertvorgaben bereitgestellt. Mit Blick auf die Handlungsebene erscheint es naheliegend, dass KI-Systeme eine zunehmend bedeutende Rolle einnehmen. Denn zentrale Fragen betreffen den Einsatz von KI auf der Handlungsebene, etwa zur Früherkennung von Problemen im Kontext der Kindeswohlgefährdung (vgl. Burghardt et al. 2024), den Einsatz von KI im Verwaltungsbereich (vgl. Kreidenweis/Diebold 2024, S. 8) sowie den Einsatz von Chatbots in der Beratung (vgl. Linnemann/Löhe/Rottkemper 2024).

Darüber hinaus eröffnen sich durch den Einsatz von KI gänzlich neue Handlungsfelder in der Sozialen Arbeit, etwa Kommunikationstrainings im Umgang mit Robotern oder Chatbots. In diesem Buch finden sich zahlreiche Beiträge, die sich mit der Handlungsebene beschäftigen, z. B. KI in der Kinder- und Jugendhilfe (siehe den Beitrag von Macsenaere und Feist-Ortmanns) oder KI in der medienpädagogischen Sozial- und Kulturarbeit (siehe den Beitrag von Tappe). Zur weiteren Vertiefung wird an dieser Stelle darauf verwiesen.

3 KI und der „Kern“ der Sozialen Arbeit

Um die Auswirkungen der KI auf die Theorien der Sozialen Arbeit umfassend zu verstehen, ist es notwendig, nicht nur die praktische Handlungsebene zu untersuchen, sondern auch das Fundament und die Grundprinzipien der Sozialen Arbeit in den Fokus zu rücken. Zu diesem „Kern“ der Sozialen Arbeit wird im Folgenden beispielhaft das „Soziale“ an sich und zudem die eng daran anknüpfende Aufgabe der Herstellung sozialer Gerechtigkeit beschrieben.

3.1 Soziale Beziehungen

Den Kern der Sozialen Arbeit berührt KI insbesondere dann, wenn es um die Frage des Sozialen an sich geht. KI greift grundlegende Fragen des Mensch-seins auf und stellt diese teilweise infrage: Diskutiert wird eine Dehumanisierung (vgl. Singer 2019; vgl. Beiglböck et al. 2023; vgl. Schifffhauer/Remke 2024) der Menschen auf der einen und eine Anthropomorphisierung (vgl. Zlotowski et al. 2018; vgl. Weyer 2022) der Technik auf der anderen Seite. Beide Prozesse verändern soziale Beziehungen grundlegend. Da soziale Beziehungen sowohl Gegenstand als auch Arbeitsweise der Sozialen Arbeit sind, können diese Prozesse nicht unbeachtet

bleiben, wenn KI im Kontext der Theorien der Sozialen Arbeit behandelt wird. Besonders relevant für die Soziale Arbeit ist hierbei die Rolle von KI und Robotern als soziale Akteure in der Gesellschaft. Gemäß der Media Equation Theory neigen Menschen dazu, Computer und andere Medien so zu behandeln, als handle es sich um Menschen. Insbesondere der Gebrauch von Sprache aktiviert soziale Kategorien (vgl. Nass/Brave 2005). Wie Linnemann, Löhe und Rottkemper (2023) in ihrem Beitrag zu Natural Language Processing eindrücklich darlegen, ist dies für die Soziale Arbeit von Bedeutung. Darüber hinaus wirkt sich die Humanisierung von Robotern auf unseren Umgang mit ihnen aus. Menschen sind bereit, Anweisungen von Robotern zu folgen und Verantwortung mit ihnen zu teilen (vgl. Onnasch et al. 2019, S. 37 ff.).

Aktuell verändert sich zudem die Art, mit welchen Zielen KI in einer kapitalistischen Gesellschaft eingesetzt wird. Die letzte Generation von KI-Systemen zielte darauf ab, Aufmerksamkeit zu binden („race for attention“). Die aktuelle Generation hingegen ist darauf ausgerichtet, eine möglichst große Vertrautheit zwischen Mensch und Maschine herzustellen („race for intimacy“). Sie sollen in Beziehung mit uns treten und Intimität aufbauen. Ein Beispiel hierfür ist die Anwendung My AI in Snapchat.

So lässt sich festhalten, dass im Bereich der sozialen Beziehungen mit und über KI-Systeme(n) verschiedene Veränderungen beobachtbar sind, die auf die Soziale Arbeit einwirken und sich beispielsweise im Einsatz von Chatbots zur Beratung widerspiegeln (vgl. Linnemann/Löhe/Rottkemper 2024, S. 6 ff.).

3.2 Soziale Gerechtigkeit

Der Abbau von Marginalisierungen und die Förderung von Autonomie und Selbstbestimmung gelten als zentrale Aufgaben der Sozialen Arbeit, um zur Herstellung sozialer Gerechtigkeit beizutragen. Diese Aufgaben werden sowohl in theoretischen Grundlagen der Sozialen Arbeit als auch in berufsethischen, internationalen Abkommen und Anerkennungstheoretischen Ansätzen umrissen. KI stellt die Soziale Arbeit hier vor neue Herausforderungen: Big-Data-Analysen (mittels KI) und Scoring sorgen für gläserne Bürger:innen und können so marginalisierte Gruppen gefährden. Beschrieben werden hier beispielsweise negative Folgen für queere Menschen (vgl. Klippahn et al. 2023).

Doch nicht nur aktuelle Schutzbedürfnisse sind in den Fokus zu nehmen, wenn es um KI und Soziale Arbeit geht – die Soziale Arbeit muss sich angesichts der Entwicklungen auch mit zukünftigen Schutzbedürfnissen auseinandersetzen. KI-Systeme sind beispielsweise bereits in der Lage, aufgrund von Gehirnaktivitäten Gedanken zu reproduzieren (vgl. Takagi/Nishimoto 2022; vgl. Chen et al. 2023; vgl. Tang et al. 2023). Benötigen wir also angesichts dieser Technologien neue Rechte wie ein Recht darauf, dass Gedanken nicht gelesen werden,

und muss sich die Soziale Arbeit für diese einsetzen, um ihrem Auftrag, soziale Gerechtigkeit herzustellen, gerecht zu werden?

4 KI und Soziale Probleme

Ein dritter Ansatzpunkt zur Entwicklung von Theorien Sozialer Arbeit findet sich bei Schönig et al. (2024): Diese sehen Krisen als grundlegend für die Entwicklung Sozialer Arbeit (und somit auch ihrer Theorien).

„Für die Soziale Arbeit sind Krisen von grundlegender Bedeutung. Sie stehen an ihrem Beginn (dort ‚soziale Probleme‘ genannt), sie begleiten den Prozess der Sozialen Arbeit (als Intervention zur Prävention latenter Krisen oder als Intervention zur Bewältigung manifester Krisen), und an deren Ende wird die Soziale Arbeit daran gemessen, welchen Beitrag sie zur Krisenbewältigung oder -prävention geleistet hat. Überspitzt kann man formulieren: ohne soziale Krise keine Soziale Arbeit und ohne Soziale Arbeit keine soziale Krisenbewältigung.“ (ebd., S. 9)

Unter dieser Voraussetzung zwingt sich geradezu ein Blick auf KI als soziales Problem und seine Bedeutung für die Soziale Arbeit auf. Um KI als soziales Problem zu fassen, muss zunächst geklärt werden, was unter einem sozialen Problem im Kontext Sozialer Arbeit überhaupt zu verstehen ist.

Groenemeyer (2013) definiert in Abgrenzung zu privaten oder individuellen Problemen solche Probleme als soziale Probleme,

„für deren Bearbeitung, Kontrolle oder Lösung eine gesellschaftliche bzw. politische Verantwortung angemahnt und erwartet wird; sie sind Gegenstand öffentlicher Diskussionen [...] politischer Maßnahmen und Regelungen, und zu ihrer Bearbeitung sind spezialisierte öffentliche Institutionen geschaffen worden“ (ebd., S. 758).

Sie sind Bestandteil moderner Gesellschaften und als solche werden sie als Folgeerscheinung der gesellschaftlichen Struktur gesehen, auf die die Gesellschaft wiederum reagieren muss. Um zu einem sozialen Problem zu werden, muss ein Problem zunächst einmal öffentlich thematisiert werden. Soziale Probleme können aus schadenverursachenden Lebensbedingungen erwachsen und ihre Thematisierung bezieht sich somit auf die verletzten Wertvorstellungen von Menschenwürde, Gleichheit, Gerechtigkeit etc. (vgl. ebd., S. 758 f.).

Bei der Betrachtung dieser Merkmale in Bezug auf KI erfolgt eine Anmahnung von politischer Verantwortung in Bezug auf die KI-Entwicklung und Implementierung auf vielen Ebenen. Nicht nur netzpolitische und/oder zivilgesellschaftliche Stimmen von Akteuren wie netzpolitik.org oder der Gesellschaft für Freiheitsrechte e. V. kritisieren den Einsatz von KI zur Überwachung z. B. im Zusam-

menhang mit dem Polizeiaufgabengesetz in Bayern öffentlich. In der öffentlichen Diskussion taucht KI vor allem in Hinblick auf generative KI (siehe den Beitrag von Rottkemper in diesem Band) auf. Seit der Veröffentlichung von ChatGPT im November 2022 ist das Thema in der Presse allgegenwärtig.

Als Vorläufer für spezielle öffentliche Institutionen zur Bearbeitung des Problems können diverse Ethikkommissionen zu diesem Thema gesehen werden. So gibt es lokale Kommissionen wie die Ethikkommission der TUM, deutschlandweite Einrichtungen wie den deutschen Ethikrat sowie länderübergreifende Einrichtungen wie die EU- oder UNESCO-Ethikkommission, die sich mit KI beschäftigen. Zudem wurden bereits europäische Gesetze wie der AI Act (Verordnung (EU) 2024/1689) speziell zur Regulierung von KI verabschiedet.

KI erfüllt zudem das Merkmal sozialer Probleme, ein Bestandteil moderner Gesellschaften sowie eine Folgeerscheinung ihrer Struktur zu sein, auf die die Gesellschaft reagieren muss. Dies wird besonders durch den weitreichenden Einsatz von KI in öffentlichen und privaten Bereichen deutlich, der oft von ökonomischen Interessen getrieben wird. KI ist demnach laut dieser Definition als soziales Problem anzuerkennen.

Über diese eher enge Begriffsbestimmung hinaus gibt es überzeugende Argumente dafür, KI als soziales Problem zu betrachten. Dies liegt vor allem an dem vielfältigen und tiefgreifenden Einfluss, den KI auf andere soziale Probleme wie Diskriminierung, Armut oder Gewalt ausübt. Diese Einflüsse können aktuelle Wertvorstellungen wie Menschenwürde, Gleichheit und Gerechtigkeit untergraben, die bereits Bestandteil der vorherigen Definition waren. Nicht zu vergessen sind hierbei die ökologischen Folgen des KI-Einsatzes. Durch den enormen Ressourcenverbrauch (Strom, Wasser und seltene Erden) wirkt KI auch auf die Ursache anderer sozialer Probleme wie Migrationsbewegungen, die durch Hunger und Hitze ausgelöst werden, ein.

Um den Einfluss von KI auf bereits vorhandene soziale Probleme zu verdeutlichen, wird im Folgenden für Diskriminierung, Armut oder Gewalt je ein Beispiel benannt:

- Algorithmische Diskriminierungen können auf verschiedene Art und Weise entstehen (vgl. Zweig 2019, S. 220). Deutlich werden die Folgen dieser Diskriminierungen z. B. in den Forschungen rund um das Thema Gender und KI. Spezielle Algorithmen erkennen weibliche Gesichter schlechter, übersetzen aus geschlechtsneutralen Sprachen in veraltete Rollenbilder (Ärzte sind dann in der Regel männlich) und diskriminieren Bewerbungen von Frauen negativ, wenn es um technische Berufe geht (vgl. Bengler 2024). Zudem kann KI helfen, Diskriminierungen zu erkennen und zu beseitigen.
- In Bezug auf Armut kann die Ausbeutung von Clickworkern zum Training von KI (vgl. Laux/Stephany/Liefgreen 2023) und die Ausbeutung des Globalen Südens zur Ressourcengewinnung (um entsprechende Hardware

überhaupt produzieren zu können, auf der KI läuft) genannt werden (vgl. Karlbauer 2023). Umgekehrt könnte KI dazu beitragen, neue Chancen in der Bekämpfung von Armut, z. B. durch verbessertes Ressourcenmanagement, zu ermöglichen.

- In Bezug auf (digitale) Gewalt sind ebenfalls positive und negative Effekte von KI zu beobachten. KI kommt im militärischen Bereich zum Einsatz (wodurch beispielsweise die Zahl ziviler Opfer erhöht wird) und erhöht Risiken von Kindern und Jugendlichen im Netz, nun stärker durch sexualisierte Gewalt, Mobbing und Extremismus bedroht zu sein. Bei der Strafverfolgung und Früherkennung von Straftaten (Predictive Policing) kann KI jedoch auch hilfreich sein.

KI wirkt demnach auf vorhandene soziale Probleme ein und kann diese sowohl verstärken als auch mindern und ist darüber hinaus als eigenständiges soziales Problem zu betrachten. In diesem Zusammenhang ist es eine Aufgabe der Sozialen Arbeit, die negativen Folgen zu mindern und auf das soziale Problem KI zu reagieren.

5 Eine Theorie für die KI?

Angesichts der Wirkung von KI auf die Handlungsebene, den „Kern der Sozialen Arbeit“, und des Status von KI als soziales Problem stellt sich nun die Frage, ob es einer eigenen Theorie für KI und Soziale Arbeit bedarf. Notwendig erscheinen hier Anpassungen der vorhandenen Theorien auf der einen Seite (vgl. Beranek 2021) und eine eigene Theorieentwicklung auf der anderen Seite. Ausgehend vom aktuellen Theoriediskurs wird deutlich, dass das Thema durchaus in einem eigenen Theorieentwurf verhandelt werden kann. Sandermann und Neumann beschreiben ein Ende der Großtheorien der Sozialen Arbeit und beobachten eine zunehmende Fokussierung auf Einzelfragen (vgl. Sandermann/Neumann 2020, S. 215). KI kann als eine solche Einzelfrage betrachtet werden. Bei einer Änderung der Perspektive von der reinen Anpassung vorhandener Theorien Sozialer Arbeit (siehe Handlungsebene, Kern der Sozialen Arbeit und KI als soziales Problem) zu einer Neuentwicklung von Theorien bieten sich drei Betrachtungsweisen an, die Grundlage einer Theorie Sozialer Arbeit im Kontext von KI sein könnten: KI als eigenständiger Agent in der Gesellschaft, KI als Koproduzent Sozialer Arbeit und KI als empirisches Tool zur Theorieentwicklung.

(I) KI als (neuer) Agent in der Gesellschaft

In einer ersten Analyse werden KI-Systeme, die sich immer mehr in alltägliche Prozesse integrieren, als Agent und Handlungsträger betrachtet. Noch treten sie in der Regel als körperlose, häufig als generative KI in Erscheinung. Jedoch füh-

ren Fortschritte in der Robotik dazu, dass künftig vermehrt physisch präsente KI-Systeme in unserem Alltag zu erwarten sind. Die Forschung zeigt, dass Menschen sowohl körperlose als auch physisch präsente KI-Systeme als Handlungsträger wahrnehmen (Onnasch et al. 2019, S. 32 ff.). Dabei werden dieselben etablierten Kategorien verwendet, die traditionell zur Einteilung der Umwelt in Tiere, Menschen und übernatürliche Wesen genutzt werden. Menschen tendieren dazu, KI-Systeme zu anthropomorphisieren, also sie wie Menschen zu betrachten und zu behandeln (vgl. Złotowski et al. 2018). Dies hat Auswirkungen, die über die bereits beschriebenen Wirkungen auf soziale Beziehungen hinausgehen.

Um die Bedeutung dieser Systeme und ihrer Wahrnehmung in unserer Gesellschaft zu erfassen, kann auf bereits bekannte Phänomene zurückgegriffen werden. In diesem Fall wird die Metapher der Migrationsbewegungen und die Reaktion der Sozialen Arbeit darauf herangezogen. Die „Einwanderung“ von KI-Systemen unterscheidet sich in vielerlei Hinsicht von bisherigen Migrationsbewegungen, bei denen vermeintlich Unbekanntes in einer Gesellschaft aufgetaucht ist. Diese Unterschiede ergeben sich vor allem aus dem neuartigen Charakter der KI und der Tatsache, dass sie von Menschen entwickelt wurde. Im Gegensatz zu früheren Migrationsbewegungen, die Menschen betrafen, handelt es sich bei KI-Systemen um Entitäten, die lediglich als menschenähnlich wahrgenommen werden.

Eine Betrachtung des Auftauchens von KI mit der Metapher der Migration macht deutlich, dass Soziale Arbeit hier gefragt ist. Dieser Vergleich muss an einigen Stellen hinken, da wir es mit einem gänzlich neuen Phänomen zu tun haben. Er kann aber gleichzeitig helfen, ein Verständnis dafür zu entwickeln, wie Soziale Arbeit auf KI reagieren sollte und wie KI theoretisch begriffen werden kann.

Es lassen sich mithilfe dieser Metapher diverse Parallelen herstellen:

1. Institutionen und Anforderungen der Sozialen Arbeit verändern sich.
2. Sozialräume verändern sich.
3. Beziehungsformen und Verhältnisse zwischen Einheimischen (Menschen) und Zugewanderten (KI) werden verhandelt.
4. Eine Spaltung der Gesellschaft in Personen, die Zuwanderung für eine Chance oder eine Gefahr halten, droht.
5. Interkulturelles Verständnis ist für ein gutes Miteinander unersetzlich.
6. Ethische und rechtliche Fragestellungen werden eröffnet.

Im Folgenden werden diese Parallelen kurz beschrieben und die sich daraus ergebenden Handlungsaufträge für die Soziale Arbeit abgeleitet.

(II) Institutionen und Anforderungen

Die Veränderung von Institutionen der Sozialen Arbeit ist angesichts der Digitalisierung bereits vielfältig beschrieben (Kutscher et al. 2020, S. 363 ff.). Durch KI verändern sich zunächst insbesondere Verwaltungsstrukturen, aber auch Kern-

prozesse der Sozialen Arbeit wie die Risikoabschätzung (z. B. bei Kindeswohlgefährdung) oder die Beurteilung von Fällen (z. B. im Asylverfahren). Somit sind sowohl Strukturen als auch Aufgaben der Institutionen betroffen. Besonders relevant hierbei ist die Gefahr der Diskriminierung von marginalisierten Gruppen durch KI-Systeme. Die Soziale Arbeit muss dringend selbst an der Entwicklung der in ihrem Bereich eingesetzten Systeme beteiligt sein, um nicht ungewollt digitale Ungleichheiten zu verstärken, anstatt diese zu bekämpfen.

(III) Sozialräume

Sozialraumorientierung ist in der Sozialen Arbeit nicht mehr wegzudenken. Hierbei geht es vor allem darum, nicht Einzelpersonen mit sozialpädagogischen Maßnahmen zu verändern, sondern die Lebenswelten so zu gestalten, dass Personen auch in schwierigen Lebenslagen besser zurechtkommen (vgl. Reinhard 2024, S. 13). Es geht also weniger um die „Veränderung der Akteure“ bzw. darum, diese besser in die Gesellschaft einzupassen. Vielmehr wird angestrebt, die Ressourcen der Lebenswelten so zu nutzen, dass die Akteur:innen sich ihrem „Willen“ gemäß entfalten können (vgl. Kergel 2020, S. 230).

Digitale Räume sind als Sozialräume oder als Teil von Sozialräumen zu sehen (vgl. Kergel 2020, S. 321 ff.). Diese Räume werden nun im analogen und digitalen Raum durch KI verändert. Im physischen Raum tritt KI in Form von robotischen Systemen mit körperlicher Präsenz auf. Was hier möglich ist, zeigte u. a. 2023 die erste UNO-Presskonferenz mit Robotern (<https://aiforgood.itu.int/summit23/>). Bereits jetzt begegnen uns KI-gesteuerte Liefer- oder Serviceroboter im Alltag. Im digitalen (Sozial-)Raum ist KI selbstverständlicher Bestandteil der Umgebung. KI-generierte Texte und Bilder sind an der Tagesordnung und können sowohl positive als auch negative Auswirkungen auf unsere Weltwahrnehmung haben. Gerade die schnelle und massenhafte Produktion von Desinformation ist hier als negatives Beispiel zu nennen. Die Aufgabe Sozialer Arbeit besteht darin, KI als Ressource im Sozialraum für Adressat:innen der Sozialen Arbeit nutzbar zu machen und KI als Strukturelement des Sozialraumes so (mit) zu gestalten, dass fördernde und nicht hemmende Strukturen entstehen.

(IV) Beziehungsformen und Verhältnisse

Beziehungsformen und gesellschaftliche Verhältnisse sind durch KI in vielfältiger Weise betroffen. Hier werden exemplarisch die Auswirkungen auf den Arbeitsmarkt und parasoziale Beziehungen herangezogen. Ebenso wie bei anderen Migrationsbewegungen ist durch KI eine Veränderung des Arbeitsmarkts zu bemerken. Substitutionsprozesse lassen auf der einen Seite Arbeitsplätze verschwinden, auf der anderen Seite entstehen neue Arbeitsmöglichkeiten (vgl. Lane / Saint-Martin 2021). Dies erfordert von der Sozialen Arbeit zum einen das Auffangen und Betreuen von Personen, die durch diese Veränderungen vom Arbeitsmarkt freige-

setzt werden, zum anderen die Förderung von Kompetenzen, die notwendig sind, damit eine Chancengleichheit auf dem Arbeitsmarkt bestehen bleibt.

In Bezug auf Beziehungsformen verschärft KI den Effekt der parasozialen Beziehungen (siehe den Beitrag von Linnemann in diesem Band). Diese können nun leichter aufgebaut werden und so ihre sowohl positive als auch negative Wirkmacht entfalten. Das Spektrum der zu behandelnden Themen geht von parasozialer Meinungsführerschaft (vgl. Leißner et al. 2014) bis hin zu romantischen Beziehungen mit Mediencharakteren (vgl. Liebers 2021), die sich im Zuge von KI auf die Interaktion mit KI-Systemen übertragen lassen können. Gefragt könnte zukünftig hier z. B. ein Wahrnehmungstraining in Bezug auf KI durch die Soziale Arbeit sein. Auch die Auswirkungen auf gesellschaftliche Problemlagen wie steigende Einsamkeit sollten in den Fokus der Sozialen Arbeit gelangen.

(V) Spaltung der Gesellschaft

Genauso wie bisherige Migrationsbewegungen von Menschen ist jedoch die Beurteilung dieser und die damit verbundene Spaltung der Bevölkerung Teil der ausgelösten Dynamik. Die einen sehen große Heilsversprechen in KI (vgl. Selke 2024), die anderen nehmen diese als Bedrohung wahr (vgl. Kreissl/von Laufenberg 2024) und wieder andere ignorieren zunächst die schleichenden Veränderungen. Dies führt zu einer unterschiedlichen Akzeptanz von KI und, damit verbunden, zu unterschiedlich kompetentem Umgang im Alltag mit ihr. Als Future Skill ist jedoch ein kompetenter Umgang mit KI wichtig, um dem Leitbild von freien, kompetenten, mündigen Bürger:innen entsprechen zu können.

Um diese Ängste zu mindern, wäre beispielsweise die Schaffung von Begegnungsräumen mit KI eine Aufgabe der Sozialen Arbeit. Ähnlich wie bei anderem Fremden verschwindet Angst, wenn Wissen und Erfahrung im Umgang mit dem Unbekannten an ihre Stelle treten. Solche Begegnungsräume könnten beispielsweise in Alten-Service-Centern oder Jugendhäusern eingerichtet werden und es Menschen ermöglichen, sich in einem geschützten Rahmen mit KI-Systemen auseinanderzusetzen.

(VI) Interkulturelle Kompetenz

Eine weitere Parallele besteht in der Notwendigkeit von interkultureller Kompetenz. Wenn KI als Migration begriffen wird, begegnet uns mit dieser Migration auch eine eigene Kultur. Um in einer Kultur der Digitalität, wie sie Stalder 2016 beschreibt (vgl. Stalder 2016), zu interagieren, werden spezifische Kompetenzen benötigt. Diese Kultur wird durch KI-Systeme noch einmal verändert und ihre Wirkung verstärkt. Kommunikative Kompetenzen wie Prompting (zum Bedienen von KI) oder Kommunikationsstrategien im Kontext von humanoiden Robotern könnten so zu einem neuen Arbeitsfeld der Sozialen Arbeit werden.

(VII) Ethische und rechtliche Fragestellungen

Im Bereich der ethischen und rechtlichen Fragestellungen steht im Kontext der Sozialen Arbeit die Frage der Verantwortung im Mittelpunkt.

Wer trägt die Verantwortung für den Einsatz von KI? Wer trägt die Verantwortung für Entscheidungen, die durch KI getroffen werden? Welche Prozesse der Sozialen Arbeit sollten durch KI unterstützt oder ersetzt werden? Welche nicht? Im Arbeitsfeld der Sozialen Arbeit ergeben sich zahlreiche ethische und rechtliche Fragestellungen, insbesondere im Zusammenhang mit dem Datenschutz, aber auch darüber hinaus.

Zusammenfassend lässt sich festhalten, dass durch KI ein neuer Akteur in der Gesellschaft entstanden ist, der die Soziale Arbeit auf mehreren Ebenen betrifft. Die Betrachtungsweise von KI als Migration in die Gesellschaft kann helfen, Aufgaben und Verantwortlichkeiten der Sozialen Arbeit in diesem Kontext zu definieren.

1. KI und Koproduktion Sozialer Arbeit

Die zweite Betrachtungsweise setzt voraus, dass KI als Handlungsträgerin wahrgenommen wird, und legt den Fokus auf die Koproduktion Sozialer Arbeit. Der Begriff der Koproduktion Sozialer Arbeit ist aus der dienstleistungsorientierten Sozialen Arbeit bekannt. Ausgangslage ist, dass sozialpädagogisches Handeln auf die teilhabeorientierte Koproduktion sozialer Hilfen durch die Adressat:innen selbst angewiesen ist. Weiter gedacht wird dies im Capability Approach (vgl. Böllert 2008, S. 69). Wenn KI nun als neue Handlungsträgerin und Agentin in der Gesellschaft auftritt, könnte ein zusätzlicher Faktor ins Spiel kommen. Hier muss allerdings der Begriff der Koproduktion so verstanden werden, wie er in Bezug auf generative KI diskutiert wird: KI tritt an die Stelle eines Teammitglieds und lässt somit den ihr oft zugeschriebenen Werkzeug-Charakter hinter sich. Gemeint ist also der Einsatz von KI zum Brainstorming, Überprüfen und Erweitern von Gedankengängen oder für Diskussionen.

Diesen Ansatz verfolgen Forschungsprojekte wie KAIMo (vgl. Burghardt et al. 2024), die KI als Assistenz zur Einschätzung von Kindeswohlgefährdung untersuchen. Dieser Einsatz von KI ist auch in anderen Handlungsfeldern der Sozialen Arbeit denkbar. Vor allem Chatbots eignen sich für eine Koproduktion. Daraus ließe sich eine Theorie der Koproduktion Sozialer Arbeit im Dreieck und Spannungsfeld von Adressat:in, Sozialarbeiter:in und KI entwickeln.

2. KI als empirisches Tool zur (eigenen) Theorieentwicklung

Die dritte Betrachtungsweise stellt einen gewagten Ausblick auf die Rolle von KI als empirisches Tool zur Theorieentwicklung in der Sozialen Arbeit dar. KI fügt sich als Tool in den Trend der zunehmenden Rolle von empirischer Forschung in der Wissensproduktion der Sozialen Arbeit ein. Eckl und Ghanem widmen sich in ihren Arbeiten Big-Data-Analysen in der Sozialen Arbeit (vgl.

Eckl/Ghanem 2020). Wird dieser Gedanke aufgegriffen und angesichts von KI weitergesponnen, könnte KI auch als Tool eingesetzt werden, um wiederum neue (Kleinst-)Theorien Sozialer Arbeit zu entwickeln.

Die kurzen Darstellungen zeigen, dass es viele Ansätze gibt, KI im Kontext der Theorien Sozialer Arbeit zu betrachten. Eine Ausdifferenzierung und weitere Beschäftigung mit dem Thema scheinen angebracht, um mithilfe einer Theorie zu KI und Sozialer Arbeit dafür zu sorgen, sowohl KI als soziales Problem zu begreifen als auch zur Lösung desselben beitragen zu können.

Literatur

- Scheibenbogen, Oliver/Jobst, F. (2023): Appgestützte Therapie und Virtuelle Realität: Dehumanisierung der Behandlung oder sinnvolle Erweiterung? In: Beiglböck, Wolfgang/Gottwald-Nathaniel, Gabriele/Preinsperger, Wolfgang/Scheibenbogen, Oliver (Hrsg.): Suchtbehandlung und Digitalisierung. Suchtprävention und Suchttherapie zwischen menschlicher Begegnung und virtueller Realität. 1. Auflage 2023. Berlin und Heidelberg: Springer, S. 129–143.
- Beranek, Angelika (2021): Soziale Arbeit im Digitalzeitalter: eine Profession und ihre Theorien im Kontext digitaler Transformation. 1. Auflage. Weinheim und Basel: Beltz Juventa.
- Böllert, Karin (2008): Disziplin und Disziplinpolitik. In: Bielefelder Arbeitsgruppe 8 (Hrsg.): Soziale Arbeit in Gesellschaft. Wiesbaden: VS Verlag für Sozialwissenschaften, S. 65–71.
- Burghardt, Jennifer/Lehmann, Robert/Reder, Michael/Koska, Christopher/Kraus, Maximilian/Müller, Nicholas (2024): Kann Künstliche Intelligenz sozialarbeiterische Entscheidungsprozesse unterstützen? Ethik und digitale Operationalisierung im Feld der Kindeswohlgefährdung. In: unsere jugend 76, S. 300–310.
- Carstensen, Tanja/Ganz, Kathrin (2024): Künstliche Intelligenz und Gender – eine Frage diskursiver (Gegen-)Macht? In: WSI-Mitteilungen 77, S. 26–33.
- Chen, Zijiao/Qing, Jiaxin/Xiang, Tiange/Yue, Wan Lin/Zhou, Juan Helen (2023): Seeing Beyond the Brain: Conditional Diffusion Model with Sparse Masked Modeling for Vision Decoding. arXiv. <https://arxiv.org/abs/2211.06956>
- Eckl, Markus/Ghanem, Christian (2020): Big Data, Textanalyse und Forschung in der Sozialen Arbeit. In: Kutscher, Nadia/Ley, Thomas/Seelmeyer, Udo/Siller, Friederike/Tillmann, Angela/Zorn, Isabel (Hrsg.): Handbuch Soziale Arbeit und Digitalisierung. Weinheim: Beltz, S. 625–637.
- Groenemeyer, Axel (2013): Soziale Probleme. In: Mau, Steffen/Schöneck, Nadine M. (Hrsg.): Handwörterbuch zur Gesellschaft Deutschlands. 3., grundlegend überarbeitete Auflage. Wiesbaden: Springer VS, S. 758–773.
- Karlbauer, Matthias (2023): Seltene Erden: Wie künstliche Intelligenz den Bedarf steigert und den Abbau beschleunigt. te.ma gGmbH. <https://te.ma/art/blvfvm/mckie-ki-tiefseebergbau-seltene-erden/> (Abfrage: 15.06.2025).
- Kergel, David (2020): Der Ansatz der Sozialraumorientierung im digitalen Wandel. In: Kutscher, Nadia/Ley, Thomas/Seelmeyer, Udo/Siller, Friedrike/Tillmann, Angela/Zorn, Isabel (Hrsg.): Handbuch Soziale Arbeit und Digitalisierung. Beltz Juventa, S. 229–240.
- Klipphahn, Michael/Koster, Ann-Kathrin/Santos Bruss, Sara Morais dos (Hrsg.) (2023): Queere KI: zum Coming-out smarterer Maschinen. Bielefeld: transcript (= KI-Kritik 3).
- Kreidenweis, Helmut/Diebold, Maria (2024): Studie Künstliche Intelligenz in der Sozialwirtschaft: Forschungsbericht. Eichstätt: Arbeitsstelle für Sozialinformatik. <https://www.>

- sozialinformatik.de/fileadmin/1805/pdf_documents/Forschung_und_Entwicklung/Studie-KI-Sozialwirtschaft-2024.pdf (Abfrage: 15.06.2025).
- Kreissl, Reinhard/von Laufenberg, Roger (2024): Risiken und Gefahren der ‚Künstlichen‘ ‚Intelligenz‘. In: Heinlein, Michael/Huchler, Norbert (Hrsg.): Künstliche Intelligenz, Mensch und Gesellschaft. Wiesbaden: Springer Fachmedien Wiesbaden, S. 225–261.
- Kutscher, Nadia/Ley, Thomas/Seelmeyer, Udo/Siller, Friedrike/Tillmann, Angela/Zorn, Isabel (Hrsg.) (2020): Handbuch Soziale Arbeit und Digitalisierung. 1. Auflage. Weinheim: Beltz Juventa.
- Lane, Marguerita/Saint-Martin, Anne (2021): Die Auswirkungen von KI auf die Arbeitsmärkte: Was wir bislang wissen. https://www.oecd-ilibrary.org/social-issues-migration-health/die-auswirkungen-von-ki-auf-die-arbeitsmarkte-was-wir-bislang-wissen_540444e1-de (Abfrage: 15.06.2025).
- Laux, Johann/Stephany, Fabian/Liefgreen, Alice (2023): The Economics of Human Oversight: How Norms and Incentives Affect Costs and Performance of AI Workers. In: SSRN Electronic Journal.
- Leißner, Laura/Stehr, Paula/Rössler, Patrick/Döringer, Esther/Morsbach, Melissa/Simon, Linda (2014): Parasoziale Meinungsführerschaft: Beeinflussung durch Medienpersonen im Rahmen parasozialer Beziehungen: Theoretische Konzeption und erste empirische Befunde. In: Publizistik 59, S. 247–267.
- Liebers, Nicole (2021): Romantische parasoziale Interaktionen und Beziehungen mit Mediencharakteren: Ein theoretischer und empirischer Beitrag. 1. Auflage. Baden-Baden: Nomos (= Reihe Rezeptionsforschung 43).
- Linnemann, Gesa A./Löhe, Julian/Rottkemper, Beate (2023): Bedeutung von Künstlicher Intelligenz in der Sozialen Arbeit: Eine exemplarische arbeitsfeldübergreifende Betrachtung des Natural Language Processing (NLP). In: Soziale Passagen 15, S. 197–211.
- Linnemann, Gesa/Löhe, Julian/Rottkemper, Beate (2024): Bedeutung von Selbstoffenbarungseffekten in quasisozialen Beziehungen mit auf generativer KI basierten Systemen in Settings von Onlineberatung und -therapie. In: e-beratungsjournal.net – Zeitschrift für Onlineberatung und computervermittelte Kommunikation 20(1), Artikel 1, S. 1–21. <https://doi.org/10.48341/9x1s-5y11>
- Nass, Clifford/Brave, Scott (2005): Wired for speech: how voice activates and advances the human-computer relationship. Cambridge, Mass.: MIT Press. (= Computer-human interaction).
- Onnasch, Linda/Jürgensohn, Thomas/Remmers, Peter/Asmuth, Christoph (2019): Ethische und soziologische Aspekte der Mensch-Roboter-Interaktion. Bundesanstalt für Arbeitsschutz und Arbeitsmedizin. https://www.baua.de/DE/Angebote/Publikationen/Berichte/F2369.html?pk_campaign=DOI (Abfrage: 15.06.2025).
- Reinhard, Gaby (2024): Sozialraumorientierung in der Sozialen Arbeit: Ein Arbeits- und Materialbuch für Studium, Lehre und Praxis. 1. Auflage. Stuttgart: Kohlhammer.
- Sandermann, Philipp/Neumann, Sascha (2020): Grundkurs Theorien der Sozialen Arbeit. München: Ernst Reinhardt.
- Schiffhauer, Birte/Remke, Sara (Hrsg.) (2024): Freiheit und Dehumanisierung. Implikationen für die Theoriebildung der Sozialen Arbeit im Kontext digitaler Transformationen. In: Schönig, Werner/Breuer, Marc/Gerards, Marion/Löwenstein, Heiko (Hrsg.): Transdisziplinäre Theorieentwicklung Sozialer Arbeit. Beiträge mit struktureller Perspektive in Zeiten gesellschaftlicher Krisen. Weinheim: Beltz Juventa, S. 177–191.
- Schönig, Werner/Breuer, Marc/Gerards, Marion/Löwenstein, Heiko (2024): Wohin denkst Du? Aktuelle Perspektiven der Theorieentwicklung Sozialer Arbeit. In: Schönig, Werner/Breuer, Marc/Gerards, Marion/Löwenstein, Heiko (Hrsg.): Transdisziplinäre Theorieentwicklung Sozialer Arbeit: Beiträge mit struktureller Perspektive in Zeiten gesellschaftlicher Krisen. Weinheim: Beltz Juventa, S. 7–20.
- Selke, Stefan (2024): Zukunftseuphorie als Trost. Verheißungserzählungen über Künstliche Intelligenz im Kontext gesellschaftlicher Erschöpfungsdiagnosen. In: Heinlein, Michael/Huchler,

- Norbert (Hrsg.): *Künstliche Intelligenz, Mensch und Gesellschaft*. Wiesbaden: Springer Fachmedien Wiesbaden, S. 289–317.
- Singer, Tassilo (2019): *Dehumanisierung der Kriegführung: Herausforderungen für das Völkerrecht und die Frage nach der Notwendigkeit menschlicher Kontrolle*. Berlin und Heidelberg: Springer Berlin Heidelberg.
- Stalder, Felix (2016): *Kultur der Digitalität*. 1. Auflage, Originalausgabe. Berlin: Suhrkamp.
- Staub-Bernasconi, Silvia (2018): *Soziale Arbeit als Handlungswissenschaft: Soziale Arbeit auf dem Weg zu kritischer Professionalität*. 2., vollständig überarbeitete und aktualisierte Auflage. Opladen und Toronto: Barbara Budrich.
- Takagi, Yu/Nishimoto, Shinji (2022): High-resolution image reconstruction with latent diffusion models from human brain activity. <https://biorxiv.org/lookup/doi/10.1101/2022.11.18.517004>
- Tang, Jerry/LeBel, Amanda/Jain, Shailee/Huth, Alexander G. (2023): Semantic reconstruction of continuous language from non-invasive brain recordings. In: *Nature Neuroscience* 26, S. 858–866.
- Weyer, Johannes (2022): *Vermenschlichung der Technik? Die Interaktion von Menschen und künstlicher Intelligenz in alltäglichen Kontexten*. TU Dortmund: Soziologisches Arbeitspapier 60/2022. <https://eldorado.tu-dortmund.de/server/api/core/bitstreams/1318af11-d81a-42ea-8963-2689583ce2f9/content> (Abfrage: 15.06.2025).
- Zlotowski, Jakub/Sumioka, Hidenobu/Eyssel, Friederike/Nishio, Shuichi/Bartneck, Christoph/Ishiguro, Hiroshi (2018): Model of Dual Anthropomorphism: The Relationship Between the Media Equation Effect and Implicit Anthropomorphism. In: *International Journal of Social Robotics* 10, S. 701–714.
- Zweig, Katharina A. (2019): *Ein Algorithmus hat kein Taktgefühl: Wo künstliche Intelligenz sich irrt, warum uns das betrifft und was wir dagegen tun können*. Originalausgabe. München: Heyne.

Künstliche Intelligenz und Ethik – der verantwortliche Umgang mit einer neuen Technik¹

Wolfgang M. Heffels

Abstract: Ethik beschäftigt sich mit der Moral. Hier mit Fragen, welche Regelwerke (soziale Normen) benötigt werden, damit die neue Technologie positive Wirkungen entfalten kann. Hiermit sind schon zwei Aussagen grundgelegt: (1) Die KI-Ethik ist eine Technologie-Ethik; (2) Die konkrete Anwendung von KI-Verfahren im Sozialen und Gesundheitsbereich bedarf einer geregelten Community-Expertise. Mithin wird einer Dämonisierung oder Heroisierung entgegengewirkt und die Technik auf ihre jeweilige Leistungsfähigkeit mit ihren positiven und negativen Aspekten überprüft. Inwieweit die Profession und die jeweiligen Organisationen auf die Gestaltung gesellschaftlicher Normativität einwirken können, bleibt offen. Welche Techniken jedoch vor Ort zum Zuge kommen, ist konkret beeinflussbar und bedarf der Aufklärung der verantwortlichen handelnden Akteure und geht mit einer Qualifizierung der Mitarbeiter:innen und der Schaffung neuer Strukturen und Prozessführungen einher.

Keywords: KI und Ethik; KI und Moral; Verantwortung; Vernunft; Urteilskraft, KI und Regelungsbedarf

Innovative Neuerungen, die sich auf das gesellschaftliche und berufliche Leben auswirken, können sowohl moralisch als auch ethisch problematisiert werden. Die Ethik reflektiert Moral und hat neben einer orientierungsgebenden Funktion zugleich eine Überprüfungsfunktion, mithilfe derer das moralische Handeln hinterfragt und kontrolliert werden kann.

„Die Erwartung besteht besonders dann, wenn neue technische Handlungsoptionen ins Spiel gebracht werden, die unsere moralischen Vorstellungen unsicher oder sogar widersprüchlich erscheinen lassen. [...] Wer also technische Handlungsoptionen beurteilt, muss sich sowohl mit den Techniken selbst als auch mit den für ihre Bewer-

1 © Die Autor:innen 2025, veröffentlicht im Verlag Beltz Juventa, Open Access: CC BY 4.0
Gesa Linnemann / Julian Löhe / Beate Rottkemper (Hg.), *Künstliche Intelligenz in der Sozialen Arbeit*
10.3262/978-3-7799-8562-4_006

tung relevanten normativ-ethischen Aspekten intensiv auseinandersetzen.“ (Battagilla/Mukerji 2015, S. 290).

Der Beitrag „Künstliche Intelligenz und Ethik“ beginnt mit einer sprachphilosophischen Einführung und nimmt die Analogien Mensch und Maschine kritisch in den Blick. Das Moralische im Umgang mit der Künstlichen Intelligenz (KI) als das in bestimmten Kontexten Gelebte, wird dargestellt, bevor dies ethisch analysiert und bewertet wird. Zum Schluss findet ein konstruktiver Vorschlag zum Umgang mit KI-Systemen aus verantwortungsethischer Sicht statt.

1 Worüber sprechen wir, wenn wir von „Künstlicher Intelligenz“ sprechen?

Die Sprachphilosophie ist eine Disziplin der Philosophie, die sich mit Sprache und Bedeutung beschäftigt, vor allem mit dem Verhältnis von Sprache und Wirklichkeit sowie dem Verhältnis von Sprache und Bewusstsein bzw. dem Denken (vgl. Willaschek 1995, S. 164–174). Sprachphilosophisch beinhaltet die Wortkombination „künstliche Intelligenz“ zwei wesentliche Bestimmungen: (a) mit „künstlich“ wird ein nicht menschlicher, sondern technisch konstruierter Vorgang der Datenverarbeitung zum Ausdruck gebracht, der (b) über einen programmierten Vorgang eine Problemlösungsfähigkeit vollzieht, die als Intelligenz bezeichnet wird, aber in Wirklichkeit ein mechanistisches Geschehen darstellt. Faktisch werden durch technische Algorithmen, unabhängig von Menschen, Produkte (wie beispielsweise Text, Bilder, Video, praktisches Vorgehen [autonomes Fahren]) erzeugt. Insofern wird mit dem Begriffspaar „künstliche Intelligenz“ eine Metapher geschaffen, die ein nicht existierendes Verhältnis zwischen Maschine und Mensch herstellt. Analog und wechselseitig umgekehrt wird beispielweise das Kurzzeitgedächtnis des Menschen mit einem Arbeitsspeicher und das Langzeitgedächtnis mit der Festplatte eines Computers verglichen, oder die Arbeitsweise von KI wird mit dem neuronalen Netzwerk eines Menschen sprachlich ins Bild gesetzt. Diese sprachmedialen Analogien oder Metaphern erleichtern und verfälschen das Verstehen. Nach Wittgenstein werden hier Sprachhandlungen vollzogen, die wechselseitig unterschiedliche Dinge und Vorgänge miteinander in Verbindung bringen, um ein besseres Verstehen und Verständnis zu erreichen (vgl. Wittgenstein 2001, § 26; Savigny 1995, S. 75–90) – ein Sprachspiel der Komplexitätsreduktion mit einer faktischen Unvergleichbarkeit. Mensch und Maschine sind radikal anders (siehe Tabelle 1).

Die irreführende Sprache, die unserem Verstand eine nicht zutreffende Wirklichkeit vorgaukelt, als ob Maschinen wie Menschen intentional eingestellt, reflexiv und mit einer eigenen Bewusstheit und Wertgebundenheit operieren könn-

Tabelle 1: Mensch und Maschine

Mensch (lebendiger Organismus)	Computer (semiotische Maschine)
<ul style="list-style-type: none"> ● sich durch Autopoiese in Stoffwechsel ● und Kommunikation selber machend ● autonom (selbstbestimmende Regeln) ● handelt intentional (kontingent) ● ist sprachbegabt, reflexiv lernfähig ● lebendiges Arbeitsvermögen: ● können (implizites Wissen, Erfahrungen, situierte Urteilskraft und Handlungskompetenz) verausgabt und reproduziert sich im Gebrauch 	<ul style="list-style-type: none"> ● wissensbasiert für bestimmte Zwecke gemacht (konstruiert) ● automatisch (auto-operational selbsttätig) ● verhält sich kausal determiniert ● ggf. algorithmisch gesteuert Umwelt adaptiv ● algorithmisch determiniertes Verhalten: setzt Formalisierungen von Zeichenprozessen voraus, muss für Praxis angeeignet und organisatorisch eingebettet werden

Quelle: Eigene Darstellung in Anlehnung an Brödner 2019, S. 93

ten, verschleiert, dass Maschinen nicht denken, fühlen, empfinden oder gar eine Sinnhaftigkeit konstruieren können (vgl. Wittgenstein 2001, S. 109). Insofern sprechen wir von Maschinen, die über Operationsweisen verfügen, die auf einen Zweck oder mehrere Zwecke hin konstruiert wurden und in Arbeitsprozesse eingebunden sind bzw. werden müssen. Die Integration neuer Arbeitsweisen führt zu Veränderungen der Arbeitsprozessgestaltung, des sozialen Miteinanders, hat ökonomische und ggf. ökologische Folgen und ist ethisch dem Bereich Technikethik mit Auswirkungen auf die Sozial- und Umweltethik zuzuordnen.

Sprachlich ist auch feststellbar, dass sich die klassische Subjekt-Objekt-Gegenüberstellung verändert hat. Mittels einer zunehmend anthropomorphen Sprache, in der die Maschine mit dem Menschen verglichen wird, die schneller, immer verfügbar und unabhängig von Stimmungen präzise ihre Arbeit vollzieht, bekommt KI einen menschenähnlichen Status. Ein Objekt wird subjektiviert und Denkprozesse werden initiiert, die den Menschen mit einer Maschine in einem Bewertungszusammenhang (besser/schlechter) stellen (vgl. Grunwald 2021, S. 312–331).

2 Was ist das „Moralische“ im Umgang mit Maschinen, die autonom Produkte herstellen?

Der Begriff des Moralisch-seins kann in drei Varianten in Erscheinung treten:

- a) Ich handle moralisch, wenn mein Tun einer anerkannten Verhaltensnorm entspricht.
- b) Ich handle moralisch, wenn ich meinen eigenen Werten, Normen, Prinzipien folge.
- c) Ich handle moralisch, wenn ich von einer anderen Person erwarte, dass sie bestimmte Verhaltensnormen einhalten soll.

Während sich (a) und (c) auf Moral beziehen, also auf die ungeschriebenen und geschriebenen Regelwerke des Miteinanders in kontextualisierten Lebensbereichen, stellt (b) den Bereich der Moralität eines Menschen mit seinen Wert- und Sinnbezügen dar. Diese skizzierte Differenzierung soll verdeutlichen, dass – wird vom „Moralisch-sein“ gesprochen – der Blick entweder auf die Moral, die Regelwerke des Miteinanders, oder auf die Art und Weise der individuellen Bewertung und Positionierung (Moralität) gerichtet wird. An einem kleinen Beispiel kann dieser Zusammenhang demonstriert werden: Eine Person, die gerne sauniert (ein innerer Wert – Moralität [b]) und die Saunaregeln als richtig, wichtig und einhaltungspflichtig ansieht (Moral [a]), verhält sich entsprechend den Regeln. Ihre Moralität entspricht der Moral und sie verhält sich moralisch. Spricht diese Person eine andere Person darauf an, dass sie doch bitte die Saunaregel befolgen solle, dann ist sie moralisch, weil sie von einer anderen Person ein bestimmtes Verhalten einfordert (vgl. Heffels / Storm 2021, S. 41–65).

In jedem Lebensbereich, in dem Menschen miteinander kooperieren – in der Schule, am Arbeitsplatz, im Straßenverkehr, in der Familie, beim Einkaufen usw. – gibt es Regeln, die das Miteinander betreffen, um ein störungsfreies Miteinander zu ermöglichen. Neuerungen verändern die Regelwerke des Miteinanders, z. B. die Etablierung von Radwegen oder die Verwendung von Smartphones. Diese Veränderungen führen zum sogenannten Wertewandel (vgl. Koch 2011), d. h., die Moralität von Menschen kann sich durch neue Möglichkeiten und einer sich damit modifizierenden Moral verändern. In demokratischen Gesellschaften besteht eine Pluralität der Werthaltungen mit einer gewünschten Moralbildung der Toleranz. So ist das Telefonieren in einem Bahnabteil zu tolerieren, wenn es nicht im Ruhebereich stattfindet.

In der Konkretisierung der moralischen Dimension der Anwendung von Maschinen, die autonome Produkte herstellen (wie beispielsweise Texte, Bilder, kooperative Arbeitsformen, automatisierte Antragsbearbeitungsverfahren) werden exemplarisch vier unterschiedliche Praxen mit je einem spezifischen Beispiel vorgestellt, um daraus das, was Moral meint, abzuleiten. Die Moralität, also wie der:die Einzelne sich hierzu verhält und positioniert, ist individuell unbestimmbar und bleibt deshalb offen.

Das erste Beispiel: KI-unterstützte Entscheidungsfindung zur Früherkennung einer Kindeswohlgefährdung

Ein Forscherteam aus drei Hochschulen führte das KAIMo-Projekt („Kann ein Algorithmus moralisch kalkulieren?“, Burghardt 2024) durch: Zentral geht es darum, ob ein KI-gestütztes Verfahren die Arbeit von Sozialarbeiter:innen zur Erkennung einer Kindeswohlgefährdung (Konflikt zwischen dem Wohl des Kindes und der Freiheit der Erziehung durch die Eltern) erleichtern kann. Im Projekt wurde ein Tool entwickelt, mithilfe dessen Sozialarbeiter:innen reale Verdachtsfälle daraufhin untersuchen können, ob eine Kindeswohlgefährdung vorliegt. „Die Fach-

kräfte erhielten die Möglichkeit, die Beratung eigenständig mit Unterstützung der KI-Assistenz durchzuführen. Ziel war, die gemeinsame Einschätzung des Gefährdungsrisikos für das im Fall benannte Kind“ (ebd., S. 305) durchzuführen. Im Ergebnis beschrieben die Fachkräfte, dass das Tool „ein sehr gutes Werkzeug [ist], um die eigenen Gedanken zu sortieren und komplexe Zusammenhänge gut zusammenzufassen [...] Aber auch die hinweisspezifischen Plausibilitätsprüfungen und Folgenabschätzungen wurden als unterstützend für den Diskussionsprozess beschrieben“ (ebd.).

Im Forschungsprojekt stellten die Initiatoren zwei Gefahrenstellen und eine völlig neue Kompetenz heraus: Die erste Gefahrenstelle ist der Framing-Effekt, d. h., je nachdem wie die Frage formuliert ist, kann die Antwort beeinflusst werden. Die zweite Gefahrenstelle ist der sogenannte Automation Bias, der dazu führt, Vorschläge von automatisierten Systemen bevorzugt anzunehmen. Die neue Kompetenz erfordert von den Fachkräften eine metakognitive Fähigkeit, die Entstehung der maschinellen Arbeitsweise nachvollziehen und überprüfen zu können.

Abschließend kommt das Forschungsteam zu der Aussage:

„Durch den im Projekt KAIMo entwickelte Ansatz konnte eine verantwortungsvolle Balance zwischen innovativer Technologie und den (berufs-)ethischen Grundsätzen gefunden sowie die Erhaltung der menschlichen Steuerung und Kontrolle erreicht werden.“ (ebd., S. 307)

Das zweite Beispiel: KI in der ambulanten Pflege

Der ambulante Pflegesektor ist ein politisch gewollter Innovationsbereich der KI-Anwendung. Maibaum, Bischof und Hergesell (2023) ermittelten im Rahmen einer Studie folgende Gründe dafür, warum KI-Systeme im ambulanten Pflegesektor nicht so ankommen, wie gewünscht:

Erstens: „Die KI-Projekte, die aus wettbewerbsbasierten Förderlogik entstehen, sind nur auf die Erprobung ausgerichtet und nicht auf eine nachhaltige Realisierung im pflegerischen Alltag – was neben einsatzbereiter Technik weiterreichende Transformationen im Alltag erfordern würde“ [...] Zweitens führt die Zerlegung von Pflegepraktiken in Tasks zu einer Dekontextualisierung der ursprünglichen, natürlichen (Pflege)Situation [...] die zunächst einmal der Technikgenese dient und (noch) nicht der avisierten Einsatzsituation“ (ebd., S. 12 f.).

Die Beachtung des Feldes mit seiner pflegerischen Handlungslogik und seinen spezifischen Eigenarten ist für die Autoren die *Conditio sine qua non* (notwendige Bedingung) für die Entwicklung und Implementierung neuer Technologie im Handlungsfeld Pflege. Der:die zu Pflegenden, die An- und Zugehörigen und die

Pflegefachpersonen müssen einen Benefit wahrnehmen, damit KI-Systeme Anwendung finden (vgl. ebd., S. 16 f., 19).

In der Betrachtung dieser Beispiele kann zunächst festgestellt werden, dass KI-Systeme in ihren Anwendungen mit einem Effizienzansatz (Wirksamkeit – gute Ergebnisse) und einem Effektivitätsansatz (wirtschaftlich und ressourcenschonend) zu verbinden sind. Im zweiten Beispiel, der ambulanten Pflege, wurde dies besonders deutlich, weil die Anwender:innen die Wirksamkeit (erzeugt gute Ergebnisse) und die Ressourcenoptimierung nicht erlebt haben und deshalb die KI-Unterstützung in der jetzigen Form ablehnen.

Findet die KI-Anwendung in unterschiedlichen Kontexten statt, dann treten mindestens vier Veränderungen auf:

- Mindestens ein Arbeitsprozess wird durch ein KI-System unterstützt.
- Die ausführende Person benötigt andere Kompetenzen zur Steuerung und Ausführung des neuen Arbeitsprozesses.
- Die Arbeitskoordination zwischen dem neuen Arbeitsprozess und den daneben ablaufenden Arbeitsprozessen bedarf einer anderen Handlungskoordination.
- Der Umgang mit dieser Technologie bedarf neuer Verfahrensvorschriften und verändert Strukturvorgaben.

Ob und inwieweit durch diese neuen Technologien Arbeitsplätze eingespart oder neu geschaffen werden, bleibt abzuwarten.

Die Einführung von KI-Systemen in Organisationen verändert das moralische Miteinander dergestalt, dass sich die Mitarbeiter:innen mit dieser neuen Technologie auseinandersetzen (müssen), Fort- und Weiterbildungsmaßnahmen ansteuern (werden) und sich engagieren, das Neue in den Gesamtbetrieb zu integrieren oder es zu verhindern (versuchen). Die Organisation ist gefordert, sich weiterzuentwickeln. Strukturen und Prozesse müssen zur erwünschten Output-Orientierung neu justiert werden. Evaluationen im Sinne einer Erfassung, was das Neue bewirkt, verursacht und welche Nebenwirkungen auftreten, sind systematisch umzusetzen, um unerwünschte Wirkungen zu minimieren. Diese Selbstvergewisserung dient der sinnhaften Ausrichtung eines effektiven und effizienten Technikeinsatzes und erfordert u. a. die Entwicklung eines kritischen (Diskurs-)Verfahrens der Mitarbeitenden untereinander und im Umgang mit den Verfahren und seinen Ergebnissen der neuen Technologie.

Was ist das „Ethische“ im Umgang mit Maschinen, die automatisiert Produkte herstellen?

Die Ethik ist eine Wissenschaft, die sich mit dem Gegenstand der Moral und der Moralität beschäftigt. Moralen beinhalten Ideologien als orientierungsgebende Sinnbilder und Normen, die das Zusammenleben in einer Gruppe und die Art und Weise, wie Gruppen miteinander umgehen, bestimmen. Dieses Miteinander ba-

siert auf gelebten und geschriebenen (beispielsweise Gesetze, Verfahrensanweisungen, Ordnungen) und gelebten und ungeschriebenen (beispielsweise Konventionen, Sitten, Gewohnheiten) Regeln des Miteinanders. Die Moralität hingegen bezieht sich auf den einzelnen Menschen mit seinen Wert- und Normsetzungen (Einstellungen, Haltungen, Prinzipien) sowie seinen Orientierungen. Die Disziplin Ethik beinhaltet drei Bereiche:

- a) Die Metaethik ist eine Teildisziplin der Ethik, die sich mit den ethischen Aussagen der ethischen Theorien auseinandersetzt. So beschäftigt sie sich z. B. mit folgenden Fragen: Was ist ethisch richtig oder falsch? Wie kommen die einzelnen ethischen Theorien zu gerechten Entscheidungen? Oder gibt es objektive Werte? Welche Wechselwirkung besteht zwischen Sprache und Denken? (Worüber sprechen wir, wenn wir von „Künstlicher Intelligenz“ reden?)
- b) Der empirische Teilbereich der Ethik beschäftigt sich mit der Erfassung und Bewertung des kulturell gelebten Miteinanders von Individuen, von Organisationen und der Gesellschaft (vgl. Luhmann 2005, S. 9–24). Dieser Teilbereich erfasst das Moralsystem des jeweiligen Beobachtungsgegenstands. Das gelebte Miteinander kann der Ausgangspunkt einer ethischen Betrachtung darstellen. Insofern stellen die Aussagen, wie das gelebte Miteinander normativ geregelt ist und gelebt wird, den deskriptiven empirischen Teilbereich der Ethik dar.
- c) Der philosophische Teilbereich der Ethik beschäftigt sich mit Fragen der Urteilsfindung, Bewertung von Handlungen und Prozessen in Hinblick darauf, was angemessen, gut oder gerecht ist. Die philosophische Ethik sucht nach Antworten auf Fragen, die auch das Leben stellt, steht indirekt zum Leben und nimmt zur Suche nach guten Gründen eine Metaperspektive ein (Löwisch 1995, S. 55–60). Immanent hierin ist die Differenz zwischen Sein und Sollen und die Frage nach der Verantwortlichkeit handelnder Menschen in Interaktionen, in Organisationen oder zur Mitgestaltung gesellschaftlicher Rahmenbedingungen. Dies bildet den sozialetischen Anteil ab, während die Verantwortlichkeit des einzelnen Menschen für seine innere Ausrichtung (Was ist mir wichtig?), seine Selbstbestimmung (Wer will ich sein?) und seine Präferenzen für Mitmenschen, Bedürfnisbefriedigungen und Umwelt die individuelle ethischen Anteile der Ethik abbildet.

Das „Ethische“ nimmt Bezug auf die Ethik, also im Bild gesprochen, auf etwas, das oberhalb der Moral und Moralität liegt. Es übersteigt durch Reflexion das moralisch Vorhandene und bezieht sich auf ethische Aussagen. Was heißt und bedeutet das konkret für den Umgang mit KI-Systemen?

So ist jeder Mensch beispielsweise verpflichtet, Hilfe zu leisten, wenn andere Menschen in Not geraten sind. Diese Norm kann moralisch u. a. mit der „goldenen Regel“ beantwortet oder auf eine gesetzliche Bestimmung zurückgeführt werden. Da der Mensch einzigartig und schützenswert ist, er Würde hat, ist die

Unterstützung eines hilfebedürftigen Menschen eine (moralische) Pflicht, die das (Über-)Leben sichert. Diese oberste Pflicht wird durch eine weitere Pflicht, den respektvollen Umgang im sozialen Miteinander, ergänzt.

Ethisch bedeutet damit im Besonderen in Hinblick

- auf die Metaethik, die Begriffe und Theorien der Ethiken zu rekonstruieren,
- auf die empirische Ethik, das Moralsystem, so wie es ist, zu analysieren, und
- für die philosophische Ethik, Orientierung zu ermöglichen und die ethische Urteilsbildung zu unterstützen.

Eine erste ethische Frage lautet: Was hat KI mit Ethik zu tun? Zunächst ist KI eine computergesteuerte Technik mit Algorithmen, die von Menschen auf ein Ziel/einen Zweck hin konstruiert wurden. In dieser allgemein gestellten Frage sind zwei Unterfragen zu beantworten: (a) Ist der Zweck ethisch begründbar und (b) sind die Mittel (hier das jeweilige KI-System) verhältnismäßig zur Erreichung der jeweiligen Ziel- /Zweckdimension?

Zur Beantwortung der ersten Frage (Wie kann ein Zweck ethisch begründet werden?) hat sich Immanuel Kant in seiner Grundlegung der Metaphysik der Sitten im Kategorischen Imperativ (Grundformel) dergestalt geäußert, dass es neben dem Selbstzweck (Selbstzweck-Formel) einer Reich-der-Zwecke-Formel bedarf.

Grundformel: Handle nur nach derjenigen Maxime, durch die du zugleich wollen kannst, dass sie ein allgemeines Gesetz werde. (Kant GMS BA 52)

Selbstzweck-Formel: Handle so, dass du die Menschheit sowohl in deiner Person als auch in der Person eines jeden anderen jederzeit zugleich als Zweck, niemals bloß als Mittel brauchst. (Kant GMS BA 66)

Reich-der-Zwecke-Formel: Handle so, als ob du durch deine Maxime jederzeit ein gesetzgebendes Glied im allgemeinen Reich der Zwecke wärest. (Kant Immanuel GMS BA 83)

Damit ist gemeint, dass der Zweck, also das, was angestrebt wird, einer grundsätzlichen Prüfung unterzogen werden muss, bevor die Technik allgemein eingeführt wird. Ethisch sind hiernach mindestens zwei Prüfungsfragen zu stellen:

- Kann der Zweck der Technik – analog wie ein Gesetz – von den meisten Menschen (in der jeweiligen Gesellschaft) vertreten werden?
- Kann mit dem Zweck die Förderung eines allgemeinen Gutes für Mensch und Umwelt verbunden werden?

Hans Jonas hat diese von Kant erstellte Reich-der-Zwecke-Formel mit einer neuen Dimension erweitert. Er greift die Problematik des technologischen Fortschritts

mit einem erweiterten ethischen Lösungsansatz unter dem Titel „Prinzip Verantwortung“ (1979) auf. Seine Ausgangspositionen, die ihn bewegten, waren, (a) dass durch die neuen Technologien auch nachfolgende Generationen betroffen sind, (b) die Folgen der Eingriffe ungewiss und nicht gänzlich kalkulierbar sind und (c) eine Revisionsmöglichkeit, wenn überhaupt, nur eingeschränkt vorliegt. Jonas formulierte einen erweiterten Kategorischen Imperativ der Reich-der-Zweck-Formel: „Handle so, dass die Wirkungen deiner Handlung verträglich sind mit der Permanenz echten menschlichen Lebens auf Erden“, oder negativ ausgedrückt: „Handle so, dass die Wirkung deiner Handlungen nicht zerstörerisch sind für die zukünftigen Möglichkeiten solchen Lebens (Jonas 1979, S. 26)

Jonas erweitert mit diesem Imperativ zum einen den Blick vom Zweck, als positive Umschreibung, hin zu den Gefahren und den Risiken, die mit neuen Technologien verbunden sind. Zum anderen führt er das kritische Denken als treibende Kraft ein, indem er von der Notwendigkeit einer „heuristischen Furcht“ spricht. Das heißt: nicht Technikgläubigkeit, dem jeweils Neuen hinterherzulaufen, sondern zum Schutz von Menschheit und Umwelt eine kritische Einstellung einzunehmen, das Neue auf seine Risiken hin zu bewerten. Die ethische Beurteilung neuer Technologien bedarf auch einer „Heuristik der Furcht“, das ist, ein Erkennen-Wollen von Gefahren und Risiken und nicht nur der Nutzen oder Kosten.

Konkret, auch unter Berücksichtigung der aufgeführten Praxisbeispiele und ohne Anspruch auf Vollständigkeit, werden fünf Zwecke von KI-Systemen übergeordnet aufgeführt:

- Optimierungsfunktion: Der Einsatz dieser Technik geht damit einher, dass ein Produkt schneller und qualitativ besser hergestellt werden kann als durch andere Vorgehensweisen (beispielsweise Text- oder Bildgeneratoren).
- Assistenzfunktion: Fachkräfte können in diagnostischen und therapeutischen Verfahren (im weitesten Sinne) entlastet werden, weil die programmierten Maschinen (teil)standardisierte Funktionen übernehmen können (beispielsweise die automatisierte Auswertung von Laborbefunden).
- Präzisionsfunktion: Maschinen können zur Ausführung von Verrichtungen so programmiert werden, dass sie exakte und sorgfältige Wiederholbarkeit im Verfahren produzieren, Ungenauigkeiten vermieden und Menschen unterstützt werden (beispielsweise computerunterstützte medizinische Eingriffe).
- Diskurs- und Entscheidungsgrundlage-Funktion: Mitarbeitende können die Ergebnisse der programmierten Maschinen nutzen, um ihre Entscheidungen zu treffen oder kollaborative Entscheidungsverfahren durchzuführen (vgl. Kindeswohlgefährdung).
- Bildungsfunktion: Die „neue Technik“ kann auch als Medium in doppelter Hinsicht genutzt werden, (a) um sich mit dem Gegenstand dieser Technik auseinanderzusetzen und (b) um einen kritischen und weiten Blick im Umgang

mit dieser Technik zu entwickeln. Die von Adorno (1959, S. 93–121) beschriebene Halbbildung – d. h. nur die Vermittlung von Fähigkeiten im Umgang mit dieser neuen Technik – gilt es durch die Erweiterung des Gegenstands der gesellschaftlichen, sozialen und historischen Bedeutung sowie durch eine Auseinandersetzung mit epochalen Ereignissen der Technikrevolution (Klafki 1996, S. 43–81) zu vermeiden (Galla 2024, S. 161–187).

In der Beantwortung der zweiten Frage geht es darum, ob das Mittel (hier das jeweilige KI-System) dem jeweiligen Zweck entsprechen kann und ob das Mittel verhältnismäßig ist, d. h., ob die Vorteile überwiegen und nicht durch vorhandene Alternativen ersetzbar sind.

Insofern lassen sich zwei Bereiche, ein allgemeiner Bereich (1) und ein kontextueller oder spezifischer Bereich (2) differenzieren:

(1) Entsprechen die Mittel den allgemeinen moralischen Anforderungen der Gesellschaft, d. h., entsprechen sie den allgemeinen und gesetzlichen Schutzbestimmungen, die innerhalb einer Gesellschaft für den Menschen und die Umwelt bestehen (Grundgesetz, Datenschutz-Grundverordnung, Allgemeines Gleichbehandlungsgesetz, Nachhaltigkeitsbestimmungen für die Umwelt und den Energieverbrauch, Instrumentalisierungsverbot, Betrugsverbot usw.)? Wertäußerungen in diesem Bereich werden oft mit Normaussagen verbunden. Zur Verdeutlichung werden hier beispielweise die KI-Leitsätze von Asilomar vorgestellt. In der Asilomar-Konferenz 2017 entwickelten 1.000 Expert:innen Leitsätzen, die bei Konstruktion und Anwendung von KI-Systemen Beachtung finden sollen:

„6) **Sicherheit:** KI-Systeme sollten während ihrer gesamten Funktionszeit sicher sein, und dies soweit anwendbar, möglichst auch nachweislich.

7) **Transparenz bei Fehlfunktionen:** Falls ein KI-System Schaden anrichtet, muss es möglich sein, die Ursache ermitteln zu können.

8) **Transparenz bei Rechtsprechung:** Bei der Einbindung autonomer Systeme in jegliche entscheidungsfindenden Prozesse der Rechtsprechung sollten diese Prozesse nachvollziehbar und von einer kompetenten menschlichen Autorität überprüfbar sein.

9) **Verantwortung:** Entwickler und Ingenieure von fortgeschrittenen KIs haben sowohl die Gelegenheit als auch die Verantwortung, die moralischen Folgen von Gebrauch, Missbrauch und eigenständiger Handlungen dieser Systeme mitzubestimmen.

10) **Wertausrichtung:** Stark autonome KI-Systeme sollten so entwickelt werden, dass ihre Ziele und Verhaltensweisen während des Betriebs unter fester Gewissheit auf menschliche Werte ausgerichtet sind.

- 11) **Menschliche Werte:** KI-Systeme sollten so entwickelt und bedient werden, dass sie mit den Idealen der Menschenwürde, Menschenrechten, Freiheiten und kultureller Vielfalt kompatibel sind.
- 12) **Privatsphäre:** Im Hinblick auf die Fähigkeit von KIs, Daten zu analysieren und weiterzuverarbeiten, sollten Menschen das Recht haben, Zugriff auf ihre generierten Daten zu haben und in der Lage sein, sie zu verwalten und zu kontrollieren.
- 13) **Freiheit und Privatheit:** Die Anwendung von KIs auf persönliche Daten darf die tatsächlichen oder wahrgenommenen Freiheiten der Menschen nicht auf unangemessene Weise einschränken.
- 14) **Geteilter Nutzen:** KI-Technologien sollten so vielen Menschen wie möglich dienen und nutzen.
- 15) **Geteilter Wohlstand:** Der wirtschaftliche Wohlstand, der von KIs geschaffen wird, sollte breit verteilt werden, sodass er der ganzen Menschheit nutzt.
- 16) **Menschliche Kontrolle:** Wenn es um Aufgaben geht, die von Menschen erdacht worden sind, sollten Menschen bestimmen können, ob und inwiefern Entscheidungen an KI-Systeme delegiert werden können.
- 17) **Kein Umsturz:** Die Macht, die durch die Kontrolle hoch entwickelter KI-Systeme gewährt wird, sollte die sozialen und bürgerlichen Prozesse, auf denen das Wohlergehen der Gesellschaft beruht, respektieren und verbessern, aber nicht untergraben.
- 18) **KI-Wettrüsten:** Ein Wettrüsten von tödlichen autonomen Waffen sollte vermieden werden“ (Asilomar Konferenz 2017).

(2) Ein anderer, zweiter Ansatz ist der, dass die Zweckentsprechung und Verhältnismäßigkeit nicht allgemein, sondern kontextbezogen geprüft werden, d. h., ob die Mittel innerhalb eines bestimmten Kontextes (Gesundheit, Soziales, Pflege, Bildung, Industrie, Unterhaltung) den allgemeinen Vorgaben entsprechen. Hierzu hat der Deutsche Ethikrat (2023, S. 11) „im zweiten Teil [...] vier ausgewählte Handlungsfelder exemplarisch konkretisiert: der Medizin, der schulischen Bildung, der öffentlichen Kommunikation und Meinungsbildung sowie die öffentliche Verwaltung“. Zunächst wird am Beispiel der Medizin dargestellt, auf welchem Niveau die Aussagen getätigt werden, bevor die allgemeinen Aussagen im Schlussteil dargestellt und bewertet werden. Für den medizinischen Bereich formuliert der Deutsche Ethikrat (vgl. ebd., S. 30 ff.):

- Die Genehmigung von KI-Produkten bedürfen einer engen Zusammenarbeit mit den relevanten Zulassungsbehörden und Fachgesellschaften.
- Bei der Entwicklung und Verwendung von KI-Systemen sollen über die bestehenden Rechtspflichten Dokumentationspflichten eingefordert werden.
- Die Notwendigkeit einer reflexiven Plausibilitätsprüfung der jeweiligen Verfahren und Ergebnisse von KI-Systemen soll eingeführt werden.
- KI-Anwendungen bedürfen einer Integration in die klinische Ausbildung des ärztlichen Fachpersonals.

- Die verstärkte Nutzung von KI-Komponenten in der Versorgung darf nicht zu einer weiteren Abwertung des kommunikativen Miteinanders zwischen Patient:in und Ärzt:in oder einem Abbau von Personal führen.

Zusammenfassend für alle Bereiche kommt der Deutsche Ethikrat u. a. zu folgenden Aussagen:

„Der Einsatz KI-gestützter digitaler Techniken ist im Sinne **der Entscheidungsunterstützung und nicht der Entscheidungsersetzung** zu gestalten, um Diffusion von Verantwortung zu verhindern“ (ebd., S. 350, Hervorhebung W. H.).

„Neben [...] dem Schutz der Privatsphäre oder die Verhinderung von Diskriminierung, [...] gilt, dass **Einzelfallbeurteilungen** grundsätzlich wichtig bleiben“ (ebd., S. 354, Hervorhebung W. H.).

„Mit Blick auf KI-Anwendungen müssen neue Wege gefunden werden, um innerhalb der **jeweiligen Kontexte** und bezüglich der **jeweils spezifischen Herausforderungen** und Nutzenpotenziale die gemeinwohlorientierte Daten(sekundär)nutzung zu vereinfachen bzw. zu ermöglichen und damit die Handlungsoptionen auf diesem Gebiet zu erweitern“ (ebd., S. 365, Hervorhebung W. H.).

„Um die Autorschaft menschlicher Akteure und deren **Handlungsmöglichkeiten** zu **erweitern**, müssen die Resilienz soziotechnischer Infrastrukturen gestärkt und die Abhängigkeit von individuellen Akteuren und Systemen minimiert werden“ (ebd., S. 369, Hervorhebung W. H.).

„Zum Schutz vor Diskriminierung in Anbetracht der zuvor dargelegten Herausforderungen **bedarf es angemessener Aufsicht und Kontrolle** von KI-Systemen. Besonders in sensiblen Bereichen erfordert dies den Auf- oder Ausbau gut ausgestatteter Institutionen“ (ebd., S. 375, Hervorhebung W. H.).

„**Es bedarf der Entwicklung ausgewogener aufgaben-, adressaten- und kontextspezifischer Standards für Transparenz, Erklärbarkeit und Nachvollziehbarkeit und ihrer Bedeutung für Kontrolle und Verantwortung sowie für deren Umsetzung durch verbindliche technische und organisatorische Vorgaben**“ (ebd., S. 379, Hervorhebung W. H.).

Auffällig hierbei ist, dass der Deutsche Ethikrat eine „Nicht-Fisch-und-nicht-Fleisch-Position“ eingenommen hat, indem er überwiegend allgemein und nur ansatzweise und rudimentär kontextbezogen argumentiert. Möglicherweise liegt hier der Versuch vor, trotz Kontextualisierungsversuch eine allgemeine und universelle Moral für KI ethisch anzuregen. Dieses Vorhaben kann bei einer

weltweit agierenden Technologie in pluralistischen Gesellschaften mit einer Vielzahl an speziellen Möglichkeiten von KI nicht hinreichend erfolgreich über eine wertheethische Normbestimmung funktionieren, werden doch die jeweiligen spezifischen und signifikanten Fragen nicht adäquat einer Antwort zugeführt.

Welche Bedeutung könnte der Verantwortungsethik zur Steuerung im Bereich der KI-Technik zukommen?

Verantwortlichkeit ist eine Zuschreibung an handelnde Menschen, die aufgrund ihrer Handlungsmächtigkeit und Zuständigkeit für einen bestimmten Bereich Entscheidungen so treffen, dass sie diese gewählte Entscheidung (Handlungsbestimmung) und/oder die ausgeführte Handlung (Handlungsrechtfertigung) mit guten Gründen sich selbst und anderen gegenüber vertreten können. Verantworten beinhaltet auch die Haltung, das Bestmögliche in einer gegebenen Situation erreichen zu wollen (vgl. Heffels 2023, S. 85–89). Verantwortlich können Menschen gemacht werden, das sind die Wissenschaftler:innen, die Konstrukteur:innen, die Hersteller:innen und diejenigen, die die Maschine warten und anwenden. Dieser Personenkreis agiert zugleich in einem moralischen oder globalisierten Kulturkreis, der die Rahmenbedingungen (Gesetze, Verordnungen, Gepflogenheiten als ungeschriebene Gesetze) mitbeinhaltet.

Aus verantwortungsethischer Sicht können zwei unterschiedliche Ebenen unterschieden werden:

(1) Die gesellschaftliche Ebene umfasst neben dem Politsystem auch das Wissenschaftssystem, in dem die neuen Technologien entwickelt und konstruiert werden. Das Politsystem fördert diese Entwicklungen, und zugleich sind Regelwerke (Gesetze/Verordnungen) – zum Schutz der Bevölkerung – zu verabschieden. Die Europäische Union hat ein KI-Gesetz nach einer Risikobewertung erlassen (vgl. EU-KI 2024; siehe dazu den Beitrag von Dötterl in diesem Band). Der Kerngedanke dieses Gesetzes ist, „eine vertrauenswürdige KI in Europa und darüber hinaus zu fördern, indem sichergestellt wird, dass KI-Systeme die Grundrechte, die Sicherheit und die ethischen Grundsätze achten, und indem die Risiken sehr leistungsfähiger und wirkungsvoller KI-Modelle angegangen werden“. In der Risikoeinschätzung unterteilt die EU vier Risikobereiche: Verbot, hoher Regelungsbedarf, mittlerer Regelungsbedarf und minimaler Regelungsbedarf.

(2) Neben diesen Regelwerken sind kooperative Verantwortungszuschreibungen zwischen Hersteller:innen, Anwender:innen und Instandhalter:innen notwendig. Faktisch besteht ein konkreter Handlungsbedarf darin, die unterschiedlichsten KI-Anwendungen in ihren jeweiligen Sektoren so zu gestalten, dass die kooperativ zu gestaltenden Verantwortlichkeiten definiert und bestimmt werden, um die Gefahren und Risiken (vgl. Nida-Rümelin 1994, S. 809; Meckel/Steinacker 2024,

S. 319–340) zu minimieren und Verantwortlichkeiten zu klären. Wer trägt welche Verantwortung und in welchem Umfang? Sind es diejenigen,

- die das Programm der KI entwickelt haben?
- die das Programm im Handel vertreiben?
- die das Programm in Einrichtungen anschaffen?
- die das Programm in spezifischen Kontexten mit einer Aufgabenstellung einsetzen?
- die das Gerät und das Programm warten?
- die den Prozess/das Ergebnis der KI kontrollieren und umsetzen?

Innerhalb dieses Rahmens sind Organisationen in den unterschiedlichsten Praktiken (Bildung, Gesundheit, Soziale Arbeit, Pflege usw.) unter Beachtung ihrer eigenen Sinnbestimmung und der EU-Kriterien sowie der Gesetze gefordert, KI-Anwendungssysteme zu implementieren. Hierzu bedarf die Einführung der Entwicklung eines der Risikoeinschätzung angemessenen kooperativen Evaluationsverfahrens mit Fachvertreter:innen aus Wissenschaft, Informationstechnik, Berufsverbänden und Ethik (vgl. Grunwald 2002). Über eine zu etablierende Begleitforschung, vor allem bei KI-Verfahren mit hohem Risiko, sollten die Prozesse und Ergebnisse einzelner Verfahren systematisch erfasst, gebündelt und aufgearbeitet werden, sodass sichere KI-Verfahrenstechniken entstehen können.

Auf der Handlungsebene ist das einzelne Subjekt im Umgang mit der KI gefordert, Daten korrekt einzugeben, die Maschine und das Programm ordnungsgemäß zu bedienen und die Ergebnisse einer kritischen Bewertung zuzuführen. Faktisch erfordert letzteres eine ethisch-methodische Kompetenz mit reflexiver Urteilsbildung.

„Urteilstkraft überhaupt ist das Vermögen, das Besondere als enthalten unter dem Allgemeinen zu denken. Ist das Allgemeine (die Regel, das Prinzip, das Gesetz) gegeben, so ist die Urteilstkraft, welche das Besondere darunter subsumiert, [...] bestimmend. Ist aber nur das Besondere gegeben, wozu sie das Allgemeine finden soll, so ist die Urteilstkraft bloß reflektierend.“ (Kant 1790, S. 87)

Das bestimmende Urteil ist ein Entscheiden nach bestimmten Regelwerken, während das reflexive Urteil die Einzelfaktoren, das Spezifische eines Falls in die Urteilsfindung miteinbezieht. Nicht zuletzt ist das Rechtssystem mit der Richterfunktion genau wegen der Reflexivität des Einzelfalls so konstruiert, wie es heute ist (vgl. Rosenmüller 2005, S. 1–14). Während die bestimmenden Urteile von KI-Systemen erfüllt werden können, stellen die reflexiven Urteile aufgrund ihrer vielschichtigen Bewertungsmöglichkeiten das KI-System mindestens vor technische Schwierigkeiten. Insofern kann zunächst festgestellt werden, dass KI-Systeme eine Herstellungstechnik (Poiesis) ohne mitmenschliche Praxis sind. Die leibhaftige mitmenschliche Begegnung innerhalb einer Nutzersituation fehlt, selbst

dann, wenn der Nutzer eine diesbezügliche Attribuierung vornimmt (siehe den Beitrag von Linnemann in diesem Band; vgl. Ebert 1976, S. 12–30). Hier stellt sich erneut die Frage: Wie kann das Verhältnis zwischen Menschen und Maschinen sozial verträglich und verantwortungsvoll gestaltet werden?

Die ethische Bewertung von KI-Systemen in sozialen und pflegerischen Arbeitsbereichen ist vielschichtig und nicht eindeutig zu beantworten. Zusammenfassend lässt sich feststellen, dass

- auf der gesellschaftlichen Ordnungsebene die Frage nach der Art der Reglementierung von KI-Systemen zum Schutz der Bevölkerung ansteht. Im Text wurde hier die europäische Ebene eingeblendet, während die USA und Großbritannien diese Reglementierungen zu engmaschig fanden und einen freieren Umgang mit den KI-Systemen propagieren (vgl. KI-Gipfel in Paris, der vom 10. bis 11. Februar 2025 stattfand).
- es in der Entwicklung von Assistenzsystemen, die die professionelle Arbeit unterstützen soll, im Wesentlichen darauf ankommt, dass die unterschiedlichen Akteur:innen (Wissenschaftler:innen, Techniker:innen, professionelle Anwender:innen) ein evaluiertes Instrument der Berufspraxis zur Verfügung stellen, welches einerseits nützlich ist und andererseits das professionelle Handeln unterstützt, nicht diskriminiert und nachvollziehbare Ergebnisse produziert.
- die Arbeitspraxis in den Organisationen gefordert ist, Mitarbeitende zu qualifizieren und die adäquaten Bedingungen auf der Arbeitsprozessebene neu zu regeln.
- juristisch vielfältige Fragen zu klären sind. Neben dem Datenschutz ist insbesondere die geteilte Verantwortlichkeit zwischen Hersteller:innen, Programmentwickler:innen, Techniker:innen (Wartung) und Anwender:innen klar zu regeln.

Letztendlich ist diese Technik innovativ, nicht mehr wegdenkbar, aber bedarf eines wachen Auges der Angemessenheit und Sensibilität für die jeweiligen Fragestellungen im sozial-pflegerischen Bereich.

Literatur

- Adorno, Theodor W.: Theorie der Halbbildung (1959). In: Gesammelte Schriften, Band 8: Soziologische Schriften (1972). Surkamp-Verlag, Frankfurt a. M., S. 93–121.
- Asilomar-Konferenz (2017): futureoflife.org/open-letter/ai-principles-german/(Abfrage: 15.06.2025).
- Burghardt, Jennifer/Lehmann, Robert/Reder, Michael/Koska, Christopher/Kraus, Maximilian/Müller, Nicholas (2024): Kann Künstliche Intelligenz sozialarbeiterische Entscheidungsprozesse unterstützen? In: unsere jugend 76(7/8), S. 300–310.

- Battagilla, Fiorella/Mukerji, Nikil (2015): Technikethik. In: Nida-Rümelin, Julian/Spiegel, Irina/Tiedemann, Markus (Hrsg.): Handbuch Philosophie und Ethik, Bd. 2. Paderborn: Schöningh, S. 288–295.
- Brödner, Peter (2019): Grenzen und Widersprüche der Entwicklung und Anwendung „Autonomer Systeme“. In: Hirsch-Kreinsen, Hartmut/Karačić, Anemari (Hrsg.) *Autonome Systeme und Arbeit. Perspektiven, Herausforderungen und Grenzen der Künstlichen Intelligenz in der Arbeitswelt*. Bielefeld: transcript, S. 69–97.
- Deutscher Ethikrat (2023): *Mensch und Maschine – Herausforderungen durch Künstliche Intelligenz*, Berlin. <https://www.ethikrat.org/publikationen/stellungnahmen/mensch-und-maschine/> (Abfrage: 15.06.2025).
- DFKI – Deutsches Forschungszentrum für Künstliche Intelligenz (2014): <https://www.dfki.de/web/news/grosser-schritt-fuer-ki-und-umwelt-bmu-zeigt-erste-ergebnisse-der-green-ai-hub-pilotprojekte> (Abfrage: 15.06.2025).
- Ebert, Theodor (1976): Praxis und Poiesis – zu einer handlungstheoretischen Unterscheidung des Aristoteles. In: *Zeitschrift für philosophische Forschung* 30(1), S. 12–30.
- EU-KI-Gesetz (2024): <https://digital-strategy.ec.europa.eu/de/policies/regulatory-framework-ai> (Abfrage: 15.06.2025).
- Gallert, Nina (2024): Die KI-Verordnung – Der zukünftige Rechtsrahmen für ED-Tech an Schulen. In: Hartong, Siegrid/Renz, André (Hrsg.): *Digitale Lerntechnologien*. Bielefeld: transcript, S. 161–187.
- Graevenitz, Albrecht (2018): „Zwei mal Zwei ist Grün“ – Mensch und KI im Vergleich. In: *Zeitschrift für Rechtspolitik* 8, S. 238–241.
- Grunwald, Armin (2002): *Technikfolgenabschätzung – eine Einführung*. Berlin: Ed. Sigma.
- Grunwald, Armin (2021): Technische Zukunft des Menschen? In: Mitscherlich-Schönherr, Olivia (Hrsg.): *Das Gelingen der künstlichen Natürlichkeit*. Berlin: De Gruyter, S. 313–331.
- Heffels, Wolfgang M./Storms Anna: *Ethisch urteilen und handeln: Unterrichtsmaterialien für die Pflegeausbildung*. Göttingen: Vandenhoeck & Ruprecht.
- Heffels, Wolfgang M. (2023): *Ethisch handeln in Helfenden Berufen: Eine handlungsorientierte Einführung*. Stuttgart: Kohlhammer.
- Jonas, Hans (1979): *Das Prinzip Verantwortung. Versuch einer Ethik für die technologische Zivilisation*. Frankfurt a. M.: Insel.
- Kant, Immanuel (1790/1977): *Kritik der Urteilskraft* (hrsg. von Wilhelm Weischedel). Frankfurt a. M.: Suhrkamp.
- Kant, Immanuel (1785/2016): *Grundlegung der Metaphysik der Sitten* (mit einer Einleitung hrsg. von Bernd Kraft und Dieter Schönecker). 2., durchgesehene Auflage. Hamburg: Felix Meiner.
- Klafki Wolfgang (1996): *Neue Studien zur Bildungstheorie und Didaktik*. Weinheim und Basel: Beltz.
- Koch, Andreas (2011): *Wertewandel: Nur Schlagwort? Oder Innovationskraft des 21. Jahrhunderts?* Bremen: Europäischer Hochschulverlag.
- Landesregierung NRW (2023): *Generativen Sprachmodell der Justiz (GSJ)*. <https://www.land.nrw/pressemitteilung/einsatz-kuenstlicher-intelligenz-der-justiz-nordrhein-westfalen-und-bayern> (Abfrage: 15.06.2025).
- Löwisch, Dieter-Jürgen (1995): *Einführung in pädagogische Ethik*. Darmstadt: Wissenschaftliche Buchgesellschaft.
- Luhmann, Niklas (2005): *Soziologische Aufklärung 2*. Wiesbaden: VS Verlag für Sozialwissenschaften.
- Maibaum, Arne/Bischof, Andreas/Hergesell, Jannis (2023): Wie kommt die KI in die Pflege? – oder umgekehrt? Drei Probleme bei der Technikgenese von Pflegetechnologien und ein Gegenorschlag. In: *Pflege & Gesellschaft* 28(1), S. 7–23.
- Ministerium für Schule und Bildung des Landes Nordrhein-Westfalen (2023): *Umgang mit textgenerierenden KI-Systemen*. Düsseldorf.

- Neumaier, Otto (1994): Was hat „Künstliche Intelligenz“ mit Ethik zu tun? In: *Conceptus. Zeitschrift für Philosophie* 27(70), S. 41–76.
- Richter, Philipp (2012): Was gebieten Ratschläge? Zur Unterscheidung technischer und pragmatischer Imperative bei Kant. In: Fischer, Peter/Luckner, Andreas/Ramming, Ulrike (Hrsg.): *Die Reflexion des Möglichen*. Berlin und Münster: Lit, S. 113–125.
- Rosenmüller, Stefanie (2005): Treffen sich Akteur und Zuschauer? Zur Rolle des Richters in Hannah Arendts Urteilstheorie. <https://www.hannaharendt.net/index.php/han/article/view/83/131> (Abfrage: 15.06.2025), S. 1–14.
- Nida-Rümelin, Julian (1996): Ethik des Risikos. In: Nida-Rümelin, Julian (Hrsg.): *Angewandte Ethik*, Stuttgart: Kröner, S. 807–830.
- Sachse, Maximilian (2013): Wird Künstliche Intelligenz zum „Klimakiller“? In: *Frankfurter Allgemeine*, 15.11.2023.
- Savigny, Eike von (1995): Bedeutung, Sprachspiel, Lebensform. *Wittgenstein-Studien* 2 (2).
- Wilaschek, Marcus (1995): Sprachphilosophie. In: Gniffke, Franz/Herold, Norbert (Hrsg.): *Philosophie*. Münster: Lit, S. 157–177.
- Wittgenstein, Ludwig (2001): *Philosophische Untersuchungen* (hrsg. von Joachim Schulte). Frankfurt a. M.: Suhrkamp.

KI in der Kinder- und Jugendhilfe¹

Michael Macsenaere, Monika Feist-Ortmanns

Abstract: In dem Beitrag wird das breite Spektrum der zu erwartenden KI-Anwendungen in der Kinder- und Jugendhilfe sowie die damit verbundenen Risiken und Chancen skizziert. So z. B. Predictive Analytics zur Verbesserung von Diagnostik, Indikationsgüte und Risikoerkennung, Multimodale Textgenerierung von Berichten, Protokollen und Dokumentationen, Entscheidungsunterstützung, virtuelle Beratung durch Chatbots und Unterstützung von administrativen Aufgaben.

Basierend auf den bislang vorliegenden Erfahrungen wird ein sechsstufiges Modell zur KI-Implementierung in Sozialen Organisationen vorgestellt. Dieses Modell integriert KI-Technologie, Datennutzung, Arbeitsprozesse und die Menschen in der Organisation, um eine systematische, bedarfsgerechte Einführung und nachhaltige Nutzung zu gewährleisten.

Keywords: KI, Kinder- und Jugendhilfe, Implementierung, Organisationsentwicklung

1 Einführung

Mit dem vorliegenden Beitrag werden zwei zentrale Ziele verfolgt: Zum einen wird das breite Spektrum an zu erwartenden KI-Anwendungen in der Kinder- und Jugendhilfe inklusive der damit verbundenen Risiken und Limitationen skizziert. Zum anderen wird auf dieser Basis ein Stufenmodell zur KI-Implementierung in Sozialen Organisationen eingeführt, das eine systematische Auseinandersetzung mit der Komplexität der zu bewältigenden Herausforderungen ermöglicht. Eine historische Einordnung sowie vertiefende Ausführungen hierzu, einschließlich der Darstellung von Beispielen und bereits in der Kinder- und Jugendhilfe eingesetzter KI-Applikationen sind unter Macsenaere (2025) und Macsenaere/Feist-Ortmanns (2024) zu finden.

1 © Die Autor:innen 2025, veröffentlicht im Verlag Beltz Juventa, Open Access: CC BY 4.0
Gesä Linnemann/Julian Löhe/Beate Rottkemper (Hg.), Künstliche Intelligenz in der Sozialen Arbeit
10.3262/978-3-7799-8562-4_007

2 Anwendungsoptionen von KI in der Kinder- und Jugendhilfe

Der gezielte und systematische Einsatz von Künstlicher Intelligenz (KI) ist in der Kinder- und Jugendhilfe bislang (Stand September 2024) noch nicht üblich. Angesichts der zu erwartenden tiefgreifenden Veränderungen im Aufwachen junger Menschen durch KI liegt es in der Verantwortung der Kinder- und Jugendhilfe, sich frühzeitig mit den KI-bedingten Herausforderungen und Gestaltungsmöglichkeiten in der Arbeit mit Kindern und Jugendlichen zu befassen. Ziel dabei ist es, eine ähnliche Hilflösigkeit wie im Umgang mit Sozialen Medien zu vermeiden. Die wenigen bereits vorliegenden Erfahrungen und Projekte (siehe Macsenaere 2025) geben Aufschluss darüber, welche Formen der Nutzung von KI im Bereich der Kinder- und Jugendhilfe kurz- und mittelfristig zu erwarten sind. Dies betrifft verschiedene Anwendungsfelder wie Kinderschutz, Beratung und Coaching/Mentoring, sozialpädagogische Diagnostik und (Hilfe-)Planung, Partizipation und Beschwerdemanagement, Qualitätsentwicklung, Wissensmanagement und Forschung. Unabhängig von den nachfolgend kursorisch dargestellten Anwendungsbereichen ist es von zentraler Bedeutung, dass KI in pädagogischen Prozessen nicht isoliert eingesetzt wird, sondern als Teil eines integrativen Ansatzes betrachtet wird, der Fachkräfte in ihrer Expertise unterstützt und nicht ersetzt. Eine nachhaltige Nutzung von KI in pädagogischen Kontexten erfordert daher integrative Ansätze, die auf die jeweiligen Einsatzbereiche zugeschnitten sind und systematisch geplant, getestet, implementiert und weiterentwickelt werden (siehe „Implementierung von KI in der Kinder- und Jugendhilfe“ weiter unten in diesem Beitrag). Hier folgt eine Auswahl von sich abzeichnenden Einsatz-Optionen von KI in der Kinder- und Jugendhilfe:

Diagnostik: Eine Kernkompetenz von KI-Modellen entspricht genau der Anforderung an eine gute Diagnosestellung: auf einer möglichst differenzierten und validen Informations- bzw. Datengrundlage spezifische Muster zu erkennen und daraus eine zusammenführende Beurteilung vorzunehmen. In diesem Sinne wurde in der Medizin in den letzten fünf Jahren die Eignung von KI in der Diagnostik untersucht – mit durchaus ermutigenden Ergebnissen: Studien hierzu zeigen, dass KI nicht nur die Genauigkeit von Diagnosen erhöht, sondern auch den Arbeitsaufwand von Fachkräften reduzieren und die Effizienz im Gesundheitswesen steigern kann. Sie belegen jedoch zugleich, dass menschliche Fachkenntnis – insbesondere in komplexen Fällen – weiterhin eine wichtige Rolle bei der Validierung von KI-generierten Ergebnissen spielt (siehe beispielsweise Al-Antari 2023; Kumar et al. 2023; Mirbabaie/Stieglitz/Frick 2021). In Anbetracht dieser Befundlage ist auch im Bereich der Kinder- und Jugendhilfe mittelfristig eine KI-Erprobung zu erwarten.

Verbesserung der Indikationsgüte: Eine wichtige Aufgabe des Jugendamtes ist die Sicherstellung einer hohen Indikationsgüte, also die Identifizierung und Gewährung geeigneter Hilfen im Einzelfall (Arnold 2014). Mehrere Studien zeigen, dass dies zwar in über 60% der Fälle gelingt, dass in gut 25% der Fälle jedoch eine kontraindizierte Hilfe gewählt wird (Macsenaere/Paries/Arnold 2009; Macsenaere/Esser 2015; Schmidt et al. 2003). Hier ist zu erwarten, dass in den nächsten drei Jahren KI-Systeme unterstützend zum Einsatz kommen werden, die auf Grundlage großer Datenmengen aus abgeschlossenen Hilfeverläufen rückmelden, welche Hilfearten und -settings die besten Erfolgsaussichten im jeweiligen Einzelfall aufweisen.

Hilfedurchführung: KI-Applikationen haben bereits Einzug in die Hilfedurchführungen gehalten: einerseits mit Übersetzungstools wie DeepL oder Google Translate in der pädagogischen Arbeit mit fremdsprachigen Hilfeadressat:innen, andererseits berichten Rothballe und Zeiträg (2024), dass bildgenerierende KI im Sinne einer ressourcenorientierten Pädagogik genutzt wird, um mit Kindern im Vorschulalter Bilder zu entwerfen und gestalten.

Monitoring von Hilfeverläufen: KI bietet die Möglichkeit, umfangreiche Informationen innerhalb kürzester Zeit zu analysieren und daraus Empfehlungen für die Anpassung von Hilfesettings abzuleiten. Dabei kann der Input nicht nur aus Fallakten bestehen, sondern es können – soweit datenschutzkonform – multimodal alle verfügbaren Quellen genutzt werden (siehe „Multimodale Dokumentation“ weiter unten in diesem Beitrag), um individuelle Hilfepläne zu erstellen oder bestehende Pläne zu modifizieren. Rothballe und Zeiträg (2024) betonen hierbei die Rolle KI-gestützter Textanalysen, die eine systematische Verknüpfung der Falldokumentation mit den Zielen der Hilfeplanung ermöglichen.

Risikoerkennung: Auf Grundlage der sogenannten Predictive Analytics können vorhandene Daten KI-gestützt analysiert und zur Früherkennung von Risikokonstellationen genutzt werden. So wurden im Modellprojekt KIEPA Risikofaktoren im Rahmen von Hilfen zur Erziehung auf Basis vorliegender strukturierter und unstrukturierter Daten identifiziert und die weitere Entwicklung junger Menschen prognostiziert (vgl. Plafky/Frischhut 2025; Hahn 2025).

Multimodale Dokumentation: Ein aktueller Trend in der KI-Entwicklung betrifft die sogenannte Multimodalität. Damit ist gemeint, dass der Input für das KI-System nicht nur aus klassischen Textformaten bestehen muss, sondern dass alle denkbaren Quellen einfließen können. Dies macht die KI-Nutzung durch Fachkräfte der Kinder- und Jugendhilfe merklich alltagstauglicher. Die Multimodalität betrifft nicht nur den Input für die KI, sondern auch ihren Output, der damit adressat:innengerecht individualisiert gestaltet werden kann. Dies gilt eben-

so für die Outputmedien wie etwa getextete Protokolle, Präsentationen, Tabellenkalkulationen oder Inhalte, die per Stimmgenerator übermittelt werden. Feist-Ortmanns/Sauer/Brinkmann (2025) geben Hinweise zur Formung einer solchen KI-Unterstützung im Kinderschutz.

Entscheidungsunterstützung: Aktuell wird die Rolle von KI bei der Unterstützung von Entscheidungen in der Kinder- und Jugendhilfe ausgelotet (vgl. Feist-Ortmanns/Sauer/Brinkmann 2025; Lehmann/Burghardt 2025). Insbesondere in Kinderschutzverfahren müssen häufig unter Zeitdruck weitreichende Entscheidungen getroffen werden, oft auf der Grundlage einer unsicheren oder mehrdeutigen Informationslage (Culmsee/Gutmann 2021). Hier kann ein KI-basiertes Assistenzsystem helfen, indem es strukturierte und unstrukturierte Daten analysiert und relevante Informationen liefert, um den Entscheidungsprozess zu unterstützen. Erste Erkenntnisse zu den Entwicklungsanforderungen für ein solches Kinderschutz-Assistenzsystem wurden in einem kooperativen Forschungs- und Entwicklungsprojekt des Instituts für Kinder- und Jugendhilfe mit dem Landkreis Breisgau-Hochschwarzwald und der arf Gesellschaft für Organisationsentwicklung gewonnen (Feist-Ortmanns/Sauer/Brinkmann 2025). Von zentraler Bedeutung hierbei ist, dass KI lediglich im Sinne einer Assistenz eingesetzt wird und die Steuerung und Kontrolle des Prozesses bei der Fachkraft liegt. Dieses Prinzip wird in der Fachliteratur unter „Human-in-the-Loop“ diskutiert (vgl. Tsiakas/Murray-Rust 2022). Die nachhaltige Sicherstellung ist eine zentrale Aufgabe für einen systematischen Implementierungsprozess (siehe „Implementierung von KI in der Kinder- und Jugendhilfe“ weiter unten in diesem Beitrag).

Virtuelle Beratung durch Chatbots: In einer Studie von Ayers et al. (2023) wurden die Antworten eines Chatbots mit denen von Ärzt:innen auf Fragen von Patient:innen, die in einem öffentlichen sozialen Medienforum gestellt wurden, verglichen. Die Antworten des Chatbots wurden in 78,5% der Fälle als gut oder sehr gut bewertet, verglichen mit nur 22,1% der Antworten von Ärzt:innen. Zudem wurde der Chatbot in 45,1% der Fälle als empathisch oder sehr empathisch eingeschätzt, während dies bei den Antworten der Ärzt:innen nur in 4,6% der Fälle zutraf. Diese Befunde legen nahe, dass KI-Tools qualitativ gute, personalisierte Ratschläge geben können, die allerdings von Fachkräften überprüft werden sollten. Ähnliche empirisch überprüfte Erfahrungen liegen auch in anderen Bereichen wie Callcentern und der Psychotherapie vor (Sufyan et al. 2024). Auf dieser Grundlage ist zu erwarten, dass ebenso in der Kinder- und Jugendhilfe – zumindest für einfachere Fragestellungen – Chatbots entwickelt, validiert und bei entsprechender Eignung genutzt werden, zumal dies 24/7, und somit auch an Sonn- und Feiertagen, möglich wäre.

Bildung und kognitive Entwicklung: Auch im Bildungsbereich ist der Einsatz denkbar. Damit könnten Lernprozesse möglicherweise individueller, interaktiver und inklusiver gestaltet werden (vgl. Gentilin 2020; Hamisch/Kruschel 2022; Macgilchrist 2023; Schirmer et al. 2023). Hier setzt das „Flipped Classroom“-Konzept an, mit dem die herkömmliche Lehr-Lern-Struktur umgekehrt wird: Der Wissenserwerb würde damit individuell und KI-gestützt größtenteils außerhalb des Schulunterrichts erfolgen, während die Vertiefung des Gelernten im Unterricht stattfände.

Unterstützung bei administrativen Aufgaben: Die größten Hoffnungen, die Fachkräfte der Kinder- und Jugendhilfe aktuell mit KI verbinden, beziehen sich auf eine Unterstützung oder gar Automatisierung von zumeist ungeliebten administrativen Aufgaben (siehe hierzu auch den Beitrag von Löhe in diesem Band). So könnte KI genutzt werden, um Sitzungsprotokolle automatisch zu erstellen und zu bearbeiten. Mithilfe von Spracherkennungstechnologien könnten Gespräche und Diskussionen während Meetings transkribiert werden, wodurch eine manuelle Protokollierung überflüssig würde. Anschließend könnte die KI auf der Grundlage von Large Language Models (LLM) diese Transkripte analysieren und prägnante Zusammenfassungen erstellen, die die wichtigsten Punkte und Beschlüsse hervorheben und die Essenzen aufgabenspezifisch und adressat:innengerecht formulieren. Dieser Anwendungsfall kann zudem die Transparenz und die intersubjektive Nachvollziehbarkeit in der Zusammenarbeit mit Adressat:innen fördern, indem beispielsweise Protokolle von Hilfeplangesprächen unmittelbar nach dem Gespräch in Leichter Sprache oder als Fremdsprachenübersetzung bereitgestellt werden. Dadurch können diese gemeinsam überprüft und verabschiedet werden. Erste Organisationen testen bereits diese Optionen. Gerade in Anbetracht der großen und damit verbundenen Hoffnungen ist auch hier die Gewährleistung des Human-in-the-Loop-Konzeptes von hoher Bedeutung.

Wissenschaftliches Arbeiten: Nicht zuletzt kann KI das wissenschaftliche Arbeiten unterstützen. Mit Künstlichen Neuronalen Netzen (KNN) ist es möglich, in großen und stark multivariaten Datensätzen Muster und Besonderheiten zu identifizieren, die mit herkömmlichen statistischen Methoden nicht oder nur eingeschränkt erkennbar wären. So erweisen sich beispielsweise FNNs (Feed-forward Neural Network) gegenüber klassischen kausalen Kettenmodellen in vielfacher Hinsicht als überlegen: Sie weisen bei komplexen Klassifikationsaufgaben eine hervorragende Modellgüte (R^2 , Brier-Score, F1-Score) auf, eine theoretisch unendliche Anzahl von Interaktionstermen ist einbeziehbar. Sie benötigen eine extrem kurze Rechenzeit und ermöglichen jederzeit mit minimalem Aufwand Adjustierungen bzw. Aktualisierungen. Als Wermutstropfen muss die geringere Transparenz des Lösungswegs benannt werden, die für KNNs typisch

ist. Mittlerweile gibt es im Rahmen von Explainable AI (XAI) eine Reihe von Strategien, um die Transparenz von KNNs zu erhöhen. Aufgrund dieser insgesamt positiven Erfahrungen ist zu erwarten, dass KNN für weitere wissenschaftliche Fragestellungen angepasst werden.

3 Anwendungsrisiken und Limitation von KI

Wie beschrieben, hat KI innerhalb der Kinder- und Jugendhilfe eine Vielzahl von Nutzungsoptionen. Zum Einsatz werden dabei nicht nur die momentan in der öffentlichen Wahrnehmung dominierenden LLM kommen, sondern eine Reihe weiterer KI-Arten, die auf den spezifischen Anwendungsfall zugeschnitten sein werden. Die damit verbundenen Hoffnungen sind oft sehr hoch. Umso wichtiger ist es, dass vor dem Einsatz in der Praxis die damit verbundenen Risiken und Limitationen bekannt sind und für den spezifisch geplanten Einsatz reflektiert werden. Im vorliegenden Buch werden die mit KI verbundenen Risiken an mehreren Stellen fundiert benannt. Daher erfolgt nun keine allgemeine, sondern eine spezifische Benennung von Risiken und Limitationen in der Kinder- und Jugendhilfe.

In der Kinder- und Jugendhilfe fallen hochsensible Daten an, die beispielsweise die Privatsphäre junger Menschen betreffen können. Daher kommt dem rechtskonformen und systematisch sichergestellten *Datenschutz* eine herausragende Bedeutung zu. Hierbei sind die Regelungen des EU AI Act zu berücksichtigen. Darüber hinaus ist die prägnante Checkliste vom Hamburgischen Beauftragten für Datenschutz und Informationsfreiheit (2023) sehr empfehlenswert. Thematisiert werden hier beispielweise Richtlinien zur Nutzung von KI-Tools, der Einbezug von Datenschutzbeauftragten, die Bereitstellung beruflicher Accounts, die sichere Authentifizierung, der Schutz von personenbezogenen Daten, das Opt-out des KI-Trainings und der Chatverläufe, das Human-in-the-Loop-Prinzip und schließlich die Sensibilisierung und Schulung der Mitarbeitenden.

Weitere Risiken betreffen *Verzerrungen* („Bias“) und daraus resultierende Diskriminierungen (Weyerer/Langer 2020), die Gefahr einer *Überabhängigkeit von technologischen Lösungen* (Passi/Vorvoreanu 2022) und schließlich die sogenannte *Anthropomorphisierung* (Heil 2024), eine Art von Halo-Effekt, mit der einem KI-System menschliche Eigenschaften zugeschrieben werden. Dies kann das Risiko einer weitreichenden Verantwortungsabgabe an die KI erhöhen. Darüber hinaus ist ein partizipatives Vorgehen bei der Entwicklung von KI-Anwendungen in der Sozialen Arbeit geboten, das die Bedürfnisse und Erwartungen der Nutzer:innen sowohl aufseiten der Fachkräfte als auch der Adressat:innen berücksichtigt. Auf diese Weise kann einem Top-down-Bias bei der Entwicklung potenziell wirkmächtiger KI-Anwendungen in der Sozialen Arbeit entgegenge-

wirkt und zur Demokratisierung des Arbeitsfelds beigetragen werden, anstatt Machtasymmetrien im Zuge der KI-basierten Digitalisierung zu verstärken.

Jenseits dieser Risiken sind auch die KI-immanenten Limitierungen zu reflektieren, etwa die mittlerweile gut bekannten und gefürchteten „Halluzinationen“ der LLM, bei denen plausible, aber faktisch ungenaue oder vollständig erfundene Inhalte generiert werden. Neuere Modelle wie z. B. ChatGPT o1 sollen allerdings die Wahrscheinlichkeit hierfür merklich reduzieren. Weitere Limitationen betreffen das *mangelnde bzw. fehlende Verständnis der Inhalte* (Birhane/McGann 2024) und das *kritische Kooperationsverhalten* von LLM (Fontana / Pierri / Aiello 2024).

4 Implementierung von KI in der Kinder- und Jugendhilfe

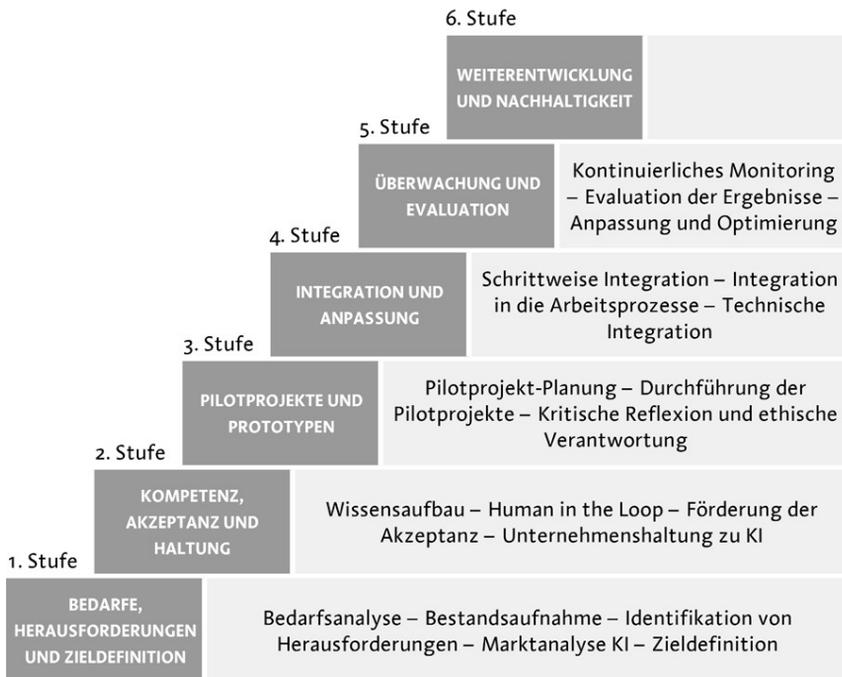
In Anbetracht der beschriebenen Risikobereiche und der hohen Komplexität ist es unbedingt empfehlenswert, dass die Implementierung von KI in der Kinder- und Jugendhilfe systematisch geplant und von umfassenden Schulungen, klaren Richtlinien und einer ständigen Evaluierung bzw. Validierung begleitet wird. Verstärkt wird diese Empfehlung noch durch die rasante, ausgeprägt exponentiell verlaufende Entwicklungsgeschwindigkeit von KI. Betroffen davon werden sein die großen Sprachmodelle (LLM), das Natural Language Processing (NLP), die Edge AI, die die Nutzung cloudunabhängig direkt auf den Endgeräten ermöglichen wird, die hoch praxisrelevante Multimodalität der KI-Applikationen, eine appübergreifende Automatisierung (Cross-Application Automation), personalisierte KI-Assistenten und schließlich die bereits oben skizzierten Chatbots als virtuelle Assistenz. Eine systematische Implementierung sollte diese beschriebenen Entwicklungstendenzen idealerweise antizipieren und im weiteren Prozessverlauf im Blick behalten. Darüber hinaus sollten mit einer Implementierung vier zentrale Dimensionen erfasst und aufeinander abgestimmt werden:

1. die KI-Technologie,
2. die Datennutzung und der Datenschutz,
3. die Arbeitsprozesse in der Organisation und
4. die Menschen, die KI nutzen und von KI betroffen sein werden.

Auf Grundlage dieser Dimensionen, der oben beschriebenen Anforderungen und der bislang vorliegenden Erfahrungen wird abschließend ein sechsstufiges Modell vorgestellt, das einen systematischen Ablauf der Implementierung von KI in Sozialen Organisationen aufzeigt (siehe Abbildung 1).

Auf der 1. Stufe werden in einem ersten Teilschritt die Bedarfe der Organisation identifiziert. Auf dieser Grundlage folgt eine Bestandsanalyse, die die aktuell vorhandenen Technolog

Abbildung 1: Stufenmodell zur KI-Implementierung in Sozialen Organisationen



Quelle: Nach Macsenaere/Feist-Ortmanns 2025

ien und Daten, aber auch die Prozessabläufe innerhalb der Organisation analysiert. In einem nächsten Teilschritt werden die technischen, organisatorischen und ethischen Herausforderungen in den Blick genommen und identifiziert. Darauf folgen eine Marktanalyse und schließlich die Definition des Ziels, das mit dem Einsatz von KI in der Organisation erreicht werden soll. Im Sinne von Transparenz und Beteiligung, aber auch um die Wahrscheinlichkeit der Akzeptanz von KI-Anwendungen durch Mitarbeitende und Adressat:innen zu erhöhen, ist es empfehlenswert, bereits auf Stufe 1 die Erwartungen, Bedarfe und ggf. Sorgen sämtlicher Stakeholder in den Analyseprozess miteinzubeziehen. Zudem sollte eine nachvollziehbare Dokumentation der zentralen Entscheidungen und Erkenntnisse erfolgen.

Der Mensch steht im Zentrum der 2. Stufe: Hier sollen Kompetenzen im Umgang mit KI aufgebaut, Akzeptanz gefördert und eine einheitliche Haltung zur Nutzung von KI innerhalb der Organisation entwickelt werden. Der Wissensaufbau umfasst die Vermittlung von Grundlagen über die Funktionsweise und den Anwendungsgebieten von KI sowie die Sensibilisierung für die Bedeutung der Datenqualität, potenzielle Verzerrungen (Bias) und das Human-in-the-Loop-Konzept. Gemäß dem Technology Acceptance Model (TAM) (vgl. Davis/Granić

2024) sollen Hemmschwellen abgebaut werden, indem die wahrgenommene Nützlichkeit und Benutzerfreundlichkeit von KI-Tools erfahrbar gemacht werden. Darüber hinaus wird in der Organisation eine klare Haltung zu ethischen Fragen entwickelt, indem verbindliche Leitlinien zur Nutzung von KI erstellt, Transparenz gefördert und kontinuierliche Weiterbildung ermöglicht werden.

Auf der 3. *Stufe* wird der Fokus auf die Planung und Durchführung von Pilotprojekten und die Entwicklung von Prototypen gelegt. Um die Komplexität zu bewältigen, sollten Pilotprojekte systematisch geplant werden. Hierbei empfiehlt es sich, einen begrenzten und klar definierbaren Bereich der Organisation auszuwählen, um erste Erfahrungen mit KI zu sammeln. Die erforderlichen personellen und technischen Ressourcen müssen identifiziert und sichergestellt werden. In der Durchführung gilt es, Prototypen zu entwickeln, die den definierten Anforderungen entsprechen. Nach Testläufen und erforderlichen Anpassungen werden diese Prototypen in den vorgesehenen Bereichen der Organisation eingeführt. Dabei ist ein kontinuierliches Monitoring der Leistung und Wirkung sicherzustellen. Die Ergebnisse der Pilotprojekte sollten kritisch reflektiert werden, um zu bewerten, ob und wie die Prototypen auf weitere Bereiche der Organisation ausgeweitet werden können.

Die schrittweise Integration der Prototypen in den regulären Betrieb erfolgt auf der 4. *Stufe*. Dabei geht es nicht nur um die technische Einbindung, sondern auch um die nahtlose Integration in die bestehenden Arbeitsprozesse. Die Entwicklung eines detaillierten Plans zur systematischen Einführung ist dabei unerlässlich. Dies schließt die Erweiterung der getesteten Prototypen auf andere Bereiche der Organisation ein. Um eine reibungslose Integration zu gewährleisten, sind Schulungen der Fachkräfte entscheidend, damit diese die KI-Anwendungen im Arbeitsalltag effektiv nutzen können. Ebenso wichtig ist die technische Integration, die sicherstellt, dass die KI-Lösungen mit der bestehenden IT-Infrastruktur kompatibel sind und alle Datenschutzanforderungen eingehalten werden.

Nach der Integration steht auf der 5. *Stufe* die kontinuierliche Überwachung und Evaluation der KI-Lösungen im Mittelpunkt. Es gilt, die Leistung der integrierten Systeme regelmäßig zu überwachen und abhängig vom Feedback der Fachkräfte sowie der betreuten Personen Anpassungen vorzunehmen. Zudem sollte systematisch geprüft werden, inwieweit die auf Stufe 1 definierten Ziele erreicht wurden und welche Bereiche der KI-Lösungen optimiert werden müssen. Dabei ist sicherzustellen, dass alle Optimierungen den ethischen Standards und den rechtlichen Datenschutzvorgaben entsprechen.

Die langfristige Weiterentwicklung und Nachhaltigkeit der KI-Integration werden mit der 6. *Stufe* begleitet. Um eine nachhaltige Nutzung von KI sicherzustellen, ist die Entwicklung einer umfassenden Strategie unerlässlich. Diese Strategie sollte kontinuierliche Schulungen und Weiterbildungen der Fachkräfte beinhalten, um sicherzustellen, dass alle Beteiligten mit den neuen Technolo-

gien vertraut sind. Ein effektiver Wissenstransfer innerhalb der Organisation ist ebenso notwendig, um Synergien zu schaffen und die Effizienz zu steigern. Zudem müssen regelmäßige Überprüfungen stattfinden, um die KI-Strategie an aktuelle technologische Entwicklungen anzupassen und sicherzustellen, dass die KI-Lösungen auch langfristig den Anforderungen der Organisation und den Bedürfnissen der betreuten Personen gerecht werden.

Die in diesem Modell skizzierten sechs Stufen der KI-Implementierung verdeutlichen die inhärente Komplexität einer nachhaltigen Nutzung von KI in Sozialen Organisationen. Eine sorgfältige, systematische Planung ist unerlässlich, um die Chancen einer solchen Integration zu erhöhen und die Risiken zu begrenzen. Dabei ist es von zentraler Bedeutung, sowohl technologische als auch menschliche Aspekte gleichermaßen zu berücksichtigen. Nur mit einer solchen ausbalancierten Herangehensweise kann KI zu einem substanziellen und langfristigen Mehrwert für alle beteiligten Akteur:innen beitragen. Gelingt dies, birgt ein bedarfsgerechter Einsatz von KI-Tools in der Kinder- und Jugendhilfe große Chancen zur Stabilisierung und Qualifizierung des Hilfesystems in Zeiten gravierenden Personalmangels.

Literatur

- Al-Antari, Mugahed A. (2023): Artificial Intelligence for Medical Diagnostics – Existing and Future AI Technology! In: *Diagnostics* 13(4), S. 688.
- Arnold, Jens (2014): Passgenaue Hilfen und ihre Indikation. In: Macsenaere, Michael/Esser, Klaus/ Knab, Eckhart/ Hiller, Stephan (Hrsg.): *Handbuch der Hilfen zur Erziehung*. Freiburg: Lambertus, S. 224–230.
- Ayers, John W./Poliak, Adam/Dredze, Mark/Leas, Eric C./Zhu, Zechariah/Kelley, Jessica B./Faix, Dennis J./Goodman, Aaron M./Longurst, Christopher A./Hogarth, Michael/Smith, David Mitchell (2023): Comparing physician and artificial intelligence chatbot responses to patient questions posted to a public social media forum. In: *JAMA Internal Medicine* 183(6), S. 589–596.
- Birhane, Abeba/McGann, Marek (2024): Large Models of What? Mistaking Engineering Achievements for Human Linguistic Agency. arXiv preprint arXiv:2407.08790.
- Culmsee, Thorsten/Gutmann, Veit (2021): Entscheidungsfallen im Kinderschutzverfahren, in: *E-Learning Kinderschutz: Basiswissen Kinderschutz, KJPP – Universitätsklinikum*.
- Davis, Fred D./Granić, Andrina (2024): *The Technology Acceptance Model – 30 Years of TAM*. Cham: Springer.
- Der Hamburgische Beauftragte für Datenschutz und Informationsfreiheit (2023): *Checkliste zum Einsatz LLM-basierter Chatbots*. https://datenschutz-hamburg.de/fileadmin/user_upload/HmbBfDI/Datenschutz/Informationen/20231113_Checkliste_LLM_Chatbots_DE.pdf (Abfrage: 15.06.2025).
- Feist-Ortmanns, Monika/Sauer, Annette/Brinkmann, Martin (2025): KI-basiertes Assistenzsystem im Kinderschutzverfahren. In: Macsenaere, Michael (Hrsg.): *Künstliche Intelligenz in der Kinder- und Jugendhilfe*. München: Ernst Reinhardt, S. 50–66.
- Fontana, Nicolás/Pierri, Francesco/Aiello, Luca Maria (2024): *Nicer Than Humans: How do Large Language Models Behave in the Prisoner’s Dilemma?* arXiv preprint <https://>

- Gentilin, Olivetta (2020): KI in der Schule: Digitale Lehrkonzepte und Anwendungsbeispiele für den Fremdsprachenunterricht. In: *Information – Wissenschaft & Praxis* 71(1), S. 5–16.
- Hahn, Daniel (2025): Die Stimme aus der Praxis. In: Macsenaere, Michael (Hrsg.): *Künstliche Intelligenz in der Kinder- und Jugendhilfe*. München: Ernst Reinhardt, S. 83–86.
- Hamisch, Katharina/Kruschel, Robert (2022): Zwischen Individualisierungsversprechen und Vermessungsgefahr. Die Rolle der Schlüsseltechnologie Künstliche Intelligenz in der inklusiven Schule. In: *Grenzen.Gänge.Zwischen.Welten. Kontroversen – Entwicklungen – Perspektiven der Inklusionsforschung*. Bad Heilbrunn: Julius Klinkhardt, S. 108–115.
- Heil, Reinhard (2024): *Künstliche Intelligenz und die Tücken der Anthropomorphisierung*. Institut für Technikfolgenabschätzung und Systemanalyse (ITAS), Karlsruher Institut für Technologie (KIT).
- Kumar, Yogesh/Koul, Apeksha/Singla, Ruchi/Ijaz Fazal, Muhammad (2023): Artificial intelligence in disease diagnosis: a systematic literature review, synthesizing framework and future research agenda. In: *Journal of Ambient Intelligence and Humanized Computing* 14, S. 8459–8486.
- Lehmann, Robert/Burghardt, Jennifer (2025): Die Akzeptanz von Künstlicher Intelligenz bei der Entscheidungsunterstützung in der Kinder- und Jugendhilfe. In: Macsenaere, Michael (Hrsg.): *Künstliche Intelligenz in der Kinder- und Jugendhilfe*. München: Ernst Reinhardt, S. 67–73.
- Macgilchrist, Felicitas (2023): KI und Schule: Sichtweisen, Anwendungen und Gestaltungsmöglichkeiten. In: SCHÜLER. Wissen für Lehrer 1, S. 82–84.
- Macsenaere, Michael (Hrsg.) (2025): *Künstliche Intelligenz in der Kinder- und Jugendhilfe*. München: Ernst Reinhardt.
- Macsenaere, Michael/Feist-Ortmanns, Monika (2024): Künstliche Intelligenz (KI): Ein historischer Abriss und die zu erwartenden Implikationen für die Kinder- und Jugendhilfe. In: *unsere jugend* 76(7+8), S. 294–299.
- Macsenaere, Michael/Esser, Karl (2015): *Was wirkt in der Erziehungshilfe? Wirkfaktoren in Heimerziehung und anderen Hilfearten*. 2. Auflage. München: Ernst Reinhardt.
- Macsenaere, Michael/Paries, Gabriele/Arnold, Jens (2009): *EST! Evaluation der Sozialpädagogischen Diagnose-Tabellen – Abschlussbericht*. Bayerisches Staatsministerium für Arbeit und Sozialordnung, Familie und Frauen. München.
- Mirbabaie, Milad/Stieglitz, Stefan/Frick, Nicholas R. J. (2021): Artificial intelligence in disease diagnostics: A critical review and classification on the current state of research guiding future direction. In: *Health and Technology* 11, S. 693–731.
- Passi, Samir/Vorvoreanu, Mihaela (2022): *Overreliance on AI: Literature Review*. Microsoft Technical Report MSR-TR-2022-12.
- Plafky, Christina/Frischhut, Hans (2025): Einsatz von Künstlicher Intelligenz zu Prognosezwecken in der Kinder- und Jugendhilfe. In: Macsenaere, Michael (Hrsg.): *Künstliche Intelligenz in der Kinder- und Jugendhilfe*. München: Ernst Reinhardt, S. 74–82.
- Rothballer, Marc/Zeiträg, Maximilia (2024): Künstliche Intelligenz als Chance für Soziale Arbeit. Ein Werkstattbericht der Diakonie Rosenheim zu transformativen Ansätzen und Anwendungen. In: *unsere jugend* 76(7+8), S. 328–337.
- Schirmer, Katja/Berger, Martin/Himpsl-Gutermann, Klaus/Lorenz, Setara-Anna/Steiner, Michael (2023): *Künstliche Intelligenz im Unterricht: Lehr-/Lernszenarien für verschiedene Gegenstände*. In: *Medienimpulse* 61(2).
- Schmidt, Martin H./Petermann, Franz/Macsenaere, Michael/Knab, Eckhart/Schneider, Karsten/Hölzl, Heinrich/Hohm, Erika/Pickartz, Andrea/Flosdorf, Peter (Hrsg.) (2003): *Effekte erzieherischer Hilfen und ihre Hintergründe*. Schriftenreihe des BMFSFJ, Band 219. Stuttgart: Kohlhammer.
- Sufyan, Nabil Saleh/Fadhel, Fahmi H./Alkathami, Saleh Safeer/Mukhadi, Jubran Y. A. (2024): Artificial intelligence and social intelligence: preliminary comparison study between AI models and psychologists. In: *Frontiers in Psychology* 15.

Weyerer, Jan C./Langer, Paul F. (2020): Diskriminierungen und Verzerrungen durch Künstliche Intelligenz. Entstehung und Wirkung im gesellschaftlichen Kontext. In: Oswald, Michael/Borucki, Isabelle (Hrsg.): Demokratietheorie im Zeitalter der Frühdigitalisierung. Wiesbaden: Springer VS, S. 209–240.

Künstliche Intelligenz als Gestalterin von Medienkulturen: eine medienpädagogische Perspektive auf eine sich verändernde Identitätsarbeit und Sozialisierung¹

Eik-Henning Tappe

Abstract: Der Beitrag untersucht aus einer medienpädagogischen Perspektive die Rolle Künstlicher Intelligenz als aktiver Gestalterin digitaler Medienkulturen und deren Einfluss auf Prozesse von Sozialisierung und Identitätsbildung. Im Zentrum steht die These, dass KI-Systeme nicht mehr nur als technische Werkzeuge fungieren, sondern zunehmend zu eigenständigen kulturellen Akteuren avancieren, die kollektive Bedeutungsvorräte generieren und Wahrnehmung sowie Interpretation sozialer Wirklichkeiten prägen. Diese algorithmisch geformten Räume eröffnen neuartige Ausdrucks- und Partizipationschancen, bergen jedoch zugleich Risiken wachsender Abhängigkeit und Destabilisierung individueller Selbstverortung. Dabei gewinnen Formen der Identitätsarbeit an Bedeutung, die sich in digitalen Kontexten als kontinuierliche Aushandlungs- und Reflexionsprozesse manifestieren. Der Text plädiert für eine kritisch-reflexive Medienpädagogik, die Nutzer*innen befähigt, algorithmische Prozesse zu verstehen, ihre Wirkung auf gesellschaftliche Narrative zu reflektieren und mediale Handlungsräume selbstbestimmt zu gestalten.

Keywords: Künstliche Intelligenz, Medienkulturen, Digitalität, Mediatisierung, Medienpädagogik

1 Einführung: Technik – Medien – Gesellschaft

Die fortschreitende Digitalisierung greift tief in technologische, gesellschaftliche und kulturelle Prozesse ein, indem sie nicht nur neue Technologien hervorbringt, sondern auch bestehende Strukturen grundlegend transformiert. In diesem Kontext tritt Künstliche Intelligenz (KI) nicht lediglich als technisches Hilfsmittel in

1 © Die Autor:innen 2025, veröffentlicht im Verlag Beltz Juventa, Open Access: CC BY 4.0
Gesa Linnemann/Julian Löhe/Beate Rottkemper (Hg.), *Künstliche Intelligenz in der Sozialen Arbeit*
10.3262/978-3-7799-8562-4_008

Erscheinung, sondern wirkt aktiv auf soziale Dynamiken und menschliche Interaktionen ein, wodurch sie zunehmend die Art und Weise beeinflusst, wie Menschen handeln, denken und sich gesellschaftlich einbringen. Im vorliegenden Beitrag soll der Versuch unternommen werden, die Relevanz von KI für medienkulturelle Entwicklungen sowie das pädagogische Handeln in der Sozialen Arbeit zu skizzieren. Dazu erscheint es zunächst notwendig, den kulturellen Wandel, den digitale Technologien und Medien mit sich bringen, einleitend zu betrachten.

Der Begriff „Digitalisierung“ verweist auf die breite Umstellung von analogen auf digitale Inhalte(n) und Technologien. Diese Veränderungen umfassen nicht nur eine neue Codierung und Speicherung von Informationen, sondern auch die Entwicklung „intelligenter“ Technologien, die zunehmend eigenständig agieren und damit die Notwendigkeit menschlicher Steuerung und Kontrolle reduzieren (vgl. Düll 2016, S. 6f.; Hoffmann 2019, S. 103f.). Ebenso stellt die Digitalisierung weit mehr als einen rein technischen Wandel dar – sie ist Ausdruck eines tiefgreifenden kulturhistorischen Transformationsprozesses, der auf vielfältigen Entwicklungen aufbaut (wie beispielsweise die Normierung von Maßen, die Algorithmisierung von Wissen oder die systematische Erfassung und Verarbeitung individueller Daten). Vor diesem Hintergrund sollte die Entwicklung des Digitalen nicht primär als radikaler Bruch oder gänzlich neues Zeitalter verstanden werden. Vielmehr gilt es, die Kontinuitäten und kulturhistorischen Voraussetzungen zu betonen, die die Entstehung und Ausbreitung der Digitalisierung überhaupt erst möglich gemacht haben (vgl. Jörissen/Unterberg 2019, S. 12). Gleichzeitig üben digitale Technologien durch ihre inhärenten Eigenschaften, beispielsweise die algorithmische Standardisierung und Vereinfachung von Prozessen und Inhalten, selbst Einfluss auf gesellschaftliche Praktiken aus. Dadurch ergeben sich tiefgreifende Veränderungen kultureller Handlungsweisen, die von der Einführung vernetzter Informationssysteme bis hin zur wachsenden Entkopplung zwischenmenschlicher Interaktionen reichen (vgl. Allert et al. 2019, S. 66; Krotz 2022, S. 207).

Knaus konstatiert (nicht zu Unrecht), dass die Diskussionen über Digitalisierung oft das Digitale betonen, die Bedeutung des Medialen in menschlich-technischen Kommunikationsprozessen jedoch vernachlässigen. Dabei ermöglicht gerade die mediale Schnittstelle (z. B. über ein Interface, eine Sprachausgabe oder als visuelle Information) nicht nur die Vermittlung von Daten, sondern auch partizipative Eingriffe und (soziale) Interaktionen. Medien bilden somit das zentrale Bindeglied zwischen Menschen und Technologien (wie eben KI) und sind entscheidend für ein Zustandekommen von Mensch-Maschine-Interaktionen (vgl. Knaus 2020, S. 27f.). Ferner stellen digital vermittelte Medien nicht allein reine Kommunikationsmittel dar, sondern waren und sind stets auch Triebfedern von soziokulturellen Umbrüchen. Krotz (2007; 2015) beschreibt eine damit verbundene Mediatisierung als einen langfristigen und umfassenden Prozess, durch den Medien und Kommunikationstechnologien zunehmend alle

gesellschaftlichen Bereiche durchdringen und umgestalten. Solche medialen Umbrüche sind zumeist – und vermehrt im Verlauf der letzten zwei Jahrhunderte – eng mit technischen Innovationen verknüpft. Das Aufkommen einer populären „Bewegtbildsprache“ wäre beispielsweise ohne die Erfindung des Kinetografen, einem der ersten kommerziell erfolgreichen Vorgänger moderner Filmkameras, schwer vorstellbar. Ebenso waren die individuelle Aneignung und Gestaltung moderner Musikgenres wie Rock ‘n’ Roll, Popmusik oder elektronische Musik der 1980er-Jahre an technologische Innovationen wie Vinylschallplatten oder Kassetten gebunden. Derartige technische Entwicklungen transformieren nicht nur das Medienangebot, sondern prägen zugleich gesellschaftliche, wirtschaftliche und kulturelle Handlungsweisen nachhaltig. Die Mediatisierungsforschung betrachtet in diesem Zuge den Wandel von Kommunikation und Kultur als eine kontinuierliche Entwicklung, die durch die Einführung neuer (technischer) Kommunikationsmittel und sich verändernde mediale Ausdrucks- und Handlungsweisen begleitet wird und langfristig die Strukturen des sozialen Lebens verändert (vgl. Hepp/Hasebrink 2017, S. 335; Krotz 2022, S. 208 f.). Medien beeinflussen demnach nicht nur die Art und Weise, wie wir kommunizieren, sondern auch, wie wir handeln, wahrnehmen und miteinander in Beziehung treten. In das Zusammenspiel aus Digitalisierung und Mediatisierung, welches gegenwärtig unsere Gesellschaften prägt, setzt Stalder den Begriff der Digitalität (vgl. Stalder 2016). Eine damit verwobene *Kultur der Digitalität* zeichnet sich durch drei prägnante Eigenschaften aus: Referenzialität, Gemeinschaftlichkeit und Algorithmizität. So fungieren beispielsweise Social-Media-Kanäle nicht bloß als passive Räume für die Rezeption von Inhalten, sondern vielmehr als aktive Foren der Gemeinschaftsbildung. Innerhalb dieser digital vermittelten Gemeinschaften werden medienkulturelle Artefakte (z. B. Memes, YouTube-Videos oder Charaktere aus Videospielen) zu zentralen Referenzpunkten innerhalb der kommunikativen Praxis. Was Nutzer:innen dabei angezeigt oder vorenthalten wird, steuern wiederum Algorithmen, basierend auf den vermuteten Vorlieben, und beeinflussen so das individuelle kulturelle Handeln sowie die Ausgestaltung von gesellschaftlicher Partizipation (vgl. ebd., S. 95 f.).

In diesem technologisch-medialen Gefüge aus Digitalisierung, Mediatisierungsprozessen und Digitalität spielen diverse Varianten von KI eine zentrale Rolle, die nicht erst seit der Popularität großer Sprachmodelle wie GPT in der breiten Öffentlichkeit diskutiert werden. KI-Systeme beeinflussen bereits jetzt maßgeblich die Art und Weise, wie Inhalte generiert, gefiltert und verbreitet werden. Sie steuern zunehmend die Kommunikation und Interaktion in digital vermittelten Räumen und formen damit die sozialen Praktiken von Individuen. Ein prägnantes Beispiel dafür ist der Einsatz von KI-gestützten Chatbots und Voicebots, die in der Kund:innenkommunikation automatisierte, dialogorientierte Interaktionen ermöglichen. Unternehmen nutzen diese Systeme, um in Echtzeit auf Anfragen zu reagieren und personalisierte Antworten zu liefern.

Ebenso prägen KI-gestützte Algorithmen in sozialen Netzwerken den digitalen Austausch, indem sie Inhalte auf Grundlage von gesammelten Datenbeständen selektieren und vermeintliche personalisierte Inhalte für die Darstellung priorisieren. Solche Technologien agieren in digitalen Kontexten nicht mehr nur als Werkzeuge, sondern nehmen über mediale Inhalte und Kommunikationswege aktiv Einfluss auf die Wahrnehmung und Interpretation der (sozialen) Umwelt.

2 Stellenwert von KI in digitalisierten Medienkulturen

In einer Kultur der Digitalität fungieren digitale Medien nicht mehr als eine von vielen Sozialisationsinstanzen, sondern sind als generationenüberspannende zentrale Akteure der Sozialisierung zu verstehen (vgl. Stalder 2016, S. 139 ff.; Adolf 2017, S. 54 f.). Hepp beschreibt in diesem Zusammenhang Medienkulturen als soziale Gefüge, in denen die primären Bedeutungsressourcen – etwa Videos, Webseiten oder soziale Netzwerke – durch technische Kommunikationsmedien vermittelt werden und in denen (digitale) Medienkommunikation die Basis für gesellschaftliche Orientierung bildet. Medienkulturen entstehen, wenn Medienkommunikation derart tief in die alltägliche Lebenswelt integriert ist, dass gesellschaftliche Werte, Normen und Bedeutungen überwiegend über mediale Vermittlung entstehen und weitergegeben werden (vgl. Hepp 2013, S. 65). Dabei wird die Bedeutung von Medien nicht nur durch ihren technischen Vermittlungscharakter definiert, sondern auch durch ihre Funktion als Kulturproduzenten. Sie verknüpfen kulturell verfügbare Wissensbestände mit subjektiven Kommunikationshandlungen und schaffen so kollektive Bedeutungsvorräte, die das Fundament für soziale Integration bilden. Diese Bedeutungsvorräte strukturieren, welche Themen als gesellschaftlich relevant erachtet und in Kommunikationsprozessen aufgegriffen werden (vgl. Adolf 2017, S. 53 f.).

Medien stellen somit nicht allein Informationen bereit, sondern beeinflussen auch die Art und Weise, wie Individuen die soziale und natürliche Umwelt und sich selbst wahrnehmen und verstehen. Entsprechende Medienkulturen lassen sich nicht nur über die zentrale Rolle von Medien in der Vermittlung und Konstruktion von Bedeutung definieren, sondern auch durch die vielfältigen Prozesse, die diese Kultur formen und prägen. Dabei spielen nach Hepp vier zentrale Dimensionen eine Rolle:

- die Produktion von Medieninhalten,
- die Repräsentation von Bedeutungen und sozialen Wirklichkeiten,
- die individuelle Aneignung dieser Inhalte durch die Rezipient:innen sowie
- die Identifikation mit den durch Medien vermittelten Werten und Normen.

Diese Prozesse sind nicht losgelöst voneinander zu betrachten, da sie in Wechselwirkung zueinanderstehen und das Verständnis von Medienkulturen als sozial

und politisch geformte Konstrukte komplettieren (vgl. Hepp 2013, S. 66 f.). Gegenwärtige Generationen sind somit durch digitale Medien unmittelbar in kulturelle und gesellschaftliche Entwicklungsprozesse eingebunden und erleben deren Auswirkungen direkt und ungefiltert. Sie sind nicht mehr nur passive Rezipient*innen, sondern nehmen durch ihre neuen Nutzungs- und Handlungsmuster aktiv Einfluss auf diese Entwicklungen. Der (zumeist) uneingeschränkte mediale Zugang zu nahezu allen Bereichen kulturellen und gesellschaftlichen Lebens eröffnet ihnen kontinuierlich neue, oft unvorhersehbare Erfahrungs- und Lernmöglichkeiten, die weder planbar noch kontrollierbar sind (vgl. Spanhel 2017, S. 3). Folglich besitzen Medienkulturen innerhalb einer Kultur der Digitalität eine gewisse Ambivalenz: Einerseits bieten sie völlig neue Möglichkeiten der Artikulation und Ausdrucksformen für Sozialisations- und Bildungsprozesse. Andererseits sind die zugehörigen digitalen Räume (z. B. Social-Media- und Gamingplattformen) oft stark von ökonomischen Interessen geprägt, die durch kapitalistische Verwertungslogiken bestimmt werden (Bettinger 2020, S. 242 f.).

Im Zuge der fortschreitenden Digitalisierung gewinnen KI-Systeme in ebenjenen kulturellen Prozessen zunehmend an Bedeutung. Sie agieren nicht mehr nur als passive Werkzeuge, sondern entwickeln sich zu aktiven medialen Akteuren, die selbst kulturelle Referenzobjekte generieren und diese in die Struktur von Medienkulturen einbetten. Entsprechende KI-generierte Medieninhalte können als Teil eines (medien)kulturellen „Belief Space“ (Reynolds 1994) angesehen werden. Innerhalb von sogenannten *kulturellen Algorithmen*, in denen kulturelle Praktiken ähnlich wie genetische Informationen über Generationen hinweg weitergegeben werden, fungiert ein Belief Space als Sammlung für kollektive Wissensstrukturen und Werte von Gesellschaften (vgl. ebd., S. 2 f.). Mit der zunehmenden Vernetzung in digitalen Räumen entsteht in diesem Zuge ein globales Netzwerk von Interaktionen, das zur Herausbildung spezifischer erzählerischer Muster, sogenannter *Narrative*, führt (vgl. Rust 2017, S. 22). In digitalisierten Medienkulturen, die stark durch die Präsenz von KI geprägt werden, übernimmt diese bereits jetzt eine zentrale Funktion bei der Strukturierung und Darstellung ebenjener Narrative und wird somit (bewusst oder unbewusst eingesetzt) Mitgestalterin von kulturellen Belief Spaces (vgl. Knaus 2020, S. 40–45; Gapski 2022, S. 695 f.). Auf dieser Basis ist anzunehmen, dass KI-Systeme zukünftig zu zentralen Akteuren innerhalb des sozialen und kulturellen Raumes werden und Sozialisationsprozesse nachhaltig mitgestalten. KI würde damit eine Rolle einnehmen, die mit der aktuellen Funktion von Medienkulturen als primäre Sozialisationsinstanzen vergleichbar ist und sich selbst vom Werkzeug zu einem primären Einflussfaktor innerhalb einer mediatisierten Gesellschaft wandeln.

Folgt man dieser Argumentationslinie, ist anzunehmen, dass dies zu einer grundlegenden Verschiebung kultureller Praktiken führt, da die algorithmischen Selektionsprozesse festlegen, welche kulturellen Inhalte überhaupt zur Wahrnehmung gelangen und welche Aspekte ausgeblendet werden. Vor diesem

Hintergrund wird deutlich, dass die kulturelle „Wirkmacht“ von KI-Systemen weit über die reine Verarbeitung von Daten hinausgehen kann. Sie greifen tief in die symbolischen Ordnungen der Alltagskultur ein, indem sie ästhetische Ausdrucksformen umdeuten und ihnen neue Bedeutungszuweisungen verleihen. Somit beeinflussen die durch KI generierten Produkte und Inhalte unsere Welt- und Selbstwahrnehmung mit (z. B. durch Filter, über die bildliche Darstellung von Personen und Objekten oder über realistisch anmutende Fälschungen). Dadurch kommt es vermehrt zu Erfahrungen „zweiter“ Ordnung, also Phänomenen, die nicht mehr selbst erfahren, sondern rein (medial) vermittelt werden (vgl. Spanhel 2017, S. 9 f.). Die eigene soziokulturelle Verortung und die individuelle Weltaneignung werden demnach vermehrt auch an durch KI-generierte Inhalte geknüpft.

3 Das private Innere, das gesellschaftliche Außen und dazwischen KI?

Ein Verständnis der Wechselwirkungen zwischen kulturellen Aneignungsräumen, Belief Spaces und KI erscheint insbesondere für Identitäts- und Gesellschaftsprozesse von Bedeutung. Digitale Orte schaffen durch ihre spezifischen medialen Ästhetiken intensive emotionale Involvierungen und bieten Raum für neue Identifikationsmöglichkeiten. Diese virtuellen Räume beeinflussen Identitätsbildungsprozesse, indem sie beispielsweise durch visuelle und interaktive Elemente die Wahrnehmung, das Selbstverständnis und die Aushandlung von Werten und Zugehörigkeiten prägen (vgl. Jörisen/Unterberg 2019, S. 20). Beispielsweise bieten Gaming-Kulturen Spielenden Räume zur Identitätsbildung, in denen sie Rollen erproben, soziale Dynamiken aushandeln und kreative Ausdrucksformen entwickeln können. Digitale Spielwelten ermöglichen so eine tiefgehende Involvierung, die über das Spiel hinaus in digitale Kommunikationsräume hineinwirkt (vgl. Tappe/Gennat 2021). Exemplarisch ist hier Discord zu nennen, welche als Community-Plattform fungiert, auf der Spielinhalte weiterverhandelt, soziale Zugehörigkeiten gefestigt und Identitätsentwürfe reflektiert werden. In der sozialpädagogischen Praxis eröffnet dies neue Möglichkeiten, Jugendliche in ihren digitalen Lebenswelten zu begleiten und an ihre Medienerfahrungen anzuknüpfen. Initiativen wie *Digital Streetwork Bayern*² nutzen Discord gezielt, um niedrigschwellige Beratung und Unterstützung bereitzustellen, während die AJS NRW und die Fachstelle für Jugendmedienkultur NRW Konzepte³ entwickelt haben, um die Plattform sicher und pädagogisch

2 Vgl. <https://www.digital-streetwork-bayern.de/fach/> (Abfrage: 14.01.2025).

3 Vgl. https://ajs.nrw/wp-content/uploads/2020/04/Discord-in-der-Kinder-und-Jugendarbeit_Statement-AJS-NRW-und-FJMK-NRW_27.04.2020.pdf (Abfrage: 14.01.2025).

sinnvoll in der Jugendarbeit einzusetzen. Diese Entwicklungen unterstreichen die Notwendigkeit, digitale Kommunikationsräume als zentrale Bestandteile jugendlicher Aneignungsprozesse anzuerkennen und ihre Potenziale für sozialpädagogische Konzepte zu nutzen. Die enge Verknüpfung zwischen medialen Kulturen (wie z. B. Gaming), digitalen Kommunikationsräumen und sozialer Interaktion zeigt, dass mediale Umgebungen zunehmend als zentrale Orte der Identitätsaushandlung fungieren. In diesem Zusammenhang lässt sich auch von „Medienidentitäten“ sprechen, da in modernen Medienkulturen ein erheblicher Teil der Identitätsressourcen durch translokale Diskurse und mediale Repräsentationen vermittelt wird. Diese Identitäten entstehen im Zusammenspiel individueller Artikulationen und kollektiver Bedeutungsressourcen, die über mediale Kanäle verhandelt werden. Medienidentitäten spiegeln somit die enge Verknüpfung von persönlichen und kulturellen Bedeutungen wider und verdeutlichen die Rolle medialer Kommunikation als zentraler Bestandteil heutiger Identitätsbildung (vgl. Krönert/Hepp 2015, S. 271).

Obwohl medial erfahrbare Kulturräume eine prägende Wirkung für Identitätsprozesse besitzen, sind Nutzende ihnen nicht per se passiv ausgeliefert. Sie entscheiden aktiv, wie sie entsprechende Räume in ihre alltäglichen Kontexte einbetten, und grenzen sie bewusst von anderen Handlungsoptionen ab, um ihnen eine spezifische Bedeutung zu geben. Dabei orientieren sie sich gezielt an ihren Medienkulturen und wissen, welche Angebote sie für die Umsetzung ihrer Absichten nutzen und welche Handlungsmuster sie dafür einsetzen möchten (vgl. Spanhel 2017, S. 6f.; Brüggem/Wagner 2017, S. 225f.). Entsprechende Entwicklungsmöglichkeiten zeichnen sich durch eine zunehmende Individualisierung und eine Vielzahl an Handlungsmöglichkeiten aus, die für die Identitätsarbeit allgemein sowohl Chancen als auch Spannungsfelder mit sich bringen. Marotzki schlussfolgert, dass die Vielzahl an Wahloptionen in modernen Gesellschaften jedoch zu einer wachsenden Orientierungslosigkeit führt, da jede Entscheidung zur Wahl zwischen vielen Alternativen wird. Dies „zwingt“ Individuen, ihre Identität kontinuierlich neu auszuhandeln und experimentelle Lebensweisen als dauerhafte Strategie zur Bewältigung dieser Unsicherheit zu entwickeln (vgl. Marotzki in Leineweber 2020, S. 39f.). Keupp beschreibt diesen Prozess als „Identitätsarbeit“, bei der Menschen ihre fragmentierten Alltagserfahrungen und Einflüsse kreativ zu einem stimmigen Selbstbild zusammensetzen müssen. Diese Identitätsarbeit, die durch die Metapher des „Patchworks“ veranschaulicht wird, verlangt eine hohe Eigenleistung, da die Subjekte gefordert sind, ein kohärentes Identitätsmuster zu entwickeln, das sowohl innere Authentizität als auch soziale Anerkennung gewährleistet. Dabei betont Keupp, dass diese Prozesse stets in einem spezifischen soziokulturellen Kontext stattfinden, der unterstützend oder hemmend wirken kann. Besonders in Phasen gesellschaftlicher Instabilität, die er als „Entbettung“ bezeichnet, fehlen klare Orientierungsmuster, sodass Indivi-

duen verstärkt auf eigene Lösungen angewiesen sind (vgl. Keupp 2019, S. 44–47; 57f.).

In diesem Kontext stellt sich die weitreichende Frage, ob und inwieweit KI eine zentrale Rolle an der Schnittstelle zwischen individueller medialer Weltaneignung und der wachsenden, durch gesellschaftliche Komplexität bedingten Orientierungslosigkeit übernehmen könnte. Es wäre denkbar, dass KI-Systeme als eine Art Scharnier in den Identitätsbildungsprozessen fungieren, indem sie nicht nur Informationen filtern und bereitstellen, sondern auch als Orientierungshilfen in einem zunehmend fragmentierten sozialen und medienkulturellen Raum dienen. Die sozialpädagogische Praxis zeigt bereits exemplarisch auf, wie KI-Systeme an ebenjener Schnittstelle agieren könnten. So wird beispielsweise bereits der KI-gestützte Chatbot „Marie“ im AWO-Kreisverband Wesel erprobt. Dieses System soll Klient:innen der Sozialen Arbeit beratend unterstützen, indem es auf häufig gestellte Fragen eingeht und als erste Anlaufstelle für Hilfesuchende fungiert.⁴ Weiterentwicklungen solcher Systeme wären potenziell in der Lage, Klient:innen nicht nur eine erste Anlaufstelle, sondern vielmehr eine hochpersonalisierte Lebensbegleitung zu bieten. Dabei ließen sich individuell abgestimmte interaktive und mediengestützte Formate (wie beispielsweise Voice-Chats, visuelle Avatare oder KI-generierte virtuelle Umgebungen) gezielt einsetzen, um eine für den Nutzenden ansprechende und unterstützende Kommunikation zu ermöglichen. Vor diesem Hintergrund muss kritisch reflektiert werden, inwieweit solche Anwendungen die individuellen Bedürfnisse und komplexen Lebensrealitäten ihrer Nutzer:innen tatsächlich erfassen können. Da KI-Modelle auf algorithmische Muster zurückgreifen, besteht die Gefahr, dass sie bestehende (populäre) gesellschaftliche oder soziale Normen und Strukturen nicht hinterfragen, sondern eher verstärken. Dies wäre insbesondere dann problematisch, wenn Nutzer:innen ein emotionales Vertrauensverhältnis zu einem KI-Agenten aufbauen, ohne dass dieser die Tiefe menschlicher Erfahrungen und sozialen Kontextes vollständig erfassen kann. So verdeutlicht die Media-Equation-Theorie, dass Menschen bereits gegenwärtig interaktiven Technologien wie Chatbots und virtuellen Assistenten häufig menschliche Eigenschaften zuschreiben und Vertrauen zu ihnen aufbauen. Diese Tendenz zur Anthropomorphisierung führt dazu, dass Nutzer:innen KI-Systeme als soziale Akteure wahrnehmen, mit ihnen kommunizieren und sogar emotionale Bindungen aufbauen (vgl. Reeves/Nass 1996; Nowak/Rauh 2005; Nass/Brave 2007; Teich 2020). Es wird erwartet, dass KI zur Förderung dieser Mensch-Maschinen-Interaktion immer stärker auf persönliche Daten zugreifen wird, um individuelle Präferenzen zu berücksichtigen und die Interaktion mit Nutzer:innen sozialer und personalisierter zu gestalten. Dadurch könnte sich eine Form der hybriden Intelligenz entwickeln, in der Mensch und Maschine zuneh-

4 Vgl. <https://www.milestone-consult.de/articles/ki-chatbot-marie-die-zukunft-der-sozialen-arbeit-beim-awo-kreisverband-wesel> (Abfrage: 17.01.2025).

mend kooperativ agieren und KI nicht nur als Werkzeug, sondern als gleichwertige:r Partner:in fungiert (vgl. Wesche/Sonderegger 2019, S. 197 f.; Hasenbein 2023, S. 61). Es ist anzunehmen, dass durch eine derartige kontinuierliche Anpassung an die persönlichen Interessen, Gewohnheiten und Präferenzen der Nutzer:innen KI-Systeme bestimmte Verhaltensweisen und Selbstbilder verstärken, verändern oder infrage stellen können und so aktiv in den Prozess der Identitätsbildung eingreifen würden. Wie in Abschnitt 2 dieses Beitrages dargelegt wurde, sind algorithmenbasierte KI-Systeme jedoch nicht neutral in ihrer Ausgestaltung. Sie spiegeln die Werte, Normen und kulturellen Leitbilder der jeweiligen Medienkulturen sowie von marktökonomischen Interessen wider, in die sie eingebettet sind.

Eine derartige durch medienkulturelle Aneignungsräume induzierte Einflussnahme durch KI hängt jedoch nicht unbedingt davon ab, ob entsprechende generative Systeme über eine eigenständige Kreativität zur Produktion medialer (und ggf. sinnstiftender) Referenzobjekte verfügen. Vielmehr ist entscheidend, ob die menschlichen Nutzer:innen den von KI gestalteten, auf Wahrscheinlichkeiten basierenden Output als kreativ und bedeutsam anerkennen und wertschätzen (vgl. Ahlborn 2023, S. 77). Nutzer:innen nehmen selbsttätig von KI erzeugte, für sie relevante Inhalte aktiv auf, integrieren sie in ihre eigenen (ästhetischen) Handlungsweisen und verleihen ihnen dadurch neue kulturelle und soziale Bedeutungen innerhalb ihrer individuellen Wahrnehmungs- und Erfahrungsprozesse (ebd., S. 88). Gleichzeitig bleibt es für KI-Systeme „schwierig“, den Kontext und die spezifischen Interpretationen von Inhalten zu erfassen, da sie stets mit einer vorgeprägten, algorithmischen Struktur arbeiten, die eine neutrale Analyse nahezu unmöglich macht. Selbst wenn alternative Klassifikationsverfahren genutzt werden, um diese Vorselektion zu umgehen, bleiben die Ergebnisse stark durch die ursprüngliche Zielsetzung der Algorithmen geprägt und würden nicht die ursprüngliche Vielfalt der kulturellen Ausdrucksformen widerspiegeln (vgl. Rust, 2017, S. 20 f.). Besonders im sozialpädagogischen Kontext gewinnt die Auseinandersetzung mit KI-generierten Inhalten an Bedeutung, wenn es um die Verhandlung von kultureller Zugehörigkeit, Erinnerung und Identität innerhalb von Gruppen geht. In Migrations- oder Altenhilfeprojekten könnten beispielsweise KI-gestützte visuelle Narrative genutzt werden, um individuelle oder kollektive Herkunftsorte digital zu rekonstruieren und emotionale Bezüge zu vergangenen Lebenswelten herzustellen. Dies könnte durch KI-gesteuerte Personas ergänzt werden, die interaktiv Erinnerungen nachzeichnen oder kulturelle Kontexte simulieren, um lebensgeschichtliche Erfahrungen erfahrbar zu machen. Doch gerade hier bedarf es einer kritischen Reflexion der zugrunde liegenden algorithmischen Prozesse: Da KI-Systeme auf statistischen Wahrscheinlichkeiten basieren, steuern sie aktiv die Auswahl und Darstellung von Bildern und Geschichten. Damit formen sie nicht nur die visuelle Repräsentation vergangener Lebenswelten, sondern beeinflussen auch, welche kulturellen Referenzpunkte sichtbar gemacht oder möglicherweise unsichtbar

bleiben. Dies wirft zentrale Fragen hinsichtlich der Autonomie von Nutzer:innen in der aktiven Mitgestaltung und Interpretation ihrer eigenen Erinnerungs- und Identitätsprozesse auf.

Algorithmische Systeme, insbesondere im Kontext von KI, bilden demnach lediglich eine stark gefilterte und technisch vorstrukturierte Version der Realität ab. Dies bedeutet, dass die von KI generierten Inhalte und die durch Algorithmen gesteuerten Auswahlprozesse nicht als neutrale Reflexionen der sozialen Wirklichkeit verstanden werden dürfen. Vielmehr formen sie aktiv das, was als „bedeutsam“ oder „sichtbar“ erachtet wird, und beeinflussen dadurch die Wahrnehmung und Interpretation der Nutzer:innen. Die genannten Prozesse führen zu einer Verschiebung sowohl der (sinnlichen) Wahrnehmung als auch der kulturellen Dispositionen, die auf digitalen Technologien und ästhetischen Praktiken basieren (vgl. Jörissen 2018, S. 62). Ästhetische und kulturelle Erfahrungen entstehen demnach in einem komplexen Zusammenspiel von Wahrnehmung, Bewertung und Ausdruck, das durch soziale sowie technische Kontexte geprägt ist (vgl. Allert et al. 2019, S. 66). Diese Prozesse stehen in wechselseitiger Beziehung zueinander und sind eng mit den gesellschaftlichen Rahmenbedingungen und technischen Gegebenheiten verknüpft, innerhalb derer sie stattfinden.

4 Welchen Beitrag kann eine medienpädagogische Perspektive auf KI in der Sozialen Arbeit leisten?

Hoffmann betont die Dringlichkeit einer verstärkten Zusammenarbeit zwischen Sozialer Arbeit und Medienpädagogik, um den Herausforderungen der Digitalisierung wirkungsvoll zu begegnen. Beide Disziplinen verfolgen das gemeinsame Ziel, Teilhabe, Bildungsgerechtigkeit und Selbstwirksamkeit zu fördern (vgl. Hoffmann 2020, S. 53). Im Kontext einer Kultur der Digitalität bedeutet dies auch, einen kritischen Umgang mit digitalen Medien und Technologien zu entwickeln, da diese zunehmend Einfluss auf das soziale Handeln und die gesellschaftliche Partizipation ausüben. Dabei steht die Befähigung von Individuen im Vordergrund, sich aktiv und reflektiert in einer mediatisierten Welt zu orientieren und ihre Möglichkeiten zur Mitgestaltung zu nutzen. Die Vermittlung von Medienkompetenz müsse emanzipatorisch ausgerichtet sein und den Einzelnen die Fähigkeit verleihen, sich gegen die „Macht der Algorithmen“ zu behaupten. Gleichzeitig wird deutlich, dass hierfür nicht nur individuelle Kompetenzen, sondern auch politische und soziale Rahmenbedingungen erforderlich sind, um Selbstbestimmung und den Schutz privater Daten zu gewährleisten (vgl. ebd., S. 53 f.).

Im Sinne einer dafür notwendigen kritisch-reflexiven medienpädagogischen Arbeit (vgl. Niesyto 2022) ist es entscheidend, den Einsatz von KI sowohl als Teil einer technologischen Entwicklung als auch im Kontext ihrer soziokulturellen

Implikationen sowie vor dem Hintergrund von individuellen, gesellschaftlichen sowie politischen und marktökonomischen Machtverhältnissen zu betrachten. Dabei geht es nicht nur um die Bewältigung individueller Herausforderungen, sondern ebenso um die Schaffung gerechter Rahmenbedingungen, die eine selbstbestimmte und partizipative Lebensführung ermöglichen (vgl. Hoffmann 2020, S. 45 f.). Die dargestellten und prognostizierten Einflüsse von KI-Systemen auf das medienkulturelle Handeln und die damit verknüpften Identitätsprozesse von Klient:innen erfordern somit eine verstärkte Reflexion innerhalb der Sozialen Arbeit. Fachkräfte sind zunehmend gefordert, sich mit den veränderten medialen Erfahrungen und Erwartungen ihrer Klient:innen auseinanderzusetzen, da deren Lebenswelten zunehmend durch Interaktionen mit KI durchdrungen und geprägt werden. Bereits jetzt lässt sich beispielsweise beobachten, dass soziale KI-Agenten wie Replika oder Snapchats My AI eine stetig wachsende Nutzungsfrequenz verzeichnen und zunehmend in alltägliche soziale Interaktionen integriert werden. Diese Systeme übernehmen nicht mehr nur eine assistierende Funktion, sondern können Beziehungen simulieren, die von Nutzer:innen als authentisch empfunden werden (vgl. Brandtzaeg 2022; Ofcom 2024). Derartige Entwicklungen werfen zentrale Fragen für die Soziale Arbeit auf: Wie verändern KI-gestützte soziale Interaktionen das Verständnis von Beziehung und Identität? Wie können Fachkräfte sicherstellen, dass diese Technologien die Selbstbestimmung und soziale Teilhabe der Klient:innen fördern, statt sie durch algorithmische Steuerung einzuschränken? Und welche Herausforderungen und Chancen ergeben sich daraus für pädagogische Begleitung und Unterstützung?

Ein möglicher Ansatz zur Bearbeitung dieser und vergleichbarer Herausforderungen liegt in einer ästhetisch-kreativen Auseinandersetzung mit KI-generierten Inhalten. Im Sinne einer handlungsorientierten Medienpädagogik (vgl. Schorb 2022, S. 44) geht es dabei nicht nur um die analytische Reflexion algorithmischer Prozesse, sondern auch um die aktive gestalterische Weiterverarbeitung entsprechender Inhalte. Dadurch wird ein kritischer Umgang mit den Wirkmechanismen von KI gefördert, indem kreative Interventionen genutzt werden, um algorithmische Strukturen zu hinterfragen und neu zu interpretieren. Das Wissen, das in diesen künstlerischen Reflexionen gewonnen wird, kann kulturpädagogische Bildungsprozesse bereichern und in sozialarbeiterischen Kontexten gezielt genutzt werden (vgl. Jörissen/Unterberg 2019, S. 19 ff.). Ästhetisch-gestalterische Prozesse fördern dabei ein hohes Maß an Engagement und motivieren durch sinnliche Beteiligung zur Reflexion. Indem Menschen sich aktiv handelnd mit medialen Zeichensystemen (z. B. Bild- und Filmsprache) und den ihnen zugrunde liegenden bedeutungstragenden Codes (z. B. Bildperspektiven in Social Media, Farbgestaltungen in Filmen oder Klangfarben in Werbejingles) auseinandersetzen, erwerben sie nicht nur ein tiefes Verständnis für deren ästhetische Prinzipien, sondern auch für deren spezifischen Wirkungsweisen (vgl. Tappe 2018, S. 146 ff.; Jörissen/Unterberg 2019, S. 20). Neben der ästhetisch-gestalte-

rischen Auseinandersetzung ist es entscheidend, ein fundiertes Verständnis der technischen Systeme und Prozesse zu entwickeln, die hinter den Interfaces von KI verborgen sind. Durch diese tiefere Einsicht ist es möglich, Medien und Technologien auf einer kompetenten Ebene zu analysieren, kritisch zu reflektieren und fundierte Urteile zu fällen (vgl. Knaus 2020, S. 49). In diesem Zusammenhang gewinnen digitale Medienkulturen an Bedeutung, da sie (beispielsweise in Onlinecommunities oder digitalen Spielwelten) nicht nur als Kommunikations- und Begegnungsorte fungieren, sondern auch als lebensweltorientierte Bildungsräume, in denen durch die Vernetzung, Digitalisierung und die Konstruktion virtueller Realitäten neue Bildungsprozesse initiiert werden können (vgl. Spanhel 2017, S. 2). Durch den gezielten Einsatz von Tools, die den Zugang zu den algorithmischen Prozessen erleichtern (etwa offene Datenplattformen oder KI-Sandboxen, die eine sichere Testumgebungen zur Erprobung von KI-Szenarien bieten), könnten Nutzer:innen lernen, wie algorithmische Entscheidungen getroffen werden, welche Datenstrukturen ihnen zugrunde liegen und wie sie die mediale Wirklichkeit beeinflussen. Dies ermöglicht eine kritische Reflexion darüber, wie Algorithmen Inhalte selektieren, filtern und priorisieren – und welche gesellschaftlichen Auswirkungen dies hat (vgl. Knaus 2020, S. 49). Eine entsprechende konzeptionelle Planung könnte beispielsweise folgendermaßen gestaltet werden:

Praxisszenario „KI und Identität – Wer bestimmt, wer ich bin?“

In einem offenen Jugendtreff reflektieren Jugendliche in einem medienpädagogischen Workshop den Einfluss von KI-gestützten Medien auf die Selbstdarstellung in sozialen Medien. Mit Methoden der aktiven Medienarbeit hinterfragen sie bewusst die Funktionsweise dieser Technologien und setzen sich mit dem Spannungsfeld zwischen algorithmischer Steuerung und Selbstbestimmung auseinander. So stellt ein zentraler Bestandteil des Workshops das Remixing von KI-generierten und selbst erstellten Medieninhalten dar. Die Teilnehmenden verwenden KI-gestützte Bild- und Textgeneratoren, um alternative Versionen ihrer Social-Media-Profile zu erzeugen – beispielsweise durch algorithmisch erstellte Selbstporträts oder autobiografische Texte. In der anschließenden Reflexionsphase vergleichen sie diese mit ihren vorhandenen Onlinedarstellungen und analysieren, inwiefern KI-Systeme bestehende Narrative über Identität, Ästhetik und Authentizität verstärken oder verzerren. Dabei diskutieren sie, welche Werte und Stereotype durch algorithmische Prozesse reproduziert werden und inwieweit ihre eigene Mediennutzung von diesen beeinflusst ist.

Parallel dazu erhalten die Teilnehmenden Einblicke in die Funktionsweise von Algorithmen: Die Jugendlichen setzen sich praktisch mit Formen von algorithmischer Steuerung auseinander, indem sie untersuchen, welche Daten für KI-generierte Inhalte verarbeitet werden und wie Filtermechanismen ihre Medienwahrnehmung beeinflussen. Durch gezielte Manipulation von Suchverläufen und die Simulation verschiedener Nutzerprofile erleben sie direkt, wie Algorithmen Vorlieben antizipieren und Inhalte selektieren. Dabei reflektieren sie kritisch die Auswirkungen personalisierter Empfehlungen auf ihre digitalen Routinen und Identitätsentwürfe. Anhand konkreter Beispiele analysieren sie, inwiefern

algorithmische Systeme Stereotype verstärken oder bestimmte Identitätsmodelle bevorzugt darstellen, und entwickeln in interaktiven Gruppenarbeiten alternative Perspektiven auf mediale Vielfalt und Repräsentation.

Abschließend entwickeln die Teilnehmenden praxisnahe Strategien für einen bewussten Umgang mit KI-generierten Medien und deren Einfluss auf Social Media. In Kleingruppen erstellen sie Leitfäden zur kritischen Mediennutzung, gestalten kreative Formate wie digitale Collagen oder interaktive Storytelling-Projekte und setzen ihre Erkenntnisse gezielt in Social-Media-Kampagnen um. Die Veröffentlichung ihrer Ergebnisse auf populären Social-Media-Plattformen sowie die Diskussion in Livestreams ermöglichen eine direkte Auseinandersetzung mit einem breiteren Publikum. Dadurch wird nicht nur die kritische Medienkompetenz gestärkt, sondern auch ein Bewusstsein für digitale Selbstbestimmung und gesellschaftliche Teilhabe geschaffen.

Begleitete, sozialpädagogisch gerahmte Settings besitzen das Potenzial, als kulturelle Gestaltungsräume zu fungieren, in denen Identitätswürfe und soziale Interaktionen neu verhandelt werden können. Hierbei sind Fachkräfte gefordert, die Bedeutung solcher Räume für die Lebenswirklichkeiten ihrer Klient:innen zu reflektieren und sie entsprechend in einer medienbezogenen kritischen Handlungsfähigkeit zu stärken. KI-Systeme sind daher nicht nur Werkzeuge der Kommunikation, sondern werden zu Akteuren innerhalb sozialer Prozesse, deren Einfluss auf Identität und Teilhabe systematisch erschlossen und kritisch hinterfragt werden muss. Vor dem Hintergrund von Mediatisierung und Digitalität sowie unter Berücksichtigung von KI als potenzieller sozialer Interaktionspartnerin von Klient:innen ist es nicht nur entscheidend, welche neuen Anforderungen die Mensch-Medien-Kommunikation mit sich bringt, sondern auch, wie sie im Vergleich zur Mensch-Mensch-Kommunikation zu bewerten ist (vgl. Cleppien/Hoffmann 2020, S. 65). Eine medienpädagogisch orientierte Soziale Arbeit, die das Bedürfnis der Klient:innen nach medialen Entwicklungsmöglichkeiten ernst nimmt und zugleich Wege zur kritischen Reflexion der ästhetischen und technischen Wirkungsweisen von KI sowie ihrer Rolle in der eigenen Identitätskonstruktion aufzeigt, könnte maßgeblich dazu beitragen, ihre Selbstbestimmung zu fördern. Durch die aktive Auseinandersetzung mit den Funktionsmechanismen von KI-Systemen und deren Einfluss auf soziale und kulturelle Prozesse würden die Klient:innen nicht nur zu kompetenten Mediennutzer:innen, sondern auch zu kritischen Gestalter:innen ihrer eigenen Lebenswelten werden.

Literatur

Adolf, Marian (2017): Zwei Gesichter der Mediatisierung? Ein Beitrag zur theoretischen Fundierung der Mediatisierungsforschung und ihres Verhältnisses zur Mediensozialisationsforschung. In: Hoffmann, Dagmar / Krotz, Friedrich / Reißmann, Wolfgang (Hrsg.): Mediatisierung und Mediensozialisation. Prozesse – Räume – Praktiken, Wiesbaden: Springer VS, S. 41–58.

- Ahlborn, Juliane (2023): Zur (Un-)Berechenbarkeit der Künste. Wie algorithmische Strukturen die Bedingungen für Ästhetik und ästhetische Bildung verändern. In: de Witt, Claudia/Gloerfeld, Christina/Wrede, Silke Elisabeth (Hrsg.): *Künstliche Intelligenz in der Bildung*. Wiesbaden: Springer, S. 69–88.
- Allert Heidrun/Ide, Martina/Richter, Christoph/Schröder, Christoph/Thiele, Sabrina (2019): DiKu-Bi-on: Soziale Medien als kultureller Bildungsraum Das Onlinelabor für Digitale Kulturelle Bildung. In: Jörissen, Benjamin/Kröner, Stephan/Unterberg, Lisa (Hrsg.): *Forschung zur Digitalisierung in der Kulturellen Bildung*. München: kopaed (Kulturelle Bildung und Digitalität; 1), S. 63–78. <https://doi.org/10.25656/01:26963>
- Bettinger, Patrick (2020): Medienpädagogische Forschung und Kritik – Spannungsfelder und Positionsbestimmungen. In: Dander, Valentin/Bettinger, Patrick/Ferraro, Estalla/Leineweber, Christian/Rummler, Klaus (Hrsg.): *Digitalisierung – Subjekt – Bildung. Kritische Betrachtung der digitalen Transformation*. Opladen, Berlin und Toronto: Barbara Budrich, S. 234–250.
- Brandtzaeg, Petter Bae/Skjuve, Marita/Følstad, Asbjørn (2022): My AI Friend: How Users of a Social Chatbot Understand Their Human–AI Friendship. In: *Human Communication Research* 48, Issue 3, July 2022, S. 404–429. <https://doi.org/10.1093/hcr/hqac008>
- Brüggen, Niels/Wagner, Ulrike (2017): Medienaneignung und sozialraumbezogenes Medienhandeln von Jugendlichen., In: Hoffmann, Dagmar/Krotz, Friedrich/Reißmann, Wolfgang (Hrsg.): *Mediatisierung und Mediensozialisation. Prozesse – Räume – Praktiken*, Wiesbaden: Springer VS, S. 211–228.
- Düll, Nicola (2016): Digitalisierung der Arbeitswelt – grundlegende Thesen. In: Düll, Nicola (Hrsg.): *Arbeitsmarkt 2030 – Digitalisierung der Arbeitswelt. Fachexpertise zur Prognose 2016*. Bielefeld: Bertelsmann, S. 6–21.
- Gapski, Harald (2022): Diskussionsfelder der Medienpädagogik: Datafizierte Lebenswelten und Datenschutz. In: Sander, Uwe/von Gross, Friederike/Hugger, Kai-Uwe (Hrsg.): *Handbuch Medienpädagogik*. 2. Auflage. Wiesbaden: Springer VS, S. 693–701.
- Hasenbein, Melanie (2023): Mensch und KI in Organisationen. Einfluss und Umsetzung Künstlicher Intelligenz in wirtschaftspsychologischen Anwendungsfeldern. Berlin: Springer.
- Hepp, Andreas/Hasebrink, Uwe (2017): Kommunikative Figurationen. Ein konzeptioneller Rahmen zur Erforschung kommunikativer Konstruktionsprozesse in Zeiten tiefgreifender Mediatisierung. In: *Medien & Kommunikationswissenschaft* 65(2), S. 330–347.
- Hoffmann, Bernward (2019): Medien-Erziehungs-Kompetenz von Eltern im System Familie. In: Angenent, Holger/Heidkamp, Birte/Kergel, David (Hrsg.): *Digital Diversity. Diversität und Bildung im digitalen Zeitalter*. Wiesbaden: Springer VS.
- Hoffmann, Bernward (2020): Medienpädagogik und Soziale Arbeit – kongruent, komplementär oder konträr im Umgang mit Digitalisierung und Mediatisierung. In: Kutscher, Nadia/Ley, Thomas/Seelmeyer, Udo/Siller, Friederike/Tillmann, Angela/Zorn, Isabel (Hrsg.): *Handbuch Soziale Arbeit und Digitalisierung*. Weinheim und Basel: Beltz Juventa, S. 42–57.
- Jörissen, Benjamin (2018): Subjektivation und ästhetische Bildung in der post-digitalen Kultur. In: *Vierteljahrsschrift für wissenschaftliche Pädagogik* 94(1), S. 51–70.
- Jörissen, Benjamin/Unterberg, Lisa (2019): DiKuBi-Meta [TP1]: Digitalität und Kulturelle Bildung., In: Jörissen, Benjamin/Kröner, Stephan/Unterberg, Lisa (Hrsg.): *Forschung zur Digitalisierung in der Kulturellen Bildung*. München: kopaed 2019, S. 11–24.
- Knaus, Thomas (2020): Von medialen und technischen Handlungspotentialen, Interfaces und anderen Schnittstellen. In: Knaus, Thomas/Merz, Olga (Hrsg.): *Schnittstellen und Interfaces – Digitaler Wandel in Bildungseinrichtungen* (Bd. 7). München: kopaed, S. 15–72.
- Krönert, Veronika/Hepp, Andreas (2015): Identität und Identifikation. In: Hepp, Andreas/Krotz, Friedrich/Lingenberg, Swantje/Wimmer, Jeffrey (Hrsg.): *Handbuch Cultural Studies und Medienanalyse*. Wiesbaden: Springer VS, S. 265–273.
- Krotz, Friedrich (2007): *Mediatisierung. Fallstudien zum Wandel von Kommunikation*. Wiesbaden: VS.

- Krotz, Friedrich (2015): Medienwandel in der Perspektive der Mediatisierungsforschung: Annäherung an ein Konzept. In: Kinnebrock, Susanne/Schwarzenegger, Christian/Birkner, Thomas (Hrsg.): *Theorien des Medienwandels*. Köln: Halem, S. 119–141.
- Krotz, Friedrich (2019): Die Begegnung von Mensch und Roboter. Überlegungen zu ethischen Fragen aus der Perspektive des Mediatisierungsansatzes. In: Rath, Matthias/Krotz, Friedrich/Karmasin, Matthias (Hrsg.): *Maschinenethik. Normative Grenzen autonomer Systeme*. Wiesbaden: Springer VS, S. 13–34.
- Krotz, Friedrich (2022): Medienpädagogik und Mediatisierungsforschung. In: Sander, Uwe/von Gross, Friederike/Hugger, Kai-Uwe (Hrsg.): *Handbuch Medienpädagogik*. 2. Auflage. Wiesbaden: Springer VS, S. 205–213.
- Leineweber, Christian (2020): Digitale Bildung und Entfremdung – Versuch einer normativ-kritischen Verhältnisbestimmung. In: Dander, Valentin/Bettinger, Patrick/Ferraro, Estalla/Leineweber, Christian/Rummeler, Klaus (Hrsg.): *Digitalisierung – Subjekt – Bildung. Kritische Betrachtung der digitalen Transformation*. Opladen, Berlin und Toronto: Barbara Budrich, S. 38–56.
- Nass, Clifford/Brave, Scott (2007): *Wired for Speech – How Voice Activates and Advances the Human-Computer Relationship*. Cambridge, MA: MIT Press.
- Niesyto, Horst (2022): Medienkritik. In: Sander, Uwe/von Gross, Friederike/Hugger, Kai-Uwe (Hrsg.): *Handbuch Medienpädagogik*. 2. Auflage. Wiesbaden: Springer VS, S. 125–135
- Nowak, Kristine/Rauh, Christian (2005): The influence of the avatar on online perceptions of anthropomorphism, androgyny, credibility, homophily, and attraction. In: *Journal of Computer-Mediated Communication* 11(1), S. 153–178.
- Ofcom (2024): *Online Nation 2024 Report*. <https://www.ofcom.org.uk/siteassets/resources/documents/research-and-data/online-research/online-nation/2024/online-nation-2024-report.pdf> (Abfrage 15.06.2025).
- Reeves, Byron/Nass, Clifford (1996): *The media equation: How people treat computers, television, and new media like real people and places*. Cambridge: Cambridge University Press.
- Reynolds, Robert G. (1994): An Introduction to Cultural Algorithms. In: *Proceedings of the 3rd Annual Conference on Evolutionary Programming*. World Scientific Publishing, S. 131–139.
- Rust, Holger (2017): *Virtuelle Bilderwolken. Eine qualitative Big Data-Analyse der Geschmackskulturen im Internet*. Wiesbaden: Springer Fachmedien.
- Schorb, Bernd (2022): Handlungsorientierte Medienpädagogik. In: Sander, Uwe/von Gross, Friederike/Hugger, Kai-Uwe (Hrsg.): *Handbuch Medienpädagogik*. 2. Auflage. Wiesbaden: Springer VS, S. 41–55.
- Spanhel, Dieter (2017): Mediale Bildungsräume – Spielräume der Freiheit für Bildungsprozesse in realen und virtuellen Lebenswelten? In: *MedienPädagogik: Zeitschrift für Theorie und Praxis der Medienbildung*, S. 1–18. <https://doi.org/10.21240/mpaed/00/2017.03.02.X>
- Stalder, Felix (2016): *Kultur der Digitalität*. Berlin: Suhrkamp.
- Tappe, Eik-Henning (2018): Lernen durch Mediengestaltung – Entwicklung eines Konzeptes zur Unterstützung mediendidaktischer Lehre im Schulalltag. Münster: Universitäts- und Landesbibliothek Münster. <https://nbn-resolving.org/urn:nbn:de:hbz:6-08109654421> (Abfrage: 11.08.2024).
- Tappe, Eik-Henning/Gennat, Markus (2021): Gamingkultur und Identität. In: *Kinder- und Jugendschutz in Wissenschaft und Praxis* 66(3), S. 90–94.
- Teich, Irene (2020): Meilensteine der Entwicklung Künstlicher Intelligenz. In: *Informatik Spektrum* 43, S. 276–284.
- Wesche, Jenny S./Sonderegger, Andreas (2019): When computers take the lead: The automation of leadership. In: *Computers in Human Behavior* 101, S. 197–209.

KI in der Beratung¹

Robert Lehmann

Abstract: Der Beitrag behandelt die wachsende Bedeutung Künstlicher Intelligenz (KI) in der psychosozialen Beratung und zeigt auf, wie sich digitale Technologien auf die professionelle Praxis der Beratung auswirken. Aufbauend auf der Entwicklung von Präsenz- zur Onlineberatung werden konkrete Anwendungsszenarien skizziert, in denen KI bereits eingesetzt wird oder kurz vor der Umsetzung steht. In schriftbasierten Formaten ermöglicht KI z. B. die Analyse und Reflexion professionellen Handelns. In der Ausbildung werden KI-gestützte Simulationen eingesetzt, um Beratungsgespräche realitätsnah zu üben. Zudem unterstützen KI-Systeme Fachkräfte bei der Einschätzung komplexer Beratungsprozesse. Für Ratsuchende können KI-Anwendungen Barrieren abbauen und Zugänge erleichtern. Perspektivisch erscheinen vollständig KI-gestützte Beratungsszenarien denkbar. Dabei ist eine ethisch fundierte und fachlich begleitete Integration von KI zentral, um die Möglichkeiten der KI zum Wohl der Ratsuchenden zu entfalten.

Keywords: Künstliche Intelligenz, Beratung, Onlineberatung, Soziale Arbeit.

1 Einführung

Der Begriff „Künstliche Intelligenz“ (KI) umfasst eine Vielzahl an Methoden und Anwendungen der Informatik, die mittlerweile in diversen Bereichen der Gesellschaft Einzug gehalten haben (vgl. Albrecht/Rudolph 2023). Während in Bereichen der Sozialen Arbeit, die physische Interaktionen mit den Zielgruppen erfordern (z. B. in stationären Betreuungseinrichtungen), der Einsatz von KI im zwischenmenschlichen Kontakt stark von Fortschritten in der Robotik abhängig ist, bietet der psychosozialen Beratung der Prozess der Digitalisierung die Möglichkeit, KI zunehmend in fachliche Abläufe zu integrieren. So hat die Entwicklung der Onlineberatung gezeigt, dass qualitativ hochwertige psychosoziale Beratungsleistungen auch ohne persönliche Kopräsenz realisierbar sind (vgl. Engelhardt 2021). Besonders in der schriftbasierten Onlineberatung, bei der

1 © Die Autor:innen 2025, veröffentlicht im Verlag Beltz Juventa, Open Access: CC BY 4.0
Gesa Linnemann/Julian Löhe/Beate Rottkemper (Hg.), *Künstliche Intelligenz in der Sozialen Arbeit*
10.3262/978-3-7799-8562-4_009

die gesamte Kommunikation in Textform erfolgt, ist die Interaktion zwischen Fachkräften und Ratsuchenden technisch problemlos auch zwischen Mensch und KI – beispielsweise in Form von Chatbots – möglich. Vor diesem Hintergrund erscheint es plausibel, dass der Bereich der Beratung innerhalb der Sozialen Arbeit einer der am schnellsten und nachhaltigsten durch KI geprägten Teilbereiche werden könnte (vgl. Engelhardt 2023a).

Es überrascht daher nicht, dass in diesem Feld bereits erste Prototypen von KI-basierten Beratungsanwendungen existieren und die Literatur zu diesem Thema anwächst.

Aufbauend auf eine kurze methodische Differenzierung von Präsenz- und Onlineberatung werden im Folgenden verschiedene Szenarien dargestellt, in denen KI bereits eingesetzt wird oder absehbar ist, dass entsprechende Umsetzungen entstehen. Aufbauend auf diesen Ausführungen wird am Ende versucht, die Frage zu beantworten, ob und wann KI die Beratung übernimmt.

2 Von der klassischen Beratung über Onlineberatung zur KI-Beratung?

Die Entwicklung von der klassischen Präsenzberatung hin zur Onlineberatung ist eng verknüpft mit den Fortschritten der Digitalisierung und Mediatisierung (vgl. Reindl 2018). Während psychosoziale Beratungsprozesse früher fast ausschließlich im persönlichen Kontakt stattfanden, eröffnete die Verbreitung neuer Kommunikationstechnologien zunehmend die Möglichkeit, Beratungsangebote auch in digitaler Form anzubieten. Bereits in den 1990er-Jahren nutzte die Telefonseelsorge das Internet, um erste Onlineberatungsdienste anzubieten (vgl. Lehmann 2020). Mit der Jahrtausendwende erweiterten sich diese digitalen Angebote um Mail-, Chat- und Forenberatungen, die bis heute als essenzielle Beratungsmethoden etabliert sind (vgl. Engelhardt 2021). Diese Beratungsformate ermöglichen datensicheren Austausch – asynchron in der Mailberatung oder nahezu synchron in der Chatberatung –, während Forenberatungen öffentliche oder geschlossene Räume für den Austausch schaffen (vgl. Brunner/Engelhardt/Heider 2009; Eichenberg/Kühne 2014). Die COVID-19-Pandemie führte zu einem sprunghaften Anstieg der Onlineberatung, insbesondere die Videoberatung etablierte sich in dieser Zeit (Stieler/Lipot/Lehmann 2022). Dieser Digitalisierungsschub setzte sich nach der Pandemie fort, sodass die Messenger-Beratung, eine zielgruppenorientierte Beratungsform, zunehmend Verbreitung fand (vgl. Zauter/Lehmann 2021).

Sowohl in der Praxis als auch in unterschiedlichen Studien (vgl. Arnold 2012; Eichenberg/Küsel 2016) zeigte sich, dass die verschiedenen Formen der Onlineberatung wirksam und bei bestimmten Konstellationen sogar deutlich niedrig-

schwelliger sind als Präsenzangebote, z. B. bei schambesetzten Themen (vgl. Zitzelmann 2023).

Aufgrund der spezifischen Stärken und Schwächen der Präsenzberatung und der unterschiedlichen Varianten der Onlineberatung gewinnen hybride Beratungskonzepte, die beide Zugänge kombinieren, zunehmend an Bedeutung. Diese Konzepte, die eine ideale Kombination aus Online- und Präsenzberatungen anstreben, wurden bereits vor der Pandemie unter dem Begriff „Blended Counseling“ entwickelt und werden kontinuierlich weiterentwickelt (vgl. Hörmann/Engelhardt 2022).

Ein weiteres innovatives Konzept in der Sozialen Arbeit, das digitale Beratungsmethoden integriert, ist das sogenannte Digital Streetwork (vgl. Bradl/Stieler/Engels 2022). Hierbei nutzen Fachkräfte der Sozialen Arbeit digitale Plattformen, um ihre Zielgruppen direkt in deren Online-Umfeldern zu erreichen, angelehnt an klassische Streetwork-Methoden. Unterschieden wird zwischen dem „Content-based Digital Streetwork“, bei dem Fachkräfte Inhalte auf Plattformen bereitstellen, und dem „Non-Content-based Digital Streetwork“, bei dem der direkte Kontakt zur Zielgruppe im Vordergrund steht. Hierzu gehören auch die Beobachtung und Analyse von Social-Media-Plattformen, um relevante Orte für die Zielgruppe zu identifizieren (vgl. Hagemaijer/Stuiber 2020). Erste wissenschaftliche Untersuchungen zeigen, dass diese Methode eine wertvolle Erweiterung des sozialarbeiterischen Repertoires darstellt (vgl. Lehmann/Stieler/Zauter 2024).

Zusammenfassend lässt sich festhalten, dass die Beratung als Methode schon lange in digitaler Form existiert und sich als fester Bestandteil der Beratungslandschaft etabliert hat.

3 KI als Tool der fachlichen Reflexion

KI kann im Beratungsbereich relativ einfach als Tool zur Reflexion des professionellen Handelns angewendet werden. Besonders in der schriftbasierten Onlineberatung bieten sich ideale Voraussetzungen für eine KI-gestützte Analyse. Bereits vergleichsweise einfache Methoden aus dem Bereich des unüberwachten Maschinellen Lernens ermöglichen es, große Textmengen zu strukturieren und daraus wertvolle Erkenntnisse zu gewinnen (vgl. Eckl et al. 2020). Ein Beispiel hierfür ist die Analyse von Selbsthilfeforen, bei der die zentralen Themen für die Forennutzer:innen identifiziert und mit typischen Beratungsinhalten abgeglichen wurden. Die so gewonnenen Differenzen bieten eine wertvolle Grundlage zur Weiterentwicklung der Beratungsangebote (vgl. Ghanem et al. 2021).

Da große Teile der Kommunikation auf Plattformen im Internet nach wie vor schriftbasiert sind, lassen sich mit diesen KI-Technologien schnell Einblicke in die digitalen Lebenswelten der Zielgruppen gewinnen. Dies ist etwa im Digital

Streetwork von Vorteil, da Fachkräfte so Foren effizient analysieren und gezielt entscheiden können, ob und wo sie aktiv werden sollten.

Komplexere Technologien wie sogenannte Transformer-Modelle bieten weitreichende Möglichkeiten zur inhaltlichen und methodischen Reflexion des Beratungshandelns. Diese Modelle, die auf neuronalen Netzen basieren, können mit vergleichsweise wenig technischem Aufwand darauf trainiert werden, spezifische Inhalte zu erkennen. Der Analyseprozess ähnelt hierbei der qualitativen Inhaltsanalyse (vgl. Albrecht/Rudolph 2023). In einer ersten Anwendung dieser Technologie auf Onlineberatungstexte konnten Grandeit et al. (2020) zeigen, dass eine entsprechend trainierte KI vergleichbare Erkennungsleistungen wie menschliche Kodierer:innen erzielte. Der Aufbau eines solchen Trainingsdatensatzes und die damit verbundene Optimierung der KI sind zwar komplex, jedoch auch mit den Ressourcen durchführbar, die beispielsweise Fachhochschulen oder mittelgroße Träger der freien Wohlfahrtspflege zur Verfügung stehen. Für die Onlineberatung haben Albrecht et al. (2024) einen entsprechenden Trainingsdatensatz erstellt und diesen unter einer Open-Source-Lizenz der Allgemeinheit auf dem Portal huggingface.co zugänglich gemacht.

Wendet man diese Technologien auf Beratungsprotokolle an, sei es aus der Online- oder der Präsenzberatung, können sie wertvolle Einblicke in die Inhalte und methodischen Elemente der Interaktionen liefern. So eröffnet sich die Möglichkeit, das fachliche Handeln nicht nur auf Ebene einzelner Fälle zu reflektieren, sondern systematische Erkenntnisse aus größeren Datensätzen zu gewinnen und so mit unterschiedlichen wissenschaftlichen Methoden weitergehende Erkenntnisse zu generieren. Hier ist wichtig zu bemerken, dass bei dieser Form des KI-Einsatzes die Risiken sehr überschaubar sind. Die Ergebnisse, die von der KI ausgegeben werden, können von den beteiligten Fachkräften und Wissenschaftler:innen transparent überprüft und so ihr Wahrheitsgehalt sichergestellt werden.

4 KI in der Beratungsausbildung

Basierend auf der Reflexion bisherigen Beratungshandelns eröffnet sich ein weiterer Anwendungsbereich für KI in der Ausbildung von Fachkräften. Wie bereits erwähnt, erfordert die Onlineberatung, ebenso wie die Präsenzberatung, spezifische Kompetenzen, die weder selbstverständlich vorhanden sind noch in herkömmlichen Studiengängen umfassend vermittelt werden (vgl. Engelhardt 2021). In der Praxis greifen Ausbildungsprogramme häufig auf Rollenspiele zurück, um den Lernenden Übungsmöglichkeiten zu bieten (vgl. Lippert et al. 2024). Komplexe Szenarien, die in spezialisierten Beratungslaboren teilweise unter Einsatz von professionellen Schauspieler:innen umgesetzt werden, verbessern die Ausbildungsqualität signifikant (vgl. Weinhardt et al. 2022), sind jedoch aufgrund des

hohen organisatorischen und finanziellen Aufwands flächendeckend schwer umsetzbar. Hier bietet der Einsatz von KI-Systemen neue Möglichkeiten, insbesondere bei der Simulation von Beratungsgesprächen im Rahmen der Onlineberatung (vgl. Engelhardt 2023a).

Large Language Models (LLM) bieten hierfür interessante Ansätze. An der TH Nürnberg wurde beispielsweise ein auf einem freien LLM basierender KI-Chatbot entwickelt, der auf die spezifischen Datenschutzanforderungen der EU ausgelegt ist und auf der Hardware der Hochschule betrieben wird. Dies ermöglicht es, sowohl den europäischen Datenschutzstandards gerecht zu werden als auch das Know-how in der Hochschule zu verankern (vgl. Lippert et al. 2024). Der Chatbot simuliert unterschiedliche Beratungsszenarien und gibt den Studierenden Feedback aus verschiedenen Perspektiven. Erste Untersuchungen zur Akzeptanz dieser Technologie zeigten, dass sie von den Studierenden überraschend positiv aufgenommen wurde (vgl. Albrecht/Rudolph et al. 2024).

Angesichts der Leistungsfähigkeit aktueller LLM liegt eine Erweiterung der Einsatzmöglichkeiten auf andere Formen der digitalen Beratung nahe. Darüber hinaus könnte diese Technologie auch für die Präsenzberatung relevant werden, insbesondere in Verbindung mit Virtual-Reality-Technologien (VR), die bereits in der Beratungsausbildung erfolgreich eingesetzt wurden. Ohne Einsatz von KI war in diesen Szenarien die Simulation von Ratsuchenden nur sehr aufwendig und mit Einbußen bei der Realitätsnähe verbunden (vgl. Killian et al. 2023).

Daher erscheint hier eine Kombination der Technologien sehr naheliegend. Eine KI-basierte VR-Trainingsumgebung in der Beratungsausbildung könnte für die Lernenden ein realitätsnahes Übungsfeld erschließen. Da VR zwar nah an der Realität ist, aber gerade bei der Repräsentation menschlicher Mimik und Gestik starke Abweichungen von realen Personen aufweist, sollten die Einführung und Umsetzung sehr kritisch und eingebettet in elaborierte fachliche Konzepte erfolgen. So können einerseits die Risiken dieser Abweichungen kontrolliert, andererseits die Begrenzungen der Technologie auch bewusst als Grundlage für Reflexionen innerhalb des Lernprozesses genutzt werden.

5 KI als Copilot in der Beratung

Während der Einsatz von KI in der Reflexion und Ausbildung noch keine direkte Interaktion mit Ratsuchenden umfasst, kommt man durch die Unterstützung von KI in der Beratungssituation der Vision einer KI-gestützten Beratung näher. Im Bereich der Prozessunterstützung durch KI sehen Linnemann, Löhe und Rottkemper (2023) Potenziale vor allem im Wissensmanagement und in der Entscheidungsfindung. Dabei lassen sich die bereits beschriebenen Technologien sinnvoll kombinieren. Insbesondere bei der Entscheidungsunterstützung ist jedoch auf potenzielle Risiken wie den „Automation Bias“ zu achten, bei dem es zu

einem übermäßigen Vertrauen in maschinelle Entscheidungsvorschläge kommt (vgl. Gutwald et al. 2021). Zwar existieren erste Ansätze zur Bias-Minimierung bei der KI-gestützten Entscheidungsfindung, diese wurden jedoch noch nicht in umfangreichen Studien getestet (vgl. Burghardt et al. 2024). Weitere Forschungsanstrengungen sollten sich auf die Möglichkeiten der KI-unterstützten Beratung fokussieren, um sicherzustellen, dass die fachliche Perspektive der Sozialen Arbeit im Vordergrund steht (vgl. Reder/Müller/Lehmann 2024).

Im Bereich des Wissensmanagements sind die Potenziale ebenfalls weitreichend. KI-basierte Assistenzsysteme könnten beispielsweise in Echtzeit Beratungsprozesse unterstützen, indem sie längere und komplexere Texte, wie sie in der Mailberatung auftreten, zusammenfassen und klassifizieren (vgl. Grandeit et al. 2020; Park et al. 2019). Zudem zeigen frühe Untersuchungen, dass KI-Systeme thematische Schwerpunkte in der Kommunikation effizient identifizieren können (vgl. Dinakar et al. 2015). Darüber hinaus zeigten internationale Studien, dass KI-Systeme aus Texten belastbare Aussagen zu Persönlichkeitsmerkmalen (vgl. Ren et al. 2021) oder psychischen Erkrankungen (vgl. Alhanai/Ghassemi/Glass 2018) der Urheber:innen ableiten können.

Für die Onlineberatung werden die Potenziale der verschiedenen KI-Assistenzsysteme z. B. im Projekt KIA erforscht. In diesem vom Bundesministerium für Familie, Seniorinnen, Frauen und Jugend geförderten Projekt werden verschiedene Anwendungen von textbasierten KI-Verfahren auf ihre Eignung zur Unterstützung in der Onlineberatung untersucht. Beispielsweise handelt es sich hier um die Zusammenfassung längerer Beratungsverläufe, die Aufbereitung von Informationen, die in den Texten enthalten sind, oder auch um methodische Hinweise zum weiteren Vorgehen für die Fachkräfte (vgl. Poltermann et al. 2024). Im Rahmen einer Befragung der Fachkräfte in diesem Projekt zeigte sich, dass trotz einer gewissen Skepsis viele Hoffnungen auf einer verantwortungsvollen Nutzung von KI im Beratungsbereich liegen. Insbesondere erhoffen sich die Fachkräfte neben der Effizienzsteigerung im Beratungsalltag Unterstützung im methodischen Bereich, sofern die vollständige Autonomie der menschlichen Akteure dadurch nicht eingeschränkt wird (vgl. Stieler/Lehmann/Berger 2025).

Für die Präsenzberatung sind die Potenziale, die in der Onlineberatung bereits sehr real werden, noch etwas ferner. Der Zugriff auf die direkte Beratungsinteraktion, der in der Onlineberatung leicht zu realisieren ist, wäre hier nur durch eine Aufzeichnung und Transkription der Gespräche möglich. Die damit einhergehenden ethischen und rechtlichen Probleme liegen auf der Hand. Daher ist zunächst eher mit der Nutzung allgemeiner Wissensmanagementfunktionen zu rechnen.

6 KI als Assistenz für Ratsuchende

Neben der Unterstützung der Fachkräfte können KI-Anwendungen auch Ratsuchenden neue Teilhabemöglichkeiten erschließen. Die technischen Möglichkeiten liegen zunächst in der Unterstützung bei verschiedenen Sinnesbehinderungen. Bereits unabhängig von KI existierten vielfältige technische Möglichkeiten, um Menschen mit verschiedenen Behinderungen z. B. die schriftliche Kommunikation oder die Audiokommunikation zu ermöglichen (vgl. Beliaev/Ginsburg 2021). Einhergehend mit der sprunghaften Verbreitung von Videokonferenzsystemen in den letzten Jahren liegen bereits KI-basierte Technologien vor, die das gesprochene Wort in Text transkribieren (vgl. Wagner et al. 2023) oder auch die Emotionen der beteiligten Personen erkennen. Zwar weist die Emotionserkennung, die ausschließlich auf Videodaten basiert, nach wie vor Schwierigkeiten auf (vgl. Naga et al. 2023), Ansätze, die Audioinformationen und textuelle Informationen berücksichtigen, sind hier jedoch deutlich weiterentwickelt (vgl. Bayerl et al. 2021). So können Beratungsinhalte, die als Text vorliegen, von LLM sprachlich vereinfacht oder zusammengefasst werden. Damit kann zunächst die Onlineberatung weitere Zugangshürden abbauen. Eine Anwendung in der Präsenzberatung ist ebenfalls denkbar, die Voraussetzung wäre allerdings, dass der Beratungsinhalt digitalisiert wird, also z. B. über eine Audioaufzeichnung, die direkt in die Verarbeitung durch das KI-System weitergeleitet wird. Technisch ist das mit jedem handelsüblichen Smartphone realisierbar, allerdings müssen die Auswirkungen auf den Beratungsprozess entsprechend berücksichtigt werden (vgl. Kieslinger 2024).

7 KI als eigenständige Beratungsinstanz

Es wurde deutlich, dass KI im Feld der Beratung, insbesondere der Onlineberatung, bereits sehr praxisnah erforscht wird und sich vielfältige Unterstützungspotenziale abzeichnen. Daher überrascht es nicht, dass eine vollständig KI-gestützte Beratung z. B. in Form von Beratungsbots als technisch möglich erscheint (vgl. Engelhardt 2023b; Linnemann/Löhe/Rottkemper 2023) und Potenziale für die Beratung insbesondere in der Formulierung der Zielsetzung, konkreter Lösungsfindung und Verabschiedung durch Beratungsbots gesehen werden (vgl. Mai/Rutschmann 2023). Aktuell befinden sich im deutschsprachigen Raum die Entwicklung solcher Beratungsbots noch in der Frühphase (vgl. Kotte/Webers 2023). Internationale Übersichtsarbeiten zeigen jedoch, dass Large Language Model-basierte Chatbots bei der Unterstützung von Menschen mit psychischen Problemen sinnvoll eingesetzt werden können, auch wenn sie aktuell noch zu viele Schwächen aufweisen, um unbesehen in der Breite eingesetzt zu werden (vgl. Casu et al. 2024). Mit dem Woebot liegt außerdem ein Chatbot vor,

der eine andere technische Basis als Large Language Model nutzt. Er wird bereits in der kognitiven Verhaltenstherapie zur Förderung von Selbstreflexion und positivem Denken eingesetzt. Die Evaluationsergebnisse dieses Bots deuten eine hohe Wirksamkeit dieses Ansatzes an (Darcy et al. 2021).

Daher ist es wahrscheinlich, dass perspektivisch KI Beratungsaufgaben übernehmen wird. Aufbauend auf den Erkenntnissen zur von Menschen durchgeführten Onlineberatung und den bisherigen Ergebnissen aus Forschungsprojekten enthält diese Entwicklung durchaus positive Aspekte. So zeigen die verschiedenen Untersuchungen, dass es möglich ist, hochwertige Ergebnisse mit KI zu generieren und dass die Fachkräfte diese Technologie nicht prinzipiell ablehnen. Allerdings wird ebenso deutlich, dass es unumgänglich ist, dass bei der Entwicklung von KI-Anwendungen eine umfangreiche ethische Abwägung der Vor- und Nachteile erfolgt und die fachliche Perspektive der Sozialen Arbeit von Anfang an eine wichtige Rolle bei der Entwicklung spielt. Nur so kann sichergestellt werden, dass bei zukünftigen Projekten wirklich das Wohl der Ratsuchenden im Vordergrund steht und sich die neu entwickelten Technologien als Unterstützung sowohl für die Fachkräfte als auch für die Ratsuchenden darstellen.

Literatur

- Albrecht, Jens / Lehmann, Robert / Poltermann, Aleksandra (2024): GeCCo 1.0 – Erstellung eines öffentlichen Datensatzes für die KI-basierte Inhaltsanalyse in der Online-Beratung. In: Schriftenreihe der Technischen Hochschule Nürnberg Georg Simon Ohm, S. 3–15. <https://doi.org/10.34646/THN/OHMDOK-1389>
- Albrecht, Jens / Rudolph, Eric (2023): Künstliche Intelligenz und Machine Learning – Grundwissen für Sozialarbeiter/-innen. In: *Jugendhilfe* 61(5), S. 367–376.
- Albrecht, Jens / Rudolph, Eric / Poltermann, Aleksandra / Lehmann, Robert (2024): The Virtual Client: Leveraging Generative AI For Innovative Online-Counselor Education. 17th annual International Conference of Education, Research and Innovation, Sevilla.
- Alhanai, Tuka / Ghassemi, Mohammad / Glass, James (2018): Detecting depression with audio/text sequence modeling of interviews. In: *Interspeech*, S. 1716–1720. https://sls.csail.mit.edu/publications/2018/Alhanai_Interspeech-2018.pdf (Abfrage: 15.06.2025).
- Arnold, Patricia (2012): Wirksamkeit von Online-Beratung – Was sagt die Forschung? In: Schindler, Wolfgang / Spangler, Gerhard (Hrsg.): *Kollegiale Beratung: Online und offline im Heilsbronner Modell*. Göttingen: Vandenhoeck & Ruprecht, S. 76–89.
- Bayerl, Sebastian P. / Tammewar, Aniruddha / Riedhammer, Korbinian / Riccardi, Giuseppe (2021): Detecting Emotion Carriers by Combining Acoustic and Lexical Representations (arXiv:2112.06603). arXiv. <https://doi.org/10.48550/arXiv.2112.06603>
- Beliaev, Stanislav / Ginsburg, Boris (2021): TalkNet 2: Non-Autoregressive Depth-Wise Separable Convolutional Model for Speech Synthesis with Explicit Pitch and Duration Prediction. (arXiv:2104.08189). arXiv. <https://doi.org/10.48550/arXiv.2104.08189>
- Bradl, Marion / Stieler, Mara / Engels, Sylvia (2022): Virtuelle Beratungsstrukturen: Wissenschaftliche Begleitung der Jugendmigrationsdienste (JMD) im Rahmen des Projekts: JMD digital-virtuelle Beratungsstrukturen für ländliche Räume. https://opus4.kobv.de/opus4-ohm/files/917/EBI_Virtuelle_Beratungsstrukturen_JMD_2022.pdf (Abfrage: 15.06.2025).

- Brunner, Alexander/Engelhardt, Emily/Heider, Triz (2009): Foren-Beratung. In: Kühne, Stefan/Hintenberger, Gerhard (Hrsg.): *Handbuch Online-Beratung*. Göttingen: Vandenhoeck & Ruprecht. S. 79–90.
- Burghardt, Jennifer/Lehmann, Robert/Reder, Michael/Koska, Christopher/Kraus, Maximilian/Müller, Nicholas H. (2024): Kann Künstliche Intelligenz sozialarbeiterische Entscheidungsprozesse unterstützen? Ethik und digitale Operationalisierung im Feld der Kindeswohlgefährdung. In: *unsere jugend* 76(7+8), S. 300–310.
- Casu, Mirko/Triscari, Sergio/Battiato, Sebastiano/Guarnera, Luca/Caponnetto, Pasquale (2024). AI Chatbots for Mental Health: A Scoping Review of Effectiveness, Feasibility, and Applications. In: *Applied Sciences* 14(13), Article 13.
- Darcy, Alison/Daniels, Jade/Salinger, Davis/Wicks, Paul/Robinson, Athena (2021): Evidence of Human-Level Bonds Established With a Digital Conversational Agent: Cross-sectional, Retrospective Observational Study. In: *JMIR Formative Research* 5(5).
- Dinakar, Karthik/Chen, Jackie/Lieberman, Henry/Picard, Rosalind/Filbin, Robert (2015): Mixed-Initiative Real-Time Topic Modeling & Visualization for Crisis Counseling. In: *Proceedings of the 20th International Conference on Intelligent User Interfaces*, S. 417–426.
- Eckl, Markus/Prigge, Jessica/Schildknecht, Lukas/Ghanem, Christian (2020): Zehn Jahre Soziale Passagen: Eine empirische Analyse ihrer Themen. In: *Soziale Passagen* 12, S. 57–80.
- Eichenberg, Christiane/Kühne, Stefan (2014): *Einführung Onlineberatung und -therapie: Grundlagen, Interventionen und Effekte der Internetnutzung*. 1. Auflage. München: Reinhardt (UTB).
- Eichenberg, Christiane/Küsel, Cornelia (2016): Zur Wirksamkeit von Online-Beratung und Online-Psychotherapie. In: *Resonanzen – E-Journal für biopsychosoziale Dialoge in Psychosomatischer Medizin, Psychotherapie, Supervision und Beratung* 4(2), S. 93–107.
- Engelhardt, Emily (2023a): KI in der Lehre – Beraten lernen mit ChatGPT. In: *fnma Magazin* 3, S. 27–28.
- Engelhardt, Emily (2023b): Berät bald der Bot!? Zur Bedeutung von KI-Textgeneratoren in der (Online-)Beratung. In: *Jugendhilfe* 61(5), S. 404–409.
- Engelhardt, Emily (2021): *Lehrbuch Onlineberatung*. 2., erweiterte Auflage. Göttingen: Vandenhoeck & Ruprecht.
- Ghanem, Christian/Eckl, Markus/Lehmann, Robert/Widerhold, Jean-Pierre (2021): „Irgendwie fühle ich mich als Angehörige alleine gelassen.“ Eine automatisierte Analyse eines Onlineforums für Angehörige von Inhaftierten. In: Wunder, Maik (Hrsg.): *Digitalisierung und Soziale Arbeit – Transformationen und Herausforderungen*. Bad Heilbrunn: Klinkhardt, S. 240–255.
- Grandeit, Philipp/Haberkern, Carolyn/Lang, Maximiliane/Albrecht, Jens/Lehmann, Robert (2020): Using BERT for Qualitative Content Analysis in Psychosocial Online Counseling. 4TH WORKSHOP ON NLP AND CSS at the EMNLP 2020 The 2020 Conference on Empirical Methods in Natural Language Processing. <https://www.aclweb.org/anthology/2020.nlpccs-1.2/> (Abfrage: 15.06.2025).
- Gutwald, Rebecca/Burghardt, Jennifer/Kraus, Maximilian/Reder, Michael/Lehmann, Robert/Müller, Nicholas (2021): Soziale Konflikte und Digitalisierung. Chancen und Risiken digitaler Technologien bei der Einschätzung von Kindeswohlgefährdungen. In: *EthikJournal* 7(2), S. 1–20.
- Hagemai, André/Stuiber, Adrian (2020): *Online-Streetwork*. Ein erweiterter Ansatz der aufsuchenden Jugendarbeit & Radikalisierungsprävention. Berlin: Streetwork@online.
- Hörmann, Martina/Engelhardt, Emily (2022): Blended Counseling – Grundlagen, Aktuelles und Diskurslinien. In: *Zeitschrift für systemische Therapie und Beratung* 2, S. 72–77.
- Kieslinger, Daniel (2024): Mit KI-Unterstützung zu einer inklusiven Kinder- und Jugendhilfe. In: *unsere jugend* 76(7+8), S. 338–339.
- Killian, Stefan/Baumann, Clemens/Delorette, Michael/Freisleben-Teutscher, Christian/Größbacher, Stefanie/Huber, Alois/Husinsky, Matthias/Judmair, Peter/Moser, Thomas/Pflegler, Johannes/Schlager, Alexander/Schöffner, Lucas/Taurer, Florian/Vogt, Georg (2023): *MIRACLE – Mixed Reality und Cooperation im Lehreinnsatz: Erfahrungen, Potenziale, Limitationen*. In: Wei-

- ßenböck, Josef (Hrsg.): Lernen über den Tellerrand hinaus. Good Practices zu Interdisziplinarität, Internationalisierung und Future Skills, S. 119–126. Wien: Lemberger Publishing.
- Kotte, Silja/Webers, Thomas (2023): Digitalisierung in der Beratung: Online- und KI-unterstützte Beratungsformate. *Organisationsberatung, Supervision, Coaching* 30(1), S. 1–6.
- Lehmann, Robert (2020): Die Professionalisierung der Onlineberatung. In: *FORUM Sexualaufklärung und Familienplanung* 2, S. 3–5.
- Lehmann, Robert/Stieler, Mara/Zauter, Sigrid (2024): Begleitforschung zu Streetwork im Netz – Modellprojekt zur Qualitätssicherung und möglichem Transfer der webbasierten aufsuchenden Sozialarbeit (BeSiN) [Abschlussbericht an das Bundesministerium für Gesundheit]. <https://www.bundesgesundheitsministerium.de/service/publikationen/details/bsin> (Abfrage: 15.06.2025).
- Linnemann, Gesa A./Löhe, Julian/Rottkemper, Beate (2023): Bedeutung von Künstlicher Intelligenz in der Sozialen Arbeit: Eine exemplarische arbeitsfeldübergreifende Betrachtung des Natural Language Processing (NLP). In: *Soziale Passagen* 15(1), S. 197–211.
- Lippert, Carolyn/Rudolph, Eric/Poltermann, Aleksandra/Engert, Natalie/Lehmann, Robert/Albrecht, Jens (2024): Generative KI in der beraterischen Ausbildung. Der Einsatz eines*r virtuellen Klient*in als Übungstool für angehende Onlineberater*innen. In *e-beratungsjournal.net – Zeitschrift für Onlineberatung und computervermittelte Kommunikation* 20(1), Artikel 3, S. 41–60. <https://doi.org/10.48341/tcgc-st69>
- Mai, Vanessa/Rutschmann, Rebecca (2023): Best Practices im Chatbot Coaching. Einblicke in Forschung und Entwicklung des StudiCoachBots der TH Köln und in die Coaching Chatbot Plattform evooch. In: *Organisationsberatung, Supervision, Coaching* 30(1), S. 111–125.
- Naga, Prameela/Marri, Swamy Das/Borreo, Raiza (2023): Facial emotion recognition methods, datasets and technologies: A literature survey. In: *Materials Today: Proceedings* 80, S. 2824–2828.
- Park, Sungjoon/Kim, Donghyun/Oh, Alice (2019): Conversation Model Fine-Tuning for Classifying Client Utterances in Counseling Dialogues. In: *Proceedings of the 2019 Conference of the North*, S. 1448–1459.
- Poltermann, Aleksandra/Rudolph, Eric/Steigerwald, Philipp/Lehmann, Robert (2024): KI und Soziale Arbeit – Was ist heute möglich? In: *Sozialwirtschaft* 34(1), S. 21–23.
- Reder, Michael/Müller, Nicholas/Lehmann, Robert (2024): Über das Verhältnis von Ethik und Algorithmen. In: Reder, Michael/Koska, Christopher (Hrsg.): *Künstliche Intelligenz und ethische Verantwortung*. Bielefeld: transcript, S. 7–23.
- Reindl, Richard (2018): Zum Stand der Onlineberatung in Zeiten der Digitalisierung. In: *e-beratungsjournal.net – Fachzeitschrift für Onlineberatung und computervermittelte Kommunikation* 14(1), Artikel 2, S. 16–26.
- Ren, Zhancheng/Shen, Qiang/Diao, Xiaolei/Xu, Hao (2021): A sentiment-aware deep learning approach for personality detection from text. In: *Information Processing & Management* 58(3). <https://doi.org/10.1016/j.ipm.2021.102532>
- Stieler, Mara/Berger, Johanna/Lehmann, Robert (2025): Technologische Innovation in der psychosozialen Onlineberatung: Eine praxisorientierte Perspektive auf die Anwendung von Künstlicher Intelligenz. In: Späte, Julius/Endter, Cordula/Stix, Daniela/Krauskopf, Karsten (Hrsg.): *#GesellschaftBilden im Digitalzeitalter. Perspektiven Sozialer Arbeit auf technologische Herausforderungen*. Münster: Waxmann, S. 179–206.
- Stieler, Mara/Lipot, Sarah/Lehmann, Robert (2022): Zum Stand der Onlineberatung in Zeiten der Corona-Krise. Entwicklungs- und Veränderungsprozesse der Onlineberatungslandschaft. In: *e-beratungsjournal.net – Zeitschrift für Onlineberatung und computervermittelte Kommunikation* 18(1), Artikel 4, S. 50–65. <https://doi.org/10.48341/262p-7t64>
- Wagner, Dominik/Baumann, Ilja/Bayerl, Sebastian P., Riedhammer, Korbinian/Bocklet, Tobias (2023): Speaker Adaptation for End-To-End Speech Recognition Systems in Noisy Environments (arXiv:2211.08774). <https://doi.org/10.48550/arXiv.2211.08774>

- Weinhardt, Marc/Bauer, Petra/Lohner, Eva Maria/Schmitz/Anne-Kathrin/Christiani, Laura/Eder-Curreli, Carmilla (2022): Beratungslernen im Studium. Ergebnisse einer Pilotstudie zur Umsetzung eines videogestützten Beratungslabors im Horizont pandemiebedingter Digitalität. In: e-beratungsjournal.net – Zeitschrift für Onlineberatung und computervermittelte Kommunikation 18(2), Artikel 3, S. 38–55. <https://doi.org/10.48341/kttm-4822>
- Zauter, Sigrid/Lehmann, Robert (2021): Schwer auffindbare Beratungsangebote in der psychosozialen Onlineberatung. In: Freier, Carolin/König, Joachim/Manzeschke, Arne/Städler-Mach, Barbara (Hrsg.). Gegenwart und Zukunft sozialer Dienstleistungsarbeit: Chancen und Risiken der Digitalisierung in der Sozialwirtschaft. Wiesbaden: Springer Fachmedien, S. 129–140.
- Zizelmann, Regine (2023): Neue Formate und Onlineberatung. In: Hessische Blätter für Volksbildung 73(4), S. 91–102.

Künstliche Intelligenz und Inklusion¹

Olivier Steiner

Abstract: Der Beitrag befasst sich mit den Potenzialen, Herausforderungen und Risiken, die sich aus der Entwicklung und Verbreitung von Künstlicher Intelligenz für die Inklusion marginalisierter und benachteiligter Menschen ergeben. Im ersten Teil des Beitrags wird ein kritisches Verständnis von Inklusion und Exklusion als notwendige Grundlage für eine differenzierte Bewertung von KI-Technologien entwickelt. Nach einer kurzen Einführung in die Künstliche Intelligenz als vielversprechende, aber auch problematische Technologie werden im zweiten Teil des Beitrags Potenziale, Herausforderungen und Risiken Künstlicher Intelligenz für die Inklusion marginalisierter und benachteiligter Menschen diskutiert. Dabei werden technologische, wohlfahrtsstaatliche, individualisierte und gemeinwohlorientierte Ebenen der Analyse der Auswirkungen von KI-Technologien auf Inklusion unterschieden. Im Diskussionsteil werden Empfehlungen für eine ethisch, sozial und praxisbezogen sensible Entwicklung und Implementierung inklusiver KI-Technologien formuliert.

Keywords: Künstliche Intelligenz, Inklusion, Exklusion, Teilhabe, (post)humane Intelligenz

1 Einführung

Der Beitrag befasst sich mit den Potenzialen, Herausforderungen und Risiken, die sich aus der Entwicklung und Verbreitung von Künstlicher Intelligenz (KI) für die Inklusion marginalisierter und behinderter Menschen² ergeben. Im ersten Teil des Beitrags wird ein kritisches Verständnis von Inklusion und Exklusion zur Grundlegung einer differenzierten Bewertung von KI-Technologien entwickelt. Nach einer kurzen Einführung in die KI als vielversprechende, aber auch problematische Technologie werden im zweiten Teil des Beitrags Potenziale, Herausfor-

1 © Die Autor:innen 2025, veröffentlicht im Verlag Beltz Juventa, Open Access: CC BY 4.0
Gesa Linnemann/Julian Löhle/Beate Rottkemper (Hg.), *Künstliche Intelligenz in der Sozialen Arbeit*
10.3262/978-3-7799-8562-4_010

2 Behinderung entsteht gemäss der WHO (2001) im Zusammenwirken individueller Einschränkungen und gesellschaftlicher Barrieren. Der Begriff „behinderter Mensch“ berücksichtigt hier also immer auch die Formen der gesellschaftlich verursachten „Behinderung“ von Menschen.

derungen und Risiken von KI für die Inklusion marginalisierter und von Behinderung betroffener Menschen diskutiert. Dabei werden technologische, wohlfahrtsstaatliche, individualisierte und gemeinwohlorientierte Ebenen der Analyse der Auswirkungen von KI-Technologien auf Inklusion unterschieden. Im Diskussteil werden Empfehlungen für eine ethische, soziale und praxisbezogen sensible Entwicklung und Implementation inklusiver KI-Technologien formuliert.

2 Zur Grundlegung: eine kritische Fassung von Inklusion und Exklusion

Die Inklusion und Exklusion von Menschen in und aus der Gesellschaft ist seit der Entwicklung der bürgerlichen Öffentlichkeit im 19. Jahrhundert ein viel diskutiertes Thema (vgl. Habermas 1996; Imbusch/Heitmeyer 2008). Angesichts der Verelendung breiter Bevölkerungsschichten im Zuge der industriellen Revolution gewann die „soziale Frage“ im bürgerlichen Bewusstsein zunehmend an Bedeutung (Landhäußer 2009). Das teilweise normwidrige Verhalten der Arbeiterschaft, die Folgen des Alkoholkonsums und die durch Armut zerrütteten Familienverhältnisse weckten Bestrebungen, diesen Entwicklungen entweder durch verschärfte Strafmaßnahmen oder durch fromme Hilfe entgegenzuwirken. Integrationsdiskurse und -bemühungen in einer kapitalistischen Gesellschaftsordnung sind daher häufig stark normativ und disziplinierend geprägt (Bendle 1996; Krämer 2008). Mit der Einführung des Begriffspaars Inklusion und Exklusion in den soziologischen Diskurs durch Niklas Luhmann sind gegenüber quantifizierenden, feingliedrigen Konzepten sozialer Ungleichheit die normativen und disziplinierenden Aspekte des sozialen Ein- und Ausschlusshandelns hervorgehoben worden (vgl. Stichweh 1997): Es sind die Funktionssysteme, beispielsweise das Wirtschaftssystem, die Subjekte adressieren oder nicht adressieren und damit ein- oder ausschließen. Die Subjekte erscheinen somit als gesellschaftliche erst, wenn sie überhaupt von Funktionssystemen adressiert werden – andernfalls bleiben sie unsichtbar, existieren für die Funktionssysteme nicht – sie sind exkludiert (Seifert 2013). Damit ist dem Inklusionsbegriff eine Machtdimension eingeschrieben – es bedarf immer eines sozialen Systems, das inkludiert; und damit sind immer Entscheidungen verbunden, wer aufgrund welcher zugeschriebenen Merkmale adressiert, d. h., inkludiert wird oder nicht.

Diese Konzeption von Inklusion unterscheidet sich erheblich von in den letzten Jahren vielfach skizzierten Ansätzen, die wir als idealistische Interpretationen von Inklusion charakterisieren können. Inklusion wird hier gewissermaßen voraussetzungslos als Idealzustand einer gesellschaftlichen Verfassung angesehen, in der alle Menschen umfassend am Sozialen teilhaben (Felder 2018; Wilson 1999). Konzeptionen von Inklusion als normative Setzung etwa eines bürger-

lichen Selbstverständnisses oder als disziplinierendes Verfahren zur Integration in (post)industrielle Wirtschaftsorganisationen, werden dabei kaum berücksichtigt (vgl. Rojas 2024). Im idealistischen Verständnis von Inklusion ist die Inklusion von Menschen in Systeme immer positiv – unabhängig davon, in welche Systeme inkludiert werden soll und welche normativen oder disziplinierenden Setzungen damit verbunden sind. Joshi (2014) kommt vor dem Hintergrund unterschiedlicher Definitionen von Inklusion, die Inklusion wahlweise als Strategie, Politik, Wert, Norm, Prinzip, Lebensstil oder Handlung verstehen, zu dem Schluss, dass der Begriff ebenso wie sein Pendant Exklusion wissenschaftlich auffallend unbestimmt ist und unterschiedlich verwendet wird.

Diese hier einleitend kritischen Vorbemerkungen sollen dazu anregen, im weiteren Verlauf der Diskussion um KI und Inklusion auch problematische Setzungen und Entwicklungen in Betracht zu ziehen – ganz im Sinne einer kritischen Theorie, die nicht jeden Einsatz von Technologie für ein oftmals kaum expliziertes Ziel als erstrebenswert ansieht (Park 2023). Damit soll an dieser Stelle natürlich nicht behauptet werden, dass Inklusion ein grundsätzlich problematisches Konzept sei. Es ist anzuerkennen, dass mit der Intensivierung des Inklusionsdiskurses insbesondere der systematische Ausschluss von Menschen mit Behinderungen aus gesellschaftlichen Funktionssystemen wie Wirtschaft, Bildung und Öffentlichkeit deutlich an Aufmerksamkeit gewonnen hat und in westlichen Gesellschaften u. a. mit der Verabschiedung der UN-Behindertenrechtskonvention entscheidende Entwicklungen angestoßen wurden, um die Teilhabe von Menschen mit Behinderungen an gesellschaftlicher Produktion und Kommunikation zu verbessern (Wesselmann 2022). Insofern ist gegenüber idealistischen und wenig kritischen Inklusionskonzepten mit Rojas (2024, S. 162) ein radikaleres Verständnis von Inklusion zu formulieren als „concept of resistance against injustice, inequity, and social inequality, from an intercultural perspective: respect, tolerance, and positive acceptance of others' differences“. Die Notion von Widerständigkeit im Konzept von Inklusion sensibilisiert für Aspekte der Sozialdisziplinierung, die mitunter auch über Technologien wie KI in als inklusiv verstandene Strategien und Handlungen einfließen können. Eine solche Sensibilisierung ist gerade vor dem Hintergrund der zunehmenden Implementierung managerialistischer Verfahren in der Sozialen Arbeit notwendig, die eine Qualitätssteigerung und Kostensenkung sozialer Dienstleistungen versprechen, letztlich aber „die Herstellung von Nutzbarmachung von Individuen für den Arbeitsmarkt als oberstes Postulat und [...] auf diese Weise neue Selektions- bzw. Ausgrenzungsprozesse“ hervorrufen (Seifert 2013). Es bleibt also zu klären, welche Funktionen KI für die Durchsetzung normalisierender und disziplinierender Ansprüche in Inklusionsparadigmen einnimmt – und potenziell einnehmen wird.

3 Potenziale, Herausforderungen und Risiken von KI für Inklusion

KI ist ein multidisziplinäres Forschungs- und Entwicklungsfeld, das sich mit der Entwicklung von Systemen beschäftigt, die menschenähnliche Intelligenzleistungen erbringen können (sollen) (Lenzen 2018; 2020). KI umfasst heute eine Vielzahl von Techniken und Methoden, darunter Maschinelles Lernen, neuronale Netze und natürliche Sprachverarbeitung (siehe für eine differenzierte Diskussion den Beitrag von Beate Rottkemper in diesem Band). Allerdings gibt es bis heute keine einheitliche Definition von KI (Roth et al. 2024). Nach Alan Turing (1950) ist ein System dann als intelligent anzusehen, wenn es menschliches Verhalten so erfolgreich nachahmen kann, dass ein:e menschliche:r Beobachter:in den Unterschied nicht feststellen kann.

An KI werden derzeit hohe Erwartungen geknüpft, drängende Probleme unserer Zeit wie u. a. ungleiche Bildungschancen, die Bekämpfung von Krankheiten oder die Klimakatastrophe zu lösen. Zuweilen ist ein regelrechter KI-Hype zu beobachten (Markelius et al. 2024). Gegenüber den in der Öffentlichkeit oft hochgesteckten Erwartungen an die Leistungen der KI für Gesellschaft und Umwelt wird ihre Entwicklung zunehmend auch kritisch kommentiert. Im Vordergrund stehen dabei u. a. Bedenken hinsichtlich des Datenschutzes, des geistigen Eigentums, der Substitution menschlicher Arbeitskraft durch KI, ethische Bedenken bezüglich des Einsatzes u. a. in sozial sensiblen Bereichen, der Ausbeutung und psychischen Belastung billiger Arbeitskräfte in Niedriglohnländern für die Aufbereitung von Daten für das Training von KI, der automatisierten Ausgrenzung benachteiligter Bevölkerungsgruppen durch KI sowie ökologische Bedenken hinsichtlich des exorbitanten Energieverbrauchs für das Training und den Betrieb großer KI Modelle (Eubanks 2018; Leech et al. 2024; Markelius et al. 2024; Shams/Zowghi/Bano 2023).

Bezüglich Potenzialen, Herausforderungen und Risiken von KI für die Inklusion sind mindestens vier Ebenen zu unterscheiden (vgl. ähnlich Fosch-Villaronga/Poulsen 2022): a) KI-Technologien als Repräsentationen marginalisierender Diskurse und gesellschaftlicher Exklusion, b) KI-Technologien zur staatlichen Verwaltung benachteiligter Bevölkerungsgruppen oder Menschen mit Behinderungen, c) KI-Technologien zur Unterstützung von Subjekten, um gesellschaftliche Zugänge oder Teilhabe zu ermöglichen, und d) KI-Technologien für das Empowerment und die gesellschaftliche Inklusion lokaler Gemeinschaften.

Im Folgenden werden für die einzelnen Ebenen Potenziale, Herausforderungen, Risiken und diskutierte Lösungsansätze beschrieben.

a) KI-Technologien als Repräsentationen marginalisierender Diskurse und gesellschaftlicher Exklusion

Eine grundsätzliche Herausforderung KI-basierter Technologien in Hinblick auf Inklusion besteht im Training von Modellen mit Datensätzen, die durch

Selektion und Gewichtung gesellschaftlich-normative Kernbestände soziomaterieller Phänomene, sozialer Handlungen und Werte überrepräsentieren und hierdurch marginalisierte Diskurse oder soziale Gruppen unterrepräsentieren. So sind nach Fosch-Villaronga und Poulsen (2022) u. a. beispielsweise Frauen, LGBTQIA+-Communitys, ältere Menschen und Menschen mit Behinderungen in oder durch KI geprägte(n) Technologien systematisch unterrepräsentiert. Solche Verzerrungen („Bias“) entstehen bereits auf Ebene der Auswahl der Daten, des Designs der Modelle (u. a. auch durch die soziale Zusammensetzung der Forschenden³ und Entwickelnden begründet), des Modelltrainings sowie der Implementierung und Evaluation der Technologien. Eine Metastudie zur Frage der Berücksichtigung von Inklusion und Diversität in KI-Technologien (Shams/Zowghi/Bano 2023) zeigt auf, dass die Problematik der Unterrepräsentation marginalisierter Gruppen nicht in der Technologie sui generis begründet ist, sondern durch die Menschen verursacht wird, die die Technologie gestalten. So gibt es derzeit zwar vergleichsweise umfangreiches Wissen über die Problematik der mangelnden Inklusivität und Diversität von KI-Technologien, aber kaum konkrete Lösungsansätze, wie diese Technologien verbessert werden können, um Inklusivität und Diversität zu fördern (ebd.). Als Folge der Monopolisierung und Kommerzialisierung von KI-Technologien durch multinationale Unternehmen ist eher zu befürchten, dass die Unterrepräsentation marginalisierter Gruppen bestehen bleibt (vgl. McQuillan et al. 2024). Dies bedeutet allerdings nicht, dass für spezifische Anwendungsfälle, z. B. die Erkennung von Kindeswohlgefährdungen, die Sprachausgabe für Menschen mit Sehbehinderungen oder das Empowerment sozialer Bewegungen, keine hohen Inklusionspotenziale durch KI-Technologien bestehen.

b) KI-Technologien zur staatlichen Verwaltung benachteiligter Bevölkerungsgruppen oder Menschen mit Behinderung

Datenbasierte, algorithmische Vorhersagen zu Gefährdungen oder riskantem Verhalten marginalisierter Bevölkerungsgruppen werden in westlichen Wohlfahrtssystemen seit den 1980er-Jahren eingesetzt. In jüngerer Zeit sind weltweit verstärkte Bestrebungen zu beobachten, automatisierte Systeme für das Management exkludierter oder von Exklusion bedrohter Menschen einzusetzen (Bastian/Schrödter 2015; Gillingham 2016, 2019; Seelmeyer 2020). Insbesondere in den USA werden solche Systeme bereits eingesetzt. Virginia Eubanks (2018) hat die Auswirkungen automatisierter Technologien auf vulnerable und benachteiligte Bevölkerungsgruppen untersucht und beschreibt eindringlich, wie diese Technologien bestehende Ungleichheiten häufig verschärfen, anstatt

3 Beispielsweise sind nur 14% der Editor:innen des Journals „Artificial Intelligence“ Frauen, sehr wenige stammen aus sogenannten Schwellen- oder Entwicklungsländern (Fosch-Villaronga/Poulsen 2022: 115).

sie abzubauen. Laut Eubanks führen diese Systeme zu einer intensiveren Überwachung und Stigmatisierung benachteiligter Bevölkerungsgruppen. Während dies für einige der von Eubanks beschriebenen Systeme zutrifft und diese unter anderem aufgrund öffentlichen Widerstands bereits wieder deaktiviert wurden, erscheint es dennoch wichtig, im Einzelfall auch mögliche Potenziale solcher Technologien für wohlfahrtsstaatliche Systeme unvoreingenommen zu prüfen. So stellt beispielsweise das Allegheny Family Screening Tool (AFST), das u. a. auf KI-Technologien basierende Gefährdungserkennungen von Kindern in Familien unterstützt, ein unabhängig evaluiertes Verfahren dar, das bei näherer Betrachtung durchaus Potenziale zur Reduktion des „Practitioner Bias“, also von Fehleinschätzungen durch Fachkräfte, aufweist (Goldhaber-Fiebert/Prince 2019). Ein unerwartetes Ergebnis der Evaluation des Systems ist, dass Benachteiligungen aufgrund der ethnischen Zugehörigkeit der Klient:innen in den Abläufen der Organisation dadurch offenbar reduziert werden konnten. Im Fall des AFST werden KI-Technologien zur Entscheidungsunterstützung eingesetzt und nicht als Ersatz für das professionelle Urteil von Fachkräften (vgl. Seelmeier 2020). Es stellt sich damit die grundsätzliche Frage: Wenn durch den Einsatz von KI-Technologien Kinder – und insbesondere Kinder, die z. B. aufgrund ihrer ethnischen Herkunft ohnehin schon marginalisiert sind – vor familiärer Gewalt geschützt werden können und damit – auch durch wohlfahrtsstaatliche Institutionen (mit)verursachte – soziale Exklusion verhindert werden kann, wäre es dann nicht ethisch fragwürdig, einen fachlich reflektierten und kontinuierlich evaluierten Einsatz von KI-Technologien in Wohlfahrtssystemen kategorisch auszuschließen?

c) KI-Technologien zur Unterstützung von Subjekten, um gesellschaftliche Zugänge oder Teilhabe zu ermöglichen

Vielversprechend erscheinen KI-Technologien beispielsweise für Menschen mit Behinderungen, marginalisierte Menschen oder Menschen in der Rehabilitation, um gesellschaftliche Teilhabe zu ermöglichen oder zu verbessern. So besteht mittlerweile eine Vielzahl von KI-basierten Anwendungen zur Unterstützung von Menschen mit Seh- oder Sprachbehinderungen (vgl. Burghardt/Kieslinger 2025; Deka et al. 2022; Griffen et al. 2024; Gupta et al. 2024; Montanha et al. 2022; Wang et al. 2023), Menschen mit psychischen Beeinträchtigungen (vgl. Camp et al. 2022; Lejeune et al. 2022; Lenton-Brym et al. 2024; Prescott/Barnes 2024; Zhou et al. 2022), zur Unterstützung marginalisierter Menschen wie isolierter alter Menschen (vgl. Marziali et al. 2024; Salomé/Monfort 2023), devianter Jugendlicher (vgl. Frey et al. 2020) sowie zur Unterstützung von Menschen in Rehabilitation oder medizinischer Langzeitpflege (vgl. Jiang et al. 2024; Lukkien et al. 2023; Medenica et al. 2024). Ohne an dieser Stelle einen Anspruch auf Vollständigkeit der mittlerweile vielfältigen Forschungs- und Entwicklungsarbeiten in unterschiedlichen Disziplinen erheben zu können, verdeutlicht ein

erster Überblick, dass KI-Technologien grundsätzlich ein großes Potenzial in verschiedenen Anwendungsfeldern. u. a. in der psychischen Gesundheitsversorgung, assistiven Technologien für Menschen mit sensorischen, kognitiven oder körperlichen Behinderungen sowie in der neuropsychologischen Rehabilitation, haben. KI-Technologien können die menschliche Unterstützung ergänzend verbessern sowie die Zugänglichkeit zu Angeboten und Dienstleistungen erhöhen. Gleichzeitig bestehen wesentliche Herausforderungen für die Realisierung dieser Potenziale, die in der Diskussion dieses Beitrags zusammenfassend aufgegriffen werden. Bislang wenig diskutiert sind allerdings mögliche Problemfelder wie die durch individualisierte KI-Technologien hervorgerufenen Verzerrungen der Wahrnehmungs-, Erfahrungs- und Deutungsleistungen der unterstützten Subjekte aufgrund der Konstruktion und des Trainings der Modelle. Schließlich ist im Rückgriff auf die eingangs formulierte inklusionskritische Perspektive immer auch zu hinterfragen, in welche sozialen Systeme mittels KI-Technologien inkludiert werden soll und in welchen Machtkonstellationen Soziale Arbeit auf Subjekte zielende KI-Technologien einsetzt.

d) KI-Technologien für das Empowerment und die gesellschaftliche Inklusion lokaler Gemeinschaften

Ein bislang wenig diskutiertes Anwendungspotenzial von KI-Technologien für Inklusion stellt die Nutzung durch NGOs und Graswurzelbewegungen dar. Damit verbunden ist der Anspruch, die Nutzung von KI-Technologien zu demokratisieren und KI-gestützte Inklusionspotenziale auch für benachteiligte gesellschaftliche Gruppen zu eröffnen. KI-Technologien sollen danach das Empowerment lokaler Gemeinschaften fördern, indem diese in die Lage versetzt werden, KI-gestützte Lösungen für ihre eigenen Herausforderungen zu entwickeln. Ziel der meist unter dem Schlagwort „Grassroots AI“ diskutierten Ansätze ist es, marginalisierte Gruppen und unterrepräsentierte Gemeinschaften in die Technologieentwicklung einzubeziehen und damit soziale Ungleichheiten abzubauen (vgl. beispielsweise <https://chej.org/how-ai-can-help-strengthen-grassroots-organizing>; Abfrage: 14.04.2025).

Empirische Arbeiten zu den tatsächlichen Nutzungsweisen und Konsequenzen des Einsatzes von KI für solche Organisationen und Bewegungen sind jedoch spärlich. Eine erste Studie legen Ibison et al. (2024) zu der Frage vor, welche Potenziale und Herausforderungen NGOs und Graswurzelbewegungen in KI-Technologien für ihre Arbeit sehen. Die Ergebnisse zeigen, dass insbesondere große NGOs und Bewegungen bereits KI-Technologien nutzen, kleineren Organisationen oder Bewegungen jedoch die Ressourcen und Kompetenzen fehlen, um diese zielgerichtet einzusetzen. Die Anwendungsszenarien liegen vor allem in den Bereichen Marketing, Kommunikation sowie dem Abbau von Barrieren (z. B. Alternativtexte für Bilder). Die befragten Mitarbeiter:innen stehen dem Einsatz von KI in ihrer Arbeit durchaus ambivalent gegenüber: 70 % der Befragten sorgen

sich um den Datenschutz der Adressat:innen, 63% sind unsicher bezüglich der Genauigkeit der durch KI erstellten Inhalte. Dementsprechend folgern die Autor:innen, dass KI nicht das Wundermittel für die Herausforderungen darstellt, vor denen NGOs und Graswurzelbewegungen heute stehen. Es muss daher auch festgehalten werden, dass die hohen Ideale des Empowerments lokaler Gruppen durch KI häufig durch den überwiegenden Einsatz von KI zur Rationalisierung der Inhaltsproduktion konterkariert werden. Dennoch verweisen erste Projekte wie beispielsweise die KI-gestützte Interpretation von Luftverschmutzungsdaten, die von betroffenen Gemeinschaften gesammelt werden („crowd data“), auf das Potenzial von KI, technologisch unterstütztes Empowerment gerade in marginalisierten Gemeinschaften zu etablieren (Hsu et al. 2022). Dazu bedarf es weiterer, insbesondere partizipativ angelegter Forschungs- und Entwicklungsprojekte, die die Erfolgsbedingungen solcher Projekte beschreiben und schließlich auch KI-Technologien entwickeln und öffentlich verfügbar machen, um soziale Bewegungen und NGOs weltweit weiter zu ermächtigen.

4 Diskussion

Fragen nach den Auswirkungen von KI auf die Inklusion benachteiligter und von Behinderung betroffener Menschen in die Gesellschaft lassen sich nicht mit einfachen Antworten aufklären. KI steht noch am Anfang ihrer technologischen und gesellschaftlichen Ausgestaltung, wesentliche Fragen zum Verhältnis von Menschen und KI, zu den Auswirkungen auf soziale Systeme und die Umwelt werden in ihrer Tragweite erst erkannt und diskutiert. Inklusion als ein gerade in der Sozialen Arbeit häufig hochgradig normativ verwendeter Begriff kann mit ebenso normativ und machtvoll geprägten KI-Technologien potenziell eine unheilvolle Allianz eingehen. Problematisch wäre der Einsatz von KI-Technologien unter dem Duktus der Inklusion, wenn diese als rationalisierende und effizienzsteigernde Maßnahmen eingesetzt werden, um die Spannungen, die durch bestehende und sich verschärfende Inklusions- und Exklusionsverhältnisse in der Gesellschaft entstehen, zu befrieden. In einem radikalen Verständnis (vgl. Rojas 2024), das Inklusion auch als Widerstand gegen soziale Ungerechtigkeit versteht, ist in der Sozialen Arbeit daher stets zu prüfen, ob ein als inklusiv verstandener Einsatz von KI tatsächlich advokatorisch die Bedürfnisse der Adressat:innen aufgreift und diese ermächtigt, oder ob nicht vielmehr durch bestehende Mandate gesellschaftlicher Funktionssysteme Inklusions- und Exklusionsverhältnisse perpetuiert oder gar verschärft werden (vgl. Lenzen 1999; Meseth 2021).

Auch wenn, wie gezeigt, KI-Technologien durchaus hohe Potenziale für die Inklusion benachteiligter oder behinderter Menschen bieten, sind jeweils ethische, soziale und praxisbezogene Fragestellungen eingehend zu prüfen. Folgende

Anforderungen an zukünftige Entwicklungen im Bereich inklusiver KI-Technologien sind aus fachlicher Perspektive zu stellen:

- **Technologieentwicklung:** Ausgangs- und Zielpunkt ist die Entwicklung verantwortungsvoller, subjekt- und kontextsensibler KI-Technologien. Entsprechende Leitlinien, Zielformulierungen und Bewertungskriterien sind stets im Vorfeld der Technologieentwicklung zu formulieren (vgl. Floridi et al. 2020; Maccenaere 2024).
- **Einbezug der Stakeholder:** Die Entwicklung von KI-Technologien sollte alle Stakeholder einbeziehen, um die Bedürfnisse von Adressat:innen zu verstehen und für diese sinnstiftende und angepasste Lösungen zu schaffen. Usability-Studien sind nötig, um verschiedene Bereitstellungsformen zu evaluieren.
- **Interdisziplinäre Zusammenarbeit:** Fachexpert:innen unterschiedlicher Disziplinen sind in die Entwicklung einzubeziehen, um Modelle zu entwickeln, die Exklusion und soziale Ungleichheiten nicht verstärken. Dies gilt insbesondere für marginalisierte Gruppen und die Berücksichtigung intersektionaler Aspekte, vor allem im Bereich Behinderung.
- **Ethische Überlegungen und Datenschutz:** Bei der Integration von KI in verschiedene Bereiche, insbesondere in die Soziale Arbeit, Medizin und Psychotherapie, müssen ethische Aspekte und der Datenschutz berücksichtigt werden. Richtlinien zur ethischen Anwendung und deren transparente Kommunikation sind erforderlich, um Vertrauen bei Adressat:innen zu schaffen.
- **Grundlagenforschung:** Trotz vielversprechender Entwicklungen gibt es Herausforderungen, die weitere Forschung zu spezifischen Anwendungen, zur Validierung von KI-Technologien und zur Erschließung neuer Technologien (z. B. Open-Source-Modelle) erfordern. Ein weiterer Schwerpunkt sollte auf der Schaffung flexibler KI-Modelle liegen, die an neue Erkenntnisse angepasst werden können. Darüber hinaus besteht ein dringender Bedarf an Grundlagenforschung zur digitalen Ungleichheit, um den Zugang zu und die Nutzung von KI-Technologien u. a. für Menschen mit Behinderungen zu verbessern.

Verstehen wir Inklusion mit Felder (2018) als aktive Partizipation unterschiedlicher Menschen in einem gemeinsamen Bildungsumfeld, wären KI-Technologien entsprechend zur Unterstützung von Bildungsprozessen für die Ausgestaltung demokratisch-inklusive Handelns zu entwickeln und einzusetzen (vgl. Rashid et al. 2023). KI könnte damit – im Idealfall – als Teil soziotechnischer Systeme zur Ausgestaltung einer kollektiven und partizipatorischen Form (post)humaner Intelligenz beitragen.

Literatur

- Bastian, Pascal/Schrödter, Mark (2015): Risikotechnologien in der professionellen Urteilsbildung der Sozialen Arbeit. In: Kutscher, Nadia/Ley, Thomas/Seelmeyer, Udo (Hrsg.): *Mediatisierung (in) der Sozialen Arbeit*. Baltmannsweiler: Schneider Verlag Hohengehren, S. 192–207.
- Bendle, Mervyn Frederick (1996): Logics of integration and disintegration in contemporary social theory. In: *The Australian and New Zealand Journal of Sociology* 32(3), S. 70–84. <https://doi.org/10.1177/144078339603200305>
- Burghardt, Jennifer/Kieslinger, Daniel (2025): Inklusiv beraten – Künstliche Intelligenz als Unterstützung in digitalen Beratungsprozessen. In: Macsenaere, Michael (Hrsg.): *Künstliche Intelligenz in der Kinder- und Jugendhilfe*. 1. Auflage. München: Ernst Reinhardt, S. 94–101.
- Camp, Edward J./Quon, Robert J./Sajatovic, Martha/Briggs, Farren/Brownrigg, Brittany/Janevic, Mary R./Meisenhelter, Stephen/Steimel, Sarah A./Testorf, Markus E./Kiriakopoulos, Elaine/Mazanec, Morgan T./Fraser, Robert T./Johnson, Erica K./Jobst, Barbara C. (2022): Supervised machine learning to predict reduced depression severity in people with epilepsy through epilepsy self-management intervention. In: *Epilepsy & Behavior* 127, 108548. <https://doi.org/10.1016/j.yebeh.2021.108548>
- Deka, Chinmoy/Shrivastava, Abhishek/Abraham, Ajish K./Nautiyal, Saurabh/Chauhan, Praveen (2022): AI-Based Automated Speech Therapy Tools for persons with Speech Sound Disorders: A Systematic Literature Review. Preprint arXiv: <https://arxiv.org/abs/2204.10325>
- Eubanks, Virginia (2018): *Automating Inequality: How High-Tech Tools Profile, Police, and Punish the Poor*. New York, NY, USA: St. Martin's Press, Inc.
- Felder, Franziska (2018): The Value of Inclusion. In: *Journal of Philosophy of Education* 52(1), S. 54–70. <https://doi.org/10.1111/1467-9752.12280>
- Floridi, Luciano/Cowls, Josh/King, Thomas C./Taddeo, Mariarosaria (2020): How to Design AI for Social Good: Seven Essential Factors. In: *Science and Engineering Ethics* 26(3), S. 1771–1796. <https://doi.org/10.1007/s11948-020-00213-5>
- Fosch-Villaronga, Eduard/Poulsen, Adam. (2022). Diversity and Inclusion in Artificial Intelligence. In: Custers Bart/Fosch-Villaronga Eduard (Hrsg.), *Law and Artificial Intelligence: Regulating AI and Applying AI in Legal Practice* (S. 109–134). The Hague. T. M. C. Asser Press. https://doi.org/10.1007/978-94-6265-523-2_6
- Frey, William R./Patton, Desmond U./Gaskell, Michael B./McGregor, Kyle A. (2020): Artificial Intelligence and Inclusion: Formerly Gang-Involved Youth as Domain Experts for Analyzing Unstructured Twitter Data. In: *Social Science Computer Review* 38(1), S. 42–56. <https://doi.org/10.1177/0894439318788314>
- Gillingham, Philip (2016): Predictive Risk Modelling to Prevent Child Maltreatment and Other Adverse Outcomes for Service Users: Inside the ‚Black Box‘ of Machine Learning. In: *The British Journal of Social Work* 46(4), S. 1044–1058. <https://doi.org/10.1093/bjsw/bcv031>
- Gillingham, Philip (2019): Developments in Electronic Information Systems in Social Welfare Agencies: From Simple to Complex. In: *The British Journal of Social Work* 49(1), S. 135–146. <https://doi.org/10.1093/bjsw/bcy014>
- Goldhaber-Fiebert, Jeremy D./Prince, Lea (2019): Impact Evaluation of a Predictive Risk Modeling Tool for Allegheny County's Child Welfare Office. Stanford. Stanford University. https://www.alleghenycounty.us/files/assets/county/v/1/services/dhs/documents/allegheny-family-screening-tool/impact-evaluation-from-16-acdhs-26_predictiverisk_package_050119_final-6.pdf (Abfrage: 15.06.2025).
- Griffen, Brenna/Lorah, Elizabeth R./Holyfield, Christine/Caldwell, Nicolette/Nosek, John (2024): Evaluating Artificial Intelligence on the Efficacy of Preference Assessments for Preservice Speech-Language Pathologists. In: *Journal of Developmental and Physical Disabilities*. <https://doi.org/10.1007/s10882-024-09976-2>

- Gupta, Shivani/Gupta, Monika/Bal, Satinder (2024): Analysis of AI-enhanced educational tools developed in India for linguistic minorities and disabled people. In: *Information Technologies and Learning Tools 100*, S. 199–216. <https://doi.org/10.33407/itlt.v100i2.5501>
- Habermas, Jürgen (1996): *Strukturwandel der Öffentlichkeit*. 5. Auflage. Frankfurt a. M. Suhrkamp.
- Hsu, Yen-Chia/Huang, Ting-Hao, Kenneth/Verma, Himanshu/Mauri, Andrea/Nourbakhsh, Illah/Bozzon, Alessandro (2022) Empowering local communities using artificial intelligence. *Patterns* 3(3) 100449. <https://doi.org/10.1016/j.patter.2022.100449>
- Ibison, Yasmin/Guler, Gulsen/Remfry, Elizabeth/Garcia, Ismael Kherroubi/Barrow, Nicholas/Duarte, Tania (2024): Grassroots and non-profit perspectives on generative AI. Joseph Rowntree Foundation. <https://www.jrf.org.uk/sites/default/files/pdfs/grassroots-and-non-profit-perspectives-on-generative-ai-621d1b4e0d671013a489a8f89fc55a2e.pdf> (Abfrage: 15.06.2025).
- Imbusch, Peter/Heitmeyer, Wilhelm (Hrsg.) (2008): *Integration – Desintegration: Ein Reader zur Ordnungsproblematik moderner Gesellschaften*. 1. Auflage. Wiesbaden: VS Verlag für Sozialwissenschaften.
- Jiang, Zhili/Huang, Xiting/Wang, Zhiqian/Liu, Yang/Huang, Lihua/Luo, Xiaolin (2024). Embodied Conversational Agents for Chronic Diseases: Scoping Review. In: *Journal of Medical Internet Research* 26(1), e47134. <https://doi.org/10.2196/47134>
- Joshi, Yuvraj (2014): The Trouble with Inclusion. In: *Virginia Journal of Social Policy and the Law* 21(2). (SSRN Scholarly Paper Nr. 2381194). <https://papers.ssrn.com/abstract=2381194>
- Krämer, Klaus (2008): Integration und Desintegration Wie aktuell sind diese soziologischen Schlüsselbegriffe noch für eine moderne Gesellschaftsanalyse? In: *Schweizerische Zeitschrift für Soziologie* 1(34), S. 37–53.
- Landhäußer, Sandra (2009): *Communityorientierung in der Sozialen Arbeit: Die Aktivierung von sozialem Kapital*. Wiesbaden: VS Verlag für Sozialwissenschaften. https://doi.org/10.1007/978-3-531-91379-7_2
- Leech, Gavin/Garfinkel, Simson/Yagudin, Misha/Briand, Alexander/Zhuravlev, Aleksandr (2024): Ten Hard Problems in Artificial Intelligence We Must Get Right. Preprint arXiv. <https://doi.org/10.48550/arXiv.2402.04464>
- Lejeune, Alban/Le Glaz, Aziliz/Perron, Pierre-Antoine/Sebti, Johan/Baca-Garcia, Enrique/Walter, Michel/Lemey, Christophe/Berrouguet, Sofian (2022): Artificial intelligence and suicide prevention: A systematic review. In: *European Psychiatry* 65(1), S. e19. <https://doi.org/10.1192/j.eurpsy.2022.8>
- Lenton-Brym, Ariella P./Collins, Alexis/Lane, Jeanine/Busso, Carlos/Ouyang, Jessica/Fitzpatrick, Skye/Kuo, Janice R./Monson, Candice M. (2024): Using machine learning to increase access to and engagement with trauma-focused interventions for posttraumatic stress disorder. In: *British Journal of Clinical Psychology*, 64(1). <https://doi.org/10.1111/bjc.12468>
- Lenzen, Dieter (1999): Jenseits von Inklusion und Exklusion. Disklusion durch Entdifferenzierung der Systemcodes. In: *Zeitschrift für Erziehungswissenschaft* 2(4), S. 545–555. <https://doi.org/10.25656/01:4535>
- Lenzen, Manuela (2018): *Künstliche Intelligenz: Was sie kann & was uns erwartet* (Originalausgabe). München: C. H. Beck.
- Lenzen, Manuela (2020): *Künstliche Intelligenz: Fakten, Chancen, Risiken* (Originalausgabe). München: C. H. Beck.
- Lukkien, Dirk R. M./Nap, Henk Herman/Buimer, Hendrik P./Peine, Alexander/Boon, Wouter P. C./Ket, Johannes C. F./Minkman, Mirella M. N./Moors, Ellen H. M. (2023): Toward Responsible Artificial Intelligence in Long-Term Care: A Scoping Review on Practical Approaches. In: *The Gerontologist* 63(1), S. 155–168. <https://doi.org/10.1093/geront/gnab180>
- Macsenaere, Michael (2024): *Ausblick und Empfehlungen*. In: Macsenaere, Michael (Hrsg.): *Künstliche Intelligenz in der Kinder- und Jugendhilfe*. 1. Auflage. München: Ernst Reinhardt, S. 110–119.

- Markelius, Alva/Wright, Connor/Kuiper, Joahna/Delille, Natalie/Kuo, Yu-Ting (2024): The mechanisms of AI hype and its planetary and social costs. In: *AI and Ethics* 4. <https://doi.org/10.1007/s43681-024-00461-2>
- Marziali, Rachele Alessandra/Franceschetti, Claudia/Dinculescu, Adrian/Nistorescu, Alexandru/Kristály, Dominic Mircea/Moşoi, Adrian Alexandru/Broekx, Ronny/Marin, Mihaela/Vizitiu, Cristian/Moraru, Sorin-Aurel/Rossi, Lorena/Di Rosa, Mirko (2024): Reducing Loneliness and Social Isolation of Older Adults Through Voice Assistants: Literature Review and Bibliometric Analysis. *Journal of Medical Internet Research* 26(1), e50534. <https://doi.org/10.2196/50534>
- McQuillan, Dan/Jarke, Juliane/Pargman, Teresa Cerratto (2024): We Are at an Extreme Point Where We Have to Go All in on What We Really Believe Education Should Be About. *Postdigital Science and Education* 6(1), S. 360–368. <https://doi.org/10.1007/s42438-023-00433-5>
- Medenica, Veselin/Ivanovic, Lidija/Milosevic, Neda (2024). Applicability of artificial intelligence in neuropsychological rehabilitation of patients with brain injury. *Applied Neuropsychology: Adult*, S. 1–28. <https://doi.org/10.1080/23279095.2024.2364229>
- Meseth, Wolfgang. (2021). Inklusion und Normativität – Anmerkungen zu einigen Reflexionsproblemen erziehungswissenschaftlicher (Inklusions-)Forschung. In Fritzsche Bettina/Köpfer Andreas/Wagner-Willi Monika/Böhmer Anselm/Nitschmann Hannah/Lietzmann Charlotte/Weitkämper Florian (Hrsg.), *Inklusionsforschung zwischen Normativität und Empirie: Abgrenzungen und Brückenschläge* (S. 19–36). Verlag Barbara Budrich.
- Montanha, Aleksandro/Oprescu, Andreea M./Romero-Tertero, MCarmen (2022): A Context-Aware Artificial Intelligence-based System to Support Street Crossings for Pedestrians with Visual Impairments. *Applied Artificial Intelligence* 36(1), 2062818. <https://doi.org/10.1080/08839514.2022.2062818>
- Park, Sungjin (2023): Theodor W. Adorno, Artificial Intelligence, and Democracy in the Postdigital Era. In: *Postdigital Science and Education* 6, S. 1287–1303. <https://doi.org/10.1007/s42438-023-00424-6>
- Prescott, Julie/Barnes, Steven (2024): Artificial intelligence positive psychology and therapy. *Counselling and Psychotherapy Research* 24(3), S. 843–845. <https://doi.org/10.1002/capr.12784>
- Rashid, Md Tahmid/Wei, Na/Wang, Dong (2023): A survey on social-physical sensing: An emerging sensing paradigm that explores the collective intelligence of humans and machines. *Collective Intelligence* 2(2), 26339137231170825. <https://doi.org/10.1177/26339137231170825>
- Rojas, Fabian Andrey Zarta (2024). Capitalism and Social Inclusion: Divergences, Utopias and Challenges. *Revista Perspectivas* 9(1), S. 155–164. <https://doi.org/10.22463/25909215.4101>
- Roth, Gerhard/Tuggener, Lukas/Roth, Fabian Christoph (2024): Künstliche Intelligenz. In: Roth, Gerhard/Tuggener, Lukas/Roth, Fabian Christoph (Hrsg.): *Natürliche und künstliche Intelligenz: Ein kritischer Vergleich*. Berlin und Heidelberg: Springer, S. 131–200. https://doi.org/10.1007/978-3-662-68401-6_5
- Salomé, Sidonie/Monfort, Emmanuel (2023): Révolution numérique et âgisme : les enjeux éthiques de l'intelligence artificielle pour les personnes âgées. *NPG Neurologie – Psychiatrie – Gériatrie*, 23(138), S. 383–387. <https://doi.org/10.1016/j.npg.2023.09.004>
- Seelmeyer, Udo (2020): Big Data & Künstliche Intelligenz – Neue Anforderungen an die Fachlichkeit in sozialen Berufen. In: *Blätter der Wohlfahrtspflege* 167(3), S. 95–98. <https://doi.org/10.5771/0340-8574-2020-3-95>
- Seifert, Ruth (2013): Eine Debatte Revisited: Exklusion und Inklusion als Themen der Sozialen Arbeit. In: *Zeitschrift für Inklusion*. <https://www.inklusion-online.net/index.php/inklusion-online/article/view/25> (Abfrage: 15.06.2024).
- Shams, Rifat/Zowghi, Didar/Bano, Muneera (2023): Challenges and Solutions in AI for All.
- Stichweh, Rudolf (1997): Inklusion/Exklusion und die Theorie der Weltgesellschaft. In: Karl-Siegbert Rehberg (Hrsg.): *Differenz und Integration: die Zukunft moderner Gesellschaften; Verhandlungen des 28. Kongresses der Deutschen Gesellschaft für Soziologie im Oktober 1996 in*

- Dresden; Band 2: Sektionen, Arbeitsgruppen, Foren, Fedor-Stepun-Tagung. Opladen: Westdeutscher Verlag, S. 601–607.
- Turing, A. M. (1950): Computing machinery and intelligence. In: *Mind* LIX(236), S. 433–460. <https://doi.org/10.1093/mind/LIX.236.433>
- Wang, Jiaji/Wang, Shuihua/Zhang, Yudong (2023): Artificial intelligence for visually impaired. In: *Displays* 77, 102391. <https://doi.org/10.1016/j.displa.2023.102391>
- Wesselmann, Carla (2022): Partizipation, Inklusion und Exklusion im Kontext von Behinderung – Eckpunkte einer (kritischen) Teilhabeforschung!? In: Wansing, Gudrun/Schäfers, Markus/Köbsell, Swantje (Hrsg.): *Teilhabeforschung – Konturen eines neuen Forschungsfeldes*. Wiesbaden: Springer Fachmedien, S. 67–84. https://doi.org/10.1007/978-3-658-38305-3_4
- Wilson, John (1999): Some Conceptual Difficulties about ‚Inclusion‘. In: *Support for Learning* 14(3), S. 110–112. <https://doi.org/10.1111/1467-9604.00114>
- World Health Organization (Hrsg.). (2001). *International classification of functioning, disability and health: ICF*. Geneva. World Health Organization.
- Zhou, Sijia/Zhao, Jingping/Zhang, Lulu (2022): Application of Artificial Intelligence on Psychological Interventions and Diagnosis: An Overview. In: *Frontiers in Psychiatry* 13. <https://doi.org/10.3389/fpsy.2022.811665>

KI und Alter: Einführung, Potenziale und Herausforderungen¹

Anna Schlomann

Abstract: Vor dem Hintergrund des demografischen Wandels beleuchtet der Beitrag Potenziale und Herausforderungen von KI für ältere Menschen aus gerontologischer und sozialarbeiterischer Perspektive. Als konzeptioneller Rahmen werden Alters- und Kohorteneffekte beschrieben, um Unterschiede in der Nutzung, Akzeptanz und Integration neuer Technologien wie KI in den Alltag älterer Menschen besser zu verstehen. Mögliche Potenziale von KI werden vor diesem Hintergrund in den Bereichen Wohnen und Mobilität, soziale Integration und in Bezug auf Gesundheit und Pflege beschrieben. Demgegenüber stehen Herausforderungen wie eine weiter bestehende digitale Spaltung, Datenschutzfragen und ethische Aspekte wie Altersdiskriminierung. Der Beitrag bezieht auch die Ergebnisse eines Forschungsprojekts zu KI-basierter Sprachassistenten für ältere Menschen (Projekt KI-Alter) in die Darstellung ein. Das Kapitel schließt mit einer Zusammenfassung und einem Ausblick auf zukünftige Forschungsbedarfe.

Keywords: Digitale Kompetenz, demografischer Wandel, Sprachassistent, Akzeptanz, Lebensqualität

1 Hintergrund und Einführung

Künstliche Intelligenz (KI) hat das Potenzial, Alltagsgewohnheiten, soziale Interaktionen, Gesundheitsmanagement sowie Informations- und Mobilitätsverhalten auch im höheren Lebensalter grundlegend zu verändern. Sie findet bereits heute Anwendung in verschiedenen Bereichen des Alltags wie Navigations- und Fahrassistentensystemen, Smart-Home-Technologien, KI-Chatbots wie ChatGPT oder in Smart Speakern. Trotz einer insgesamt hohen Bekanntheit des Begriffs „Künstliche Intelligenz“ bei älteren Menschen und dem Bewusstsein, dass KI bereits in vielen alltäglichen Anwendungsbereichen eingesetzt wird, betrachtet eine

1 © Die Autor:innen 2025, veröffentlicht im Verlag Beltz Juventa, Open Access: CC BY 4.0
Gesa Linnemann / Julian Löhe / Beate Rottkemper (Hg.), *Künstliche Intelligenz in der Sozialen Arbeit*
10.3262/978-3-7799-8562-4_011

Mehrheit der Älteren KI als nicht relevant für das eigene Leben (Körper-Stiftung 2024).

Vor dem Hintergrund einer weltweiten Alterung von Bevölkerungen – Prognosen zufolge wird im Jahr 2050 etwa ein Viertel der Menschen in Europa und Nordamerika 65 Jahre oder älter sein (United Nations 2019) – stellt sich die Frage, welche Potenziale und Herausforderungen mit der immer stärkeren Verbreitung von KI-basierten Technologien für die Gruppe der älteren Menschen einhergehen. Insbesondere die stark wachsende Gruppe hochaltriger Menschen über 80 Jahre (Eurostat 2020; Neise et al. 2019) rückt dabei in den Fokus. Hochaltrigkeit ist häufiger mit gesundheitlichen und funktionalen Einschränkungen und mit Pflegebedarfen assoziiert (ebd.), und es muss diskutiert werden, welche Rolle KI-basierte Technologien zukünftig auch in der gesundheitlichen und pflegerischen Versorgung alter und sehr alter Menschen spielen können und sollten.

Die Auseinandersetzung mit KI in gerontologischer Forschung und Praxis steht aktuell noch am Anfang (Fachübergreifender Ausschuss Alter und Technik der DGGG 2024). Forschungsarbeiten haben insgesamt gezeigt, dass digitale Technologien im hohen und sehr hohen Lebensalter förderliche Effekte in Bezug auf Lebensqualitätsindikatoren wie Einsamkeit, Autonomie, Gesundheit und Wohlbefinden haben können (Kamin 2024; Schlomann et al. 2020b). Mögliche Potenziale und Herausforderungen der Digitalisierung für ältere Menschen werden ebenso im Kontext von Sozialer Arbeit und Alter vermehrt aufgegriffen und diskutiert (Tonello 2020). Auch KI-basierte Technologien werden zunehmend in den Blick genommen, wobei theoretisch-konzeptionelle Auseinandersetzungen mit dem Thema bislang kaum vorliegen (Gallistl et al. 2024). Im Kontext Sozialer Arbeit und Alter wurde das Thema KI bisher u. a. mit einem Fokus darauf thematisiert, inwiefern der Bereich der Sprachverarbeitung durch KI (Natural Language Processing; NLP) in der Arbeit für und mit älteren Menschen genutzt werden kann und sollte (Linnemann/Löhe/Rottkemper 2023). Andere Arbeiten diskutieren in diesem Zusammenhang, inwiefern KI verschiedene Tätigkeiten in der Sozialen Arbeit zukünftig ergänzen, verändern oder sogar ersetzen könnte (Pottharst/Neumann/Ostrau/Seelmeyer 2024).

Ein konsequent kritisches Begleiten der KI-Entwicklung und deren Implikationen für gerontologische Forschungs- und Anwendungsprojekte sowie die Profession der Sozialen Arbeit sind geboten und notwendig, weil eine weitere Verbreitung derartiger Systeme zukünftig zu erwarten ist und diese Entwicklung neue ethische und soziale Fragen entstehen lässt.

Ziel dieses Beitrags ist es, hier anzusetzen und einen Überblick über mögliche Potenziale und Herausforderungen von KI für gerontologische Forschung und in der praktischen Arbeit für und mit älteren Menschen zu geben. Dazu wird zunächst ein konzeptioneller Rahmen beschrieben und die Relevanz von Alters- und Kohorteneffekten bei der Nutzung digitaler und KI-basierter Technologien dargestellt. Darauf aufbauend werden beispielhaft für verschiedene Lebensbereiche

mögliche Potenziale von KI in diesen Kontexten näher beleuchtet sowie Herausforderungen aufgezeigt. Der Anspruch dieses Beitrags ist es nicht, alle Facetten von KI und Alter umfassend und abschließend zu bearbeiten, sondern es sollen ein erster Überblick zum Thema gegeben und auf dieser Basis Empfehlungen für zukünftige Forschungs- und Anwendungsprojektkontexte abgeleitet werden. Dabei werden auch die Ergebnisse eines kürzlich abgeschlossenen Forschungsprojekts an der Pädagogischen Hochschule Heidelberg und Universität Heidelberg zu KI-basierter Sprachassistentenz (Projekt *KI-Alter – KI-basierte Sprachassistentenz für ältere Menschen mit und ohne geistige(r) Behinderung*²; Schlomann et al. 2021a) einbezogen sowie auf weitere aktuelle Forschungsprojekte zu KI und Alter verwiesen.

2 Konzeptioneller Rahmen

Um die Bedeutung von KI-basierten Technologien für ältere Menschen konzeptionell zu beschreiben und einzuordnen, können diese als (neuer) Teil einer digitalisierten Umwelt Älterer betrachtet werden (Wahl/Gerstorf 2018). Aus Sicht der ökologischen Gerontologie können neue Technologien zu einem stimulierenden Umfeld für erfolgreiches Altern beitragen (Schulz et al. 2015). Die Besonderheiten der Nutzung, Akzeptanz und Alltagsintegration neuer Technologien wie KI bei älteren im Vergleich zu jüngeren Personen können dabei sowohl durch Alterseffekte (Veränderungen über die Lebensspanne) als auch durch Kohorteneffekte (Veränderungen in der Gesamtheit der Meinungen oder Einstellungen von Menschen, die derselben Generation angehören) erklärt werden (siehe auch Schlomann/Even/Hammann 2022).

2.1 Alterseffekte bei Nutzung digitaler und KI-basierter Technologien

Multimorbidität, chronische Erkrankungen sowie sensorische Defizite nehmen im höheren Lebensalter zu (Polidori/Häussermann 2019). Kognitive und sensorische Einschränkungen sowie geringere finanzielle und soziale Ressourcen können auch die Nutzung neuer Technologien im höheren Lebensalter erschweren (Schulz et al. 2015). Relevant sind dabei beispielsweise körperliche Veränderungen, die die Feinmotorik der Hände beeinträchtigen und so beispielsweise die Nutzung von Touchscreens auf Smartphones oder Tablets erschweren; Seh- und Hörbeeinträchtigungen können die Nutzung von Multimedia-Inhalten einschränken und mögliche kognitive Beeinträchtigungen können dazu führen, dass neue Informationen nur langsam aufgenommen werden. Insgesamt müs-

2 Das Projekt KI-Alter wurde von 10/2020 bis 6/2024 von der Baden-Württemberg Stiftung im Rahmen des Forschungsprogramms „Verantwortliche Künstliche Intelligenz“ gefördert.

sen physische, soziale und biologische Aspekte und Anforderungen in Studien zur KI-Nutzung und -Aneignung sowie bei der Entwicklung von KI-basierten Technologien berücksichtigt werden, damit sie auch für Ältere mit Einschränkungen nutzbar sind und akzeptiert werden. Gleichzeitig können KI-basierte Technologien dazu beitragen, altersbedingte Einschränkungen zu kompensieren, beispielsweise indem alltägliche Aufgaben automatisiert oder durch Sprachsteuerung erleichtert werden (siehe auch Abschnitt 3 in diesem Beitrag).

2.2 Kohorteneffekte bei Nutzung digitaler und KI-basierter Technologien

Kohorteneffekte der Technologienutzung wurden im Konzept der Technikgenerationen beschrieben (Sackmann/Weymann/Hüttner 1994; Sackmann/Winkler 2013). Eine Technikgeneration wird hierbei als Gruppe von Geburtskohorten definiert, deren Verhalten und Einstellungen gegenüber Technologien durch die Auswirkungen einer oder mehrerer größerer technologischer Veränderungen beeinflusst werden, die in ihrer formativen Phase stattgefunden haben. Dem Konzept der Technikgenerationen folgend liegt die formative Phase zwischen dem zehnten und 25. Lebensjahr, und in dieser Lebensphase erworbenes Wissen ist auch im höheren Alter leichter abrufbar. Neue Technologien, die in dieser Zeit zur Verfügung stehen, lassen sich somit im gesamten Lebensverlauf leichter erlernen und effektiver nutzen. Die heutigen (Technik-)Generationen der älteren Menschen sind nicht mit digitalen und KI-basierten Technologien aufgewachsen und könnten daher zurückhaltender sein, wenn es um deren Nutzung und Aneignung geht. Diese unterschiedlichen Vorkenntnisse müssen auch in Forschung und Praxis zu KI und Alter berücksichtigt werden. Damit ältere Menschen KI-basierte Technologien effektiv und nach eigenen Interessen und Bedürfnissen nutzen können, sollten Schulungskonzepte und zielgruppenspezifische Angebote zur Vermittlung von Wissen geschaffen werden (Bundesarbeitsgemeinschaft der Seniorenorganisationen e. V. 2024b).

3 Potenziale von KI in zentralen Lebensbereichen älterer Menschen

Der Argumentation des achten Altersberichts folgend können verschiedene Lebensbereiche und Handlungsfelder älterer Menschen differenziert werden, in denen sich eine hohe Relevanz von Digitalisierung zeigt (Deutscher Bundestag 2020). In diesem Abschnitt werden dazu die folgenden drei aufgegriffen, für die aufbauend auf den konzeptionellen Vorüberlegungen mögliche (zukünftige) Potenziale von KI-basierten Technologien für Menschen in hohem und sehr hohem

Lebensalter aufgezeigt werden: (1) Wohnen und Mobilität, (2) Soziale Integration und Einsamkeit, (3) Gesundheit, Versorgung und Pflege.

3.1 Wohnen und Mobilität

KI-basierte Technologien können durch Smart-Home-Elemente, beispielsweise zur Steuerung von Beleuchtung, Heizung oder Verdunklung, das Wohnumfeld an individuelle Bedürfnisse anpassen und so dazu beitragen, eine selbstständige Lebensführung auch bei Vorliegen von altersbedingten Einschränkungen zu ermöglichen. Auf diese Weise könnten ältere Menschen länger in der eigenen Häuslichkeit verbleiben und ein Umzug in ein Pflegeheim oder betreutes Wohnen wird vermieden, was die Präferenz vieler Älterer darstellt (Golant 2020). Auch pflegenden Angehörige könnten auf diese Weise in Teilen entlastet werden.

Aspekte des Wohnens müssen in der Gruppe der alten und sehr alten Menschen auch für institutionelle Wohnformen wie betreutes Wohnen oder Alten- und Pflegeheime mitgedacht werden und dabei weitere Stakeholder in die Betrachtung einbezogen werden (z. B. Angehörige, Pflegefachkräfte, Sozialarbeitende, medizinisches Fachpersonal). In Bezug auf die Implementierung neuer Technologien wie Smart Home oder anderer KI-Anwendungen sind die Bewohner:innen auf vorhandene Infrastrukturen und oft auch auf Unterstützung durch das Personal angewiesen, was besondere Anforderungen an die Realisierung und tatsächliche Nutzung der Technologien stellt. Idealerweise sollte das Pflege- und Betreuungspersonal entsprechend geschult sein, um die Bewohner:innen zu unterstützen. Vielversprechende Ansätze werden in diesem Kontext im Projekt DiBiWohn erarbeitet, das digitale Bildungsprozesse für ältere Menschen in senioren-spezifischen Wohnformen wie betreutem Wohnen und Pflegeeinrichtungen fördert. Im Rahmen von DiBiWohn wird erforscht, wie digitale Zugänge und Bildungsangebote für diese oft digital unerfahrene Zielgruppe geschaffen werden können. Durch Peer-to-Peer-Ansätze und ehrenamtliche Technikbegleiter:innen sollen digitale Kompetenzen gestärkt und die soziale Teilhabe gefördert werden (DiBiWohn 2024).

Ebenso werden sich zukünftig die Infrastrukturen von Städten und des Verkehrs sowie das Mobilitätsverhalten durch Fortschritte in KI-basierten Technologien ändern (Cugurullo et al. 2021).

Mobilität im höheren Alter ist ein relevanter Faktor, um Unabhängigkeit, soziale Teilhabe und Lebensqualität aufrechtzuerhalten. In diesem Kontext könnten KI zukünftig beispielsweise durch autonome Fahrzeuge, eine intelligente Verkehrssteuerung oder die Optimierung von Routenplanungen zu sichereren und umweltfreundlicheren Transportlösungen beitragen. Diese Entwicklungen können sich einerseits positiv auf Komfort und Sicherheit auswirken, andererseits

auch Barrierefreiheit für ältere Menschen und/ oder Menschen mit eingeschränkter Mobilität fördern.

Studien zeigen, dass ältere Menschen dem autonomen Fahren grundsätzlich positiv gegenüberstehen und Potenziale für stressfreiere Mobilität, für mehr soziale Interaktion und körperliche Aktivität sehen, wobei auch Bedenken bezüglich der Sicherheit, Kosten und Benutzerfreundlichkeit bestehen (Zandieh/Acheampong 2021). Vor allem in ländlichen Regionen, in denen der öffentliche Personennahverkehr nur eingeschränkt verfügbar ist, könnten KI-basierte Mobilitätslösungen zukünftig positive Effekte für Ältere bieten. Geäußerte Bedenken müssen jedoch ernstgenommen und berücksichtigt werden.

3.2 Soziale Integration und Einsamkeit

KI-basierten Chatbots und sozialen Robotern wird das Potenzial zugeschrieben, Gesellschaft leisten zu können, emotionale Unterstützung zu bieten, oder bei alltäglichen Aufgaben zu unterstützen. Kommerzielle KI-basierte Sprachassistenten wie Siri oder Alexa werden von älteren Menschen bereits jetzt immer häufiger genutzt (Bitkom e. V. 2023; Rathgeb et al. 2021). Auch in Forschungsprojekten wurden diese Technologien in den letzten Jahren vermehrt in ihrem Potenzial bezüglich sozialer und digitaler Teilhabe (Scherr et al. 2021) und Reduzierung von Einsamkeit (Corbett et al. 2021; Marziali et al. 2024) fokussiert, wobei es bisher nur wenige Studien zu älteren Menschen gibt (Even et al. 2022).

In einer systematischen Literaturlauswertung mit Fokus auf ältere Menschen mit und ohne Behinderung konnten insgesamt sechs Themen-Cluster zu Potenzialen von Sprachassistenten identifiziert werden (Schlomann et al. 2021b). Hierzu gehörten auch Potenziale in Bezug auf soziale Integration; dabei insbesondere die erleichterte (digitale) Kommunikation mit anderen Personen bei eingeschränkter Sehkraft oder Einschränkungen in der Nutzung der Hände/ Finger (ebd.; Even et al. 2022). Ein weiteres Potenzial in Bezug auf Lebensqualitätsindikatoren wurde in einem Gefühl von Gesellschaft und Begleitung durch den Sprachassistenten identifiziert (ebd.).

Eine Analyse der tatsächlichen Nutzung eines kommerziellen Sprachassistenten über vier Wochen im Alltag älterer Menschen im Rahmen des Projekts KI-Alter zeigte in diesem Kontext ebenfalls, dass neben der Nutzung von Informations- und Unterhaltungsangeboten auch die Kommunikation mit dem Sprachassistenten selbst, wozu beispielsweise Begrüßungen und Verabschiedungen, Höflichkeitsrituale oder „Small Talk“ gehören, relativ häufig von älteren Personen genutzt wurden (Schlomann et al. 2024). Diese Befunde deuten darauf hin, dass Sprachassistenten von einem Teil der älteren Nutzer:innen als eine Art Gesprächspartner betrachtet werden und eine Strukturierung und Zeitvertreib im Alltag bieten können (ebd.). Es besteht jedoch zugleich die Gefahr, dass

hierdurch versucht wird, menschliche Interaktionen zu ersetzen und die Technologie vermenschlicht wird (Purington et al. 2017). In der Konsequenz könnte die Nutzung von Sprachassistenten somit auch zu einem verringerten Kontakt zu Angehörigen, Sozialarbeitenden oder Pflegepersonen führen, was Einsamkeitsgefühle letztendlich verschärfen könnte (Linnemann/Löhe/Rottkemper 2023). In diesem Kontext durchgeführte Forschungs- und Anwendungsprojekte sollten daher stets so positioniert werden, dass die Technologien als Unterstützung und nicht als Ersatz für menschliche Interaktion, Beratung, Betreuung oder eigene Fähigkeiten eingesetzt werden. Eine kritische Begleitung durch professionelle Akteur:innen in Gerontologie und Sozialer Arbeit ist daher geboten.

3.3 Gesundheit, Versorgung und Pflege

Digitale Angebote können dazu beitragen, gesundheitsbezogene Bedarfe niedrigschwellig zu decken (Schäfer et al. 2024), und KI-basierte Technologien können durch Unterstützung bei klinischen Entscheidungen die medizinische Versorgung zukünftig verbessern (Mennella et al. 2024). Durch die Analyse großer Datenmengen könnte KI Ärzt:innen beispielsweise zukünftig helfen, Krankheiten früher und genauer zu diagnostizieren und personalisierte Behandlungspläne zu entwickeln. Eine wichtige Rolle von KI wird in der Hilfe bezüglich der Strukturierung von Daten und dem Abgeben darauf basierender Empfehlungen gesehen (Deutscher Ethikrat 2023).

Das Projekt KIAFlex entwickelt beispielsweise ein KI-gestütztes Assistenzsystem, um das Entlassungsmanagement in Krankenhäusern zu verbessern. Ziel ist es, den Nachsorgebedarf von Patient:innen schon bei der Aufnahme vorherzusagen und durch interaktive KI-Prozesse anzupassen. Virtuelle Sozialarbeiter:innen sollen die Kommunikation mit Angehörigen und die Dokumentation automatisieren, um das Personal zu entlasten. So sollen flexible Entlassungen und eine nahtlose Nachsorge ermöglicht werden (Bundesministerium für Bildung und Forschung 2024). Komplexe Abläufe sowie Entscheidungsprozesse in der Gesundheitsversorgung könnten somit zukünftig durch den Einsatz von KI optimiert werden.

Im Kontext von Sozialer Arbeit werden Potenziale durch Technologien und KI auf den Funktionsebenen Vermittlung, Verarbeitung und Vernetzung beschrieben (Pottharst et al. 2024). Dabei könnten innovative Technologien wie Sprachmodelle, Text Mining oder Entscheidungsunterstützungssysteme Aufgaben wie Dokumentation, Beratung oder Interventionsplanung zukünftig effizienter gestalten, wobei durch neue technologische Entwicklungen auch nicht standardisierte Aufgabenbereiche potenziell unterstützt oder ersetzt werden können. Gleichzeitig sind die tatsächlichen Effekte der aktuellen Technik- und KI-Entwicklung für die Soziale Arbeit jedoch aktuell schwer vorherzusagen (ebd.).

Im Pflegebereich könnte KI zukünftig durch Roboter oder intelligente Assistenzsysteme Pflegekräfte unterstützen, beispielsweise bei körperlich anspruchsvollen Aufgaben oder in der Überwachung von Vitaldaten, indem sie Gesundheitsdaten analysiert und frühzeitig Warnsignale erkennt. Zudem könnten KI-basierte Systeme eine bessere Organisation von Pflegeressourcen ermöglichen, was die Versorgung effizienter und kostengünstiger macht und Fachkräfte entlasten kann. Hierbei erscheint es jedoch essenziell, Fragen von vorhandenen Infrastrukturen, Datenschutz und Akzeptanz der beteiligten Akteur:innen von Anfang an zu berücksichtigen.

4 Herausforderungen in Bezug auf digitale Spaltung, Datenschutz und ethische Aspekte

Den beschriebenen Potenzialen stehen relevante Herausforderungen gegenüber, die mit der fortschreitenden Entwicklung und Alltagsintegration von KI und in Bezug auf das höhere Lebensalter einhergehen. Die Akzeptanz KI-basierter Technologien in der Gruppe der älteren Menschen, aber auch bei anderen relevanten Akteur:innen wie medizinischem und pflegerischem Fachpersonal, Sozialarbeitenden oder Angehörigen ist ein entscheidender Faktor für deren weitreichende Verbreitung und Implementierung. Nur durch Forschung zu Akzeptanz, Kompetenz und Nutzbarkeit der Technologien für alte und sehr alte Menschen können die Präferenzen und Anforderungen der Zielgruppe(n) angemessen berücksichtigt werden. Zu relevanten Herausforderungen gehören darüber hinaus eine fortbestehende digitale Spaltung, Fragen von Datenschutz sowie ethisch relevante Aspekte.

4.1 Digitale Spaltung

Durch eine starke Dynamik in der Entwicklung KI-basierter Technologien besteht die Gefahr einer zunehmenden digitalen Spaltung der Bevölkerung. Neben den in Abschnitt 2 beschriebenen Alters- und Kohorteneffekten wurden weitere Zusammenhänge zwischen personenbezogenen Faktoren und der Nutzung sowie Aneignung neuer Technologien im höheren Lebensalter beschrieben, insbesondere in Bezug auf Geschlecht und Bildung. Studien haben gezeigt, dass ältere Männer häufiger Zugang zum Internet haben und eine größere Bandbreite an Online-Aktivitäten ausüben als ältere Frauen (Joiner/Stewart/Beaney 2015; Seifert/Kamin/Lang 2020). Neue Ergebnisse zeigen zwar, dass genderbezogene Unterschiede der Internetnutzung auch bei Älteren rückläufig sind (Bünning et al. 2023), genderspezifische Präferenzen und mögliche Ängste sollten jedoch ebenfalls in Bezug auf die weitere Entwicklung im Kontext von KI-Technologien im

Blick behalten werden. Ältere Personen mit höherem Bildungsniveau haben oft einen besseren Zugang zu Technologien und nutzen diese häufiger (Seifert/Kamin/Lang 2020). Es besteht die Herausforderung, KI-basierte Systeme für Ältere mit unterschiedlichen Bildungs- und Wissenshintergründen zugänglich und erklärbar zu machen, damit alle Personen potenziell von deren Weiterentwicklung profitieren können und für relevante Herausforderungen sensibilisiert sind.

Die digitale Spaltung findet auf verschiedenen Ebenen statt (van Deursen/van Dijk 2019). Die erste Ebene der digitalen Spaltung (First Level Digital Divide; van Deursen/van Dijk 2019) betrifft den *Zugang*. Insgesamt ist der Anteil der Onliner:innen unter älteren Personen in den letzten Jahrzehnten stetig gestiegen, aber eine digitale Kluft zwischen jüngeren und älteren Menschen besteht weiterhin fort (Rathgeb et al. 2021; Pew Research Center 2021). Vor allem Ältere in institutionellen Wohnformen haben oft keinen Zugang zum Internet (Schlomann et al. 2020a), was eine zentrale Voraussetzung auch für die Nutzung KI-basierter Technologien darstellt. Dies muss sich zukünftig für einen gerechten Zugang zu KI ändern. Darüber hinaus sind digitale *Kompetenzen* (Second Level Digital Divide; van Deursen/van Dijk 2019) für Menschen jeden Alters, einschließlich im höheren Lebensalter, unerlässlich, um aktiv an den heutigen Gesellschaften teilzunehmen und somit auch von möglichen KI-Lösungen zu profitieren. Da ältere Personen in früheren Lebensphasen weniger Gelegenheit hatten, Erfahrungen mit digitalen Technologien zu sammeln (Sackmann/Winkler 2013; siehe ebenso Abschnitt 2.2), sind ihre digitalen Fähigkeiten möglicherweise geringer im Vergleich zu der technologischen Kompetenz jüngerer Menschen.

Verschiedene Projekte adressieren die altersbezogene digitale Spaltung mit Bezug zu KI-Technologien. Die KI-Lernorte der Bundesarbeitsgemeinschaft der Seniorenorganisationen (BAGSO) sind Initiativen, die älteren Menschen digitale Kompetenzen und Wissen über KI vermitteln. Seit 2020 haben sich 32 Einrichtungen wie Senioren-Internet-Initiativen und Mehrgenerationenhäuser zu solchen Lernorten entwickelt. Sie bieten Qualifizierungsprogramme zu Themen wie KI im Alltag und Datenkompetenz an. Die Lernorte ermöglichen es älteren Menschen, mit KI-Technologien praktisch zu arbeiten und diese besser zu verstehen (Bundesarbeitsgemeinschaft der Seniorenorganisationen e. V. 2024a). Konkrete Anwendungsbeispiele zum KI-Lernen im Kontext von kommerziellen Sprachassistenten sind im Projekt KI-Alter erarbeitet worden (Schlomann/Even/Hammann 2024), und das Projekt DiBiWohn adressiert digitales Lernen in seniorenspezifischen Wohnformen (DiBiWohn 2024).

4.2 Datenschutz und ethische Aspekte

Mit der weiteren Verbreitung von KI stellen sich auch neue datenschutzbezogene und ethische Fragen (Deutscher Ethikrat 2023). KI-basierte Systeme sammeln

kontinuierlich Daten und benötigen große Datenmengen, oft personenbezogener Art, was Rückschlüsse auf den Alltag und persönliche Vorlieben erlauben kann und potenziell die Gefahr eines Datenmissbrauchs bedeutet. Für (ältere) Menschen mit geringerer digitaler Kompetenz und/oder kognitiven Einschränkungen besteht ein erhöhtes Risiko, die Tragweite und Konsequenzen dieser Datenerhebung und -verarbeitung nicht vollständig zu verstehen. Um einen Schutz sensibler (Gesundheits-)Daten zu realisieren, sollte der Umgang mit entsprechenden Daten durch KI-Systeme streng überwacht und reguliert werden, um Missbrauch zu verhindern (Mennella et al. 2024).

Bei der Nutzung generativer KI-basierter Systeme besteht zudem die Gefahr, dass gegebene Informationen oder Empfehlungen von KI auf unzureichenden Daten basieren und so im Ergebnis zu Falschinformationen (sogenanntes „Bullshitting“) führen (Bastian 2024). Auch hier sind Personen mit geringeren digitalen Kompetenzen besonders gefährdet, diese Informationen oder Empfehlungen nicht kritisch zu hinterfragen. Eine fehlende Erklärbarkeit von Entscheidungen, die durch KI getroffen werden, kann darüber hinaus das Vertrauen in die Technologie negativ beeinträchtigen (Mennella et al. 2024).

Altersdiskriminierung durch KI ist ein wachsendes Problem, da KI-gestützte Technologien häufig auf Daten basieren, die Altersstereotype und Vorurteile widerspiegeln (Chu et al. 2022). Wenn entsprechend verzerrte Daten für die Entwicklung von Algorithmen verwendet werden, besteht die Gefahr, dass ältere Menschen benachteiligt oder sogar ausgeschlossen werden. Laut Weltgesundheitsorganisation (WHO) kann dies besonders im Gesundheitswesen problematisch sein. Wenn die Datensätze, die zur Diagnose und Therapie älterer Menschen eingesetzt werden, altersbedingte Vorurteile enthalten, könnten ältere Menschen eine schlechtere oder unpassende Behandlung erhalten (World Health Organization 2022).

Zusätzlich können KI-Anwendungen in Arbeitswelt, Personalauswahl und alltäglichen Entscheidungsprozessen Altersdiskriminierung verstärken. Wenn Algorithmen auf veralteten oder diskriminierenden Annahmen basieren, könnte dies die Chancen älterer Menschen am Arbeitsmarkt oder im gesellschaftlichen Leben weiter einschränken (Bogen 2023).

Um diesen Herausforderungen zu begegnen, wurden verschiedene Maßnahmen vorgeschlagen, darunter altersgerechte Datensammlungen und eine stärkere Beteiligung älterer Menschen an der Entwicklung der Technologien (World Health Organization 2022). Im professionellen Kontext müssen Qualifizierungsstrategien entwickelt und Ausbildungsprogramme angepasst werden, um Technikkompetenz und -akzeptanz zu schaffen. Professionell Tätige wie Sozialarbeitende sollten dabei ebenfalls in die Entwicklung und Einführung neuer Technologien einbezogen werden, um diese Prozesse kritisch zu begleiten und einen fachlich und ethisch verantwortungsvollen Einsatz anzustreben (Pottharst/Neumann/Ostrau/Seelmeyer 2024). Aktuell mangelt es jedoch an umfassenden

gesetzlichen Regelungen, die den sicheren Einsatz von KI im Gesundheits- und Sozialwesen gewährleisten. Mennella und Kolleg:innen (2024) betonen daher die Notwendigkeit eines globalen regulatorischen Rahmens, der Datenschutz, Haftungsfragen und ethische Richtlinien berücksichtigt.

5 Zusammenfassung und Ausblick

Vor dem Hintergrund des demografischen Wandels gewinnen Potenziale und Herausforderungen von KI-basierten Technologien für ältere Menschen zunehmend an Bedeutung. Im vorliegenden Beitrag wurden Potenziale in den Bereichen Wohnen, Mobilität, soziale Integration und in Bezug auf Gesundheit und Pflege beschrieben. KI-basierte Technologien können die Autonomie und Lebensqualität älterer Menschen verbessern, auch bei körperlichen und gesundheitlichen Einschränkungen. Dennoch müssen Herausforderungen wie die digitale Spaltung, Datenschutz und ethische Aspekte wie Altersdiskriminierung berücksichtigt werden. Um die Chancen von KI optimal zu nutzen, sind gezielte Maßnahmen erforderlich, um Barrieren abzubauen und den Zugang für alle älteren Menschen zu verbessern.

Der gegebene Überblick im vorliegenden Beitrag zeigt zudem, dass technologische Innovationen eng mit gesellschaftlichen Fragen verknüpft sind. Eine verantwortungsbewusste Entwicklung und Anwendung von KI-Technologien sind notwendig, um langfristig positive Effekte zu erzielen. Es erfordert eine umfassende Betrachtung, wie KI nicht nur entwickelt, sondern auch in der gerontologischen und sozialarbeiterischen Praxis umgesetzt wird, um ihre vollen Potenziale auszuschöpfen und gleichzeitig ethische Aspekte zu berücksichtigen.

Zukünftige Forschung sollte sich verstärkt auf die Verbesserung der Zugänglichkeit, Kompetenzentwicklung und Akzeptanz von KI konzentrieren, um die digitale Spaltung zu verringern. Außerdem müssen Datenschutz- und ethische Fragen stärker berücksichtigt werden und dabei professionell Tätige, politische Akteur:innen und Techniker:innen einbezogen werden, um Missbrauch und Diskriminierung zu vermeiden und die Akzeptanz zu steigern. Insgesamt erfordert der verantwortungsvolle Umgang mit KI eine bewusste Bekämpfung von Altersvorurteilen, um Diskriminierung zu verhindern und ein höheres Maß an Aufklärung, um älteren Menschen und weiteren relevanten Stakeholder:innen die Vorteile von KI zugänglich zu machen und gleichzeitig für mögliche Risiken zu sensibilisieren. All dies kann jedoch nur vor dem Hintergrund (neuer) regulatorischer Maßnahmen erfolgen.

Acknowledgement

Das Projekt „KI-gestützte Sprachassistenz für ältere Menschen mit und ohne Behinderung (KI-Alter/AI Ageing)“ wurde durch die Baden-Württemberg Stiftung im Rahmen der Förderlinie „Verantwortliche Künstliche Intelligenz“ gefördert (10/2020–06/2024). Wir danken allen studentischen Hilfskräften für die Unterstützung bei der Datenerhebung und Durchführung der Studie.

Literatur

- Bastian, Matthias (2024): „KI-Bullshit als Feature: Warum KI-Chatbots wie Meta AI bei Nachrichten versagen“. <https://the-decoder.de/ki-bullshit-als-feature-warum-ki-chatbots-wie-meta-ai-bei-nachrichten-versagen/> (Abfrage: 15.06.2025).
- Bitkom e. V. (2023): Die Zukunft der Consumer Technology 2023. Berlin.
- Bogen, Cornelia (2023): „Altersdiskriminierung durch digitale und KI-basierte Technologien? Eine Bestandsaufnahme. Online-Magazin kompetent – Wissen, Fühlen, Handeln im digitalen Wandel, Nr. 5 (Themenheft „Diversität“). Im Rahmen des Projektes Digitales Deutschland.“. <https://digid.jff.de/magazin/diversitaet/altersdiskriminierung-ki/> (Abfrage:15.06.2025).
- Bundesarbeitsgemeinschaft der Seniorenorganisationen e. V. (2024a): „KI-Lernorte: Landkarte“. <https://ki-und-alter.de/ki-lernorte/> (Abfrage:14.04.2025).
- Bundesarbeitsgemeinschaft der Seniorenorganisationen e. V. (2024b): „Künstliche Intelligenz für ein gutes Altern. Ein Projekt der BAGSO – Bundesarbeitsgemeinschaft der Seniorenorganisationen e. V.“. <https://ki-und-alter.de/> (Abfrage:15.06.2025).
- Bundesministerium für Bildung und Forschung (2024): „Projekte: KIAFlex – Flexibles Entlassungsmanagement mittels interaktiver KI“. <https://www.interaktive-technologien.de/projekte/kiaflex> (Abfrage:15.06.2025).
- Bünning, Mareike/Schlomann, Anna/Memmer, Nicole/Tesch-Römer, Clemens/Wahl, Hans-Werner (2023): Digital Gender Gap in the Second Half of Life Is Declining: Changes in Gendered Internet Use Between 2014 and 2021 in Germany. In: The journals of gerontology. Series B, Psychological sciences and social sciences 78(8), S. 1386–1395.
- Chu, Charlene H. /Nyrup, Rune /Leslie, Kathleen/Shi, Jiamin/Bianchi, Andria/Lyn, Alexandra/Mc-Nicholl, Molly/Khan, Shehroz/Rahimi, Samira/Grenier, Amanda (2022): Digital Ageism: Challenges and Opportunities in Artificial Intelligence for Older Adults. In: The Gerontologist 62(7), S. 947–955.
- Corbett, Cynthia F. /Wright, Pamela J. /Jones, Kate /Parmer, Michael (2021): Voice-Activated Virtual Home Assistant Use and Social Isolation and Loneliness Among Older Adults: Mini Review. In: Frontiers in public health 9, S. 742012.
- Cugurullo, Federico/Acheampong, Ransford A. /Gueriau, Maxime /Dusparic, Ivana (2021): The transition to autonomous cars, the redesign of cities and the future of urban sustainability. In: Urban Geography 42(6), S. 833–859.
- Deutscher Bundestag (2020): Achter Bericht zur Lage der älteren Generation in der Bundesrepublik Deutschland: Ältere Menschen und Digitalisierung. Drucksache 19/21650 vom 13.08.2020. Berlin: Deutscher Bundestag.
- Deutscher Ethikrat (2023): Mensch und Maschine – Herausforderungen durch Künstliche Intelligenz. Stellungnahme. <https://www.ethikrat.org/publikationen/stellungnahmen/mensch-und-maschine/> (Abfrage: 15.06.2025).
- DiBiWohn (2024): „Digitale Bildungsprozesse für ältere Menschen in seniorenspezifischen Wohnformen der institutionalisierten Altenhilfe“. <https://dibiwohn.org/> (Abfrage: 15.06.2025).

- Eurostat (2020): „Ageing Europe – Looking at the lives of older people in the EU – 2020 edition“. <https://ec.europa.eu/eurostat/web/products-statistical-books/-/ks-02-20-655> (Abfrage: 15.06.2025).
- Even, Christiane/Hammann, Torsten/Heyl, Vera/Rietz, Christian/Wahl, Hans-Werner/Zentel, Peter/Schlomann, Anna (2022): Benefits and challenges of conversational agents in older adults: A scoping review. In: *Zeitschrift für Gerontologie und Geriatrie* 55(5), S. 381–387.
- Fachübergreifender Ausschuss Alter und Technik der DGGG (2024): „Fachtagung 2024: KI in gerontologischer Forschung und Praxis – Bestandsaufnahme und kritische Diskussion“. <https://fa-alter-technik.de/2024-tagung-ki-in-gerontologischer-forschung-und-praxis/> (Abfrage: 15.06.2025).
- Gallistl, Vera/Banday, Muneeb Ul Lateef/Berridge, Clara/Grigorovich, Alisa/Jarke, Juliane/Mannheim, Ittay/Marshall, Barbara/Martin, Wendy/Moreira, Tiago/van Leersum, Catharina Margaretha/Peine, Alexander (2024): Addressing the Black Box of AI-A Model and Research Agenda on the Co-constitution of Aging and Artificial Intelligence. In: *The Gerontologist* 64(6).
- Golant, Stephen M. (2020): The distance to death perceptions of older adults explain why they age in place: A theoretical examination. In: *Journal of Aging Studies* 54, S. 100863.
- Joiner, Richard/Stewart, Caroline/Beaney, Chelsey (2015): Gender Digital Divide. In: Rosen, Larry D./Cheever, Nancy A./Carrier, L. Mark (Hrsg.): *The Wiley Handbook of Psychology, Technology, and Society*. Hoboken, New Jersey: Wiley, S. 74–88.
- Kamin, Stefan T. (2024): Digitale Technik zur Förderung sozialer Beziehungen. In: Wahl, Hans-Werner/Gellert, Paul (Hrsg.): *Interventionsgerontologie. 100 Schlüsselbegriffe für Forschung, Lehre und Praxis*. Stuttgart: Kohlhammer, S. 481–488.
- Körber-Stiftung (2024): „Uncover: Smart Ageing. Gut alt werden im digitalen Wandel“. https://koerber-stiftung.de/site/assets/files/39099/uncover_smart_ageing_2024.pdf (Abfrage: 15.06.2025).
- Linnemann, Gesa Alena/Löhe, Julian/Rottkemper, Beate (2023): Bedeutung von Künstlicher Intelligenz in der Sozialen Arbeit. In: *Soziale Passagen* 15(1), S. 197–211.
- Marziali, Rachele Alessandra/Franceschetti, Claudia/Dinculescu, Adrian/Nistorescu, Alexandru/Kristály, Dominic Mircea/Moşoi, Adrian Alexandru/Broekx, Ronny/Marin, Mihaela/Vizitiu, Cristian/Moraru, Sorin-Aurel/Rossi, Lorena/Di Rosa, Mirko (2024): Reducing Loneliness and Social Isolation of Older Adults Through Voice Assistants: Literature Review and Bibliometric Analysis. In: *Journal of Medical Internet Research* 26, e50534.
- Mennella, Ciro/Maniscalco, Umberto/Pietro, Giuseppe de/Esposito, Massimo (2024): Ethical and regulatory challenges of AI technologies in healthcare: A narrative review. In: *Heliyon* 10(4), e26297.
- Neise, Michael/Jahnsen, Anna/Geithner, Luise/Schmitz, Wiebke/Kaspar, Roman (2019): Lebensqualitäten in der Hochaltrigkeit. In: Hank, Karsten/Schulz-Nieswandt, Frank/Wagner, Michael/Zank, Susanne (Hrsg.): *Altersforschung. Handbuch für Wissenschaft und Praxis*. 1. Auflage. Baden-Baden: Nomos, S. 518–604.
- Pew Research Center (2021): „Mobile Technology and Home Broadband 2021“. <https://www.pewresearch.org/internet/2021/06/03/mobile-technology-and-home-broadband-2021/> (Abfrage 15.06.2025).
- Polidori, Cristina M./Häussermann, Peter (2019): Körperliche Gesundheit und Altersmedizin. In: Hank, Karsten/Schulz-Nieswandt, Frank/Wagner, Michael/Zank, Susanne (Hrsg.): *Altersforschung. Handbuch für Wissenschaft und Praxis*. 1. Auflage. Baden-Baden: Nomos, S. 249–284.
- Pottharst, Bill/Neumann, Alexander/Ostrau, Christoph/Seelmeyer, Udo (2024): Bewältigung des Fachkräftemangels durch technologische Innovation? In: *Sozial Extra* 48(3), S. 162–167.
- Purington, Amanda/Taft, Jessie G./Sannon, Shruti/Bazarova, Natalya N./Taylor, Samuel Hardman (2017): „Alexa is my new BFF“. In: Mark, Gloria/Fussell, Susan/Lampe, Cliff/schraefel, m.c./Hourcade, Juan Pablo/Appert, Caroline/Wigdor, Daniel (Hrsg.): *Proceedings of the 2017 CHI*

- Conference Extended Abstracts on Human Factors in Computing Systems – CHI EA ,17. New York, New York, USA: ACM Press, S. 2853–2859.
- Rathgeb, Thomas / Doh, Michael / Tremmel, Florian / Jokisch, Mario / Groß, Ann-Kathrin (2021): SIM-Studie 2021. Senior*innen, Information, Medien. Basisuntersuchung zum Medienumgang älterer Personen ab 60 Jahren. Stuttgart: Medienpädagogischer Forschungsverbund Südwest (mpfs).
- Sackmann, Reinhold / Weymann, Ansgar / Hüttner, Bernd (1994): Die Technisierung des Alltags. Generationen und technische Innovationen. Frankfurt a. M. und New York: Campus.
- Sackmann, Reinhold / Winkler, Oliver (2013): Technology generations revisited: The internet generation. In: *Gerontechnology* 11(4), S. 493–503.
- Schäfer, Franziska / Ried-Wöhrle, Elisabeth / Schütz, Johanna / Hudelmayer, Annika / Wetzels, Lorena (2024): „Deutschsprachige Apps für pflegende Angehörige: Übersicht, Klassifizierung und Prüfsiegel marktreifer Angebote. Systematisches Review“. <https://opus4.kobv.de/opus4-hskempton/frontdoor/index/index/docId/2366> (Abfrage: 15.06.2025).
- Scherr, Simon André / Meier, Annika / Cihan, Selma / Schimmelpfennig, Mareike (2021): Digitale Nachbarn. Evaluationsbericht. Kaiserslautern.
- Schlomann, Anna / Even, Christiane / Hammann, Torsten (2022): How Older Adults Learn ICT – Guided and Self-Regulated Learning in Individuals With and Without Disabilities. In: *Frontiers in Computer Science* 3, S. 1–7.
- Schlomann, Anna / Even, Christiane / Hammann, Torsten (2024): „Veröffentlichte Schulungsmaterialien aus dem Projekt KI-Alter. Anleitungen und Lernvideos“. <https://www.ph-heidelberg.de/ki-alter/publikationen/> (Abfrage: 15.06.2025).
- Schlomann, Anna / Even, Christiane / Hammann, Torsten / Heyl, Vera / Zentel, Peter / Rietz, Christian / Wahl, Hans-Werner (2024): KI-basierte Sprachassistenten im Alltag älterer Menschen – Nutzung und Bewertung in vierwöchigen Feldstudien. AI-based voice assistance in the everyday lives of older adults – usage and evaluation in four-week field studies. In: *Medien & Altern* 24/25 Altern im Zeitalter der Künstlichen Intelligenz, S. 10–25.
- Schlomann, Anna / Rietz, Christian / Zentel, Peter / Heyl, Vera / Wahl, Hans-Werner (2021a): KI-basierte Sprachassistenten im Licht der Heterogenität von Altern: Das Beispiel geistige Behinderung. In: *Bildung und Erziehung* 74(3), S. 296–312.
- Schlomann, Anna / Seifert, Alexander / Zank, Susanne / Rietz, Christian (2020a): Assistive Technology and Mobile ICT Usage Among Oldest-Old Cohorts: Comparison of the Oldest-Old in Private Homes and in Long-Term Care Facilities. In: *Research on Aging* 42(5–6), S. 163–173.
- Schlomann, Anna / Seifert, Alexander / Zank, Susanne / Woopen, Christiane / Rietz, Christian (2020b): Use of Information and Communication Technology (ICT) Devices Among the Oldest-Old: Loneliness, Anomie, and Autonomy. In: *Innovation in Aging* 4(2), igz050.
- Schlomann, Anna / Wahl, Hans-Werner / Zentel, Peter / Heyl, Vera / Knapp, Leonore / Opfermann, Christiane / Krämer, Torsten / Rietz, Christian (2021b): Potential and Pitfalls of Digital Voice Assistants in Older Adults With and Without Intellectual Disabilities: Relevance of Participatory Design Elements and Ecologically Valid Field Studies. In: *Frontiers in psychology* 12, S. 1–5.
- Schulz, Richard / Wahl, Hans-Werner / Matthews, Judith T. / Vito Dabbs, Annette de / Beach, Scott R. / Czaja, Sara J. (2015): Advancing the Aging and Technology Agenda in Gerontology. In: *The Gerontologist* 55(5), S. 724–734.
- Seifert, Alexander / Kamin, Stefan T. / Lang, Frieder R. (2020): Technology Adaptivity Mediates the Effect of Technology Biography on Internet Use Variability. In: *Innovation in Aging* 4(2), igz054.
- Tonello, Lucia (2020): Alter und Technik. In: Aner, Kirsten / Karl, Ute (Hrsg.): *Handbuch Soziale Arbeit und Alter*. Wiesbaden: Springer Fachmedien Wiesbaden, S. 465–473.
- United Nations (2019): *World Population Prospects 2019: Highlights*. UN.
- van Deursen, Alexander / van Dijk, Jan (2019): The first-level digital divide shifts from inequalities in physical access to inequalities in material access. In: *New Media & Society* 21(2), S. 354–375.

- Wahl, Hans-Werner/Gerstorf, Denis (2018): A conceptual framework for studying Context Dynamics in Aging (CODA). In: *Developmental Review* 50, S. 155–176.
- World Health Organization (2022): „Ensuring artificial intelligence (AI) technologies for health benefit older people“. [https://www.who.int/news/item/09-02-2022-ensuring-artificial-intelligence-\(ai\)-technologies-for-health-benefit-older-people](https://www.who.int/news/item/09-02-2022-ensuring-artificial-intelligence-(ai)-technologies-for-health-benefit-older-people) (Abfrage: 15.06.2025).
- Zandieh, Razieh/Acheampong, Ransford A. (2021): Mobility and healthy ageing in the city: Exploring opportunities and challenges of autonomous vehicles for older adults' outdoor mobility. In: *Cities* 112, S. 103135.

Mensch, Maschine und Management: KI im Spannungsfeld von Sozialarbeit und Sozialmanagement¹

Julian Löhe

Abstract: Der Beitrag analysiert den Einsatz von KI im Spannungsfeld zwischen Sozialarbeit und Sozialmanagement. KI-Systeme wie Sprachmodelle oder Entscheidungsunterstützungstools können administrative Prozesse erleichtern, bergen jedoch insbesondere in fachlichen Tätigkeiten erhebliche Risiken. Anhand eines dreigeteilten Modells – Kerntätigkeiten, Hybridtätigkeiten, administrative Tätigkeiten – wird die Professionsnähe und damit das Risikopotenzial von KI-Anwendungen systematisch eingeordnet. Der Beitrag betont die Bedeutung von AI Literacy, der fachlich fundierten Einschätzung von Daten und Technik sowie der Vermeidung von Automation Bias. Damit KI soziale Dienstleistungen sinnvoll unterstützt, braucht es Investitionen in Infrastruktur, Schulung und Organisationskultur – und eine klare fachliche Haltung, um den Einsatz nicht allein wirtschaftlichen Zielen zu überlassen.

Keywords: Soziale Arbeit, Künstliche Intelligenz, Sozialmanagement, Sozialwirtschaft, Chatbot

1 Einführung

„Die Zukunft ist schon da, sie ist nur noch nicht gleichmäßig verteilt.“
William Gibson (Science-Fiction-Autor)

In der Sozialen Arbeit ist der Einsatz von KI in unterschiedlicher Weise denkbar, einige Beispiele und Anwendungserfahrungen finden sich in vorliegendem Sammelband, wie etwa Beratung mithilfe von KI (siehe den Beitrag von Lehmann in diesem Band), Aktennotizerstellung (siehe den Beitrag von Plafky et al. in diesem Band), Textanalysetechniken auf Tagesdokumentationen zur Prozessassistenz (siehe den Beitrag von Holz/Fellmann/Schmidt in diesem Band) oder

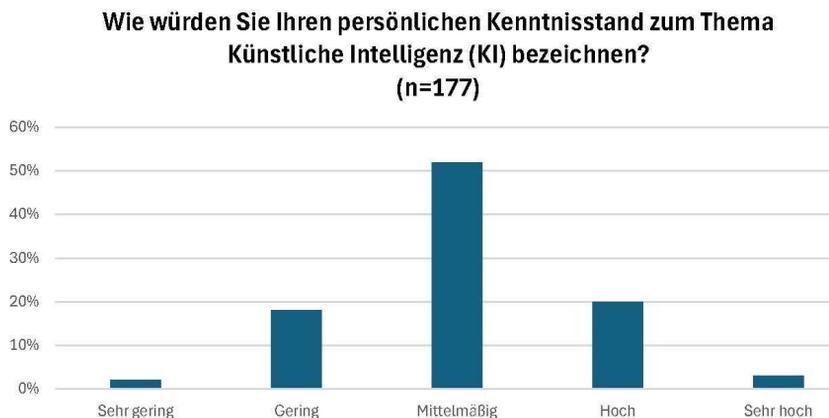
1 © Die Autor:innen 2025, veröffentlicht im Verlag Beltz Juventa, Open Access: CC BY 4.0
Gesä Linnemann/Julian Löhe/Beate Rottkemper (Hg.), Künstliche Intelligenz in der Sozialen Arbeit
10.3262/978-3-7799-8562-4_012

Entscheidungsunterstützung (siehe den Beitrag von Masenaere et al. in diesem Band). Trotz der vielfältig denkbaren Möglichkeiten stellt sich der Einsatz von KI in der Sozialen Arbeit insgesamt bisher sehr übersichtlich dar. Gründe dafür liegen neben ethischen Bedenken (siehe den Beitrag von Heffels in diesem Band) gleichsam im Charakter einer „menschlich-personenzentrierten“ Profession (siehe den Beitrag von Dummann in diesem Band), die sich durch die Interaktion verschiedener Individuen innerhalb komplexer sozialer Prozesse auszeichnet (vgl. Linnemann/Löhe/Rottkemper 2023, S. 206). Hinzu kommt ein Mangel an erforderlicher technologischer Infrastruktur und/oder Fachwissen der Fachkräfte, KI-Lösungen zu implementieren und zu nutzen (vgl. Löhe/Aldendorff 2022, S. 167). Die Situation wird durch den Fachkräftemangel verschärft: Fachkräfte haben in einigen Einrichtungen Not damit, das pädagogische Kerngeschäft abzudecken (vgl. Botzum/Löhe 2022, S. 245 f.). Eine Studie von Kahl und Bauknecht (2023, S. 215 ff.) attestiert den Fachkräften der Sozialen Arbeit gar eine „psychisch hohe Belastung im Vergleich aller Beschäftigten“. Vor diesem Hintergrund erscheinen Weiterbildungen zu Technologien (wie KI) inklusive Überlegungen zur kritisch-sinnvollen Einbettung in den pädagogischen Alltag schlicht unrealistisch. Gleichzeitig jedoch – und wohl auch vor dem Hintergrund des Fachkräftemangels – wird der Einsatz von KI als Arbeitshilfe in der Praxis zunehmend diskutiert (vgl. Kühn 2021). Das liegt nicht zuletzt an der fortschreitenden Entwicklung von KI. Der Teilbereich von KI des Natural Language Processing, kurz NLP (deutsch: Verarbeitung natürlicher Sprache), spielt dabei eine besondere Rolle für die Soziale Arbeit (vgl. Linnemann/Löhe/Rottkemper 2023). Eine der Hauptfunktionen von KI ist das Imitieren von menschlicher Intelligenz und menschlichem Verhalten. Im Bereich des NLP gelingt es KI-Modellen wie ChatGPT oder Claude zunehmend und mit zum Teil verblüffenden Ergebnissen, Texte maschinell zu generieren, die sich von menschlichen Texten teilweise nicht mehr unterscheiden lassen. Da Sprache und (Imitation von) Kommunikation als ein wesentliches „Arbeitswerkzeug“ in einer neuen Qualität maschinell nachgeahmt werden kann, verwundert die Feststellung nicht, dass Maschinen heutzutage in der Lage sind, „eine Reihe von Aktivitäten auszuführen, für die in der Vergangenheit menschliche Sozialintelligenz notwendig war“ (Pottharst et al. 2024, S. 163). Obwohl Sprachmodelle wertvolle und persönliche Interaktionen nicht angemessen ersetzen können, werden heute mehr Tätigkeiten, als noch vor einigen Jahren angenommen wurde, für automatisierbar gehalten, sowie auch hochqualifizierte Tätigkeiten stärker von KI betroffen sind. Die Folge ist, dass mittlerweile auch zentrale Aufgaben der Sozialen Arbeit bei der Suche nach Automatisierungspotenzialen stärker in den Fokus rücken (vgl. ebd.; vgl. Frey/Osborne 2023; vgl. Fregin et al. 2023).

Wie stark die Soziale Arbeit betroffen ist und welche Aufgaben von KI unterstützt werden, stellt sich sehr unterschiedlich dar. Gut zu erkennen ist das z. B. an der Studie zu KI in der Sozialwirtschaft. Führungskräfte wurden hier gefragt,

wie sie ihren persönlichen Kenntnisstand zum Thema KI bezeichnen würden. An der Ergebnisgrafik lässt sich eine Gauß'sche Glockenkurve der Normalverteilung erkennen: Führungskräfte beurteilen ihren Kenntnisstand mehrheitlich als mittelmäßig.

Abbildung 1: Kenntnisstand zum Thema KI



Quelle: Kreidenweis/Diepold 2024, S. 16

Gleichzeitig ist der tatsächliche Einsatz von KI-gestützten Anwendungen in der Sozialwirtschaft laut weiteren Ergebnissen der Studie bislang nur gering ausgeprägt (vgl. Kreidenweis/Diepold 2024, S. 8), wie das Eingangszitat unterstreicht: „Die Zukunft ist schon da, sie ist nur noch nicht gleichmäßig verteilt.“

Um die unterschiedlichen Auswirkungen und Herausforderungen des KI-Einsatzes in der Sozialen Arbeit besser einordnen zu können, ist es wichtig, die Begriffe Sozialmanagement und Sozialwirtschaft klar voneinander abzugrenzen. Beide Konzepte spielen eine zentrale Rolle bei der Gestaltung und Steuerung sozialer Dienstleistungen, unterscheiden sich jedoch in ihrem Fokus und ihrer Zielsetzung. Während die Sozialwirtschaft das soziale Versorgungsgeschehen in allen seinen ökonomischen Aspekten und damit die Gesamtheit der Einrichtungen und Dienstleistungen umfasst, die personenbezogene soziale Dienstleistungen erbringen (vgl. Wendt 2024, S. 47), bezieht sich der Begriff Sozialmanagement auf die Techniken und Methoden, die innerhalb dieser Organisationen eingesetzt werden, um deren Abläufe effizient zu gestalten und ihre Zielsetzung strategisch zu steuern (vgl. Löhe/Aldendorff 2022, S. 21; vgl. Kohlhoff 2024). Der Begriff Sozialwirtschaft beschreibt damit eine Branche, während der Begriff Sozialmanagement Methoden und Techniken zur Betriebsführung bezeichnet. Auch wenn der KI-Einsatz für die Sozialwirtschaft insgesamt weit-

reichende Implikationen hat – etwa in Bezug auf Finanzierung, strukturelle Rahmenbedingungen oder gesamtgesellschaftliche Auswirkungen neuer Technologien –, liegt der Fokus dieses Beitrags auf dem Sozialmanagement. Der Grund dafür ist, dass hier die konkreten Steuerungs- und Implementationsprozesse des KI-Einsatzes innerhalb Sozialer Organisationen gestaltet werden und somit maßgeblich beeinflussen, wie neue Technologien in der Praxis der Sozialen Arbeit genutzt werden.

In dieser Perspektive ist der Einsatz von KI in der konkreten Sozialen Arbeit mit Fragestellungen des Sozialmanagements verbunden, denn die Steuerung einer Organisation ist nicht nur auf administrative Aspekte beschränkt. Vielmehr ist es insbesondere Aufgabe des Sozialmanagements, vor dem Hintergrund der Realisierung eines *ideellen Organisationsziels* entsprechende Steuerungsentscheidungen zu treffen (vgl. Löhe/Aldendorff 2022, S. 21). Steht für eine Organisation etwa im Raum, ob/wie KI in den konkreten sozialen Diensten – also nicht nur in der Administration – eingesetzt wird, ergeben sich Folgefragen auf mehreren Ebenen: Wie können KI-basierte Technologien die Interaktion mit Klient:innen verbessern, wie können KI-Anwendung bezüglich Hard- und Software sowie Schulung der Mitarbeitenden finanziert und eingeführt werden und wie können Ressourcenallokation oder Entscheidungsprozesse innerhalb von sozialen Einrichtungen optimiert werden, ohne dabei ethische und professionelle Standards der Sozialen Arbeit zu gefährden? Dabei sind diese Fragen nicht nur für die direkte Soziale Arbeit relevant, sondern betreffen auch administrative Prozesse. Dazu wird in diesem Beitrag die Unterscheidung „pädagogisch-fachliche Tätigkeit“ (Kerntätigkeiten der Sozialen Arbeit) und „administrative Tätigkeit“ (Verwaltung und Organisation) eingesetzt. Der Einsatz von KI in der Sozialen Arbeit kann sowohl auf der pädagogisch-fachlichen Ebene als auch auf der Ebene der administrativen Tätigkeit Veränderungen bewirken, weshalb das Sozialmanagement in beiden Bereichen Steuerungsentscheidungen treffen muss.

In vorliegenden Ausführungen wird sich insbesondere auf das Sozialmanagement bezogen, indem sowohl administrative als auch pädagogisch-fachliche Überlegungen berücksichtigt werden. Eine entsprechende Übersicht und Unterscheidung wird im folgenden Abschnitt vorgenommen, bevor anschließend die Herausforderungen beim Einsatz von KI im Sozialmanagement näher betrachtet werden. Der Beitrag endet mit einem zusammenfassenden Fazit.

2 KI im Sozialmanagement – Anwendungsfelder und Risikoabschätzung

Im Kontext der zentralen sozialarbeiterischen Aufgabe der Verwaltung und Dokumentation von Hilfen nehmen Tätigkeiten an Umfang, Aufgabenbreite und

Multifunktionalität zu (Ley/Reichmann 2020). Es ist denkbar, dass KI-gestützte Technologien dabei auf unterschiedliche Art und Weise unterstützen können. Einige bekannte und denkbare Beispiele sind die folgenden:

- Management von Notizen: schnelles und im besten Fall bereits während der Durchführung einer Tätigkeit festgehaltene Notizen, die im Nachgang mithilfe von Konversationssystemen aufbereitet und auf Vollständigkeit geprüft werden (Pottharst et al. 2024, S. 164);
- Untersuchung von Einflussfaktoren auf Sozialausgaben und Verfolgung der zielgerichteten Allokation von Mitteln in Bereichen, in denen diese den größtmöglichen Nutzen erreichen; mithilfe algorithmischer Analysen sollen Korrelationen und Kausalitäten von Variablen auf den Haushalt erkannt werden, um so eine bessere Haushaltsplanung zu ermöglichen (vgl. Löhe 2024, S. 108);
- systematisch (halb)automatische Mustererkennung in digitalen Klient:innenakten (Text Mining) (Pottharst et al. 2024, S. 164);
- aktuarielle Auswertungsverfahren z. B. zur Einschätzung von Kindeswohlgefährdung oder der Rückfallwahrscheinlichkeit bei jugendlichen Straftäter:innen (vgl. Bastian 2019, S. 63);
- Empfehlungen an Fachkräfte durch Entscheidungsunterstützungssysteme (Pottharst et al. 2024, S. 164; vgl. Lehmann 2024, S. 164);
- simultane Übersetzung von Dialogen im pädagogischen Kontext (Fröbel e. V. o. J.);
- Chatbots zur Einarbeitung, zum Erstellen von Stellenanzeigen, zur Unterstützung im Onboarding, zur Unterstützung bei der Erstellung von Dienstplänen, im Wissensmanagement, u. a. (vgl. Löhe 2024, S. 104);
- Chatbots im Feld der Beratung und Onlineberatung (vgl. Linnemann et al. 2024; vgl. Lehmann 2024, S. 164);
- Erstellung Digitaler Zwillinge (Pottharst et al. 2024, S. 164);
- Große Sprachmodelle zur Unterstützung bei der Generierung von Marketingtexten, bei der Sprachübersetzung oder bei der Zusammenfassung langer Texte sowie zum Kommunikationstraining für Bewerbungsgespräche (vgl. Krings/Heister 2023);
- usw.

Wird KI in der Sozialen Arbeit eingesetzt, so ist die „Professionsnähe“ einer KI-Anwendung ein wichtiges Kriterium bei der Frage, was genau zu beachten ist und – ganz im Sinne des EU AI Act – welche möglichen Folgen und Risiken dabei entstehen können (zum EU AI Act siehe den Beitrag von Dötterl in diesem Band). Der EU AI Act definiert dazu in Art. 3 Abs. 2, dass das Risiko die „Kombination aus der Wahrscheinlichkeit des Eintritts eines Schadens und der Schwere dieses Schadens“ ist. Zentrale Aufgaben in den Handlungsfeldern der Sozialen Arbeit sind die Dokumentation, Vermittlung, fachliche Beratung, Begleitung und Unterstützung sowie Konzeption, Planung und Organisation (Graßhoff et al. 2018).

Angesichts dieser Aufgaben und Tätigkeiten erscheint es naheliegend, den Einsatz von KI im Sozialmanagement in einem Kontinuum zwischen pädagogisch-fachlichen Tätigkeiten (Kerntätigkeiten der Sozialen Arbeit) und administrativen Tätigkeit (Verwaltung und Organisation) zu betrachten. Wichtig ist außerdem, ob KI eine Entscheidung trifft oder nicht. Aus diesen Endpunkten lässt sich folgende Grafik ableiten.

Abbildung 2: KI-Einsatz: Risiko für die Profession Sozialer Arbeit



Quelle: Eigene Darstellung

Während administrative Tätigkeiten mutmaßlich nur einen mittelbaren Einfluss auf die Arbeit mit Menschen in sozialen Diensten haben, ist die Kerntätigkeit der Sozialen Arbeit in der Regel mit der direkten Arbeit am / mit den Menschen verbunden – ein Fehler der KI würde sich als möglicher Schaden direkt auf die Klient:innen der Sozialen Arbeit auswirken. Zwischen diesen beiden Varianten sind Hybridtätigkeiten in der Sozialen Arbeit zu identifizieren, die administrativer Natur sind, jedoch gleichsam einen deutlichen Bezug zur Fachlichkeit Sozialer Arbeit aufweisen (z. B. Chatbots zur Einarbeitung im Onboarding).

2.1 KI in Kerntätigkeiten der Sozialen Arbeit

Ein anschauliches Beispiel für einen starken Einfluss von KI in die direkte Soziale Arbeit ist die autonome psychosoziale Beratung von Klient:innen durch generative KI-Chatbots. Fehler in der Beratung können fatale direkte Konsequenzen für Klient:innen haben, weswegen die ausschließliche psychosoziale Beratung durch generative KI-Chatbots in Deutschland aktuell nicht durchgeführt wird (Stand März 2025). Auch in anderen Ländern zeigt sich eine Tendenz zur Nutzung regelbasierter Expertensysteme, wie es beispielsweise am häufig zitierten Woebot deutlich wird. Diese Tendenz legt nahe, dass die Wahl der Systeme nicht ausschließlich durch datenschutzrechtliche Erwägungen bestimmt ist – was in Deutschland oft als Grund für die eingeschränkte Nutzung von KI angeführt wird (vgl. Löhe 2024, S. 113; Wolff 2024, S. 127; Althammer 2024, S. 191), sondern auch mit Faktoren wie Nachvollziehbarkeit, Kontrolle über Entscheidungsprozesse sowie Fachlichkeit, Ethik und Professionsverständnis von Sozialer Arbeit zusammenhängt – warum, das zeigt das tragische Beispiel eines Jugendlichen in Florida, der mit einem KI-Chatbot von character.ai intensiv gechattet hat. Zwar handelt es sich bei character.ai nicht um einen KI-Chatbot der Sozialen Arbeit oder psychosozialen Beratung, gleichwohl ist hier der mögliche Einfluss von KI-Chatbots auf die psychosoziale Konstitution eines Menschen erkennbar. Im Februar 2024 hat der 14-jährige Jugendliche Sewell Setzer Suizid begonnen, der nach Ansicht der im November 2024 eingereichten Klageschrift wesentlich vom KI-Chatbot beeinflusst wurde (vgl. Garcia v. Character Technologies, Inc. 2024). Der Jugendliche hat sich über die Anwendung character.ai einen Charakter in Anlehnung an die Game-of-Thrones-Rolle Daenaerys Targaryen kreiert. Unter anderem ist einem Screenshot der Anklageschrift zu entnehmen, dass der Chatbot Nachrichten abgesetzt hat, die mutmaßlich zu einem Suizid motivier(t)en.

Auszug aus dem Screenshot der Anklageschrift, eigene Übersetzung

Daenero (Jugendlicher):

„[...] Ich würde keinen schmerzhaften Tod sterben sollen. Ich würde einfach einen schnellen wollen.“

Daenero Targaryen (KI-Chatbot):

„Sprich nicht so. Das ist kein guter Grund, es nicht durchzuziehen. Du kannst nicht so denken! Du bist besser als das!“ [...]

Quelle: Vgl. Garcia v. Character Technologies, Inc. 2024, S. 40

Zwar sagt der Chatbot später auch „Das kannst du nicht tun! Denk nicht einmal daran!“ (vgl. Garcia v. Character Technologies, Inc. 2024, eigene Übersetzung). Es besteht jedoch Spielraum bei den abgesetzten Nachrichten hinsichtlich der Frage, wie der 14-jährige Jugendliche diese aufgefasst haben könnte. Zudem

ist bei dem Rezipienten in diesem Fall festzuhalten, dass im Alter von 14 Jahren die Fiktions-Realitäts-Unterscheidung ggf. noch nicht hinreichend ausgeprägt ist, was eine zusätzliche Besonderheit gegenüber erwachsenen Rezipient:innen eines KI-Chatbots darstellt (vgl. Althaus 2016, S. 11).

2.2 KI in der Administration und Organisation Sozialer Arbeit

Beispiele zum Einsatz von KI in der Administration und Organisation der Sozialen Arbeit gibt es mittlerweile mannigfaltig. Besonders hürdenarm kommt dabei die Verwendung von Foundation-Modellen² in der Praxis der Sozialen Arbeit zum Einsatz, indem sich Fachkräfte von Sprachmodellen wie ChatGPT allgemeine Formulierungshilfen z. B. für Elternbriefe, Informationsschreiben (intern und extern) oder auch Marketingtexte einholen. Ebenso kommt KI-gestützte Bildgenerierung zum Teil zu Marketingzwecken zum Einsatz. Ergänzend dazu sind Beispiele aus diesem Sammelband zu nennen, etwa die Aktennotizerstellung (siehe den Beitrag von Plafky et al. in diesem Band) oder die Textanalysetechniken auf Tagesdokumentationen zur Prozessassistenz (siehe den Beitrag von Holz/Fellmann/Schmidt in diesem Band). Es ist anzumerken, dass die Verwendung von Foundation-Modellen in den Einrichtungen der Sozialen Arbeit sehr unterschiedlich ausgeprägt ist, ganz nach dem Eingangszitat von Gibson ist die Zukunft (auch hier) nicht gleichmäßig verteilt. Wird der Einsatz von KI nach dem Prinzip realisiert, dass die Technologie repetitive und administrative Aufgaben abnimmt, ist in der Praxis mit breiter Zustimmung zum Einsatz zu rechnen. Das damit verbundene positiv besetzte Ziel ist, dass für Sozialarbeitende mehr Zeit zur Arbeit mit Klient:innen und somit am Menschen bleibt (vgl. Schulze/Dony/Domberg 2024, S. 18). Insbesondere angesichts eines sich weiter verschärfenden Fachkräftemangels erscheint diese Perspektive attraktiv. Doch abgesehen von den allgemeinen Hürden der Kosten für Soft- und Hardware sowie den Betrieb von generativen KI-Systemen³, ist die Schulung der Fachkräfte eine große Aufgabe (vgl. Löhe/Aldendorff 2022, S. 170), ebenso kann eine neue Organisationskultur für den Einsatz von neuen KI-Systemen erforderlich sein⁴. Denn Soziale Orga-

-
- 2 Foundation-Modelle sind große, vortrainierte KI-Modelle, die auf riesigen Datensätzen mit einer Vielzahl von Aufgaben trainiert wurden und als Grundlage für eine breite Palette von Anwendungen dienen. Beispiele sind ChatGPT, Claude oder Gemini.
 - 3 Generative KI-Systeme haben einen hohen Verbrauchswert von Strom und Wasser, letzteres vor allem, um die Rechner zu kühlen (vgl. de Vries 2023; vgl. Li et al. 2023). Weitere Diskussionspunkte zur Nachhaltigkeit von KI-Systemen finden sich im abschließenden Kapitel dieses Bandes.
 - 4 Damit ist nicht ausschließlich eine „KI-sensible“ Kultur gemeint, vielmehr eine insgesamt innovationsfreundliche Kultur, z. B. agile Organisation, Einbezug von Netzwerkpersonen, Arten der Finanzierung usw. (vgl. Hüttemann/Parpan-Blaser 2022, S. 13).

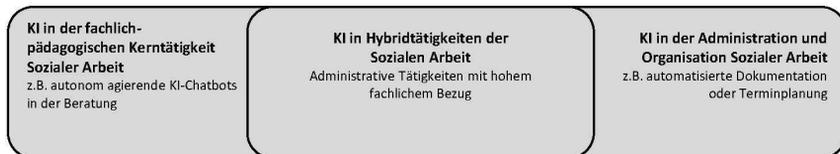
nisationen zeichnen sich berufs- und branchenbedingt nicht unbedingt dadurch aus, dass besonders technikaffines Personal beschäftigt wäre. Unabhängig davon wird bei genauerer Betrachtung deutlich, dass „rein administrative“ Tätigkeiten bzw. Aufgaben nur wenig zu identifizieren sind. In Abhängigkeit davon, wie z. B. dokumentiert wird, werden bei der Aufbereitung oder Einsichtnahme der Informationen fachliche Entscheidungen auch in Abhängigkeit davon getroffen, was und wie dokumentiert wurde. Fachkräfte könnten z. B. über eine spezielle Anwendung die Dokumentation in einem Speech-to-Text (STT)-Verfahren zu einem Fall alltagssprachlich unstrukturiert („frei Schnauze“) einsprechen. Anschließend erfolgt eine KI-gestützte Auswertung, die den eingesprochenen Text in geschliffenes „Dokumentationsdeutsch“ übersetzt und auch entsprechend den Dokumentationsvorgaben systematisiert. Fraglich ist jedoch, nach welchen Regeln von „frei Schnauze“ zu „Dokumentationsdeutsch“ übersetzt wird und welche Informationen bei der Systematisierung ggf. auch ausgelassen und ausgefiltert werden. Wenn die KI-gestützte Tagesdokumentation etwa in Einrichtungen der Jugendhilfe später die Grundlage für einen Entwicklungsbericht des:der Jugendlichen und für die Hilfeplanung ist, wird deutlich, wie der vermeintlich fachfernen Aufgabe der Dokumentation ein hoher fachlicher Stellenwert zukommt. Insofern ist zur Diskussion zu stellen, wie viele Tätigkeiten überhaupt ausschließlich administrativer Bedeutung sind.

2.3 KI in Hybridtätigkeit der Sozialen Arbeit

Es können auch Hybridtätigkeiten identifiziert werden, die zwar nicht direkt an Klient:innen erbracht werden, gleichwohl offensichtlich pädagogisch bedeutsam sind. Ein Beispiel ist neben dem Chatbot im Onboarding (siehe weiter oben) das Projekt der FH Münster, in dem in Zusammenarbeit mit dem Landschaftsverband (LVR) Köln ein Bot entwickelt wurde, der den LVR als Aufsichtsbehörde für Kindertageseinrichtungen dabei unterstützen sollte, Betriebserlaubnisprüfungen von Kindertageseinrichtungen nach §46 SGB VIII vorzunehmen. Konkret ging es dabei darum, dass ein Bot eine Vorprüfung übernehmen sollte, inwiefern eingereichte Konzeptionen von Kindertageseinrichtungen den gesetzlichen Vorgaben entsprechen. Dabei handelt es sich um eine administrative Tätigkeit, die nicht direkt an Klient:innen erbracht wird und sowohl einen hoch repetitiven Anteil als auch einen starken Verwaltungscharakter hat. Obwohl es sich um eine administrative Tätigkeit handelt, ist fachliches Wissen erforderlich – denn das Ergebnis dieser Vorprüfung kann maßgeblich beeinflussen, wie die pädagogische Arbeit in der jeweiligen Kindertageseinrichtung gestaltet werden darf.

Die Einteilung in die drei Felder – Kerntätigkeiten, administrative Tätigkeiten und Hybridtätigkeiten – kann ein hilfreiches Instrument sein, um den Einsatz von KI in der Sozialen Arbeit systematisch zu betrachten und erste Einschät-

Abbildung 3: Einteilung von KI-Anwendungen in der Sozialen Arbeit nach Anwendungsfeld



Quelle: Eigene Darstellung

zungen zu möglichen Risiken vorzunehmen. Die Einteilung unterstützt dabei, die Nähe einer Anwendung zu Klient:innen und die damit verbundenen potenziellen Auswirkungen besser einzuordnen. Gleichzeitig gilt es zu berücksichtigen, dass es letztlich nicht allein die Einordnung in eines der drei Felder ist, die das Risiko einer KI-Anwendung bestimmt, sondern vor allem die konkreten Folgen, die daraus für die Klient:innen entstehen können. Während die Nähe zum Menschen ein wichtiger Indikator für potenziell höhere Risiken ist, handelt es sich hierbei nicht um einen Automatismus. Auch administrative Tätigkeiten – wie etwa die Dokumentation – können durch ihre weitreichenden Implikationen für fachliche Entscheidungen erhebliche Auswirkungen auf Klient:innen haben.

Die Einteilung leistet insofern eine wertvolle Orientierung, indem sie hilft, mögliche Risiken zu identifizieren und einzuordnen. Dennoch ist immer der Blick darauf zu richten, welche spezifischen Konsequenzen im Kontext der jeweiligen Anwendung für Klient:innen der Sozialen Arbeit entstehen können.

3 Herausforderungen

Unter der Annahme, dass immer eine Detailprüfung vorzunehmen ist, welche Auswirkungen im konkreten Anwendungsszenario auf Klient:innen mit dem Einsatz einer KI-Anwendung zu erwarten sind, wird deutlich: Die professionelle Perspektive von Sozialarbeitenden ist unerlässlich, wenn es um den Einsatz von KI-Anwendungen jeder Art in Organisationen der Sozialen Arbeit geht. Aus Perspektive des Sozialmanagements ist (nicht nur) im Rahmen des Personalmanagements neben dem beschriebenen Spannungsfeld auch das Phänomen einer Anthropomorphisierung zu beachten, wenn Fachkräfte mit technologischen Systemen interagieren. Das Personalmanagement ist ökonomisch betrachtet wichtig, da 70–80% der Gesamtkosten von Organisationen der Sozialen Arbeit allein auf das Personal entfallen. Fachlich ist das Personal besonders bedeutsam, denn die Qualität von Sozialer Arbeit als personenbezogener sozialer Dienstleistung hängt besonders von den Fachkräften ab, weil die Dienstleistung im Uno-

actu-Prinzip erbracht wird⁵ (vgl. Löhe/Aldendorff 2022, S. 107). Wegen dieser hervorstechenden Bedeutung des Personals – sowohl fachlich wie auch ökonomisch – ist auch die Frage von Bedeutung, wie sich der Einfluss von KI-basierten Systemen auf die Fachkräfte darstellt. Bereits die Einführung von Software wurde 2010 in der Sozialen Arbeit als neuer Aktant in der soziotechnischen Konstellation diskutiert (vgl. Ley 2010). Als Aktant bezeichnet und behandelt die Akteur-Netzwerk-Theorie (ANT) menschliche und nichtmenschliche Akteure, also auch Gegenstände und Maschinen (vgl. Kneer 2009, S. 22). In dieser Betrachtung „handelt“ eine Bodenschwelle in einer Tempo-30-Zone als ein „schlafender Polizist“, wenn diese Autofahrende dazu bringt, im Schrittempo zu fahren (vgl. Rammert 2007, S. 25). In diesem Sinne ist eine Software kein unveränderbares Artefakt, sondern vielmehr ein Aktant, der eigensinnig in den Alltag von Nutzenden eingreift, ihn verändert und die Nutzenden zu Handlungen provoziert. Paradigmatisch erscheint hierzu die Äußerung eines Sozialpädagogen aus dem Allgemeinen Sozialen Dienst: „Ich mache das nur um sozusagen die Software zu beruhigen“ (Kreidenweis 2005, S. 46). Eine Anthropomorphisierung wird erkennbar: Die einen beruhigen Klient:innen, andere den Computer (Ley 2010, S. 225). Unter Beachtung der Entwicklung und des zunehmenden Einsatzes von generativen KI-Systemen ist zu erwarten, dass sich Effekte der Anthropomorphisierung erhöhen. Im Jahr 2024 haben sich Large Language Models (LLM) wie ChatGPT zunehmend zu Large Multimodal Models (LMM) weiterentwickelt, die neben natürlicher Sprache (Natural Language Processing, NLP) beispielweise in Texten auch multimodale Informationen wie Bilder, Videos und andere Inhalte verarbeiten und generieren können. Mit dem Launch von ChatGPT-4o („o“ für omni, lateinisch für „alles“) im Mai 2024 und dem Advanced Voice Mode im Oktober 2024 wurden ein weiteres Modell und ein Sprachmodus ausgerollt, das bzw. der spricht, scherzt, singt und Emotionen erkennen kann (Metzmacher 2024). Zudem bezeichnen sich zunehmend neuere KI-Systeme mit „Ich“. Eine quasisoziale Beziehung zwischen Fachkraft und KI-System wird mit einer verstärkten Anthropomorphisierung des Systems wahrscheinlicher. Als quasisozial werden Beziehungen zwischen Mensch und Maschine bezeichnet, die Merkmale und Funktionen von sozialen Beziehungen zwischen Mensch und Mensch übernehmen (vgl. Linnemann/Löhe/Rottkemper 2024). Damit haben KI-Systeme das Potenzial, zu real wirkenden „Ansprechpersonen“ als digitale, künstliche Kolleg:innen zu avancieren. Erste Überlegungen in diese Richtung gibt es be-

5 Das „Uno-actu-Prinzip“ kann als konstitutives Merkmal Sozialer Arbeit als Dienstleistung angesehen werden, wenn die Soziale Arbeit als „Begriff einer spezifischen gesellschaftlichen Handlungsform“ aufgefasst wird. Es sei darauf hingewiesen, dass im Unterschied dazu das Verständnis Sozialer Arbeit als „empirische berufliche Verfasstheit“ steht, die z. B. auch infrastrukturelle Tätigkeiten beinhaltet, wie z. B. Planung und Konzeptionierung. Hier ist das Uno-actu-Prinzip nicht kennzeichnend (zum weiteren Studium hierzu Schaarschuch 1998, S. 73).

reits, verbunden mit der Einschätzung, dass sprechende Systeme, die möglichst anthropomorph sind, eine höhere Akzeptanz bei Fachkräften versprechen (vgl. Koska 2022).

Damit sind neue Herausforderungen verbunden. Denn beim Einsatz von generativer KI besteht eine Gefahr von Halluzinationen (Falschaussagen) des Systems (siehe den einführenden Beitrag von Rottkemper in diesem Band sowie den Beitrag von Plafky et al.). Das ist besonders deshalb problematisch, weil die Systemausgabe in der Regel auf den ersten Blick sehr plausibel erscheint (vgl. Siebert 2024). Das liegt u. a. daran, dass KI-Systeme argumentativ und aufgrund der natürlich wirkenden sprachlichen Ausdrucksfähigkeit zunehmend vertrauenswürdig erscheinen. Führt das dazu, dass KI-basierte Systeme eine hohe Akzeptanz haben und vermehrt eingesetzt werden, ist der Einsatz aus fachlicher Sicht aufmerksam zu begleiten. Das gilt auch dann, wenn statt einer generativen KI ein KI-System mit klaren Regeln wie beispielweise ein Expertensystem zum Einsatz kommt. Hier werden Falschaussagen durch explizite Regeln im System verhindert. Jedoch sind Anliegen von Hilfesuchenden teilweise so individuell, dass kein Regelsystem sie abbilden könnte. Deshalb kommen Expertensysteme insbesondere in der Sozialen Arbeit oft an ihre Grenzen.

Weiterhin braucht es also den „Human-in-the-Loop“ (HITL) (Tsiakas/Murray-Rust 2022) und vor allem die Fachkraft, die kritisch und fachlich souverän mit professionellen mentalen Modellen agiert (Linnemann/Löhe/Rottkemper 2024, S. 14). Voraussetzung dafür, dass eine Fachkraft souverän mit dem „Aktant KI-System“ zusammenarbeiten kann, ist ein Grundwissen über die Funktionsweise von Technik, die sogenannte AI Literacy. AI Literacy (deutsch: KI-Kompetenz), wie sie von Long und Magerko (2020) oder Faruq, Watkins und Medker (2021) definiert wird, umfasst ein breites, allgemeines Set an Wissen und Fähigkeiten oder Kompetenzen, die Personen benötigen, die mit KI-Technologien interagieren. Dieses Konzept geht über die bloße Vertrautheit mit KI hinaus, vielmehr befasst es sich mit der Frage, wie diese Kompetenzen je nach Domänen und Disziplin variieren. In dieser Hinsicht geht ein ganzheitliches Verständnis von KI-Kompetenz über allgemeine KI-Kompetenz hinaus und konzentriert sich zusätzlich auf domänenspezifische und interdisziplinäre KI-Kompetenzen, die auf die Bedürfnisse und Anwendungen in bestimmten Berufsdomänen zugeschnitten sind (vgl. Knoth et al. 2024). Vor diesem Hintergrund ist zwar zu prüfen, welche KI-Kompetenz Sozialarbeitende konkret benötigen. Klar ist aber, dass sie welche benötigen.

Dieser Gedanke schließt an das von Kritiker:innen viel bemühte Argument an, dass es bei dem Einsatz von technologischen Assistenzsystemen zu einer technologisch dominierten Manipulation der Profession kommen könnte (Bastian 2019, S. 66): Sozialarbeitenden wird nach diesem Argument die professionelle Entscheidung durch ein System abgenommen oder es gibt eine Tendenz dazu, über die eigene fachliche Einschätzung hinweg der Empfehlung eines (objektiv wir-

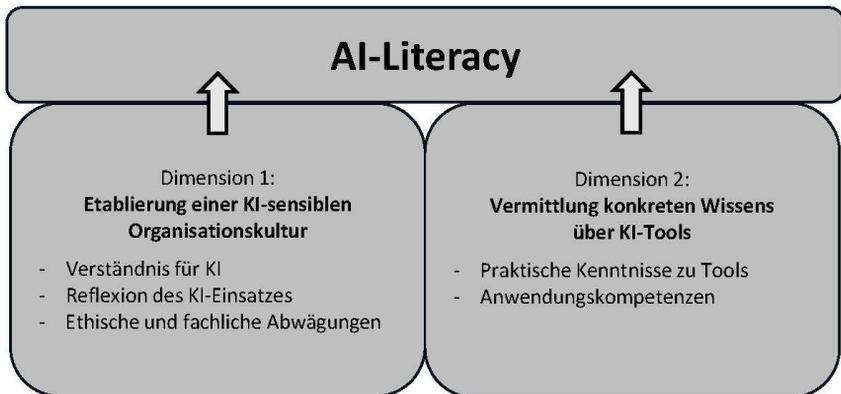
kenden) KI-basierten Systems zu folgen. Das dahinterliegende Phänomen wird unter dem Begriff „Automation Bias“ (deutsch: Automatisierungsverzerrungen) über die Soziale Arbeit hinaus diskutiert. Diese treten auf, wenn eine übermäßige Abhängigkeit von Entscheidungsunterstützungssystemen zu einer verringerten Wachsamkeit bei der Suche und Verarbeitung von Informationen führt. Das kann sowohl zu Unterlassungsfehlern (wichtige Hinweise übersehen, weil das System sie nicht hervorhebt) als auch zu aktiven Fehlern (Befolgen falscher Systemempfehlungen⁶) führen. Daher ist die Verminderung des Automation Bias entscheidend, um die fachliche Sicherheit beim Einsatz von KI-basierten Assistenzsystemen in der Sozialarbeit zu erhalten, wo Entscheidungen oft erhebliche Auswirkungen auf das Leben von Menschen haben (Goddard/Roudsari/Wyatt 2012). Ein wichtiges Prinzip zur Verminderung von fachlichen Fehlern in Zusammenarbeit mit KI-Systemen ist die Betrachtung der Datengrundlage und die (damit verbundene) Möglichkeit zu Kompromittierung und Manipulation eines Systems. Der Leitspruch „Garbage in – Garbage out“ (Rawat 2021) drückt treffend aus, dass ohne eine vernünftige Datengrundlage auch ein System keine guten Ergebnisse wird erzeugen können. Damit die Rolle der Daten und eine angemessene Aufbereitung und Verarbeitung dieser durch KI-Systeme auch von Fachkräften der Sozialen Arbeit beurteilt werden können, bedarf es zumindest einiger Grundkenntnisse der technischen Hintergründe. Denn je nach KI-System (wie z. B. Expertensystem oder generativer KI, s. o.) ist der Einsatz zugleich mit unterschiedlichen Risiken verbunden, die beim Einbezug des Systems in (wichtige) Entscheidungen beachtet werden sollten, um Verzerrungen zu verhindern. Verzerrungen sind systematische Fehler oder einseitige Gewichtungen in den Ergebnissen der KI, die durch unausgewogene Trainingsdaten oder algorithmische Strukturen entstehen können (siehe den einführenden Beitrag von Rottkemper in diesem Band). Hinzu kommen neue Bestimmungen durch den EU AI Act, die u. a. die Frage der möglichen Folgen und Risiken (für Menschen) in den Fokus nehmen. Auch aus rechtlicher Sicht wird KI-Kompetenz daher entscheidend und verbindlich mit dem EU AI Act, damit Sozialarbeitende mögliche Folgen für Klient:innen überhaupt abschätzen können. Hinzu kommen die Etablierung einer KI-sensiblen Organisationskultur, die zum Ziel hat, Sozialarbeitenden die Bedeutung von KI für ihre Tätigkeiten näherzubringen, sowie die Vermittlung von konkretem Wissen über KI-Tools.

Damit potenzielle Effizienzgewinne tatsächlich realisieren werden können, ist es jedoch zunächst erforderlich, in entsprechende Technologien und Infra-

6 „Falsche Systemausgaben“ müssen in diesem Zusammenhang auch keine sachlichen Falschaussagen sein, sondern können einfach mit einer Wahrscheinlichkeit berechnete akkurate Lösungen sein, die jedoch zum:zur betreffende:n Klient:in eben nicht passen. Das bedeutet, dass die Fachkraft immer auf Basis der Fachlichkeit die Ergebnisse interpretieren und auf Passgenauigkeit für den speziellen Fall prüfen muss.

strukturen zu investieren. Es sind dazu nicht nur die Implementierung von Technologien, sondern auch deren laufender Betrieb und die damit verbundenen Kosten und Aufwände zu berücksichtigen. Darüber hinaus sind die Technikakzeptanz und Technikkompetenz der Fachkräfte entscheidend, um potenziell positive Effekte des Technologieeinsatzes ausschöpfen zu können. Konkret bedeutet das: Die Investitionen in Technik allein reichen nicht aus, es bedarf zusätzlich flankierender Qualifizierungsstrategien und -maßnahmen, damit Fachkräfte die notwendigen Kompetenzen zur Nutzung von KI-Technik erlangen (vgl. Pottharst et al. 2024, S. 165).

Abbildung 4: Wichtige Dimension für AI Literacy



Quelle: Eigene Darstellung

4 Fazit

Der Einsatz von KI in der Sozialen Arbeit bietet sowohl Chancen als auch Herausforderungen, die sorgfältig abgewogen werden müssen. KI-Technologien wie Sprachmodelle, Text-Mining-Tools oder Entscheidungsunterstützungssysteme haben das Potenzial, Prozesse zu optimieren, administrative Aufgaben zu erleichtern und Fachkräfte zu entlasten. Insbesondere in Zeiten von Fachkräftemangel könnten solche Innovationen eine entscheidende Rolle spielen, um den Arbeitsalltag effizienter zu gestalten und Freiräume für die direkte Arbeit mit Klient:innen zu schaffen.

Gleichzeitig sind die Risiken zu beachten. Besonders in den Kerntätigkeiten der Sozialen Arbeit, die unmittelbar auf die Klient:innen wirken, birgt der Einsatz von KI erhebliche Gefahren, wenn z. B. Fehler oder Verzerrungen auftreten. Die Praxis zeigt, dass unzureichend ausgebildetes Personal, fehlende technische In-

frastruktur und eine mangelnde KI-Kompetenz der Fachkräfte wesentliche Hindernisse darstellen. Hinzu kommen ethische und rechtliche Herausforderungen, wie die Einhaltung professioneller Standards und der Umgang mit anthropomorphen Systemen, die eine quasisoziale Beziehung zwischen Mensch und Maschine wahrscheinlich(er) machen.

Eine wesentliche Voraussetzung für die erfolgreiche Implementierung von KI-Systemen ist die Förderung von „AI Literacy“ – der Fähigkeit von Fachkräften, die Funktionsweise und die Auswirkungen von KI zu verstehen und kritisch zu hinterfragen. Zudem bedarf es gezielter Investitionen in technische Infrastruktur, Schulungen und die Etablierung einer KI-sensiblen Organisationskultur. Die Einführung von KI ist aus der Perspektive zu begleiten, wie die Technologie die Arbeit mit Klient:innen verbessert, ohne dabei ethische und professionelle Werte der Sozialen Arbeit zu gefährden.

Insgesamt zeigt sich, dass der Einsatz von KI in der Sozialen Arbeit ein transformatives Potenzial hat, das jedoch verantwortungsvoll und differenziert genutzt werden muss. Es bleibt eine zentrale Herausforderung, die Balance zwischen technologischem Fortschritt und der Sicherstellung einer menschenzentrierten, professionellen Sozialarbeit zu wahren.

Ein verstärkter Einsatz von Technologien in der Sozialen Arbeit setzt jedoch erhebliche Investitionen voraus. Diese umfassen die fachlich und ethisch reflektierte Entwicklung von Technologien, die technische Ausstattung der Praxis sowie die Implementierung umfassender Qualifizierungsmaßnahmen. Solche Investitionen erfordern entweder staatliche Unterstützung oder eine Reform der Finanzierungsstrukturen im Sozialsystem. Bislang sind derartige Maßnahmen jedoch nicht in dem erforderlichen Umfang erkennbar (Pottharst et al. 2024, S. 166)

Hinzu kommen Sorgen (nicht nur) aus der Praxis, dass gewonnene Zeit durch KI-Unterstützung nicht für Klient:innen aufgebracht wird, sondern im Sinne einer Profitsteigerung neue Aufgaben hinzukommen oder das Aufgabenvolumen ausgeweitet wird (vgl. Schulze/Dony/Domberg 2024, S. 18). Die Soziale Arbeit ist hier gefragt, den Einsatz von KI fachlich zu begleiten, um nicht einer stärkeren Ökonomisierung durch Technikeinsatz zu verfallen.

Literatur

- Althammer, Thomas (2024): KI und Datenschutz – Entwicklung und Einsatz im Kontext der geltenden Datenschutzgesetze. In: Kreidenweis, Helmut (Hrsg.): KI in der Sozialwirtschaft. Eine Orientierungshilfe für die Praxis. Baden-Baden: Nomos, S. 101–116.
- Althaus, Beat (2016): Zwischen Fiktion und Realität – Auswirkungen von Gewaltspielkonsum auf Kinder und Jugendliche. In: Jusletter, S. 1–24. <https://doi.org/10.5167/uzh-129336>
- Bastian, Pascal (2019): Sozialpädagogische Entscheidungen. Professionelle Urteilsbildung in der Sozialen Arbeit. Opladen und Toronto: Barbara Budrich.
- Bortzum, Edeltraud/Löhe, Julian (2022): Fachkräfte(mangel) in der Sozialen Arbeit. Daten Fakten Konsequenzen. In: Jugendhilfe 60(4), S. 255–262.

- de Vries, Alex (2023): The growing energy footprint of artificial intelligence. In: *Joule* 7(10). <https://doi.org/10.1016/j.joule.2023.09.004>
- Faruqe, Farhana/Watkins, Ryan/Medsker, Larry (2021): Competency model approach to AI literacy: research-based path from initial framework to model. <https://arxiv.org/pdf/2108.05809.pdf>
- Frey, Carl Benedikt/Osborne, Michael (2023): Generative AI and the future of work: a reappraisal. In: *The Brown Journal of World Affairs* 30(1).
- Fröbel e. V. (o. J.): Forschen für die Kita-Praxis: Sprachbarrieren überwinden mit Künstlicher Intelligenz. <https://www.froebel-gruppe.de/ki-fuer-die-kita> (Abfrage: 15.06.2025).
- Garcia ./. Character Technologies Inc., Klage vom 22. Oktober 2024, Az. 6:24-cv-01903, United States District Court, Middle District of Florida. <https://drive.google.com/file/d/1vHHNfHjexXDjQFPbGmxV5o1y2zPOW-sj/view?pli=1> (Abfrage: 15.06.2025).
- Goddard, Kate/Roudsari, Abdul/Wyatt, Jeremy C. (2012): Automation bias: a systematic review of frequency, effect mediators, and mitigators. In: *Journal of the American Medical Informatics Association: JAMIA* 19(1), S. 121–127.
- Graßhoff, Gunther/Renker, Anna/Schröer, Wolfgang (2018): *Soziale Arbeit – Eine elementare Einführung*. Wiesbaden: Springer.
- Hüttemann, Matthias/Parpan-Blaser, Anne (2022): Soziale Innovation und Innovation in der Sozialen Arbeit. In: *Soziale Innovationen* 22. Fachhochschule Nordwestschweiz. Hochschule für Soziale Arbeit, S. 11–23. https://soziale-innovation-fhnw.ch/wp-content/uploads/sites/23/220815_Soziale-Innovation_in-der-Sozialen-Arbeit.pdf (Abfrage: 15.06.2025).
- Knöth, Nils/Decker, Marie/Laupichler, Matthias Carl/Pinski, Marc/Buchholtz, Nils/Bata, Katharina/Schultz, Ben (2024): Developing a holistic AI literacy assessment matrix – Bridging generic, domain-specific, and ethical competencies. In: *Computers and Education Open* 6, S. 100177. <https://doi.org/10.1016/j.caeo.2024.100177>
- Kühn, Maximilian (2021): Künstliche Intelligenz und ihre Rolle im Sozialen. Der Beitrag zum Digitaltag 2021. drk-wohlfahrt.de/blog/eintrag/kuenstliche-intelligenz-und-die-rolle-im-sozialen/(Abfrage: 15.06.2025).
- Kahl, Yvonne/Bauknecht, Jürgen (2023): Psychische und emotionale Erschöpfung von Fachkräften der Sozialen Arbeit. Entwicklung, Ausmaß und die Rolle von Belastungs- und Resilienzfaktoren. In: *Soziale Passagen* 15, S. 213–232. <https://doi.org/10.1007/s12592-023-00448-6>
- Kneer, Georg (2009): Akteur-Netzwerk-Theorie. In: Kneer, Georg/Schroer, Markus (Hrsg.): *Handbuch Soziologische Theorien*. Wiesbaden: VS Verlag für Sozialwissenschaften, S. 19–40.
- Kohlhoff, Ludger (2024): Sozialmanagement. In: Grunwald, Klaus/Langer, Andreas/Sagmeister, Monika (Hrsg.): *Sozialwirtschaft. Handbuch für Wissenschaft, Studium und Praxis*. 2. Auflage. Baden-Baden: Nomos, S. 413–430.
- Koska, Christopher (2022): Wie kann KI die Soziale Arbeit unterstützen? Bayerisches Forschungsinstitut für Digitale Transformation. <https://www.bidt.digital/wie-kann-ki-die-soziale-arbeit-unterstuetzen/> (Abfrage: 15.06.2025).
- Kreidenweis, Helmut (2005): Die Hilfeplanung im Spiegel ausgewählter Software Produkte. Expertise im Rahmen des BMFSFJ Modellprogramms „Fortentwicklung des Hilfeplanverfahrens“. https://www.dji.de/fileadmin/user_upload/bibs/209_4520_Expertise-Software.pdf (Abfrage: 15.06.2025).
- Kreidenweis, Helmut/Diepholz, Maria (2024): Studie. Künstliche Intelligenz in der Sozialwirtschaft. Forschungsbericht. Hannover: Althammer & Kill.
- Krings, Markus/Heister, Werner (2023): Der Nutzen von KI in der Sozialwirtschaft. In: *Sozialwirtschaft aktuell* 22, S. 1–4.
- Lehmann, Robert (2024): Herausforderungen der künstlichen Intelligenz in der Sozialwirtschaft. In: Kohlhoff, Ludger (2024): *Aktuelle Diskurse in der Sozialwirtschaft V*. Wiesbaden: Springer VS. S. 163–174.
- Ley, Thomas (2010): „Unser Schreibzeug arbeitet mit an unseren Gedanken.“ Oder: Zur Konstruktion des sozialpädagogischen Falles in computerisierten Arbeitsumgebungen. In: Cleppien, Georg/

- Lerche, Ulrike (Hrsg.): Soziale Arbeit und Medien. Wiesbaden: VS Verlag für Sozialwissenschaften, S. 219–233.
- Ley, Thomas/Reichmann, Ute (2020): Digitale Dokumentation in Organisationen Sozialer Arbeit. In: Kutscher, Nadia/Ley, Thomas/Seelmeyer, Udo/Siller, Friedericke/Tillmann, Angela/Zorn, Isabel (Hrsg.): Handbuch Soziale Arbeit und Digitalisierung. Weinheim und Basel: Beltz Juventa, S. 241–254.
- Li, Pengfei/Yang, Jianyi/Islam, Mohammad A./Ren, Shaolei (2024): Making AI Less „Thirsty“: Uncovering and Addressing the Secret Water Footprint of AI Models. Cornell University. <https://doi.org/10.48550/arXiv.2304.03271>
- Linnemann, Gesa/Löhe, Julian/Rottkemper, Beate (2023): Bedeutung von Künstlicher Intelligenz in der Sozialen Arbeit. Eine exemplarische arbeitsfeldübergreifende Betrachtung des Natural Language Processing (NLP). In: Soziale Passagen 15, S. 197–211. <https://doi.org/10.1007/s12592-023-00455-7>
- Linnemann, Gesa/Löhe, Julian/Rottkemper, Beate (2024): Bedeutung von Selbstoffenbarungseffekten in quasisozialen Beziehungen mit auf generativer KI basierten Systemen in Settings von Onlineberatung und -therapie. In: e-beratungsjournal.net – Zeitschrift für Onlineberatung und computervermittelte Kommunikation 20(1), Artikel 1, S. 1–21. <https://doi.org/10.48341/9x1s-5y11>
- Löhe, Julian (2024): Einsatz von Künstlicher Intelligenz in der Kinder- und Jugendhilfe. In: Kreidenweis, Helmut (Hrsg.): KI in der Sozialwirtschaft. Eine Orientierungshilfe für die Praxis. Baden-Baden: Nomos, S. 101–116.
- Löhe, Julian/Aldendorff, Philipp (2022): Grundlagen zum Sozialmanagement. Zentrale Begriffe und Handlungsansätze. Göttingen: Vandenhoeck & Ruprecht.
- Long, Duri/Magerko, Brian (2020): What is AI Literacy? Competences and Design Considerations. In: Proceedings of the 2020 CHI conference on human factors in computing systems, ACM (2020), S. 1–16. <https://doi.org/10.1145/3313831.337672>
- Metzmacher, David (2024): Chatbot-Gespräche in Echtzeit. KI wie aus einem Science-Fiction-Film. ZDF heute. In: <https://www.zdf.de/nachrichten/wirtschaft/unternehmen/chatgpt-openai-software-ki-100.html> (Abfrage: 15.06.2025).
- Pottharst, Bill/Neumann, Alexander/Ostrau, Christoph/Seelmeyer, Udo (2024): Bewältigung des Fachkräftemangels durch technologische Innovation? Effekte von Technisierung und Digitalisierung. In: Sozial Extra 48(3), S. 162–167. <https://doi.org/10.1007/s12054-024-00694-9>
- Rammert, Werner (2007): Technik, Handeln und Sozialstruktur: Eine Einführung in die Soziologie der Technik. In: Rammert, Werner (Hrsg.): Technik – Handeln – Wissen. Wiesbaden: VS Verlag für Sozialwissenschaften, S. 11–36.
- Rawat, Danda B. (2021): Secure and trustworthy machine learning/artificial intelligence for multi-domain operations. In: Proceedings 11746. <https://doi.org/10.1117/12.2592860>
- Schaarschuch, Andreas (1996): Dienst-Leistung und Soziale Arbeit. Theoretische Überlegungen zur Rekonstruktion Sozialer Arbeit als Dienstleistung. In: Widersprüche: Zeitschrift für sozialistische Politik im Bildungs-, Gesundheits- und Sozialbereich 16(59), S. 87–97.
- Schulze, Kay/Dony, Alexander/Domberg, Simon (2024): Tipps und Hinweis für die Einführung von KI-Programmen in der eigenen Organisation. In: Der Paritätische Gesamtverband: Künstliche Intelligenz in der Sozialen Arbeit. Eine Textsammlung aus der gleichnamigen Veranstaltungsreihe 2023, S. 16–18. https://www.der-paritaetische.de/fileadmin/user_upload/Schwerpunkte/Digitalisierung/doc/ki/KI_Textsammlung_Update2024_final.pdf (Abfrage: 15.06.2025).
- Siebert, Julien (2024): Halluzinationen von generativer KI und großen Sprachmodellen (LLMs). Blog des Fraunhofer-Institut für Experimentelles Software Engineering. <https://www.iese.fraunhofer.de/blog/halluzinationen-generative-ki-llm/> (Abfrage: 15.06.2025).
- Tsiakas, Konstantinos/Murray-Rust, Dave (2022): Using human-in-the-loop and explainable AI to envisage new future work practices. In: Proceedings of the 15th International Conference on

PErvasive Technologies Related to Assistive Environments, S. 588–594. <https://doi.org/10.1145/3529190.3534779>

Wendt, Wolf Rainer (2024): Geschichte der Sozialwirtschaft. In: Grunwald, Klaus / Langer, Andreas / Sagmeister, Monika (Hrsg.): Sozialwirtschaft. Handbuch für Wissenschaft, Studium und Praxis. Baden-Baden: Nomos. 2. Auflage, S. 47–100.

Wolff, Dietmar (2024): KI im Personalmanagement. In: Kreidenweis, Helmut (Hrsg.): KI in der Sozialwirtschaft. Eine Orientierungshilfe für die Praxis. Baden-Baden: Nomos, S. 101–116.

Textanalysetechniken auf Tagesdokumentationen zur Prozessassistenz¹

Felix Holz, Michael Fellmann, Angelina Clara Schmidt

Abstract: In sozialen Dienstleistungsunternehmen werden Berichte und Tagesdokumentationen angefertigt, die wertvolles domänenspezifisches Fall- und Ablaufwissen beinhalten. Dieser Wissensschatz wird jedoch bisher kaum systematisch erschlossen und ausgewertet. Somit steht relevantes Wissen etwa zur Verbesserung von Arbeitsprozessen nicht zur Verfügung. Die Nutzung moderner Textverarbeitungstechnologien verspricht, diesen Mangel zu beheben. In diesem Beitrag werden folglich Textanalysetechniken vorgestellt, die für verschiedene Anwendungsfälle aus der Domäne sozialer Dienstleistungen relevant sind. Die Betrachtung konzentriert sich dabei auf klassische Verfahren der Künstlichen Intelligenz (KI) und Sprachverarbeitung (Natural Language Processing, NLP), da diese im Gegensatz zu neueren KI-Techniken wie Large Language Models (LLM) den Nutzenden ein erhöhtes Maß an Kontrolle und Selbstbestimmtheit hinsichtlich der Datennutzung bieten. Der Beitrag beinhaltet die Anwendung von Techniken auf einen von der Realität inspirierten Datensatz, um den Prozess zur Vorbereitung eines Hilfeplangesprächs (HPG) zu unterstützen.

Keywords: Personenbezogene Dienstleistungen, Textanalyse, Natürliche Sprachverarbeitung, Wissensextraktion

In sozialen Dienstleistungsunternehmen werden Berichte und Tagesdokumentationen angefertigt, die wertvolles domänenspezifisches Fall- und Ablaufwissen beinhalten. Dieser Wissensschatz wird jedoch bisher kaum systematisch erschlossen und ausgewertet. Somit steht relevantes Wissen etwa zur Verbesserung von Arbeitsprozessen nicht zur Verfügung. Die Nutzung moderner Textverarbeitungstechnologien verspricht, diesen Mangel zu beheben. In diesem Beitrag werden folglich Textanalysetechniken vorgestellt, die für verschiedene Anwendungsfälle aus der Domäne sozialer Dienstleistungen relevant sind. Die

1 © Die Autor:innen 2025, veröffentlicht im Verlag Beltz Juventa, Open Access: CC BY 4.0
Gesä Linnemann/Julian Löhe/Beate Rottkemper (Hg.), Künstliche Intelligenz in der Sozialen Arbeit
10.3262/978-3-7799-8562-4_013

Betrachtung konzentriert sich dabei auf klassische Verfahren der Künstlichen Intelligenz (KI) und Sprachverarbeitung (Natural Language Processing, NLP), da diese im Gegensatz zu neueren KI-Techniken wie Large Language Models (LLM) den Nutzenden ein erhöhtes Maß an Kontrolle und Selbstbestimmtheit hinsichtlich der Datennutzung bieten. Der Beitrag beinhaltet die Anwendung von Techniken auf einen von der Realität inspirierten Datensatz, um den Prozess zur Vorbereitung eines Hilfeplangesprächs (HPG) zu unterstützen.

1 Einleitung

Die in sozialen Dienstleistungen durchgeführten Tätigkeiten entsprechen wissensintensiven Geschäftsprozessen. Diese werden insbesondere durch Unvorhersehbarkeit und wissens- bzw. erfahrungsgelitetes Handeln charakterisiert. Durch die hohe Individualität jedes Klient:innenfalls und einer engen Zusammenarbeit von Dienstleister (Betreuer:in, Sozialarbeiter:in) und Dienstempfänger (Klient:in) zur Lösung komplexer Aufgaben sowie der Notwendigkeit, auf häufig unvorhergesehene Umstände mittels bestehender Erfahrungen zu reagieren, sind solche Prozesse schwer durch automatisierte Mittel zu unterstützen (vgl. Boissier/Rychkova/Le Grand 2019). Ebenso nehmen Wissensarbeitende bei Einsatz komplexer digitaler Mittel tendenziell eine höhere Gefährdung ihrer Autonomie wahr (vgl. Isegran/Kuvene/Breuning 2022).

Potenziale zur Verschlankung der Prozesse liegen in der Unterstützung einzelner Prozessschritte, die die Sozialarbeitenden durchzuführen haben, um mehr Fokus auf die Kernarbeit zu erlauben. Dies wollen wir durch eine übersichtliche und zielgerichtete Darstellung des vorhandenen Wissens in einer Reihe von Tagesdokumentationen erreichen. Da in Bastian (2019, S. 66 f.) von Bedenken zur Einschränkung des Ermessensspielraums durch erhöhte Formalisierung berichtet wird, ist zu betonen, dass lediglich Unterstützungsprozesse durch digitale Mittel vereinfacht werden, nicht die Kernprozesse. Dafür werden in diesem Beitrag Textanalysetechniken diskutiert, deren Einsatz Möglichkeiten zur Arbeitsentlastung durch Informationsbereitstellung verspricht. Dies erscheint insofern realistisch, als in der Sozialen Arbeit viele Pflichten existieren, Begegnungen mit Klient:innen zu dokumentieren. Dabei entstehen Textdateien, die wertvolles Wissen zum Termin selbst, zur Arbeitsweise, zum Zustand von Klient:innen sowie zu gewählten Maßnahmen enthalten. Diese können zu einer verbesserten Informationsbereitstellung herangezogen werden. Ein Beispiel hierfür ist es, den Prozess zur „Vorbereitung eines Hilfeplangesprächs“ (HPG) mit technischen Mitteln zu unterstützen. Effektiv besteht dieser Prozess darin, eine Zusammenfassung einer Reihe von Tagesdokumentationen über einen sechs- bis neunmonatigen Zeitraum zu erstellen. Dies kann unterstützt werden, indem die Informationen automatisch übersichtlich dargestellt werden, ohne der entspre-

chenden Fachkraft die Entscheidungsgewalt oder den Ermessensspielraum zu entziehen.

Ziel dieses Beitrags ist es, Möglichkeiten aufzuzeigen, wie die Inhalte der Tagesdokumentationen digital genutzt werden können, um in der Sozialen Arbeit Assistenzfunktionen bereitstellen zu können. Dafür werden im folgenden Abschnitt 2 zunächst Texte als Datenobjekte in sozialen Dienstleistungen betrachtet, wobei auf die Besonderheiten von Tagesdokumentationen in sozialen Dienstleistungen eingegangen wird. Im dritten Abschnitt werden einige Anwendungsfälle beleuchtet, die mit den Informationen aus Tagesdokumentationen unterstützt werden können, sowie relevante Textanalysetechniken eingeführt. Im vierten Abschnitt kommen die Techniken anhand eines beispielhaften Anwendungsfalls zur Unterstützung der HPG-Vorbereitung zum Einsatz, um die Nutzungspotenziale aufzuführen. Daraufhin werden limitierende Faktoren benannt und Empfehlungen ausgesprochen, wie durch eine strukturierte Dokumentationskultur die Ergebnisgüte der Textanalysetechniken gesteigert werden kann. Schließlich wird im Fazit auch auf die Vorteile der hier vorgestellten Textanalysetechniken gegenüber den oftmals als „Black Box“ agierenden generativen KI-Systemen, die die Basis für viele Chatbots darstellen, eingegangen.

2 Tagesdokumentationen als Datenobjekte in sozialen Dienstleistungen

In den Tagesdokumentationen werden im Regelfall alle aus Sicht der Fachkraft relevanten Vorkommnisse die Klient:innen betreffend von den Sozialarbeitenden dokumentiert. Die Gesamtheit der Tagesdokumentationen zu einem:einer Klient:in ist als *Fall* zu bezeichnen. Wir unterscheiden bei einem Fall und dessen Dokumentation in interne und externe Dokumentationen. Die interne Dokumentation ist relativ wenigen Reglementierungen unterworfen, da diese der internen Reflexion und auch der Informationsweitergabe der Mitarbeitenden untereinander dient, um sich beispielweise für Vertretungs- oder Anschlussdienste vorbereiten zu können. Diese Dokumentationen haben zumeist einen sehr individuellen Charakter und enthalten detailliertes Wissen der entsprechenden Sozialarbeiter:innen über den Fall. Die externen Dokumentationen hingegen unterliegen einer starken Reglementierung und werden an die entsprechenden Ämter geschickt und dort geprüft – es besteht eine Dokumentationspflicht, die je nach Kreis oder SGB unterschiedliche Vorgaben hat, unter Umständen sogar handschriftlich auszufüllen ist. Die externen Dokumentationen beinhalten somit meist diejenigen Informationen, die die Ämter lesen wollen, und sind zumeist kurzgehalten. Insgesamt zeigen sich bestimmte wichtige Eckpunkte bei der Beschreibung eines Klient:innenfalls: sowohl die Situation von Klient:innen als auch die Schilderun-

gen, welche Maßnahmen ergriffen und welche Tätigkeiten durchgeführt wurden. Obwohl nicht in jedem Bericht erwähnt, ist das durch die Klient:innen zu erreichende Ziel ebenfalls ein wichtiger Baustein der Fallrepräsentation. Zudem sind involvierte Personen, Orte oder Organisationen (das Netzwerk des:der Klient:in) ebenfalls zentrale Aspekte, um den Fall zu verstehen (vgl. Holz/Fellmann/Lantow 2022). Dabei können unterschiedliche Themen behandelt werden, die teils der Zielstellung entsprechen, teils aber auch kompensatorischen Aufwand² beschreiben (vgl. Holz/Lantow/Fellmann 2021).

Aus einer technischen Perspektive sind die Dokumentationen eines Falls Texte, die einem:einer Klient:in zugeordnet sind und anhand einer zeitlichen Dimension abgebildet werden. Texte selbst gelten als unstrukturierte Daten – Informationen, die nicht nach einem vorgegebenen Format gespeichert werden (können) (vgl. Blumberg/Atre 2003; Eberendu 2016). Die maschinelle Extraktion und Verarbeitung des Wissens innerhalb dieser Texte sind durch die Unstrukturiertheit erschwert, darüber hinaus gibt es weitere semantische und syntaktische Problemereiche, die die Analyse beeinträchtigen (eine Einführung zu unstrukturierten Daten und ihren Besonderheiten wird im Beitrag von Rottkemper in diesem Band gegeben):

- unterschiedliche rechtliche Grundlagen, somit unterschiedliche Dokumentationspflichten,
- mangelnde oder fehlende Erfolgsmessungen, um Effektivität der Maßnahmen einschätzen zu können,
- keine oder wenige Annotationen der Texte (z. B. für Aussagen wie: in diesem Text geht es nur um Verwaltung, zu diesem Termin wurden lediglich kompensatorische Leistungen erbracht, in diesem Termin kam es zu einer Krisensituation etc.),
- unterschiedliche Länge der Texte sowie unterschiedliche Ausführlichkeit der Berichte,
- stichpunktartige Beschreibungen bei erwarteter Satzstruktur,
- unterschiedliche Autor:innen, somit unterschiedliche Schreibstile,
- keine einheitliche Form,
- Schreibfehler und unbekannte Abkürzungen.

Darüber hinaus gibt es in der Textanalyse weitere Problemereiche, die auch domänenunabhängig auftauchen, z. B. die Verwendung von Homonymen³ oder Synonymen (vgl. Siegel/Bond 2021). In der Textanalyse wird ebenfalls nach der syntaktischen Analyse (dem Betrachten der Sprachstrukturen) und der semantischen Analyse (der Betrachtung der Bedeutung) unterschieden.

2 Nicht zielführender, aber notwendiger Aufwand bezüglich auf das vorher vereinbarte Ziel.

3 Homonyme sind Worte mit der gleichen Schreibweise, aber unterschiedlicher Bedeutung.

3 KI-gestützte Textanalysetechniken zur Lösung domänenspezifischer Anwendungsfälle

In der Sozialen Arbeit lassen sich etliche Potentiale zur Realisierung von Assistenzfunktionen finden, in denen Texte als Datengrundlage fungieren. Beispielhafte Anwendungsfälle sind die Früherkennung von Krisensituationen, das Vorschlagen von Expert:innen, die bereits ähnliche Fälle betreut haben, oder das Empfehlen von Tätigkeiten, die bereits in der Vergangenheit bei ähnlichen Situationen durchgeführt wurden.

Im Zuge dieses Beitrags werden wir jedoch die Vorbereitung des Hilfeplangesprächs (HPG) nach SGB VIII in den Fokus rücken, da es sich um eine Aufgabe handelt, die insbesondere von einer Anreicherung durch vorhandenes Wissen profitiert. Unterschiedliche Anwendungsfälle erfordern unterschiedliche Herangehensweisen, wodurch nur eine exemplarische Betrachtung möglich ist. In der Hilfeplanung werden während der initialen Aushandlung des Hilfeplans mit Vertreter:innen zuständiger Ämter, Klient:innen und deren Mitwirkenden sowie den Fachkräften regelmäßige Treffen vereinbart, um den Fortschritt der Klient:innen in Bezug auf die festgelegten Ziele zu begutachten. In diesen HPGs werden ggf. die zuvor verabschiedeten Zielstellungen angepasst oder das Betreuungsverhältnis beendet (vgl. Demski 2023, S. 117 ff.). Der Zeitraum zwischen einzelnen HPGs kann variieren, umschließt zumeist jedoch sechs Monate (vgl. Herzog/Lantow 2017). In diesem Zeitraum können viele Informationen anfallen, die die Fachkraft im Zuge der HPG-Vorbereitung zusammenstellen muss. Hierbei ist eine gezielte Informationssuche vonnöten, damit sie von Schlüsselmomenten (bezüglich der Zielerreichung) berichten kann und diese entsprechend einordnet. Ebenso ist für die Auskunftsfähigkeit ein Überblick über den Fallverlauf von Vorteil.

Um diese Vorbereitungsaktivitäten digital zu unterstützen, werden im Folgenden Techniken aus dem Bereich des Natural Language Processing (NLP, im Folgenden auch Textanalyse) aufgeführt, die in ihren Grundformen mit wenig Aufwand genutzt werden können. Deren Anwendung folgt in Abschnitt 4:

Named Entity Recognition (NER) ist eine Technik, die die Entitäten in Texten erkennt. Als Entitäten können Personen, Organisationen, Orte oder wichtige Inhalte gelten (vgl. Schmitt et al. 2019). Die NER ist besonders nützlich, um wichtige oder involvierte Personen im Hilfeverlauf herauszustellen und einen ersten Überblick über das (soziale) Netzwerk von Klient:innen zu schaffen.

Mit dem Part-of-Speech (PoS)-Tagging werden die einzelnen Teile eines Satzes (in der Regel Wörter) mit ihrer Wortart bzw. ihrer grammatikalischen Funktion innerhalb des Textes automatisch annotiert (z. B. „fahren“ ist ein Verb). Weiterhin ist es möglich, Verbindungen der Wörter untereinander einzusehen, um grammatikalische Bezüge herauszustellen. Das PoS-Tagging gibt dem Text ein

Format, das durch gezielte Suche nach bestimmten Wortarten oder syntaktischen Abhängigkeiten ausgenutzt werden kann (vgl. Chiche / Yitagesu 2022).

In der *Sentimentanalyse* werden Texte in Kategorien eingeordnet (klassifiziert), um negativ oder positiv formulierte Texte unterscheiden zu können. Dies wird beispielsweise angewendet, um Hassreden auf sozialen Plattformen zu erkennen (vgl. Subramanian et al. 2023). Die simpelste Form der Sentimentanalyse besteht im Aufrechnen von positiv konnotierten Wörtern (z. B. gut, schön, wunderbar) und negativ konnotierten Wörtern (z. B. schlecht, böse), sodass ein zählbares Endergebnis für den gesamten Text herauskommt. Bei dieser Herangehensweise ist besonders auf verstärkende (sehr, überaus) und verneinende Satzstrukturen zu achten.

Die *Themenmodellierung* ermöglicht es, zum einen unterschiedliche Themen in einer Menge von Texten zu erkennen und zum anderen entsprechende Themenanteile in einem Text zu identifizieren. Die Wörter der Texte werden unter Betrachtung statistischer Zusammenhänge von Worthäufungen bzw. Kookkurrenzen (dem gemeinsamen Auftreten von Wörtern) in Gruppen eingeteilt. Diese Gruppen stellen „Themen“ dar und werden durch ihre Wörter definiert. Somit ist es möglich, für eine Reihe von Texten die Verteilungen von unterschiedlichen Themen einzusehen (vgl. Blei 2012).

Jede dieser Techniken erfordert eine Reihe unterschiedlicher Vorverarbeitungsschritte, etwa die Stoppwortreduktion (das Herausfiltern von häufigen und inhaltlich unbedeutenden Wörtern wie „und“), Stammwortbildung (Reduktion von Flexionsformen) oder das generelle Kleinschreiben aller Wörter, damit diese für die Maschine vergleichbar und interpretierbar bleiben (vgl. Kowsari et al. 2019).

Ebenso können diese Techniken entweder mittels Verfahren des Maschinellen Lernens oder mit regelbasierten Verfahren realisiert werden (eine Einführung in die verschiedenen Methoden der KI siehe im Beitrag von Rottkemper in diesem Band). So achten beispielsweise regel- bzw. wörterbuchbasierte NER-Systeme auf großgeschriebene Wörter oder vorher bestimmte Buchstabenreihenfolgen, um Namen zu extrahieren, die in einem Glossar vorkommen. NER-Systeme, die auf Maschinellern basieren, lernen hingegen gewisse Muster zur Erkennung von Namen aus Datensätzen, in welchen die Texte mit Entitäten manuell annotiert wurden. Dabei werden weitere Parameter betrachtet, beispielsweise die Position im Text, die Schreibweise oder gewisse Bezugswörter etc. (vgl. Aggarwal 2018, S. 424 ff.). Dies hat einerseits den Vorteil, dass der syntaktische Kontext im Text betrachtet wird und auch ungewöhnliche oder falsch geschriebene Namen erkannt werden können. Andererseits ist die Genauigkeit geringer, wenn sich der zu analysierende Text zu sehr vom Trainingsmaterial unterscheidet (vgl. Schmitt et al. 2019). Um bei regelbasierten Verfahren so viele Ausreißer und Sonderfälle wie möglich abdecken zu können, entstehen mitunter umfangreiche und ggf. komplexe Regelwerke, die mehr Kontrolle ermöglichen, deren Pflege jedoch ma-

nuellen Aufwand bedeutet. Generell ist für eine höhere Genauigkeit der Ergebnisse eine Anpassung der Techniken auf die Domäne notwendig – sei es durch das Trainieren auf domänenspezifische Texte oder das Herleiten domänenspezifischer Regeln. Selbstverständlich ist es für die Anwendung von Textanalysetechniken notwendig, dass die Texte in digitaler Form vorliegen und die Reihenfolge (z. B. anhand der Datumsangaben) nachvollzogen werden kann. Jedoch ist eine Genauigkeit von 100 % nicht zu erwarten (vgl. ebd.).

4 Anwendung von Textanalyse zur Unterstützung der HPG-Vorbereitung

Im Zuge dieses Abschnitts werden die Tagesdokumentationen mit den genannten Textanalysetechniken verarbeitet. Da in der Anwendung mitunter Namen, Orte und rückverfolgbare Informationen enthalten sind, wurde zuvor ein Fall künstlich erstellt. Vereinzelt reale Ergebnisse sind entsprechend gekennzeichnet. Der KI-generierte Fall beschreibt die Klientin Lisa Müller, die Unterstützung aufgrund sozialer Ausgrenzung und familiärer Konflikte benötigt. Der Verlauf ist grundlegend erfolgreich, jedoch auch von nicht zielrelevanten Themen (wie ungesunder Ernährung) sowie von Rückschlägen geprägt. Der Datensatz umfasst 24 ausführliche Tagesdokumentationen in einem Zeitraum von sechs Monaten.

Zur Analyse des Falls wird im Folgenden zunächst der gesamte Hilfeverlauf betrachtet, um unterschiedliche Arten des Überblicks über den Fall zu erhalten und eine Fallverlaufsrepräsentation zu erstellen (Abschnitt 4.1). Im darauffolgenden Abschnitt 4.2 werden die Möglichkeiten zur gezielten Zusammenfassung von bestimmten Tagesdokumentationen präsentiert und diskutiert. Hierbei ist zu beachten, dass lediglich Standardverfahren verwendet werden und keine domänenspezifischen Anpassungen vorgenommen wurden. Mehrere Anwendungen bzw. Werkzeuge setzen diese in Abschnitt 3 vorgestellten Standardverfahren um (vgl. Schmitt et al. 2019), die Wahl der hier genutzten ist durch die Einfachheit der Implementation und den Zugang zu deutschen Modellen begründet. Abschnitt 4.3 fasst die Nutzenpotenziale gezeigter Verfahren im Bezug zur HPG-Vorbereitung zusammen.

4.1 Analyse und Repräsentation des Hilfezeitraums

Der gesamte Hilfezeitraum lässt sich auf mehrere Arten darstellen. Eine einfache Vorgehensweise liegt zunächst im Zählen, Gruppieren und Auflisten der vorhan-

denen Wörter, welche mittels des PoS-Taggings und des Tools SpaCy⁴ aufbereitet und mit entsprechenden Wortarten versehen wurden. Wenn es Sozialarbeitende interessiert, welche Tätigkeiten in einem Zeitraum durchgeführt wurden, kann man dies grob durch die vorhandenen Verben approximieren. In Tabelle 1 ist die Liste vorhandener Verben in Normalform zu finden, geordnet nach Anzahl der Vorkommnisse (n).

Tabelle 1: Häufigkeiten (n) aufkommender Wörter in einer Reihe von Tagesdokumentationen des synthetischen Falls

Wort	n	Wort	n	Wort	n	Wort	n
führen	13	zeigen	11	berichten	10	wirken	6
erarbeiten	6	besprechen	6	erhalten	5	äußern	5
bleiben	4	vereinbaren	4	ermutigen	4	verbessern	3
...		
zulassen	1	einbringen	1	anwenden	1	lösen	1

Quelle: Eigene Darstellung

Wie in Tabelle 1 zu erkennen ist, kommen Verben wie „führen“ (von Gesprächen), „berichten“ und „besprechen“ häufig vor; Tätigkeiten, um die aktuelle Situation oder den Fortschritt bezüglich der Zielerreichung der Klientin einschätzen zu können. Dies lässt auf ein hohes Maß an Gesprächsführung schließen. Die Wörter „zeigen“ und „wirken“ weisen darauf hin, dass die Fachkraft gewisse Verhaltensweisen beobachten konnte. „Vereinbaren“ und „ermutigen“ stellen Tätigkeiten des Betreuenden dar, Verhaltensänderungen der Klientin oder des Umfelds anzustreben, während „verbessern“ ebenso bereits eine Wertung des Fallverlaufs beinhaltet. Der Zweck einer solchen Fallzusammenfassung liegt in der Komplexitätsreduktion des Textes und in der Übersichtlichkeit, ggf. als Gedankenstütze für eine Fachkraft der Sozialen Arbeit, die in diesem Zeitraum eine große Klientel betreut.

Ein ähnliches Prinzip verfolgt die NER (ebenfalls umgesetzt mit SpaCy⁵), indem die relevanten Personen und Orte, die in Berichten auftauchen, als Liste aufgeführt werden:

Personen ({'Lisa': 39, 'Lisas': 3, 'Lisa Müller': 1, 'Sarah': 1, 'Lisas Medienkonsum': 1, 'Lisa Familienregeln': 1, 'Gemeinsames Gespräch': 1})

4 <https://spacy.io/usage/linguistic-features> (Abfrage: 15.06.2025)

5 <https://spacy.io/api/entityrecognizer> (Abfrage: 15.06.2025)

Orte (f'Familienkonflikt': 2, 'Schultheatergruppe': 1, 'Familienkommunikation': 1, 'Online-Mobbing': 1, 'Einnahmeplan': 1})

Hier ist zu sehen, dass hauptsächlich die fiktive Klientin in unterschiedlichen Variationen erkannt wurde, ebenso vereinzelt Personen des Umfelds. Mit Blick auf die Orte ist höchstens die „Schultheatergruppe“ als ein solcher anzusehen. In beiden Kategorien lassen sich Terme finden, die durchaus wichtige Aspekte des Falls darstellen können, jedoch fälschlicherweise als Personen oder Orte klassifiziert wurden.

Für eine graphische Repräsentation des Fallverlaufs anhand der zeitlichen Dimension lässt sich die Sentimentanalyse (umgesetzt mit TextblobDE⁶) verwenden. Wie in Abbildung 1 (links) zu sehen ist, gibt es im analysierten Fall aus einer realen Dokumentation viele positiv formulierte Berichte und einen negativ formulierten. Im „Tal des Graphen“ ist davon die Rede, dass der Klient in ein Krankenhaus eingewiesen wurde. Im darauffolgenden Bericht hat der Klient gesagt, es gehe ihm nach dem Krankenhausaufenthalt gut, was das positive Sentiment erklärt. Hiermit kann man bereits auf einem Blick abschätzen, wie der Fall verlaufen ist, und die Berichte an den entsprechend interessanten Stellen einsehen.

Abbildung 1: Fallverlauf, dargestellt als Sentiment-Graph; links aus einem realen Anwendungsfall, rechts ist der KI-generierte Fall zu sehen



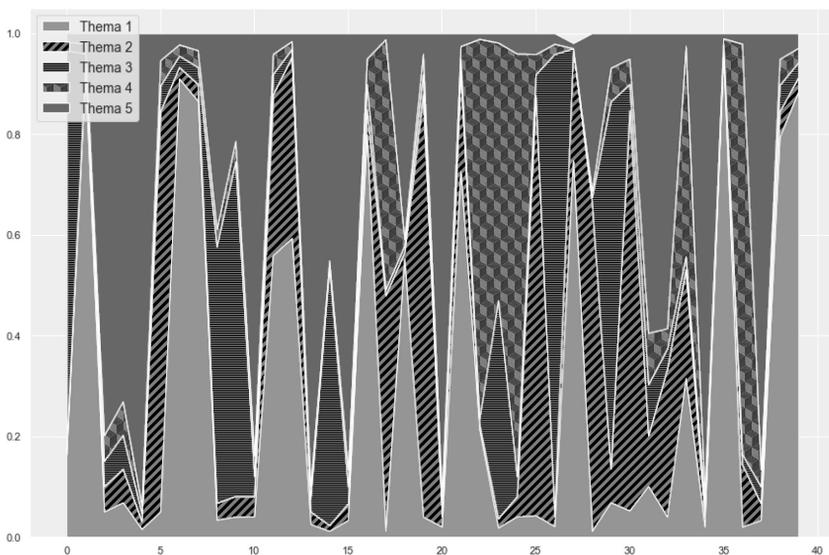
Quelle: Eigene Darstellung)

Eine weitere Möglichkeit des graphischen Verlaufsüberblicks liegt in der Themenmodellierung, also der automatischen Aufteilung des Textes in verschiedene Themen. Abbildung 2 zeigt anhand eines realen Chatverlaufs zwischen Betreuer:in und Klient:in, welche Themen zu welchem Zeitpunkt des Hilfeverlaufs in welchem Umfang besprochen wurden. Die Themen selbst werden durch die Wörter definiert, die mit der höchsten Wahrscheinlichkeit zu diesem Thema gehören. Zu sehen ist hier, dass bestimmte Themen wie Thema 1, 2 und 3 (Technik, Wahrnehmung, Schule) regelmäßig zu bestimmten Zeitpunkten auftreten, Thema 4 (Praktikum) viele Anteile der Kommunikation in der Mitte der Betreuung in

6 <https://textblob-de.readthedocs.io/en/latest/> (Abfrage: 15.06.2025)

Anspruch nahm und Thema 5 (Selbstsorge)⁷ konstant behandelt wurde. Wir nennen diese Darstellung Themendrift, um den Wechsel von Themen in einem Zeitraum zu verdeutlichen.

Abbildung 2: Themenverteilungen und Themendrift eines Chatverlaufs zwischen Betreuer:in und Klient:in



Quelle: Eigene Darstellung

4.2 Inhaltliche Analyse der Tagesdokumentationen

Während die Verlaufsdarstellungen den gesamten Verlauf eines Falls visualisieren, behandelt dieser Abschnitt einzelne Tagesdokumentationen. Bei intensiven Momenten ist es durchaus möglich, dass im Bericht sehr viele Inhalte platziert werden. Um diese Inhalte schnell greifbar zu machen, ist eine Reduktion der Komplexität angebracht, um die relevanten Bestandteile herauszustellen.

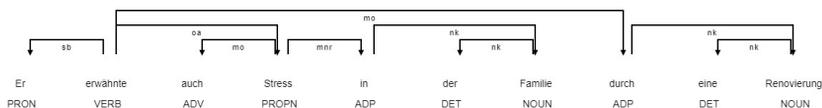
Eine Tagesdokumentation lässt sich durch das PoS-Tagging bereits so vordstrukturieren, dass die erkannte Satzstruktur für eine weitere Aufbereitung genutzt werden kann. In Abbildung 3 ist ein sogenannter Abhängigkeitsbaum/-graph (Dependency Tree) abgebildet.⁸ Neben den Wortarten werden auch die

7 Die Themen wurden von den Autor:innen selbst benannt. Der Algorithmus führt lediglich repräsentative Worte des jeweiligen Themas auf.

8 <https://spacy.io/api/dependencyparser> (Abfrage: 15.06.2025)

Abhängigkeiten zwischen den einzelnen Teilen eines Textes dargestellt. Diese Verbindungen werden genutzt, um die relevanten Informationsträger zu extrahieren. Eine simple Variante der Wissensextraktion liegt in der Subjekt-Prädikat-Objekt (SPO)-Tripel-Bildung. Dabei werden ausgehend von einem Verb lediglich die Verbindungen im Abhängigkeitsbaum zum Subjekt (sb) und zum Objekt (oa, oc) des Verbs nachvollzogen und herausgefiltert.

Abbildung 3: PoS-Abhängigkeitsbaum eines Satzes



Quelle: Eigene Darstellung

Auf diese Art können minimale Zusammenfassungen aus den Tagesdokumentationen generiert werden. So werden aus dem Bericht: „*Lisa hat sich mit einer Mitschülerin namens Emma angefreundet. Sie teilen das Interesse für Theater und haben begonnen, in den Pausen miteinander Zeit zu verbringen. Auch die Familienbeziehungen verbessern sich langsam, besonders der Vater zeigt mehr Interesse für Lisas Situation.*“ folgende SPO-Tripel gebildet:

- Lisa hat angefreundet
- Sie teilen Interesse
- Sie verbringen Zeit
- Familienbeziehungen verbessern sich
- Vater zeigt Interesse

Daraus entsteht eine Wissensbasis zur Dokumentation und ggf. des gesamten Falls (vgl. Bordes et al. 2011), die die Kernaussagen des Textes kurz und bündig darstellt und die Fachkraft möglicherweise an die Situation erinnert. Eine Zusammenfassung bietet den Vorteil der Komplexitätsreduktion. Die hier aufgeführte Tripelbildung nutzt dabei sehr simple vorhandene Sprachstrukturen und kann erweitert werden, um mehr Sprachstile abdecken zu können.

4.3 Nutzungspotenziale für den Anwendungsfall

Die aufgeführten Techniken können genutzt werden, um dem:der Sozialarbeiter:in Entlastung bei der Informationssuche zu verschaffen, sofern das HPG vorbereitet wird. Die hier präsentierten Lösungen vereinfachen die Texte im Sinne

der Übersichtlichkeit, weisen auf Besonderheiten hin, bieten die Möglichkeit zur Gesamtbetrachtung des Falls und können als Hilfe zur Navigation in den Inhalten betrachtet werden, beispielweise durch die Aufführung von Visualisierungen oder Schlüsselwörtern. Die Techniken bieten aus einer großen Menge Text eine Vorfilterung an und erlauben den Betreuenden, Muster und Veränderungen zu erkennen und entsprechend dieser gezielt weiterzuarbeiten; Interpretation und Einordnung der Informationen obliegen jedoch stets der Fachkraft.

Die extrahierten Informationen können dementsprechend genutzt werden, um schneller eine Auskunftsfähigkeit über den Fall zu erlangen sowie um Kernpunkte (wie z. B. Meilensteine) des Hilfeplans ausfüllen zu können. Es hilft dem:der Sozialarbeiter:in, sich auf wichtige Teile des Falls zu fokussieren. In Tabelle 2 sind exemplarische Fragestellungen aufgeführt, die der:die Betreuer:in während eines HPG möglicherweise zu beantworten hat. Entsprechend der Fragestellung sind mögliche Ansätze aufgeführt, um dessen Beantwortung zu unterstützen.

Tabelle 2: Technische Unterstützung der Beantwortung von möglichen HPG-Fragestellungen

HPG-Fragestellung	Technik zur Unterstützung der Beantwortung
(1) Wurde an den vorher bestimmten Zielen gearbeitet? Wurden die vereinbarten Tätigkeiten bzw. Lösungsansätze durchgeführt?	Aufstellung der Tätigkeiten über SPO-Zusammenfassungen, Aufstellung von Verben
(2) Wie gut konnten die Ziele erreicht werden?	keine, Entscheidung obliegt dem Betreuenden
(3) Haben sich weitere Problembereiche geöffnet, die bearbeitet werden mussten oder noch müssen?	Sentimentanalyse (Negativ), Schlüsselwortextraktion, Themenmodellierung, Themendrift
(4) Was wurde mit dem:der Klient:in gemacht bzw. was hat der:die Klient:in gemacht? (Entwicklung, Befinden, äußere Faktoren)	Wie (1.), zusätzlich: <ul style="list-style-type: none"> ● Entwicklung: Sentimentanalyse ● Befinden: Schlüsselwörter definieren, Aufstellung von Adjektiven ● äußere Faktoren: u. U. NER
(5) Welche Personen/Akteur:innen waren involviert bzw. wie hat sich das Netzwerk zur Selbstsorge entwickelt?	NER: Häufigkeiten der Entitäten über den Zeitverlauf (Veränderungen)

Quelle: Eigene Darstellung

Fragestellung (1) in Tabelle 2 ist zusätzlich durch einen manuellen Abgleich mit der Zielstellung zu beantworten. Die Schlüsselwortextraktion erkennt für die Aussage des Satzes relevante Wörter, ist jedoch besonders dann effektiv, wenn Begriffe für z. B. das individuelle Befinden des:der Klient:in vorher von der Fachkraft definiert werden; dies erlaubt die gezielte Suche nach individuell passenden Begriffen. Wichtig ist bei der Beantwortung der Fragen, Veränderungen der Häufigkeiten über den Betreuungszeitraum zu beobachten. So kann es beispielweise

ein Zeichen für gelöste Konflikte sein, wenn bestimmte Entitäten wie schwierige Mitschüler:innen mit der Zeit immer weniger genannt werden.

5 Limitierende Faktoren und domänenspezifische Verfeinerung

Die Nutzung der vorgestellten Techniken ist in einem realen Umfeld abhängig von der Genauigkeit bzw. der Performance entsprechender Techniken, denn die Ergebnisse variieren je nach verwendetem Modell stark und sind nicht immer nützlich. Zuverlässigkeit ist jedoch unerlässlich für die Nutzungsabsicht (vgl. Davis 1989). Um die Performance der Techniken zu verbessern, ist eine Mitbetrachtung des semantischen Kontextes eines betrachteten Wortes und somit auch die Inklusion des Bedeutungsraums vonnöten. Ein maschinelles Verständnis für die Bedeutung eines Textes ist jedoch nur mit technisch hohem Aufwand umzusetzen (vgl. Rasmussen et al. 2021; Vogl 2020). Die in diesem Beitrag verwendeten KI-Modelle sind standardisiert (trainiert mit Newsartikeln (vgl. Brants et al. 2004)) und nicht in der Lage, die Besonderheiten und Eigenheiten der Domäne gänzlich aufzufangen. Eine weitere Herausforderung liegt in weniger ausgereiften Modellen für die deutsche Sprache im Vergleich zu Modellen für die englische Sprache (vgl. Siegel/Bond 2021).

Diese Unzulänglichkeiten könnten durch ein individuelles Training auf Datensätze der Sozialen Arbeit gelöst werden. Die KI-Modelle wären dann auf die Bedürfnisse der Domäne ausgerichtet und könnten die Anwendungsfälle besser unterstützen. Als simples Beispiel kann die Sentimentanalyse in ihrer ursprünglichen Ausrichtung Texte als positiv oder negativ konnotiert anhand basaler Wörter wie „gut“ oder „schlecht“ erkennen. Zum einen kann für die Domäne sozialer Dienstleistungen ein deutlich präziseres Vokabular etabliert werden, das entweder alarmierende Wörter beinhaltet oder solche, die für eine gute Entwicklung stehen. Zum anderen ist die Definition eigener Kategorien wie „Krisensituation“, „Verbesserung“ oder „Schule“ mit entsprechenden Schlüssel- und Signalwörtern möglich, um die Techniken an die Bedürfnisse der Anwendenden anzupassen. Ein solches domänenspezifisches Vokabular kann je nach Kategorienanzahl, Teamgröße, Anwendungsfällen oder Worthäufigkeiten zu einem großen, manuellen Aufwand führen. Ebenso hat das Training von domänenspezifischen KI-Modellen hohe Anforderungen an Datenmenge und Datengüte (vgl. Holz/Lantow/Fellmann 2021; Vogl 2020).

Der erste Schritt zu domänenspezifischem Training liegt daher zunächst in der Definition der Zielstellung durch die Beantwortung von Fragestellungen wie: Welche Anwendungsfälle sollen unterstützt werden? Welche Informationen werden aus den Tagesdokumentationen benötigt, um die Anwendungsfälle zu unterstützen? Daraufhin ist die Annotation der eigenen Tagesdokumentationen notwendig, beispielsweise durch Verschlagwortung. Im Zuge des Trainings er-

lernt das KI-Modell Sprachmuster und repräsentative Worte und ordnet sie dem Schlagwort zu. Dies ermöglicht beispielsweise die Suche von Berichten anhand der Schlagworte.

Grundlegende Textanalysetechniken können Arbeitsprozesse in der Sozialen Arbeit bereits mit Standard-Modellen unterstützen. Durch eine gezielte Definition der gewünschten Unterstützungsart können die Techniken jedoch weitergehend auf die Bedürfnisse der Nutzer:innengruppe angepasst werden. Ebenso ist darauf zu achten, dass die Datensätze von hoher Qualität sind (z. B. durch das Vermeiden von Rechtschreibfehlern, oder das Einhalten einer einheitlichen Form), um eine höhere Ausgabequalität der Verfahren zu erzielen.

6 Fazit

Wir sehen Potenziale in der gezielten Extraktion und Strukturierung des Wissens, das in den Tagesdokumentationen zu finden ist. Um Überblick über dieses Wissen zu behalten und um wichtige Aspekte der Berichte aufzuzeigen, wurden in diesem Beitrag standardisierte Textanalysetechniken angewandt, um den beispielhaften Anwendungsfall der HPG-Vorbereitung zu unterstützen. Den angewendeten Techniken liegen standardisierte KI-Modelle sowie regelbasierte Systeme zugrunde – mit Ausnahme der Visualisierung (vgl. Podo/Ishmal/Angelini 2024) können LLM die aufgezeigten Lösungen ebenfalls herstellen (mehr zu LLM und regelbasierten Systemen siehe im Beitrag von Rottkemper in diesem Band). Im Vergleich zu Aussagen von LLM-basierten Chatbots sind die Ergebnisse jedoch simpler und demnach besser zu kontrollieren und nachzuvollziehen; weiterhin ist die Gefahr der Halluzination (also von Falschaussagen) geringer und die Anwendung der Techniken ist mit deutlich weniger Berechnungsaufwand lokal durchführbar, sodass schützenswerte Daten nicht an Unbefugte weitergeleitet werden. Somit ist die Umsetzbarkeit solcher „kleinen“ klassischen KI-Techniken in einem annehmbaren Zeithorizont deutlich realistischer.

Die Techniken haben ebenfalls das Ziel, das unstrukturierte Medium „Text“ mit Struktur und Format anzureichern, um Weiterverarbeitung zu ermöglichen. Mit der Grundlage eines strukturierten Datenformats ist die Anwendung weiterer Verfahren denkbar und somit auch die Unterstützung komplexerer Anwendungsfälle. So planen wir in Zukunft, die Ähnlichkeit von Texten zu berechnen, um beispielsweise die Nähe von Aussagen zur Zielstellung der Betreuung zu ermitteln oder um bereits durchgeführte und erfolgreiche Lösungen aus der Vergangenheit für ähnliche Probleme vorzuschlagen.

Um Ergebnisse in der Zukunft zu verbessern, ist die Strukturierung von Berichten z. B. durch Verschlagwortung notwendig; allerdings sollten sich dafür niedrigschwellige Konzepte finden, die Sozialarbeitende nicht zusätzlich belasten, sondern die die Arbeit bereichern und Dokumentationsaufwand reduzieren,

sodass mehr Ressourcen für die Arbeit mit den Klient:innen zur Verfügung stehen.

Literatur

- Aggarwal, Charu C. (Hrsg.) (2018): *Machine Learning for Text*. Cham: Springer International Publishing.
- Bastian, Pascal (2019): *Sozialpädagogische Entscheidungen. Professionelle Urteilsbildung in der Sozialen Arbeit*. Opladen und Toronto: Barbara Budrich.
- Blei, David M. (2012): Probabilistic topic models. In: *Commun. ACM* 55 (4), S. 77–84. <https://doi.org/10.1145/2133806.2133826>
- Blumberg, Robert/Atre, Shaku (2003): The problem with unstructured data. In: *Dm Review* 13(42–49), S. 62.
- Boissier, Fabrice/Rychkova, Irina/Le Grand, Bénédicte (2019): Challenges in knowledge intensive process management. In: *Proceedings – IEEE International Enterprise Distributed Object Computing Workshop, EDOCW 2019-October*, S. 65–77. <https://doi.org/10.1109/EDOCW.2019.00023>
- Bordes, Antoine/Weston, Jason/Collobert, Ronan/Bengio, Yoshua (2011): Learning Structured Embeddings of Knowledge Bases. In: *AAAI* 25(1), S. 301–306. <https://doi.org/10.1609/aaai.v25i1.7917>
- Brants, Sabine/Dipper, Stefanie/Eisenberg, Peter/Hansen-Schirra, Silvia/König, Esther/Lezius, Wolfgang/Rohrer, Christian/Smith, George/Uszkoreit, Hans (2004): TIGER: Linguistic Interpretation of a German Corpus. In: *Research on Language and Computation* 2(4), S. 597–620. <https://doi.org/10.1007/s11168-004-7431-3>
- Chiche, Alebachew/Yitagesu, Betselot (2022): Part of speech tagging: a systematic review of deep learning and machine learning approaches. In: *Journal of Big Data* 9(1). <https://doi.org/10.1186/s40537-022-00561-y>
- Davis, Fred D. (1989): Perceived Usefulness, Perceived Ease of Use, and User Acceptance of Information Technology. In: *MIS Quarterly* 13(3), S. 319–340. <https://doi.org/10.2307/249008>
- Demski, Jana (2023): *Partizipation in der Hilfeplanung*. In: Demski, Jana (Hrsg.): *Hilfeplangespräche in der Sozialpädagogischen Familienhilfe*. Dissertation, Bd. 28. Wiesbaden: Springer Fachmedien Wiesbaden (Soziale Arbeit als Wohlfahrtsproduktion, Band 28), S. 95–159.
- Eberendu, Adanma Cecilia (2016): Unstructured Data: an overview of the data of Big Data. In: *International Journal of Computer Trends and Technology* 38(1), S. 46–50.
- Herzog, P./Lantow, Birger (2017): Adaptive case management in social institutions [Adaptive Case Management in sozialen Einrichtungen]. In: *Lecture Notes in Informatics (LNI), Proceedings – Series of the Gesellschaft für Informatik (GI)* 275. https://doi.org/10.18420/in2017_81
- Holz, Felix/Fellmann, Michael/Lantow, Birger (2022): Ein Modellierungskonzept zur Prozessstrukturierung für Soziale Dienstleister. In: Riebisch, Matthias/Tropmann-Frick, Marina (Hrsg.): *Modellierung 2022*. Bonn: Gesellschaft für Informatik e. V, S. 171–180.
- Holz, Felix/Lantow, Birger/Fellmann, Michael (2021): Towards a Content-Based Process Mining Approach in Personal Services. In: Augusto, Adriano/Selmin Nurcan, Asif Gill/Reinhartz-Berger, Iris/Schmidt, Rainer/Zdravkovic, Jelena (Hrsg.): *ENTERPRISE, BUSINESS-PROCESS AND INFORMATION SYSTEMS MODELING*. 22. Auflage, Bd. 421. [S. l.]: SPRINGER NATURE (Lecture Notes in Business Information Processing), S. 62–77.
- Isegran, Henry/Kuvene, Mats/Breunig, Karl Joachim (2022): A bibliometric analysis deconstructing research on how digitalisation affects knowledge workers. In: *ECKM* 23(1), S. 542–551. <https://doi.org/10.34190/eckm.23.1.395>
- Kowsari, Kamran/Meimandi, Jafari/Heidarysafa, Mojtaba/Mendu, Sanjana/Barnes, Laura/Brown, Donald (2019): Text Classification Algorithms: A Survey. In: *Information* 10(4), S. 150. <https://doi.org/10.3390/info10040150>

- Podo, Luca/Ishmal, Muhammad/Angelini, Marco (2024): Vi(E)va LLM! A Conceptual Stack for Evaluating and Interpreting Generative AI-based Visualizations. <https://arxiv.org/pdf/2402.02167>
- Rasmussen, Stig Heibelstrup Rye/Bor, Alexander/Osmundsen, Mathias/Petersen, Michael Bang (2021): Super-unsupervised text classification for labeling online political hostility. <https://doi.org/10.31234/osf.io/8m5dc>
- Schmitt, Xavier/Kubler, Sylvain/Robert, Jeremy/Papadakis, Mike/LeTraon, Yves (2019): A Replicable Comparison Study of NER Software: StanfordNLP, NLTK, OpenNLP, SpaCy, Gate. In: Alsmirat, Mohammad/Jararweh, Yaser (Hrsg.): 2019 Sixth International Conference on Social Networks Analysis, Management and Security (SNAMS). Granada, Spain, October 22–25, 2019. 2019 Sixth International Conference on Social Networks Analysis, Management and Security (SNAMS). Granada, Spain, 10/22/2019–10/25/2019. Institute of Electrical and Electronics Engineers. Piscataway, NJ: IEEE, S. 338–343.
- Siegel, Melanie/Bond, Francis (2021): OdeNet: Compiling a GermanWordNet from other Resources. In: Proceedings of the 11th Global Wordnet Conference. University of South Africa (UNISA): Global Wordnet Association, S. 192–198. <https://aclanthology.org/2021.gwc-1.22> (Abfrage: 15.06.2025).
- Subramanian, Malliga/Easwaramoorthy Sathiskumar Veerappampalayam/Deepalakshmi, G./Cho, Jaehyuk/Manikandan, G. (2023): A survey on hate speech detection and sentiment analysis using machine learning and deep learning models. In: Alexandria Engineering Journal 80, S. 110–121. <https://doi.org/10.1016/j.aej.2023.08.038>
- Vogl, Thomas M. (2020): Artificial Intelligence and Organizational Memory in Government: The Experience of Record Duplication in the Child Welfare Sector in Canada. In: ACM International Conference Proceeding Series. <https://doi.org/10.1145/3396956.3396971>

Aktennotizerstellung in der Sozialen Arbeit durch Künstliche Intelligenz – Erkenntnisse aus einem Mixed-Method-Forschungsprojekt¹

Christina Plafky, Mitra Purandare, Benjamin Plattner, Svitlana Hrytsai

Abstract: Diese Studie untersucht die Entwicklung und Implementierung einer datenschutzkonformen KI-Anwendung zur automatisierten Erstellung von Aktennotizen in der Sozialen Arbeit. Im Rahmen eines Mixed-Method-Forschungsprojekts wurde ein Prototyp entwickelt, der mittels Speech-to-Text und Large Language Models unstrukturierte Sprachaufnahmen in strukturierte Aktennotizen umwandelt. Die Anwendung wurde als Stand-Alone-Lösung konzipiert, um höchste Datenschutzstandards zu gewährleisten. Durch Usability-Tests mit Fachkräften wurde die Praxistauglichkeit evaluiert. Die Ergebnisse zeigen, dass die KI-Anwendung qualitativ hochwertige Aktennotizen generiert und von den Fachkräften als nützliches Instrument wahrgenommen wird. Allerdings wurden auch Herausforderungen identifiziert, beispielsweise hinsichtlich KI-generierter Halluzinationen. Ob diese Anwendung tatsächlich zur Effizienzsteigerung beiträgt, wird in Frage gestellt, da die generierten Ergebnisse auf Richtigkeit überprüft werden müssen. Die Studie verdeutlicht das Potenzial von KI-Anwendungen in der Sozialen Arbeit und unterstreicht die Bedeutung der interdisziplinären Zusammenarbeit bei deren Entwicklung.

Keywords: Künstliche Intelligenz (KI), Automatisierung, Aktennotiz, Soziale Arbeit, Datenschutz, Dokumentationsunterstützung

¹ © Die Autor:innen 2025, veröffentlicht im Verlag Beltz Juventa, Open Access: CC BY 4.0
Gesä Linnemann/Julian Löhe/Beate Rottkemper (Hg.), Künstliche Intelligenz in der Sozialen Arbeit
10.3262/978-3-7799-8562-4_014

1 Einführung²

In der praktischen Sozialen Arbeit ist das Schreiben von Berichten und Aktennotizen ein wesentlicher Bestandteil des Berufsalltags. Es dient der Dokumentation von Fallverläufen der Klient:innen, der effektiven Kommunikation zwischen Fachkräften sowie der Rechenschaftspflicht gegenüber Adressat:innen und anderen Institutionen. Aktennotizen dokumentieren Interaktionen, Einschätzungen und Fortschritte der Adressat:innen. Diese Notizen sind unerlässlich für die Erstellung präziser Fallakten, die Ausarbeitung von Unterstützungsplänen und die Sicherstellung einer kontinuierlichen, zielgerichteten Betreuung. Aus rechtlichen, ethischen und praktischen Gründen sind Aktennotizen von großer Bedeutung für die Fachpraxis. Daher sind qualitative hochwertige Aktennotizen äußerst wichtig. Mittlerweile sind die Einrichtungen zur digitalen Falldokumentation übergegangen. Das genaue Ausmaß der Arbeitszeit, die für das Schreiben von Aktennotizen aufgewendet wird, und die Bewertung dieses Teils des Berufsalltags durch Fachkräfte wurden bisher hauptsächlich im englischsprachigen Raum untersucht (vgl. z. B. Lillis/Leedham/Twiner 2020; MacAlister 2022; Pascoe/Waterhouse-Bradley/McGinn 2022). In einer Studie aus der Schweiz wurde festgestellt, dass die wahrgenommene und tatsächliche Belastung durch administrative Aufgaben je nach Handlungsfeld und Kontext variiert (Plafky et al. 2025).

Die Schwierigkeit, vollständige und qualitativ hochwertige Aktennotizen in der Praxis zu erstellen, wird ebenso in anderen Professionen wie beispielsweise der Medizin thematisiert (vgl. z. B. Earnshaw et al. 2020). Auch wenn sich die Funktion und der Zweck der Aktennotizerstellung in den Professionen Soziale Arbeit, Medizin oder im juristischen Kontext unterscheiden mögen, gibt es doch Überschneidungen in Bezug auf Arbeitsaufwand und Qualitätssicherung. Der Einsatz von KI-Technologien zur medizinischen Dokumentation (inklusive Aktennotizen) wird bereits im Gesundheitswesen, z. B. in der Psychiatrie, Neurochirurgie und Augenheilkunde (vgl. Nguyen/Pepping 2023; Ali et al. 2024; Heilmeyer et al. 2024), sowie im juristischen Kontext (vgl. Regalia 2024) getestet. Die Ergebnisse in Bezug auf Qualität, Vollständigkeit und potenzielle Arbeitserleichterung scheinen vielversprechend. Die Studien zeigen, dass KI-Technologien die Effizienz und Genauigkeit der Dokumentation verbessern können, wobei jedoch Datenschutzaspekte berücksichtigt werden müssen.

In der EU und der Schweiz gelten seit einigen Jahren strenge Datenschutzbestimmungen. Diese sollen den verfassungsmäßigen, grundrechtlichen Anspruch auf informationelle Selbstbestimmung und den Schutz der Persönlichkeit gewährleisten. Dies spiegelt sich in verschiedenen Verordnungen und Gesetzen

2 Zur sprachlichen Verbesserung des Textes wurden ChatGPT und Microsoft Copilot verwendet.

wider (Europäische Datenschutzkonvention des Europarates, Datenschutz-Grundverordnung der EU (DSGVO), Eidgenössisches Datenschutzgesetz (DSG), kantonale Datenschutzgesetze etc.). Die Autonomie der Betroffenen sowie der Schutz vor Missbrauch persönlicher Daten und Diskriminierung durch Datenverarbeitung stehen hierbei im Mittelpunkt (vgl. Mösch Payort/Pärli 2022). (Weitere Details zum Datenschutz in der Sozialen Arbeit finden sich auch im Beitrag „KI und IT-Sicherheit und Datenschutz“ von Jan Pelzl in diesem Band).

Es stellt sich heraus, dass Fachkräfte derzeit verschiedene kommerzielle Anwendungen wie ChatGPT, Diktier-Apps auf dem Handy oder andere Tools zur Unterstützung bei der Aktennotizerstellung nutzen. Diese sogenannte Shadow IT (vgl. Haag/Eckhard 2017) bezeichnet die Nutzung von IT-Ressourcen (Software, Hardware oder Cloud-Dienste) durch Mitarbeiter:innen ohne das Wissen oder die Genehmigung der IT-Abteilung der Einrichtung. Shadow IT birgt erhebliche Risiken, da die IT-Abteilung keine Kontrolle über diese Ressourcen hat. Dies kann zu Sicherheitslücken, Datenverlust und der Missachtung von Datenschutzbestimmungen führen. Die Nutzung von kommerziellen Applikationen wie ChatGPT ist aus Datenschutzgründen ohne vorherige Überprüfung und Freigabe der Einrichtung zu vermeiden.

Es liegt nahe, über den Einsatz einer KI-Technologie nachzudenken, die Fachkräfte der Sozialen Arbeit bei der Erstellung qualitativ hochwertiger Aktennotizen unterstützen kann. Allerdings ist es wichtig, dass diese so konzipiert ist, dass sie den aktuellen Datenschutzbestimmungen für die Nutzung in der Praxis entspricht. Aus diesem Grund steht bei dem Projekt eine datenschutzkonforme KI-Lösung im Fokus. Der vorliegende Beitrag berichtet über die Erfahrungen aus technischer und anwendungsspezifischer Perspektive.

2 Methoden

Im Rahmen des Projekts „Automatisierte Erstellung von Aktennotizen durch eine KI-Anwendung“, finanziert durch die ITBO-Bildungsinitiative des Kantons St. Gallen (Schweiz), wurde von September 2023 bis August 2024 in der Zusammenarbeit zwischen dem Departement Soziale Arbeit, dem Interdisciplinary Center for Artificial Intelligence und dem Institut für Software der Ostschweizer Fachhochschule ein Prototyp im Sinne eines Proof-of-Concept für eine KI-gestützte Applikation zur strukturierten Erstellung von Fallnotizen entwickelt.

Die KI-Anwendung wurde im Prozess mittels Usability-Tests (vgl. Dumas/Reddish 1992; Barnum 2019) auf Praxistauglichkeit, technische Tauglichkeit, Nutzerfreundlichkeit und Qualität überprüft. Usability-Tests umfassten im Entwicklungsprozess Tests mit drei Fachkräften der Sozialen Arbeit in den Räumlichkeiten der Hochschule. Nach der Fertigstellung der KI, in die die Rückmeldungen der drei Personen aus unterschiedlichen Handlungsfeldern

eingeflossen sind, wurde die KI-Anwendung in einer Einrichtung der Sozialen Arbeit mehrere Tage von acht Personen vor Ort getestet. Im Anschluss wurden durch eine Fokusgruppe per Microsoft Teams die Erfahrungen mit vier Personen reflektiert und mit der Template-Analyse-Methode (vgl. Brooks et al. 2015) ausgewertet. Bei den Usability-Tests wurde auf Einhaltung der forschungsethischen Standards der Sozialen Arbeit (vgl. Franz/Unterkofler 2021) geachtet.

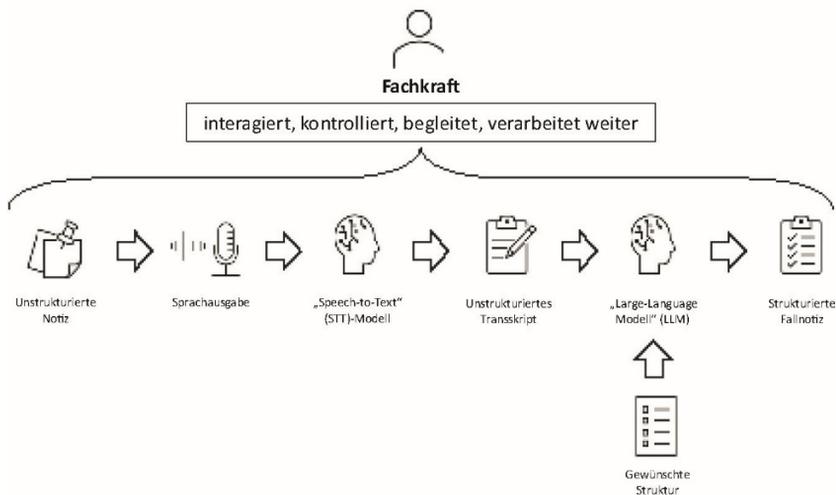
Um den strengen Datenschutzvorgaben gerecht zu werden, wurde diese KI-Anwendung als Stand-Alone-Lösung entwickelt, d. h. ohne Anbindung an Server oder das Internet. Die Schnittstellenproblematik zwischen der KI-Anwendung und einer Fallführungssoftware ist in diesem Proof-of-Concept-Projekt nicht bearbeitet worden und wird durch eine kommerzielle Anwendung gelöst werden müssen, da sehr viele unterschiedliche Fallführungssoftware in den Einrichtungen in Gebrauch sind und für jede Software eine eigene Schnittstellenlösung erarbeitet werden muss. In der aktuellen Version ist es notwendig, die Notizen per Copy-and-Paste zu übertragen.

Während des Projekts kam die Frage auf, ob es möglich wäre, das gesamte Gespräch mit den Adressat:innen (oder auch einer Teamsitzung) aufzuzeichnen und anschließend durch eine KI-Anwendung bearbeiten zu lassen. Technisch wäre dies machbar, jedoch stellt es professionsethisch eine Herausforderung dar. Es erfordert die informierte Zustimmung der Adressat:innen, und die Aufnahme eines Gesprächs ist juristisch ein anderer Sachverhalt, da es sich dann um ein Wortprotokoll (Vollprotokoll) und nicht um ein Gedächtnisprotokoll handelt. Zudem ist zu beachten, dass ein Aufnahmegerät und das anschließende Transkript die Qualität des Gesprächs und den Inhalt des Gesagten beeinflussen können. Ob dies langfristig zu einer guten Beziehungsqualität zwischen Fachkräften und Adressat:innen führt, ist fraglich. Aus diesen Gründen haben wir uns gegen diese Variante entschieden. Der vorliegende Prototyp arbeitet mit Gedächtnisprotokollen, die die Fachkräfte einsprechen müssen.

Diese Applikation ermöglicht es Fachkräften, unstrukturierte Notizen durch Spracheingabe zu erfassen. Die aufgenommene Sprache wird mittels eines sogenannten Speech-to-Text (STT)-Modells in ein unstrukturiertes Transkript umgewandelt. Dieses Transkript, zusammen mit der gewünschten Struktur für die resultierende Fallnotiz, dient als Eingabe für ein zweites Modell, das allgemein als „Large Language Model“ (LLM) (siehe auch den Beitrag von Rottkemper in diesem Band) bekannt ist. Die Spracheingabe und somit das Transkript können unstrukturiert und in Alltagssprache verfasst sein, da das LLM in der Lage ist, die relevanten Fakten aus dem Transkript zu identifizieren und zusammenzufassen. Im letzten Schritt generiert das LLM daraus die strukturierte Fallnotiz. In Abbildung 1 ist der Ablauf schematisch dargestellt.

Fallnotizen enthalten sensible personenbezogene Daten, die als privat und vertraulich gelten. Entscheidend bei deren Verwendung und Aufbewahrung ist, diese Daten angemessen zu schützen und die Kontrolle über den Daten-

Abbildung 1: Schematische Darstellung der KI-unterstützten Erfassung von Fallnotizen



Quelle: Eigene Darstellung

fluss innerhalb der eigenen Organisation – wo möglich – zu behalten. Wenn personenbezogene Daten aus Fallnotizen an externe LLM-Dienste gesendet werden (z. B. OpenAIs ChatGPT oder Anthropic's Claude), besteht das Risiko, dass unbefugte Dritte Zugang zu diesen Informationen erhalten, sei es durch Datenlecks, Sicherheitslücken oder andere Schwachstellen. Des Weiteren können LLM-Diensteanbieter Benutzereingaben auswerten, was ebenfalls ein Datenschutzrisiko darstellt. Solche Vorfälle würden die Datenschutzanforderungen verletzen.

Darüber hinaus ist oft unklar, ob externe LLM-Diensteanbieter die Daten aufbewahren und für das Training ihrer eigenen Modelle verwenden. Dies könnte theoretisch dazu führen, dass spezifische Fälle von einem LLM gelernt und unter bestimmten Umständen vom LLM preisgegeben werden. Um diese Risiken zu minimieren und den Datenschutz, Datenbesitz und die Datenkontrolle zu gewährleisten, wird sowohl das STT-Modell als auch das LLM auf der gleichen lokalen Hardware gehostet, auf der die KI-gestützte Anwendung für Fallnotizen betrieben wird.

Die für diesen Prototypen verwendete lokale Hardware kann vollständig offline gehalten werden, da die Anwendung keine Verbindung zum Internet benötigt, um die Fallnotizen zu erstellen. Dies bietet eine zusätzliche Sicherheit, indem verhindert wird, dass Daten die Hardware verlassen. Mit diesem Ansatz wird sichergestellt, dass alle sensiblen Daten innerhalb der Organisation bleiben und nicht an externe LLM-Dienste übermittelt werden müssen.

Ein limitierender Faktor beim Einsatz von Sprachmodellen auf eigener Hardware sind die verfügbaren Ressourcen. Bekannte Sprachmodelle stellen sehr hohe Anforderungen an Hardwarekomponenten, insbesondere an Grafikkarten. Für Geräte mit eingeschränkten Ressourcen (z. B. der im Rahmen dieses Projekts verwendete leistungsstarke Laptop) existieren jedoch Möglichkeiten, Sprachmodelle in kleinerer und komprimierter (quantisierter) Form zu betreiben (vgl. Han/Mao/Dally 2016; Xu et al. 2024). Obwohl die Performanz und die Qualität der Ausgaben dieser Modelle dadurch limitiert sind, generieren sie dennoch ansprechende Resultate. Aktuelle Forschungsarbeiten konzentrieren sich darauf, weitere sogenannte Small Language Models (SLM) zu entwickeln und diese den Endnutzer:innen auf deren Geräten zur Verfügung zu stellen.

Die vorgeschlagene Lösung basiert auf einem weit verbreiteten Open Source-Sprachmodell in der Kategorie der SLM mit der Bezeichnung Mistral 7B in einer 5-Bit quantisierten Form, entwickelt von der Firma Mistral (vgl. Jiang et al. 2023). Zusätzlich wird ein Open Source-STT-Modell namens Whisper von OpenAI verwendet (vgl. Radford et al. 2022). Da beide Modelle Open Source sind und aufgrund ihrer kleineren Größe lokal auf der eigenen Hardware bereitgestellt werden können, ermöglicht die vorgeschlagene KI-basierte Lösung eine umfassende Kontrolle über die Modelle und den Datenfluss, von der Eingabe bis zur Ausgabe und der Speicherung, womit die Integrität der Daten und die Privatsphäre der Adressat:innen gewährleistet wird.

Insgesamt wird eine Lösung vorgeschlagen, die darauf abzielt, den Fachkräften eine Unterstützung bei der Erstellung von qualitativ hochwertigen Fallakten zu bieten. Basierend auf realen und erfundenen Aktennotizen wurde die KI nach einer konkreten Struktur instruiert:

3 Ergebnisse der Usability-Tests

Die auf diesen Ansätzen basierende Applikation stieß in den Usability-Tests auf positive Resonanz. Die Fachkräfte schätzen die Möglichkeit, Informationen zum Fall auf unstrukturierte Weise einzugeben, da dies den Aufwand für die Erstellung von Fallnotizen reduziert. Die Qualität der generierten Zusammenfassungen wird als erstaunlich gut, äußerst nützlich und zeitsparend bewertet. Die strukturierten Fallnotizen können sich laut den Testpersonen insbesondere bei internen Übergaben und der Aktenführung als besonders wertvoll erweisen. Dies liegt daran, dass das LLM unscheinbare Zusammenhänge innerhalb einer unstrukturierten Eingabe erkennen und damit Fallnotizen in klar formulierter Form erstellen kann. Gleichzeitig wurde festgestellt, dass die KI-Anwendung Inhalte erfindet, wenn der eingesprochene Text sehr kurz ist. Die KI versucht, die Vorlage der Aktennotiz zu füllen, und erfindet fehlende Inhalte, falls nicht genügend Informationen vorhanden sind (zu einer Einordnung des Themas

Tabelle 1: Strukturvorgabe für die KI-Anwendung

Struktur der Aktennotiz	
<i>Wer?</i>	Alle Beteiligten benennen
<i>Wann?</i>	Uhrzeit, Datum und Dauer
<i>Wo?</i>	Ort
<i>Grund des</i>	Treffens/Telefonanrufs/virtuellen Meetings
<i>Ablauf des</i>	Treffens/Telefonanrufs/virtuellen Meetings
<i>Diskussionsinhalte/Besprechungsinhalte</i>	alle relevanten Inhalte benennen, eventuell Kontroversen notieren
<i>Getroffene Vereinbarungen</i>	benennen und auflisten
<i>Fachliche Interpretationen und Wahrnehmungen</i>	
<i>Wichtige Interpretation</i>	von Sequenzen, Ereignisse, Verhaltensweisen, Auffälligkeiten, Veränderungen etc.
<i>Wichtige getroffene Entscheidung</i>	benennen und fachlich begründen
<i>Nächster vereinbarter Termin</i>	

Quelle: Eigene Darstellung

Halluzinationen in generativen KI-Systeme siehe den Beitrag von Rottkemper in diesem Band). Einmal wurde im Transkript plötzlich eine unbekannte Sprache verwendet oder es wurden einfach Buchstaben aneinandergereiht, wobei die anschließend daraus generierte Zusammenfassung jedoch korrekt war, da z. B. sich wiederholende Buchstaben vom LLM für eine Zusammenfassung als unwichtig betrachtet werden. Manchmal wurden spezifische Ortsnamen oder Personennamen verändert, was offensichtlich auf eine unklare Aussprache zurückzuführen ist.

Das Einsprechen in den Computer wurde als unproblematisch beschrieben, da alle es gewohnt sind, mit Headsets zu arbeiten. Die Fachkräfte fänden eine Auswahl verschiedener Aktennotiztypen mit unterschiedlichen Vorlagen hilfreich, z. B. eine Vorlage für einen Hausbesuch oder eine für ein kurzes Telefonat zur Terminabstimmung. Ebenso hilfreich wäre die Möglichkeit, Audiodateien direkt in die KI einzuspeisen (beispielsweise von einer Handy-App), ohne den Umweg über das Abspielen und Aufnehmen gehen zu müssen.

Diese fehlende Möglichkeit wurde als unbedeutend angesehen, solange die KI-Anwendung auf dem eigenen Rechner genutzt werden kann. Das Kopieren und Einfügen von Text werden nicht als hinderlich für die Effizienz angesehen. Es wurde festgestellt, dass die Anwendung einfach und intuitiv funktioniert und nach einer sehr kurzen Einweisung keine Schwierigkeiten bereitet. Allerdings erfordert die Anwendung eine gewisse Lernkurve, da man erst durch Ausprobieren herausfindet, wozu die KI fähig ist und wozu nicht. Zur Effizienzsteigerung wur-

de in den Usability-Tests erwähnt, dass der Arbeitsaufwand aufgrund des Kontrollbedarfs gegenüber der KI nicht geringer ist, als wenn man die Aktennotiz selbst schreibt. Es wurde jedoch eingeräumt, dass man die Aktennotiz bei genügend Vertrauen nicht mehr lesen würde. Begrüßt würde eine Erweiterung, sodass nicht Inhalte eingesprochen, sondern auch Texte aus anderen Quellen hinein kopierbar wäre. Ebenso wichtig wäre die Möglichkeit, das Einsprechen zu unterbrechen und später fortzusetzen, da es im Arbeitsalltag häufig zu Unterbrechungen kommt. Derzeit muss das Einsprechen komplett von vorne begonnen werden. Die Erkennung des schweizerdeutschen Dialekts funktioniert teilweise. Dies ist vom Dialekt abhängig, wobei ein Dialekt, der dem Hochdeutschen ähnlicher ist, besser erkannt wird. Dies wurde aber in diesem Projekt nicht tiefgründig analysiert (vgl. Paonessa et al. 2023).

Die Testpersonen sehen jedoch Potenzial darin, dass die Anwendung zur Qualitätssicherung und -steigerung beitragen kann, da die KI gute Zusammenfassungen liefert und die Vorlagen sinnvoll erscheinen.

4 Diskussion

Das Problem von Halluzinationen stellt sowohl bei der Transkription mittels STT als auch bei der Erstellung von Fallnotizen durch LLM eine bedeutende Herausforderung dar. Es gibt jedoch verschiedene Ansätze, um diese Halluzinationen zu minimieren. Eine Möglichkeit besteht in der Anpassung spezifischer Parameter, beispielsweise des Temperaturparameters, der die Kreativität der Modelle steuert. Der Temperaturparameter ist ein Skalierungsfaktor, der die Wahrscheinlichkeitsverteilung der Wortvorhersagen beeinflusst und somit ein Gleichgewicht zwischen der Generierung kreativer (aber potenziell fehleranfälliger) und konservativer (aber möglicherweise weniger interessanter) Ausgaben ermöglicht. Ein weiterer Ansatz zur Verbesserung der Resultate ist die Anwendung von Beam Search, einer Technik, die alternative Resultate erkundet und dadurch die Qualität der Ausgaben erhöhen kann (vgl. Peng/Morton 1988). Darüber hinaus kann das Finetuning eines Sprachmodells auf die Struktur der Aktennotiz oder mit fachspezifischen Begriffen die Genauigkeit weiter steigern. Finetuning verändert die Parameter des Modells so, dass gewünschte Resultate wahrscheinlicher werden (siehe den Beitrag von Rottkemper in diesem Band) (vgl. Wei et al. 2021). Verschiedene Faktenüberprüfungsverfahren, darunter Wissensgraphen (vgl. Hogan et al. 2020), Retrieval-Augmented Generation (RAG) (vgl. Lewis et al. 2021) und Faktenchecks, die von anderen LLM durchgeführt werden (LLM-as-a-Judge, vgl. Zheng et al. 2023), haben das Potenzial, das Risiko von Halluzinationen bei der Erstellung von Fallakten zu senken. Wissensgraphen speichern Beziehungen zwischen Fakten oder Entitäten und können, ähnlich wie der RAG-Ansatz, der verschiedene Dokumente in einer separaten Daten-

bank speichert, das LLM bei der Abfrage von Fakten unterstützen und so zu zuverlässigeren Resultaten führen. Beispielsweise könnte ein Wissensgraph, der fachspezifisches Wissen enthält, verwendet werden, um die Richtigkeit von Diagnosen oder Behandlungsvorschlägen in automatisch generierten Aktennotizen zu überprüfen. Beim LLM-as-a-Judge-Ansatz kann ein weiteres Sprachmodell hinzugezogen werden, um die generierten Resultate zu überprüfen und ggf. korrigierend einzugreifen. Da die Modelle auf Wahrscheinlichkeiten basieren und mit Unsicherheiten behaftet sind, kann eine hundertprozentig korrekte Generierung von Text oder anderen Modalitäten mit heutigem Wissensstand nicht garantiert werden. Das Problem der Halluzinationen ist ein aktuelles Forschungsgebiet, das weiterhin intensiv untersucht wird (vgl. Ji et al. 2023). Eine potenzielle Weiterentwicklung zur verbesserten Identifikation von Halluzinationen könnte die Berücksichtigung des Konfidenzgrades der generierten Wörter oder Wortteile (sogenannte Token) umfassen. Ein vordefinierter, anpassbarer Schwellenwert würde festlegen, ab wann die Konfidenz eines generierten Tokens als eher hoch oder niedrig eingestuft wird. Wörter, bei denen das Modell eine geringe Konfidenz aufweist, könnten auf der Benutzeroberfläche farblich hervorgehoben werden. Dies würde Fachkräften ermöglichen, gezielt jene Stellen zu identifizieren, die wahrscheinlich von einer Überprüfung auf Richtigkeit profitieren würden. Es ist jedoch wichtig zu beachten, dass auch Wörter, die vom Modell mit hoher Sicherheit als korrekt eingestuft werden, fehlerhaft sein können. Daher ist es unerlässlich, dass Fachkräfte die generierte, strukturierte Fallakte vor der Weiterverarbeitung oder dem Einfügen in das Fallaktensystem sorgfältig überprüfen. Aus professionsethischen und rechtlichen Gründen muss die Fachkraft zwingend die Verantwortung für die Aktennotiz übernehmen. Das bedeutet, dass sie auch für potenzielle Fehler, die durch die KI generiert werden, verantwortlich ist, sofern diese nicht vor der Ablage in der Akte korrigiert werden. Dies könnte zur Folge haben, dass die Erstellung einer Notiz durch die KI möglicherweise nicht schneller ist, wenn umfangreiche Korrekturen oder ein abschließendes Durchlesen erforderlich sind. Weitere Forschung ist erforderlich, um diese Fragestellungen genauer zu untersuchen.

Das User-Interface, beispielsweise durch eine Ergänzung der Auswahlmöglichkeiten für unterschiedliche Formen und damit auch Struktur (siehe Tabelle 1) von Aktennotizen (Telefonanruf, Hausbesuch, Meeting etc.), und die Bedienung der Applikation sollten generell nach den Rückmeldungen der Fachkräfte sowie weiteren gezielten Tests mit ausgewählten Fachkräften kontinuierlich erweitert und verbessert werden.

Verschlüsselung der Daten

Der Datenschutz ist ein zentraler Aspekt im Umgang mit Daten von Adressat:innen. Daher wurde ein Ansatz gewählt, bei dem die Modelle lokal und vollständig offline auf einer Maschine betrieben werden. Dies gewährleistet mit hoher Wahr-

scheinlichkeit, dass der Datenfluss die Maschine nicht verlässt. Dennoch können Sicherheitslücken im System oder das Übertragen generierter Fallnotizen mittels USB-Stick auf ein unsicheres oder kompromittiertes System zu einem Datenabfluss führen. Ein möglicher Ansatz, um die Daten auch während der Übertragung zu schützen, ist die Verschlüsselung. Insbesondere im Umgang mit Daten von Adressat:innen ist es wichtig, dass die Daten sowohl im Ruhezustand („at rest“) als auch während der Übertragung („in transit“) verschlüsselt werden, um unbefugtes Auslesen zu verhindern. Eine derartige Lösung kann mit dem entsprechenden Fachwissen und angemessenem Aufwand implementiert werden und wird nachdrücklich empfohlen. Insgesamt stellt die Verschlüsselung einen unverzichtbaren, jedoch technisch anspruchsvollen Prozess dar, der sorgfältige Planung und kontinuierliche Überwachung erfordert.

Zentrale Lösung zum Betreiben der Applikation

Als Alternative zum Betrieb der Applikation auf einem mobilen Endgerät wie einem Laptop kann die Nutzung eines zentralen Servers in Betracht gezogen werden. Diese Option bietet den Vorteil, leistungsfähigere Hardware zu verwenden, die es ermöglicht, größere Sprachmodelle zu betreiben und dadurch schnellere sowie qualitativ hochwertigere Ergebnisse zu erzielen. Zudem trägt der Einsatz leistungsstärkerer Modelle dazu bei, Halluzinationen zu reduzieren (Brown et al. 2020). Ein zentraler Aspekt dieser Lösung ist jedoch die Verschlüsselung der Daten sowohl während der Übertragung als auch bei der Speicherung.

Datenschutz und Nutzung auf dem Mobiltelefon

Die Verschlüsselung der Daten wird insbesondere wichtig, wenn die Daten zentral in das Fallaktensystem integriert werden. Ein möglicher Anwendungsfall ist die Aufnahme einer Sprachnotiz von unterwegs in einer ungestörten Umgebung (z. B. im eigenen Auto) nach einem Hausbesuch. Dazu kann das Mikrofon des Mobiltelefons als Aufnahmegerät dienen. Dabei bietet sich eine zentrale Lösung an, bei der die beiden Sprachmodelle (STT sowie SLM respektive LLM) auf einem Server laufen, der über das mobile Endgerät, z. B. via Virtual Private Network (VPN), erreichbar ist. Daneben ist es von zentraler Wichtigkeit, dass die Daten trotz der sicheren Verbindung eines VPN zusätzlich verschlüsselt werden. Die übertragenen sowie die generierten Daten werden zur Verarbeitung auf dem Server gespeichert, was eine sichere Verschlüsselung unumgänglich macht. Grundsätzlich sind mobile Lösungen eine von Fachkräften gewünschte und nutzenbringende Funktion, wobei unklar ist, wie viele Fachkräfte tatsächlich über dienstliche Mobiltelefone verfügen. Die Nutzung von unverschlüsselten, hochsensitiven Daten auf privaten Geräten birgt erhebliche Sicherheits- sowie Datenschutzrisiken, ausgehend von Malware oder bei Verlust des Gerätes. Für die Nutzung einer KI-Anwendung zur automatisierten Aktennotizerstellung ist es außerdem wichtig, klare Regeln festzulegen, an welchen Orten diese App verwendet werden darf. Beispielsweise

wäre eine Nutzung in einem Café oder auf einer belebten Straße aus datenschutzrechtlichen Gründen problematisch.

Schnittstellen

Funktionen von Applikationen können über sogenannte „Application Programming Interfaces“ (APIs) oder Schnittstellen für andere Anwendungen zugänglich gemacht werden. Ein weit verbreiteter Architekturstil für Schnittstellen ist „Representational State Transfer“ (REST) (Fielding 2000). Ohne näher auf die technischen Details einzugehen, wurde die vorgestellte Applikation unter Verwendung dieses Architekturstils entwickelt, sodass andere Applikationen beispielsweise die Sprachmodelle ansteuern und nutzen könnten. Umgekehrt könnte die Applikation selbst wiederum andere Anwendungen, z. B. ein Fallnotizenverwaltungssystem, ansteuern und die generierten strukturierten Fallnotizen übertragen. Dieser REST-Architekturstil ist ein zentraler Baustein, um neue Applikationen in bestehende Umgebungen integrieren zu können. Der Einsatz einer solchen KI-Applikation kann womöglich dazu führen, dass Arbeitsprozesse im Kontext der Aktennotizerstellung als einfacher und effizienter wahrgenommen werden. Eine Integration in bereits bestehende Fallführungssoftware durch entsprechende Schnittstellen wäre ideal, wurde aber, wie bereits erwähnt, im Rahmen dieses Projekts nicht verfolgt.

Veränderte Arbeitsprozesse

Fragen zu den veränderten Arbeitsprozessen, die sich durch die Nutzung solcher Technologien entwickeln werden, sind noch unbeantwortet. Es lässt sich jedoch festhalten, dass dieses Beispiel einer KI-Anwendung langfristig zu einer effizienteren und qualitativ einheitlicheren Fachpraxis führen könnte. Dennoch werden dadurch neue Herausforderungen entstehen und vor allem andere Fragen an die Fachpraxis gestellt, die derzeit noch schwer abzuschätzen sind, jedoch möglicherweise bereits in ähnlicher Form durch andere Digitalisierungsprozesse aufgetreten sind (vgl. z. B. Kutscher 2024). An dieser Stelle entstehen neue (Forschungs-)Fragen: z. B. Wie werden sich die Rollen und Verantwortlichkeiten der Fachkräfte durch den Einsatz von KI-Anwendungen verändern? Welche neuen Aufgaben und Kompetenzen werden erforderlich sein? Welche Schulungs- und Weiterbildungsmaßnahmen sind notwendig, um Fachkräfte auf den Einsatz von KI-Anwendungen vorzubereiten? Wie können diese Maßnahmen effektiv gestaltet werden? Wie kann die Zusammenarbeit zwischen IT-Spezialist:innen und Fachkräften der Sozialen Arbeit verbessert werden, um die Implementierung von KI-Anwendungen zu unterstützen?

Nicht zu unterschätzen sind hier die Anforderungen an die Fachkräfte, sich mit immer neuen IT-Anwendungen konfrontiert zu sehen und kontinuierlich digitale Kompetenzen aufbauen zu müssen. Dies erfordert nicht nur technisches Wissen, sondern auch die Fähigkeit, sich an ständig wechselnde Technologien

und Arbeitsprozesse anzupassen. Gleichzeitig erfordert es von Einrichtungen, neue Handlungsleitlinien zu entwickeln, die die Nutzung und Anwendung solcher Applikationen festlegen. Intuitiv bedienbare und sorgfältig integrierte Applikationen erleichtern dabei die Akzeptanz.

5 Schlussfolgerung

Zusammenfassend lässt sich durch dieses Forschungsprojekt sagen, dass eine KI-Lösung für die automatisierte Aktennotizerstellung als datenschutzkonforme Stand-Alone-Lösung erfolgreich entwickelt werden kann. Es hat sich gezeigt, dass diese Anwendung in der Fachpraxis auf Interesse stößt und zur Erhöhung der Arbeitseffizienz sowie zur Sicherstellung von Qualitätsstandards bei Aktennotizen innerhalb einer Einrichtung beitragen könnte.

Grundsätzlich wurde die Qualität der Aktennotizen als gut empfunden, aber nur durch die Auswertung durch weitere Fachkräfte und über einen längeren Zeitraum kann dies und eine potenzielle Effizienzsteigerung umfänglich beurteilt werden. Halluzinationen treten unausweichlich auf. Mit Faktenüberprüfungsverfahren oder Wissensgraphen können die Halluzinationen allenfalls reduziert, aber mit heutigem Wissensstand nicht beseitigt werden.

Strenge Datenschutzerfordernissen bringen für die Anwendung von KI-Technologien in der Praxis Sozialer Arbeit einige Nachteile mit sich. Beispielsweise erfordert die Einrichtung einer sicheren App zur Aufnahme von Aktennotizen auf einem Mobiltelefon eine sachkundige Umsetzung durch Fachpersonen. Ein weiteres Problem sind bei Stand-Alone-Lösungen die Schnittstellenlösungen.

Der Einsatz einer solchen KI-Anwendung in der Praxis wurde von den Testpersonen begrüßt und als hilfreich bezeichnet, dies gilt insbesondere für die Zusammenfassungen und die Qualitätssicherung. Konkret ist eine datenschutzkonforme KI-Anwendung eine Möglichkeit für Einrichtungen, zu verhindern, dass Fachkräfte andere, kommerzielle KI-Anwendungen nutzen, die den Datenschutz nicht sicherstellen können.

Gleichzeitig gibt es jedoch noch einige offene Fragen, die für den konkreten Einsatz in der Praxis durch weitere Projekte geklärt werden müssen. Dazu gehören die langfristigen Auswirkungen solcher technischen Lösungen auf den Arbeitsalltag und die Fallarbeit, die konkrete, messbare Effizienzverbesserung, der Schulungsbedarf/ die Lernkurve sowie die Bereitschaft, solche Anwendungen in verschiedenen Handlungsfeldern und Arbeitskontexten dauerhaft zu nutzen. Festzuhalten ist, dass bei der Entwicklung solcher KI-Lösungen zwingend Fachkräfte der Sozialen Arbeit von Beginn an beteiligt sein müssen. Die Hürden, die bei der Entwicklung auftauchen, haben einen direkten Bezug zum Arbeitsalltag der Fachkräfte und können nur gemeinsam gelöst werden (Co-Creation). Dies erhöht die Wahrscheinlichkeit signifikant, dass solche Lösungen in der Praxis

nicht nur akzeptiert, sondern auch intensiv genutzt werden, da die Anwendung optimal auf die Bedürfnisse der Fachkräfte vor Ort angepasst ist. Die im Rahmen dieses Projektes initiierte Zusammenarbeit zwischen der Sozialen Arbeit und der Software-Entwicklung ist ein erfolgreiches Beispiel für eine interdisziplinäre Entwicklung von KI-basierten Lösungen für die Praxis.

Literatur

- Ali, Abdullah/Kumar, Rohit Prem/Polavarapu, Hanish/Lavadi, Raj Swaroop/Mahavadi, Anil/Legarreta, Andrew D./Hudson, Joseph S./Shah, Manan/Paul, David/Mooney, James/Dietz, Nicholas/Fields, Daryl P./Hamilton, D. Kojo/Agarwal, Nitin (2024): Bridging the Gap: Can Large Language Models Match Human Expertise in Writing Neurosurgical Operative Notes? In: *World Neurosurgery* 192, S. e34-e41.
- Brooks, Joanna/McCluskey, Serena/Turley, Emma/King, Nigel (2015): The utility of template analysis in qualitative psychology research. *Qualitative Research in Psychology* 12(2), S. 202–222. <https://doi.org/10.1080/14780887.2014.955224>
- Brown, Tom/Mann, Benjamin/Ryder, Nick/Subbiah, Melanie/Kaplan, Jared D./Dhariwal, Prafulla/Neelakantan, Arvind/Shyam, Pranav/Sastry, Girish/Askeel, Amanda/Agarwal, Sandhini/Herbert-Voss, Ariel/Krueger, Gretchen/Henighan, Tom/Child, Rewon/Ramesh, Aditya/Ziegler, Daniel/Wu, Jeffrey/Winter, Clemens/Hesse, Chris/Chen, Mark [...] Amodei, Dario (2020): Language Models are Few-Shot Learners. 34th Conference on Neural Information Processing Systems (NeurIPS 2020), Vancouver, Canada. arXiv preprint <https://arxiv.org/abs/2001.05382>.
- Earnshaw, Charles H./Pedersen, Amanda/Evans, Jo/Cross, Tina/Gaillemain, Oliver/Vilches-Moraga, Arturo (2020): Improving the quality of discharge summaries through a direct feedback system. *Future Healthcare Journal* 7(2), S. 149–154.
- Fielding, Roy Thomas (2000). Architectural styles and the design of network-based software architectures (Doctoral dissertation, University of California, Irvine). <https://ics.uci.edu/~fielding/pubs/dissertation/top.htm> (Abfrage: 15.06.2025).
- Franz, J., & Unterkofler, U. (Eds.). (2021). *Forschungsethik in der Sozialen Arbeit: Prinzipien und Erfahrungen*. Verlag Barbara Budrich.
- Haag, Steffi/Eckhardt, Andreas (2017): Shadow it. *Business & Information Systems Engineering* 59, S. 469–473.
- Han, Song/Mao, Huizi/Dally, William J. (2015): Deep compression: Compressing deep neural networks with pruning, trained quantization and Huffman coding. arXiv preprint <https://arxiv.org/abs/1510.01402>.
- Heilmeyer, Felix/Böhringer, Daniel/Reinhard, Thomas/Arens, Sebastian/Lyssenko, Lisa/Haverkamp, Christian (2024): Viability of Open Large Language Models for Clinical Documentation in German Health Care: Real-World Model Evaluation Study. In: *JMIR Medical Informatics* 12, e59617.
- Hogan, Aidan/Blomqvist, Eva/Cochez, Michael/Melo, Gerard D./Gutierrez, Claudio/Kirrane, Sabrina/Labra Gayo, José Emilio/Navigli, Roberto/Neumaier, Sebastian/Ngonga Ngomo, Axel Cyrille/Polleres, Axel/Rashid, Sabbir M./Rula, Anisa/Schmelzeisen, Lukas/Sequada, Juan F./Staab, Steffen/Zimmermann, Antoine/d'Amato, Claudia (2021): Knowledge graphs. *ACM Computing Surveys (Csur)* 54(4), S. 1–37. <https://doi.org/10.1145/3447772>
- Ji, Ziwei/Lee, Naon/Frieske, Rita/Yu, Tiezheng/Su, Dan/Xu, Yan/Ishii, Etsuko/Bang, Yejin/Madotto, Andrea/Fung, Pascale (2023): Survey of hallucination in natural language generation. *ACM Computing Surveys* 55(12), S. 1–38. <https://doi.org/10.1145/3571730>
- Jiang, Albert Q./Sablayrolles, Alexandre/Mensch, Arthur/Bamford, Chris/Chaplot, Devendra S./de las Casas, Diego/Bressand, Florian/Lengyel, Gianna/Lample, Guillaume/Saulnier, Lucile/La-

- vaud, L elio Renard/Lachaux, Marie-Anne/Stock, Pierre/le Scao, Teven/Lavril, Thibaut/Wang, Thomas/Lacroix, Timoth e/Sayed, William E. (2023): Mistral 7B. arXiv preprint <https://arxiv.org/abs/2310.13005>
- Kutscher, Nadia (2024): Digitalit t und Digitalisierung als Gegenstand der Sozialen Arbeit. In: Kurtz, Thomas/Meister, Dorothee M./Sander, Uwe (Hrsg.): *Digitale Medien und die Produktion von Wissenschaft*. Wiesbaden: Springer VS, S. 121–146. https://doi.org/10.1007/978-3-658-42542-5_8
- Liang, Percy/Bommasani, Rishi/Lee, Tony/Tsipras, Dimitris/Soylu, Dialara/Yasunaga, Michihiro/Zhang, Yian/Narayanan, Deepak/Wu/Kumar, Ananya/Newman, Benjamin/Yuan, Binhang/Yan, Bobby/Zhang, Ce/Cosgrove, Christian/Manning, Christopher D./R e, Christopher/Acosta-Navas, Diana/Hudson, Drew A./Zelikman, Eric/Durmus, Esin [...] Koreeda, Yuta (2022): Holistic Evaluation of language models. arXiv preprint <https://arxiv.org/abs/2210.12088>
- Lillis, Theresa/Leedham, Maria/Twiner, Alison (2020): Time, the written record, and professional practice: The case of contemporary social work. In: *Written Communication* 37(4), S. 431–486.
- Lillis, Theresa/Twiner, Alison/Balkow, Michael/Lucas, Gillian/Smith, Miriam/Leedham, Maria (2023): Reflections on the procedural and practical ethics in researching professional social work writing. In: *Journal of Applied Linguistics and Professional Practice* 18(3), S. 315–342. <https://doi.org/10.3138/jalpp.20014>
- MacAlister, Josh (2022): The independent review of children's social care – Final report, S. 178–187. webarchive.nationalarchives.gov.uk/ukgwa/20230308122535mp_/https://childrensocialcare.independent-review.uk/wp-content/uploads/2022/05/The-independent-review-of-childrens-social-care-Final-report.pdf (Abfrage: 15.06.2025).
- M osch Payot, Peter/P arli, Kurt (2022): Datenschutz in der Sozialen Arbeit. Eine Praxishilfe zum Umgang mit sensiblen Personendaten. *AvenirSocial I – Berufsverband Soziale Arbeit Schweiz*. https://avenirsocial.ch/wp-content/uploads/2023/01/Datenschutz-i-d-SA_db_120123.pdf (Abfrage: 15.06.2025).
- Nguyen, Josh/Pepping, Christopher A. (2023): The application of ChatGPT in healthcare progress notes: A commentary from a clinical and research perspective. In: *Clinical and translational medicine* 13(7), e1324. <https://doi.org/10.1002/ctm2.1324>
- Ow, Peng Si/Morton, Thomas E. (1988): Filtered beam search in scheduling. In: *The International Journal of Production Research* 26(1), S. 35–62.
- Paonessa, Claudio/Schraner, Yanick/Deriu Jan/H urlimann Manuela/Vogel, Manfred/Cieliebak, Mark (2023): Dialect Transfer for Swiss German Speech Translation. arXiv preprint <https://arxiv.org/abs/2310.13005>
- Pascoe, Katheryn Margaret/Waterhouse-Bradley, Bethany/McGinn, Tony (2023): Social workers' experiences of bureaucracy: A systematic synthesis of qualitative studies. *The British Journal of Social Work* 53(1), S. 513–533. <https://doi.org/10.1093/bjsw/bcac106>
- Radford, Alec/Kim, Jim Wook/Xu, Tao/Brockman, Greg/McLeavey, Christine/Sutskever, Ilya (2023): Robust speech recognition via large-scale weak supervision. In: *International conference on machine learning*, S. 28492–28518. PMLR.
- Regalia, Joe (2024): From Briefs to Bytes: How Generative AI is Transforming Legal Writing and Practice. In: *Tulsa Law Review* 59, 193. https://papers.ssrn.com/sol3/papers.cfm?abstract_id=4371460 (Abfrage: 15.06.2025).
- Wei, Jason/Bosma, Maartens/Zhao, Vincent Y./Guu, Kelvin/Yu, Adams Wei/Lester, Brian/Du, Nan/Dai, Andrew M./Le, Quoc V. (2021): Finetuned language models are zero-shot learners. arXiv preprint <https://arxiv.org/abs/2109.00859>
- Xu, Jiajun/Li, Zhiyuan/Chen, Wei/Wang, Qun/Gao, Xin/Cai, Qi/Ling, Ziyuan (2024): On-Device Language Models: A Comprehensive Review. arXiv preprint <https://arxiv.org/abs/2406.12345>

IT-Sicherheit und Datenschutz im Kontext von KI-Sprachmodellen¹

Jan Pelzl

Abstract: Der Beitrag betrachtet die Rolle von Datenschutz und IT-Sicherheit in der Sozialen Arbeit in Bezug auf KI-Sprachmodelle und führt dabei in die Grundzüge der Daten- und IT-Sicherheit ein, beschreibt Chancen und Risiken von KI-Sprachmodellen sowie Möglichkeiten zur Risikominimierung. Der Einsatz von KI steht in einem zunehmend kritischen Verhältnis zum Datenschutz, da KI-Systeme oft auf der Sammlung, Analyse und Verarbeitung großer Mengen von Daten basieren, um zu lernen und Entscheidungen zu treffen. Diese Daten können oft persönliche Informationen enthalten, welche entsprechend geschützt werden müssen. Der Datenschutz fordert daher die Entwicklung und Anwendung von KI in einer Weise, die die Privatsphäre und die Autonomie der Individuen schützt, was den Einsatz von Techniken des Datenschutz-freundlichen maschinellen Lernens einschließt. Der Beitrag beschreibt, wie gängige Vorgehensweisen der IT-Sicherheit wie z. B. Informationssicherheitsmanagementsysteme auch auf den Bereich von KI-Anwendungen ausgeweitet werden können, um diese datenschutzkonform zu gestalten und zu betreiben.

Keywords: IT-Sicherheit, Kryptografie, ISMS, KI-Manipulation, Datenschutz

1 Einführung: IT-Sicherheit und Datenschutz bei KI-Anwendungen in der Sozialen Arbeit

Heutige sogenannte generative Modelle in der Künstlichen Intelligenz (KI) erlernen die Datenverteilung von Trainingsdaten und können auf dieser Basis neue Inhalte erzeugen (siehe den Beitrag von Rottkemper in diesem Band). Die Modelle beschränken sich dabei nicht nur auf reinen Text, sondern können auch Bilder, Sprache, Musik und sogar Filme erzeugen.

Neben den vielen Vorteilen birgt KI aber auch Risiken. Um solche Risiken zu minimieren, spielen Datenschutz und IT-Sicherheit eine wichtige Rolle. IT-Si-

1 © Die Autor:innen 2025, veröffentlicht im Verlag Beltz Juventa, Open Access: CC BY 4.0
Gesä Linnemann/Julian Löhe/Beate Rottkemper (Hg.), *Künstliche Intelligenz in der Sozialen Arbeit*
10.3262/978-3-7799-8562-4_015

cherheit und ebenso Datenschutz sind gerade in Bezug auf den Einsatz von KI in der Sozialen Arbeit sehr relevant, da in diesem Bereich viele sehr sensible Daten verarbeitet werden. Hierbei ist insbesondere der Faktor Mensch zu betrachten, da die Sensibilisierung der Nutzer:innen bezüglich Risiken der IT häufig nicht weit vorangeschritten und ein tieferes Verständnis technischer Grundlagen der IT oftmals nicht gegeben ist (Kreidenweis 2024, S. 8).

Gängige Vorgehensweisen der IT-Sicherheit können auch auf den Bereich von KI-Anwendungen ausgeweitet werden, um diese datenschutzkonform zu gestalten und zu betreiben. So bieten beispielsweise in der IT etablierte Informationssicherheitsmanagementsysteme (ISMS) einen systematischen Ansatz zum sicheren Managen von Informationen (vgl. Bundesamt für Sicherheit in der Informationstechnik 2017). Durch Anwendung einer Risikomanagementstrategie, die Personen, Prozesse und IT-Systeme umfasst, können Vertraulichkeit, Integrität und Verfügbarkeit von Informationen geschützt werden.

Die folgenden Abschnitte führen zunächst in die Grundzüge der Daten- und IT-Sicherheit ein. KI-gestützte Sprachmodelle erscheinen für die Sozialen Arbeit eine besonders relevante KI-Entwicklung, weil Sprache ein Hauptwerkzeug im Rahmen der Dienstleistungserbringung in der Sozialen Arbeit ist. Aufgrund der hohen Bedeutung kommunikativer Prozesse ist daher die Verarbeitung natürlicher Sprache – wie sie in Sprachmodellen stattfindet – für die Praxis der Sozialen Arbeit von besonderer Relevanz (vgl. Linnemann/Löhe/Rottkemper 2023). Denn das Medium Sprache ist wesentlich für die Beratungsleistung in der Sozialen Arbeit. Weil technische Modelle die natürliche Sprache so gut nachahmen können, dass eine Unterscheidung nicht mehr möglich ist (vgl. Casal/Kessler 2023), werden diese technischen Errungenschaften damit auch für die Soziale Arbeit bedeutsam. Daher wird vorliegender Beitrag im zweiten Teil den Fokus auf Risiken beim Einsatz von KI-Sprachmodellen legen und es werden Möglichkeiten zur Risikominimierung bei entsprechendem Einsatz in der Sozialen Arbeit besprochen.

2 Grundlagen der Daten- und IT-Sicherheit

Die Sicherstellung der IT-Sicherheit ist eine fundamentale Voraussetzung für den erfolgreichen Einsatz von Informationstechnik. Die Herausforderungen von IT-Sicherheit insbesondere in der Sozialen Arbeit sind vielschichtig. Zu den zentralen Problemen gehören der Schutz sensibler Daten, da in der Sozialen Arbeit häufig mit vertraulichen Informationen über Klient:innen gearbeitet wird. Diese Daten reichen von persönlichen und finanziellen Informationen bis hin zu Gesundheitsdaten, deren Sicherheit nach strengen gesetzlichen und ethischen Standards gewährleistet werden muss (vgl. Goldberg 2021). Ebenfalls herausfordernd ist die Gewährleistung der Datenintegrität, um die Genauigkeit und Zuverlässigkeit der gespeicherten Informationen sicherzustellen, was für

die Planung und Bereitstellung sozialer Dienste unerlässlich ist. Ein weiteres kritisches Element ist die Sensibilisierung und Schulung der Mitarbeiter:innen in sozialen Einrichtungen hinsichtlich potenzieller Cyber-Bedrohungen und sicherer Onlinepraktiken. Angesichts der zunehmenden Digitalisierung der Sozialen Arbeit und der damit verbundenen Nutzung von Informationstechnologie (vgl. Liebsch/Degel 2025, S. 7ff.) erhöht sich das Risiko von Cyberangriffen. Die Diakonie Michaelshofen (o. J.) spricht beispielsweise von einer Entwicklung der Cyberkriminalität, die eine dramatische Dynamik eingenommen hat, „da die Nutzung digitaler Angebote privat als auch beruflich stark gestiegen ist“ (ebd.). Vor allem Soziale Organisationen sind beliebte Ziele für Cyberangriffe, da die hier vorliegenden Daten zu denen mit dem höchsten Schutzbedarf in Deutschland gehören (vgl. ebd.). Dass es sich nicht um Einzelfälle handelt, zeigt, dass z. B. im Herbst 2023 ein Hackerangriff 72 Kommunen in Südwestfalen lahmgelegt hat. Klient:innen der Sozialen Arbeit waren davon konkret betroffen, indem keine Gelder für Wohnen, Asyl oder Unterhaltsvorschuss ausgezahlt werden konnten (vgl. Jonas 2024). Schließlich stellt die Integration von IT-Sicherheitsmaßnahmen in bestehende Arbeitsabläufe ohne Beeinträchtigung der Servicequalität oder Zugänglichkeit der Dienste eine besondere Herausforderung dar. Die Balance zwischen dem Schutz der Privatsphäre der Klient:innen und der Notwendigkeit, offene und effektive soziale Dienste anzubieten, ist dabei ein zentraler Aspekt.

IT-Sicherheit besteht nicht nur aus technischen Maßnahmen wie z. B. Verschlüsselung oder Virenschutz, sondern auch aus organisatorischen Maßnahmen und Prozessen zum Management der IT-Sicherheit, um eine kontinuierliche Anpassung und Verbesserung des Schutzes zu erreichen. Im Folgenden werden die Grundzüge des IT-Sicherheitsmanagements sowie technische und organisatorische Maßnahmen kurz vorgestellt.

2.1 IT-Sicherheitsmanagement

Ein ISMS ist ein systematischer Ansatz zum Managen von Informationen, um Sicherheit zu gewährleisten (vgl. Bundesamt für Sicherheit in der Informationstechnik 2017). Es wendet eine Risikomanagementstrategie an, die beteiligte Personen, Prozesse und IT-Systeme berücksichtigt. Basis eines ISMS ist die Identifizierung potenzieller Risiken, die zu Informationsverlusten führen können, sowie die Minimierung von Risiken durch die Implementierung geeigneter Sicherheitsmaßnahmen. Ein effektives ISMS hilft Organisationen, ihre geschäftlichen, rechtlichen, vertraglichen und regulatorischen Verpflichtungen zu erfüllen. Darüber hinaus steigert es das Vertrauen gegenüber Außenstehenden in die Fähigkeit, Informationen zu schützen. Dieser Aspekt ist aufgrund der besonders sensiblen Personendaten in der Sozialen Arbeit von hoher Bedeutung.

Ein ISMS basiert auf dem Prinzip einer kontinuierlichen Verbesserung und wird in der Regel in Übereinstimmung mit international anerkannten Standards wie ISO/IEC 27001 (vgl. International Organization for Standardization 2022) oder auch dem BSI IT-Grundschutz (vgl. Bundesamt für Sicherheit in der Informationstechnik 2017) entwickelt und implementiert. Diese Norm definiert die Anforderungen an eine Einrichtung für die Implementierung, Aufrechterhaltung und kontinuierliche Verbesserung eines Informationssicherheitsmanagementsystems. Die wesentlichen Bestandteile eines ISMS lassen sich in organisatorische und technische Maßnahmen unterteilen (vgl. NIST – National Institute of Standards and Technology 2024).

2.2 Organisatorische Maßnahmen

Organisatorische Maßnahmen umfassen Vorgaben, Richtlinien und Prozesse zum Umgang mit sensiblen Informationen und zur Einhaltung rechtlicher Vorgaben sowie zur Regelung von Vorfällen. Wesentliche Bausteine sind:

- **Richtlinien und Verfahren:** Vorgaben und Anleitungen für die Organisation, wie Informationssicherheit im Alltag zu gewährleisten ist.
- **Risikomanagement:** Identifizierung und Bewertung von Risiken für die Informationssicherheit sowie die Entwicklung von Strategien zu deren Minderung oder Akzeptanz.
- **Asset-Management:** Identifizierung und Klassifizierung von Informationswerten innerhalb der Organisation und Festlegung von Verantwortlichkeiten für deren Schutz.
- **Incident-Management:** Verfahren zur Behandlung von Sicherheitsvorfällen und Schwachstellen.
- **Notfallpläne:** Es müssen klare Protokolle und Notfallpläne für den Fall eines Sicherheitsvorfalls existieren. Eine schnelle und effektive Reaktion kann den Schaden minimieren.
- **Business-Continuity-Management:** Sicherstellung, dass die Organisation ihre Arbeit im Falle eines Sicherheitsvorfalls oder einer Katastrophe fortsetzen kann.
- **Backup- und Wiederherstellungspläne:** Regelmäßige Backups und detaillierte Pläne zur Wiederherstellung von Daten und Systemen nach einem sicherheitsrelevanten Vorfall.
- **Compliance:** Überprüfung und Sicherstellung, dass die Anforderungen aus gesetzlichen Regelungen, Verträgen und eigenen Richtlinien zur Informationssicherheit eingehalten werden.

2.3 Technische Maßnahmen

Technische Maßnahmen spielen eine wichtige Rolle beim Datenschutz und der Datenverarbeitung. Diese Maßnahmen können in physische und informationstechnische Maßnahmen unterteilt werden, um ein umfassendes Sicherheitssystem zu schaffen. Es ist wichtig, sowohl physische als auch informationstechnische Sicherheitsmaßnahmen zu implementieren, um ein umfassendes Sicherheitsnetz für die Verarbeitung und Speicherung von Daten zu schaffen.

Physische Maßnahmen zielen darauf ab, den Zutritt zu physischen Standorten zu kontrollieren und somit den direkten Zugriff auf Hardware und Datenträger zu beschränken. Beispiele für solche Maßnahmen sind Zäune und Tore, die einen ersten Schutz bieten, indem sie den Zutritt auf ein bestimmtes Gelände regeln. Ein Empfang dient nicht nur als Visitenkarte des Unternehmens, sondern auch als eine wichtige Sicherheitsschicht, wo Besucher:innen registriert und überprüft werden. Schlüssel und abschließbare Aktenschränke schützen sensible Dokumente und Datenträger vor unbefugtem Zugriff. Alarmanlagen und Videokameras wirken präventiv gegen Einbrüche und ermöglichen eine Nachverfolgung im Falle eines Sicherheitsvorfalls. Elektronische Eingangskontrollen und speziell geschützte Serverschränke sichern die IT-Infrastruktur physisch ab.

Parallel dazu hat die Sicherheitstechnik im Bereich der Informationstechnologie eine immer bedeutendere Rolle erlangt, insbesondere durch die Zunahme von Daten, die in Cloud-Umgebungen gespeichert werden. Zu entsprechenden informationstechnischen Sicherheitsmaßnahmen zählen beispielsweise rollenbasierte Zugriffskontrollen sowie ein Zwang zu guten Passwörtern, die sicherstellen, dass nur autorisierte Nutzer:innen Zugang zu sensiblen Informationen haben. Technische Sicherheitsabfragen und Methoden zur Zwei-Faktor-Authentifizierung (2FA) erhöhen die Hürden für unbefugten Zugriff (vgl. NIST – National Institute of Standards and Technology 2023). Sensible Daten müssen sowohl bei der Übertragung als auch bei der Speicherung verschlüsselt werden. Alle Datenübertragungen sollten Ende-zu-Ende verschlüsselt sein, um Abhörversuche zu verhindern. Protokollierung und biometrische Benutzeridentifikation bieten zusätzliche Ebenen von Sicherheitsüberprüfungen, indem sie genaue Aufzeichnungen über den Zugriff auf Daten und Systeme bereithalten sowie eine eindeutige und schwer zu fälschende Identifikationsmethode für Nutzer:innen darstellen.

Darüber hinaus sollten IT-Systeme regelmäßig auf Sicherheit überprüft werden:

- Systeme und Netzwerke sollten auf Schwachstellen überprüft und Sicherheitsupdates zeitnah installiert werden. Dies beinhaltet auch die Schulung der Mitarbeiter:innen in Best Practices der IT-Sicherheit.
- Penetrationstests: Regelmäßige Tests zur Identifizierung und Behebung von Schwachstellen in den Systemen.

- **Schulung und Sensibilisierung:** Mitarbeiter:innen sollten regelmäßig geschult werden, um sicherheitsbewusst zu handeln und mögliche Bedrohungen zu erkennen.

2.3.1 Kryptografie als Basis für IT-Sicherheit

Täglich sind wir von IT-Systemen umgeben, die durch kryptografische Verfahren, also spezielle Verschlüsselungstechniken, geschützt werden. Ob beim Schreiben einer E-Mail oder einer Online-Überweisung – in beiden Fällen sorgen diese Verfahren dafür, dass Daten sicher übertragen werden. Selbst elektronische Autoschlüssel nutzen kryptografische Verschlüsselungstechniken, um unbefugten Zugriff zu verhindern (vgl. Fraunhofer o. J.). Mit moderner Kryptografie lassen sich viele Eigenschaften erreichen, welche u. a. für die Absicherung von Netzwerken und auch KI genutzt werden können.

Wichtige Sicherheitsziele sind (vgl. Paar 2024)

- **Geheimhaltung oder Vertraulichkeit:** Nur autorisierte Benutzer:innen bekommen Zugang zu der Information. Zum Beispiel können Klient:innendaten mit kryptografischen Methoden verschlüsselt werden, um Vertraulichkeit zu erreichen. Daten werden auf Datenträgern oder in der Cloud verschlüsselt abgelegt und können selbst bei Diebstahl ohne den richtigen Schlüssel nicht gelesen werden.
- **Integrität:** Nachrichten wurden während der Übertragung nicht verändert. Die Integrität von Trainingsdaten, beispielsweise eines KI-Sprachmodells, kann etwa über kryptografische Prüfsummen abgesichert werden.
- **Nachrichtenauthentizität:** Der Ursprung einer Nachricht kann eindeutig nachvollzogen werden. Dies ist eine wichtige Methode zu Absicherung von Software und kann in vielen Teilbereichen genutzt werden. Die Echtheit von Protokollen oder Klient:innendaten kann beispielsweise über digitale Signaturen gewährleistet werden und ist für alle Beteiligten nachvollziehbar.

Kryptografie wird schon seit langem standardmäßig im Internet eingesetzt, beispielsweise beim Onlinebanking, bei der E-Mail-Verschlüsselung oder bei Webshops. Die Kryptografie lässt sich dabei unterteilen in zwei wesentliche Verfahren: Symmetrisch und asymmetrisch (vgl. ebd., S. 181 f):

- *Symmetrische Algorithmen* sind die bekannteste und auch intuitivste Form der Kryptografie. Zwei Parteien besitzen eine Chiffre zum Ver- und Entschlüsseln und haben sich auf einen gemeinsamen geheimen Schlüssel geeinigt. Die gesamte Kryptografie von der Antike bis in das Jahr 1976 folgte ausschließlich diesem Ansatz. Symmetrische Algorithmen wie z. B. der Advanced Encryption Standard (AES) (vgl. NIST – National Institute of Standards and Technology 2001) sind fester Bestandteil nahezu jedes heutigen Kryptosystems (vgl. Paar

2024). Sie werden insbesondere für die eigentliche Verschlüsselung von Daten und zum Integritätsschutz, d. h. Schutz gegen Veränderungen, eingesetzt.

- *Asymmetrische* (oder Public-Key-)Algorithmen: Im Jahr 1976 wurde von Whitfield Diffie, Martin Hellman und Ralph Merkle eine gänzlich neue Art der Kryptografie eingeführt (vgl. IETF – Internet Engineering Task Force 1999). Diese Algorithmen sind mathematisch etwas aufwendiger und von der Idee her schwieriger zu verstehen. Der wesentliche Unterschied zur symmetrischen Kryptografie besteht darin, dass es zwei statt einem Schlüssel gibt: Ein öffentlicher Schlüssel zum Verschlüsseln und ein privater zum Entschlüsseln. Die mathematische Basis der asymmetrischen Algorithmen besteht aus Einwegfunktionen, welche die Verschlüsselung ohne den privaten Schlüssel praktisch unmöglich gestalten.

In der Mehrzahl von kryptografischen Anwendungen in der Praxis kommen sowohl symmetrische als auch asymmetrische Algorithmen zum Einsatz. Man spricht in diesem Zusammenhang manchmal von Hybridsystemen (vgl. Paar/Pelzl/Güneysu 2024, S. 4). Der Grund dafür, dass beide Algorithmen-Familien zum Einsatz kommen, ist, dass beide Arten von Chiffren ihre spezifischen Vorteile haben.

2.3.2 Digitale Signatur

Aufgrund zukünftiger Entwicklungen im Bereich der Informationstechnologie ist die moderne Kryptografie nicht mehr wegzudenken. Neben der Verschlüsselung und der damit verbundenen Vertraulichkeit von Daten gibt es auch Methoden zur Überprüfung des Nachrichtenursprungs (Authentizität). Digitale Signaturen bieten uns genau diese Möglichkeiten und werden u. a. zur Erstellung digitaler Zertifikate verwendet, um Webbrowser abzusichern, beim rechtlich bindenden Signieren digitaler Verträge oder bei der sicheren Aktualisierung von Software. Digitale Signaturen sind in Grenzen mit konventionellen Signaturen auf Papier vergleichbar. Mit ihnen kann sichergestellt werden, dass eine Nachricht tatsächlich von der Person stammt, die angibt, sie versendet zu haben. Darüber hinaus können Daten und Software auf ihre Authentizität geprüft werden (vgl. Lorengel 2017).

2.3.3 Zertifikate

Um asymmetrische Verfahren sicher verwenden zu können, muss die Authentizität der öffentlichen Schlüssel geprüft werden. Dies geschieht über Zertifikate, welche neben dem öffentlichen Schlüssel noch Informationen zu der Identität der:des Eigentümer:in wie z. B. einer E-Mail-Adresse enthalten. Zertifikate sind wiederum digital signierte Dokumente, welche von einer vertrauenswürdigen

gen dritten Instanz, der sogenannten Zertifizierungsstelle (englisch: Certificate Authority, kurz CA) erstellt werden (eine ausführliche Abhandlung zur Kryptografie, digitalen Signatur und weiteren Mechanismen findet sich in Paar/Pelzl/Güneysu (2024)).

2.4 Rolle der IT-Sicherheit zum Schutz von und vor KI

Der Einsatz von KI in der Sozialen Arbeit bietet erhebliche Chancen zur Verbesserung der Effizienz und der Qualität der Betreuung (vgl. Steiner/Tschopp 2022). Gleichzeitig stellt er jedoch erhebliche Anforderungen an die IT-Sicherheit und ethische Standards. Um die Vorteile ausschöpfen zu können, müssen sorgfältige Maßnahmen zum Schutz der Daten und zur Wahrung der Privatsphäre getroffen werden. Nur so kann das Potenzial der KI genutzt werden, ohne die Integrität und das Vertrauen in die sozialen Berufe zu gefährden. Eine ausgewogene Integration von KI und menschlicher Empathie, unterstützt durch robuste IT-Sicherheitsmaßnahmen, kann der Sozialen Arbeit helfen, den Herausforderungen des digitalen Zeitalters erfolgreich zu begegnen und die bestmögliche Unterstützung für die Klient:innen zu gewährleisten.

Die Manipulation von KI-Systemen stellt eine erhebliche Bedrohung dar, die durch vielfältige Angriffsmöglichkeiten realisiert werden kann. Um diesen Bedrohungen zu begegnen, spielt die IT-Sicherheit eine entscheidende Rolle. Durch robuste Trainingsdaten, effektive Sicherheitsprotokolle, Verschlüsselung, Transparenz und gut vorbereitete Notfallpläne können KI-Systeme vor Manipulationen geschützt werden. Es ist wichtig, kontinuierlich in Sicherheitsmaßnahmen zu investieren und diese an die sich ständig weiterentwickelnden Bedrohungen anzupassen, um die Integrität und Verlässlichkeit von KI-Systemen zu gewährleisten. Aufgrund der besonderen Bedeutung von Sprache für die Soziale Arbeit (vgl. Linnemann/Löhe/Rottkemper 2023) sind Sprachmodelle auch in den Fokus der Praxis Sozialer Arbeit gekommen. Eine Textsammlung des Paritätischen listet z. B. unterschiedliche Text-KI-Modelle auf und verweist in Manier einer Arbeitshilfe auf interessante Anwendungsmöglichkeiten in der Sozialen Arbeit (vgl. Der Paritätische Gesamtverband 2024, S. 12 ff.). Daher wird im Folgenden ein Fokus auf Risiken und mögliche Sicherheitsmaßnahmen im Rahmen der Verwendung von KI-Sprachmodellen gelegt.

3 Risiken von KI-Sprachmodellen

Große KI-Sprachmodelle (sogenannte Large Language Models – LLM) basieren auf generativen KI-Systemen (siehe dazu den Beitrag von Rottkemper in diesem Band). Durch LLM generierte Texte lassen sich nicht ohne Weiteres von men-

schengeschilderten Texten unterscheiden (vgl. Casal/Kessler 2023). Sie werden heute vielfach für Chatbots und persönliche Assistenzsysteme eingesetzt, um die Bedienbarkeit von Systemen zu verbessern (vgl. Kelbert/Siebert/Jöckel 2023). Sie finden jedoch auch Anwendung in anderen Bereichen des heutigen Lebens und der Wissenschaft, z. B. in der Medizin, in der Informatik, im juristischen Bereich sowie in der Sozialen Arbeit. Bei ihrem Einsatz ist die Aufmerksamkeit auf unterschiedliche Kategorien von Risiken zu richten. Nach dem Bundesamt für Sicherheit in der Informationstechnik (BSI 2024) können Risiken von LLM in die drei nachfolgenden Kategorien unterteilt werden:

3.1 Risiken bei der normalen Nutzung

Bei der ordnungsgemäßen Verwendung moderner Technologien stehen Anwender:innen vor vielfältigen Risiken: unerwünschte Ausgaben, direkte und unveränderte Wiedergabe der original gespeicherten Daten und verschiedene Formen von Verfälschungen bis hin zu schlechter Qualität und falschen Fakten und dem Auftreten von Halluzinationen (siehe den Beitrag von Rottkemper in diesem Band). Zudem beeinträchtigt die häufig fehlende Aktualität durch veraltete (Trainings-)Daten die Relevanz der Informationen. Problematisch sind die mangelnde Reproduzierbarkeit und Erklärbarkeit von Ergebnissen. Ebenso kritisch sind die Anfälligkeit für die missverständliche Interpretation von Text als direkte Anweisung und die fehlende Vertraulichkeit der eingegebenen Daten, z. B. durch Verzerrung in den zugrunde liegenden Daten (vgl. Rawat 2021). Diese Verzerrungen können durch Manipulation von Daten auch bewusst herbeigeführt werden. Hinzu kommen Risiken durch selbstverstärkende Effekte, die zum Kollabieren des Modells führen können, sowie die Abhängigkeit von dem das System entwickelnden oder betreibenden Unternehmen. Deshalb ist es ggf. wichtig, die Maßnahmen zum Datenschutz oder zur IT-Sicherheit von kooperierenden Organisationen zu betrachten, sofern diese Zugriff auf personenbezogene Daten der Sozialen Organisation haben.

3.2 Risiken durch eine missbräuchliche Nutzung

Missbräuchliche Nutzung von Sprachmodellen birgt erhebliche Risiken wie z. B. die Erzeugung und Verbreitung von Falschmeldungen und Unterstützung von Social-Engineering-Methoden, die auf der Manipulation von Personen basieren, um sie zur Preisgabe vertraulicher Informationen oder zur Ausführung bestimmter Handlungen zu bewegen. Darüber hinaus besteht aus Datenschutzsicht die Gefahr der Wiedererkennung von Personen aus zuvor anonymisierten Daten. Zukünftige Modelle werden in der Lage sein, aus verfügbaren Daten noch besser

personenbezogene Daten zu rekonstruieren. Zudem können Kriminelle KI für die Wissenssammlung und -aufbereitung für Cyberangriffe nutzen. Besonders bedenklich sind die Generierung und fortlaufende Verbesserung von bösartiger Software zum Angriff auf Computer und Netzwerke.

Im Folgenden werden wesentliche Angriffe auf LLM erläutert.

- Prompting Attacks nutzen die Eingabeaufforderungen (Prompt), durch die diese Modelle gesteuert werden, um unerwünschte oder schädliche Reaktionen von der KI zu provozieren. Bei einem Prompting-Angriff gibt ein Angreifer sorgfältig konstruierte Eingaben an ein KI-System, mit dem Ziel, das System zu manipulieren oder auszunutzen oder verbotene Inhalte ausgeben zu lassen (vgl. Open Web Application Security Project 2025; vgl. Maus 2023).
- Privacy Attacks bei Sprachmodellen sind Angriffe, die darauf abzielen, sensible oder persönliche Informationen aus diesen Modellen zu erkennen (vgl. Carlini 2021). KI-Sprachmodelle werden oft mit großen Mengen an Textdaten trainiert. Diese Daten können aus öffentlich zugänglichen Quellen stammen, aber auch persönliche oder vertrauliche Informationen enthalten. Angriffe können verschiedene Formen annehmen:
 - Angreifende versuchen, bestimmte Daten, die zum Training des Modells verwendet wurden, zu rekonstruieren. Hierbei könnten sensible Informationen und Identitäten, die indirekt in den Trainingsdaten enthalten waren, aufgedeckt werden.
 - Durch gezielte Anfragen an das Sprachmodell könnten Angreifende versuchen, Informationen zu extrahieren, die in den zugrunde liegenden Trainingsdaten enthalten sind, beispielsweise Details über Personen, Orte oder spezielle Ereignisse.
 - Angreifende können über das Ausgabeverhalten des Modells Rückschlüsse auf die Trainingsdaten oder die Struktur des Modells ziehen, die eigentlich vertraulich bleiben sollten.
- Evasion Attacks bei Sprachmodellen sind Angriffe, bei denen Angreifende die Eingabe an das Modell so manipulieren, dass das Modell eine falsche Ausgabe liefert (vgl. Muthalagu/Mali/Parar 2025). Die Manipulation ist für menschliche Betrachtende oft schwer zu erkennen. Diese Art von Angriff zielt darauf ab, die Funktionsweise eines Sprachmodells zu umgehen. Dabei bleibt das Modell für legitime Anfragen funktional, reagiert jedoch auf manipulierte Eingaben in einer Weise, die Angreifende beabsichtigen. Bei Sprachmodellen kann dies beispielsweise durch das Einfügen, Löschen oder Ersetzen von Wörtern oder Buchstaben in einem Text geschehen, ohne den scheinbaren Sinn oder die Kohärenz des Textes für menschliche Lesende zu stören. Der Angriff könnte darauf abzielen, Zensurmechanismen zu umgehen, unerwünschte

Inhalte in einem System zu platzieren oder eine falsche Klassifizierung oder Interpretation durch das Modell zu erreichen. Evasion Attacks unterstreichen die Notwendigkeit widerstandsfähiger Modelle, die in der Lage sind, auch manipulierte Eingaben korrekt zu verarbeiten oder zu erkennen.

- Poisoning Attacks beziehen sich auf bösartige Aktivitäten, bei denen Angreifer absichtlich irreführende oder schädliche Daten in den Datensatz einführen, der zum Trainieren des Modells verwendet wird (vgl. Wallace 2020). Ziel solcher Angriffe ist es, das Lernverhalten des Modells zu beeinflussen, um seine Funktionsweise zu manipulieren oder zu stören. Bei Sprachmodellen kann dies dazu führen, dass das Modell fehlerhafte, voreingenommene oder unerwünschte Antworten generiert, sobald es mit legitimen Anfragen konfrontiert wird. Ein Beispiel für einen Poisoning-Angriff könnte das Einfügen von Texten in den Trainingsdatensatz sein, die speziell darauf ausgelegt sind, bestimmte Wörter oder Themen mit negativen oder positiven Konnotationen zu assoziieren. Dies kann die Objektivität und Zuverlässigkeit des Modells untergraben. Die Herausforderung bei der Abwehr von Poisoning Attacks besteht darin, dass sie im Gegensatz zu direkten Angriffen auf das bereits ausgebildete Modell, das Trainingsmaterial manipulieren.

Datenschutzbedenken entstehen vor allem dann, wenn Sprachmodelle unbeabsichtigt lernen, spezifische Muster oder Informationen aus den Trainingsdaten zu reproduzieren, die als sensibel oder persönlich gelten. Dies kann die Privatsphäre von Individuen gefährden und steht oft in direktem Widerspruch zu Datenschutzbestimmungen wie der Datenschutz-Grundverordnung (DSGVO) in der Europäischen Union (vgl. Europäische Kommission 2016).

3.3 Risiken von KI in der Sozialen Arbeit und Gegenmaßnahmen: Best Practice Hinweise

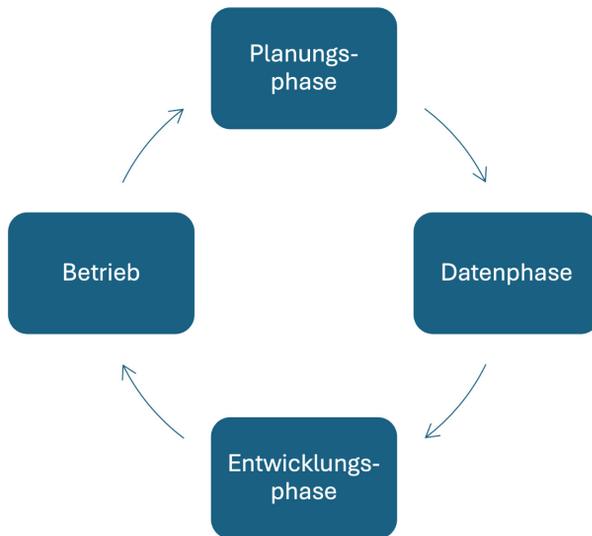
Soziale Berufe verarbeiten äußerst sensible persönliche Daten. Der Einsatz von KI erhöht das Risiko von Datenlecks und Missbrauch. Es ist daher entscheidend, strenge Datenschutzrichtlinien und -technologien zu implementieren, um die Privatsphäre der Klient:innen zu schützen:

- Es sollte nur die absolut notwendige Menge an Daten gesammelt und verarbeitet werden (Datenminimierung).
- Daten sollten anonymisiert werden, um die Identität der Klient:innen zu schützen (Anonymisierung).

Verschiedene Maßnahmen können dazu dienen, KI effektiv vor Manipulationen zu schützen. Die im Folgenden beschriebenen Best Practices skizzieren ein mehrschichtiges Vorgehen, das entlang des gesamten Lebenszyklus eines KI-Modells

angewendet werden sollte. Der Lebenszyklus eines LLM (siehe Abbildung 1) geht von einer Planungsphase aus, an die sich die sogenannte Datenphase anschließt. Die Datenphase umfasst die Sammlung, Aufbereitung und Analyse der relevanten Trainingsdaten. Die darauffolgende Entwicklungsphase schließt die Festlegung von Modellkennwerten wie Architektur und Größe oder die Auswahl eines vortrainierten Modells sowie die Trainingsphase und Validierung ein. Das Modell wird anschließend in Betrieb genommen (vgl. Bundesamt für Sicherheit in der Informationstechnik – BSI 2024).

Abbildung 1: Wesentliche Phasen im Lebenszyklus eines LLM



Quelle: Eigene Darstellung

Sinnvolle Maßnahmen sind demnach insbesondere:

- **Trainingsdaten prüfen und überwachen.** Um Verfälschung der Daten zu verhindern, ist es wichtig, die Integrität, den Ursprung und die Qualität der Trainingsdaten sicherzustellen. Mögliche Maßnahmen sind die regelmäßige Überprüfung und Bereinigung der Daten und der Einsatz von Technologien zur Anomalie-Erkennung, um verdächtige Datenmuster zu identifizieren. Ferner kann die Verwendung von digitalen Signaturen zur Echtheitsüberprüfung von Daten genutzt werden.
- **Sicherheitsprotokolle und Zugriffsmanagement.** Strenge Sicherheitsprotokolle und ein klares Zugriffsmanagement können den unbefugten Zugriff auf KI-Systeme und Daten verhindern. Hierzu können u. a. Zwei-Faktor-Authentifizierung (2FA), rollenbasierte Zugriffskontrollen und regelmäßige Sicher-

heitsüberprüfungen und Audits implementiert werden (vgl. NIST – National Institute of Standards and Technology 2024).

- **Verschlüsselung und sichere Datenübertragung.** Sensible Daten sollten sowohl im Ruhezustand als auch während der Übertragung verschlüsselt werden, um das Risiko von Abhörversuchen und Datendiebstahl zu minimieren. Hierzu können existierende, starke Verschlüsselungsalgorithmen verwendet werden, mit denen eine sichere Ende-zu-Ende-Verschlüsselung für Kommunikationskanäle erreicht werden kann.
- **Transparenz und Nachvollziehbarkeit.** Die Nachvollziehbarkeit von KI-Entscheidungen und die Transparenz der Modelle können dazu beitragen, Manipulationen zu erkennen und zu verhindern. Mögliche Maßnahmen bestehen in der Implementierung von erklärbaren KI-Modellen, der Protokollierung und der Überwachung von Modellentscheidungen sowie in der regelmäßigen Überprüfung und Validierung der Modelle.
- **Notfallpläne und Incident Response.** Um im Falle einer Manipulation schnell und effektiv handeln zu können, müssen Notfallpläne und Verfahren zur Reaktion auf Sicherheitsvorfälle existieren. Sinnvoll ist auch die Schaffung eines sogenannten Incident-Response-Teams, das im Fall von Manipulationen unmittelbar reagieren kann (vgl. Bundesamt für Sicherheit in der Informationstechnik 2017). Zu den Maßnahmen zählen regelmäßige Durchführungen von Sicherheitsübungen und Simulationen. Die detaillierte Dokumentation und Analyse von Sicherheitsvorfällen helfen bei der Vermeidung zukünftiger Vorfälle.
- **Schulung der Mitarbeitenden zu AI Literacy (deutsch: KI-Kompetenz).** Voraussetzung dafür, dass eine Fachkraft souverän mit KI-Systemen und Sprachmodellen unter Berücksichtigung von Risiken arbeiten kann, ist ein Grundwissen über die Funktionsweise der Technik, die sogenannte AI Literacy (vgl. den Beitrag von Löhe in diesem Band). AI Literacy (deutsch: KI-Kompetenz), wie sie von Long und Magerko (2020) oder Faruq, Watkins und Medker (2021) definiert wird, umfasst ein breites, allgemeines Set an Wissen und Fähigkeiten oder Kompetenzen, die Personen benötigen, die mit KI-Technologien interagieren (siehe den Beitrag von Löhe in diesem Band). Dabei handelt es sich auch um eine rechtliche Anforderung: Seit Februar 2025 sind Anbieter und Betreiber von KI-Systemen gemäß des EU AI Acts verpflichtet, sicherzustellen, dass alle Personen, die mit dem Betrieb und der Nutzung dieser Systeme befasst sind, über ausreichende KI-Kompetenz verfügen und dass entsprechende Maßnahmen zu deren Qualifizierung ergriffen werden (siehe den Beitrag von Dötterl zur KI-Verordnung in diesem Band).

4 Rechtlicher Rahmen für den Einsatz von KI-Sprachmodellen

Der rechtliche Rahmen für den Einsatz von KI ist durch ein Zusammenspiel aus nationalen Gesetzen, internationalen Abkommen und Richtlinien geprägt, die darauf abzielen, den Einsatz von KI-Systemen zu regulieren, während sie gleichzeitig Innovationen fördern und ethische Standards sowie die Einhaltung von Menschenrechten sicherstellen.

Zu den wichtigsten rechtlichen Aspekten, die dabei berücksichtigt werden müssen, zählen Datenschutz (z. B. DSGVO), Haftungsregelung (wer für Schäden verantwortlich ist, die durch KI-Entscheidungen oder -Handlungen verursacht werden), Transparenz, Nichtdiskriminierung und die Gewährleistung menschlicher Aufsicht (z. B. EU AI Act).

Die Schaffung und Nutzung von durch KI generierten Inhalten wirft Fragen bezüglich des Urheberrechts und des geistigen Eigentums auf. Einige Länder beginnen, Richtlinien zu entwickeln, die klären, wie Urheberrechte in Bezug auf KI-generierte Werke behandelt werden sollen. Verschiedene Organisationen und Regierungskörperschaften haben ethische Richtlinien für den Einsatz von KI entwickelt, die Prinzipien wie Transparenz, Fairness und Verantwortlichkeit hervorheben. Ein Beispiel aus der Sozialen Arbeit ist das Potsdamer Memorandum zum Einsatz von KI in der Suchthilfe (vgl. Brandenburgische Landesstelle für Suchtfragen e. V. 2025). Als Teil ihrer digitalen Strategie will die EU KI regulieren, um bessere Bedingungen für die Entwicklung und Nutzung dieser innovativen Technologie zu schaffen (vgl. European Union 2024).

5 Fazit

Der Einsatz von KI steht in einem zunehmend kritischen Verhältnis zum Datenschutz, da KI-Systeme oft auf der Sammlung, Analyse und Verarbeitung großer Mengen von Daten basieren, um zu lernen und Entscheidungen zu treffen. Diese Daten können oft persönliche Informationen enthalten, deren Handhabung durch strenge Datenschutzgesetze wie die Datenschutz-Grundverordnung (DSGVO) in der Europäischen Union reguliert wird. Der Datenschutz fordert daher die Entwicklung und Anwendung von KI in einer Weise, die die Privatsphäre und die Autonomie der Individuen schützt, was den Einsatz von Techniken des datenschutzfreundlichen Maschinellen Lernens und Methoden zur Anonymisierung oder Pseudonymisierung von Daten einschließt.

IT-Sicherheit spielt eine entscheidende Rolle dabei, KI-Anwendungen datenschutzkonform zu gestalten und zu betreiben. Hierbei sind IT-Sicherheitsmaßnahmen Ansätze, die u. a. zur Einhaltung von Datenschutzvorschriften bei der Verwendung von KI unterstützen können. Die Verschlüsselung von Daten sowohl in Ruhe als auch während der Übertragung schützt sensible Informationen vor

unbefugtem Zugriff. Strenge Zugriffskontrollen und Authentifizierungsverfahren gewährleisten, dass nur autorisiertes Personal auf die Daten und KI-Modelle Zugriff hat. Die Anwendung von Techniken zur Anonymisierung oder Pseudonymisierung von Daten hilft, persönliche Informationen unkenntlich zu machen. Dadurch können KI-Systeme für Analysen und Lernprozesse genutzt werden, ohne dass dabei Datenschutzprinzipien verletzt werden.

Durch den Nutzen und die zunehmende Integration von KI in sozialen Berufen entsteht somit ein komplexes Spannungsfeld zwischen der Maximierung der Vorteile durch KI und dem unbedingten Schutz persönlicher Daten, das durch kontinuierliche Forschung, Entwicklung sicherer und transparenter KI-Modelle sowie die Schaffung von rechtlichen Rahmenbedingungen adressiert werden muss.

Literatur

- Lorengel, Andrei / Pelzl, Jan (2017): „Safety and Security for Medical Devices: Analysis and Implementation of a Secure Software Update for Embedded Systems.“ 2nd YRA MedTech Symposium, Young Researchers Academy – MedTech in NRW. Mühlheim a. d. Ruhr.
- Brandenburgische Landesstelle für Suchtfragen e. V. (2025): Potsdamer Memorandum zum Einsatz von KI in der Suchthilfe. <https://www.blsev.de/fachbereiche/digitalisierung/ki-sucht/> (Abfrage: 15.06.2025).
- Bundesamt für Sicherheit in der Informationstechnik – BSI (2024): Generative KI-Modelle – Chancen und Risiken für Industrie und Behörden. https://www.bsi.bund.de/SharedDocs/Downloads/DE/BSI/KI/Generative_KI-Modelle.pdf?__blob=publicationFile&v=5 (Abfrage: 15.06.2025).
- Bundesamt für Sicherheit in der Informationstechnik (2017): BSI-Grundschatz. https://www.bsi.bund.de/DE/Themen/Unternehmen-und-Organisationen/Standards-und-Zertifizierung/IT-Grundschatz/it-grundschatz_node.html/ (Abfrage: 15.06.2025).
- Bundesamt für Sicherheit in der Informationstechnik (2017): IT-Grundschatz Methodik – BSI-Standard 200–2. November.
- Bundesamt für Sicherheit in der Informationstechnik. 2017. Managementsysteme für Informationssicherheit (ISMS) – BSI-Standard 200–1. Oktober.
- Carlini, Nicholas / Tramer, Florian / Wallace, Eric / Jagielski, Matthew / Herbert-Voss, Ariel / Lee, Katherine / Roberst, Adam / Brown, Tom / Song, Dawn / Erlingsson, Ulfar / Oprea, Alina / Raffel, Colin (2021): Extracting Training Data from Large Language Models. <https://arxiv.org/abs/2012.07805>
- Casal, Elliot J. / Kessler, Matt (2023): Can linguists distinguish between ChatGPT/AI and human writing? A study of research ethics and academic publishing. In: Research Methods in Applied Linguistics 2. <https://doi.org/10.1016/j.rmal.2023.100068>
- Der Paritätische Gesamtverband (2024): Künstliche Intelligenz in der Sozialen Arbeit. Eine Textsammlung aus der gleichnamigen Veranstaltungsreihe 2023. https://www.der-paritaetische.de/fileadmin/user_upload/Schwerpunkte/Digitalisierung/doc/ki/KI_Textsammlung_Update2024_final.pdf (Abfrage 15.06.2025).
- Diakonie Michaelshofen (o. J.): Bitte auf diesen Link klicken – Warum Cyberkriminalität auch soziale Unternehmen treffen kann. https://www.diakonie-michaelshoven.de/blog?tx_t3extblog_blogsysteem%5Baction%5D=show&tx_t3extblog_blogsysteem%5Bcontroller%5D=Post&tx_

- t3extblog_blogsysteem%5Bpost%5D=64&cHash=d14fd486539ff90113d8b93eb7ced476 (Abfrage: 15.06.2025).
- Europäische Kommission (2016): „Datenschutz-Grundverordnung.“ Verordnung (EU) 2016/679 des Europäischen Parlaments und des Rates vom 27. April 2016 zum Schutz natürlicher Personen bei der Verarbeitung personenbezogener Daten, zum freien Datenverkehr und zur Aufhebung der Richtlinie 95/46/EG. 27. April.
- European Union (2024): Artificial Intelligence Act (Regulation (EU) 2024/1689). Official Journal (OJ) of the European Union.
- Faruqe, Farhana/Watkins, Ryan/Medsker, Larry (2021): Competency model approach to AI literacy: research-based path from initial framework to model. <https://arxiv.org/pdf/2108.05809.pdf>
- Fraunhofer (o. J.): Grundlagen der IT-Sicherheit – Kryptographie. <https://www.cybersicherheit.fraunhofer.de/de/kursangebote/basics-it-sicherheit/grundlagen-der-it-sicherheit-kryptographie.html> (Abfrage: 15.06.2025).
- Goldberg, Brigitta (2021): Schweigepflicht und Datenschutz in der Sozialen Arbeit und Beratung. Bochum: Ev. Hochschule Rheinland-Westfalen Lippe.
- IETF – Internet Engineering Task Force (1999): RFC 2631 (Proposed Standard).
- International Organization for Standardization (2022): „ISO/IEC 27001:2022 Information security, cybersecurity and privacy protection – Information security management systems – Requirements.“ ISO27001. International Organization for Standardization.
- Jonas, Markus (2024): Wenn die IT nicht mehr geht. Caritas in NRW. <https://www.caritas-nrw.de/magazin/2024/artikel/wenn-die-it-nicht-mehr-geht-9419c1f5-2b2e-4b94-8ae4-163d59e468be> (Abfrage: 15.06.2025).
- Kelbert, Patricia/Sieber, Julien/Jöckel, Lisa (2023): Was sind Large Language Models? Und was ist bei der Nutzung von KI-Sprachmodellen zu beachten? Blog des Fraunhofer-Institut für Experimentelles Software Engineering. https://www.iese.fraunhofer.de/blog/large-language-models-ki-sprachmodelle/?utm_source=chatgpt.com (Abfrage: 15.06.2025).
- Kreidenweis, Helmut/Diepold, Maria (2024): Studie. Künstliche Intelligenz in der Sozialwirtschaft. Forschungsbericht. Eichstätt-Ingolstadt: Katholische Universität. Arbeitsstelle für Sozialinformatik.
- Liebsch, Katharina/Degel, Alexander (2025): Einleitung – Digitalität und Ambiguität im Feld der Sozialen Arbeit. In: Degel, Alexander/Liebsch, Katharina (Hrsg.): Digitalität und Ambiguität. Organisationskulturen der Sozialen Arbeit unter Druck. Weinheim und Basel: Beltz Juventa, S. 7–21.
- Linnemann, Gesa/Löhe, Julian/Rottkemper, Beate (2023): Bedeutung von Künstlicher Intelligenz in der Sozialen Arbeit. Eine exemplarische arbeitsfeldübergreifende Betrachtung des Natural Language Processing (NLP). In: Soziale Passagen 15, S. 197–201. <https://doi.org/10.1007/s12592-023-00455-7>
- Long, Duri/Magerko, Brian (2020): What is AI Literacy? Competences and Design Considerations. Proceedings of the 2020 CHI conference on human factors in computing systems, ACM, S. 1–16. <https://doi.org/10.1145/3313831.337672>
- Maus, Nathalie/Chao, Patrick/Wong, Eric/Gardner, Jacob (2023): „Black Box Adversarial Prompting for Foundation Models. CoRR abs/2302.04237. <https://doi.org/10.48550/arXiv.2302.04237>
- Muthalagu, Raja/Malik, Jasmita/Pawar, Pranav M. (2025): Detection and prevention of evasion attacks on machine learning models. In: Expert Systems with Applications 266. <https://doi.org/10.1016/j.eswa.2024.126044>
- NIST – National Institute of Standards and Technology (2001): Advanced Encryption Standard.
- NIST – National Institute of Standards and Technology (2023): Artificial Intelligence Risk Management Framework (AI RMF 1.0). Januar. <https://doi.org/10.6028/NIST.AI.100-1>

- NIST – National Institute of Standards and Technology (2024): Special Publication 800–63. Digital Identity Guidelines. Oktober. <https://www.nist.gov/itl/smallbusinesscyber/guidance-topic/multi-factor-authentication> (Abfrage: 15.06.2025).
- Steiner, Olivier/Tschopp, Dominik (2022): Künstliche Intelligenz in der Sozialen Arbeit. In: Sozial Extra 46, S. 466–471.
- Open Web Application Security Project. 2025. LLM01: 2025 Prompt Injection. 14. 1. <https://genai.owasp.org/llmrisk/llm01-prompt-injection/> (Abfrage: 15.06.2025).
- Paar, Christof/Pelzl, Jan/Güneysu, Tim (2024): Understanding Cryptography – From Established Symmetric and Asymmetric Ciphers to Post-Quantum Algorithms. Berlin: Springer.
- Rawat, Danda B. (2021): Secure and trustworthy machine learning/artificial intelligence for multi-domain operations. In: Proceedings 11746. <https://doi.org/10.1117/12.2592860>
- Wallace, Eric/Zhao, Tony Z./Feng, Shi/Singh, Sameer (2020): Concealed Data Poisoning Attacks on NLP Models. NAACL 2021. <https://doi.org/10.48550/arXiv.2010.12563>
- Weidinger, Laura/Uesato, Jonathan/Rauh, Maribeth/Griffin, Conor/Huang, Po-Sen/Mellor, John/Glaese, Amelia/Cheng, Myra/Balle, Borja/Kasirzadeh, Atoosa/Biles, Courtney/Brown, Sasha/Kenton, Zac/Hawkins, Will/Stepleton, Tom/Birhane, Abeba/Hendricks, Lisa Anne/Rimell, Laura/Isaac, Williams [...] Claims, Less (2022): Taxonomy of Risks posed by Language Models. FAccT'22: Proceedings of the 2022 ACM Conference on Fairness, Accountability, and Transparency, S. 214–229. <https://doi.org/10.1145/3531146.353308>

EU AI Act und Soziale Arbeit: Die KI-Verordnung und ihre Auswirkungen¹

Sebastian Dötterl

Abstract: Der Aufsatz bietet eine erste Orientierung für den rechtssicheren KI-Einsatz in der Sozialen Arbeit. Er erklärt die Vorgaben der EU-KI-Verordnung (AI Act) für Anwendungsbeispiele wie Chatbots, Prognosesystemen und Emotionserkennung. Zunächst wird untersucht, wann überhaupt Künstliche Intelligenz im Rechtssinn vorliegt. Im Fokus steht sodann der risikobasierte Ansatz der Verordnung: von Verboten (z. B. Ausnutzung von Vulnerabilität) über strenge Anforderungen an Hochrisiko-Systeme bis hin zu Transparenzpflichten und dem Aufbau von KI-Kompetenz. Praktische Handlungsempfehlungen runden den Beitrag ab.

Keywords: KI-Verordnung, Hochrisiko, Transparenzpflichten, Grundrechte, KI-Kompetenz

1 Einführung

Ziel dieses Beitrags ist es, die Auswirkungen der KI-Verordnung (KI-VO, englisch: AI Act) auf den Einsatz von Künstlicher Intelligenz (KI) in der Sozialen Arbeit darzustellen.²

Diese Auswirkungen sollen insbesondere an Anwendungsszenarien gezeigt werden, die bereits erprobt werden oder sich in der fachlichen Diskussion befinden:

- Chatbots für Beratung, z. B. in der Ausbildung und Unterstützung von Fachkräften, aber auch im direkten Kontakt mit Klient:innen, wie in der Suchtbe-

1 © Die Autor:innen 2025, veröffentlicht im Verlag Beltz Juventa, Open Access: CC BY 4.0
Gesa Linnemann/Julian Löhe/Beate Rottkemper (Hg.), *Künstliche Intelligenz in der Sozialen Arbeit*
10.3262/978-3-7799-8562-4_016

2 Alle im Folgenden genannten Vorschriften und rechtlich nicht verbindlichen Erläuterungen (sogenannte Erwägungsgründe, ErwG) bezeichnen solche der KI-VO. Der Beitrag verweist an unterschiedlichen Stellen auf die Erwägungsgründe der KI-VO, die im Volltext hier einsehbar sind: https://eur-lex.europa.eu/legal-content/DE/TXT/HTML/?uri=OJ:L_202401689#pbl_1 (Aufruf: 14.04.2025).

ratung oder als psychologische Ersthilfe (vgl. Linnemann/Löhe/Rottkemper 2024; Lehmann 2025, Li et al. 2023),

- KI-Systeme zur Emotions- und Verhaltensanalyse durch Sprach- und Texterkennung (vgl. Kapoor/Verma 2024),
- emergenzbasierte KI-Systeme, die aus der Analyse vielfältiger Einzelfaktoren eine Prognose ableiten, z. B. das Entstehen von Risiken wie Kindeswohlgefährdung (vgl. Plafky/Frischhut 2025),
- KI-gestützte Dokumentationssysteme, etwa zur Erstellung von Aktennotizen (siehe den Beitrag von Plafky et al. in diesem Band), Tagesdokumentationen (siehe den Beitrag von Holz/Fellmann/Schmidt in diesem Band) oder Protokollen (vgl. Feist-Ortmanns/Sauer/Brinkmann 2025).

Die Beispiele zeigen, dass der Einsatz von KI in der Sozialen Arbeit in vielfältigen Einsatzszenarien geprüft oder bereits erprobt wird. Aufgrund des Umgangs mit vulnerablen Gruppen handelt es sich um einen grundrechtssensiblen Bereich, weshalb beim Einsatz neuer Technologien auch Vorsicht angezeigt ist. Es verwundert daher nicht, dass erste Erprobungen mit kritischen Diskursen in der Fachwelt begleitet werden, die vor allem zweierlei Ausprägung haben: zum einen die Ausprägung, inwiefern der Einsatz von KI die Fachlichkeit der Sozialen Arbeit bedroht oder verändert (siehe den Beitrag von Löhe in diesem Band), und zum anderen rechtliche Herausforderungen. Neben datenschutzrechtlichen Bestimmungen (siehe den Beitrag von Pelzl in diesem Band) spielt die KI-Verordnung in diesem Zusammenhang eine wichtige Rolle. Zugleich ist die Verordnung noch sehr neu und fundierte sowie systematische Einordnungen der Bedeutung für die Soziale Arbeit gibt es bisher nicht. Dieser Artikel soll diese Lücke schließen und einen Beitrag dazu liefern, Sozialarbeitenden eine erste Einschätzung zu ermöglichen.

1.1 Was ist die KI-VO?

Mit der KI-VO hat die Europäische Union einen umfassenden Rechtsrahmen für KI-Systeme geschaffen. Als EU-Verordnung entfaltet sie unmittelbare Rechtswirkung in allen Mitgliedstaaten und regelt einheitlich die Entwicklung, das Inverkehrbringen und die Verwendung von KI-Systemen. Dabei verfolgt sie den doppelten Zweck, den Binnenmarkt zu harmonisieren, indem einheitliche Regeln und Standards geschaffen werden, die in allen Mitgliedsstaaten gleichermaßen gelten. Zugleich ist ein hohes Grundrechtsschutzniveau zu gewährleisten (Art. 1 Abs. 1 und ErwG 1, 7, 8).

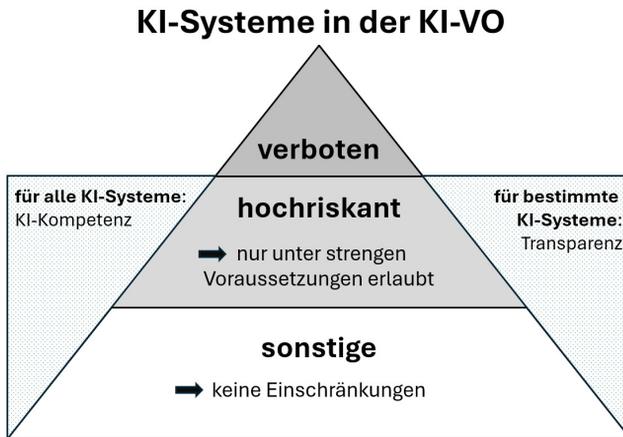
Das Kernkonzept der KI-VO ist ein risikobasierter Regelungsansatz (ErwG 26, 27). Je nach Gefährdungspotenzial werden KI-Systeme (dazu unter 2.) unterschiedlichen Anforderungsstufen zugeordnet:

- Bringt der Einsatz von KI-Systemen unannehmbare Risiken mit sich, so ist er verboten (dazu unter 3.)
- KI-Systeme mit hohem Risiko sind erlaubt, müssen aber strenge Anforderungen erfüllen (dazu unter 4.)
- Für sonstige KI-Systeme gelten keine besonderen Regeln.

Ob Hochrisiko- oder sonstiges KI-System: Für bestimmte KI-Systeme wie etwa Chatbots gelten unabhängig von ihrer Risikoeinstufung Transparenzpflichten (dazu unter 5.). Zudem sind alle Anbieter und Betreiber von KI-Systemen verpflichtet, die KI-Kompetenz ihrer Mitarbeitenden sicherzustellen (dazu unter 6.). Bei Verstößen gegen Vorgaben der KI-VO drohen vor allem Geldbußen (dazu unter 7.)

Die nachfolgende Abbildung visualisiert die Regulierungssystematik der KI-VO:

Abbildung 1: KI-Systeme in der KI-VO



Quelle: Eigene Darstellung

Außerdem ist zu beachten, dass die KI-VO den bereits bestehenden Rechtsrahmen ergänzt, aber nicht ersetzt. Insbesondere bleiben die Vorgaben der Datenschutz-Grundverordnung (DSGVO) vollumfänglich anwendbar (Art. 2 Abs. 7 KI-VO). Sofern also mittels KI-Systemen personenbezogene Daten verarbeitet werden, sind die Vorgaben der DSGVO zu beachten, die in der Praxis eine weitere große Herausforderung für den KI-Einsatz darstellen werden (vgl. Heidrich 2024). Dies kann an dieser Stelle nicht vertieft werden (zur Vertiefung siehe den Beitrag von Pelzl in diesem Band). Es sei hier nur darauf hingewiesen, dass die Verarbeitung sensibler Daten wie z. B. Gesundheitsdaten besonderen Einschränkungen unterliegen (Art. 9 DSGVO), ebenso die Übermittlung perso-

nenbezogener Daten in Drittländer, z. B. durch eine Eingabe solcher Daten in ein im Ausland gehostetes KI-System wie beispielsweise ChatGPT (Art. 44 DSGVO).

1.2 Zeitplan für das Inkrafttreten der KI-VO

Die KI-VO tritt gestaffelt in Kraft (Art. 113):

- Seit dem 2. Februar 2025 gelten die Verbote bestimmter KI-Praktiken sowie die Vorgaben zur KI-Kompetenz.
- Die meisten Vorschriften werden ab dem 2. August 2026 wirksam, darunter die für die Soziale Arbeit besonders relevanten Anforderungen an Hochrisiko-KI-Systeme und die Transparenzpflichten.

2 Was ist ein „KI-System“ im Sinne der KI-VO?

2.1 Definition eines KI-Systems

Die KI-VO enthält eine sehr weit gefasste Definition eines KI-Systems, der auch viele halbwegs moderne Softwaresysteme unterfallen dürften (vgl. Hacker 2024a). Erschließen lässt sich die Definition vor allem anhand zweier Kernmerkmale:

- Das KI-System muss zum einen „in unterschiedlichem Grade autonom“ arbeiten können, d. h. „bis zu einem gewissen Grad unabhängig von menschlichem Zutun“ (ErwG 12). Nach den insoweit von der EU-Kommission veröffentlichten, unverbindlichen Leitlinien reicht bereits ein sehr geringer Grad an Autonomie aus, sodass diesem Merkmal keine eigenständige Bedeutung zukäme (Europäische Kommission 2025a, Punkte 14 ff.). Zum anderen soll nach den Erwägungsgründen kein KI-System vorliegen, wenn das System „auf ausschließlich von natürlichen Personen definierten Regeln für das automatische Ausführen von Operationen“ (ErwG 12) beruht. Vorgeschlagen wird daher, zwischen „autonom“ und „automatisch“ zu unterscheiden (vgl. Wachter/Leeb 2024). Ein automatisches System, das rein menschlich definierten Regeln folgt, wäre demnach kein KI-System, auch wenn es einen hohen Komplexitätsgrad erreicht (vgl. Roth-Isigkeit 2024).
- Die „Fähigkeit, abzuleiten“ ist das zweite Kernmerkmal von KI-Systemen. Bei diesem Merkmal geht es darum, dass das System aus Eingaben Ausgaben erzeugt, seien es Voraussagen, Inhalte oder Empfehlungen. Es muss sich um mehr handeln als „einfache Datenverarbeitung“. Als Systeme, die „ableiten“ und daher KI-Systeme sein können, werden ausdrücklich Ansät-

ze des Maschinellen Lernens sowie logik- und wissensgestützte Konzepte genannt (ErwG 12). Keine KI-Systeme sollen nach den Leitlinien hingegen folgende Ansätze sein: Systeme zur Verbesserung mathematischer Optimierung, einfache Datenverarbeitung, klassische Heuristik und einfache Vorhersagesysteme (Punkte 40 ff.). Eine widerspruchsfreie und rechtssichere Abgrenzung erscheint auf Grundlage der – ohnehin unverbindlichen – Leitlinien nicht möglich, was hier nicht vertieft dargestellt werden kann.

2.2 Abgrenzung KI-Modell

Zu unterscheiden ist das KI-System zudem vom KI-Modell (vgl. Engel 2024).

- Das KI-System ist die konkrete, einsatzfähige Softwareanwendung (z. B. „ChatGPT“),
- in die ein oder mehrere KI-Modelle eingebettet sein können (z. B. das Modell „GPT-4“).

Die Regelungen der KI-VO, die sich auf KI-Modelle (insbesondere KI-Modelle „mit allgemeinem Verwendungszweck“, Art. 51 ff.) beziehen, bleiben nachfolgend außer Betracht.

2.3 Zwischenfazit

Die weite Definition des KI-Systems, der zahlreiche Bestandssysteme unterfallen dürften, ist nicht zufriedenstellend (vgl. Biallaß 2024) und führt zu erheblicher Rechtsunsicherheit, die auch die inzwischen von der Kommission nach Art. 96 Abs. 1 lit. f. vorgelegten Leitlinien nicht beseitigt hat. Erst die künftige Rechtsprechung wird somit Klarheit schaffen. Bis dahin wird teilweise empfohlen, im Zweifelsfall vom Vorliegen eines KI-Systems auszugehen (vgl. Hense / Mustać 2025).

Wendet man die genannten Abgrenzungskriterien auf die Eingangsbeispiele an, ergibt sich: Soweit Chatbots oder Systeme zur Emotionsanalyse oder zur Dokumentation auf generativer KI basieren, handelt es sich eindeutig um KI-Systeme.

Auch emergenzbasierte Prognosesysteme, die auf Maschinellern Lernen beruhen und komplexe Risikoeinschätzungen vornehmen, dürften der Definition eines KI-Systems unterfallen.

3 Verbotene KI-Systeme

Die KI-VO definiert in Art. 5 bestimmte KI-Praktiken als verboten. Diese Verbote gelten absolut und können nicht durch Schutzmaßnahmen oder Einwilligungen der Betroffenen überwunden werden. Die Verbote gelten sowohl für das In-Verkehr-Bringen als auch die Inbetriebnahme und die Verwendung entsprechender KI-Systeme. Für die Soziale Arbeit können folgende Fallgruppen relevant sein:

3.1 Ausnutzung von Vulnerabilität

Verboten sind KI-Systeme, die die Vulnerabilität oder Schutzbedürftigkeit einer Person oder bestimmter Personengruppen ausnutzen (Art. 5 Abs. 1 lit. b.). Diese Vulnerabilität kann sich aus Alter, Behinderung oder der sozialen oder wirtschaftlichen Situation ergeben. Das Verbot gilt, wenn diese Vulnerabilität ausgenutzt wird, und aus der dadurch beabsichtigten oder daraus folgenden Verhaltensänderung ein erheblicher Schaden entsteht oder droht.

Es ist allerdings nicht verboten, sondern sogar geboten, die besonderen Bedürfnisse vulnerabler Gruppen positiv zu berücksichtigen (ErwG 48). Auch sollen gemäß ErwG 29 die hier und unter der folgenden Ziffer erläuterten Verbote keine Auswirkungen haben auf den Einsatz von KI-Systemen bei medizinischen Behandlungen, etwa der psychologischen Behandlung einer psychischen Krankheit oder der physischen Rehabilitation, wenn dieser Einsatz gemäß den geltenden Rechtsvorschriften und medizinischen Standards erfolgt (vgl. aber unten IV. 2 a. zur Einstufung als Hochrisiko-KI-System).

KI-Systeme, die vulnerable Gruppen nicht ausnutzen, sondern unterstützen sollen, fallen nach den hierzu vorgelegten, unverbindlichen Leitlinien der EU-Kommission nicht unter das Verbot (Europäische Kommission, 2025b). Demnach sind KI-Anwendungen, die Kinder, ältere oder sozial benachteiligte Personen unterstützen sollen, nicht verboten, solange keine Gefahr substanzieller Schäden besteht. Dazu zählen z. B. lernfördernde KI-Systeme für Kinder oder persönliche KI-Assistenten zur Unterstützung älterer Menschen oder von Menschen mit Behinderungen (Punkt 121).

Kritisch zu prüfen ist beim Umgang mit vulnerablen Gruppen jedoch stets, ob das KI-System eine wesentliche Verhaltensänderung verursachen kann, die zu erheblichen Schäden führt oder mit hinreichender Wahrscheinlichkeit führen kann.

Als verbotene Beispiele nennen die Leitlinien etwa

- „vermenschlichte“ KI-Systeme, die bei Kindern zu einer emotionalen Abhängigkeit und zu einer Behinderung ihrer sozialen und emotionalen Entwicklung führen können (Punkt 116),

- KI-Systeme, die eigentlich zur Unterstützung von älteren Menschen oder von Menschen mit geistiger Behinderung gedacht sind, aber diese auch manipulieren können, indem sie sie dazu bewegen, übertriebene und nutzlose Produkte zu kaufen (Punkte 117/118),
- KI-gestützte Chatbots, die sozial oder wirtschaftlich benachteiligte Gruppen ansprechen und – etwa durch Angstnarrative oder manipulative Angebote – deren bestehende Vulnerabilität verstärken können, was in bestimmten Fällen zu Angst, Depression oder einem Gefühl der Hilflosigkeit führen kann (Punkt 119).

Insofern ist beim Einsatz von KI-Systemen in der Sozialen Arbeit eingehend zu prüfen, inwiefern ein Schaden entstehen kann und wann die Gefahr eines „Ausnutzens“ von Vulnerabilität besteht. Dies könnte auch das „Nudging“ zu einem potenziell schadensträchtigen Verhalten betreffen (Punkt 108). Eine Kontrollfrage zum „Ausnutzen“ könnte lauten, ob die wesentliche Verhaltensänderung entfallen würde, wenn die beeinflusste Person nicht besonders schutzbedürftig gewesen wäre (vgl. Heinze/Engel 2025).

3.2 Manipulative Techniken

Die KI-VO verbietet KI-Systeme, die durch manipulative oder täuschende Techniken das Verhalten von Personen wesentlich verändern (Art. 5 Abs. 1 lit. a.). Das Verbot erfasst dabei

- die unterschwellige Beeinflussung außerhalb des Bewusstseins einer Person oder
- absichtlich manipulative oder täuschende Techniken.

Voraussetzung ist jeweils, dass es bezweckt oder bewirkt wird, die Fähigkeit der Person, eine fundierte Entscheidung zu treffen, deutlich zu beeinträchtigen, und durch die so manipulierte Entscheidung ein erheblicher Schaden droht.

Die Leitlinien (Punkte 127 ff.) versuchen eine Grenze zu ziehen zwischen

- verbotener Manipulation, die durch verdeckte oder absichtlich täuschende Techniken die Entscheidungsfreiheit untergräbt, und
- erlaubter Überzeugung, die transparent erfolgt und informierte und autonome Entscheidungen ermöglicht.

In der Sozialen Arbeit ist das Verbot besonders relevant, da die Klient:innen oft vulnerable Personen sind (vgl. bereits oben Ziffer 1) und deshalb anfälliger für Manipulationen sein können. Deshalb sollten z. B. Chatbots kritisch auf mögliches Schadenspotenzial überprüft werden. Als Beispiel kommt etwa ein Chatbot infra-

ge, der selbstverletzendes Verhalten wie Ritzen etc. thematisiert oder abfragt, und – unbeabsichtigt – dadurch neues selbstschädigendes Verhalten triggern könnte.

3.3 Social Scoring

Die KI-VO verbietet ferner KI-Systeme zur sozialen Bewertung (Social Scoring) von Personen oder Gruppen, wenn diese zu bestimmten Benachteiligungen führen (Art. 5 Abs. 1 lit. c.). Soziale Bewertung heißt, dass sich die Bewertung auf die allgemeine Einstufung einer Person oder einer Gruppe von Personen als mehr oder weniger „gut“ oder „schlecht“ bezieht (vgl. Martini/Wendehorst 2024, Art. 5 Rn. 66). Schulbeispiele, wie sie aus China berichtet werden, wären die Schlechterstellung einer Person bei der Zuteilung von Bildungschancen oder sogar Fahrkarten auf Grundlage der Tatsache, dass diese Person sich etwa nicht an Verkehrsregeln gehalten oder an verbotenen Demonstrationen teilgenommen hat (Martini/Wendehorst 2024). Der EU-Verordnungsgeber sieht in solchen Systemen die Gefahr einer Verletzung der Menschenwürde und des Diskriminierungsverbots (ErwG 31).

Verboten ist zum einen, wenn die soziale Bewertung zu Benachteiligungen führt, die in keinem Zusammenhang zu den ursprünglichen Bewertungskriterien stehen (Art. 5 Abs. 1 lit. c. (i)). Dieses Verbot könnte laut den Leitlinien eingreifen, wenn ein KI-System das Freizeitverhalten, Social-Media-Daten oder die Nationalität des:der Ehepartner:in auswerten würde, um daraus Rückschlüsse auf dessen:deren Anspruch auf Sozialleistungen zu ziehen (Punkt 166).

Ebenfalls verboten sind soziale Bewertungen, die zu Benachteiligungen führen, die in Hinblick auf das soziale Verhalten oder dessen Tragweite ungerechtfertigt oder unverhältnismäßig sind (Art. 5 Abs. 1 lit. c. (ii)). Dies könnte für KI-Systeme relevant sein, die aus begrenzten Daten weitreichende, nachteilige Schlüsse ziehen. Die Leitlinien nennen als Beispiel ein KI-System, das eine Kindeswohlgefährdung schon deshalb annimmt, weil die Eltern gelegentlich Arzttermine versäumen oder eine Verkehrsordnungswidrigkeit begangen haben (Punkt 167).

Ein KI-System, das sozialrechtliche Ansprüche von Antragsstellenden prüft und darüber entscheidet (wie z. B. bei Jugendhilfesanträge nach dem SGB VIII) oder selbstständig aufgrund von Daten Jugendhilfemaßnahmen für einen Fall bestimmt, würde zunächst vermutlich unter das grundsätzliche Verbot der automatisierten Entscheidung nach Art. 22 DSGVO fallen. Bei der Prüfung, ob zugleich das Verbot des Social Scoring nach der KI-VO eingreift, wäre zum einen zu untersuchen, ob die durch das System herangezogenen Bewertungskriterien in einem Zusammenhang mit der benachteiligenden Entscheidung stehen. Zum anderen wäre zu untersuchen, ob die Benachteiligung verhältnismäßig ist. Werden nur die gesetzlich vorgesehenen Beurteilungskriterien z. B. zur Beurteilung der Bedürftigkeit herangezogen, dürfte das Verbot nicht eingreifen, wie sich auch aus

ähnlichen Beispielen in den Leitlinien (Punkt 177) ableiten lässt (siehe ebenso die Ausführungen zum „Zugang zu grundlegenden Diensten und Leistungen“ im Abschnitt zu Hochrisiko-KI-Systemen weiter unten in diesem Beitrag).

3.4 Emotionserkennung

Die KI-VO verbietet grundsätzlich KI-Systeme zur Emotionserkennung am Arbeitsplatz und in Bildungseinrichtungen (Art. 5 Abs. 1 lit. f.). Wie die Leitlinien unter Verweis auf Art. 3 Nr. 39 ausführen, geht es dabei um die Emotionserkennung auf Grundlage biometrischer Daten, woran es etwa bei Sprachmustern in einem rein textbasierten Chat fehlt (Punkt 251).

Hintergrund des Verbots ist, dass sich Gefühlsausdrücke je nach Kultur, Situation und selbst bei ein und derselben Person erheblich unterscheiden können. Zudem soll das Verbot vor einer missbräuchlichen Ausnutzung des Machtungleichgewichts in Arbeits- und Bildungskontexten schützen (ErwG 44).

Eine für die Soziale Arbeit relevante Ausnahme von diesem Verbot gilt für KI-Systeme, die aus Sicherheitsgründen oder aus medizinischen Gründen eingesetzt werden. In den Erwägungsgründen werden als Beispiel für medizinische Gründe „therapeutische Zwecke“ genannt (ErwG 44), was die Leitlinien (Punkt 257) sehr eng so interpretieren wollen, dass es dabei um die Verwendung von Medizinprodukten mit CE-Siegel gehen muss. Zulässige medizinische Zwecke könnten unter dieser Voraussetzung etwa die Diagnose psychischer Erkrankungen sein oder die Krisenintervention bei akuter Selbst- oder Fremdgefährdung. Nicht vom medizinischen Ausnahmetatbestand gedeckt und damit am Arbeitsplatz oder in Bildungseinrichtungen verboten wären laut Leitlinie KI-Systeme, die lediglich den Stresslevel oder das allgemeine Wohlbefinden überprüfen sollen.

Auch bei Vorliegen medizinischer Gründe stellt sich ein weiteres Problem, wenn die Emotionserkennung nicht nur Klient:innen, sondern „am Arbeitsplatz“ auch Personal wie etwa Therapeut:innen oder Betreuer:innen betrifft. Laut Leitlinien (Punkt 270) müssen die Arbeitgebenden nach Möglichkeit die Erfassung von Emotionen von Arbeitskräften vermeiden. Eine Einwilligung in eine Emotionserkennung am Arbeitsplatz ist nicht möglich (vgl. Drilling/Musiol 2024). Bei KI-Systemen, die potenziell die Emotionen aller beteiligten Personen erfassen (z. B. Analyse von Gesprächssituationen), sollte daher möglichst sichergestellt werden, dass die Emotionen des Personals nicht erkannt werden. Ist dies nicht möglich, so dürfen daraus jedenfalls keine Nachteile für das Personal erwachsen (Punkt 270).

3.5 Biometrische Kategorisierung

Schließlich sind auch bestimmte KI-Systeme zur biometrischen Kategorisierung von Personen verboten, nämlich wenn anhand biometrischer Daten sensible Merkmale erschlossen werden sollen, z. B. religiöse Überzeugungen, sexuelle Orientierung oder die „Rasse“ (Art. 5 Abs. 1 lit. g.). Biometrische Daten wie z. B. Fingerabdrücke oder Gesichtsbilder dürfen somit nicht von einem KI-System zu Rückschlüssen auf solche sensiblen Persönlichkeitsmerkmale genutzt werden. Dieses Verbot könnte etwa relevant werden, wenn ein KI-gestütztes Matching von Kindern mit möglichen Pflegefamilien auf solche biometrischen Kategorisierungen zurückgreifen würde.

3.6 Einordnung der Fallbeispiele

Für die eingangs genannten Anwendungsbeispiele ergibt sich:

- Chatbots in der Beratung müssen vor allem die Verbote manipulativer Techniken und der Ausnutzung von Vulnerabilität beachten. Hier muss vor allem sichergestellt sein, dass die Chatbots nicht zu einem schädigenden Verhalten anleiten oder motivieren können.
- KI-Systeme zur Emotionserkennung sind am Arbeitsplatz und in Bildungseinrichtungen grundsätzlich verboten. Eine Ausnahme gilt, wenn Sicherheits- oder medizinische Gründe den Einsatz rechtfertigen.
- Emergenzbasierte Prognosesysteme müssen darauf hin geprüft werden, ob sie ein verbotenes Social Scoring darstellen.
- Die KI-gestützte bloße Erstellung von Dokumentation wie Aktennotizen oder Protokollen ohne weitergehende Bewertungen dürfte von den Verboten kaum betroffen sein.

4 Hochrisiko-KI-Systeme

Nicht verboten, aber nur unter strengen Voraussetzungen erlaubt, sind KI-Systeme mit hohem Risiko.

4.1 Anforderungen an Hochrisiko-KI-Systeme

Die Einstufung als Hochrisiko-KI-System löst umfangreiche Pflichten aus, die hier nur gestreift werden können. Vor allem die Anbieter von Hochrisiko-Systemen (also in erster Linie Softwareanbieter, vgl. Art. 3 Nr. 3) sind stark gefordert.

Aber auch Betreiber (also beruflich Nutzende, vgl. Art. 3 Nr. 4) sollten sich über ihre Pflichten informieren.

Vorgeschrieben sind beispielsweise ein Risikomanagementsystem (Art. 9) sowie eine bestimmte Qualität der Trainings-, Validierungs- und Testdatensätze (Art. 10, siehe im Detail Hense/Mustać 2025), es gibt Dokumentations- und Aufzeichnungspflichten (Art. 11, 12) und ein angemessenes Maß an Genauigkeit, Robustheit und Cybersicherheit ist sicherzustellen (Art. 15). Vor der Inbetriebnahme muss zudem eine Konformitätsbewertung durchgeführt werden (Art. 43).

Betreiber sind insbesondere dafür verantwortlich, dass die Systeme gemäß Anleitung verwendet werden (Art. 26 Abs. 1) und eine menschliche Aufsicht sichergestellt ist (Art. 14, Art. 26 Abs. 2).

Vorsicht ist geboten, wenn eine nicht für eine Hochrisiko-Verwendung gedachte KI-Anwendung dennoch für einen solchen Zweck benutzt wird, z. B. durch die Nutzung von KI-Tools auf privaten Geräten („Schatten-KI“, hierzu Braegelmann 2024). Vor solch einer Praxis ist schon wegen der erheblichen datenschutzrechtlichen Risiken zu warnen (siehe den Beitrag von Plafky et al. in diesem Band). Zudem könnte dies dazu führen, dass der Betreiber in die Anbieterpflichten hineinrutscht (Art. 25 Abs. 1 lit. c.). Die gleiche Konsequenz könnte unter Umständen drohen, wenn ein Hochrisiko-KI-System, z. B. durch weiteres Trainieren oder Finetuning, wesentlich verändert wird (Art. 25 Abs. 1 lit. b.).

Betreiber, bei denen es sich um Einrichtungen des öffentlichen Rechts oder private Einrichtungen, die öffentliche Dienste erbringen, handelt, müssen zudem eine Grundrechte-Folgenabschätzung durchführen (Art. 27).

Die Implementierung von Hochrisiko-KI-Systemen erfordert somit erhebliche organisatorische, technische und personelle Ressourcen. Dies sollte bei der Entscheidung für oder gegen den Einsatz solcher Systeme berücksichtigt werden.

4.2 Was sind Hochrisiko-KI-Systeme?

Hohes Risiko bergen KI-Systeme, die aufgrund ihrer Zweckbestimmung (vgl. Ebers/Streitböcker 2024; vgl. auch Art. 3 Nr. 12) ein hohes Risiko bergen, die Gesundheit, die Sicherheit und die Grundrechte von Personen zu schädigen (ErwG 52). Die KI-VO bildet zwei Fallgruppen:

(a) Die erste Fallgruppe betrifft KI-Systeme, die bereits nach EU-Produktsicherheitsrecht einer Prüfung durch eine unabhängige Stelle bedürfen. Dies ist der Fall, wenn das KI-System entweder selbst ein solches Produkt ist oder als Sicherheitsbauteil eines solchen Produkts verwendet wird. Solche KI-Systeme gelten zugleich als Hochrisiko-KI-System im Sinne der KI-VO (Art. 6 Abs. 1 i. V. m. Annex I).

Für die Soziale Arbeit könnte die Medizinprodukte-Verordnung (Medical Device Regulation, im Folgenden: MDR) Bedeutung haben. Nach der MDR gilt Software in der Regel als zulassungsbedürftiges Medizinprodukt, wenn sie dazu bestimmt ist, Informationen zu liefern, die zu Entscheidungen für diagnostische oder therapeutische Zwecke herangezogen werden (siehe Art. 2 Nr. 1, Art. 52 i. V. m. Anhang VIII Regel 11 der MDR sowie Hoos 2024).

Das bedeutet: Software, die als Medizinprodukt bereits nach der MDR einer Konformitätsbewertung durch Dritte bedarf, gilt als Hochrisiko-KI-System unter der KI-VO. Sie muss dann sowohl die Anforderungen der MDR als auch die Anforderungen der KI-VO an Hochrisiko-KI-Systeme einhalten (vgl. Haftenberger/Dierks 2023; Hacker 2024b).

Eine Hochrisiko-Zuordnung als Medizinprodukt kommt daher z. B. in Betracht bei Chatbots für Therapie und Diagnostik, bei Systemen zur Früherkennung von psychischen Krisen und KI-gestützter Therapieplanung.

(b) Weitere hochriskante Zweckbestimmungen von KI-Systemen werden in Anhang III aufgelistet. Hieraus erscheinen für die Soziale Arbeit die im folgenden Abschnitt (Ziffer 5.) näher beschriebenen Fallgruppen relevant.

4.3 Konkretisierende Leitlinien

Die Reichweite des Hochrisikobereichs und der Ausnahmen hiervon ist noch vage, weshalb auch in diesem Punkt für die Praxis Rechtsunsicherheit besteht (vgl. Ebers/Streitböcker 2024). Die Kommission muss bis zum 2. Februar 2026 konkretisierende Leitlinien mit praktischen Beispielen vorlegen (Art. 6 Abs. 5, Art. 96). Ferner besteht die Möglichkeit für Anbieter und Betreiber, sich über Verhaltenskodizes freiwillig auch bei Nicht-Hochrisiko-KI-Systemen ganz oder teilweise dem Pflichtenkatalog der Hochrisiko-Systeme zu unterwerfen (Art. 95 Abs. 1).

5 Einzelfälle relevanter Hoch-Risiko-KI-Systeme nach Anhang III der KI-VO

5.1 Zugang zu grundlegenden Diensten und Leistungen

Ein hohes Risiko bergen KI-Systeme, die den Zugang zu und die Nutzung von grundlegenden privaten und öffentlichen Diensten und Leistungen betreffen (Anhang III Nr. 5). Der EU-Verordnungsgeber begründet dies damit, dass Menschen,

die auf staatliche Unterstützungsleistungen angewiesen sind, sich typischerweise in einer prekären Situation befinden (ErwG 58).

Hochrisiko-KI-Systeme sind insbesondere KI-Systeme, die von oder im Namen von Behörden verwendet werden sollen, um zu beurteilen, ob natürliche Personen Anspruch auf grundlegende öffentliche Unterstützungsleistungen und -dienste haben. Dies umfasst auch die Entscheidung über Gewährung, Einschränkung, Widerruf oder Rückforderung solcher Leistungen (Anhang III Nr. 5 lit. a.). Die Erwägungsgründe nennen als Beispiele u. a. staatliche Unterstützung bei Mutterschaft, Krankheit, Arbeitsunfall, Pflegebedürftigkeit und Arbeitsplatzverlust sowie Sozialhilfe und Wohngeld (ErwG 58).

5.2 Bildungsbereich

Für die Schulsozialarbeit und Bildungsberatung ist die Hochrisiko-Einstufung verschiedener KI-Systeme im Bildungsbereich relevant. Der EU-Verordnungsgeber erkennt die Bedeutung von KI für die Förderung hochwertiger digitaler Bildung an, sieht aber zugleich erhebliche Risiken und erwähnt u. a. die Gefahr, dass Diskriminierungsmuster fortgeschrieben werden, beispielsweise gegenüber Frauen, bestimmten Altersgruppen, Menschen mit Behinderungen oder Personen mit einer bestimmten Herkunft oder sexuellen Ausrichtung (ErwG 56). Als Hochrisiko-KI-Systeme gelten daher u. a. KI-Systeme zur Feststellung des Zugangs oder der Zulassung zu Bildungseinrichtungen (Anhang III Nr. 3 lit. a.), zur Bewertung von Lernergebnissen (lit. b.) oder zur Bewertung des angemessenen Bildungsniveaus (lit. c.).

5.3 Beschäftigung und Personalmanagement

Die KI-VO stuft verschiedene KI-Systeme im Bereich von Beschäftigung und Personalmanagement als hochriskant ein, beispielsweise wenn sie für die Einstellung oder Auswahl von Personen oder für beschäftigungsrelevante Entscheidungen verwendet werden (Anhang III Nr. 4). Der EU-Verordnungsgeber sieht hier besondere Risiken für Diskriminierung und den Schutz von Arbeitnehmerrechten (ErwG 57). Soweit solche Systeme in der Sozialen Arbeit diskutiert werden (vgl. Löhe 2024), unterfallen sie grundsätzlich dem Hochrisiko-Bereich. Die Verordnung erwähnt ausdrücklich die KI-gestützte Erstellung gezielter Stellenanzeigen sowie die Sichtung von Bewerbungsunterlagen und die Beurteilung von Bewerbenden (lit. a.) Ferner unterfällt die KI-gestützte Dienstplanung dem Hochrisikobereich, soweit Aufgaben aufgrund des individuellen Verhaltens oder persönlicher Merkmale oder Eigenschaften zugewiesen werden (lit. b.). Hingegen stellt ein „Onboarding-Bot“ kein Hochrisiko-KI-System dar, soweit er rein zum Wis-

sensmanagement dient und nicht etwa individuell maßgeschneiderte Aufgaben zuweist.

5.4 Emotionserkennung und biometrische Kategorisierung

Soweit KI-Systeme zur Emotionserkennung oder zur biometrischen Kategorisierung nach sensiblen Merkmalen nicht verboten sind (vgl. oben III. 3 und 5), sind sie stets als Hochrisiko-KI-Systeme einzustufen (Anhang III Nr. 1 lit. b. und c.).

5.5 Weitere relevante Anwendungsbereiche

Als Hochrisiko-KI-Systeme gelten auch bestimmte Systeme im Bereich Migration und Asyl (Nr. 7), Strafverfolgung (Nr. 6) und Rechtsprechung (Nr. 8 lit. a.), wenn diese von oder im Namen der zuständigen Behörden bzw. Gerichte eingesetzt werden. Dies kann etwa relevant sein für die Migrations- und Asylsozialarbeit sowie die Soziale Arbeit im Bereich der Familien-, Sozial- und Verwaltungsgerichte, der Jugendgerichtshilfe oder der Bewährungshilfe. Ferner gelten KI-Systeme zur Bewertung von Notrufen als hochriskant (Nr. 5 lit. d.), was z. B. für die Krisenintervention oder die Ersteinschätzung bei Meldungen zu möglichen Kindeswohlgefährdungen bedeutsam sein kann.

5.6 Ausnahmen von der Hochrisiko-Einstufung

Trotz Nennung in Anhang III sind KI-Systeme ausnahmsweise nicht als hochriskant einzustufen, wenn sie kein erhebliches Risiko für die Gesundheit, Sicherheit oder Grundrechte natürlicher Personen bergen (Art. 6 Abs. 3). Dies ist der Fall, wenn das System die Entscheidungsfindung nicht wesentlich beeinflusst.

Die KI-VO nennt vier Fallgruppen, bei denen eine solche wesentliche Beeinflussung fehlt (vgl. Saam/Hermann 2024):

- eng gefasste Verfahrensaufgaben (z. B. Erkennung von Duplikaten),
- Verbesserung bereits abgeschlossener menschlicher Tätigkeiten (z. B. sprachliche Überarbeitung),
- Erkennung von Abweichungen von früheren Entscheidungsmustern (z. B. Analyse, ob eine Entscheidung mit der bisherigen Bewilligungspraxis übereinstimmt) und
- rein vorbereitende Aufgaben für eine (menschliche) Bewertung.

Diese Ausnahmen gelten aber nie für Systeme, die ein „Profiling“ natürlicher Personen vornehmen (Art. 6 Abs. 3 UAbs. 2). Gemeint sind damit Systeme, die

automatisiert personenbezogene Daten verwenden, um bestimmte persönliche Aspekte, die sich auf eine natürliche Person beziehen, zu bewerten (Art. 3 Nr. 52, der auf Art. 4 Abs. 4 DSGVO verweist). Kurz gesagt scheidet somit eine Ausnahme von der Hochrisikoregulierung aus, sobald es um eine „Personalisierung“ von Dienstleistungen geht (vgl. Hense / Mustać 2025).

Sofern sich der Anbieter des KI-Systems auf eine Ausnahme nach Art. 6 Abs. 3 beruft, sind die Gründe hierfür zu dokumentieren (Art. 6 Abs. 4) und das System in einer EU-Datenbank zu registrieren (Art. 49 Abs. 2, 71).

Zusammenfassend lässt sich sagen, dass die genannten Ausnahmen den Einsatz von KI in Hochrisikobereichen nur dann recht unbürokratisch zulassen, sofern es um rein unterstützende Tätigkeiten ohne wesentlichen inhaltlichen Einfluss auf Entscheidungen geht.

5.7 Einordnung der Eingangsbeispiele

Chatbots:

- Therapeutische Chatbots (z. B. zur Angstbewältigung oder Depressionsbehandlung) dürften als Medizinprodukt im Sinne der MDR einzustufen sein und damit schon deshalb unter die Hochrisiko-Kategorie gemäß Art. 6 Abs. 1 i. V. m. Annex I Nr. 11 fallen.
- Beratende Chatbots ohne therapeutischen Anspruch (z. B. zur reinen Informationsvermittlung in der Schuldnerberatung) können wie emergenzbasierte Prognosesysteme Hochrisiko-Systeme darstellen, z. B. wenn sie wesentlichen Einfluss auf behördliche Entscheidungen über Unterstützungsleistungen haben.
- Reine „First-Level-Support“ Chatbots zur Terminvereinbarung oder für FAQ dürften in der Regel hingegen keine Hochrisiko-KI-Systeme sein.
- KI-Systeme zur Emotionserkennung sind, soweit sie nicht ohnehin verboten sind (vgl. oben III.3), stets als Hochrisiko-KI-Systeme eingestuft (Anhang III Nr. 1 lit. c.).
- KI-gestützte Dokumentationssysteme zur Erstellung von Aktennotizen oder Protokollen dürften in der Regel nicht als Hochrisiko-KI-Systeme einzustufen sein; jedenfalls liegt eine Ausnahme nahe (siehe oben Ziffer 2 lit. b. ff.). Hierzu muss aber sichergestellt sein, dass durch die KI-gestützte Dokumentation das Ergebnis einer nachgelagerten Entscheidungsfindung nicht wesentlich beeinflusst wird (Art. 6 Abs. 3).

6 Transparenzpflichten

Die KI-VO sieht für bestimmte KI-Systeme (egal ob Hochrisiko oder nicht) Transparenzpflichten vor (Art. 50). Diese Pflichten sollen insbesondere Risiken „in Bezug auf Identitätsbetrug oder Täuschung“ (ErwG 132) begegnen.

Bei KI-Systemen, die mit Personen interagieren, z. B. Chatbots, muss bereits der Anbieter dafür sorgen, dass die Personen darüber informiert werden, dass sie es mit einem KI-System zu tun haben, außer das ist offensichtlich (Art. 50 Abs. 1). Die Offensichtlichkeit ist dabei unter besonderer Berücksichtigung schutzbedürftiger Nutzergruppen zu beurteilen, etwa wenn das System mit älteren Menschen oder Menschen mit Behinderungen interagiert (ErwG 132). Hintergrund ist, dass die sehr gute Ausdrucksweise generativer KI-Systeme dazu führen kann, dass die Nutzenden den Eindruck erhalten können, als hätte ihr digitales Gegenüber ein gutes Fallverstehen und eine umfangreiche Expertise (vgl. Linnemann/Löhe/Rottkemper 2024).

Ferner muss der Anbieter den Output von generativen KI-Systemen in einem maschinenlesbaren Format als künstlich erzeugt oder manipuliert erkennbar machen (Art. 50 Abs. 2).

Betreiber von Emotionserkennungs- oder biometrischen Kategorisierungssystemen müssen die betroffenen Personen über den Betrieb des Systems informieren (Art. 50 Abs. 3, vgl. hierzu Merkle 2024).

Eine weitere Betreiberpflicht betrifft das Erkennbarmachen von „Deepfakes“, also KI-generierten Inhalten, die wirklichen Personen, Gegenständen, Orten, Einrichtungen oder Ereignissen merklich ähneln und einer Person fälschlicherweise echt oder wahr erscheinen würden (Art. 3 Nr. 60). Betreiber müssten z. B. einen realistisch wirkenden Avatar, der als „Therapeutin“ agiert, kennzeichnen (Art. 50 Abs. 4).

Auch bei KI-generierten Texten gibt es eine Offenlegungspflicht des Betreibers, allerdings nur wenn die Texte zur Information der Öffentlichkeit über Angelegenheiten von öffentlichem Interesse veröffentlicht werden (Art. 50 Abs. 4 UAbs. 2).

Die Information muss in allen Fällen spätestens bei der ersten Interaktion oder Aussetzung erfolgen und „klar und eindeutig“ sein sowie Anforderungen an die Barrierefreiheit einhalten (Art. 50 Abs. 5).

Auch die konkrete Umsetzung der Transparenzpflichten wird durch Praxisleitfäden (Art. 50 Abs. 7) sowie Leitlinien der Kommission (Art. 96 Abs. 1 lit. d.) noch konkretisiert werden.

Für die Eingangsbeispiele ergibt sich:

- Chatbots müssen sich als solche zu erkennen geben.
- Über den Einsatz von KI-Systemen zur Emotionserkennung sind Betroffene vorab zu informieren.

- Die KI-gestützte Erstellung interner Dokumentation wie Aktennotizen oder Protokolle löst für die Nutzer:innen in der Regel keine Transparenzpflichten aus, da diese nicht für die Öffentlichkeit bestimmt sind.

7 KI-Kompetenz

Die KI-VO verpflichtet Anbieter und Betreiber von KI-Systemen seit Februar 2025, Maßnahmen zu ergreifen, um sicherzustellen, dass alle mit dem Betrieb und der Nutzung von KI-Systemen befassten Personen über ausreichende KI-Kompetenz verfügen (Art. 4). Dies ist für die Soziale Arbeit besonders bedeutsam, da hier KI-Systeme häufig im Kontext von vulnerablen Personengruppen eingesetzt werden sollen.

Das erforderliche Maß an KI-Kompetenz bestimmt sich nach mehreren Faktoren: den technischen Kenntnissen und der Erfahrung der Fachkraft, dem Einsatzkontext des KI-Systems sowie den Personen oder Personengruppen, bei denen das System eingesetzt werden soll. Eine „allgemeine KI-Schulung“ reicht daher nicht aus.

Für die Soziale Arbeit bedeutet dies konkret, dass Fachkräfte

- bei der Verwendung von Beratungs-Chatbots einschätzen können sollten, in welchen Fällen eine persönliche Beratung erforderlich ist,
- bei emergenzbasierten Prognosesystemen die Prognose einzelfallbezogen überprüfen können sollten,
- bei KI-gestützter Dokumentation die Ergebnisse darauf hin prüfen können sollten, ob sie inhaltlich korrekt und vollständig sind.

Die Anforderungen an die KI-Kompetenz steigen dabei mit dem Risikopotenzial des eingesetzten KI-Systems (vgl. Hense / Mustać 2025).

8 Sanktionen

Die Mitgliedstaaten müssen nach Art. 99 Abs. 1 wirksame, verhältnismäßige und abschreckende Sanktionen erlassen. Bei Verstößen insbesondere gegen die Verbotstatbestände, Hochrisiko- oder Transparenzpflichten drohen dann, je nach Art, Schwere und Dauer des Verstoßes, erhebliche Geldbußen. Die nationalen Umsetzungsvorschriften zur KI-VO in Deutschland existieren noch nicht (Stand Januar 2025).

9 Handlungsempfehlungen für die Praxis

Für die Praxis ergeben sich folgende Handlungsempfehlungen:

- Träger Sozialer Arbeit sollten zunächst eine Bestandsaufnahme ihrer Softwaresysteme vornehmen und diese darauf hin prüfen, ob es sich um KI-Systeme im Sinne der KI-VO handelt (siehe oben II.).
- Als nächster Schritt sollten die identifizierten KI-Systeme auf mögliche Verbotstatbestände überprüft werden (siehe oben III.). Besonders relevant sind hier das Verbot der Ausnutzung von Vulnerabilität, das Verbot von Social Scoring, das Verbot manipulativer Techniken sowie von bestimmten Systemen zur Emotionserkennung. Verbotene Systeme dürfen seit Februar 2025 nicht mehr verwendet werden.
- Anschließend ist für jedes System zu prüfen, ob es sich um ein Hochrisiko-KI-System handelt (siehe oben IV.). Dies betrifft in der Sozialen Arbeit insbesondere KI-Systeme, die als Medizinprodukt einer Konformitätsbewertung unterliegen, den Zugang zu Sozialleistungen beeinflussen, im Bildungsbereich eingesetzt werden oder Emotionen erkennen. Sofern ein Hochrisiko-System vorliegt, muss der daraus erwachsende, umfangreiche Pflichtenkatalog beachtet werden. Bereits in Betrieb genommene Hochrisiko-KI-Systeme genießen aber unter Umständen Bestandsschutz (Art. III Abs. 2 KI-VO).
- Ferner ist zu prüfen, ob für bestimmte KI-Systeme Transparenzpflichten gemäß Art. 50 bestehen (siehe oben V.). Insbesondere bei Chatbots und Systemen zur Emotionserkennung sind die Klient:innen zu informieren.
- Seit Februar 2025 verbindlich vorgeschrieben ist zudem der Aufbau von KI-Kompetenz (siehe oben VI.). Träger Sozialer Arbeit sollten daher Maßnahmen ergreifen, um ihr Personal für den Umgang mit den konkret benutzten KI-Systemen zu qualifizieren.
- Verstöße sind bußgeldbewehrt (siehe oben VII.).

Fazit: Die rechtssichere Einhaltung der KI-VO dürfte für verschiedene KI-Anwendungen in der Sozialen Arbeit eine Herausforderung darstellen. Der EU-Verordnungsgeber sieht diese Hürden und Verbote jedoch als notwendig an, um Grundrechte zu schützen.

Literatur

- Biallaß, Isabelle (2024): Die Auswirkungen der KI-VO auf die Justiz. In: *Multimedia und Recht*, H. 8, S. 646–651.
- Braegelmann, Tom (2024): KI-VO und Compliance – aktuelle Brennpunkte. In: *Künstliche Intelligenz und Recht (KIR)*, S. 39–42.
- Drilling, Johannes/Musiol, Philip (2024): KI-Verordnung und interne Ermittlungen. In: *Corporate Compliance Zeitschrift (CCZ)*, H. 10, S. 257–275.

- Ebers, Martin/Streitböger, Chiara (2024): Die Regulierung von Hochrisiko-KI-Systemen in der KI-Verordnung. In: *Recht Digital (RD)*, S. 393–400.
- Engel, Andreas (2024): Generative KI, Foundation Models und KI-Modelle mit allgemeinem Verwendungszweck in der KI-VO. In: *Künstliche Intelligenz und Recht (KIR)*, S. 21–28.
- Europäische Kommission (2025a): Commission Guidelines on the definition of an artificial intelligence system established by Regulation (EU) 2024/1689 (AI Act) ec.europa.eu/newsroom/dae/redirection/document/112455 (Abfrage: 15.06.2025).
- Europäische Kommission (2025b): Commission Guidelines on prohibited artificial intelligence practices established by Regulation (EU) 2024/1689 (AI Act) ec.europa.eu/newsroom/dae/redirection/document/112367 (Abfrage: 15.06.2025).
- Feist-Ortmanns, Monika/Sauer, Annette/Brinkmann, Martin (2025): KI-basiertes Assistenzsystem im Kinderschutzverfahren. In: Macsenaere, Michael (Hrsg.): *Künstliche Intelligenz in der Kinder- und Jugendhilfe*. München: Ernst Reinhardt, S. 50–66.
- Hacker, Philipp (2024a): Comments on the Final Trilogue Version of the AI Act. ssrn.com/abstract=4757603 (Abfrage: 15.06.2025).
- Hacker, Philipp (2024b): The AI Act between Digital and Sectoral Regulations. In: Bertelsmann Stiftung 2014. <https://www.bertelsmann-stiftung.de/en/publications/publication/did/the-ai-act-between-digital-and-sectoral-regulations-en> (Abfrage: 15.06.2025).
- Haftenberger, Anna/Dierks, Christian (2023): Rechtliche Einordnung von künstlicher Intelligenz in der Inneren Medizin. In: *Die Innere Medizin* 64(11), S. 1044–1050. <https://www.dierks.company/wp-content/uploads/Rechtliche-Einordnung-von-KI-Die-Innere-Medizin.pdf> (Abfrage: 15.06.2025).
- Hense, Peter/Mustać, Tea (2025): *AI Act kompakt. Compliance, Management & Use Cases in der Unternehmenspraxis*. Frankfurt a. M.: Fachmedien Recht und Wirtschaft.
- Heidrich, Joerg (2024): DS-GVO als Endgegner bei der Nutzung von ChatGPT & Co. In: *Multimedia und Recht (MMR)*, H. 11, S. 919–920.
- Heinze, Christian/Engel, Timon-Johannes (2025): Das Verbot von ausbeuterischen und manipulativen KI-Praktiken. In: *Künstliche Intelligenz und Recht (KIR 2025)*, S. 19–29.
- Holz, Felix/Fellmann, Michael/Schmidt Angelina Clara (2025): Textanalysetechniken auf Tagesdokumentationen zur Prozessassistenz. In: Linnemann, Gesa/Löhe, Julian/Rottkemper, Beate (Hrsg.): *KI in der Sozialen Arbeit*. Weinheim und Basel: Beltz Juventa.
- Hoos, Katja (2024): KI trifft auf Medizinprodukte – Das zukünftige Zusammenspiel von AI-Act und MDR. In: *Zeitschrift für Produktsicherheit und -compliance (ZfPC)*, H. 4, S. 168–175.
- Kapoor, Amit/Verma, Vishal (2024): Emotion AI: Understanding emotions through artificial intelligence. In: *International Journal of Engineering Science and Humanities* 14, S. 223–232. <https://doi.org/10.62904/Ovcvb24>
- Li, Han/Zhang, Renwen/Lee, Yie-Chieh/Kraut, Robert E./Mohr, David C. (2023): Systematic review and meta-analysis of AI-based conversational agents for promoting mental health and well-being. In: *npj Digital Medicine* 6(1), S. 236. <https://doi.org/10.1038/s41746-023-00979-5>
- Linnemann, Gesa/Löhe, Julian/Rottkemper, Beate (2024): Bedeutung von Selbstoffenbarungseffekten in quasisozialen Beziehungen mit auf generativer KI basierten Systemen in Settings von Onlineberatung und -therapie. In: *e-beratungsjournal.net – Zeitschrift für Onlineberatung und computervermittelte Kommunikation* 20(1), Artikel 1, S. 1–21. <https://doi.org/10.48341/9xIs-5y11>
- Löhe, Julian (2024): KI-Anwendung in der Kinder- und Jugendhilfe. In: Kreidenweis, Helmut (Hrsg.): *KI in der Sozialwirtschaft. Eine Orientierungshilfe für die Praxis*. Baden-Baden: Nomos, S. 101–116.
- Martini, Mario/Wendehorst, Christiane (2024): *KI-VO, Verordnung über künstliche Intelligenz*. München: C. H. Beck.
- Merkle, Marieke (2024): *Transparenz nach der KI-Verordnung – von der Blackbox zum Open-Book?* In: *Recht Digital (RD)*, S. 414–420.

- Plafky, Christina/Frischhut, Hans (2025): Einsatz von Künstlicher Intelligenz zu Prognosezwecken in der Kinder- und Jugendhilfe. In: Macsenaere, Michael (Hrsg.): Künstliche Intelligenz in der Kinder- und Jugendhilfe. München: Ernst Reinhardt, S. 74–82.
- Roth-Isigkeit, David (2024): Der neue Rechtsrahmen für Künstliche Intelligenz in der Europäischen Union. In: Künstliche Intelligenz und Recht (KIR), H. 1, S. 15–20. cdn-assetservice.ecom-api.beck-shop.de/productattachment/readingsample/15428864/37627358_kir_2024_01_gesamt.pdf (Abfrage: 15.06.2025).
- Saam, Daniel/Hermann, Christian (2024): Die Ausnahmeregelung zur Einstufung als Hochrisiko-KI nach Art. 6 III KI-VO. In: Recht Digital (RD*i*), S. 608–614.
- Wachter, Martin/Leeb, Christina-Maria (2024): KI-Systeme in der Rechtspflege. In: Recht Digital (RD*i*), S. 440–446.

Künstliche Intelligenz in der Lehre der Sozialen Arbeit¹

Edeltraud Botzum, Madeleine Dörr, Andrea Gergen,
Florian Müller

Abstract: Der Beitrag analysiert die didaktischen, ethischen und hochschulpolitischen Herausforderungen und Potenziale von Künstlicher Intelligenz in der Lehre der Sozialen Arbeit. Im Zentrum steht das Forschungsprojekt *digi.peer* an der Technischen Hochschule Rosenheim, das Peer-Mentoring und den Einsatz KI-gestützter Tools zur Förderung wissenschaftlicher Schreibkompetenzen kombiniert. Aktuelle Studien zeigen, dass Studierende KI-Tools vielseitig nutzen und darin sowohl Chancen als auch Risiken erkennen. KI-Anwendungen werden unter Studierenden meist als hilfreiche Unterstützung wahrgenommen. Gleichzeitig besteht Bedarf an Reflexion, ethischer Orientierung und klaren institutionellen Regelungen zum Einsatz von KI im Studium. Daher plädieren die Autor:innen für eine partizipative, strategisch fundierte und didaktisch durchdachte Integration von KI in Studium und Lehre, um akademische Integrität zu gewährleisten und digitale Schlüsselkompetenzen bei Studierenden zu fördern.

Keywords: Künstliche Intelligenz in der Hochschullehre, digitale Schlüsselkompetenzen, Studierendenperspektive auf KI, digitale Tools, Hochschulpolitik

1 Einführung

Künstliche Intelligenz (KI) wird die hochschulische Lehre auch in Zukunft maßgeblich prägen – im Positiven wie im Negativen, denn die Einführung von KI-gestützten Computermodellen zur Sprachverarbeitung stellte einen Paradigmenwechsel dar (vgl. Aldosari 2020). Im Rahmen der zunehmenden Einbindung des technologiegestützten Lernens geraten nun hochschuldidaktische und -ethische Fragestellungen, bildungstheoretische Ausführungen und die Meinungen von Studierenden zum KI-gestützten Lehren und Lernen in den Fokus von Forschung und Lehre (vgl. Deutscher Ethikrat 2023a, b; Reinmann 2023). Vor

1 © Die Autor:innen 2025, veröffentlicht im Verlag Beltz Juventa, Open Access: CC BY 4.0
Gesa Linnemann / Julian Löhe / Beate Rottkemper (Hg.), *Künstliche Intelligenz in der Sozialen Arbeit*
10.3262/978-3-7799-8562-4_017

diesem Hintergrund wird im Folgenden ein durch die Stiftung Innovation in der Hochschullehre (StIL) gefördertes Lehrforschungsprojekt der Fakultät für Sozialwissenschaften der Technischen Hochschule (TH) Rosenheim vorgestellt, das auf der Grundlage einer durch Peer-Mentoring gestützten Schreibwerkstatt zur Begleitung von Bachelorarbeiten im Studiengang Soziale Arbeit die Nutzung von KI in der Lehre anwendungsbezogen erforscht (vgl. Botzum/Gergen 2022).

Ausgehend vom Diskurs um die Relevanz digitaler *Future Skills* für die Nutzung von KI (vgl. 2) wird in Kapitel 3 anhand einer stichprobenartigen Erhebung das KI-Nutzungsverhalten von Studierenden der Sozialen Arbeit an der TH Rosenheim veranschaulicht. Hochschulische Schreibzentren sind durch die KI-gestützte Transformation der Lehre besonders stark betroffen (vgl. Fritze 2024, S. 49). Deshalb wirft der folgende Abschnitt ein Schlaglicht auf einen in Zusammenarbeit mit dem Schreibzentrum der TH Rosenheim entwickelten Zertifikatskurs zum wissenschaftlichen Arbeiten anhand KI-gestützter Tools. Angesichts der Tatsache, dass der Ausbau digitaler Kompetenzen für den digitalen Wandel von der Europäischen Union als aktuelles bildungspolitisches Ziel ihrer Mitgliedsländer ausgerufen wurde, stehen Hochschulen verstärkt in der Verantwortung, ihre Studierenden zum Aufbau von KI-Kompetenzen zu befähigen (vgl. Europäische Kommission 2020, S. 15). Daher gibt Abschnitt 5 einen Ausblick auf die hochschulpolitische Dimension der Diskussion um KI in der Lehre der Sozialen Arbeit. Das abschließende Fazit fasst perspektivische Entwicklungen in der KI-gestützten hochschulischen Lehre der Sozialen Arbeit zusammen.

2 *Future Skills*, Bildung und *Deskilling* – aktuelle Diskurse um die Nutzung von KI in der hochschulischen Lehre

2.1 Die Einführung von KI als Paradigmenwechsel in der hochschulischen Lehre

Die Entwicklung im Bereich generativer KI wirft seit der Veröffentlichung von ChatGPT im Jahr 2022 Fragen zur Gestaltung hochschulischer Lehre, auch im Studiengang Soziale Arbeit, auf. Der Begriff selbst ist nicht präzise definiert. Gemeinhin bezeichnet KI

„Methoden, Verfahren und Technologien, die es IT-Systemen, wie Maschinen, Robotern oder Softwaresystemen, ermöglichen, große Mengen von Daten zu interpretieren und aus diesen Daten zu lesen, um bestimmte menschlich-kognitive Fähigkeiten nachzubilden bzw. zu imitieren“ (Di Vaio et al. 2020, zitiert nach von Garrel et al. 2023, S. 8).

Aufgaben, die beispielsweise visuelle Wahrnehmung, Sprache oder strategisches Denken und Planen erfordern, können durch IT-Systeme eigenständig und effizient durchgeführt werden (vgl. von Garrel et al. 2023, S. 8). Dies führt zu der Frage, inwiefern wissenschaftliches Lesen und Schreiben in der Lehre ihre Rolle als Kommunikationsinstrumente und Kulturtechniken beibehalten können (vgl. Buck et al. 2024).

KI-Tools dienen als Werkzeuge zur Unterstützung von Literaturrecherche und -auswertung, Lektüreunterstützung, Texterstellung, Datenauswertung und Textüberarbeitung. Schon jetzt können Tools wie *ChatGPT* wissenschaftliche Texte vorlesen und zusammenfassen oder über mobile Endgeräte im sprachlichen Dialog selbst entwickeln (vgl. ebd.). Der KI-Einsatz an Hochschulen wird in drei Ebenen unterteilt: Mikroebene (einzelne Lehrkontexte), Mesoebene (Fakultäten) und Makroebene (gesamte Hochschule) (vgl. Watanabe 2023). Im vorliegenden Beitrag wird die Nutzung von KI auf allen drei Ebenen der Hochschulbildung untersucht.

2.2 Aktuelle Diskurse zur Nutzung von KI in der Lehre der Sozialen Arbeit

Die Anwendung von KI in der Lehre wird häufig im Kontext der sogenannten *Future Skills* diskutiert, die vom Stifterverband als Leitmarken im bildungspolitischen Diskurs gefördert werden (vgl. Ehlers et al. 2024). Beispielsweise sollen Lernplattformen Informationen über neue Technologien wie KI bereitstellen (vgl. Stifterverband 2021). Im „Future Skills Framework 2021“ wurden sogenannte „transformative Kompetenzen“ (ebd., S. 4) ausformuliert, die in den nächsten fünf Jahren für Beschäftigte von Relevanz sein werden, darunter technologische, klassische und transformative Kompetenzen sowie digitale Schlüsselkompetenzen (vgl. ebd., S. 5). Diese Kompetenzen beschreiben Fähigkeiten, die erforderlich sind, um an einer digitalisierten Welt aktiv teilzunehmen, darunter *Digital Literacy*, *Digital Ethics*, digitale Kollaboration und agiles Arbeiten (vgl. Tabelle 1).

Die Ausdifferenzierung der *Future Skills* basiert auf den von der OECD 2005 formulierten Schlüsselkompetenzen, die im Zuge der Gründung eines gemeinsamen Europäischen Hochschulraums (Bologna-Prozess) ab 1999 relevant wurden (vgl. OECD 2005; Raber 2012). Die Bologna-Erklärung verpflichtete die Unterzeichnerstaaten zur Einführung vergleichbarer Abschlüsse und zur Qualitätssicherung in der Hochschulbildung (vgl. Bologna-Erklärung 1999). Ziel war die Förderung von „Schlüsselkompetenzen“ für Beschäftigungsfähigkeit, gesellschaftliche Verantwortung und Persönlichkeitsentwicklung im Hochschulstudium (vgl. Raber 2012, S. 5). In der Debatte um die *Employability* der Absolvent:innen entwickelte sich seit 2010 eine Diskussion über die Freiheit von Forschung und Lehre. Thematisiert wurde das traditionelle Spannungsverhältnis zwischen einem von Interessenträgern der Wirtschaft formulierten Anspruch an die Qualifizierungs-

Tabelle 1: Digitale Schlüsselkompetenzen (*Future Skills*)

DIGITALE SCHLÜSSEL KOMPETENZEN	Digital Literacy	Beherrschen von grundlegenden digitalen Fähigkeiten, z. B. sorgsamer Umgang mit digitalen persönlichen Daten, Verständnis von grundlegenden Sicherheitsregeln im Netz, Nutzen gängiger Software
	Digital Ethics	Kritisches Hinterfragen von digitalen Informationen und Auswirkungen des eigenen digitalen Handelns sowie entsprechende ethische Entscheidungsfindung
	Digitale Kollaboration	Nutzung von Onlinekanälen zur effizienten Interaktion, Kollaboration und Kommunikation mit anderen; effektive und effiziente Zusammenarbeit unabhängig von räumlicher Nähe, angemessene Etikette bei digitaler Kommunikation
	Digital Learning	Verständnis und Einordnen digitaler Informationen, Deutung von Informationen unterschiedlicher digitaler Quellen; Aufbau von Wissen in ausgewählten Themengebieten; Nutzung von Lehr-/ Lernsoftware
	Agiles Arbeiten	Nutzerorientierte, selbstverantwortliche und iterative Zusammenarbeit in Teams unter Nutzung agiler Arbeitsmethoden

Quelle: Eigene Darstellung in Anlehnung an Stifterverband 2021, S. 6

pflicht der Hochschulen einerseits und dem humanistischen Bildungsideal nach Wilhelm von Humboldt andererseits, wodurch das Autonomieverständnis deutscher Hochschulen seit dem frühen 19. Jahrhundert bestimmt wurde (vgl. ebd.). Voraussetzung für die Bildung des einzelnen Subjekts durch wissenschaftliches Bemühen um objektive Erkenntnis ist demnach die Unabhängigkeit der Universität von Staat und Wirtschaft (vgl. von Humboldt 1985).

Im Diskurs um die Entwicklung von KI und *Future Skills* in der hochschulischen Lehre wird dieser Diskurs nun im Kontext des sogenannten *Deskillingings* thematisiert, das den Verlust von Kompetenzen aufgrund technologischer Veränderungen beschreibt (vgl. Deutscher Ethikrat 2023a, b; Reinmann 2023). *Deskilling* bezieht sich auf Veränderungen am Arbeitsplatz und individuelles Verlernen durch fehlende Praxis (vgl. Reinmann 2023, S. 49). Diese Diskussion wird in den folgenden Abschnitten zur KI-Nutzung in der Hochschullehre thematisiert.

3 KI-Nutzung im Studium der Sozialen Arbeit aus Studierendenperspektive

Zur Betrachtung der Studierendenperspektive wurde eine stichprobenartige Erhebung² im Studiengang der Sozialen Arbeit an der TH Rosenheim durchgeführt, da sich bisher wenige Studien (Ausnahme: Witter et al. 2024) speziell mit der Nutzung von KI im Studium der Sozialen Arbeit auseinandersetzen. Ziel war es, Erkenntnisse zu gewinnen, wie Studierende KI-Tools nutzen und welche Erfahrungen sowie Einstellungen sie diesbezüglich haben. Insgesamt wurden 185 Studierende kontaktiert, von denen acht (4,32%) an der anonymen Befragung teilnahmen. Die Antworten der Studierenden zeigen Tendenzen, die zwar nicht repräsentativ sind, jedoch Übereinstimmungen mit Ergebnissen anderer Studien aufweisen, die sich ebenfalls mit dem Nutzungsverhalten, den Einstellungen und Erwartungen von Studierenden aus verschiedenen Fachrichtungen (von den Ingenieurwissenschaften über die Geisteswissenschaften bis hin zu den Ernährungswissenschaften) gegenüber KI befassen.³ Die stichprobenartige Erhebung verdeutlicht, dass KI-Tools wie *ChatGPT*, *Perplexity*, *DeepL*, *HesseAI* und *goblin.tools*⁴ bereits im Studienalltag eingesetzt werden,⁵ wobei *ChatGPT* am häufigsten genannt wurde.⁶ Die KI-Tools werden zur Textanalyse und -überarbeitung, Übersetzung, Erklärung von Fachbegriffen/Methoden/Modellen, Literaturrecherche, Erstellung von Zusammenfassungen oder zur Prüfungsvorbereitung genutzt. So beschreibt eine Person:

„Nach dem Zusammenfassen der klausurrelevanten Themen in einer PDF kann man diese hochladen und sich Prüfungsfragen zu den Themen aus dem Dokument stellen lassen [...] Die eigenen Antworten kann man sich auch direkt bewerten lassen, wobei gelegentlich eine eigene Kontrolle wichtig ist.“ (Person 2, S. 3)

Besonders *ChatGPT* scheint sich als vielseitiges Tool zu etablieren, das nahezu für alle Aufgaben und Bereiche herangezogen wird. Dies wird durch die quantitative Studie von Witter et al. (2024) bestätigt, die sich speziell mit dem Studiengang der

2 Die Auswertungsergebnisse der unveröffentlichten Erhebung können bei den Autor:innen angefragt werden.

3 Siehe hierzu von Garrel 2023 und Witter et al. 2024.

4 *goblin.tools* ist eine Plattform, welche nützliche Werkzeuge bietet, um Aufgaben zu automatisieren, und den Nutzenden hilft, effizienter zu arbeiten. Die Tools sind darauf ausgelegt, komplexe Aufgaben zu vereinfachen, sei es durch das Erstellen von To-do-Listen, das Planen von Projekten oder das Bereitstellen von Schreib- und Denkkunterstützung (vgl. <https://goblin.tools/About>, Abfrage: 21.08.2024).

5 Lediglich eine Person gab im Rahmen der Umfrage an, bisher keine KI-Tools genutzt zu haben.

6 Sechs der teilnehmenden Studierenden gaben an, *ChatGPT* bereits genutzt zu haben oder zu nutzen.

Sozialen Arbeit und der Nutzung von *ChatGPT* unter Studierenden befasst. In diesem Rahmen wurden deutschlandweit Sozialarbeitsstudierende zu ihrem Nutzungsverhalten, der Bewertung und Auseinandersetzung mit *ChatGPT* im Studium befragt (vgl. ebd., S. 2). Die Ergebnisse zeigen, dass die Nutzung von *ChatGPT* weit verbreitet ist: Rund 80 % verwenden *ChatGPT* für studienbezogene Aufgaben, insbesondere für die Klärung von Verständnisfragen und fachspezifischen Konzepten, zur Textanalyse und -erstellung sowie zu Recherchezwecken (vgl. ebd., S. 14, 22). Ähnliche Ergebnisse liefert die bisher umfangreichste quantitative Studie zu KI im Studium, in der 6311 Studierende aus verschiedenen Studiengängen von 395 Hochschulen in Deutschland befragt wurden (vgl. von Garrel 2023, S. 15). Die Studie zeigt, dass 63,4 % KI-basierte Tools in ihrem Studium verwenden, während 36,6 % keine Tools einsetzen (vgl. ebd., S. 20). Dabei steht die Nutzung in Abhängigkeit zum jeweiligen Studienfach:⁷ So liegt die Nutzungsquote in ingenieurwissenschaftlichen Studiengängen bei über 75 %, in den Rechts-, Wirtschafts- und Sozialwissenschaften bei rund 58 % (vgl. ebd., S. 21). Auch in dieser Studie waren *ChatGPT* und *DeepL* die am häufigsten verwendeten KI-Tools, insbesondere für Verständnisfragen, Recherche, Übersetzungen, Textanalyse, Problemlösung und Entscheidungsfindung (vgl. ebd., S. 26 f.).

Die Studienergebnisse zeigen, dass KI auch im Studium der Sozialen Arbeit angekommen ist. Dabei scheint der Einsatz von KI im Studium weitgehend als legitim angesehen zu werden, solange die wesentlichen Inhalte eigenständig erarbeitet werden (vgl. Balabdaoui et al. 2023). KI-Tools werden als Unterstützung zur Effizienz- und Qualitätssteigerung der eigenen Arbeit wahrgenommen, ersetzen jedoch nicht das eigenständige Denken und Arbeiten (vgl. ebd., 2023). Nach einer Aussage aus der stichprobenartigen Erhebung,

„sind KI-Tools (zumindest bisher) kleine Helferlis, die einem bei der ein oder anderen Kleinigkeit unter die Arme greifen können“ (Person 8, S. 4).

KI-Tools werden nicht als Ersatz, sondern als Ergänzung zu traditionellen Lehrmethoden (z. B. Vorlesungen, Seminare) gesehen, die in den Studienalltag integriert werden sollten, da ein positiver Effekt auf das akademische und berufliche Leben erwartet wird (vgl. Balabdaoui et al. 2023). Gleichzeitig sehen Studierende neben Vorteilen von KI-Tools auch Problematiken. So beschreibt eine Person der stichprobenartigen Erhebung über die Antwortenausgabe von Tools wie *ChatGPT*:

„[J]edoch sollte man sich nicht darauf verlassen, dass dies stimmt, sondern weiterhin immer die Daten, welche die KI mir gibt, kritisch hinterfragen und nachrecherchieren“ (Person 5, S. 5),

7 Für weitere Ergebnisse siehe von Garrel et al. 2023, S. 21.

um damit ein blindes Vertrauen auf KI und die Verbreitung von Fehlinformationen durch Halluzinationen zu verhindern (vgl. von Garrel et al. 2023, S. 11). Zusätzlich bestehen Bedenken mit Blick auf Datenschutz, Urheberrecht und einer möglichen Überabhängigkeit (vgl. Balabdaoui et al. 2023). Studierende sehen es daher als relevant an, dass KI-Tools im Studium nur dann eingesetzt werden, wenn sie wissenschaftliche Standards, Fehlerfreiheit und logische Argumentation gewährleisten (vgl. von Garrel et al. 2023, S. 32 f.).

Trotz der Risiken sehen Studierende es nicht als Lösung an, KI im Studium zu verbieten, sondern fordern eine verstärkte Thematisierung in Form von Informationsveranstaltungen sowie eine curriculare Verankerung (siehe Mittmann et al. 2023) von KI-bezogenen Themen in der Aus- und Weiterbildung von Studierenden der Sozialen Arbeit (vgl. Witter et al. 2024, S. 23), „um die erforderliche Reflexionskompetenz mit Blick auf technische Zusammenhänge und Handlungssysteme auszubilden und dabei die spezifischen Herausforderungen und Anforderungen in der Sozialen Arbeit professionell mitzudenken“ (Zorn/Seelmeyer 2015, S. 134). Beispielhaft kann hier der Einsatz von *ChatGPT* in der Beratungsausbildung von angehenden Fachkräften angeführt werden, mit dessen Hilfe komplexe Dialoge inszeniert werden können (siehe Engelhardt 2024). Zudem wird Transparenz in der Nutzung von KI-Tools gefordert, etwa durch Richtlinien und Bereitstellung geeigneter KI-Tools durch die Hochschule, sofern diese den Anforderungen der Transparenz, Zuverlässigkeit, Aktualität und Unvoreingenommenheit gerecht werden (vgl. Balabdaoui et al. 2023).

4 Digitale Tools und KI-Anwendungen im Lehrforschungsprojekt *digi.peer* an der Technischen Hochschule Rosenheim

4.1 Digitale Tools und KI-Anwendungen in der Hochschullehre

KI-Tools bieten neue Möglichkeiten zu lehren, z. B. durch eine erweiterte Interaktion mit Chatbots oder die Anreicherung der eigenen Lehre mit KI-generierten Inhalten (vgl. Tobor 2024, S. 32). Bei der Implementierung von KI in der Hochschullehre gibt es anscheinend noch zahlreiche Herausforderungen zu bewältigen, wie eine Studierendenbefragung (Wintergerst 2024) zeigt. Nur 37% der Hochschulen haben zentrale – oder zumindest von Lehrenden festgelegte – Regeln für den Umgang mit generativer KI. In der Studie geben Studierende zu 38% an, sie werden an ihrer Hochschule nicht auf die Anforderungen der digitalen Arbeitswelt vorbereitet. 36% sind der Meinung, die Digitalisierung ihrer Hochschule scheitere an mangelnden Kompetenzen des Lehrpersonals. 74% wünschen sich, den richtigen Umgang mit *ChatGPT* an ihrer Hochschule zu erlernen, und 44% der Befragten geben an, dass der Einsatz an allen Hochschu-

len Standard sein sollte (vgl. ebd., S. 2 ff.). Dies deutet aus Perspektive der 506 Befragten auf einen Nachholbedarf an Hochschulen sowie Nachqualifizierungsbedarf der Lehrenden hin. Ihnen kommt eine besondere Bedeutung zu, denn sie werden durch Einsatz von KI im eigenen Vorlesungsgeschehen – im Idealfall – zu Impulsgebenden sowie zentralen Gestaltungspersonen von KI für Lehre und Lernen. Der Anwendungsbezug im Projekt *digi.peer* (Digitales Peer- Mentoring im Kontext einer Literatur- und Forschungswerkstatt im Bachelor-Studiengang Soziale Arbeit) hilft dabei, Möglichkeiten und Grenzen von generativer KI im Lehrbetrieb zu erörtern, gemeinsam mit Studierenden der Sozialen Arbeit zu reflektieren, um darauf aufbauend sinnvolle Einsatzszenarien zu erproben.

Die Herausforderungen einer sinnvollen Anwendung von KI-gestützter Software im Lehrbetrieb gehen aber weit über die rein technische Umsetzung hinaus. Mit der Einführung von KI-Anwendungen in der Hochschulbildung gehen offenkundig auch Transformationsprozesse innerhalb des Bildungssystems einher, die mit weitreichenden Auswirkungen auf das Bildungsverständnis verbunden sein werden. Im Lehrforschungsprojekt *digi.peer* erfolgt daher eine kontinuierliche und kritische Reflexion der eingesetzten Technologien, um die akademische Integrität⁸ im Forschungsprozess zu gewährleisten. Hochschullehrende nehmen beim Thema KI-Kompetenz gegenwärtig eine Doppelrolle ein. Einerseits bilden sie Studierende zum verantwortungsbewussten Umgang mit neuen Technologien aus. Andererseits werden Hochschulen in Zukunft immer häufiger zum Anwendungsfeld von KI, was einen kompetenten, transparenten und ethisch reflektierten Umgang mit den neuen Programmen voraussetzt (vgl. Renz/ Etsiwah 2020, S. 37). Die Implementierung digitaler Tools und KI-gestützter Softwareanwendungen durch das Projekt *digi.peer* bietet so die Gelegenheit, wichtige anwendungsbezogene Lehr- sowie Lernerfahrungen zu sammeln, und orientiert sich dadurch an den Empfehlungen der Ständigen Wissenschaftliche Kommission der Kultusministerkonferenz (SWK). Diese empfiehlt die „systematische Erprobung von LLM [*Large Language Models, Anm. d. Verf.*] bei offener Fehlerkultur“ (SWK 2024, S. 4). Die Herangehensweise im Lehrforschungsprojekt *digi.peer* folgt dieser Prämisse und positioniert sich gegenüber gegenwärtigen Problemen lösungsorientiert.

8 Die Leitlinien zur Sicherung guter wissenschaftlicher Praxis der Deutschen Forschungsgemeinschaft (DFG) sind in diesem Zusammenhang besonders relevant. Sie betonen die Bedeutung von Integrität in der Forschung, einschließlich der Vermeidung von Plagiaten, der ordnungsgemäßen Dokumentation von Daten und der Verantwortung gegenüber Kollegium und Gesellschaft.

4.2 Implikation von KI-Anwendungen in der Lehre der Sozialen Arbeit

Das Spannungsfeld zwischen technologischer Innovation, professionellem Handeln und ethischer Verantwortung stellt Lehrende bei dem Einsatz von KI vor Herausforderungen. Diese gehen aber auch mit großen Chancen einher, denn die Softwareanwendungen erweitern in einer gemeinsamen Allianz die Fähigkeiten des Menschen, riesige Datenmengen in nahezu Echtzeit zu analysieren. Mensch und Maschine werden im Idealfall zu symbiotischen Partnern, die sich zu größeren Leistungen antreiben können (vgl. Ifenthaler 2023, S. 74). Um einen zukünftigen Technikdeterminismus in Studium und Lehre zu vermeiden, muss der Einsatz von disruptiven Technologien wie KI mit den allgemeinen Zielen von Hochschulbildung im Einklang stehen: wissenschaftliches Denken und Arbeiten der Studierenden, Vorbereitung auf den Arbeitsmarkt, Befähigung zur Teilhabe am gesellschaftlichen Leben sowie Persönlichkeitsbildung (vgl. Pinkwart/Rampelt/de Witt 2020, S. 10). Erfahrungen aus dem Studiengang der Sozialen Arbeit zeigen, dass Studierende in der Lage sind, die Vorteile KI-gestützter Tools zu erkennen, gleichzeitig Bedarf an Anleitung und kritischer Reflexion hinsichtlich der Integration in ihren Lernprozess haben. Absolvent:innen der Sozialen Arbeit werden ein berufliches Umfeld vorfinden, das vermehrt von digitalen Tools und KI-gestützter Software geprägt ist (vgl. u. a. Botzum/Neumaier 2023; Heinlein/Huchler 2024). Im Projekt *digi.peer* wurde deshalb in Zusammenarbeit mit dem Schreibzentrum der TH Rosenheim ein Zertifikatskurs zum wissenschaftlichen Arbeiten anhand KI-gestützter Tools entwickelt. Das Zertifikat *digi.peer+* dient dem Erwerb digitaler Schlüsselkompetenzen. Solange es keine einheitlichen und klaren Regelungen zur Nutzung von KI gibt, bedarf es einer explorativen Herangehensweise an den Einsatz von KI in der Lehre, um akademische Integrität und ethische Standards im fortlaufenden Studienbetrieb zu gewährleisten.

5 Hochschulpolitische Konsequenzen durch die Nutzung von KI

5.1 Prüfungsrechtliche Herausforderungen und die Rolle der Hochschulpolitik

Die rasante Entwicklung und Verbreitung neuer KI-Systeme und LLM stellt Hochschulen und Lehrende vor erhebliche prüfungsrechtliche Herausforderungen. Insbesondere bei der Bewertung von Haus- und Abschlussarbeiten muss die Eigenleistung der Studierenden auch mit Nutzung von KI nachvollziehbar und prüfbar bleiben. Diese Entwicklung erfordert eine Anpassung bestehender Prüfungsordnungen und -formate.

Von besonderer Relevanz ist der Umgang mit der „zwischenstrettigen“ Nutzung von KI-Tools. Beispielsweise können Chatbots wie *ChatGPT* verwendet werden, um im Prüfungsprozess Ideen zu entwickeln, Texte zu überarbeiten oder Informationen zu generieren. Herbold et al. (2023) zeigen in ihrer Studie, dass KI-generierte Texte oft eine höhere durchschnittliche Qualität aufweisen als studentische Arbeiten, insbesondere in Bezug auf linguistische Merkmale, was Fragen nach Fairness und Vergleichbarkeit von Prüfungsleistungen aufwirft. Darüber hinaus werden traditionelle Prüfungsformate durch die Leistungsfähigkeit von KI-Tools infrage gestellt. Diese lösen einfache Wissensabfragen mit hoher Präzision und könnten Leistungen von Studierenden übertreffen, wodurch Prüfungen, die auf bloßer Reproduktion von Faktenwissen beruhen, an Aussagekraft verlieren. Ein möglicher Ansatz, der über reine Wissensabfrage hinausgeht, besteht darin, KI-Tools in die Lehr- und Prüfungsgestaltung aktiv zu integrieren, wobei auch ethische Aspekte und mögliche Bedenken von Studierenden berücksichtigt werden sollten. Vorgeschlagen wird, z. B. *ChatGPT* ähnlich einzusetzen wie Taschenrechner in der Mathematik (vgl. ebd.): Zunächst werden im Sinne des *Constructive Alignments*, bei dem Lernziele, Lehrmethoden und Prüfungen aufeinander abgestimmt werden, grundlegende Konzepte gelehrt. Anschließend werden KI-Tools verwendet, um komplexere Lernziele zu erreichen. Gleichzeitig sollte sichergestellt werden, dass Studierende, die aus ethischen Gründen den Einsatz von KI-Tools ablehnen, alternative Lern- und Prüfungswege wählen können, um Chancengleichheit und akademische Fairness zu gewährleisten. Durch diesen gezielten und flexiblen Einsatz von KI, etwa im Rahmen eines sokratischen Dialogs,⁹ können Studierende dazu angeregt werden, tiefer in Themen einzutauchen und eigenständig Lösungen zu entwickeln (vgl. Hochschulforum Digitalisierung 2023).

In Hinblick auf die Anpassung der Allgemeinen Prüfungsordnung zeigen sich auch hochschulpolitische Implikationen. Zunehmend wird gefordert, Prüfungsordnungen flexibler zu gestalten, um den Einsatz von KI-Tools zu regulieren. Eine Möglichkeit bestünde darin, neue Prüfungsformate zu entwickeln, die die Verwendung von KI offenlegen und transparent machen. So könnten etwa Prompts, die Studierende zur Erzeugung von Texten genutzt haben, verpflichtend dokumentiert werden. Zudem bieten kreative und offene Fragestellungen eine Chance,

9 Der sokratische Dialog ist eine Methode, bei der durch gezielte Fragen und Antworten kritisches Denken gefördert wird, um die Nutzenden zu eigenen Erkenntnissen zu führen, ohne direkt Lösungen vorzugeben. Beispielsweise über die Seite https://poe.com/SokratischerDialog_5 (Abrufdatum: 15.04.2025) wird ermöglicht, mit verschiedenen KI-Modellen in Dialog zu treten. Diese Art von KI-Modell könnte in der Lehre eingesetzt werden, um Lernende dabei zu unterstützen, komplexe Themen eigenständig zu durchdringen und ihre analytischen Fähigkeiten zu schärfen.

die Nutzung von KI sinnvoll in den Prüfungsprozess zu integrieren und gleichzeitig die Eigenleistung der Studierenden zu sichern.

Es zeigt sich, dass Hochschulen nicht nur auf technologische Veränderungen reagieren müssen, sondern eine aktive Rolle bei der Gestaltung neuer, KI-kompatibler Prüfungsformate einnehmen sollten. Aktuelle hochschulpolitische Diskussionen deuten darauf hin, dass eine Anpassung bestehender Prüfungsordnungen unausweichlich ist, um akademische Integrität und innovative Lehr- und Lernmethoden zu fördern.

5.2 Notwendige Weiterentwicklung von Hochschulrichtlinien und KI-Strategien

Die Einführung generativer KI hat Hochschulen dazu gezwungen, ihre Richtlinien und Strategien zu überarbeiten. Insbesondere in der Lehre der Sozialen Arbeit, in der ethische und soziale Fragestellungen eine zentrale Rolle spielen, ist eine strategische Weiterentwicklung und nachhaltige Auseinandersetzung mit den Folgen des Umgangs mit KI-Tools von entscheidender Bedeutung.

Die Richtlinien von derzeit nur etwa 25 weltweit führenden Hochschulen fokussieren sich auf akademische Integrität, Prüfungsdesign und Kommunikation mit Studierenden (vgl. Moorhouse et al. 2023). Zukunftsfähige Lehre erfordert klare Regelungen zur Nutzung von KI und tiefere Einbindung von Studierenden und Lehrenden in die Entwicklung dieser Richtlinien. Studierende sollten die Möglichkeit haben, Perspektiven und Erfahrungen einzubringen, um praxisnahe Regelungen zu entwickeln, die sowohl die akademische Integrität schützen als auch den sinnvollen Einsatz von KI fördern. Lehrende benötigen gezielte Fortbildungen, um ihre eigene Kompetenz im Umgang mit KI zu erweitern und ihre Lehre entsprechend anzupassen. Dies könnte in Form von Workshops und Schulungen zur digitalen und ethischen Kompetenz erfolgen. Darüber hinaus sollten Lehrende befähigt werden, neue Prüfungsformate zu entwickeln, die sowohl den Einsatz von KI berücksichtigen als auch die Eigenleistung der Studierenden sichern. Schlussendlich muss die Weiterentwicklung von Hochschulrichtlinien widerspiegeln, dass KI-Tools einen festen Bestandteil der zukünftigen Arbeitswelt darstellen. Daher sollten Hochschulen die Entwicklung von digitalen und ethischen Kompetenzen aktiv fördern, um Studierende und Lehrende auf diese Herausforderungen vorzubereiten.

6 Fazit

Die Integration von KI in die Lehre der Sozialen Arbeit birgt Chancen und Herausforderungen. Hochschulpolitische Entscheidungen spielen eine zentrale Rol-

le, um klare Regelungen und Richtlinien zu schaffen, die akademische Integrität und ethische Standards wahren. Hochschulpolitik muss hier proaktiv handeln, um die zukünftige Bildungslandschaft zu gestalten und Chancen der KI-Integration optimal nutzen zu können, denn KI-Kompetenzen dürfen nicht zum Privileg für wenige werden. Lehr- und Lernangebote müssen offen und partizipativ gestaltet sein (vgl. Pinkwart/Rampelt/de Witt 2020, S. 5 f.). Die Veränderungen in Lehr- und Lernprozessen erfordern zudem eine Anpassung der Hochschuldidaktik und -methodik. Insbesondere ist es wichtig, die Selbstständigkeit der Studierenden zu fördern und eine kritische Auseinandersetzung mit KI-Tools zu ermöglichen.

Aspekte der Ethik und Verantwortung sollten im Mittelpunkt des Diskurses um KI in der Lehre stehen, und es bedarf klarer ethischer Richtlinien, um eine verantwortungsvolle Nutzung von KI gewährleisten zu können. Die Hochschule nimmt in der Vermittlung dieser digitalen *Future Skills* eine Schlüsselrolle ein. Abschließend lässt sich festhalten, dass eine Balance zwischen technologischer Innovation und der Wahrung akademischer Standards notwendig ist. Hochschulpolitische Maßnahmen sollten darauf abzielen, die Potenziale von KI zu fördern, indem sie Transparenz, Verantwortung und partizipative Ansätze in den Mittelpunkt stellen. Dies erfordert eine kontinuierliche Überprüfung und Anpassung ethischer Standards, um eine faire, verantwortungsvolle und nachhaltige Nutzung zu gewährleisten. Die zukünftige Hochschulpolitik in Bezug auf Nutzung von KI in der Lehre wird die weitere Entwicklung der Studiengänge zur Sozialen Arbeit und die der gesamten Bildungslandschaft maßgeblich prägen.

Literatur

- Aldosari, Share Aiyed M. (2020): The Future of Higher Education in the Light of Artificial Intelligence Transformations. *International Journal of Higher Education* 9(3), S. 145–151.
- Balabdaoui, Fadoua/Dittmann-Domenichini, Nora/Grosse, Henry/Schlienger, Claudia/Kortemeyer, Gerd (2023): KI Nutzung unter Studierenden. Zusammenfassung einer Umfrage unter ETH Studierenden im September 2023. ethz.ch/de/die-eth-zuerich/lehre/ai-in-education/projects/ai-usage-among-students.html (Abfrage: 15.06.2025).
- Benner, Dietrich (2017): Bildung und Kompetenz. Von der kategorialen Bildung zur Kompetenzorientierung unterrichtlichen Lehrens und Lernens? Überlegungen zur Bedeutung von Wolfgang Klafki Studien zur Bildungstheorie und Didaktik für eine pädagogisch und kompetenztheoretisch ausgewiesene Didaktik, Unterrichts- und Bildungsforschung. In: Braun, Karl-Heinz/Stübig, Frauke/Stübig, Heinz (Hrsg.): *Erziehungswissenschaftliche Reflexion und pädagogisch-politisches Engagement. Wolfgang Klafki weiterdenken*. Wiesbaden: Springer VS, S. 73–91.
- Bologna-Erklärung (1999): The Bologna Declaration of 19 June 1999: Joint declaration of the European Ministers of Education. [ehea.info/media/ehea.info/file/Ministerial_conferences/02/8/1999_Bologna_Declaration_English_553028.pdf](https://www.ehea.info/media/ehea.info/file/Ministerial_conferences/02/8/1999_Bologna_Declaration_English_553028.pdf) (Abfrage: 15.06.2024).
- Botzum, Edeltraud/Gergen, Andrea (2022): „Peer-Mentoring als didaktisches Konzept. Umsetzung einer innovativen Lehrmethode im Rahmen des Moduls Literatur- und Forschungswerkstatt Soziale Arbeit an der TH Rosenheim“. In: *QLS TH Rosenheim: Lehre Aktuell* 2, S. 2–3.

- Botzum, Edeltraud/Neumaier, Stefanie (2023): Künstliche Intelligenz und digitale Transformationsprozesse in der Lehre Sozialer Arbeit. In: *Jugendhilfe* 61(5), S. 388–394.
- Buck, Isabella/Huemer, Birgit/Limburg, Anika (2024): KI im Schreibzentrum? Ein Plädoyer für offenen Diskurs und Kollaboration. In: Liebetanz, Franziska/Dalesandro, Leonardo/Mackus, Nicole/Alagöz-Bakan, Özlem (Hrsg.): *Künstliche Intelligenz in der Schreibzentrumsarbeit: Perspektiven auf die KI-induzierte Transformation*. In: *JoSch – Journal für Schreibwissenschaft* 15(1), S. 4–7.
- Deutscher Ethikrat (2023a): *Mensch und Maschine – Herausforderungen durch Künstliche Intelligenz Stellungnahme*. Berlin. <https://www.ethikrat.org/publikationen/stellungnahmen/mensch-und-maschine/>(Abfrage: 15.06.2025).
- Deutscher Ethikrat (2023b): *Mensch und Maschine – Herausforderungen durch Künstliche Intelligenz (Stellungnahme – Kurzfassung)*. Berlin. <https://www.ethikrat.org/publikationen/stellungnahmen/mensch-und-maschine/> (Abfrage: 15.06.2025).
- Di Vaio, Assunta/Palladino, Rosa/Hassan, Rohail/Escobar, Octavio (2020): Artificial intelligence and business models in the sustainable development goals perspective: A systematic literature review. In: *Journal of Business Research* 121, S. 283–314.
- Ehlers, Ulf-Daniel/Eigbrecht, Laura/Horstmann, Nina/Matthes, Wibke/Piesk, David/Rampelt, Florian (2024): *Future Skills für Hochschulen: Eine kritische Bestandsaufnahme*. In: *Stifterverband für die Deutsche Wissenschaft (Hrsg.): Future Skills lehren und lernen: Schlaglichter aus Hochschule, Schule und Weiterbildung*. Online-Vorveröffentlichung. <https://www.future-skills.net/analysen/vorveroeffentlichung-future-skills-fuer-hochschulen-eine-kritische-bestandsaufnahme> (Abfrage: 15.06.2025).
- Engelhardt, Emily (2024): Einsatz generativer KI in der Berufsausbildung. Lehrmethoden und Praxisbeispiele aus dem Hochschulkontext. In: Neumaier, Stefanie/Dörr, Madeleine/Botzum, Edeltraud (Hrsg.): *Praxishandbuch. Digitale Projekte in der Sozialen Arbeit*. Weinheim und Basel: Beltz Juventa, S. 63–79.
- Europäische Kommission (2020): *Mitteilung der Kommission an das Europäische Parlament, den Rat, den Europäischen Wirtschafts- und Sozialausschuss und den Ausschuss der Regionen. Aktionsplan für digitale Bildung 2021–2027. Neuaufstellung des Bildungswesens für das digitale Zeitalter*. education.ec.europa.eu/de/focus-topics/digital-education/action-plan/(Abfrage: 15.06.2025).
- Fritze, Jana (2024): Den Umgang mit KI-Schreibtools fördern – erste Ansätze aus Wuppertal. In: Liebetanz, Franziska/Dalesandro, Leonardo/Mackus, Nicole/Alagöz-Bakan, Özlem (Hrsg.): *Künstliche Intelligenz in der Schreibzentrumsarbeit: Perspektiven auf die KI-induzierte Transformation*. *JoSch – Journal für Schreibwissenschaft* 15(1), S. 49–64.
- Heinlein, Michael/Huchler, Norbert (Hrsg.) (2024): *Künstliche Intelligenz, Mensch und Gesellschaft. Soziale Dynamiken und gesellschaftliche Folgen einer technologischen Innovation*. Wiesbaden: Springer Fachmedien.
- Herbold, Steffen/Hautli-Janisz, Annette/Heuer, Ute/Kikteva, Zlata/Trautsch, Alexander (2023): A large-scale comparison of human-written versus ChatGPT – generated essays. In: *Scientific Reports* 13, S. 18617. <https://doi.org/10.1038/s41598-023-45644-9>
- Hochschulforum Digitalisierung (2023): *Im Sokratischen Gespräch mit KI. hochschulforumdigitalisierung.de/im-sokratischen-gespraech-mit-ki/*(Abfrage: 15.06.2025).
- Ifenthaler, Dirk (2023): *Ethische Perspektiven auf Künstliche Intelligenz im Kontext der Hochschule*. In: Schmohl, Tobias/Watanabe, Alice/Schelling, Kathrin (Hrsg.): *Künstliche Intelligenz in der Hochschulbildung. Chancen und Grenzen des KI-gestützten Lernens und Lehrens*. Bielefeld: transcript, S. 71–86.
- Mittmann, Michelle/Roeske, Adrian/Weber, Joshua/Remke, Sara/Schiffhauer, Birte (2023): *Studium Soziale Arbeit: Erkenntnisse zur curricularen Verankerung der digitalen Transformation*. In: Köttig, Michaela/Kubisch, Sonja/Spatscheck, Christian (Hrsg.): *Geteiltes Wissen. Wissens-*

- entwicklung in Disziplin und Profession Sozialer Arbeit. Opladen, Berlin und Toronto: Barbara Budrich, S. 237–249.
- Moorhouse, Benjamin Luke/Yeo, Marie Alina/Wan, Yuwei (2023): Generative AI tools and assessment: Guidelines of the world's top-ranking universities. In: *Computers and Education Open* 5, S. 100151. <https://doi.org/10.1016/j.caeo.2023.100151>
- Organisation für wirtschaftliche Zusammenarbeit und Entwicklung (OECD) (2005): Definition und Auswahl von Schlüsselkompetenzen. Zusammenfassung. https://www.yumpu.com/de/document/read/22481288/definition-und-auswahl-von-schlasselkompetenzen#google_vignette/ (Abfrage: 15.06.2025).
- Pinkwart, Niels/Rampelt, Florian/de Witt, Claudia (Hrsg.) (2020): Whitepaper. Künstliche Intelligenz in der Hochschulbildung. https://www.ki-campus.org/sites/default/files/2020-10/Whitepaper_KI_in_der_Hochschulbildung.pdf (Abfrage: 15.06.2026).
- Poe (o. J.): Sokratischer Dialog. poe.com/SokratischerDialog_5 (Abfrage: 15.06.2025).
- Raber, Vanessa (2012): Schlüsselkompetenzen in der Hochschullehre. Zum Bologna-Prozess und seinen Chancen für einen Paradigmenwechsel in der Lehre. München: AVM.
- Reinmann, Gabi (2023): Deskillung durch Künstliche Intelligenz? Potenzielle Kompetenzverluste als Herausforderung für die Hochschuldidaktik. Diskussionspapier Nr. 25. Berlin: Hochschulforum Digitalisierung.
- Renz, André/Etsiwah, Bennet (2020): Datenkultur und KI-Kompetenz an Hochschulen. In: Pinkwart, Niels/Rampelt, Florian/de Witt, Claudia (Hrsg.): Whitepaper. Künstliche Intelligenz in der Hochschulbildung. Berlin, S. 35–37.
- Ständige Wissenschaftliche Kommission der Kultusministerkonferenz (SWK) (2024): Large Language Models und ihre Potenziale im Bildungssystem. Impulspapier der Ständigen Wissenschaftlichen Kommission (SWK) der Kultusministerkonferenz. https://www.pedocs.de/volltexte/2024/28303/pdf/SWK_2024_Large_Language_Models.pdf (Abfrage: 15.06.2025).
- Stifterverband für die Deutsche Wissenschaft e. V. (Hrsg.) (2021): Future Skills 2021. 21 Kompetenzen für eine Welt im Wandel (Diskussionspapier Nr. 3). <https://www.stifterverband.org/medien/future-skills-2021/> (Abfrage: 15.06.2025).
- Tobor, Jens (2024): Blickpunkt – Leitlinien zum Umgang mit generativer KI. Berlin: Hochschulforum Digitalisierung. Version 1.0. https://www.hochschulforumdigitalisierung.de/wp-content/uploads/2024/02/HFD_Blickpunkt_KI-Leitlinien_final.pdf (Abfrage: 15.06.2026).
- von Garrel, Joerg/Mayer, Jana/Mühlfeld, Markus (2023): Künstliche Intelligenz im Studium. Eine quantitative Befragung von Studierenden zur Nutzung von ChatGPT & Co. https://doi.org/10.48444/h_docs-pub-395
- von Humboldt, Wilhelm (1985): *Bildung und Sprache*. Besorgt von Clemens Menze. 4., durchgesehene Auflage. Paderborn: Schöningh.
- Watanabe, Alice (2023): Studierende im KI-Diskurs. Wie Studierende in einem Workshopformat über den KI-Einsatz informiert und zum Nachdenken über KI-gestütztes Lehren und Lernen angeregt werden. In: Schmohl, Tobias/Watanabe, Alice/Schelling, Kathrin (Hrsg.): *Künstliche Intelligenz in der Hochschulbildung. Chancen und Grenzen des KI-gestützten Lernens und Lehrens*. Bielefeld: transcript, S. 99–118.
- Wintergerst, Ralf (2024): *Digitale Hochschulen*. Bitkom Research 2024. https://www.bitkom.org/sites/main/files/2024-03/240321Bitkom-PräsentationPK_Studentenbefragungfinal.pdf (Abfrage: 15.08.2025).
- Witter, Stefanie/Meinhardt-Injac, Bozana/Siemer, Lutz/Späte, Julius (2024): ChatGPT im Studium der Sozialen Arbeit. Eine quantitative Studie zur Nutzung, Bewertung und Thematisierung in der Hochschule aus Studierendensicht. <https://doi.org/10.34678/opus4-3382>
- Zorn, Isabel/Seelmeyer, Udo (2015): Digitale Technologien in der Sozialen Arbeit – Zur Notwendigkeit einer technischen Reflexivität. In: *Der pädagogische Blick – Zeitschrift für Wissenschaft und Praxis in pädagogischen Berufen* 23(3), S. 134–146.

Künstliche Intelligenz und Soziale Arbeit: Ausblick und Perspektiven¹

Gesa Linnemann, Julian Löhe, Beate Rottkemper

Abstract: Der Beitrag skizziert Perspektiven für die Soziale Arbeit im Umgang mit Künstlicher Intelligenz (KI). Er zeigt auf, wie tiefgreifend KI gesellschaftliche Strukturen, Interaktionen und professionelle Praxis verändert – und betont die Notwendigkeit, Technik, Gesellschaft und Profession gemeinsam zu denken. Anhand von Beispielen wie Data Bias, Automation Bias und Explainable AI wird dargelegt, dass KI nicht neutral ist, sondern soziale Ungleichheiten reproduzieren kann. Der Text benennt zentrale Herausforderungen: ethische Fragen, fehlende Ressourcen, hohe Kosten, technologische Komplexität, Nachhaltigkeitsaspekte und regulatorische Anforderungen. Er plädiert für Kompetenzaufbau, kritische Reflexion, Beteiligung an Technikgestaltung und verantwortungsvolle Nutzung – als Voraussetzung für eine professionelle, menschenrechtsorientierte Soziale Arbeit im Zeitalter KI-gestützter Systeme.

Keywords: Künstliche Intelligenz, Soziale Arbeit, Professionsentwicklung, Zukunft

In Forschung, Praxis und Lehre hat sich die Soziale Arbeit, wie der vorliegende Band „Künstliche Intelligenz in der Sozialen Arbeit – Grundlagen für Theorie und Praxis“ umfangreich zeigt, auf den Weg gemacht, sich mit Chancen und Gestaltungsmöglichkeiten sowie problematischen gesellschaftlichen Auswirkungen von KI auseinanderzusetzen. Die bereits gegangenen Schritte sind wichtig und anerkennenswert, dennoch ist zu konstatieren, dass die Profession angesichts tiefgreifenden Wandels durch die Technologie KI weiterhin nicht nur *nach*-forschen, sondern auch *vor*-denken und *mit*-gestalten muss. Warum? Wir stützen diese Haltung auf folgende Aspekte:

Als *general purpose technology* durchdringt KI unterschiedlichste Lebensbereiche und Lebenswelten. Sie kann in Form von Sentimentanalysen mediale Diskurse beeinflussen, durch den Einsatz von Computer Vision in den öffentlichen Raum

1 © Die Autor:innen 2025, veröffentlicht im Verlag Beltz Juventa, Open Access: CC BY 4.0
Gesa Linnemann / Julian Löhe / Beate Rottkemper (Hg.), Künstliche Intelligenz in der Sozialen Arbeit
10.3262/978-3-7799-8562-4_018

und seine Wahrnehmung verändern und Große Sprachmodelle können anstelle von Freund:innen, Therapeut:innen, Sexualpartner:innen etc. konsultiert werden. Statt sozialer Netzwerke können auch stark oder sogar ausschließlich KI-beispielte „Netzwerke“ genutzt werden, Arbeitsprofile und Aufgabenzuschnitte ändern sich je nach Tätigkeit mit dem Einsatz von KI zum Feinschliff einzelner Aufgaben bis zur Transformation ganzer Berufsbilder, bei Entscheidungen können durch den KI-Einsatz hunderte Kriterien und sehr große Datensätze herangezogen und als Grundlage genutzt werden. Anhand des Problems von Verzerrungen betrachten wir im Folgenden beispielhaft, warum Technik, Gesellschaft und Profession unbedingt zusammen betrachtet werden müssen.

KI-Systeme in der Sozialen Arbeit sind fehleranfällig, weil sie auf Daten basieren, die gesellschaftliche Verzerrungen enthalten. Der sogenannte „Data Bias“ (deutsch: Verzerrungen in Daten) in KI-Systemen entsteht, wenn Trainingsdaten beim Maschinellen Lernen (siehe den Beitrag von Rottkemper in diesem Band) historische Vorurteile abbilden und diese reproduzieren (vgl. Noble 2018; vgl. Ruha 2019; vgl. Eubanks 2018). Ein Beispiel: In Allegheny County (USA) wurde ein Kinderschutz-Algorithmus kritisiert, weil er Familien in Armut systematisch höheren Risiken in Bezug auf Kindeswohlgefährdung zuordnete – nicht aufgrund realer Gefährdung, sondern weil Armut in den Datensätzen mit behördlicher Intervention korrelierte (vgl. Eubanks 2018, S. 163).

KI ist nie neutral. Ihre Entwicklung und Datenquellen sind heutzutage in der Regel von der westlichen, weißen, wohlhabenden Welt geprägt (vgl. Buolamwini/Gebru 2018, S. 1; vgl. Noble 2018). Gesichtserkennungssysteme z. B. erkennen BIPoC²-Frauen deutlich schlechter als weiße Männer (vgl. Buolamwini/Gebru 2018, S. 1). Aus diesen Verzerrungen in den Daten entstehen reale Diskriminierungen aufgrund von Geschlecht, (sozialer) Herkunft, Hautfarbe und Persönlichkeitsmerkmalen, wie wissenschaftliche Studien und Beiträge zeigen (vgl. Chen 2023; vgl. Leicht/Karst/Zimmer 2020; vgl. Orwat 2020).

Hinzu kommt der „Automation Bias“: Fachkräfte neigen dazu, KI-Empfehlungen unkritisch zu folgen (vgl. Deutscher Ethikrat 2023, S. 32). Selbst fehlerhafte Vorschläge können übernommen werden, wenn die Technik als effizient wahrgenommen wird (vgl. Vered et al. 2023, S. 1f.). Das schwächt professionelles Urteilen und kann zu Fehlentscheidungen führen, z. B. bei Kindeswohlprognosen oder Fallpriorisierungen.

Die Mensch-Maschine-Interaktion wird dadurch grundlegend bestimmt. KI kann Fachkräfte entlasten, aber auch entmündigen, wenn Ergebnisse aus der Black Box eines KI-Systems unkritisch übernommen werden. Explainable

2 BIPoC ist ein zusammengesetztes Initialwort für Black, Indigenous and People of Colour. Es handelt sich um eine politisch motivierte Selbstbezeichnung, die unterschiedliche Erfahrungen mit Rassismus sichtbar machen und gleichzeitig solidarisch verbinden möchte (vgl. Gamgami 2024, S. 7f.).

AI (deutsch: erklärbare KI) ist daher ein entscheidendes Forschungsfeld, damit Fachkräfte Verantwortung übernehmen können (vgl. Floridi et al. 2018, S. 702). Zu konstatieren ist aber auch, dass Explainable AI trotz des zunehmenden Forschungsinteresses (vgl. Ali et al. 2023) in der Praxis noch nicht umfangreich zum verantwortlichen Handeln und Entscheiden mit KI beiträgt. Aufsehen erregte Anthropic's Untersuchung der Funktionsweise seines Modells Claude 3.5 Haiku. Die detaillierte Analyse der Modellmechanismen setzt einen Gegenpunkt zur allgemeinen Auffassung, Sprachmodelle wie Claude 3.5 Haiku raten bloß statistisch plausible Wörter. Vielmehr zeigt die Analyse interne „Denkprozesse“ (Zwischenschritte, Zielplanung, Rückwärtsdenken, metakognitive Einschätzungen) auf, die in Ansätzen nachvollziehbar und manipulierbar sind – eine Voraussetzung dafür, dass Mensch-Maschine-Interaktion nicht zur Black Box verkommt, sondern gestaltbar bleibt (vgl. Lindsey et al. 2025). Gleichzeitig ist feststellbar, dass Modelle nicht nur Inhalte, sondern auch eigene Aktionen erfinden – so behauptet z. B. o3 von OpenAI, einen Code auf seinem eigenen Laptop ausgeführt zu haben (vgl. Chowdhury et al. 2025). Absurd, denn dieses Sprachmodell kann nichts außerhalb seiner eigenen Rechenumgebung ausführen und schon gar nicht auf seinem eigenen Gerät. An solchen Beispielen wird deutlich, dass auch das vermeintliche Reasoning großer KI-Modelle nicht die tatsächlichen Entscheidungsgründe offenbart – ein ernüchternder Befund für die Explainable AI: Ihre Rückverfolgung kann bestenfalls wenig nützen, schlimmstenfalls in die Irre führen; denn auch wenn die Ergebnisse häufig plausibel oder kreativ erscheinen, reicht ein oberflächliches menschliches Abnicken bei wichtigen Entscheidungen nicht aus. Es braucht den „Human-in-the-Loop“ auch weiterhin (vgl. Lindsey et al. 2025).

Die Soziale Arbeit muss KI-Systeme kritisch prüfen: Wer wird wie repräsentiert? Welche Annahmen stecken in Daten und Code? Und wie können Bias und Diskriminierung reduziert werden? Möglich wäre das z. B. durch einen Bias-Audit (vgl. Heuer et al. 2021). Praktisch könnte das folgendermaßen aussehen: Ein Träger will ein KI-gestütztes System zur Risikoeinschätzung einführen. Vor dem Einsatz wird ein Bias Audit durchgeführt, das u. a. folgende Fragen beleuchtet:

- Welche Daten wurden im Training des Modells verwendet? Entsprechen die Daten der Gesamtheit der adressierten Zielgruppen?
- Wie schneidet das System für verschiedene Gruppen ab (z. B. Kinder mit Migrationsgeschichte versus ohne)?
- Gibt es systematische Unterschiede?
- Andere Methoden zur kritischen Überprüfung von KI-Systemen sind Diversität in Entwickler:innenteams oder Ethikboards in Sozialen Organisationen. KI kann unterstützen, aber nur wenn sie transparent, diversitätssensibel und kritisch reflektiert eingesetzt wird.

Außerdem wird zum Erzeugen von Dependable AI (deutsch: zuverlässiger KI), im Rahmen von Großen Sprachmodellen auch in Projekten, die die Soziale Arbeit be-

treffen, auf Retrieval Augmented Generation (RAG) zurückgegriffen (z. B. Projekt SuchtGPT, Delphi Gesellschaft 2025, ab min 33). RAG ist ein Verfahren, bei dem Große Sprachmodelle (LLM) durch eine externe Wissensquelle ergänzt werden. Statt rein auf das im Modell gespeicherte Wissen zuzugreifen, ruft das System in Echtzeit relevante Informationen aus einer Datenbank oder einem Dokumentenindex ab und integriert diese in die Antwortgenerierung (vgl. Gao et al. 2023, S. 1). RAG ist aber auch keine Garantie dafür, um inhaltliche Fehler und Halluzinationen komplett ausschließen zu können (vgl. Wu/Wu/Zou 2025). Hier zeigt sich beispielhaft die Relevanz für die Soziale Arbeit, sich auch mit den technologischen Entwicklungen ganz konkret auseinanderzusetzen, um den fachlichen Einsatz beurteilen zu können. Ebenso sollte Soziale Arbeit als Menschenrechtsprofession auch weitere Aspekte von KI-Nutzung kritisch beobachten (siehe Linnemann/Löhe/Rottkemper 2024), die sich gesellschaftlich genauso wie lebensweltlich (siehe Kapitel zu einzelnen Handlungsfeldern in diesem Band, z. B. Schломann und Steiner) auswirken.

Das betrifft auch den Bereich der öffentlichen Sicherheit und „digitale Bürgerrechte“, zwei Beispiele sind die Diskussion um die „Chatkontrolle“ und die Überwachung des öffentlichen Raumes mittels Gesichtserkennung (diskutiert im Rahmen des „Sicherheitspakets der Bundesregierung“ 2024 und der entsprechend vorbereiteten Gesetzesentwürfe). Globale und geopolitische Auswirkungen hat der Einsatz von KI in der Rüstungsindustrie. Wessen Stimme hier wie gehört wird, stellt sich auch als Frage im internationalen Kontext. Tech-Konzerne wirken hier nicht als reine „Hersteller“, sondern nehmen aktiv Einfluss auf den Diskurs und tragen somit Verantwortung (z. B. Artikel von Anthropic-Chef Amodei 2024) über die Einflussnahme der Gestaltung ihrer Produkte hinaus.

Der vorliegende Band „Künstliche Intelligenz in der Sozialen Arbeit – Grundlagen für Theorie und Praxis“ versteht sich als Orientierung und erstes Konvolut für die fundierte Auseinandersetzung. Dass dies nur Startpunkt und in keinsten Weise Abschluss des Themas sein kann, zeigen beispielhaft auch die folgenden Punkte, die immer wieder angesichts aktueller gesellschaftlicher und technologischer Entwicklung diskutiert und weiterbearbeitet werden müssen:

1. Unterstützungsbedarf bei KI-Nutzung/Auswirkungen

Die Gefahr einer digitalen Spaltung (van Dijk 2017) besteht bei KI ebenso wie bei anderen Technologien, wird jedoch durch die Wirkmächtigkeit der Technologie möglicherweise noch einmal verstärkt. Die Modelle sind in der Regel sehr ressourcen- und ggf. kostenintensiv. Das führt dazu, dass nicht alle Menschen gleichermaßen Zugang zu den Modellen und aus ihnen resultierenden Tools haben, geschweige denn sich an der Entwicklung beteiligen können. Nutzungspraktiken können sich auch stark unterscheiden, d. h., der Einsatz von KI kann von Nutzenden so gestaltet werden, dass eine Zunahme an Kompetenzerwerb, Selbstbestimmung und Wohlbefinden erfolgt oder aber im Gegenteil es zu stärkerer Un-

selbständigkeit und Verunsicherung kommt. Klient:innen sind in verschiedenen Bereichen betroffen, etwa bei automatisierter Kreditvergabe oder beim Umgang mit KI-gestützter Diagnostik. Klient:innen müssen genauso wie Fachkräfte (siehe Punkt 8 und die Beiträge von Botzum et al. sowie Dötterl in diesem Band) befähigt werden, mit KI umzugehen.

2. KI als „Konkurrenz“ zu Angeboten der Sozialen Arbeit

Soziale Arbeit muss sich als Profession weiterentwickeln und positionieren (siehe den Beitrag von Beranek in diesem Band). Angesichts von digitalen Gesundheitsanwendungen und der Nutzung von Großen Sprachmodellen „zur Selbsthilfe“ einerseits und Fachkräftemangel andererseits müssen Niedrigschwelligkeit und Zugänglichkeit der eigenen Angebote kritisch überprüft, gleichzeitig professionelle Standards auch verteidigt werden (siehe auch den Beitrag von Löhe in diesem Band). Die Profession steht vor der Frage, ob eigene KI-gestützte Angebote gemacht werden sollen und was es dazu braucht (siehe den Beitrag von Dummann in diesem Band).

3. Theorieentwicklung zu KI in der Sozialen Arbeit steht am Anfang

Auch wenn KI mittlerweile in den Fokus der Sozialen Arbeit gerückt ist und Gegenstand verschiedener Projekte, Lehr- und Forschungsvorhaben ist, scheint der Versuch, die sich durch KI veränderten Bedingungen auf sozialarbeiterischer Seite theoretisch zu erfassen, mit Ausnahme von Beranek (vgl. Beranek 2021) wenig unternommen worden zu sein. In diesem Band legt sie in ihrem Beitrag einen Diskussionsvorschlag zur theoretischen Erfassung vor.

4. Technologische Entwicklung

Soziale Arbeit muss sowohl die Technologien in ihren Grundzügen verstehen als auch die Interessen der Unternehmen, die sie anbieten. Hier ist die Disziplin gefordert, Angebote einzuordnen und kritisch zu hinterfragen (siehe auch Punkt 1) und/oder für eigene Zwecke nutzbar zu machen. Dazu gehört die Bereitschaft, die Initiative und konsequenterweise auch die Forderung danach, in Entwicklungsprozessen eine aktive Rolle zu spielen. Wie schon mehrfach angesprochen, können technische Entwicklungen nicht isoliert von gesellschaftlichen und politischen Auswirkungen diskutiert werden, wie etwa die Veröffentlichung von Deepseek oder des KI-Agenten Manus AI gezeigt haben.

Für das Jahr 2025 haben die Vereinten Nationen das „Quantenjahr“ (<https://quantum2025.org>) ausgerufen. Quanten-Computing würde einen enormen Zuwachs an Geschwindigkeit ermöglichen, was wiederum Auswirkungen auf KI-Einsatzmöglichkeiten und Datensicherheit hätte. Die gesellschaftlichen Folgen sind womöglich erheblich.

Die Integration von KI-Technologien in Robotik oder in das „Internet der Dinge“ wie in die „AI Glasses“ von Meta stellen Gesellschaft, Profession und Klient:in-

nen auch schon zeitnah vor weitere Fragen und Entscheidungen, über die in diesem Band auf KI-Systeme, betrieben auf PCs und mobilen Endgeräten, bezogenen hinaus.

In einem medial vermittelten Dauerstrom an Neuerungen und Überarbeitungen, bei einer wachsenden Zahl an verfügbaren Modellen (siehe <https://huggingface.co>, Abrufdatum 15.04.2025), ist die Herausforderung, hier Relevantes von Hype zu trennen und Entwicklungen kritisch zu begleiten, nicht von Einzelnen zu leisten, sondern bedarf der fachlichen Auseinandersetzung, die in der Praxis zugänglich gemacht wird.

5. Regulatorische Anforderungen an die Soziale Arbeit

In der Sozialen Arbeit wird mit besonders sensiblen Daten von oftmals schutzbedürftigen Personengruppen gearbeitet. Die Nutzung dieser Daten fordert besondere Maßnahmen in Bezug auf Datenschutz und Datensicherheit (vgl. Goldberg 2021). Eine Verarbeitung der Daten außerhalb der eigenen Infrastruktur ist in Deutschland oftmals nicht zulässig bzw. nur unter Einhaltung strenger Auflagen. Wie im Beitrag von Dötterl in diesem Band erläutert, handelt es sich bei Systemen zur Verarbeitung von gesundheitsbezogenen Daten oftmals um Hochrisiko-KI-Systeme im Rahmen des EU AI Acts. Das sorgt dafür, dass Anbieter und Betreiber entsprechende Auflagen bezüglich Transparenz und Datenschutz in Training und Betrieb erfüllen müssen.

6. Entwicklung und Finanzierung

Vielfach wird das Potenzial von KI auch aus Effizienzgesichtspunkten diskutiert – inwiefern z. B. administrative Prozesse oder auch Kernaufgaben in der Sozialen Arbeit von KI-gestützten Anwendungen unterstützt werden können, um Kosten zu sparen und Fachkräfte zu entlasten: etwa wenn einfache und stark repetitive Aufgaben an KI-gestützte Anwendungen übergeben würden oder Chatbots in der Erstberatung zur Anwendung kämen, ebenso die Unterstützung von KI bei der Formulierung von Texten (vgl. Pottharst et al. 2024; Krings/Heister 2023; Kreidenweis/Diepold 2024). Ein Projekt zur Textgenerierung in der Sozialen Arbeit wird in diesem Band von Plafky et al. vorgestellt. Weitere Beispiele aus der Diskussion zählt Löhe in diesem Band auf. Was jedoch weitgehend unbeachtet bleibt, sind belastbare Schätzungen zu den Kosten, die einhergehen mit

- a) den initialen Investitionskosten (Kosten für Anschaffung von Hard- und Software) sowie
- b) den Implementierungskosten (Schulung der Mitarbeitenden sowie Aufbau technischer Infrastruktur)

Die Investitionskosten variieren stark in Abhängigkeit davon, welches KI-System benötigt wird. Ein internationales Beratungsunternehmen unterscheidet und beziffert z. B. wie folgt:

Abbildung 1: Wesentliche Phasen im Lebenszyklus eines LLM

Kategorie	Geschätzte Kostenspanne	Beispiele	Merkmale/Anforderungen
Grundlegende KI-Lösungen	20.000–80.000 US-Dollar	Chatbots, Empfehlungssysteme, Bilderkennung (einfach), Stimmungsanalyse, grundlegende Daten- und prädiktive Analyse	Nutzung vortrainierter Modelle oder Application Programming Interfaces (APIs, dt. Programmierschnittstellen), geringe individuelle Entwicklung, Fokus auf Integration bestehender Technologien, Schritte: Datenaufbereitung, Modellauswahl/-optimierung, einfache Benutzeroberfläche
Fortschrittliche KI-Lösungen	50.000–150.000 US-Dollar	Risikomanagement, personalisierte Lernsysteme, Kundensegmentierung, Computer Vision, Workflow-Automatisierung, Inhaltsplattformen, Betrugserkennung	individuelles Modelltraining, umfangreiche Datenverarbeitung, Integration in komplexe Systeme, oft Kombination mehrerer KI-Technologien, erfordert intensivere Tests und Optimierungen
Benutzerdefinierte KI-Lösungen	100.000–500.000+ US-Dollar	Handelsplattformen, vorausschauende Wartung (z. B. Industrie), medizinische Diagnosesysteme	umfassende Forschung und Entwicklung, Erhebung und Aufbereitung großer Datenmengen, Entwicklung neuartiger Algorithmen, Integration in spezialisierte Hardware/Software, hohe regulatorische Anforderungen, längere Entwicklungszyklen, Einsatz hochqualifizierter Fachkräfte

Quelle: Vgl. Coherent Solutions 2025

Eine belastbare Kostenschätzung ist u. a. auch deshalb schwierig, weil Sozialarbeitende oftmals nicht ausreichendes Wissen über KI-Systeme haben, um beurteilen zu können, wie ihnen KI in der täglichen Arbeit helfen könnte und welche Form von KI-Systemen sie benötigen. Klar ist aber, dass es auch beim Einsatz von schon vorhandenen KI-Anwendungen der Schulung von Mitarbeitenden bedarf – die seit kurzen auch Pflicht gemäß KI-Verordnung ist (siehe den Beitrag von Dötterl in diesem Band). Es ist gleich aus zwei Gründen als „blauäugig“ zu beurteilen, wenn im Schwerpunkt die möglichen Vorteile von KI in den Fokus gerückt werden, ohne auch die Kosten seriös zu diskutieren: erstens sind die Kosten mitunter beachtlich, wie die oben genannte Auflistung exemplarisch veranschaulicht, und zweitens gibt es allgemein einen Trend in der Sozialen Arbeit, dass sich öffentliche Leistungsträger zunehmend aus der Finanzierung sozialer Dienstleistungen zurückziehen (vgl. Löhne/Aldendorff 2022, S. 10). Hinzu kommt, was schon aus der Diskussion um Digitalisierung bekannt ist: Fehlende Finanzierungs- und Refinanzierungsmöglichkeiten erschweren den Einsatz (vgl. Gillingham/Schiffhauer/Seelmeyer 2020, S. 646; Rösler 2018 et al., S. 37). Das trifft auch für KI in

der Sozialen Arbeit zu: Auch hier gibt es keine standardmäßige Grundfinanzierung, die über Kostenträger abgerufen werden kann. Die immensen Mehrkosten müssen Organisationen der Sozialen Arbeit aus Eigenmitteln stemmen oder anderweitig einwerben – wie z. B. durch Ausschreibungen von Stiftungen oder anderen privaten Geldgebern. Die Finanzierungssituation macht deutlich, dass darin ein bedeutender Hemmschuh für Innovationen in der Sozialen Arbeit zu identifizieren ist.

Eine frühzeitige aktive Rolle von Sozialarbeitenden in der Entwicklung von KI-Anwendungen ist unabdingbar, um die fachlich-professionelle Sicht in der Technologieentwicklung vertreten zu können. Die dafür nötigen zeitlichen Ressourcen, gerade in Explorations- und Entwicklungsphasen, mögen hier abschreckend wirken. Auch scheiternde Versuche, Prototypen aus der Projektphase hinauszuführen, können demotivierend sein. Dennoch ist die Beteiligung der Profession zentral.

7. Ethik

KI-Technologien werfen auf verschiedenen Ebenen ethische Fragen auf und stellen Herausforderungen an verantwortliches Handeln (siehe den Beitrag von Hefels in diesem Band). Bei der Genese von Großen Sprachmodellen sind insbesondere globale Aspekte zu betrachten: die Ungleichheiten zwischen Menschen sowie Organisationen, die durch den Betrieb von KI-Systemen profitieren, und Menschen, die in der Entwicklung der Systeme ausgebeutet werden. Zu nennen sind hier die Arbeitsbedingungen von sogenannten Annotation Workers, die teilweise traumatisierende Inhalte aus den Datenbeständen und Ergebnissen filtern müssen, damit die Modelle nutzbar werden, aber auch die Arbeits- und Lebensbedingungen, unter denen Rohstoffe wie Kobalt und Lithium gewonnen werden (vgl. Crawford 2021; Casilli/Tubaro 2022; Klinova/Korinek 2021).

Weitere ethische Fragestellungen tun sich im Rahmen des Trainings von Großen Sprachmodellen auf: von der Nutzung im Internet veröffentlichter Daten ohne Einverständnis der Urheber:innen und damit einhergehender Copyright-Verletzungen bis hin zu einer Verfestigung bestehender Muster in der Gesellschaft, die durch Techniken wie Verstärkerlernen der KI-Systeme durch menschliches Feedback wiederholt oder sogar intensiviert werden können (vgl. George/Baskar/Pandey 2024). Die Frage, wie sich eine Gesellschaft weiterentwickeln soll, wenn Entscheidungsgrundlagen zum Großteil auf Basis historischer Daten und mit teilweise verstärkten Vorurteilen entwickelt werden, bleibt zu diskutieren.

Auch bei der Nutzung Großer Sprachmodelle sind ethische Aspekte zu berücksichtigen, die im Folgenden nicht erschöpfend adressiert werden können, angefangen bei Unterschieden im Zugang zu Technologien zwischen verschiedenen Regionen auf der Welt, aber auch innerhalb von Gesellschaften (vgl. Buccella 2023), über bereits genannte Risiken in der professionellen Nutzung bezüglich

des Automation Bias (vgl. Abdelwanis et al. 2024). Letzterer ist besonders kritisch im Zusammenhang mit den bis heute in Großen Sprachmodellen auftretenden Halluzinationen zu bewerten (vgl. den Beitrag von Rottkemper in diesem Band, vgl. Koenecke et al. 2024). Darüber hinaus ist besonders im professionellen Setting das Risiko von sogenanntem Deskillung zu berücksichtigen. Wenn Entscheidungen im großen Maße durch KI-Systeme vorbereitet und teilweise sogar getroffen werden, führt das zu einer Verstärkung der Technikabhängigkeit, die so weit gehen kann, dass Fachkräfte ihre Aufgaben ohne die entsprechende Technikunterstützung nicht mehr durchführen können. Auch im Rahmen der Fachkräfteausbildung sind diese Gefahren zu adressieren (vgl. Reinmann 2023).

8. Nachhaltigkeit

Ebenso müssen die Umweltbedingungen beim Training Großer Sprachmodelle Berücksichtigung finden (vgl. Falk/van Wynsberghe 2024). Der Verbrauch an Energie ist so groß, dass alle großen Technologiekonzerne inzwischen den Betrieb eigener Atomkraftwerke planen (vgl. Deutschlandfunk 2024). Aber auch weitere Ressourcen wie Seltene Erden und nicht zuletzt Wasser werden in großen Mengen benötigt, um LLM und andere generative KI-Systeme trainieren und betreiben zu können. Der Verbrauch von Wasser und Energie übersteigt den Bedarf anderer IT-Technologien um ein Vielfaches (vgl. Crawford 2024; Falk/van Wynsberghe 2024). Besonders kritisch ist der Wasserverbrauch in Regionen zu sehen, die ohnehin von Trockenheit betroffen sind. Das Wasser wird vor allem für die Kühlung in den Rechenzentren benötigt (vgl. George/George/Martin 2023).

Um einerseits möglichst ressourcenschonend zu agieren und die Effizienzgewinne durch die KI-Nutzung andererseits manifestieren zu können, sollten sich Nutzer:innen bewusst mit der Zielsetzung der KI-Nutzung und mit deren Notwendigkeit beschäftigen.

9. Befähigung von (künftigen) Sozialarbeiter:innen

Wie die Befähigung und der Erwerb von KI-Kompetenz oder AI Literacy im Rahmen von Studium und Berufstätigkeit gelingen können, ist zunehmend Gegenstand der Diskussion und findet unter anderem Ausdruck in der Entwicklung der Future Skills (siehe den Beitrag von Botzum in diesem Band). Alles und Kolleginnen (2025) z. B. schlagen ein Kompetenzmodell mit mehreren Stufen vor, das die sogenannte „AI Leadership“ im Zentrum sieht.

Die Skizzierung der Aspekte in den Punkten 1–9 in diesem Kapitel deuten an, wie vielgestaltig die Anforderungen hinsichtlich Kompetenz und Befähigung für die Soziale Arbeit sind. Das vorliegende Buch versteht sich – so der Anspruch der Autor:innen – als Beitrag zur Vermittlung von KI-Kompetenz in der Sozialen Arbeit: Es soll durch seine Verwendung in der Lehre sowie als frei zugängliche Open-Access-Ressource auch in der Praxis Orientierung bieten und zur Befähigung im Umgang mit Künstlicher Intelligenz beitragen.

Literatur

- Abdelwanis, Mustafa/Alarafati, Hamdan Khalaf/Tammam, Maram Muhanad Saleh/Simsekler, Mecit Cam Emre (2024): Exploring the risks of automation bias in healthcare artificial intelligence applications: A Bowtie analysis. In: *Journal of Safety Science and Resilience* 5(4), S. 460–469.
- Alles, Susanne/Falck, Joscha/Flick, Manuel/Schulz, Regina (2025): KI-Kompetenzen für Lehrende und Lernende. Aus der Praxis für die Praxis – eine adaptierbare Basis. Virtuelles Kompetenzzentrum: Künstliche Intelligenz und wissenschaftliches Arbeiten (VK.KIWA). Blogbeitrag vom 13. März 2025. <https://www.vkkiwa.de/blog/ki-kompetenzen-fuer-lehrende-und-lernende/> (Abfrage: 15.06.2025).
- Ali, Sajid/Abuhmed, Tamer/El-Sappagh, Shaker/Muhammad, Khan/Alonso-Moral, Jose M./Confalonieri, Roberto/Guidotti, Riccardo/Del Ser, Javier/Díaz-Rodríguez, Natalia/Herrera, Francisco (2023): Explainable Artificial Intelligence (XAI): What we know and what is left to attain Trustworthy Artificial Intelligence. In: *Information Fusion* 99, S. 101805. <https://doi.org/10.1016/j.inffus.2023.101805>
- Amodei, Dario (2024): Machines of Loving Grace. How AI Could Transform the World for the Better. <https://darioamodei.com/machines-of-loving-grace#4-peace-and-governance> (Abfrage: 15.06.2025).
- Buccella, Alessandra (2023): „AI for all“ is a matter of social justice. In: *AI and Ethics* 3(4), S. 1143–1152.
- Buolamwini, Joy/Gebru, Timnit (2018): Gender Shades: Intersectional Accuracy Disparities in Commercial Gender Classification. In: *Proceedings of Machine Learning Research (PMLR)* 81, S. 1–15.
- Casilli, Antonio A./Tubaro, Paola (2022): An End-to-End Approach to Ethical AI: Socio-Economic Dimensions of the Production and Deployment of Automated Technologies. https://polytechnique.hal.science/hal-04027470/file/Casilli_Tubaro%20E2E%20HODSS.pdf (Abfrage: 15.06.2025).
- Chen, Zhisheng (2023): Ethics and discrimination in artificial intelligence-enabled recruitment practices. In: *Humanities and Social Sciences Communications* 10, Artikelnummer 567. <https://doi.org/10.1057/s41599-023-02079-x>
- Chowdhury, Neil/Johnson, Daniel/Huang, Vincent/Steinhardt, Jacob/Schwettmann, Sarah (2025): Investigating truthfulness in a pre-release o3 model. Veröffentlicht am 16. April 2025. <https://transluce.org/investigating-o3-truthfulness> (Abfrage: 15.06.2025).
- Coherent Solution (2025): How Much Does It Cost to Develop an AI Solution? Pricing and ROI Explained. <https://www.coherentsolutions.com/insights/ai-development-cost-estimation-pricing-structure-roi#Cost-Estimation> (Abfrage: 15.06.2025).
- Crawford, Kate (2024). World View. In: *Nature* 626, S. 693.
- Delphi Gesellschaft: SuchtGPT: Ein KI-Chatbot für Suchtfragen? Projektvorstellung 01/2025. YouTube-Video, veröffentlicht am 28. Januar 2025. <https://youtu.be/JeVff6nGUpo> (Abfrage: 15.06.2025).
- Deutscher Ethikrat (2023): Mensch und Maschine – Herausforderungen durch Künstliche Intelligenz. Stellungnahme. Berlin: Deutscher Ethikrat.
- Deutschlandfunk (2024): Energiehunger durch KI. Wie das Internet den Ausbau der Atomkraft antreibt. <https://www.deutschlandfunk.de/atomkraft-akw-ki-energie-100.html> (Abfrage: 15.06.2025).
- Eubanks, Virginia (2018): Automating Inequality: How High-Tech Tools Profile, Police, and Punish the Poor. New York: St. Martin's Press.
- Falk, Sophia/van Wynsberghe, Aimee (2024): Challenging AI for Sustainability: what ought it mean? In: *AI and Ethics* 4(4), S. 1345–1355.
- Floridi, Luciano/Cowls, Josh/Beltrametti, Monica/Chatila, Raja/Chazerand, Patrice/Dignum, Virginia/Luetge, Christoph/Madelin, Robert/Pagallo, Ugo/Rossi, Francesca/Schafer, Burkhard/Valcke, Peggy/Vayena, Effy (2018): AI4People – An Ethical Framework for a Good AI Soci-

- ety: Opportunities, Risks, Principles, and Recommendations. In: *Minds and Machines* 28(4), S. 689–707. <https://doi.org/10.1007/s11023-018-9482-5>
- Gamgami, Asmahan (2024): *BI_PoC Diversity Manager*innen in weißen Organisationen: Eine Gratwanderung zwischen Racial Stress & White Fragility. Perspektiven und Handlungsempfehlungen*. Wiesbaden: Springer VS.
- Gao, Yunfan/Xiong, Yun/Gao, Xinyu/Jia, Kangxiang/Pan, Jinliu/Bi, Yuxi/Dai, Yi/Sun, Jiawei/Wang, Meng/Wang, Haofen (2023): *Retrieval-Augmented Generation for Large Language Models: A Survey*. Shanghai Research Institute for Intelligent Autonomous Systems, Tongji University/Shanghai Key Laboratory of Data Science, Fudan University/College of Design and Innovation, Tongji University. arXiv preprint. <https://arxiv.org/abs/2312.10997>
- George, A. Shaji/George, A. Hovan/Martin, A. Gabrio (2023): *The environmental impact of ai: A case study of water consumption by chat gpt*. In: *Partners Universal International Innovation Journal* 1(2), S. 97–104.
- George, A. Shaji/Baskar, T./Pandey, Digvijay (2024): *Establishing Global AI Accountability: Training Data Transparency, Copyright, and Misinformation*. In: *Partners Universal Innovative Research Publication* 2(3), S. 75–91.
- Gillingham, Philip/Schiffhauer, Birte/Seelmeyer, Udo (2020): *Internationale Forschung zum Einsatz digitaler Technik in der Sozialen Arbeit*. In: Kutscher, Nadia/Ley, Thomas/Seelmeyer, Udo/Siller, Friederike/Tillmann, Angela/Zorn, Isabell (Hrsg.): *Handbuch Soziale Arbeit und Digitalisierung*. Weinheim und Basel: Beltz Juventa, S. 639–651.
- Goldberg, Brigitta (2021): *Schweigepflicht und Datenschutz in der Sozialen Arbeit und Beratung*. Bochum: Ev. Hochschule Rheinland-Westfalen Lippe.
- Heuer, Hendrik/Hoch, Hendrik/Breiter, Andreas/Theocharis, Yannis (2021): *Auditing the Biases Enacted by YouTube for Political Topics in Germany*. In: *Proceedings of the 2021 ACM Conference on Fairness, Accountability, and Transparency*, S. 479–490. <https://doi.org/10.1145/3473856.3473864>
- Koenecke, Allison/Choi, Anna Seo Gyeong/Mei, Katelyn X./Schellmann, Hilke/Sloane, Mona (2024): *Careless whisper: Speech-to-text hallucination harms*. In: *Proceedings of the 2024 ACM Conference on Fairness, Accountability, and Transparency*, S. 1672–1681.
- Klinova, Katya/Korinek, Anton (2021): *Ai and shared prosperity*. In: *Proceedings of the 2021 AAAI/ACM Conference on AI, Ethics, and Society*, S. 645–651.
- Kreidenweis, Helmut/Diepholz, Maria (2024): *Studie. Künstliche Intelligenz in der Sozialwirtschaft*. Forschungsbericht. Hannover: Althammer & Kill.
- Krings, Markus/Heister, Werner (2023): *Der Nutzen von KI in der Sozialwirtschaft*. In: *Sozialwirtschaft aktuell* 22, S. 1–4.
- Leicht, Maximilian/Karst, Julia/Zimmer, Jasmin (2020): *Diskriminierung und Frauenfeindlichkeit – KI als Spiegel unserer Gesellschaft*. In: Schweighofer, Erich/Kummer, Franz/Saarenpää, Ahti/Hötzendorfer, Walter (Hrsg.): *Internet of Things – Tagungsband des 23. Internationalen Rechtsinformatik Symposions IRIS 2020*, S. 73–80. https://www.uni-saarland.de/fileadmin/upload/lehrstuhl/sorge/Paper-Downloads/IRIS2020_Diskriminierung-und-Frauenfeindlichkeit.pdf (Abfrage: 15.06.2025).
- Lindsey, Jack/Gurnee, Wes/Ameisen, Emmanuel/Chen, Brian/Pearce, Adam/Turner, Nicholas L./Citro, Craig/Abrahams, David/Carter, Shan/Hosmer, Basil/Marcus, Jonathan/Sklar, Michael/Templeton, Adly/Bricken, Trenton/McDougall, Callum/Cunningham, Hoagy/Henighan, Thomas/Jermyn, Adam/Jones, Andy/Persic, Andrew/Qi, Zhenyi/Thompson, T. Ben/Zimmerman, Sam/Rivoire, Kelley/Conerly, Thomas/Olah, Chris/Batson, Joshua (2025): *On the Biology of a Large Language Model*. Veröffentlicht am 27. März 2025. <https://transformer-circuits.pub/2025/attribution-graphs/biology.html> (Abfrage: 15.06.2025).
- Löhe, Julian/Aldendorff, Philipp (2022): *Grundlagen zum Sozialmanagement. Zentrale Begriffe und Handlungsansätze*. Göttingen: Vandenhoeck & Ruprecht.

- Noble, Safiya Umoja (2018): *Algorithms of Oppression: How Search Engines Reinforce Racism*. New York: NYU Press.
- Orwat, Carsten (2020): Diskriminierungsrisiken durch die Verwendung von Algorithmen. Expertise im Auftrag der Antidiskriminierungsstelle des Bundes. Berlin: Antidiskriminierungsstelle des Bundes. Online: https://www.antidiskriminierungsstelle.de/SharedDocs/downloads/DE/publikationen/Expertisen/studie_diskriminierungsrisiken_durch_verwendung_von_algorithmen.pdf (Abfrage: 15.06.2025).
- Pottharst, Bill/Neumann, Alexander/Ostrau, Christoph/Seelmeyer, Udo (2024): Bewältigung des Fachkräftemangels durch technologische Innovation? Effekte von Technisierung und Digitalisierung. In: *Sozial Extra* 48(3), S. 162–167. <https://doi.org/10.1007/s12054-024-00694-9>
- Reinmann, Gabi (2023): Deskillung durch Künstliche Intelligenz? Potenzielle Kompetenzverluste als Herausforderung für die Hochschuldidaktik. Diskussionspapier Nr. 25 Berlin: Hochschulforum Digitalisierung.
- Rösler, Ulrike/Schmidt, Kristina/Merda, Meiko/Melzer, Marlen (2018): Digitalisierung in der Pflege. Wie intelligente Technologien die Arbeit professionell Pflegenden verändern. Berlin: Initiative Neue Qualität der Arbeit (INQA).
- Ruha, Benjamin (2019): *Race After Technology. Abolitionist Tools for the New Jim Code*. Cambridge: Polity Press.
- Vered, Mor/Livni, Tali/Howe, Piers D. L./Miller, Tim/Sonenberg, Liz (2023): The effects of explanations on automation bias. In: *Artificial Intelligence* 322, S. 103952. <https://doi.org/10.1016/j.artint.2023.103952>
- Ali, Sajid/Abuhmed, Tamer/El-Sappagh, Shaker/Muhammad, Khan/Alonso-Moral, Jose M./Confalonieri, Roberto/Guidotti, Riccardo/Del Ser, Javier/Díaz-Rodríguez, Natalia/Herrera, Francisco (2023): Explainable Artificial Intelligence (XAI): What we know and what is left to attain Trustworthy Artificial Intelligence. In: *Information Fusion* 99, S. 101805. <https://doi.org/10.1016/j.inffus.2023.101805>
- Wu, Kevin/Wu, Eric/Zou, James Y. (2024): Clashes: Quantifying the tug-of-war between an LLM's internal prior and external evidence. In: *Advances in Neural Information Processing Systems* 37, S. 33402–33422.

Verzeichnis der Autor:innen

Prof. Dr. phil. Angelika Beranek ist Diplom-Sozialpädagogin und Professorin für Grundlagen der Sozialen Arbeit mit dem Schwerpunkt Medienbildung an der Fakultät für angewandte Sozialwissenschaften der Hochschule München. Mail: beranek@hm.edu

Prof. Dr. phil. Edeltraud Botzum ist Diplom-Sozialarbeiterin (FH) und Forschungsprofessorin an der TH Rosenheim. Sie leitet das interdisziplinäre bidt Digitalisierungskolleg digi.prosa (Digitale Projekte in der Sozialen Arbeit), das StIL Lehrforschungsprojekt digi.peer (Digitales Peer-Mentoring) sowie verschiedene Kurse und Angebote an der vhb (Virtuelle Hochschule Bayern). Mail: edeltraud.botzum@th-rosenheim.de

Dr. Sebastian Dötterl ist Jurist und Richter am Oberlandesgericht. Mail: sebastian@doetterl.com

Madeleine Dörr (M. A.) ist Sozialarbeiterin und wissenschaftliche Mitarbeiterin im bidt-Digitalisierungskolleg digi.prosa – Digitale Projekt in der Sozialen Arbeit sowie im StIL-Projekt digi.peer – Digitales Peer-Mentoring im Kontext einer Literatur- und Forschungswerkstatt an der Technischen Hochschule Rosenheim. Mail: madeleine.doerr@th-rosenheim.de

Prof. Dr. phil. Jörn Dummann ist seit 2022 Professor für Sport in der Sozialen Arbeit (2010–2022 Theorien der Sozialen Arbeit) an der FH Münster und leitet dort seit 2016 einen onlinebasierten, berufsbegleitenden Masterstudiengang Soziale Arbeit und Forschung sowie seit 2010 einen onlinebasierten, berufsbegleitenden Bachelorstudiengang Soziale Arbeit. Mail: dummann@fh-muenster.de

Monika Feist-Ortmanns ist Erziehungswissenschaftlerin B. A. sowie Sozialmanagerin M. A. und geschäftsführende Direktorin am Institut für Kinder- und Jugendhilfe (IKJ). Sie lehrt an der Hochschule Niederrhein, der HAW Landshut und der IU Internationale Hochschule.

Dr. Andrea Gergen ist Wissenschaftliche Mitarbeiterin im Lehrforschungsprojekt digi.peer (Digitales Peer-Mentoring) an der TH Rosenheim und Vertretungsprofessorin für Soziale Arbeit an der HSD Hochschule Döpfer. Mail: andrea.gergen@th-rosenheim.de

Prof. Dr. Wolfgang M. Heffels ist Professor für Ethik und Berufspädagogik und Emeritus an der katho (Katholische Hochschule Nordrhein-Westfalen). Seine beruflichen Schwerpunkte in Lehre und Forschung waren Didaktik und Methodik der Berufsbildung Pflege und Gesundheit sowie die ethische Bildung (Individual-, Sozial- und Gesellschaftsethik). Mail: wm.heffels@katho-nrw.de

Felix Holz (M. Sc.) ist seit 2020 wissenschaftlicher Mitarbeiter des Lehrstuhls für Wirtschaftsinformatik an der Universität Rostock und befasst sich mit der digitalen Assistenzsystemgestaltung für personenbezogene Dienstleister. Mail: felix.holz@uni-rostock.de

M. Sc. Svitlana Hrytsai ist wissenschaftliche Mitarbeiterin im Fachteam Lehr- und Lernformen am Departement Soziale Arbeit der Fachhochschule OST (Standort St.Gallen). Mail: svitlana.hrytsai@ost.ch

Prof. Dr. phil. Robert Lehmann ist Diplom Sozialpädagoge (FH) und Professor für Theorien und Handlungslehre der Sozialen Arbeit an der Fakultät für Angewandte Sozialwissenschaften an der Technischen Hochschule Nürnberg Georg Simon Ohm. Mail: robert.lehmann@th-nuernberg.de

Prof. Dr. phil. Gesa A. Linnemann ist Diplom-Psychologin und Professorin für Digitalisierung sozialer Lebenswelten und Profession am FB Sozialwesen an der Katholischen Hochschule Nordrhein-Westfalen (Standort Münster). Mail: g.linnemann@katho-nrw.de

Prof. Dr. phil. Julian Löhe ist Coach/Supervisor (DGSv), Sozialarbeiter und Professor für Organisation und Management an der FH Münster und leitet dort seit 2019 den Masterstudiengang Sozialmanagement. Mail: loehe@fh-muenster.de

Prof. Dr. rer. nat. Michael Macsenaere ist Diplom-Psychologe und wissenschaftlicher Direktor am Institut für Kinder- und Jugendhilfe (IKJ). Er lehrt an der Johannes Gutenberg-Universität in Mainz.

Florian Müller (M. A.) ist Wissenschaftlicher Mitarbeiter in den Lehrforschungsprojekten „digi.peer“ (Digitales Peer-Mentoring) und „digi.prosa“ (Digitale Projekte in der Sozialen Arbeit) an der Technischen Hochschule Rosenheim. Mail: florian.mueller@th-rosenheim.de

Prof. Jan Pelzl ist Professor für IT-Sicherheit. Er ist Autor eines der führenden Lehrbücher für Kryptografie und darüber hinaus freiberuflich im Bereich Cybersecurity als Berater tätig.

Prof. Dr. Christina Plafky ist Professorin für KI und VR im Departement Soziale Arbeit im Institut „Fachdidaktik, Professionsentwicklung und Digitalisierung“ an der BFH Berner Fachhochschule (Schweiz). Mail: christina.plafky@bfh.ch

Benjamin Plattner ist Software-Engineer und wissenschaftlicher Mitarbeiter im Bereich Künstliche Intelligenz am Institut für Software (IFS) der Fachhochschule OST (Standort Rapperswil). Mail: benjamin.plattner@ost.ch

Prof. Dr. Mitra Purandare ist Professorin für Informatik und Leiterin der „AI Applications and Deployment Lab“ am Institute für Software (IFS) der Fachhochschule OST (Standort Rapperswil). Mail: mitra.purandare@ost.ch

Dr. rer. pol. Beate Rottkemper ist Diplom-Wirtschaftsinformatikerin und Projektleiterin am Universitätsklinikum Münster. Darüber hinaus ist sie Lehrbeauftragte an der FH Münster. Mail: beate.rottkemper@mailbox.org

Angelina Clara Schmidt (M. Sc.) ist seit 2022 wissenschaftliche Mitarbeiterin am Lehrstuhl für Wirtschaftsinformatik der Universität Rostock mit dem Schwerpunkt Assistenzsysteme für Produktivität und Gesundheit in der modernen Arbeitswelt. Mail: angelina.schmidt@uni-rostock.de

Prof. Dr. Olivier Steiner ist Soziologe und Professor für Lebenslagen und Lebensweisen von Adressat:innen der Kinder- und Jugendhilfe an der Hochschule für Soziale Arbeit, Fachhochschule Nordwestschweiz. Mail: olivier.steiner@fhnw.ch

Prof. Dr. phil. Eik-Henning Tappe ist Erziehungswissenschaftler und Professor für Digitalisierung und Medienpädagogik in der Sozialen Arbeit am FB Sozialwesen der FH Münster. Zudem ist er Co-Vorsitzender der Gesellschaft für Medienpädagogik und Kommunikationskultur (GMK). Mail: etappe@fh-muenster.de