

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

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

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 Beitragsautor:innen
Publikation: Beltz Juventa in der Beltz Verlagsgruppe GmbH & Co. KG

Werderstraße 10, 69469 Weinheim

service@belitz.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.belitz.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 Schlamann</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 Plafsky, 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 Handbuchs 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/Hershcovich, 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): Alptraum Computer. <https://www.zeit.de/1972/03/alptraum-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 einstehen 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 Maschinellem 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 Chancengleichheit 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 Schłomann 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 Wenn gleich 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 Technologien.

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 eingordnet 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 *Wolfgang 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 Gerger* und *Florian Müller* das Thema „Künstliche Intelligenz in der Lehre der Sozialen

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. Linemann/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) wieder gefunden 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 semistrukturiertener 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 Klimaanlagen 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

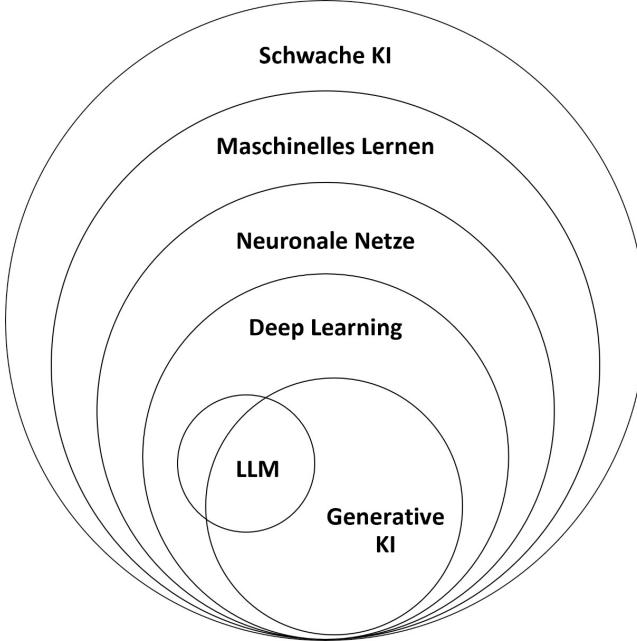
2024). Algorithmen, die zu den Expertensystemen gehören, sind z. B. Heuristiken, Entscheidungsbäume oder wissensbasierte Netzwerke.

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

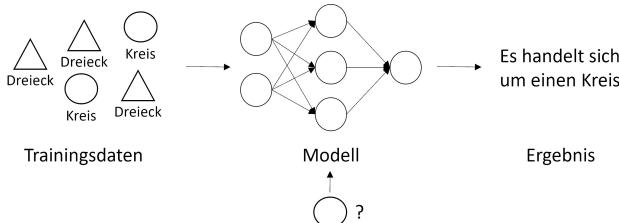
Im Maschinellen Lernen wird zwischen überwachten und unüberwachten Verfahren unterschieden. Darüber hinaus werden im weiteren Verlauf des Beitrags Begriffe wie Neuronale Netze, Deep Learning, generative KI und Large Language Models (LLM) eingeordnet und erläutert (siehe Abbildung 2).

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



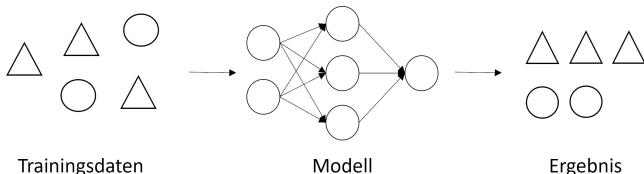
Quelle: Eigene Darstellung nach Choo et al. 2021

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 (Abbildung 4). 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).

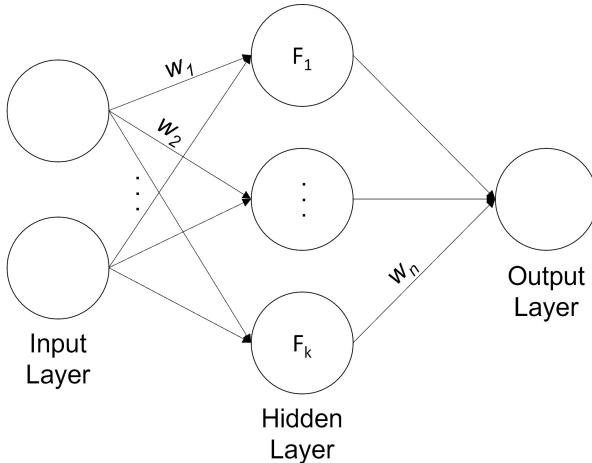
Beispiele dafür gibt es viele, eines aus der Sozialen Arbeit stammt aus den Niederlanden. Hier wurde ein Algorithmus eingesetzt, um Kindergeldansprüche zu prüfen. Aufgrund der Ergebnisse des Algorithmus wurden Familien mit Migrationsgeschichte systematisch die Kindergeldansprüche gestrichen. Infolgedessen 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 Neuronalen 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. Raparthi 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 zwangsläufig gelabelte) Datenmengen benötigt (vgl. Rahali/Akhloifi 2023).