Sabrina Schork *Hrsg.*

Vertrauen in Künstliche Intelligenz

Eine multi-perspektivische Betrachtung

Springer Vieweg

Vertrauen in Künstliche Intelligenz

Sabrina Schork
(Hrsg.)

Vertrauen in Künstliche Intelligenz

Eine multi-perspektivische Betrachtung

Springer Vieweg

Hrsg.
Prof. Dr. Sabrina Schork
esn. Institut und TH Aschaffenburg
München und Aschaffenburg, Bayern, Deutschland

Mit Beiträgen von
Prof. Dr. Rainer Hofmann
Aschaffenburg, Deutschland

Prof. Dr. Ines Langemeyer
Karlsruhe, Deutschland

Daniel Glinz
Zürich, Schweiz

Prof. Dr. Derya Gür-Seker
Bonn-Rhein-Sieg, Deutschland

Dr. Johannes Schrumpf
Heilbronn, Deutschland

Janine Strotherm
Bielefeld, Deutschland

Prof. Dr. Anders Madsen
Aalborg, Dänemark

Prof. Dr. Galia Weidl
Aschaffenburg, Deutschland

Alissa Müller
Bielefeld, Deutschland

Dr. Katharina Weitz
Berlin, Deutschland

Eva Bartaguiz
Kaiserslautern, Deutschland

Prof. Dr. Wolfgang Kemmler
Erlangen-Nürnberg, Deutschland

Prof. Dr. Peter Rötzel
Aschaffenburg, Deutschland

Dr. Franz-Josef Schmitt
Halle, Deutschland

Janne Mesenhöller
Potsdam, Deutschland

Dr. Katharina Weitz
Augsburg, Deutschland

Prof. Dr. Benjamin Paaßen
Berlin, Deutschland

Prof. Dr. Katrin Böhme
Potsdam, Deutschland

Prof. Dr. Sandra Leaton Gray
London, England

Prof. Dr. Barbara Hammer
Bielefeld, Deutschland

Stefan Slembrouck
Arnsberg, Deutschland

Dr. Carlo Dindorf
Kaiserslautern, Deutschland

Prof. Dr. Michael Fröhlich
Kaiserslautern-Landau, Deutschland

Prof. Dr. Andrea Pieter
Saarbrücken, Deutschland

ISBN 978-3-658-43815-9 ISBN 978-3-658-43816-6 (eBook)
https://doi.org/10.1007/978-3-658-43816-6

Die Deutsche Nationalbibliothek verzeichnet diese Publikation in der Deutschen Nationalbibliografie; detaillierte bibliografische Daten sind im Internet über https://portal.dnb.de abrufbar.

© Der/die Herausgeber bzw. der/die Autor(en), exklusiv lizenziert an Springer Fachmedien Wiesbaden GmbH, ein Teil von Springer Nature 2024

Das Werk einschließlich aller seiner Teile ist urheberrechtlich geschützt. Jede Verwertung, die nicht ausdrücklich vom Urheberrechtsgesetz zugelassen ist, bedarf der vorherigen Zustimmung des Verlags. Das gilt insbesondere für Vervielfältigungen, Bearbeitungen, Übersetzungen, Mikroverfilmungen und die Einspeicherung und Verarbeitung in elektronischen Systemen.
Die Wiedergabe von allgemein beschreibenden Bezeichnungen, Marken, Unternehmensnamen etc. in diesem Werk bedeutet nicht, dass diese frei durch jede Person benutzt werden dürfen. Die Berechtigung zur Benutzung unterliegt, auch ohne gesonderten Hinweis hierzu, den Regeln des Markenrechts. Die Rechte des/der jeweiligen Zeicheninhaber*in sind zu beachten.
Der Verlag, die Autor*innen und die Herausgeber*innen gehen davon aus, dass die Angaben und Informationen in diesem Werk zum Zeitpunkt der Veröffentlichung vollständig und korrekt sind. Weder der Verlag noch die Autor*innen oder die Herausgeber*innen übernehmen, ausdrücklich oder implizit, Gewähr für den Inhalt des Werkes, etwaige Fehler oder Äußerungen. Der Verlag bleibt im Hinblick auf geografische Zuordnungen und Gebietsbezeichnungen in veröffentlichten Karten und Institutionsadressen neutral.

Planung/Lektorat: David Imgrund
Springer Vieweg ist ein Imprint der eingetragenen Gesellschaft Springer Fachmedien Wiesbaden GmbH und ist ein Teil von Springer Nature.
Die Anschrift der Gesellschaft ist: Abraham-Lincoln-Str. 46, 65189 Wiesbaden, Germany

Das Papier dieses Produkts ist recycelbar.

Vorwort

Die in den letzten Jahrzehnten rasant wachsende Rechenleistung von Großrechnern, aber auch von mobilen Endgeräten wie dem Smartphone hat die Voraussetzungen geschaffen, um durch das Zusammenführen von Algorithmen des Maschinellen Lernens (ML) bzw. der Künstlichen Intelligenz (KI) sowie umfassender digitaler Datenbestände das Potenzial automatisierter Datenanalysen in Theorie und Anwendung zu erforschen und zu nutzen. Die maschinelle Erkennung von Mustern und Korrelationen auch in komplexen und umfangreichen Datensätzen ermöglicht Anwendungen, die die Leistungsfähigkeit bisheriger Werkzeuge erheblich überschreiten. Besonders sichtbar ist dies beispielsweise im Bereich der Objekt- oder Gesichtserkennung, aber auch bei generativen Funktionen wie dem Hervorbringen von automatisierter Kunst.

Die TH Aschaffenburg hat daher in den vergangenen Jahren wiederholt Ringvorlesungen zu unterschiedlichen gesellschaftlichen Bezügen von KI abgehalten, so zu KI in der industriellen Praxis (2020), der Medizin (2021) oder der Kunst (2022).

Besondere gesellschaftliche Relevanz erlangen KI-Systeme, wenn sie aktiv Prozesse steuern, Empfehlungen vorbereiten oder solche gar selbst abgeben, und in Produkte integriert werden, die direkt dem Verbraucher zum Erwerb oder als Service angeboten werden. Je direkter Menschen in ihrer Alltagsumgebung mit Ergebnissen von KI-Verfahren konfrontiert werden, desto stärker stellt sich die Frage von Nutzerakzeptanz und Vertrauen in diese Technologie. Gleichzeitig ist es für einen weiteren Fortschritt von KI-Modellen oft unerlässlich, dass Individuen private Nutzerdaten zur technischen Auswertung zur Verfügung stellen. In einer bürgerlichen Kultur, die durch das Bewusstsein für die eigene Privatsphäre und einen umfassenden Datenschutz geprägt ist, kann nur die aufgeklärte Autonomie des Menschen als Eigentümer seiner persönlichen Daten, z.B. im Gesundheitsbereich, Grundlage für die freiwillige und selbstbestimmte Preisgabe von Daten sein. Damit verbunden ist die Notwendigkeit, für eine Akzeptanz von KI-Technologien in der Gesellschaft und bei Verbrauchern zu werben.

Das vorliegende Buch ist ein erster Ansatz, das Themenfeld „Vertrauen in KI“ aus multiplen Perspektiven zu beleuchten. Eine Annäherung findet aus der wirtschafts- und sozialwissenschaftlichen, informationstechnischen sowie interdisziplinären Perspektive mit insgesamt fünfzehn Beiträgen statt. Über die Auseinandersetzungen sollen Lernende,

Lehrende, Forschende und Vertreterinnen sowie Vertreter aus Politik und Wirtschaft in die Lage versetzt werden, sich eine auf Fakten gestützte Meinung zu bilden und, darauf aufbauend, fundierte persönliche Entscheidungen im Umgang mit KI zu treffen.

Unser Dank für dieses Buchprojekt gilt insbesondere der Herausgeberin Prof. Dr. Sabrina Schork sowie den Autorinnen und Autoren und dem Springer Team.

Aschaffenburg
im Herbst 2023

Prof. Dr.-Ing. Konrad Doll
Prof. Dr. Michael Möckel
Sprecher des Kompetenzzentrums
Künstliche Intelligenz an der
TH Aschaffenburg

Inhaltsverzeichnis

Teil I
Buchvorspann

Einleitung

1

Sabrina Schork und Peter Gordon Rötzel

1.1 Historische Entwicklung

Nachdem Alan Turing und andere in den 1930er Jahren die formalen Grundlagen für digitale Maschinen gelegt hatten, die Daten mit Hilfe von Algorithmen verarbeiten, stellte sich Alan Turing 1950 in einer Veröffentlichung die Frage, ob Maschinen denken können und wie man - mit dem bekannten ‚Turing-Test' - herausfinden kann, inwieweit sie das tun [1]. Der Begriff ‚KI' wurde 1955 zum ersten Mal von John McCarthy am MIT benutzt [2]. Im Sommer 1956 trafen sich Wissenschaftlerinnen und Wissenschaftler zu einer Konferenz am Dartmouth College im US-Bundesstaat New Hampshire, um dort über den Begriff KI zu diskutieren [1].

Künstliche Intelligenz ist ein Teilgebiet der Informatik, das sich mit der Entwicklung von Maschinen befasst, die Probleme lösen, für die normalerweise menschliche Intelligenz erforderlich ist [3]. Je nach Art des Trainings und der Daten, auf die KI-Systeme zugreifen, können sie ein breites Spektrum von Problemen lösen. Aufgrund dieser Vielfalt sind Versuche, KI-basierte Systeme eindeutig zu definieren, oft uneinheitlich und mitunter vage. Eine allgemein anerkannte wissenschaftliche Definition von KI existiert bislang nicht [4].

Die Europäische Kommission definiert KI wie folgt: „AI is the ability of a machine to display human-like capabilities such as reasoning, learning, planning and creativity. AI

S. Schork (✉)
Wirtschaft und Recht, Technische Hochschule Aschaffenburg, Aschaffenburg, Deutschland
E-Mail: sabrina.schork@th-ab.de

P. G. Rötzel
Ingenieurwissenschaften, Technische Hochschule Aschaffenburg, Aschaffenburg, Deutschland
E-Mail: peter.roetzel@th-ab.de

© Der/die Autor(en), exklusiv lizenziert an Springer Fachmedien Wiesbaden GmbH, ein Teil von Springer Nature 2024

S. Schork (Hrsg.), *Vertrauen in Künstliche Intelligenz,*
https://doi.org/10.1007/978-3-658-43816-6_1

enables technical systems to perceive their environment, deal with what they perceive, solve problems and act to achieve a specific goal. The computer receives data - already prepared or gathered through its own sensors such as a camera - processes it and responds. AI systems are capable of adapting their behaviour to a certain degree by analysing the effects of previous actions and working autonomously.“ [5]

Eine erste deutschsprachige Publikation mit den Titelstichworten ‚Vertrauen in Künstliche Intelligenz‘ findet sich über OPACplus im Jahr 2019. Es handelt sich dabei um einen Aufsatz von Cremers et al. [6]. Die Publikation beschreibt die interdisziplinäre Entwicklung einer KI-Zertifizierung. Ein erster englischsprachiger Beitrag mit den Titelstichworten ‚Trust in Artificial Intelligence‘ findet sich über OPACplus bereits im Jahr 1984. Veröffentlicht wurde dieser durch Jon Doyle im ‚The AI Magazine‘ [7]. Der Autor fordert, dass IT-Expertensysteme auf Eis gelegt werden sollten, bis ihr Verhalten formal analysiert werden könne.

Insgesamt wurden am 18.09.2023 in OPACplus 38 deutsch- und 3.392 englischsprachige Publikationen zu den oben genannten Titelstichwörtern gefunden. Die Verteilung nach Jahren ist in Abb. 1.1 und 1.2 dargestellt. Zu erkennen ist ein Anstieg von englischsprachigen Publikationen im Jahr 1998 und ein Anstieg der deutschsprachigen Publikationen im Jahr 2020. Im Jahr 2020 wurden 18 deutschsprachige und 225 englischsprachige Publikationen zum gesuchten Thema in OPACplus veröffentlicht. Der große zahlenmäßige Unterschied könnte darauf zurückzuführen sein, dass die Wissenschaftssprache ‚Englisch‘ ist. Im Jahr 2006 ist ein extremer Anstieg englischsprachiger Fachartikel zum Thema ‚Trust in Artificial Intelligence‘ zu verzeichnen. Gerade in diesem Jahr erschien ein sehr wichtiger Beitrag zum Deep Learning von Geoffrey E. Hinton, Simon Osindero und Yee-Whye Teh mit dem Titel ‚A Fast Learning Algorithm for Deep Belief Nets‘. In diesem Artikel wird ein neuer Typ von neuronalen Netzen vorgestellt, der es ermöglicht, mehrschichtige neuronale Netze effizient zu trainieren.

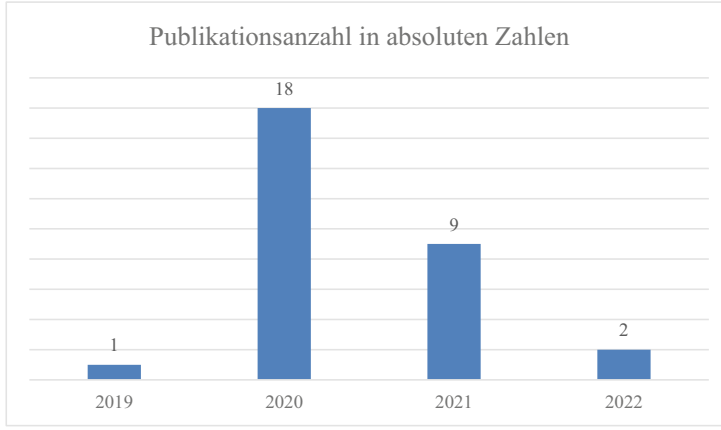

Abb. 1.1 Ergebnis der Titelstichwortsuche ‚Vertrauen in Künstliche Intelligenz' auf OPACplus. (Eigene Darstellung)

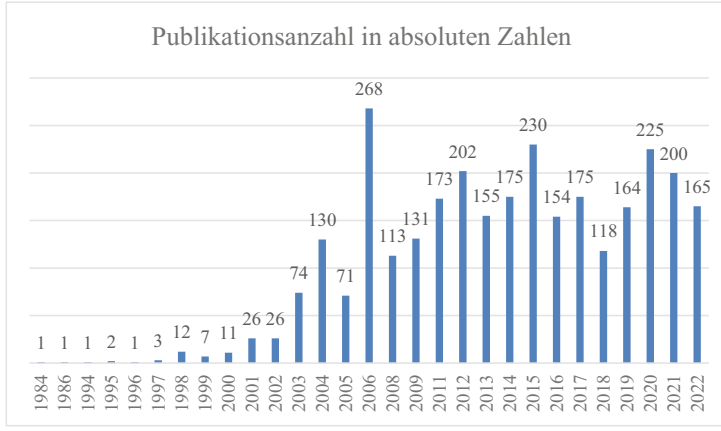

Abb. 1.2 Ergebnis der Titelstichwortsuche ‚Trust in Artificial Intelligence' auf OPACplus. (Eigene Darstellung)

1.2 Gesellschaftliche, wissenschaftliche und wirtschaftliche Bedeutung von Vertrauen in KI

Vertrauen in die KI ist in der **Wissenschaft** wichtig, damit KI-Systeme entwickelt werden, die den Bedürfnissen der Gesellschaft entsprechen. Vertrauenswürdige KI kann dazu beitragen, neue wissenschaftliche Erkenntnisse zu gewinnen oder neue Technologien zu entwickeln, die grundlegende und notwendige Veränderungen mit sich bringen.

Wenn Wissenschaftler KI-Systemen nicht vertrauen, werden sie es eher vermeiden, diese Systeme zu benutzen. Dies kann zu Verzögerungen führen und die Qualität der Forschung beeinträchtigen. Durch den Aufbau von Vertrauen in die KI in der **Wirtschaft** können Unternehmen die Vorteile dieser Technologie nutzen, um neue Produkte und Dienstleistungen zu entwickeln, ihre Geschäftsprozesse zu verbessern und ihre Wettbewerbsfähigkeit zu steigern. Durch den Einsatz von KI können insbesondere Kosten und Zeit eingespart werden. Durch den Aufbau von Vertrauen in KI in der **Politik** können Bürgerinnen und Bürger die Vorteile dieser Technologie nutzen, um bessere politische Entscheidungen zu treffen und die Qualität der Demokratie zu verbessern [8].

1.2.1 Gesellschaftliche Bedeutung von Vertrauen in KI

Datenbeschaffung und -verarbeitung waren in der Vergangenheit eine große Herausforderung, insbesondere die Beschaffung und systematische Aufarbeitung von Daten und deren Bereitstellung für Entscheidungsprozesse stellten Unternehmen vor erhebliche Herausforderungen [9]. Durch zunehmend automatisierte Datenmanagementsysteme und integrierte Reportingtools haben sich die Herausforderungen im Bereich der Datenbeschaffung verlagert. Zunehmend geht es nicht mehr um die reine Beschaffung von relevanten Daten, sondern um die Bewältigung der großen, verfügbaren Datenmengen [10]. Wachsende Datenmengen und zunehmende Fähigkeiten von KI-Algorithmen führen dazu, dass immer mehr Entscheidungen an automatisierte Prozesse delegiert werden [11]. Die Wichtigkeit der Anwendungsbereiche von KI nimmt dabei zu, bspw. zur Verkehrssteuerung, für medizinische Beurteilungen, zur Betrugsprävention, bei der Kreditvergabe etc. [12]. Dies unterstreicht die Relevanz von Diskussionen zur Rechenschaftspflicht im KI-Kontext. Dabei wird zunehmend deutlich, dass die Nachvollziehbarkeit von KI-generierten Entscheidungsvorschlägen bzw. Handlungsempfehlungen einen wichtigen Faktor in der gesellschaftlichen Debatte einnimmt [13]. Sind die dem Algorithmus zugrunde liegenden Daten fehlerhaft, kann dies zu falschen Ergebnissen, Verzerrungen und Unfairness führen, siehe z. B. die CalGang-Datenbank, ein Datenpool zur Vorhersage von Gewaltverbrechen, der sich als verzerrt und fehlerhaft herausstellte [14]. Weitere Beispiele sind auch die Diskriminierung von Menschen mit dunkler Hautfarbe bei der Festlegung einer medizinischen Behandlungsreihenfolge [15] oder ein fehlerhafter Amazon-Algorithmus, der Frauen im Recruiting-Prozess benachteiligte [16]. Welche Faktoren dazu führen, dass KI Entscheidungen trifft, die aus menschlicher Sicht ‚falsch' sind, ist (noch nicht eindeutig geklärt. Nach Araujo et al. [11] bestehen in der Gesellschaft große Bedenken aufgrund von Unkenntnis über die Risiken und möglichen Chancen von KI. Darüber hinaus gibt es keinen gesellschaftlichen Konsens über die Fairness und den Nutzen von KI-generierten Entscheidungen in der Gesellschaft.

Die Entwicklung der KI hat sich in den letzten Jahren beschleunigt und beeinflusst zunehmend unser tägliches Leben und Arbeiten [14, 17]. Der gesellschaftliche Diskurs scheint mit der KI-Entwicklung nicht Schritt zu halten. Die Akzeptanz von KI ist jedoch eine Voraussetzung für ihren Nutzen. Dennoch scheint KI für die breite Gesellschaft derzeit noch eine Art unbekanntes, unerforschtes Terrain zu sein. Umso wichtiger ist es, Vertrauen in KI durch Transparenz in der KI-Entscheidungsfindung sowie durch vereinbarte KI-Regeln und KI-Verantwortlichkeiten zu schaffen.

Um dieser Entwicklung bzw. der gesellschaftlichen Skepsis und Intransparenz zu begegnen, wurden bereits erste Schritte unternommen. So wurde z. B. der *EU AI Act* verabschiedet [17]. Dieser umreist einen rechtlichen Rahmen, um die KI zu fördern und gleichzeitig die Risiken von KI zu minimieren. Der Einsatz von KI soll sich an vereinbarten Regeln orientieren. Auch in der Datenschutzgrunverordnung wurden bereits erste Schritte zur rechtlichen Auseinandersetzung mit KI unternommen [18]. Auf internationaler Ebene befasst sich z. B. die ISO/IEC TR 24.028:2020 mit der KI und deren Haftung. Auch in der Wissenschaft werden die Rolle der KI und ihre Auswirkungen auf die Gesellschaft diskutiert. Einen Überblick über die aktuelle Literatur und den wissenschaftlichen Diskurs bieten Kempton et al. [12].

1.2.2 Wissenschaftliche Bedeutung von Vertrauen in KI

Der wissenschaftliche Diskurs beschäftigt sich neben den zahlreichen Anwendungsmöglichkeiten für die KI auch mit der Etablierung eines Rahmenwerkes, das den potenziell sicheren Einsatz der KI ermöglicht. Eine Übersicht bieten unterschiedliche Fachbeiträge [14, 23]. Darin werden Akzeptanz, Transparenz und Verantwortlichkeit hervorgehoben, z. B. von Kaur [13, S. 1] „Therefore, it has become essential to make these systems safe, reliable, and trustworthy.“ Neben der Nutzerakzeptanz und dem Nutzervertrauen wird die Notwendigkeit der Regulierung von KI gefordert [26–28]. Dabei geht es um die Diskussion, wie es gelingen kann menschliche Vorbehalte gegenüber KI-generierten Entscheidungsvorlagen und letztlich auch Entscheidungen abzubauen und Vertrauen in KI zu schaffen. Die automatisierte Entscheidungsfindung und damit verbundene Blackbox wird von Wachter et al. [19] thematisiert. Sehr relevante Publikationen zum Thema ‚Trust in AI‘ finden sich in den Referenzen [19–25 und 29]. Nicodeme [24] beschreibt verschiedene Faktoren zur Steigerung des Vertrauens, Theodorou und Dignum [26] fordern Gesetze zur Steigerung der Akzeptanz und Kocielnik et al. [27] untersuchen, wie überzogene Erwartungen an KI-Systeme die Wahrnehmung und Akzeptanz solcher Systeme negativ beeinflussen. Duan et al. [28] und Smeets et al. [20] beschäftigen sich mit der Kooperation zwischen Mensch und KI und dem dabei entstehenden Vertrauen. Lockey et al. [29] identifizieren fünf zentrale KI-Vertrauensherausforderungen: 1. Transparenz und Erklärbarkeit, 2. Genauigkeit und Zuverlässigkeit, 3. Automatisierung, 4. Anthropomorphismus (menschliche Eigenschaft an nicht menschlichen Wesen) und 5. Massendatenextraktion.

1.2.3 Wirtschaftliche Bedeutung von Vertrauen in KI

Der Entscheidungsfindungsprozess in Organisationen beruht auf der Systematisierung, Operationalisierung und Analyse von immer größer werdenden Datenmengen. Die Bewältigung dieser Datenmengen wird zunehmend automatisiert und mithilfe von KI vorgenommen. Das Vertrauen in KI-generierte Entscheidungen und Entscheidungsvorschläge spielt hier eine zentrale Rolle [21, 22, 25]. Darüber hinaus wird auch der Bereich ‚Verantwortlichkeit der KI' diskutiert. Dabei geht es um die Zuordnung von sowohl negativen als auch positiven Konsequenzen von Entscheidungen zu Entscheidungsträgern. Ein weiterer Aspekt im Kontext des Vertrauens in KI ist das Phänomen der Algorithmus-Abneigung. Dieses Phänomen beschreibt die fehlende Akzeptanz gegenüber KI-generierten Entscheidungsgrundlagen. Menschen verlassen sich gerne auf von Menschen gemachte Vorhersagen (obgleich diese oft verzerrt sind) und weniger auf KI-generierte Vorhersagen (obgleich diese eher objektiv sind). In der wissenschaftlichen Forschung existieren bereits zahlreiche Ansätze, um das Vertrauen bzw. die Akzeptanz von KI bei Entscheidungsträgern zu steigern bzw. KI mit innerorganisationalen Prozessen zu verzahnen und die Effektivität der KI sicherzustellen [23]. Glikson und Wooley arbeiten die unterschiedlichen Formen des Vertrauens in eine KI heraus (kognitiv versus emotional) [25].

1.3 Autorenverzeichnis mit Beitragszusammenfassung

Im folgenden Unterabschnitt werden die Autorinnen und Autoren des Buchs vorgestellt und deren Fachbeiträge in aggregierter Form dargestellt.

Kapitel 2: Wirtschaftswissenschaftliche Perspektive

Prof. Dr. Peter Gordon Rötzel ist Professor an der Technischen Hochschule Aschaffenburg und habilitierte am Lehrstuhl für Controlling an der Universität Stuttgart. In seinem Beitrag „Künstliche Intelligenz (KI) – unser bester Freund?" setzt er sich mit den Dynamiken, Herausforderungen und den Chancen der Mensch-KI-Interaktion auseinander. Im Mittelpunkt steht dabei die Rolle von Vertrauen in der Interaktion. Dabei werden kognitive, emotionale und soziale Faktoren berücksichtigt.

Prof. Dr. Georg Rainer Hofmann ist Informatiker und Ökonom. Er arbeitet als Professor und Direktor am Information Management Instituts IMI der Technischen Hochschule in Aschaffenburg. Für die Fragestellung „Kann Vertrauen eine Beziehung zwischen Menschen und Maschinen sein?" entwickelt er einen Ansatz des Institutionenvertrauens, da er die Möglichkeit eines persönlichen Vertrauens zwischen Mensch und Maschine verneint. Für ein Institutionenvertrauen in KI identifiziert sein Beitrag Komponenten des klassischen Markenvertrauens und überträgt sie auf KI-Anwendungen.

Daniel Glinz ist Berater und Unternehmer im Bereich der datengetriebenen digitalen Transformation. Seine Expertise erstreckt sich über die Bereiche Wirtschaft, Design und Technologie, wodurch er in der Lage ist, wegweisende, datenzentrierte Lösungen zu konzipieren, die das Unternehmenswachstum durch die verantwortungsvolle Integration von Technologie vorantreiben. Bei der Entwicklung von Lösungen legt Daniel Glinz einen besonderen Schwerpunkt auf die sorgfältige Gestaltung der Schnittstelle zwischen Mensch und Maschine. Großen Wert legt er dabei auf einen verantwortungsvollen und nachhaltigen Umgang mit Daten. In seinem Beitrag erörtert er aus systemtheoretischer Sicht, wie Nutzerinnen und Nutzer in KI-Systeme Vertrauen fassen können und dadurch die notwendigen Daten zur Verfügung stellen, die über den Erfolg oder Misserfolg von KI entscheiden. Vorgeschlagen werden Designprinzipien, die die Entwicklung von KI-Systemen anleiten sollen.

Kapitel 3: Sozialwissenschaftliche Perspektive

Prof. Dr. Katrin Böhme ist Professorin für Inklusionspädagogik an der Universität Potsdam und **Janne Mesenhöller** ist ihre akademische Mitarbeiterin im gemeinsamen Projekt Künstliche Intelligenz im Schulkontext. Der Beitrag mit dem Titel „Meine Kollegin, die KI –Wie die Nutzung von Künstlicher Intelligenz das schulische Lehren und Lernen verändert“ wird die Frage beantwortet, unter welchen Bedingungen KI-basierte Systeme zukünftig eine Bereicherung für das schulische Lehren und Lernen darstellen können. Eine empirische Untersuchung mit 141 Lehrkräften zeigt, dass diese der Nutzung von KI positiv gegenüberstehen, aber auch viele Bedenken und Ängste haben. Damit KI als Entlastung wahrgenommen wird, schlagen die Autorinnen die Entwicklung von professionalisierungsbezogenen Angeboten vor. Unklar bleibt, ob durch die Einführung von KI-Anwendungen die aktuelle Überlastung und Überforderung im Bildungsbereich bedingt durch eine zunehmende Heterogenität und sinkende Lehrkraftanzahl minimiert werden kann. Auch die schlechte Medien- und Infrastruktursituation wird thematisiert.

Prof. Dr. Ines Langemeyer ist Professorin für Lehr-Lernforschung, Allgemeine Pädagogik und Berufspädagogik am Karlsruher Institut für Technologie. In ihrem Beitrag „Vertrauensvolle KI – eine Diskussion aus psychologischer und pädagogischer Sicht“ wird KI als ein sozialer Akteur diskutiert, mit dem Menschen in eine Beziehung treten. Die Autorin regt eine Reflextion der Gebrauchsweise digitaler Technologien an und fordert das Verantworten-Können von Entscheidungen ein. Aktuell beantwortet KI menschliche Fragen aus großen Datensätzen des Internets heraus, aus denen kommunikative Sätze heraus generiert werden. Auf diese Weise schafft die KI soziale Realitäten. KI kann jedoch keine Verantwortung für sich selbst und andere übernehmen und damit als Instrument der Beeinflussung in Entscheidungssituationen genutzt werden. Im Fokus der Auseinandersetzung stehen die

beruflich-professionellen Beziehungen in Wissensinstitutionen und die durch KI bedingten unreflektierten Praktiken und Strukturen.

Dr. Sandra Leaton Gray ist Professorin und Lehrstuhlinhaberin für Bildungszukunft am University College London Institute of Education. Der Beitrag „Die Ethik der KI in Universitäten: Qualität, Identität und Privatsphäre im Spannungsfeld" befasst sich mit den ethischen Dilemmata, die mit der Einführung von KI an Hochschulen verbunden sind. Analysiert werden die komplexen Beziehungen zwischen den jüngsten technologischen Fortschritten, Datenschutzbedenken und der wissenschaftlichen Praxis. Thematisiert werden soziologische und philosophische Aspekte. Und es werden algorithmische Voreingenommenheit, Datenqualität, Fairness, Rechenschaftspflicht, Datenschutz, Erklärbarkeit und Transparenz als Themen internationaler Datenschutzgesetze erörtert. Erarbeitet werden erkenntnistheoretische Veränderungen der Hochschulbildung durch den Einsatz von KI-Systemen sowie Handlungsempfehlungen für datenschutzfreundliche Rahmenbedingungen.

Kapitel 4: Computerwissenschaftliche Perspektive

Prof. Dr. Anders Madsen ist Professor der Computerwissenschaften an der Aalborg Universität in Dänemark. Er forscht zu verteilten, eingebetteten und intelligenten Systemen. Zusammen mit **Prof. Dr. Galia Weidl** vom Kompetenzzentrum KI der Technischen Hochschule Aschaffenburg befasst er sich mit der Verwendung von Bayes'schen Netzen als Methode zur Implementierung transparenter, erklärbarer und vertrauenswürdiger Künstlicher Intelligenz.

Prof. Dr. Benjamin Paaßen ist Juniorprofessor für Wissensrepräsentation und Maschinelles Lernen an der Universität Bielefeld. Gemeinsam mit **Janine Strotherm** und **Barbara Hammer** von der Technischen Fakultät sowie **Alissa Müller** von der Medizinischen Fakultät untersucht er in dem Beitrag „Fairness in KI-Systemen" verschiedene Fairness-Definitionen und Strategien zur Erkennung von Unfairness anhand praktischer Anwendungsbeispiele.

Dr. Franz Josef Schmitt ist Wissenschaftler der Physik an der Martin-Luther-Universität Halle-Wittenberg und Vorsitzender der Initiative für Hochbegabung e. V. Er widmet sich der Frage „Kann man ChatGPT aus der Nutzerperspektive in der physikalischen Forschung und Lehre trauen?". Es findet eine Diskussion entlang der Leitlinien zur Sicherung der guten wissenschaftlichen Praxis statt, die im Bereich der Generativen KI noch keine Anwendung finden. Fokussiert wird die Nutzung von KI als Unterstützungssystem von Lernenden und Lehrenden. Lernende haben, so der Beitrag, eine hohe Einstiegshürde, um KI effizient zu nutzen. Chat Leerzeichen entfernen GPT ist derzeit nicht kontrollierbar und damit nicht erklärbar. Die fehlende Prognostizierbarkeit führt nach Ansicht des Autors zu einer geringen Vertrauenswürdigkeit.

Prof. Dr. Sabrina Schork ist Forschungsprofessorin an der Technischen Hochschule Aschaffenburg und ist Mitglied des Kompetenzzentrums KI. In ihrem Beitrag „Vertrauensbildende Maßnahmen am Beispiel von KI-Anwendungen in der Hochschulbildung" beantwortet die beiden Fragen „Wie werden KI-Anwendungen in der Bildung eingesetzt?" und „Wie kann (berechtigtes) Vertrauen in bestehende KI-Anwendungen aufgebaut werden?". Sie zeigt auf, wie KI-Systeme im Hochschulkontext eingesetzt und welche Maßnahmen ergriffen werden, um kognitives bzw. emotionales Vertrauen aufzubauen. Durch eine vertiefende Auseinandersetzung kommt sie zu dem Ergebnis, dass die Qualität des Lehrens und Lernens durch den Einsatz von KI-Systemen verbessert werden kann. Entscheidend ist dabei die datengestützte Individualisierung von Lerninhalten und Kommunikationsweisen. Ausgereifte maschinelle Intelligenz und eine KI-Darstellung erhöhen das Vertrauen in KI-Systeme. Insbesondere wirken kognitive Faktoren auf das Vertrauen in KI-Anwendungen, die durch emotionale Faktoren verstärkt werden können.

Kapitel 5: Kulturwissenschaftliche Perspektive

Stefan Slembrouck ist Ökonom und Philosoph. Er arbeitet als Berater für digitale Energiestrategien und promoviert kooperativ an der Technischen Hochschule Aschaffenburg bei Prof. Dr. Sabrina Schork. In seinem Beitrag beleuchtet er, wie vertrauensvolle Beziehungen zwischen Mensch und Technik entstehen können und bezieht sich dabei auf konkrete Handlungskontexte in der lebensweltlichen Praxis. KI wird als sozialer Agent mit eingebauter Moralität verstanden. Als vertrauensbildende Maßnahme wird ein kritischer öffentlicher Diskurs gefordert, weniger die Einhaltung technischer Kriterien.

Prof. Dr. Derya Gür-Şeker ist Professorin an der Hochschule Bonn-Rhein-Sieg. In ihrem Beitrag „Vertrauen in KI – kulturwissenschaftlich mediale Perspektiven" geht die Autorin der Frage nach, wie sich die mediale Repräsentation von KI-Systemen in unterschiedlichen Kontexten auf Basis von Social-Media- und Zeitungsdaten erschließen lässt. KI-Diskurse werden aus gesellschaftlicher Perspektive beleuchtet. Herausgearbeitet werden Muster und Regelhaftigkeiten der Mediendarstellung, aber auch medial repräsentierte Einstellungen sowie kulturelle Praktiken im Kontext von KI und Vertrauen.

Kapitel 6: Interdisziplinäre Perspektive

Dr. Katharina Weitz ist Mitarbeiterin im Team Applied Machine Learning Group beim Fraunhofer-Institut für Nachrichtentechnik, Heinrich-Hertz-Institut, HHI. In ihrem Beitrag betrachtet die Autorin die Gestaltung menschenzentrierter KI aus einer interdisziplinären Perspektive. Ein besonderer Fokus liegt dabei auf der Gestaltung nachvollziehbarer Systeme mithilfe erklärbarer KI. Dazu werden die Konzepte des Vertrauens und der mentalen Modelle als Messgrößen für die Wirkung von KI vorgestellt. Anhand von Forschungsarbeiten in drei exemplarischen Anwendungsszenarien (Bildung, Industrie und Medizin) wird gezeigt, wie

erklärbare KI auf Nutzerinnen und Nutzer wirkt. Damit gibt dieses Kapitel einen Einblick in die aktuelle Forschung der nachvollziehbaren und menschzentrierten KI-Entwicklung.

Dr. Carlo Dindorf, Eva Bartaguiz und **Prof. Dr. Michael Fröhlich** vom Fachgebiet Sportwissenschaft der Rheinland-Pfälzischen Technischen Universität Kaiserslautern beschäftigen sich in ihrem Beitrag gemeinsam mit dem Sportökonom **Prof. Dr. Wolfgang Kemmler** von der Friedrich-Alexander-Universität Erlangen-Nürnberg mit der Vertrauensbildung beim Einsatz von KI in der Trainingslehre. Herausgearbeitet werden zentrale Themen wie Datenschutzrichtlinien, gesellschaftliche Diskussionen und erklärbare Modelle.

Dr. Johannes Schrumpf promovierte am Institut für Kognitionswissenschaft an der Universität Osnabrück und arbeitet derzeit bei applied AI in München. In seinem Beitrag setzt er sich mit der Überwindung von Herausforderungen beim Einsatz von digitalen Lernassistenten für Studierende an Hochschulen auseinander. Vorgestellt werden technische, organisatorische und nutzerbezogene Maßnahmen, die im Rahmen des vom BMBF geförderten Projekts SIDDATA (Studienindividualisierung durch datengestützte, digitale Assistenten). Schrumpf zeigt verschiedene Strategien auf, wie digitale Studienassistentensysteme zukünftig bestmöglich in die Hochschule integriert werden können.

1.4 Roter Faden

In Tab. 1.1 werden die unterschiedlichen Perspektiven der einzelnen Buchbeiträge auf das Thema ‚Vertrauen in KI' dargestellt. Die jeweiligen Kernaussagen der Beiträge werden zusammengefasst.

Deutlich wird, dass jeder Beitrag ein Stück mehr zum Verständnis des komplexen und vielschichtigen Themenfeldes ‚Vertrauen in KI' beiträgt. Lediglich bei den interdisziplinären Forschungsbeiträgen kommt es vereinzelt zu Überschneidungen zwischen Disziplinen.

Tab. 1.1 Wie Vertrauen in KI erreicht werden kann. (Eigene Darstellung)

Perspektive	Autorinnen und Autoren	Vertrauen in KI entsteht durch
Wirtschaftswissenschaftlich	Rötzel Hofmann Glinz	• emotionale, kognitive und soziale Faktoren • eine starke Marke und Vertrauen in Institutionen • das Einhalten von Designprinzipien zur Steigerung von Nutzungsraten
Sozialwissenschaftlich	Böhme und Mesenhöller Langemeyer Leaton Gray	• die Entwicklung professionalisierungsbezogener Angebote • einen verantwortungsbewussten Diskurs • einen datenschutzfreundlichen institutionellen Rahmen
Computerwissenschaftlich	Madsen und Weidl Strotherm et al. Schmitt Schork	• den Einsatz Bayes'sche Netze bei der Implementierung von transparenter, erklärbarer und vertrauenswürdiger KI • faire KI-Systeme • kontrollier- und prognostizierbare KI • ausgereifte maschinelle Intelligenz und KI-Darstellungen
Kulturwissenschaftlich	Slembrouck Gür-Şeker	• öffentliche Diskurse über KI • die Reflexion von Mustern, Einstellungen und kulturellen Praktiken über KI im Mediendiskurs
Interdisziplinär	Weitz Dindorf et al. Schrumpf	• deren Nachvollzieharbeit und Menschenzentriertheit • das Einhalten von Datenschutzrichtlinien, gesellschaftliche Diskussionen über KI und erklärbare Modelle • technische, organisatorische und nutzerbezogene Maßnahmen

Dies lässt vermuten, dass dieses Buch nur der Anfang einer multi-perspektivischen Auseinandersetzung mit dem Thema sein kann. Tab. 1.1 ist daher als eine begonnene, noch nicht abgeschlossene und keinesfalls vollständige Auflistung zu verstehen.

Betrachtet man die einzelnen Wissenschaftsdisziplinen genauer, so wird deutlich, dass sich die *wirtschaftliche Perspektive* mit Organisationen und den darin agierenden Menschen beschäftigt. Von besonderem Interesse scheint dabei die Beziehung zwischen Mensch und Maschine zu sein, welche von sozialen, emotionalen und kognitiven Faktoren bestimmt wird. Die Nutzungsrate von KI-Anwendungen und das Vertrauen in Institutionen spielen ebenso eine Rolle wie die Designprinzipien, die bei der Entwicklung verfolgt werden.

Die *sozialwissenschaftlichen* Beiträge fokussieren auf den Diskurs zwischen Menschen über KI und die sich dadurch verändernden Verantwortungsräume. Thematisiert werden auch Professionalisierungsangebote, die Nutzerinnen und Nutzer bei der Anwendung von KI unterstützen sollen und damit deren Ängste und Bedenken minimieren sollen. Für den Schutz der Daten von Nutzerinnen und Nutzern werden konkrete Rahmenbedingungen auf institutioneller Ebene eingefordert.

Die *computerwissenschaftliche* Perspektive befasst sich mit der Implementierung von fairen, erklärbaren, transparenten und vertrauenswürdigen KI-Systemen. Die Implementierungsmethode Bayes'sche Netze wird ebenso diskutiert wie die Kontrollierbarkeit und Prognostizierbarkeit generativer KI. Ausgereifte maschinelle Intelligenz und KI-Darstellungen werden eingefordert.

Die *Kulturwissenschaften* streben einen öffentlichen Diskurs über Vertrauen in KI an, der Muster, Einstellungen und kulturelle Praktiken von Medien thematisiert. Die Perspektive lässt sich am ehesten mit sozialwissenschaftlichen Beiträgen vergleichen.

Interdisziplinäre Forschungsprojekte weisen Schnittmengen zwischen den Computer-, Kultur-, Wirtschafts- und Sozialwissenschaften auf. Gefragt sind Erklärbarkeit, Nachvollziehbarkeit und Menschenzentriertheit (vgl. Computerwissenschaften), Datenschutzrichtlinien (vgl. Sozialwissenschaften), gesellschaftliche Diskussionen (vgl. Kulturwissenschaften) sowie organisatorische und nutzerbezogene Maßnahmen (vgl. Wirtschaftswissenschaften).

Literatur

1. Turing, A.M.: Computing machinery and intelligence. Mind 59(236), 433–460 (1950).
2. McCarthy, J., Minsky, M.L., Rochester, N., Shannon, C.E.: A Proposal for the Dartmouth Summer Research Project on Artificial Intelligence. Dartmouth College, Hanover (1955).
3. Görz, G., Rollinger, C.-R., Schneeberger, J.: Handbuch der künstlichen Intelligenz, 6. Auflage. Wissenschaftsverlag, Oldenbourg (2021).
4. Scheuer, D.: Akzeptanz von Künstlicher Intelligenz – Grundlagen intelligenter KI-Assistenten und deren vertrauensvolle Nutzung. Springer Vieweg, Wiesbaden (2020).

5. Europäische Kommission, What is artificial intelligence and how is it used? https://www.europarl.europa.eu/topics/en/article/20200827STO85804/what-is-artificial-intelligence-and-how-is-it-used (2023), letzter Zugriff am 17.08.2023.
6. Cremers, A.B., Englander, A., Gabriel, M., Hecker, D., Mock, M., Poretschkin, M., ..., Wrobel, S.: Vertrauenswürdiger Einsatz von Künstlicher Intelligenz: Handlungsfelder aus philosophischer, ethischer, rechtlicher und technologischer Sicht als Grundlage für eine Zertifizierung von künstlicher Intelligenz. Fraunhofer IAIS, Sankt Augustin (2019).
7. Doyle, J.: Expert Systems without Computers or Theory and Trust in Artificial Intelligence. The AI Magazine 5(2), 59–62 (1984).
8. Google Bard, Home https://bard.google.com/?utm_source=sem&utm_medium=paid-media&utm_campaign=q3deDE_sem7 (2023), letzter Zugriff am 21.08.2023.
9. Müller, J.: Datenbeschaffung für das Data Warehouse. In: Chamoni, P., Gluchowski, P. (Hrsg.) Analytische Informationssysteme. Springer Verlag, Berlin und Heidelberg (1999).
10. Günther, W.A., Mohammad H., Mehrizi, R., Huysman, M., Feldberg, F.: Debating big data: A literature review on realizing value from big data. The Journal of Strategic Information Systems 26(3), 191–209 (2017).
11. Araujo, T., Helberger, N., Kruikemeier, S., Vreese, C. H.: In AI we trust? Perceptions about automated decision-making by artificial intelligence. AI & Society 35(3), 611–623 (2020).
12. Kempton, A.M.; Parmiggiani, E.; Vassilakopoulou, P.: Accountability in Managing Artificial Intelligence: State of the Art and a way forward for Information Systems Research. In: The 31st European Conference on Information Systems (ECIS). University of Agder, Agder (2023).
13. Kaur, D., Uslu, S., Rittichier, K.J., Durresi, A.: Trustworthy artificial intelligence: a review. ACM Computing Surveys 55(2), 1–38 (2022).
14. Crawford, K.: The Atlas of AI. University Press, Yale (2021).
15. Martin, K.: Ethics of Data and Analytics – Concepts and cases, 1. Auflage. Auerbach Verlag, Leipzig (2022).
16. Reuters, Insight - Amazon scraps secret AI recruiting tool that showed bias against women https://www.reuters.com/article/us-amazon-com-jobs-automation-insight-idUSKCN1MK08G (2018), letzter Zugriff am 21.08.2023.
17. Europäisches Parlament, EU AI Act: first regulation on artificial intelligence, https://t1p.de/a7dhh (2024), letzter Zugriff am 17.08.2023.
18. BMJ, Datenschutz-Grundverordnung, https://www.bmj.de/DE/themen/digitales/digitale_buergerrechte/dsgvo/dsgvo_artikel.html (2023), letzter Zugriff am 21.08.2023.
19. Wachter, S., Mittelstadt, B., Russell, C.: Counterfactual Explanations Without Opening the Black Box: Automated Decisions and the GDPR. Harvard Journal of Law & Technology 31, 841 (2017).
20. Smeets, M.R.; Roetzel, P.G.; Ostendorf, R.J.: AI and its Opportunities for Decision-Making in Organizations: A Systematic Review of the Influencing Factors on the Intention to use AI. DU 75 (3), 432–460 (2021).
21. Omrani, O., Rivieccio, G., Fiore, U., Schiavone, F., Agreda, S.G.: To trust or not to trust? An assessment of trust in AI-based systems: Concerns, ethics, and contexts. Technological Forecasting and Social Change 181 (2022).
22. Yang, R., Wibowo, S.: User trust in artificial intelligence: A comprehensive conceptual framework. Electron Markets 32, 2053–2077 (2022).
23. Li, B., Qi, P., Liu, B., Di, S., Liu, J., Pei, J., Yi, J., Zhou, B.: Trustworthy AI: From Principles to Practices. ACM Computing Surveys 55(9), 1–46 (2023).
24. Nicodeme, C.: Build confidence and acceptance of AI-based decision support systems – Explainable and liable AI. In: 13th International Conference on Human System Interaction (HSI), pp. 20–23. IEEE, Tokyo (2020).

25. Glikson, E., Wooley, A.W.: Human trust in artificial intelligence: Review of empirical research. Academy of Management Annals 14(2), 627–6660 (2020).
26. Theodorou, A., Dignum, V.: Towards ethical and socio-legal governance in AI. Nature Machine Intelligence 2(1), 10–12 (2020).
27. Kocielnik, R., Amershi, S., Bennett, P.N.: Will You Accept an Imperfect AI? In: Brewster, S., Fitzpatrick, G., Cox, A., Kostakos, V. (eds.): Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems, pp. 1–14. ACM, Glasgow and New York (2019).
28. Duan, Y., Edwards, J.S., Dwivedi, Y.K.: Artificial intelligence for decision making in the era of Big Data – evolution, challenges and research agenda. International Journal of Information Management 48, 63–71 (2019).
29. Lockey, S., Gillespie, N., Holm, D., Someh, I.A.: A review of trust in artificial intelligence: Challenges, vulnerabilities and future directions. In: Proceedings of the 54th Hawaii International Conference on System Sciences, pp. 5463–5472. IEEE, Hawaii (2021).

Teil II
Wirtschaftswissenschaftliche Perspektive

2 Künstliche Intelligenz (KI) – unser bester Freund?

Wie Menschen auf KI-Entscheidungsempfehlungen reagieren

Peter Gordon Rötzel

2.1 Einführung

In einer Zeit, die durch den rasanten technologischen Fortschritt gekennzeichnet ist, hat sich die Künstliche Intelligenz (KI) als transformative Kraft erwiesen, die verschiedene Aspekte des Arbeitsalltags verändert [1]. Mit der Integration von KI-Systemen in die tägliche Routine stellt sich eine Frage, die zum Nachdenken anregt: Können Menschen freundschaftliche Beziehungen zu KI-Entscheidungsunterstützungssystemen aufbauen oder werden diese Systeme als Werkzeuge wie Plug-ins betrachtet?

Von virtuellen Assistenten und autonomen Fahrzeugen bis hin zu Empfehlungssystemen und medizinischer Diagnostik – KI-Technologien sind allgegenwärtig und spielen eine immer größere Rolle bei der Gestaltung unserer Gesellschaft [2]. Da KI weiterhin unseren Alltag durchdringt, ist das Verständnis der Feinheiten der Interaktion zwischen Menschen und intelligenten Maschinen von entscheidender Bedeutung.

Dieses Kapitel zielt darauf ab, in das Gebiet der Mensch-KI-Interaktionen (MKI) einzutauchen und ihre potenziellen Auswirkungen auf Individuen und Organisationen zu beleuchten. Mithilfe eines interdisziplinären Ansatzes, der Psychologie, Betriebswirtschaft und Wirtschaftsinformatik umfasst, versucht der Autor, die Dynamik, die Herausforderungen und die Chancen zu skizzieren, die sich aus der Interaktion zwischen Menschen und KI-Begleitern ergeben, wobei der Schwerpunkt auf der entscheidenden Rolle des Vertrauens in dieser Interaktion liegt.

P. G. Rötzel (✉)
Ingenieurwissenschaften, Technische Hochschule Aschaffenburg, Aschaffenburg, Deutschland
E-Mail: peter.roetzel@th-ab.de

© Der/die Autor(en), exklusiv lizenziert an Springer Fachmedien Wiesbaden GmbH, ein Teil von Springer Nature 2024

S. Schork (Hrsg.), *Vertrauen in Künstliche Intelligenz,*
https://doi.org/10.1007/978-3-658-43816-6_2

MKI umfasst die Entwicklung, Bewertung und Untersuchung von Schnittstellen und Interaktionsmodalitäten, die eine effektive Zusammenarbeit zwischen Menschen und intelligenten Maschinen ermöglichen [3]. Während sich die technologischen Fähigkeiten von KI-Systemen exponentiell entwickelt haben, bestehen die Herausforderungen im Zusammenhang mit ihrer Integration in menschenzentrierte Umgebungen fort.

Dieses Kapitel gibt einen kurzen Überblick darüber, wie Vertrauen im Umgang mit KI entsteht, und geht dabei auf die kognitiven, emotionalen und sozialen Faktoren ein, die die Entstehung und Aufrechterhaltung von Vertrauen beeinflussen. Zu den kognitiven Faktoren gehören Aspekte wie die Transparenz, Erklärbarkeit und Interpretierbarkeit von KI-Systemen [4, 5]. Emotionale Faktoren umfassen die emotionale Bindung und Beziehung zwischen Menschen und KI-Agenten [6, 7]. Soziale Faktoren umfassen die gesellschaftlichen Normen, Erwartungen und kulturellen Einflüsse, die das Vertrauen in KI-Systeme prägen [8, 9]. Darüber hinaus berichtet dieses Kapitel über Herausforderungen, die mit dem Vertrauen in KI verbunden sind, wie z. B. Verzerrungen, Fehler und Unsicherheiten, die das Vertrauen untergraben können. Das Verständnis dieser Herausforderungen und die Erforschung möglicher Lösungen sind entscheidend für die Entwicklung vertrauenswürdiger KI-Systeme, die mit den menschlichen Bedürfnissen und Werten übereinstimmen.

Frühere Forschungen haben empirisch belegt, dass Vertrauen ein wesentlicher Faktor für die Beurteilung von KI-Entscheidungshilfen ist [10]. Jüngste Entwicklungen im Bereich der natürlichen Sprachverarbeitung, der emotionalen Intelligenz und des Deep Learning haben KI-Systeme in die Lage versetzt, anspruchsvolle Gespräche zu führen, menschliche Emotionen zu erkennen und darauf zu reagieren sowie sich an individuelle Vorlieben und Verhaltensweisen anzupassen [11]. In dem Maße, in dem KI in ihren Fähigkeiten dem Menschen immer ähnlicher wird, wird die Möglichkeit, echte Beziehungen zu diesen intelligenten Wesen aufzubauen, zu einer verlockenden Perspektive.

Auch wenn der Gedanke an eine ‚Freundschaft' zwischen Mensch und KI auf Skepsis oder Widerstand stoßen mag, ist es wichtig, die potenziellen Vorteile zu bedenken. Die Forschung hat gezeigt, dass soziale Verbundenheit und Kameradschaft eine zentrale Rolle für das menschliche Wohlbefinden spielen und die psychische Gesundheit und die allgemeine Lebenszufriedenheit fördern [12]. Angesichts der weit verbreiteten Einsamkeit und sozialen Isolation in der modernen Gesellschaft [13] könnte die KI-Begleitung eine einzigartige Möglichkeit bieten, diese Herausforderungen anzugehen, indem sie den Menschen eine Quelle emotionaler Unterstützung, intellektueller Anregung und Begleitung bietet.

Zwei gegensätzliche Phänomene, die in dieser Interaktion auftauchen, sind die Automatisierungsneigung und die Tendenz zur Algorithmusvermeidung. Voreingenommenheit bei der Automatisierung bezeichnet die Neigung von Menschen, sich unhinterfragt auf KI-Empfehlungen oder -Entscheidungen zu verlassen, während die Tendenz zur Algorithmusvermeidung die Neigung beschreibt, KI-Empfehlungen zugunsten eines menschlichen Urteils abzulehnen oder außer Kraft zu setzen. Das Zusammenspiel dieser beiden Tendenzen stellt eine komplexe Herausforderung in der Mensch-KI-Interaktion dar.

Voreingenommenheit gegenüber der Automatisierung tritt auf, wenn Menschen KI-Systemen unangemessenes Vertrauen entgegenbringen und sich auf sie verlassen, was dazu führt, dass sie Entscheidungen treffen oder Maßnahmen ergreifen, die ausschließlich auf den von der KI generierten Ergebnissen beruhen, ohne eine ausreichende kritische Bewertung oder Überprüfung vorzunehmen. Diese Voreingenommenheit kann durch Faktoren wie die vermeintliche Unfehlbarkeit von KI, den Wunsch nach Effizienz oder den Glauben, dass KI dem menschlichen Urteilsvermögen überlegen ist, entstehen. Die Voreingenommenheit bei der Automatisierung kann erhebliche Folgen haben, die von kleinen Unannehmlichkeiten bis hin zu schwerwiegenden Fehlern reichen, insbesondere wenn KI-Systeme unvollkommen sind oder Fehler machen [14].

Andererseits manifestiert sich die Tendenz zur Algorithmusvermeidung als Widerstand gegen die vollständige Übernahme von KI-Empfehlungen oder -Entscheidungen, was dazu führt, dass Menschen KI-Ergebnisse zugunsten ihres eigenen Urteils oder ihrer Intuition ignorieren oder außer Kraft setzen. Diese Tendenz kann auf mangelndes Vertrauen in KI-Systeme, auf Bedenken hinsichtlich der Transparenz und Interpretierbarkeit von KI-Algorithmen oder auf den Wunsch zurückzuführen sein, die menschliche Kontrolle und Autonomie zu wahren. Die Tendenz zur Algorithmusvermeidung kann die potenziellen Vorteile der KI behindern, die Effizienz der Entscheidungsfindung beeinträchtigen und dazu führen, dass Gelegenheiten zur effizienten Nutzung der KI-Funktionen verpasst werden [15].

Das Spannungsverhältnis zwischen der Voreingenommenheit gegenüber der Automatisierung und der Tendenz zur Vermeidung von Algorithmen verdeutlicht das empfindliche Gleichgewicht, das bei der Interaktion zwischen Mensch und KI entscheidend sein kann. Das Streben nach einer besseren Zusammenarbeit zwischen Menschen und KI-Systemen erfordert die Bewältigung dieses Spannungsverhältnisses, um die Stärken beider Parteien zu nutzen und gleichzeitig die mit blindem Vertrauen oder ungerechtfertigter Skepsis verbundenen Risiken zu mindern.

Zu den kognitiven Faktoren gehören die Wahrnehmung der KI-Kompetenz, die Erklärbarkeit und Transparenz von KI-Algorithmen und das Verständnis für die Grenzen der KI. Emotionale Faktoren umfassen vertrauensbildende Mechanismen, Nutzererfahrungen und die Auswirkungen von KI auf Emotionen und Vertrauen [4, 16, 17]. Soziale Faktoren umfassen den Einfluss gesellschaftlicher Normen, kultureller Faktoren und ethischer Überlegungen auf die Einführung und Akzeptanz von KI-Systemen [18].

2.2 Zwei Seiten einer Medaille – Wirtschaftliche Vorteile, Hypes und Ängste zu MKI

Während sich KI-Anwendungen rasant entwickelt haben, haben die Vorbehalte der Menschen gegenüber KI Schritt gehalten. Die Hauptbedenken gegen KI sind, dass KI Informationen verzerrt und massiv in die Entscheidungsfreiheit der Menschen eingreift.

Wie sollen Menschen echte Informationen von KI-generierten Informationen unterscheiden können? Welche Auswirkungen könnten KI-generierte Informationen auf die politischen und sozial-ökologischen Netzwerke haben? Woher wissen die Menschen, ob KI sie manipuliert? Stephen Hawking äußerte Vorbehalte, als er sagte, dass eine vollständige KI „das Ende der menschlichen Rasse bedeuten könnte“ [19].

Um einen Hinweis auf aktuelle Suchtrends zu erhalten, wurden bei der Eingabe von „wird KI“ in Google die häufigsten Suchanfragen aufgelistet, wie in Abb. 2.1 dargestellt:

Obwohl KI bei den Menschen viele Vorbehalte hervorruft und zunehmend eine wichtige Rolle in unserer Wirtschaft spielt, gibt es keine einheitliche Definition des Begriffs ‚KI‘ [19]. KI wird allgemein als die Fähigkeit von Maschinen beschrieben, durch Erfahrung zu lernen und auf neue Reize zu reagieren. Glickson und Woolley [20, S. 3] definieren KI wie folgt: „Künstliche Intelligenz (KI) ist eine hochleistungsfähige und komplexe Technologie, die darauf abzielt, menschliche Intelligenz zu simulieren.“ Wenn KI darauf abzielt, die menschliche Intelligenz zu simulieren, und dies gelingt, wo liegen dann die Grenzen, die verhindern, dass KI den Menschen ersetzt? Diese Überlegungen könnten zu einem Misstrauen gegenüber KI und KI-generierten Empfehlungen oder Entscheidungsgrundlagen führen. Das könnte der Kipppunkt sein: Sollten Menschen Angst vor KI haben, wenn KI vermeintlich langweilige Routineaufgaben ersetzt?

Um KI in einer Volkswirtschaft zu implementieren und die Akzeptanz der Menschen für KI zu entwickeln, ist eine Definition der Mensch-Maschine-Kollaboration notwendig. In den Wirtschaftswissenschaften wurde in den letzten Jahren der Begriff ‚Komplementäre, kollaborative Intelligenz‘ (KKI) entwickelt [21–23]. KKI bedeutet eine kooperative Arbeitsweise von Menschen und Maschinen. Wilson und Daugherty [21] geben einige Beispiele für KI in der Wirtschaft, insbesondere in den Bereichen IT und traditionelle High-Tech-Industrie, Finanzen, Medizin, soziale Medien, Personalwesen, Dienstleistungen, Sicherheitsaspekte und tägliches Leben, die in Abb. 2.2 zusammengefasst sind.

Abb. 2.1 Gefunden über eine Google Suchanfrage am 07.07.2023 um 14:23 Uhr

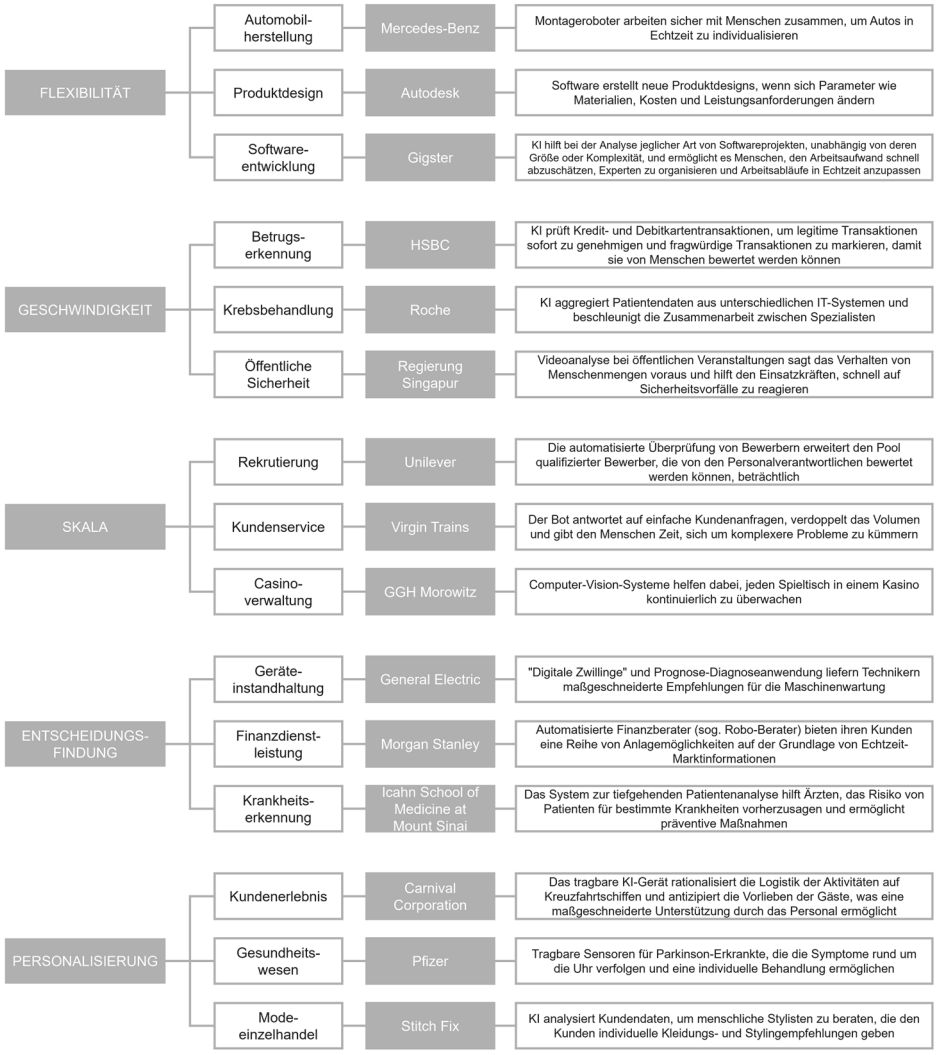

Abb. 2.2 Co-Working-Dimensionen von KI [21, S. 9]

Die KKI hat bereits einige Bereiche der Wirtschaft und insbesondere die Mensch-Maschine-Schnittstelle verbessert und zu größerer Flexibilität, besserer Arbeitsleistung und besserer Entscheidungsfindung beigetragen. Weitere Bereiche, in denen KI Einzug gehalten hat, sind die Wirtschaftsprüfung [1], die Prozessentwicklung [10] und in Projektteams [24]. Die aktuelle Diskussion über den Wettbewerb zwischen Mensch und Maschine muss daher zu einer besseren Verflechtung von KI-gestützter und menschlicher Arbeit führen.