



Quo Vadis: Auf dem Weg zu Ethik-Guidelines für den Einsatz KI-basierter Lern-Companions in der Lehre?

Ricarda Schlimbach · Bijan Khosrawi-Rad ·
Susanne Robra-Bissantz

Eingegangen: 1. Oktober 2021 / Angenommen: 3. Februar 2022 / Online publiziert: 2. März 2022
© Der/die Autor(en) 2022

Zusammenfassung Begünstigt durch die wachsende Relevanz digitaler Unterstützungsangebote für das Lehren und Lernen, bspw. in Schulen und Universitäten, gewinnt der Forschungszeitung der Conversational Agents im Bildungskontext an Bedeutung. Sogenannte *Lern-Companions*, also virtuelle Lern-Gefährten, bieten das Potenzial einer nutzeradaptiven sowie örtlich und zeitlich flexiblen Lernbegleitung. Da mit der Anwendung derart neuartiger Zukunftstechnologien ein hohes ethisches Verantwortungsbewusstsein einhergeht, diskutiert dieser Beitrag Chancen und Herausforderungen durch den Einsatz von Lern-Companions aus ethischer Perspektive. Basierend auf einer qualitativen Studie mit sechs Experten sowie der Diskussion der gewonnenen Erkenntnisse mit bestehender Literatur zeigt dieser Artikel sieben praktische Handlungsempfehlungen zur Gestaltung virtueller Lern-Companions aus einem ethischen Blickwinkel auf und adressiert dabei unter anderem den Umgang mit Ethik-Guidelines, aktive Nutzereinbindung und Bildungsgerechtigkeit.

Schlüsselwörter Lern-Companion · Ethik · Ethische Gestaltung · Guidelines · Conversational Agents

Ricarda Schlimbach (✉) · Bijan Khosrawi-Rad · Susanne Robra-Bissantz
Institut für Wirtschaftsinformatik, Lehrstuhl für Informationsmanagement, Technische Universität
Braunschweig, Braunschweig, Deutschland
E-Mail: r.schlimbach@tu-bs.de

Bijan Khosrawi-Rad
E-Mail: b.khosrawi-rad@tu-bs.de

Susanne Robra-Bissantz
E-Mail: s.robra-bissantz@tu-bs.de

Quo Vadis: Towards Ethical Guidelines for the Use of AI-based Learning Companions in Teaching?

Abstract The growing relevance of digital support services for teaching and learning, e.g., in schools and universities, has led to an increase in the importance of Conversational Agents in educational contexts. So-called *Learning Companions* offer the potential of user-adaptive learning support that is flexible in terms of location and time. Since the application of such novel future technologies is accompanied by a strong sense of ethical responsibility, this paper discusses opportunities and challenges posed by the use of Learning Companions from an ethical perspective. Based on a qualitative study with six experts and the discussion of the findings with existing literature, this article presents seven practical recommendations for the design of virtual learning companions from an ethical perspective and addresses, among other aspects, the handling of ethical guidelines, active user involvement, and educational equity.

Keywords Learning companion · Ethics · Ethical design · Guidelines · Conversational agents

1 Einleitung

Im Lernalltag müssen Studierende, Schüler¹ und Berufstätige vielfältige Hindernisse überwinden und lösen, sei es die Orientierung auf dem (Online-) Campus, die Lernorganisation oder Herausforderungen in puncto Motivation und Zeitmanagement. Oft mangelt es an individueller Begleitung, was sich insb. an Universitäten zeigt, da die Anzahl von Lehrenden und Lernenden in einem unausgeglichenen Verhältnis steht (Winkler et al. 2019). Neben diesem Mangel an Unterstützungsleistungen ist auffällig, dass viele Bildungseinrichtungen das Potenzial der Nutzung digitaler Technologien noch nicht ausreichend ausgeschöpft haben, was nicht zuletzt in der Corona-Pandemie deutlich geworden ist. *Pedagogical Conversational Agents* (PCAs) stellen eine digitale Hilfestellung für Lernende dar und bilden eine kontextspezifische Untergruppe der Conversational Agents (CAs). Sie zeichnen sich dadurch aus, dass sie dialogbasiert (als Chatbots oder sprachbasiert ähnlich Siri oder Alexa) mit dem Nutzer kommunizieren, den Lernenden begleiten, und durch die Simulation menschlicher Sprache sowie oft auch ergänzt um einen Avatar, menschenähnliches Verhalten zeigen (Winkler et al. 2019; Wellnhammer et al. 2020; Bendel 2021). Sie bieten den Vorteil, leicht skalierbar, ortsunabhängig und (zeitlich) dauerhaft verfügbar zu sein. Daraus ergibt sich die Chance, gefördert durch den Einsatz von *Künstlicher Intelligenz* (KI), individualisiert auf die Lernenden einzugehen (Hobert und Meyer von Wolff 2019). Erste Entwicklungen von PCAs begannen bereits in den 1980-er sowie 90-er Jahren (Bendel 2003), wobei das Themengebiet in den letzten Jahren zunehmende Aufmerksamkeit erlangt, wie die steigende Zahl

¹ Zugunsten der Lesbarkeit verwenden wir im gesamten Beitrag das generische Maskulinum und weisen ausdrücklich darauf hin, dass stets alle Geschlechter gemeint sind.

an Veröffentlichungen zeigt (Hobert und Meyer von Wolff 2019; Karrenbauer et al. 2021). Eine besondere Form der PCAs sind *Lern-Companions* (LCs), welche in ihrer Funktion als Gefährten eine enge soziale Bindung, ähnlich einer freundschaftlichen Beziehung, zu dem Lernenden aufbauen (Bendel 2003, 2021). Diese zeichnet sich dadurch aus, dass der Companion die individuellen Bedürfnisse des Nutzers adressiert und hilfsbereit agiert (Rawlins 2017; Strohmann 2021). LCs besitzen zudem das Potenzial, mittels Verfahren der KI sowie des Machine Learnings intelligent, proaktiv und adaptiert an den Nutzer zu kommunizieren, sowie sich Lernstände und -fortschritte zu merken, (Bendel 2003, 2021; Kim et al. 2006). Die Nutzung und gezielte Analyse von Daten spielt dabei eine große Rolle, um die Bedürfnisse der Nutzer individuell zu erfassen und auf deren Wünsche einzugehen (Stucki et al. 2020; Strohmann 2021). Während frühere Ansätze von CAs sehr eingeschränkte Funktionalitäten und ein mangelndes Sprachverständnis aufwiesen, entwickelte sich das Themenfeld in den letzten Jahren stark weiter, sodass durch neuartige Technologien (z.B. neuronale Netze) menschliche Sprache weitaus besser interpretiert werden kann (Stucki et al. 2020). Für die Ermöglichung solch intelligenter Services werden vielfältige Nutzerdaten einbezogen, zum einen um den Dialog auf der Basis von Trainingsdaten menschenähnlicher zu gestalten, zum anderen um auf Basis von Informationen über den Lernprozess des Nutzers individuelle Hilfestellungen anbieten zu können, bspw. durch Anwendung von Machine-Learning-Algorithmen (Gubareva und Lopes 2020).

In Forschung und Praxis gewinnt ein ethisch reflektierter Umgang mit KI sowie die Analyse von Nutzerdaten an Bedeutung (Bendel 2021), wie sich u.a. an vermehrten Tracks auf wissenschaftlichen Konferenzen oder der wachsenden Zahl an Ethik-Richtlinien erkennen lässt (Enquete-Kommission Künstliche Intelligenz 2020). Allerdings sind ethische Aspekte bei der Konzeption sowie Gestaltung von CAs bislang zweitrangig oder sogar unberücksichtigt geblieben (Ruane et al. 2019). Nur vereinzelte Wissenschaftler adressieren explizit Aspekte wie die Datensammlung, Verantwortung und Haftung, das Ersetzen von Lehrkräften oder Inklusion durch PCAs aus einem ethischen Blickwinkel (Bendel 2021; Wambsgans et al. 2021). Es mangelt daher an richtungweisenden Empfehlungen für die praktische Ausgestaltung. Um die genannten Potenziale von LCs für den Einsatz in realen Lehr- und Lernszenarien auszuschöpfen, ist ein praxisorientierter ethischer Orientierungsrahmen allerdings unverzichtbar, um z. B. diskriminierende oder unfaire Entscheidungen durch KI-Algorithmen zu vermeiden (Casas-Roma und Conesa 2021). Aufgrund der rasanten technologischen Entwicklung von KI-Systemen sowie ihrer zunehmend erlangten Autonomie, besteht der Bedarf, sich frühzeitig – im Sinne der Proaktivität – mit ethischen Aspekten in Hinblick auf den Einsatz von KI zu befassen, sowohl aus Sicht der Gesellschaft als auch aus Perspektive der Forschung und Praxis. Ethik ist „eine philosophische Disziplin, die die Moral untersucht, hinterfragt und begründet“ (Bendel 2021, S. 154). In diesem Beitrag steht die Informationsethik im Vordergrund, welche sich u.a. mit KI-Systemen und der ethischen Sicht auf Datenverarbeitung und Algorithmen befasst (Bendel 2021), da für den Einsatz von LCs die intelligente Datenverarbeitung essenziell ist. Dass die Auseinandersetzung mit KI aus ethischer Sicht relevant ist, zeigt sich nicht nur auf nationaler Ebene durch die Einrichtung einer Enquete-Kommission im Bundestag zur Verabschiedung

eines Orientierungsrahmens zum verantwortungsvollen Umgang mit KI (Enquete-Kommission Künstliche Intelligenz 2020), sondern auch die EU hat sich bereits im Rahmen der „Ethik-Leitlinien für eine vertrauensvolle künstliche Intelligenz“ (Hochrangige Expertengruppe für künstliche Intelligenz 2019) mit der Thematik befasst und Guidelines für den Einsatz von KI formuliert. Diese Leitlinien sind an dem Grundrecht nach Privatsphäre sowie am Diskriminierungsverbot ausgerichtet, welchen im Zuge der Zunahme an gespeicherten sowie verarbeiteten personenbezogenen Daten im Digitalzeitalter eine besondere Bedeutung zukommt. Zentral ist hier die Sicherstellung der Vertrauenswürdigkeit, so soll KI z. B. auf den Menschen ausgerichtet sein, diesem keinen Schaden zufügen, im guten Sinne handeln und auf den Grundsätzen von Autonomie und Gerechtigkeit basieren. Diese Grundätze bieten zwar ein Rahmenwerk, es mangelt allerdings an Substanz und Detailtiefe, insb. in Hinblick darauf, was unter dem „Handeln im guten Sinne“ oder der „Vermeidung von Schaden“ zu verstehen ist. Folglich besitzen die bestehenden Ethik-Leitlinien eine mangelnde Übertragbarkeit auf den Einsatz von KI in der Praxis (Luchterhandt 2021).

Ziel des Beitrags ist daher, LCs aus ethischer Perspektive zu betrachten und Orientierung für ihre Entwicklung und Integration im Praxisalltag zu generieren. Dies erfolgt durch Interviews mit Experten aus Forschung sowie Praxis, deren reflektierte Ergebnisse zu Handlungsempfehlungen für die Gestaltung von LCs zu Lehr- und Lernzwecken führen. Der Artikel ist wie folgt gegliedert: Im folgenden Kap. 2 wird das methodische Vorgehen des Beitrags dargestellt, woraufhin in Kap. 3 die wesentlichen Ergebnisse in Hinblick auf Herausforderungen und Chancen von LCs aus ethischer Sicht sowie die daraus abgeleiteten Handlungsempfehlungen als Implikationen für die Praxis dargelegt werden. Abschließend werden die gewonnenen Erkenntnisse zusammengefasst und kritisch reflektiert (Kap. 4).

2 Methodisches Vorgehen

Mit dem Ziel, vielfältige Perspektiven für den Einsatz von LCs in der Lehre aufzeigen zu können, wurden in den Kalenderwochen 33–35 im Jahr 2021 Interviews mit sechs per E-Mail akquirierten, deutschsprachigen Experten aus Forschung und/oder Praxis geführt (vgl. Tab. 1). Bei der interdisziplinären Auswahl der Experten wurde auf fachliche Expertise (Ethik, CAs, Analyse der Lerndaten), Funktionsexpertise (z. B. Promotion oder Professur im Fachbereich, Bundestagsreferent), demographische Diversität (Alter, Geschlecht, Wohnort) und auf den Einbezug von Perspektiven aus Forschung und Praxis geachtet (vgl. Tab. 1). Infolgedessen konnten trotz der geringen Stichprobengröße ($n=6$) facettenreiche Expertenmeinungen generiert werden (Brüsemeyer 2008). Als Orientierungsrahmen für den Interviewverlauf wurde vorab ein Interviewleitfaden erstellt, welcher für alle Befragungen in einer standardisierten Form verwendet wurde. Dieser Leitfaden beinhaltet größtenteils offene Fragen, welche sich auf die Chancen sowie Herausforderungen des Einsatzes von LCs aus ethischer Sicht, notwendige Kontrollinstanzen für den vertrauensvollen Umgang mit KI in der Lehre sowie abzuleitenden Gestaltungsempfehlungen und -richtlinien beziehen. Ein besonderer Fokus wurde auf die Erfragung

Tab. 1 Übersicht über die befragten Experten

Experten	Hintergrund und Qualifikationen
Experte 1 (E1)	<i>Forscher & assoziierter Professor</i> Forschung zur Gestaltung von CAs und Virtuellen Companions
Experte 2 (E2)	<i>Forscher</i> Forschung zur Gestaltung von CAs und Virtuellen Companions
Experte 3 (E3)	<i>Berater & Forscher</i> Chatbot-Beratung & Forschung zu den Einsatzzwecken von Chatbots
Experte 4 (E4)	<i>Wirtschaftsinformatiker</i> Software-Entwicklung
Experte 5 (E5)	<i>Forscher & Gründer</i> Forschung im Bereich IT-Sicherheit sowie Lerndatenanalyse & Lern-Coach
Experte 6 (E6)	<i>Bundestagsreferent</i> Fokus auf KI & Ethik

von Empfehlungen für einen verantwortungsvollen Umgang mit der Sammlung sowie gezielten Auswertung von Lernendendaten gelegt, weil diese essentiell für die individuelle Begleitung des Lernenden und somit für den neuen Forschungsstrang der LCs sind. Zudem wurden die Befragten zum Schluss des Interviews gebeten, kurze Einschätzungen zur zukünftigen Entwicklung von KI in der Lehre und den damit verbundenen Risiken zu geben, z. B. dahingehend, inwiefern diese die soziale Ungerechtigkeit oder die Diskriminierung von Minderheiten in der Gesellschaft verschärfen könnten. Um tiefere Erkenntnisse zu erlangen und den Befragten die Möglichkeit zur detaillierten Erläuterung genannter Aspekte zu geben, wurden bei Bedarf Rückfragen in den Interviewverlauf eingestreut. Die Interviews hatten eine Dauer von 25–50 min und wurden aufgezeichnet, vollständig transkribiert sowie mithilfe der Software *MAXQDA* analysiert, wobei die Autoren insgesamt mehr als 500 Kodierungen vergaben und sich zur Sicherstellung der korrekten wissenschaftlichen Auswertung in einem Peer-Review gegenseitig überprüften.

3 Ergebnisse

3.1 Herausforderungen und Chancen von LCs aus ethischer Sicht

Die Interviewteilnehmer thematisierten verschiedene ethische Herausforderungen, welche mit dem Einsatz von LCs in der Lehre einhergehen. Der Fokus der Förderung von interpersonellen Beziehungen durch virtuelle Companions (Krämer et al. 2011) könnte dazu führen, dass Nutzer sich in virtuellen Welten „verlieren“ oder von LCs abhängig werden (E1, E4, E6). E1 führt hierzu die Beispiele der bereits in der Praxis eingesetzten virtuellen Companions Replika AI² sowie der von Microsoft entwickelten Anwendung XiaoIce³ an: Da sich sogar bereits Menschen in die dahinterstehende KI verliebt haben, sind auch im Lernkontext unerwünschte Zukunfts-

² <https://replika.ai/>, Zuletzt eingesehen: 28.09.2021.

³ <https://www.xiaoice.com/>, Zuletzt eingesehen: 28.09.2021.

szenarien denkbar, z. B. wenn Studierende sich von ihren Kommilitonen entfremden (E1, E3, E4). Eine zunehmende Menschenähnlichkeit könnte sich zudem negativ auf die Empfindungen der Nutzer auswirken (E5, E6), was auch als „Uncanny Valley“ bzw. „unheimliches Tal“ bezeichnet wird. Dieses von Mori (1970) erstmals für Roboter aufgezeigte Phänomen zeichnet sich dadurch aus, dass ab einem gewissen Grad der Menschenähnlichkeit intelligenter Systeme die Gefahr besteht, dass diese Erscheinung als beängstigend wahrgenommen wird, und Nutzer der Technologie folglich weniger vertrauen. Auch wenn diese Theorie aufgrund nicht ausreichender empirischer Belege umstritten ist, heben mehrere Forscher hervor, dass es sich um ein auch im Kontext von CAs zu beachtendes Problem handelt (z. B. Strohmann 2021; Diederich et al. 2022), was aus Sicht der Autoren berücksichtigt werden sollte, um die Akzeptanz von LCs zu steigern. Ein weiteres Risiko besteht darin, dass LCs Diskriminierung verschärfen könnten, bspw. falls diese auf Basis rassistischer Input-Daten trainiert werden (E2, E4, E6). E2 nannte hierzu das Negativbeispiel des Chatbots *Microsoft Tay*, welcher ungewollt, durch von den Nutzern eingespeiste Interaktionsdaten, rassistische Handlungsweisen erlernte und daher binnen von Stunden wieder offline gestellt wurde. Eine Gefahr bestehe darüber hinaus darin, dass fehlende Zugänglichkeit zu LCs aufgrund mangelnder technischer Ausrüstung, Affinität oder finanziellen Mitteln zu Chancenungleichheit führen könnte (E5). Auch der verantwortungsvolle Umgang mit den Lernendendaten wurde als Herausforderung gesehen (E2, E4, E5). So handle es sich nach E5 hierbei um hochsensitive Daten, vor allem, da Lernende sich in den Interaktionen mit LCs bewusst Schwächen eingestehen müssen. Beachtet werden sollte jedoch, dass von mehreren Befragten vor einer Überregulierung im Rahmen des Datenschutzes aufgrund einer drohenden Hemmung der Innovationskraft gewarnt wurde (E1, E2, E3, E6). Dies ist allerdings aus Sicht der Autoren sowie mehrerer Experten nicht damit gleichbedeutend, dass der Datenschutz im Zusammenhang mit neuen KI-Technologien zweitrangig wäre – ganz im Gegenteil: Die Verwendung sicherer Verschlüsselungsalgorithmen (E5), die proaktive Vermeidung von Cyberangriffen (E4) oder die Transparenz gegenüber den Nutzern in Hinblick auf die Datenverarbeitung (E1, E2, E4, E5, E6) sind sogar wichtiger denn je. Vielmehr gilt es, das Schüren unbegründeter Ängste zu vermeiden (E1, E6).

Neben den genannten Herausforderungen wurden von den Experten auch Chancen in Bezug auf ethische Gesichtspunkte genannt. So bestünde eine Hoffnung auf „echte Chancengleichheit zwischen verschiedenen sozialen Gruppen“ (E4), indem man Bildungsangebote allen Menschen zugänglich machen könne, auch denen, „die sich sonst keine persönlichen Nachhilfelehrer leisten könnten“ (E4). Der Einsatz von KI biete nach E2, E3 und E4 z. B. die Chance, dem Problem entgegenzuwirken, dass Studierende aufgrund des Charakters von Massenvorlesungen an Universitäten selten die Möglichkeit haben, individuelle Unterstützung durch die Lehrenden zu erhalten (Lehmann et al. 2015). Die hohe Skalierbarkeit von LCs könnte es möglich machen, dieses Betreuungsverhältnis zu verbessern, sodass E2 zufolge sogar eine 1:1-Betreuung von Lernenden zukünftig möglich erscheint (Winkler et al. 2019). In Bezug auf die Chancengleichheit widersprechen sich demnach die Aussagen. E5 erläutert hierzu, dass ein gewisses „Tal der Ungerechtigkeit“ bestünde: Zunächst würde die Gerechtigkeit aufgrund technologischer Hürden bei der Umsetzung und

der ungenügenden Zugänglichkeit der neuartigen Technologie fallen, wobei sich ab einem gewissen „Tipping Point“ – sobald Bildungseinrichtungen genügend Ressourcen in die Ausgestaltung sowie gerechte Verteilung der Technologie investiert haben – die soziale Gerechtigkeit verbessern könnte.

3.2 Abgeleitete Handlungsempfehlungen

Synthetisiert aus der Literatur und den Positionen der Experten ergeben sich sieben von den Autoren dieses Beitrags abgeleitete Handlungsempfehlungen zur ethisch verantwortungsvollen Gestaltung von LCs:

1. Sparsamer Umgang mit Ethik-Guidelines und dynamische Einzelfallprüfung

Das AI Global Ethics Inventory⁴ verzeichnet im September 2021 in seiner Datenbank 173 Ethik-Guidelines – eine Verdopplung innerhalb von zwei Jahren (Enquete-Kommission Künstliche Intelligenz 2020). Wenngleich sich daraus die steigende Relevanz und Notwendigkeit von Ethik-Guidelines ableiten lässt, so birgt die hohe Dichte der Guidelines gleichzeitig die Gefahr der Unübersichtlichkeit und Überregulierung. Infolgedessen schlägt sogar der Deutsche Bundestag vor, von weiteren Ethik-Katalogen abzusehen und verweist auf „das Grundgesetz der Bundesrepublik bzw. die Grundrechtecharta der EU mit dem Begriff der Menschenwürde als Maßgabe für alle politische Gestaltung“ (ebd., S. 83). Diese Einschätzung deckt sich mit dem Credo der befragten Experten, so viel wie nötig und so wenig wie möglich zu regulieren. Mehrere Experten warnen vor einer Überregulierung, welche die Innovationskraft beeinträchtigen würde (E1, E2, E3, E6). Zudem besteht ein Konsens, dass Ethik-Guidelines dynamisch und zu jeder Zeit anpassbar sein müssten (E1–E6). E5 warnt vor der Gefahr, dass Ethik-Guidelines als „Kompromisslösung“ einen niedrigeren Standard haben könnten als dieser eigentlich notwendig wäre, um verantwortungsbewusst mit dem Einsatz von KI umzugehen. Sinnvoller sei es nach E4 und E5, wenn einzelne Marktakteure bzw. Bildungsträger selbst jeweils dynamisch anpassbare Standards für den Einsatz von KI in der Lehre konzipieren. Auch in Hinblick auf die Vermeidung diskriminierender Praktiken durch CAs schlägt E6 eine kontextbezogene Einzelfallprüfung anstelle generischer Richtlinien vor. Zudem helfen sogenannte Risikostufenmodelle (European Commission 2021) einzuschätzen, in welchen Kontexten – je nach Höhe des Risikos negativer Folgen durch den Einsatz von KI – eine stärkere Regulierung notwendig wäre (E6). Bzgl. der Kontrolle ethischer Gesichtspunkte raten mehrere Experten ebenfalls von einem zu starken Korrektiv durch eine übergeordnete Instanz ab. Viel wichtiger sei es, die Nutzer als Empfänger dieser zukunftsweisenden Technologie darin einzubinden, basierend auf ihren Anforderungen, Ethik-Standards festzulegen (E1, E2, E6). E1 schlägt hierzu bspw. eine „Crowd-basierte“ Prüfinstanz vor, welche aus Nutzern besteht, die LCs auf eine ethisch vertretbare Gestaltung validieren.

⁴ <https://inventory.algorithmwatch.org>, Zuletzt eingesehen: 21.12.2021.

2. Aktive Einbindung von Nutzern in den Gestaltungsprozess

Diese aktive Einbindung (potenzieller) Nutzer in den Gestaltungsprozess neuartiger KI-Systeme sowie entsprechender Guidelines ist ein in den bestehenden Ethik-Guidelines kaum adressierter Aspekt, der weder im Ergebnis der Untersuchung des Bundestags zu häufig genannten Aspekten in Ethik-Guidelines (Enquete-Kommission Künstliche Intelligenz 2020), noch im o. g. AI Global Ethics Inventory adressiert wird. Dieser Aspekt wurde von allen befragten Personen hervorgehoben. E2 betont, ethische Verantwortung müsse „aus der Gesamtheit der Nutzer kommen“. Daher plädiert E2 für einen ethischen Kodex, der dynamisch auf Basis der Nutzer gepflegt wird und die Werte sowie Moralvorstellungen der Lernenden berücksichtigt. Da der Einsatz von KI-Technologie das Risiko birgt, dass die Analyse von Lernendendaten zu Fehlentscheidungen führt, ist es zudem wichtig, kontinuierliches Feedback von den Lernenden zu erhalten. So sollten Lernende dazu befragt werden, ob ihre Lernpräferenzen tatsächlich korrekt durch den LC abgebildet werden (E5). Die kontinuierliche Erhebung der Nutzerbedürfnisse ist folglich von zentraler Bedeutung (E3, E5, E6). Zusätzlich sollten LCs eine individuelle Anpassbarkeit ermöglichen, z. B. in Hinblick auf die optische Ausgestaltung der Haut- und Augenfarbe des Avatars (E2, E4, E6), das Alter und Geschlecht (E1, E3), den Kommunikationsstil (E2, E3) oder die Nationalität (E4). Aus ethischer Betrachtungsweise empfehlen die Autoren, in den Auswahlmöglichkeiten auch unterrepräsentierte Merkmale abzubilden und dadurch bewusst Vielfalt zu integrieren.

3. Diskriminierung gezielt entgegenwirken

Diese Abbildung von Vielfalt kann zudem dadurch befördert werden, dass diskriminierende Praktiken durch die KI präventiv verhindert werden. KI-Systeme, die mit personenbezogenen Daten arbeiten, bergen nämlich das Risiko der ungerechtfertigten Benachteiligung bestimmter Personengruppen. Da sie regelbasiert entscheiden, können sich diskriminierende Muster fortpflanzen. Dieses auch *Algorithmic Bias* (E1) genannte Fehlverhalten von KI-Systemen beruht auf einer systematischen Verzerrung der Input-Daten (Casas-Roma und Conesa 2021; Wambsganss et al. 2021). Dennoch ist das Erzeugen eines solchen *Bias* notwendig, um KI-Systemen das Generalisieren von Zusammenhängen überhaupt zu ermöglichen. Die technischen Systeme stehen in einem ständigen Zielkonflikt zwischen Voraussagequalität und Fairness, denn besonders Minderheiten sind in den Datensätzen meist unterrepräsentiert (Enquete-Kommission Künstliche Intelligenz 2020). Neuerdings wird nach Maßgabe des europäischen „AI Acts“ (European Commission 2021) sowie des deutschen Bundestags erstmalig KI-basierte, systematische Diskriminierung rechtlich verfolgt und eingeschränkt. Maßnahmen umfassen insb. die Vermeidung von Diskriminierung aufgrund fehlender Diversität in den Trainingsdaten für die KI, indem sich reproduzierende Diskriminierungsmuster begrenzt werden. Zudem werden sogenannte „Fairnessmaße“ eingeführt (Enquete-Kommission Künstliche Intelligenz 2020, S. 63). Dabei handelt es sich um Metriken, welche messen, inwiefern richtige bzw. falsche Entscheidungen eines (KI-) Algorithmus gerecht auf verschiedene Personengruppen verteilt sind. Hierzu werden zur Ermittlung unterschiedliche Krite-

rien wie die Fehlertoleranz bei Entscheidungen herangezogen (Enquete-Kommission Künstliche Intelligenz 2020; Orwat 2019). Die Experten befürworten technische Kontrollmechanismen, um die Gefahr der systematischen Diskriminierung, wie im o. g. Beispiel von *Microsoft Tay*, abzuwehren (E1, E2, E4, E6). E6 schlägt vor, durch technische Barrieren das Einspeisen vulgärer und diskriminierender Inhalte zu verhindern. E5 regt an, in den Datensätzen zum Anlernen der KI auch bewusst Minderheiten zu berücksichtigen und E4 befürwortet die Etablierung eines Ethikrates, der Unternehmen und ihre Kommunikation stärker kontrollieren soll. E6 sieht das größte Potenzial der Diskriminierungsvermeidung in der Dateneinspeisung, denn ein KI-System sei per se „neutral, das heißt, alle Biases, die wir einspeisen, speisen wir als Menschen ein“, sodass die ethische Hauptverantwortung bei den Menschen liege, die für eine ausreichend diversifizierte Datengrundlage des Trainings einer KI sorgen müssen.

4. Transparenz schaffen und Nutzer aufklären

Zudem ist eine maximale Transparenz gegenüber den (potenziellen) Nutzern erforderlich, um die Akzeptanz von LCs sicherzustellen (Wambsganss et al. 2021). Insb. gilt es, im Sinne von *Explainable AI (Erklärbarer KI)* das Vertrauen der Nutzer in die neuartige Technologie zu gewinnen, indem diesen die Hintergründe der KI-Algorithmen sowie der zugrundeliegenden anschaulich erklärt werden (Ruane et al. 2019; Wambsganss et al. 2021), z. B. durch Aufzeigen von Potenzialen und Mehrwerten (E5, E6). Diese Transparenz zeigt sich zum einen in der Gestaltung des LCs selbst, zum anderen aber auch in Hinblick auf die Gestaltung der Lernszenarien, in denen LCs eingesetzt werden. Wie E5 ausführte, ist es von entscheidender Bedeutung, dass Lernende den Empfehlungen und Instruktionen, welche sie von dem Companion erhalten, vertrauen. Dies gelingt dadurch, dass LCs auf Grundlage einer ganzheitlichen Analyse der Lernendendaten entscheiden, wobei Lernenden dieser maschinelle Lernprozess transparent dargestellt werden sollte. Dies inkludiert, eindeutig aufzuzeigen, was die KI über den Nutzer weiß und auf welchen Grundlagen sie zu bestimmten Entscheidungen gelangt ist (E2, E5). Da ein solcher maschineller Lernprozess nur durch die Sammlung und gezielte Auswertungen von Lernendendaten möglich ist, gilt auch hier: Es sollten Mehrwerte aufgezeigt werden, anstatt Ängste zu schüren (E6). Lernende sollten über die Potenziale der Datenanalyse aufgeklärt werden, ohne potenzielle Risiken zu verschweigen. Der faire Umgang mit den Daten und das Aufzeigen dessen, z. B. durch Vergabe von Fairness Labels oder durch die Kontrolle unabhängiger Instanzen, trägt zu einem solchen Vertrauensgewinn bei (E5). Transparenz bewegt sich zudem auch auf Ebene der Kommunikation nach außen: Lernende sollten zu jeder Zeit bewusst wahrnehmen, dass es sich um ein digitales Tool handelt, welches sie in einer symbiotischen Mensch-Maschine-Beziehung begleitet. Dies ist wichtig, um der o. g. Gefahr, dass Lernende sich in digitalen Welten „verlieren“ können, entgegenzuwirken (E1, E3, E5).

5. Bedürfnisse und Lernendendaten ganzheitlich erfassen

Eine besondere Chance hinter dem Einsatz KI-basierter LCs ist die durch die Auswertung von Lernenden- und Interaktionsdaten ermöglichte Adaption an die Nutzer (E1, E2, E3, E5). So offeriert der Einsatz von KI die Chance, individuelle Lernpräferenzen zu erfassen und daran angepasste Lehre zu ermöglichen (Sharef et al. 2020; Filho et al. 2021). Wenn die Bedürfnisse der Lernenden allerdings nicht ganzheitlich erfasst werden, könnte diese vermeintliche Individualisierung nach E5 ein „zweischneidiges Schwert“ sein. Viele auf KI basierende adaptive Lern-Systeme nutzen bspw. Klassifikationen, um Lernende in gewisse Lerntypen oder -stile einzuteilen und ihnen zu ihrem Lerntyp oder -stil passende Lerninhalte oder Empfehlungen vorzuschlagen (Adel et al. 2017). Da Klassifikationen Input-Daten komprimieren, besteht die Gefahr, dass Individualität als „Datenausreißer“ wahrgenommen wird, und Lernende im Sinne eines „Schubladen-Denkens“ in Kategorien eingeordnet werden, die nicht ihren wirklichen Bedürfnissen entsprechen. Gerade dieses „Schubladen-Denken“ sollte verhindert werden, sodass die Nutzung von KI zu einer Pluralität an Lernmöglichkeiten führt, anstatt Lernenden auf Grundlage undurchschaubarer Entscheidungen einer KI lediglich beschränkte Lernpfade vorzuschreiben. Die ganzheitliche Erfassung und Interpretation der Lernendendaten sowie die Rückspiegelung dieser Interpretation an die Anwender sind erforderlich, um zuvor genanntes Vertrauen zu gewinnen und Transparenz sicherzustellen (E5). Auch auf Forschungsseite ist es nach E1 und E5 notwendig, eine wissenschaftliche Grundlage für die Analyse von Lerndaten sowie adaptivem Lernen auf Basis fundierter Lern-Theorien zu schaffen. Forscher sollten hierzu zukünftig die Gestaltung intelligenter Lern-Systeme tiefergehend untersuchen, damit ihr Einsatz in der Praxis zu einer tatsächlichen Verbesserung des Lernerfolgs beiträgt (Hobert und Meyer von Wolff 2019).

6. Zugänglichkeit für alle schaffen

Die Enquete-Kommission für den ethischen Umgang mit KI des deutschen Bundestags hat bei einer Untersuchung von 84 KI-Ethik-Dokumenten eine Rangliste der relevantesten ethischen Prinzipien im Kontext von KI, gemessen an der Häufigkeit ihres Vorkommens, erstellt. Hierbei erreichen die *Gerechtigkeit und Fairness durch Zugänglichkeit* mit 68 von 84 möglichen Nennungen den zweithöchsten Stellenwert nach der *Schaffung von Transparenz* mit einer Erwähnung in 73 der 84 untersuchten Guidelines (Enquete-Kommission Künstliche Intelligenz 2020). Auch alle befragten Experten heben hervor, LCs möglichst allen Menschen zugänglich zu machen, um dadurch die Bildungsgerechtigkeit zu erhöhen und die Potenziale von LCs auszuschöpfen. Als Lösungsansatz zur Umsetzung der Barrierefreiheit wurde im Experteninterview bspw. empfohlen, sehbehinderten Menschen LCs durch eine Vorlesefunktion zugänglich zu machen (E2, E4). Darüber hinaus bemängelt E4 die ungerechte Verteilung technischer Geräte und stellt infrage, inwiefern die Kenntnisse im Umgang mit moderner Technologie gerecht ausgeprägt sind. Eine sozialverträgliche Verteilung von Geräten und Schulungen zum Umgang mit Lern-Software sind demnach erforderlich. Zudem besteht ein Interessenskonflikt, einer-

seits eine möglichst flächendeckende Nutzung zu gewährleisten, gleichzeitig jedoch unternehmerische Ziele wie die Gewinnmaximierung nicht außer Acht zu lassen (E3, E4, E5). Deshalb betont E4, dass die Kosten für die Nutzer erschwinglich gehalten werden sollten, bspw. durch eine öffentliche (Co-)Finanzierung.

4 LCs sollten die Lehre begleiten, nicht ersetzen

Letztendlich ist entscheidend, dass LCs Lehrende nicht ersetzen sollen, sondern als Unterstützungs- und Begleitwerkzeug einzusetzen sind, um eine Symbiose aus dem Menschen und (intelligenten) Maschinen zu ermöglichen (E2, E4). Zu der Umsetzung praxistauglicher LCs äußerten mehrere der Befragten Bedenken (E1, E3, E4, E5). So wurde hinterfragt, ob eine vollumfängliche Analyse und Interpretation von Lernendendaten bereits technologisch machbar wären, ohne zu gefährden, dass intelligente Maschinen falsche Entscheidungen treffen (E5). Fehlschlüsse und falsche Entscheidungen könnten gravierende Konsequenzen haben, z. B., falls diese dazu führen, dass Lernende ihr Studium abbrechen (E4, E5). Dies verdeutlicht, dass weiterhin die Lehrkräfte federführend für die Gestaltung von Lehre verantwortlich sein sollten. So ist es, wie E2 ausführt, „Teil der menschlichen Natur, Wissen weiterzugeben“. Darüber hinaus sollten wir uns darüber im Klaren sein, dass wir Menschen – ähnlich wie es in der Debatte über selbstfahrende Autos häufig angepriesen wird – letztendlich immer die endgültige Entscheidung treffen sollten, da wir uns nicht vollständig auf die KI verlassen können. Bildungseinrichtungen und Lehrende dürfen beim Einsatz von LCs ferner nicht zu dem Trugschluss kommen, dass diese eine Möglichkeit darstellen, Kosten und Ressourcen zu sparen, indem auf die eigene Lehre verzichtet wird. Ganz im Gegenteil: Um durch LCs eine höhere Individualisierung und das Aufzeigen individueller Lernpfade zu ermöglichen, wird es notwendig sein, Lerninhalte auf deutlich vielfältigere Weise aufzubereiten, um eine Vielzahl von Lerntypen und -stilen zu befriedigen (E5). Als Leitsatz gilt hier: Unternehmerisches Denken sollte nicht über die inhaltlichen Anforderungen des Einsatzes von LCs sowie über das übergeordnete Ziel, Lernerfolg zu fördern, gestellt werden. Aber nicht nur aus Sicht der Lehrenden ist eine symbiotische Mensch-Maschine-Beziehung entscheidend: Auch Lernende sollten sich bewusstwerden, dass das Lernen mit einem Companion nicht das eigenständige, aktive Lernen ersetzt (E1). Ein Szenario, wie das von Spitzer (2012) beschriebene der „digitalen Demenz“ könnte drohen, wenn Lernende ihre Eigenverantwortung über den Lernprozess an eine KI abgeben. Es muss demnach klar kommuniziert werden, dass eigenständige Lernkompetenzen weiterhin von Studierenden und Schülern zu kultivieren sind (E1).

5 Zusammenfassung und Ausblick

Das Digitalzeitalter befördert das Aufkommen innovativer Lernkonzepte und erfordert die geschickte Nutzung neuartiger technologischer Möglichkeiten. LCs bieten das Potenzial, Bildungsformate neu zu denken und die Mensch-Maschine-Interaktion fachübergreifend fortschrittlich zu gestalten. Schon heute hat der Bundestag

daher eine Zukunftsvision für KI im Bildungssektor auf ethischer Grundlage, eine staatliche Förderung für „Smart Schools“ sowie eine Grundbildung für die Verwendung von KI-Systemen in der Lehre entworfen (Enquete-Kommission Künstliche Intelligenz 2020). LCs stellen eine zukunftssträchtige Möglichkeit dar, um Studierende und Schüler in ihrem Lernprozess zu begleiten. Auch aus ethischer Perspektive bestehen Potenziale für den Einsatz von LCs, so könnte eine individuelle und passgenaue Förderung von Lernenden erreicht werden. Dennoch birgt dies auch mehrere Herausforderungen, bspw. in Bezug auf die Gewährleistung von Chancengerechtigkeit, die Vermeidung von Diskriminierung oder den verantwortungsvollen Umgang mit Lernendendaten. Ethik-Guidelines für LCs könnten grundsätzlich dazu beitragen, diese Herausforderungen zu bewerkstelligen und die Potenziale des Einsatzes von KI in der Lehre zu entfalten. Der Beitrag offenbart jedoch, dass Guidelines nicht ausschließlich von übergeordneten Instanzen entwickelt werden sollten, sondern dass Nutzer aktiv in ihre Gestaltung eingebunden werden müssen. Der Artikel zeigt insgesamt sieben Handlungsempfehlungen für die Gestaltung von LCs aus ethischer Sicht auf, welche sich aus sechs Interviews mit Experten verschiedener Disziplinen ergeben. Es sollte hervorgehoben werden, dass diese Empfehlungen dynamisch anpassbar sind und „lediglich“ eine Momentaufnahme aufzeigen. Zukünftige Forschungsvorhaben könnten außerdem die Praxistauglichkeit der aufgezeigten Empfehlungen untersuchen. Darüber hinaus zeigen die Ergebnisse Spannungsfelder auf, welche Ansatzpunkte für eine nähere Untersuchung in weiteren Forschungsarbeiten liefern. So konnte z. B. keine abschließende Antwort auf die Frage gefunden

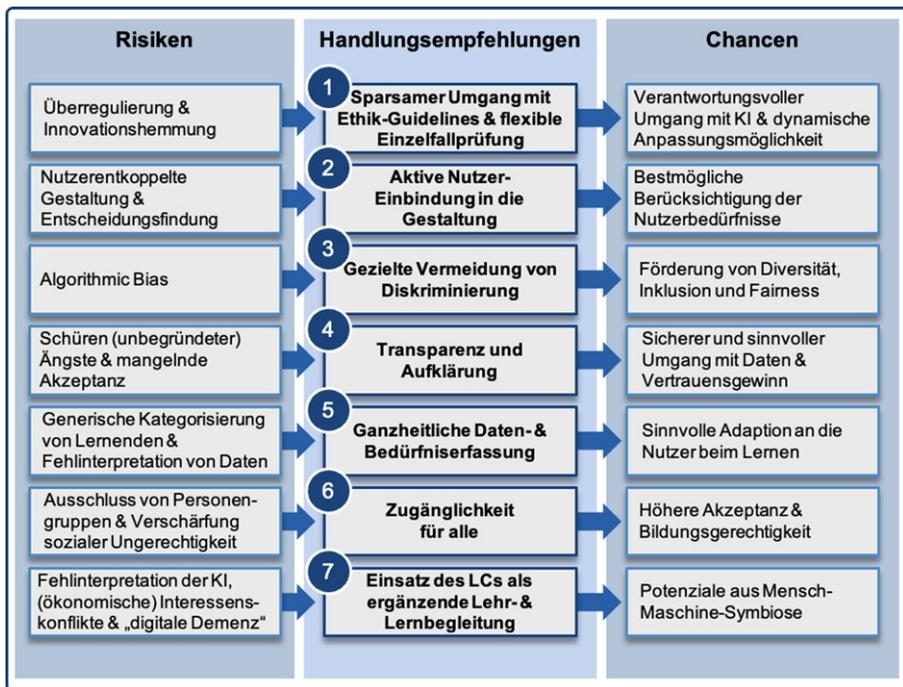


Abb. 1 Risiken, Handlungsempfehlungen und Chancen für den Einsatz von LCs in der Lehre

werden, inwiefern der Einsatz von LCs die soziale Gerechtigkeit beeinflusst. Abb. 1 fasst die diskutierten Risiken sowie die daraus abgeleiteten Handlungsempfehlungen und Chancen von LCs aus ethischer Perspektive komprimiert zusammen, indem sie die sieben diskutierten Handlungsempfehlungen aus Abschn. 3.2 illustrativ ihren jeweiligen Risiken und Chancen zuordnet.

Die in diesem Beitrag herausgearbeiteten Erkenntnissen bieten Praktikern mit einem Bezug zu neuartigen Bildungstechnologien (z. B. Lehrenden, Software-Entwickler, etc.) Anhaltspunkte sowie Empfehlungen zum Design und Einsatz von LCs. Dennoch ergeben sich auch Limitationen, beispielsweise durch die kleine Stichprobengröße der befragten Experten, die Beschränkung auf den deutschsprachigen Raum ohne Berücksichtigung kultureller oder rechtlicher Unterschiede auf internationaler Ebene und die eigene Schwerpunktsetzung bei der Ausgestaltung der Befragung. Daher erscheint es notwendig, die empirische Grundlage weiter auszubauen und zukünftige Szenarien und Schwerpunktthemen für den verantwortungsvollen Einsatz von KI und Machine Learning zu Lehr- und Lernzwecken weiter zu untersuchen. Dieser Artikel leistet einen ersten Beitrag dazu, bei der Gestaltung von LCs stärker auf ethische Gesichtspunkte einzugehen und offeriert Chancen, um in Folgestudien tiefergehende Erkenntnisse zu erlangen.

Förderung Dieser Beitrag ist Teil des Projekts StuBu, gefördert durch das Bundesministerium für Bildung und Forschung (BMBF); Förderkennzeichen 21INVI06.

Funding Open Access funding enabled and organized by Projekt DEAL.

Open Access Dieser Artikel wird unter der Creative Commons Namensnennung 4.0 International Lizenz veröffentlicht, welche die Nutzung, Vervielfältigung, Bearbeitung, Verbreitung und Wiedergabe in jeglichem Medium und Format erlaubt, sofern Sie den/die ursprünglichen Autor(en) und die Quelle ordnungsgemäß nennen, einen Link zur Creative Commons Lizenz beifügen und angeben, ob Änderungen vorgenommen wurden.

Die in diesem Artikel enthaltenen Bilder und sonstiges Drittmaterial unterliegen ebenfalls der genannten Creative Commons Lizenz, sofern sich aus der 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.

Weitere Details zur Lizenz entnehmen Sie bitte der Lizenzinformation auf <http://creativecommons.org/licenses/by/4.0/deed.de>.

Literatur

- Adel N, Latham A, Crockett KA (2017) Towards socially intelligent automated tutors: predicting learning style dimensions from conversational dialogue, S 315–320
- Bendel O (2003) Pädagogische Agenten im Corporate E-Learning. Dissertation
- Bendel O (2021) Soziale Roboter in der Moral: Ethische Betrachtungen am Beispiel des Unterrichts. In: Bendel O (Hrsg) Soziale Roboter. Springer, Wiesbaden, S 149–167
- Brüsemeyer T (2008) Qualitative Forschung. VS, Wiesbaden
- Casas-Roma J, Conesa J (2021) Towards the design of ethically-aware pedagogical conversational agents. In: Barolli L, Takizawa M, Yoshihisa T et al (Hrsg) Advances on P2P, parallel, grid, cloud and Internet computing. Springer, Cham, S 188–198

- Diederich S, Brendel A, Morana S, Kolbe L (2022) On the design of and interaction with conversational agents: an organizing and assessing review of human-computer interaction research. *J Assoc Inf Syst* 23(1):96–138. <https://doi.org/10.17705/1jais.00724>
- Enquete-Kommission Künstliche Intelligenz (2020) Unterrichtung der Enquete-Kommission Künstliche Intelligenz – Gesellschaftliche Verantwortung und wirtschaftliche, soziale und ökologische Potenziale (Bericht der Enquete-Kommission Künstliche Intelligenz – Gesellschaftliche Verantwortung und wirtschaftliche, soziale und ökologische Potenziale)
- European Commission (2021) Proposal for a regulation of the European parliament and of the council laying down harmonised rules on artificial intelligence (artificial intelligence ACT) and amending certain union legislative acts
- Filho AH, Viecelli K, do Prado HA et al (2021) An adaptive and proactive interface agent for Interactivity and decision-making improvement in a collaborative virtual learning environment. Springer, Heidelberg, Berlin, New York
- Gubareva R, Lopes RP (2020) Virtual assistants for learning: A systematic literature review, S 97–103
- Hobert S, Meyer von Wolff R (2019) Say hello to your new automated tutor—A structured literature review on pedagogical conversational agents. In: *Wirtschaftsinformatik 2019 Proceedings*
- Hochrangige Expertengruppe für künstliche Intelligenz (2019) Europäische Kommission, Generaldirektion Kommunikationsnetze, Inhalte und Technologien, (2019). Ethik-leitlinien für eine vertrauenswürdige KI
- Karrenbauer C, König CM, Breiter MH (2021) Individual digital study assistant for higher education institutions: status quo analysis and further research agenda. In: *Proceedings 16th International Conference on Wirtschaftsinformatik*
- Kim Y, Baylor AL, PALS Group (2006) Pedagogical agents as learning companions: the role of agent competency and type of interaction. *Educ Tech Res Dev* 54:223–243. <https://doi.org/10.1007/s11423-006-8805-z>
- Krämer NC, Eimler S, von der Pütten Payr A, Payr S (2011) Theory of companions: what can theoretical models contribute to applications and understanding of human-robot interaction? *Appl Artif Intell* 25:474–502. <https://doi.org/10.1080/08839514.2011.587153>
- Lehmann K, Oeste S, Janson A et al (2015) Flipping the Classroom—IT-unterstützte Lerneraktivierung zur Verbesserung des Lernerfolges einer universitären Massenlehrveranstaltung. *HMD* 52:81–95. <https://doi.org/10.1365/s40702-014-0102-4>
- Luchterhandt O (2021) Individuelle und kollektive Verantwortung. Reichweiten und Rechtsfolgen. In: Haux R, Gahl K, Jipp M et al (Hrsg) *Zusammenwirken von natürlicher und künstlicher Intelligenz*. Springer, Wiesbaden, S 121–150
- Mori M (1970) The uncanny valley. *Energy* 7:33–35
- Orwat C (2019) Diskriminierungsrisiken durch Verwendung von Algorithmen: Eine Studie erstellt mit einer Zuwendung der Antidiskriminierungsstelle des Bundes. *Nomos*, Baden-Baden
- Rawlins WK (2017) *Friendship matters communication, dialectics, and the life course*. Routledge, London
- Ruane E, Birhane A, Ventresque A (2019) Conversational AI: social and ethical considerations. In: *Proceedings for the 27th AIAI Irish Conference on Artificial Intelligence and Cognitive Science*. Galway, Ireland, S 104–115
- Sharef NM, Azmi Murad MA, Mansor EI et al (2020) Learning-analytics based intelligent simulator for personalised learning. In: *Int. Conf. Adv. Data Sci., E-Learn. Inf. Syst., ICADEIS*. Institute of Electrical and Electronics Engineers Inc
- Spitzer M (2012) *Digitale Demenz: wie wir uns und unsere Kinder um den Verstand bringen*. Droemer, München
- Strohmann T (2021) *From assistance to companionship – designing virtual companions* (Universitätsbibliothek Braunschweig)
- Stucki T, D’Onofrio S, Portmann E (2020) Chatbots gestalten mit Praxisbeispielen der Schweizerischen Post: HMD Best Paper Award 2018. Springer, Wiesbaden
- Wambsgans T, Höch A, Zierau N, Söllne M (2021) Ethical design of conversational agents: towards principles for a value-sensitive design. In: *Wirtschaftsinformatik 2021 Proceedings*
- Wellnhammer N, Dolata M, Steigler S, Schwabe G (2020) Studying with the help of digital tutors: design aspects of conversational agents that influence the learning process
- Winkler R, Söllner M, Neuweiler ML et al (2019) Alexa, can you help us solve this problem?: How conversations with smart personal assistant tutors increase task group outcomes. In: *Extended Abstracts of the 2019 CHI Conference on Human Factors in Computing Systems*. ACM, Glasgow Scotland U, S 1–6