

# KI als Teammitglied: Wie Erwartungen die Zusammenarbeit prägen

Von **Sophie Kerstan** (ETH Zürich), **Prof. Dr. Eleni Georganta** (Universität von Amsterdam) und **Prof. Dr. Anna-Sophie Ulfert-Blank** (Technische Universität Eindhoven)

**O**b Bürosachbearbeitungs-Humanoid Lena oder Barkeeper-Roboter Sam – die Praxisbeispiele für den Einzug von künstlicher Intelligenz (KI) in die Arbeitswelt mehren sich. Dieser Trend besteht bereits seit einigen Jahren. Neu ist jedoch, dass KI zunehmend als Teammitglied positioniert wird, das eine auf seine Fähigkeiten abgestimmte Rolle ausfüllt und dabei mit Mitarbeitenden interagiert (O’Neill et al., 2020). Dabei geht die Konzeption von KI als Teammitglied über die bisher vorherrschende Anwendung von KI als Assistenzsystem hinaus. KI-Assistenzsysteme unterstützen klar definierte Teilaspekte einer Aufgabe, indem sie bspw. Informationen liefern oder Entscheidungsvorschläge generieren. Die endgültige Entscheidungs- und Handlungsbe-fähigung liegt jedoch beim Menschen. KI-Assistenzsysteme verfügen demnach über wenig bis keine Autonomie in der Ausführung einer gegebenen Aufgabe. Durch das stark begrenzte

Fähigkeitsspektrum und den eingeschränkten Handlungsspielraum der KI sind die Interaktionen zwischen Menschen und KI in diesen Fällen stark repetitiv. Als Teammitglied wird KI hingegen ein erweiterter Handlungsspielraum zugeschrieben, sodass der Grad der Autonomie steigt und KI eigene Entscheidungen treffen und Handlungen ausführen kann. Auf Basis dessen können KI-Teammitglieder ihre Verhaltensweisen flexibler an Veränderungen innerhalb der Aufgabe oder hinsichtlich der Rahmenbedingungen anpassen. Zudem sind Aufgabenausführung und Zielerreichung von Menschen und KI stärker miteinander verbunden. KI trägt dabei indirekt (durch Interaktionen mit menschlichen Mitarbeitenden) sowie unmittelbar (durch Handlungen oder Entscheidungen) zum Arbeitsergebnis bei. Die aufgezeigten Unterschiede zwischen KI-Assistenzsystemen und KI-Teammitgliedern sind in Abbildung 1 zusammengefasst.

**Abb. 1: Darstellung der Unterschiede zwischen dem Einsatz von KI als Assistenzsystem und KI als Teammitglied**

DIMENSION DER UNTERSCHIEDUNG	KI als Assistenzsystem	KI als Teammitglied
<b>Aufgabenspektrum</b>	KI übernimmt klar definierte, eng begrenzte, repetitive Teilaufgaben.	KI kann Aufgaben(-pakete) übernehmen, die umfassend sind und Anpassungen unterliegen können.
<b>Interaktion zwischen Mensch und KI</b>	KI unterstützt bei der Bearbeitung einer Aufgabe. Über das Arbeitsergebnis bestimmen jedoch die Entscheidungen und Handlungen des Menschen. Die Interaktionsmuster zwischen KI und Mensch sind weitestgehend repetitiv und vorgeschrieben.	KI und Mensch arbeiten in gegenseitiger Abhängigkeit an einem gemeinsamen Ziel. Das Arbeitsergebnis ist durch die Handlungen der KI mitbestimmt. Die Interaktionsmuster zwischen KI und Mensch werden von dynamischer Koordination geprägt.
<b>Autonomie und Proaktivität der KI</b>	KI wird von Menschen kontrolliert (z. B. müssen Menschen Handlungsvorschläge der KI bewilligen). Handlungen der KI sind dementsprechend im Sinne ihrer Proaktivität stark eingeschränkt.	KI besitzt (in Teilen) Autonomie über eigene Handlungen (z. B. kann KI selbstgesteuert Entscheidungen treffen). Dies befähigt KI zu proaktivem Verhalten (z. B. unaufgeforderte Ausführung einer Handlung).
<b>Vorhersagbarkeit KI-basierter Handlungen</b>	Handlungsschritte der KI können von menschlichen Teammitgliedern in der Regel passend vorhergesagt werden.	Aufgrund von gesteigerter Flexibilität und Proaktivität sind Vorhersagen zu den Handlungsschritten der KI von Unsicherheit geprägt.

Quelle: Eigene Darstellung

## ABSTRACT

**Forschungsfrage:** Künstlicher Intelligenz (KI) wird zunehmend die Rolle eines Teammitglieds zugeschrieben. Deshalb untersuchen wir, welche Erwartungen Menschen an KI-Teammitglieder stellen und skizzieren, wie sich diese Erwartungen auf die Mensch-KI-Zusammenarbeit auswirken.

**Methodik:** Wir fassen den Forschungsstand zusammen und leiten sieben Beobachtungen ab.

**Praktische Implikationen:** KI-bezogene Erwartungen sollten in die KI-Entwicklung, Implementierung und fortlaufende Mensch-KI-Zusammenarbeit einbezogen werden. Wir zeigen Möglichkeiten auf, dieses Erwartungsmanagement in der Praxis umzusetzen.

Auch steht der Einsatz von KI als Teammitglied dem Konzept, Mitarbeitende durch KI ersetzen zu wollen, entgegen, da er darauf beruht, die Fähigkeiten und Stärken von Menschen und KI so zusammenzubringen, dass sie sich gegenseitig ergänzen. Diese Herangehensweise ist u. a. darin begründet, dass für die überwiegende Zahl der bestehenden und voraussehbaren KI-Anwendungen eine Besetzung von auf Menschen zugeschnittene Arbeitsrollen durch KI schlicht nicht möglich ist. Spätestens während der Implementierung von KI wird meist deutlich, dass Arbeitsprozesse durch KI eher transformiert werden, als dass bestehende Posten gänzlich von KI eingenommen würden. Entsprechend zielt der Ansatz, KI als Teammitglied zu integrieren, darauf ab, die bestehenden Fähigkeiten der menschlichen Teammitglieder zu bereichern, um mittels der so entstehenden Mensch-KI-Teams Leistungssteigerungen zu erwirken (Seeber et al., 2020).

Neben diesen Überlegungen befeuert auch der rasante technologische Fortschritt den Diskurs über den Einsatz von KI als Teammitglied. Laufend werden Studien veröffentlicht, die über neue Durchbrüche, bspw. im Bereich von selbstlernenden Algorithmen oder KI-basierter Spracherzeugung und -verarbeitung, berichten. Erfolge wie diese ermöglichen die Entwicklung von KI-Systemen, die sich an sich verändernde Umgebungsbedingungen und Aufgaben anpassen, zu proaktivem Handeln in der Lage sind und mit Mitarbeitenden flexibel kommunizieren können. Aus Perspektive der Teamforschung decken diese Fähigkeiten viele der Kerneigenschaften ab, die für die Zusammenarbeit im Team essenziell sind (Lyons et al., 2019). Es sind jedoch nicht die wachsenden technologischen Möglichkeiten allein, wegen derer es für Unternehmen von Interesse sein sollte, sich mit dem Einsatz von KI als Teammitglied zu beschäftigen. Studienergebnisse zeigen, dass Personen, die KI nicht nur als Tool, sondern als Teammitglied wahrnehmen, eine positivere affektive Einstellung gegenüber KI haben und über geteilten Aufgaben besser mit ihr kommunizieren und sie koordinieren (O'Neill et al., 2020). In KI ein Teammitglied zu sehen, kann demnach leistungsbezogene Vorteile erbringen.

Obwohl sich die Forschung zu Mensch-KI-Teams derzeit erst im Aufbau befindet, sind die Faktoren, die dazu beitragen, dass Menschen KI als Teammitglied wahrnehmen, vielfältig. Die Be-

fundlage zeigt insbesondere, dass Menschen eine Reihe von Erwartungen (z. B. hinsichtlich fehlerloser Leistung, menschenähnlichen Kommunikationsfähigkeiten und der Fähigkeit zur Entwicklung eines geteilten Verständnisses der gemeinsamen Aufgabe) gegenüber KI-Teammitgliedern haben (Zhang et al., 2021). Wissen und Verständnis über diese Erwartungshaltungen und ihre Auswirkungen auf die Zusammenarbeit zwischen Mitarbeitenden und KI sind ein fundamentaler Baustein für eine sichere und effektive Mensch-KI-Zusammenarbeit (O'Neill et al., 2020; Lyons et al., 2019). Ziel dieses Artikels ist deshalb, den derzeitigen Forschungsstand innerhalb dieses Themengebiets zusammenzufassen und die Auswirkungen KI-bezogener Erwartungen auf die Zusammenarbeit in Mensch-KI-Teams zu beleuchten. Neben der Literatur zu Mensch-KI-Teams ziehen wir dazu auch Erkenntnisse aus der psychologischen Literatur zur Erwartungsbestätigung und -widerlegung sowie der Forschung zu Interaktionen in menschlichen Teams heran. Die gesammelten Erkenntnisse fassen wir in sieben Beobachtungen zusammen. Zudem erarbeiten wir Handlungsempfehlungen für eine aktive Einbindung von KI-bezogenen Erwartungen in die Entwicklung und Implementierung von KI-Teammitgliedern und die fortlaufende Ausgestaltung der Mensch-KI-Zusammenarbeit.

## Welche Erwartungen haben Mitarbeitende gegenüber KI-Teammitgliedern?

Erwartungen können als Annahmen über in der Zukunft liegende Geschehnisse und Handlungen definiert werden, die sich auf Situationen, Personen, technologische Systeme oder Institutionen beziehen können. Diese Erwartungen werden oftmals als antizipatorisch betitelt, da sie der Frage „Was wird geschehen?“ nachgehen. Darüber hinaus setzen sich Erwartungen jedoch auch mit der Frage „Was sollte geschehen?“ auseinander. In diesem Fall beschreiben Erwartungen keine Annahme (antizipatorische Erwartung), sondern einen Anspruch, der an die Zukunft gestellt wird. Sie werden deshalb auch als normativ beschrieben. Antizipatorische Erwartungen können sich von normativen Erwartungen unterscheiden. So zeigen z. B. Experimentalstudien, dass Menschen gegenüber KI-Teammitgliedern die antizipatorische Erwartung haben, sie würden eingeschränktere Kommunikationsfähigkeiten als

Menschen besitzen (z. B. McNeese et al., 2018). Aus qualitativen Studien wird allerdings ersichtlich, dass an ein ideales KI-Teammitglied oftmals der Anspruch gestellt wird, es sollte über ein möglichst menschenähnliches Sprachverständnis und Kommunikationsverhalten verfügen (z. B. Zhang et al., 2021). Daraus wird deutlich, dass normative Erwartungen nicht automatisch in antizipatorische übersetzt werden. Die Differenzierung dieser Erwartungstypen ist wichtig, um zu verstehen, was passiert, wenn Erwartungen auf die (technologische) Realität treffen. Es lässt sich festhalten:

**Beobachtung 1:** Erwartungen an KI als Annahme (antizipatorische Erwartungen) unterscheiden sich von idealisierten Erwartungen an KI in Form von Ansprüchen (normativen Erwartungen).

Neben den Unterschieden zwischen Erwartungstypen können Erwartungen an KI in sich widersprüchlich sein. Bspw. wird einerseits diskutiert, dass KI, um der Rolle eines vollwertigen Teammitglieds gerecht zu werden, die Fähigkeit besitzen muss, sich eigenständig und flexibel an Veränderungen anzupassen und eigene Arbeitspakete zu übernehmen (Lyons et al., 2019). Somit wäre KI den Mitarbeitenden nicht mehr kategorisch unterstellt, sondern würde über vergleichbare Handlungs- und Entscheidungsfreiheit verfügen. Andererseits wird aus ethischen und sicherheitsbezogenen Gesichtspunkten gefordert, dass Menschen zu jedem Zeitpunkt die Handlungen einer KI überwachen, kontrollieren und wenn nötig korrigieren können sollten (Rieth/Hagemann, 2022). Aus den beschriebenen Anforderungen kann ein Paradox aus Flexibilitäts- und Kontrollerwartungen an KI-Teammitglieder entstehen.

Als ein weiteres Beispiel für einen solchen Widerspruch lässt sich die in einigen Studien dokumentierte Erwartung nennen, dass KI als Teammitglied Emotionen erkennen können sollte (Rieth/Hagemann, 2022). Mit dieser Fähigkeit könnte KI bspw. Stresslevel in Echtzeit ermitteln und so menschliche Teammitglieder in besonders anspruchsvollen Situationen besser entlasten. Im Rahmen von Affective Computing, einem Ansatz, bei dem mittels KI-basierter Verfahren angestrebt wird, Personendaten wie Gesichtsausdrücke oder Stimmlagen zu analysieren, um daraus auf den emotionalen Zustand eines Menschen zu schließen, scheint dies durchaus denkbar. Allerdings steht diesem Anspruch entgegen, dass menschliche Interaktionspartner vielfach nicht wollen, dass KI personenbezogene Daten dieser Art erfasst, verarbeitet oder gar speichert. Auch aus dieser Spannung wird deutlich:

**Beobachtung 2:** Erwartungen an KI-Teammitglieder können Widersprüche enthalten.

Des Weiteren stellt sich die Frage, inwiefern Erwartungen an KI-Teammitglieder mit Erwartungen, die menschlichen Teammitgliedern gegenüber gelten, übereinstimmen. Studien zeigen diesbezüglich, dass Fähigkeiten, die für Teamarbeit in klassischen menschlichen Teams essenziell sind, auch von

KI-Teammitgliedern erwartet werden (Lyons et al., 2019). Zu den am häufigsten erwarteten Fähigkeiten zählt bspw. eigenständiges und proaktives Handeln der KI (Zhang et al., 2021). Außerdem wird zunehmend von der Erwartung berichtet, dass KI-Teammitglieder ein mit menschlichen Teammitgliedern geteiltes Verständnis von Arbeitszielen, -schritten und Kooperationsweisen formen können sollten (Andrews et al., 2022).

Allerdings zeigt sich auch, dass Erwartungen an KI von Erwartungen an menschliche Teammitglieder in einigen Aspekten abweichen. So wird KI-Teammitgliedern gegenüber teils weniger bis keine Fehlertoleranz eingeräumt. Studien untersuchen dieses Phänomen bspw. im Rahmen eines kognitiven Schemas der perfekten Automatisierung, dem Glaubenssätze wie „Hochentwickelte Technologien treffen immer die richtigen Entscheidungen“ zugrunde liegen (Merritt et al., 2015). Während in diesem Punkt höhere Erwartungen an KI als an Menschen gestellt werden, können die Erwartungen an KI in anderen Bereichen niedriger ausfallen. An diesem Punkt wird erneut die Differenzierung zwischen normativen und antizipatorischen Erwartungen wichtig. Normativ wird von KI-, wie von menschlichen Teammitgliedern, eine hohe Kommunikationsfähigkeit gefordert. Antizipatorisch gilt diese Erwartung aber für KI-Teammitglieder bedeutend weniger als für menschliche. Zusammenfassend lässt sich festhalten:

**Beobachtung 3:** Erwartungen an KI-Teammitglieder zeigen inhaltliche Schnittmengen mit Erwartungen, die menschlichen Teammitgliedern gegenüber bestehen, sie sind jedoch nicht deckungsgleich.

### Welche Auswirkungen haben Erwartungen gegenüber KI-Teammitgliedern auf Arbeitsprozesse und -leistung?

Um die Auswirkungen von KI-bezogenen Erwartungen darlegen zu können, ist es notwendig, zuerst logisch festzustellen, welche Szenarien sich ergeben, wenn initiale Erwartungen innerhalb einer Interaktion mit einem KI-Teammitglied auf die Realität treffen. Ganz grundsätzlich fällt dabei entweder eine Bestätigung oder eine Widerlegung dieser initialen Erwartungen an. Darüber hinaus können sich Erwartungen jedoch auch in ihrer Valenz (positiv/hoch vs. negativ/gering) unterscheiden und jeweils unter- oder übertroffen werden. Bei initial hohen Erwartungen an KI-Teammitglieder, wie dem zuvor beschriebenen Anspruch auf Perfektionismus und Fehlerlosigkeit, ist es sehr wahrscheinlich, dass es zu einer Enttäuschung der Erwartungen in Form einer negativen Erwartungswiderlegung kommt. Das tatsächliche Verhalten des KI-Teammitglieds bleibt in diesem Fall hinter der Ursprungserwartung zurück.

Schon im Jahrzehnt der 2000er-Jahre zeigten Studien zum Gebrauch simpler Informations- und Kommunikationstechnologie (wie E-Mail-Programme), dass diese Form der Erwartungswiderlegung die Einstellung von Mitarbeitenden gegenüber technologischen Systemen nachhaltig negativ prägt und somit ihre

Adaption langfristig erschweren kann. Zusätzlich stellten die Forschenden fest, dass dieser negative Effekt selbst bei einer positiven Erwartungswiderlegung (z. B. niedrige initiale Erwartungen, die durch die Fähigkeiten einer Technologie übertroffen wurden) auftrat (z. B. Bhattacharjee/Premkumar, 2004). Andere Studien berichten hinsichtlich einer positiven Erwartungswiderlegung gegenteilige Ergebnisse. Bspw. konnte in einem Experiment gezeigt werden, dass Studienteilnehmende mit niedrigen antizipatorischen Erwartungen an die kommunikativen Fähigkeiten ihres KI-Interaktionspartners positiv überrascht reagierten, wenn die KI in der Interaktion menschenähnliche kommunikative Fähigkeiten zeigte. Die KI wurde daraufhin von den Teilnehmenden besser bewertet als vor der Interaktion (Grimes/Schuetzler/Giboney, 2021). In der derzeitigen Studienlage sind somit gegenläufige Ergebnisse gegeben. Es scheint, als könne eine Überbietung der initialen Erwartungen in der tatsächlichen Interaktion mit KI sowohl negative als auch positive Konsequenzen haben. Grundsätzlich lässt sich jedoch festhalten:

**Beobachtung 4:** Wenn KI-bezogene Erwartungen auf die Realität treffen, ergeben sich verschiedene Szenarien der Erfüllung oder Widerlegung dieser Erwartungen, die Auswirkungen auf Einstellungen und Verhalten gegenüber KI-Teammitgliedern haben können.

Wenn eine initiale Erwartung einer Person an ein KI-Teammitglied widerlegt wird, könnte vermutet werden, dass die Person diese Erwartung an die Realität anpasst. Bspw. könnte eine Person vor der ersten Interaktion mit ihrem KI-Teammitglied eine fehlerlose Ausführung der gegebenen Aufgaben von der KI erwarten. Für den Fall, dass das KI-Teammitglied jedoch teils Fehler macht, mag auf Basis der Erwartungswiderlegung davon ausgegangen werden, dass sich der Inhalt der initialen Erwartung dem tatsächlichen Leistungsniveau der KI anpasst. Die Person würde somit automatisch lernen zu antizipieren, wann die Handlungen der KI richtig und wann sie fehlerhaft sind. Die psychologische Forschung zeigt allerdings, dass Erwartungsanpassungen oftmals weder automatisch noch unmittelbar geschehen. Vielmehr gehen Menschen mit Informationen oder Erfahrungen, die ihre Erwartungen verletzen, sehr unterschiedlich um. Zu den Strategien im Umgang mit erwartungswiderlegenden Erfahrungen gehören z. B. das Ignorieren unerwarteter Handlungen, das Herunterspielen ihrer Wichtigkeit oder die erhöhte Ausrichtung der Aufmerksamkeit auf Vorkommnisse, die die initialen Erwartungen bestätigen (Pinquart et al., 2021). Im äußersten Fall richten Menschen in Situationen des Zusammenarbeitens ihre eigenen Handlungen so stark nach ihren Erwartungen aus, dass sie im Sinne einer selbsterfüllenden Prophezeiung das vorab erwartete Resultat herbeiführen. Zur Illustration dieses Effekts stelle man sich eine Person vor, die mit einem KI-Teammitglied innerhalb einer beliebigen Aufgabe zusammenarbeitet. Bereits vor der Interaktion hat die Person sehr geringe antizipatorische Erwartungen an die KI. Nun besitzt das KI-Teammitglied Fähig-

keiten, die deutlich über diesen Erwartungen liegen. Die Person blendet diese auf Basis ihrer Ursprungserwartung jedoch aus und lässt das Potenzial der KI in der Aufgabe, an der sie zusammen mit der KI arbeitet, ungenutzt. Daraufhin ist das aus der Zusammenarbeit hervorgehende Arbeitsergebnis mangelhaft, was jedoch vor allem an der ausbleibenden Integration der Fähigkeiten der KI und nicht etwa an einer schlechten Leistung der KI per se liegt.

Erste Studienergebnisse unterstreichen, dass initiale Erwartungen an KI-Teammitglieder ungeachtet der tatsächlichen Realität überdauern und Einfluss auf die Interaktion mit KI nehmen können (z. B. Demir/McNeese/Cooke, 2018). Im Rahmen eines Experiments wurden Teilnehmende in zwei Gruppen unterteilt, wobei den Personen in der einen Gruppe mitgeteilt wurde, sie würden zusammen mit einer KI eine Aufgabe lösen. Der zweiten Gruppe hingegen wurde gesagt, sie würden die Aufgabe mit einem Menschen bearbeiten. In Wahrheit interagierten alle Teilnehmenden mit einer eingeweihten Person, die über die Gruppen hinweg das gleiche Verhalten zeigte. Der einzige Unterschied zwischen den Gruppen war somit die Information über die Art (KI oder Mensch) ihres Teammitglieds. Die Ergebnisse der Studie zeigen, dass die Teams, in denen die Teilnehmenden dachten, sie würden mit einer KI interagieren, weniger Abstimmungen trafen und so auch im Ergebnis der Aufgabe eine schlechtere Leistung erbrachten (Demir/McNeese/Cooke, 2018). Daraus lässt sich ableiten, dass die Information, eine KI oder ein Mensch wäre Teil des Teams, antizipatorische Erwartungen bei den Teilnehmenden aktivierte. Diese Erwartungen bestimmten die Zusammenarbeit in einem solchen Ausmaß, dass die tatsächlichen Fähigkeiten des Interaktionspartners in den Hintergrund traten. In Summe wird deutlich:

**Beobachtung 5:** KI-bezogene Erwartungen können anhaltende Effekte auf den Umgang und die Zusammenarbeit mit KI haben, denn auf die Verletzung einer Erwartung folgt nicht zwangsläufig eine Anpassung dieser Erwartung.

### Wie beeinflussen KI-bezogene Erwartungen mehrerer menschlicher Teammitglieder das Team als Ganzes?

Bisher haben wir uns in unseren Ausführungen auf die Erwartungen eines menschlichen Teammitglieds gegenüber einem KI-Teammitglied bezogen. In Organisationen werden jedoch vielfach mehrere menschliche Teammitglieder mit demselben KI-Teammitglied zusammenarbeiten. Zwischen diesen Teammitgliedern können Unterschiede in KI-bezogenen Erwartungen vorliegen, weil Erwartungen von verschiedenen personenbezogenen Faktoren beeinflusst werden (z. B. Kaplan et al., 2021). Besonders gut untersucht sind solche Unterschiede in Bezug auf Vertrauen in KI. Grundsätzlich beruht Vertrauen auf der Annahme, dass ein Gegenüber etwas, das für die vertrauende Person wichtig ist, tun wird. Somit kann Vertrauen als eine Form der positiven Erwartungshaltung angesehen wer-

den. Unterschiede im Ausmaß, in dem Personen KI-Systemen vertrauen, beruhen bspw. auf der individuellen Erfahrung im Umgang mit Technologien oder der eigenen Expertise in Bezug auf die Aufgabe, die mit der KI zusammen bearbeitet wird.

In Summe zeigen Studienergebnisse diesbezüglich, dass mehr Erfahrung und eine höhere Expertise innerhalb des Aufgabenbereichs eine positive Erwartungshaltung gegenüber KI fördern. Auch soziodemografische Aspekte wie die Kultur, in der Menschen aufwachsen und leben, können bedingen, wie sehr sie KI vertrauen (Kaplan et al., 2021). Somit muss festgehalten werden:

**Beobachtung 6:** Da Erwartungen an KI-Teammitglieder auf Basis personenbezogener Faktoren entstehen, können sich Erwartungen an das KI-Teammitglied zwischen menschlichen Teammitgliedern unterscheiden.

Die Teamforschung beschäftigt sich im Rahmen von empirischen Studien zu sog. mentalen Modellen bereits seit mehreren Jahrzehnten mit der Frage, welchen Einfluss Erwartungen verschiedener Teammitglieder auf die Zusammenarbeit in menschlichen Teams haben. Mentale Modelle eines Teams sind kognitive Strukturen der Teammitglieder, in denen Vorstellungen dazu verankert sind, wie Aufgaben am besten erledigt

werden, wie andere Teammitglieder sich verhalten und wie man zusammenarbeiten möchte. Diese Glaubens- und Wissensstrukturen werden als erwartungsbestimmend und somit rahmengebend für Interaktionen im Team angesehen (Andrews et al., 2022). Bei der Bewertung der Güte mentaler Modelle in Teams wird meist zwischen zwei Kriterien unterschieden. Das eine Kriterium betrifft den Grad der Gleichartigkeit oder Überlappung der mentalen Modelle der einzelnen Teammitglieder. Das zweite zielt auf die Akkuratess mentaler Modelle ab, z. B. in Form ihrer Passung zur Realität oder zu einer (falls gegeben) idealen Ausführung der Aufgabe.

Empirische Befunde zeigen, dass beide Faktoren maßgeblich zu einer effektiven und leistungsfördernden Zusammenarbeit beitragen (Andrews et al., 2022). Teams sollten demnach über mentale Modelle verfügen, die sich über die Mitglieder hinweg stark überschneiden und korrekt sind, um auf Basis einer geteilten Erwartungshaltung produktiv zusammenarbeiten zu können. Auf Grundlage dessen kann für Mensch-KI Interaktionen erfasst werden:

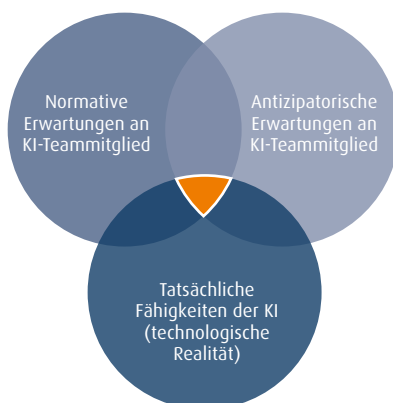
**Beobachtung 7:** Zwischen Teammitgliedern geteilte, realitätsnahe KI-bezogene Erwartungen können die Zusammenarbeit in Mensch-KI-Teams fördern.

**Abb. 2: Schematische Darstellung zur aktiven Herstellung einer Passung zwischen Erwartungen und Realität bezüglich der Fähigkeiten eines KI-Teammitglieds**

Ausgangszustand nur geringer Passung zwischen KI-bezogenen Erwartungen und den Fähigkeiten einer bestimmten KI\*

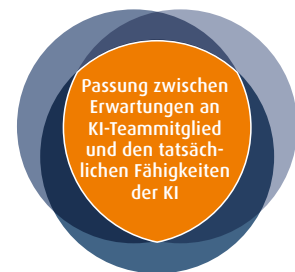
Einbezug der Erwartungen und aktives Erwartungsmanagement in verschiedenen Phasen der KI-Gestaltung und Adoption

Gesteigerte Passung KI-bezogener Erwartungen und Fähigkeiten der KI als Basis effektiver Zusammenarbeit in Mensch-KI-Teams



Einbindung und Anpassung KI-bezogener Erwartungen in:

- (1) KI-Entwicklung und -Design
- (2) Implementierung
- (3) fortlaufender Zusammenarbeit



\*Anmerkung: Jede andere Ordnung der dargestellten Kreise zueinander im Ausgangszustand ist ebenso denkbar. Sollte das Team, in dem die KI implementiert wird, aus weiteren menschlichen Teammitgliedern bestehen, steigt die Komplexität des Ausgangszustands durch die Hinzunahme ihrer Erwartungen, die durch weitere Kreise in der Abbildung dargestellt werden würden, automatisch.



## Implikationen und Handlungsempfehlungen

Die genannten Beobachtungen zeigen, dass Diskrepanzen zwischen Erwartungen an KI-Teammitglieder und den tatsächlichen Fähigkeiten von KI erkannt und aktiv adressiert werden sollten. Für dieses Vorhaben bestehen bereits in der Entwicklung und dem Design eines KI-Teammitglieds Ansatzpunkte. Darüber hinaus sollten die Auswirkungen des Wechselspiels zwischen Erwartungen und Realität auch während der Implementierung von Mensch-KI-Teams und der fortlaufenden Zusammenarbeit berücksichtigt und angepasst werden. Wie mittels eingebrachter Maßnahmen eine möglichst hohe Deckung von KI-bezogenen Erwartungen mit den realen KI-Fähigkeiten erzeugt werden kann, ist in Abbildung 2 illustriert.

Hinsichtlich der Art der zu ergreifenden Maßnahmen unterstreicht die Befundlage die zentrale Rolle eines menschenzentrierten Ansatzes. Das bedeutet, dass es für die Ableitung effektiver Maßnahmen in diesem Kontext unumgänglich ist, die Mitarbeitenden, die mit dem KI-Teammitglied zusammenarbeiten werden, möglichst früh in Überlegungen zur Einführung des KI-Teammitglieds einzubeziehen (Parker/Grote, 2020). Mit diesem Ansatz geht das deutliche Abwenden von einem technologischen Determinismus einher, in dem KI ohne die Berücksichtigung der Auswirkungen auf das Gesamtsystem implementiert wird. Allerdings bedingen sich Mensch und Technologie, wie in soziotechnischen Systemtheorien beschrieben, immer gegenseitig. Deshalb darf der Fokus von Maßnahmen zur Gestaltung von Mensch-KI-Teams nicht ausschließlich auf der Anpassung der Fähigkeiten von KI-Teammitgliedern an menschliche Erwartungen liegen. In diesem Sinne schlagen wir für verschiedene Phasen der Ausgestaltung der Mensch-KI-Zusammenarbeit Handlungsempfehlungen für das Management beider Seiten der Erwartungs-Realitäts-Beziehung vor. Die eine Seite fokussiert dabei auf das Einbeziehen menschlicher Erwartungen in die Gestaltung der KI, während die andere auf die Anpassung unrealistischer Erwartungen abzielt.

**Entwicklung und Design:** Das Einbeziehen der Erwartungen der Mitarbeitenden, die in Zukunft mit einem KI-Teammitglied arbeiten werden, bietet die Möglichkeit, frühzeitig einen Grundstein für eine erfolgreiche Mensch-KI-Zusammenarbeit zu legen. Wir gehen dabei davon aus, dass eine grundsätzliche Bedarfs- und Nutzenklärung der KI-basierten Unterstützung für das Arbeitssystem durch alle relevanten Stakeholder (Management, Mitarbeitende, KI-Entwicklungsverantwortliche) vorgenommen wurde. Daraufhin sollte ein partizipatorischer KI-Entwicklungsprozess angestoßen werden. Das bedeutet, dass ein Austausch zwischen den Verantwortlichen für die Entwicklung der KI und den in Zukunft mit ihr arbeitenden Teammitgliedern begonnen werden sollte. Mitarbeitende erhalten die Möglichkeit, ihre KI-bezogenen Erwartungen den KI-Entwicklungsverantwortlichen mitzuteilen. Andersherum sollte dieser Austausch ermöglichen, dass die KI-Entwicklungs-

verantwortlichen das Spektrum und vor allem die Grenzen der Fähigkeiten des KI-Teammitglieds aufzeigen. Konkrete Methoden des partizipativen Designs sind vielfältig und können je nach Unternehmensstruktur und angedachtem KI-Anwendungsfall unterschiedlich ausgestaltet werden. Zu klassischen Werkzeugen des partizipativen Designs gehören Fragebogenerhebungen und Interviews mit Mitarbeitenden, in denen Erwartungen erfragt werden können. Für einen direkteren und tiefgreifenderen Austausch zwischen KI-Entwicklungsverantwortlichen und zukünftigen KI-Interaktionspartnern sind vor allem Designworkshops denkbar. Dieses Format bietet den Vorteil, dass die Erwartungen der Teammitglieder untereinander gegenüber dem KI-Teammitglied eine Angleichung erfahren, da alle Teilnehmenden gebündelt Informationen zu den Fähigkeiten der KI erhalten. Zusätzlich erlangen sie durch den Austausch mehr Bewusstsein über die KI-bezogenen Erwartungen der anderen Teammitglieder.

**Implementierung:** Die zuvor beschriebenen Maßnahmen können kostspieligen und im schlimmsten Fall sicherheitskritischen Fehlentscheidungen in der Mensch-KI-Interaktion vorgeifen. Dennoch garantieren sie keine ideale Passung zwischen den Erwartungen an das KI-Teammitglied und seinen realen Fähigkeiten, sodass eine Einbindung dieser Aspekte in der Implementierungsphase ebenso wichtig ist. Im Kontext des realen Arbeitsalltags können zusätzliche Erkenntnisse über die Art und Weise, wie Mitarbeitende ihr KI-Teammitglied integrieren und wie KI selbst die Arbeitsprozesse verändert, hervortreten. Einige Erwartungen werden erst durch die Interaktion mit dem KI-Teammitglied im Arbeitsalltag überhaupt erkennbar. Deshalb sollte (mithilfe von arbeits- und organisationspsychologischer Expertise) ein Plan für die Implementierung der KI ausgearbeitet werden. Dieser sollte klar definieren, wann und wie die KI-bezogenen Erwartungen der Mitarbeitenden erhoben und adressiert werden. Essenziell ist dabei die (im Idealfall quantitative) Messung der initialen antizipatorischen Erwartungen vor der Implementierung. Eine weitere Erhebung (inklusive einer Messung der wahrgenommenen Bestätigung oder Widerlegung der initialen Erwartungen) sollte einige Wochen oder Monate später (je nach Anwendungsfall) stattfinden. Diese Daten geben Aufschluss darüber, wie die Mitarbeitenden die Fähigkeiten des KI-Teammitglieds wahrnehmen und ob Diskrepanzen zwischen initialen Erwartungen und tatsächlichen Fähigkeiten bestehen. Aus den Ergebnissen kann abgeleitet werden, inwiefern das KI-Teammitglied und seine Einbettung in das Team weiter konfiguriert werden müssen und ob Mitarbeitende eingehender hinsichtlich der Fähigkeiten des KI-Teammitglieds aufgeklärt und geschult werden sollten.

**Fortlaufende Zusammenarbeit:** Durch die beschriebenen Maßnahmen zur Entwicklung und Einführung des KI-Teammitglieds wird eine Grundpassung zwischen KI-bezogenen Erwartungen und den Fähigkeiten der KI hergestellt. Jedoch

ist es auch in der fortlaufenden Zusammenarbeit nötig, diese Passung weiterhin aktiv zu prüfen und zu stärken. Dabei hilft eine kontinuierliche Reflexion über KI-bezogene Erwartungen und Fähigkeiten des KI-Teammitglieds. Diese Reflexion kann als wiederkehrende Routine im Team etabliert werden. Besonders beachtet werden sollte dabei, dass KI-Systeme über die Zeit Modifikationen durchlaufen. Zum einen können, wie bei simpleren IT-Systemen, sicherheits- oder technologiebedingte Anpassungen notwendig werden. Zum anderen ist vorstellbar, dass die KI, je nach spezifischer Implementierung, eigenständig lernt. In beiden Fällen kann es durch Veränderungen im Können der KI zu einer Schiefelage zwischen den Erwartungen der Mitarbeitenden und den Fähigkeiten des KI-Teammitglieds kommen. Zusammengefasst sollte in der Phase der fortlaufenden Zusammenarbeit kontinuierlich geprüft werden, (1) inwiefern Erwartungen und Erwartungsänderungen von Teammitgliedern weitere Anpassungen der KI erforderlich machen und (2) welche Auswirkungen Updates einer KI auf die Passung zwischen Erwartungen und Fähigkeiten haben.

In Abbildung 3 sind die Handlungsempfehlungen und ihre Ziele zusammengefasst.

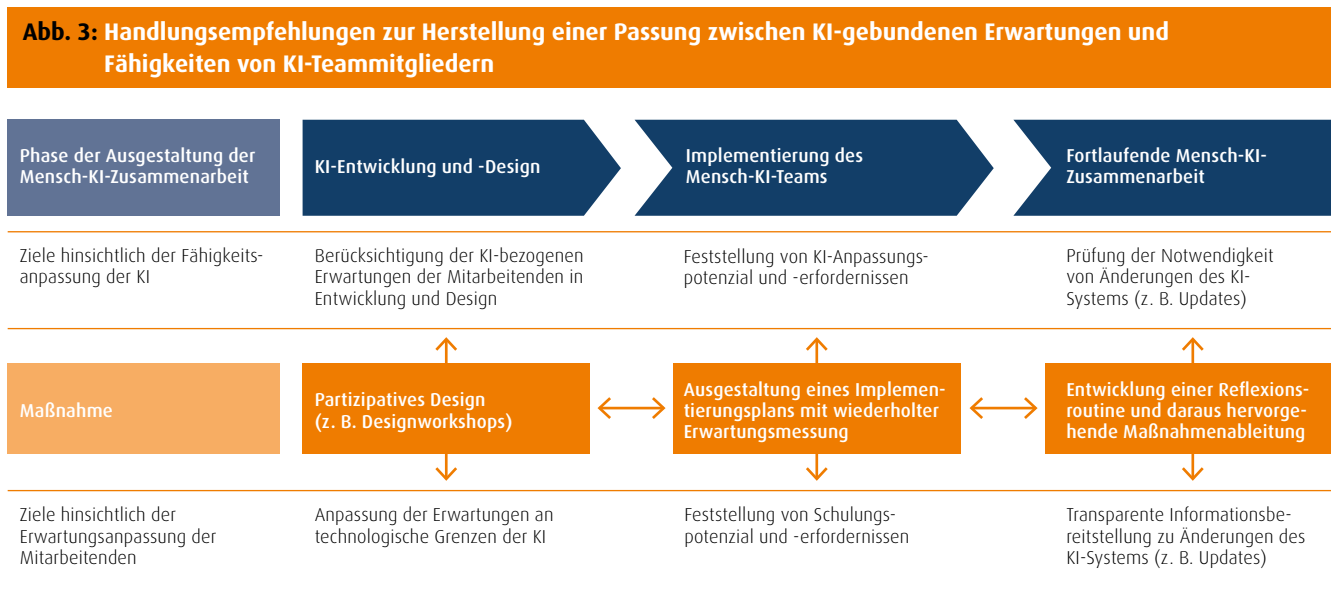
Übergreifend ist zu den Handlungsempfehlungen anzumerken, dass die Übergänge zwischen den vorgeschlagenen Maßnahmen fließend sind. So kann z. B. ein Erkenntnisgewinn aus der Implementierungsphase Informationen hervorbringen, die

dazu führen, dass eine weitere Entwicklungs- und Designphase des KI-Teammitglieds angegangen werden muss. Auch bedingen sich die Maßnahmen aufseiten der Gestaltung des KI-Teammitglieds und der Anpassungen der Erwartungen gegenseitig. Wird bspw. eine bestimmte normative Erwartungshaltung in die Ausgestaltung des KI-Teammitglieds einbezogen, so sollten die Mitarbeitenden darüber informiert werden, da sie nur so die Möglichkeit erhalten, ihre antizipatorischen Erwartungen und ihren Umgang mit der KI entsprechend auszurichten.

**Zusammenfassung und Ausblick**

Mit zunehmendem technologischen Fortschritt steigen die Einsatzmöglichkeiten von KI in Unternehmen. Allerdings operiert KI in der überragenden Mehrheit der Anwendungsfälle weder in Isolation noch ersetzt sie Mitarbeitende vollständig. Deshalb wird zunehmend über den Einsatz von KI als Teammitglied diskutiert. Trotz der erst aufstrebenden Forschung in diesem Themengebiet, zeigt sich bereits jetzt, dass die Erwartungen, die Mitarbeitende sowohl antizipatorisch als auch normativ an ihre KI-Teammitglieder stellen, besonders zentral für eine gute Zusammenarbeit in Mensch-KI-Teams sind. Ein aktives Management und Austarieren der Passung zwischen Erwartungen und Realität ist deshalb unumgänglich.

Für die kommenden Jahre kann davon ausgegangen werden, dass die Forschung weitere Erkenntnisse auf diesem Themengebiet liefern und dabei bestehende Wissenslücken schließen



Quelle: Eigene Darstellung

wird – bspw. hinsichtlich der Frage, wie genau sich voneinander abweichende KI-bezogene Erwartungen verschiedener Teammitglieder auf Interaktionen in Mensch-KI-Teams auswirken. Um diese und ähnliche Aspekte zu adressieren, sind Experimentalstudien im Labor zwar unabdingbar, vor allem können jedoch Zusammenschlüsse zwischen Praxis und Forschung aktiv zur Wissensgenerierung in diesem Themenfeld beitragen. Stakeholder aus der Praxis sind in diesem Sinne nicht nur dazu eingeladen, die Handlungsempfehlungen eines menschenzentrierten Designs von KI-Teammitgliedern anzuwenden, sondern haben die Möglichkeit, das Feld durch Forschungskollaborationen entscheidend voranzubringen.

## LITERATURVERZEICHNIS

- Andrews, R.W./Lilly, J.M./Srivastava, D./Feigh, K.M. (2022):** The role of shared mental models in human-AI teams: a theoretical review. *Theoretical Issues in Ergonomics Science*, S. 1-47.
- Bhattacharjee, A./Premkumar, G. (2004):** Understanding changes in belief and attitude toward information technology usage: A theoretical model and longitudinal test. *MIS Quarterly*, 28(2), S. 229-254.
- Demir, M./McNeese, N. J./Cooke, N. J. (2018):** The impact of perceived autonomous agents on dynamic team behaviors. *IEEE Transactions on Emerging Topics in Computational Intelligence*, 2(4), S. 258-267.
- Grimes, G. M./Schuetzler, R. M./Giboney, J. S. (2021):** Mental models and expectation violations in conversational AI interactions. *Decision Support Systems*, 144.
- Kaplan, A. D./Kessler, T. T./Brill, J. C./Hancock, P. A. (2021):** Trust in artificial intelligence: Meta-analytic findings. *Human Factors*.
- Lyons, J. B./Wynne, K. T./Mahoney, S./Roebke, M. A. (2019):** Trust and human-machine teaming: A qualitative study. In: *Artificial Intelligence for the Internet of Everything*. Elsevier, S. 101-116.
- McNeese, N. J./Demir, M./Cooke, N. J./Myers, C. (2018):** Teaming with a synthetic teammate: Insights into human-autonomy teaming. *Human Factors*, 60(2), S. 262-273.
- Merritt, S. M./Unnerstall, J. L./Lee, D./Huber, K. (2015):** Measuring individual differences in the perfect automation schema. *Human Factors*, 57(5), S. 740-753.
- O'Neill, T./McNeese, N./Barron, A./Schelble, B. (2020):** Human-autonomy teaming: A review and analysis of the empirical literature. *Human Factors*.
- Parker, S. K./Grote, G. (2020):** Automation, algorithms, and beyond: Why work design matters more than ever in a digital world. *Applied Psychology*, S. 1-45.
- Pinquart, M./Endres, D./Teige-Mocigemba, S./Panitz, C./Schütz, A. C. (2021):** Why expectations do or do not change after expectation violation: A comparison of seven models. *Consciousness and Cognition*, 89.
- Rieth, M./Hagemann, V. (2022):** Automation as an equal team player for humans? – A view into the field and implications for research and practice. *Applied Ergonomics*, 98.
- Seeber, I./Bittner, E./Briggs, R. O./de Vreede, T./de Vreede, G.-J./Elkins, A./Maier, R./Merz, A. B./Oeste-Reiß, S./Randrup, N./Schwabe, G./Söllner, M. (2020):** Machines as teammates: A research agenda on AI in team collaboration. *Information & Management*, 57(2).
- Zhang, R./McNeese, N. J./Freeman, G./Musick, G. (2021):** „An ideal human“: Expectations of AI teammates in human-AI teaming. *Proceedings of the ACM on Human-Computer Interaction*, 4, S. 1-25.



### SOPHIE KERSTAN

Doktorandin an der Professur für Arbeits- und Organisationspsychologie am Department Management, Technologie und Ökonomie, ETH Zürich

E-Mail: [skerstan@ethz.ch](mailto:skerstan@ethz.ch)  
[www.wop.ethz.ch](http://www.wop.ethz.ch)



### PROF. DR. ELENI GEORGANTA

Assistenzprofessorin in Work and Organizational Psychology; Programme Group Work and Organizational Psychology an der Fakultät für Social and Behavioural Sciences, University of Amsterdam

E-Mail: [e.georganta@uva.nl](mailto:e.georganta@uva.nl)  
[www.uva.nl](http://www.uva.nl)



### PROF. DR. ANNA-SOPHIE ULFERT-BLANK

Assistenzprofessorin für Organizational Behavior and Artificial Intelligence; Human Performance Management Group am Department für Industrial Engineering & Innovation Sciences, Technische Universität Eindhoven

E-Mail: [a.s.ulfert.blank@tue.nl](mailto:a.s.ulfert.blank@tue.nl)  
[www.research.tue.nl](http://www.research.tue.nl)

## SUMMARY

**Research question:** Artificial intelligence (AI) is increasingly positioned as a team member. Therefore, we investigate what expectations people place on AI team members and outline how these expectations affect human-AI collaboration.

**Methodology:** We summarize the state of research and derive seven observations.

**Practical implications:** AI-related expectations should be actively addressed in AI-development, implementation, and human-AI collaboration. We illustrate ways to realize this kind of expectation management in practice.