

Human Autonomy Teaming

Die Teamarbeit der Zukunft

Michèle Rieth & Vera Hagemann *Hrsg.*



Human Autonomy Teaming – Die Teamarbeit der Zukunft

Herausgegeben von:

Michèle Rieth, M.Sc.

Prof. Dr. Vera Hagemann

Universität Bremen, Fachbereich Wirtschaftswissenschaft

Fachgebiet Wirtschaftspsychologie und Personalwesen

Enrique-Schmidt-Straße 1

28359 Bremen/ Germany

<https://www.uni-bremen.de/perso>

<http://dx.doi.org/10.26092/elib/481>

Dieses Werk ist lizenziert unter einer Creative Commons Namensnennung 4.0 International Lizenz.

<https://creativecommons.org/licenses/by/4.0/>

Wintersemester 2020/2021

Seminar: Projektmodul Verhaltensorientierte Wirtschaftswissenschaft

– Personal- und Wirtschaftspsychologie

Julian Breden; Laura Kaja; Vincent-Konstantin Kapp; Lukas Kowalewski;

Leonie Küster; Jana Masch; Regina Prediger; Simon-Tolga Sambale;

Jessica Schröder; Rafael Stasiak; Mark Sturhann; Jasper Wagner;

Lea Wenzel; Benedikt Wiemann

Zusammenfassung

Der Zusammenarbeit von Mensch und Technik kommt angesichts technologischer Fortschritte eine immer größere Bedeutung zu. Das Human Autonomy Teaming (HAT) birgt in diesem Zusammenhang als neue Form der Teamarbeit zwischen menschlichen Teammitgliedern und technischen Einheiten, sogenannten autonomen Agenten, ein großes Potenzial. Der Mensch kooperiert mit seinem technischen Teammitglied und wird von diesem bei gemeinsamen Aufgaben im Team unterstützt. Beide Akteure ergänzen sich mit ihren individuellen Stärken gegenseitig im Team. Dieses E-Book soll einen Beitrag zum besseren Verständnis des HAT-Konzepts leisten und die Potenziale sowie Herausforderungen der Anwendung des HAT aufzeigen.

In *Kapitel 1* wird ein Grundverständnis für das HAT geschaffen. Dabei wird auf die Geschichte, grundlegende Begriffe, wichtige Automationsmodelle sowie die Potenziale und Risiken des HAT eingegangen. In *Kapitel 2* werden die Anforderungen für erfolgreiches HAT behandelt. Als erstes werden die Anforderungen aufseiten des Menschen und des autonomen Agenten einzeln betrachtet, wie beispielsweise Vertrauen, Akzeptanz und Systemwissen sowie die Programmierung des autonomen Agenten, dessen Design und dessen Anpassungsfähigkeit. Danach werden die Anforderungen an die Menschen und an die autonomen Agenten gemeinsam betrachtet. Ferner werden die bidirektionale Kommunikation, die Transparenz sowie die Koordination und die Teamkognition thematisiert. Die Interaktion im HAT als gleichgestellte Mitglieder entspricht nicht immer der Interaktion in rein menschlichen Teams. Autonome Agenten interagieren und kommunizieren hinsichtlich ihrer physischen Konstruktionsmerkmale und ihres Funktionsverhaltens anders als Menschen. Das Verhältnis zwischen Menschen und autonomen Agenten ist abhängig von verschiedenen Konstrukten wie Autonomie, Transparenz, Verlässlichkeit, Vertrauen und Persönlichkeit des Menschen. Dies wird im *Kapitel 3* thematisiert. In *Kapitel 4* werden Einblicke in einige Anwendungsgebiete und Praxisbeispiele des HAT gewährt, wie beispielsweise das autonome Fahren, unbemannte Fahrzeuge im Militär und autonome Roboter im Weltraum. Zusätzlich werden vier Interviews mit Experten autonomer Technologien präsentiert. *Kapitel 5* widmet sich der Zukunft von HAT. Dabei wird die zukünftige Rolle des Menschen im HAT diskutiert, Verbesserungspotentiale im HAT aufgezeigt sowie soziale und ethische Dilemmata thematisiert.

Das Thema Human Autonomy Teaming wird in Zukunft weiteren Einzug in der Wissenschaft und Praxis finden. Als neue Form der Teamarbeit zwischen autonomen Agenten und menschlichen Teammitgliedern wird dieses Konzept in Zukunft für einen großen Fortschritt der Zusammenarbeit von Mensch und Technik sorgen.

Abstract

As a result of technological advances, collaboration between humans and technology is becoming increasingly important. In this context, Human Autonomy Teaming (HAT), as a new form of teamwork between humans and technology, so-called autonomous agents, has great potential and offers many possibilities in research and application. Both team members complement each other with their individual strengths striving to achieve a common goal. This e-book is intended to contribute to a better understanding of the HAT concept and to show its potential and challenges.

Chapter 1 provides a basic understanding of HAT. The history, basic terms, important automation models as well as the potentials and risks of HAT are discussed. In *Chapter 2*, the requirements for successful HAT are presented. First, the requirements on the side of humans and autonomous agents are considered individually, such as trust, acceptance, and system knowledge as well as the programming of the autonomous agent, its design and its adaptability. Then, the requirements for humans and autonomous agents are considered together. Furthermore, bidirectional communication, transparency, coordination, and team cognition are addressed. The interaction of autonomous agents and human team members does not always correspond to the interaction in purely human teams. Autonomous agents interact and communicate differently from humans in terms of physical design features and functional behavior. The relationship between humans and autonomous agents depends on various constructs such as autonomy, transparency, reliability, trust, and the human's personality. This is discussed in *Chapter 3*. *Chapter 4* provides insights into some application areas and practical examples of HAT, such as autonomous driving, unmanned vehicles in the military, and autonomous robots in space. Additionally, four interviews with experts of autonomous technologies are presented. *Chapter 5* is dedicated to the future of HAT. The future human role in HAT is discussed, potential for improvements in HAT are shown and social and ethical dilemmas are illustrated.

In the future, the topic of human autonomy teaming will quickly find its way into science and practice. As a new form of teamwork between human team members and autonomous agents, this concept will ensure great progress in the collaboration between humans and technology.

Vorwort

Liebe Leserin, lieber Leser,

im Rahmen des Projektmoduls „Verhaltensorientierte Wirtschaftswissenschaft – Personal- und Wirtschaftspsychologie“ am Fachgebiet Wirtschaftspsychologie und Personalwesen der Universität Bremen beschäftigten wir uns im Wintersemester 2020/ 2021 mit der Thematik des *Human Autonomy Teamings* (HAT).

„Siri, wie wird das Wetter heute?“ oder „Alexa, ist ein Stau auf der A1 zwischen Bremen und Hamburg?“ Vielleicht haben Sie selbst solche Sätze, die mit „Siri“ oder „Alexa“ beginnen, schon einmal gesagt. Was Ihnen in diesem Zusammenhang bisher vielleicht nicht bewusst war, ist, dass Sie mit einer Technik interagieren und diese relativ ähnlich zu einem menschlichen Gesprächspartner behandeln. Sie fragen, was Sie gerade wissen möchten, und vielleicht antworten Sie auch mit einem „danke“. Man könnte sagen, Sie sind hier vielleicht schon in einem kleinen Human Autonomy Team unterwegs.

In diesem Buch folgen wir stets der Definition von O’Neill et al. (2020), wonach sich ein HAT aus mindestens einer Person und einer technischen Einheit, dem sogenannten autonomen Agenten, zusammensetzt. Diese arbeiten interdependent zusammen, um eine gemeinsame Aufgabe erfolgreich zu bewältigen. Technische Fortschritte im Bereich des maschinellen Lernens und der künstlichen Intelligenz ermöglichen diese neue Form der Teamarbeit. So werden technische Systeme nicht mehr als reine Werkzeuge oder Unterstützungstools betrachtet, sondern können zunehmend als Teammitglieder eigenständig agieren (Demir et al., 2019; McNeese et al., 2018; O’Neill et al., 2020). Entsprechend widmet sich die Team- und Automationsforschung der letzten Jahre zunehmend diesem Thema, woran auch wir mit diesem Buch anschließen.

Dieses E-Book ist das Ergebnis des Projektmoduls. Die Studierenden haben sich dabei in Gruppen eigenverantwortlich verschiedenen Aspekten des HAT gewidmet und ihre Erkenntnisse gemeinschaftlich in den Kapiteln festgehalten. Eine intensive wissenschaftliche Literaturrecherche ging voraus. Wir haben die Kapitelinhalte im Rahmen des Seminars gemeinsam diskutiert und reflektiert. Die Autor/innen der Kapitel tragen die Verantwortung für die Inhalte.

In *Kapitel 1* wird in das Thema eingeleitet, grundlegende Definitionen und Modelle für das gesamte Werk vorgestellt sowie die Potentiale des HAT aufgezeigt. *Kapitel 2* thematisiert menschliche und technische Anforderungen für erfolgreiches HAT, bevor in *Kapitel 3* näher auf die Zusammenarbeit zwischen Mensch und Technik und die damit einhergehenden Stärken und Schwächen eingegangen wird. *Kapitel 4* liefert Einblicke in aktuellen Anwendungsgebiete des HAT. Abschließend werden in *Kapitel 5* zukünftige Entwicklungen des HAT diskutiert.

Mit diesem E-Book möchten wir aktuelle Themen im Rahmen des HAT für Forscher/innen und Praktiker/innen übersichtlich aufbereiten und zugänglich machen sowie Anregungen liefern, um gemeinsam zur erfolgreichen Umsetzung autonomer Agenten als Teammitglied des Menschen im Sinne eines HAT beitragen zu können.

Michèle Rieth & Vera Hagemann
Februar 2021

Literaturverzeichnis

- Demir, M., McNeese N. J. & Cooke, N. J. (2019). The evolution of human-autonomy teams in remotely piloted aircraft systems operations. *Frontiers in Communication*, 4(50), 1–12.
- 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), 262–273.
- O’Neill, T., McNeese, N., Barron, A. & Schelble, B. (2020). Human–Autonomy Teaming: A review and analysis of the empirical literature. *Human Factors*. Advance Online Publication. <https://doi.org/10.1177/0018720820960865>

Kapitelübersicht

- 1 Von der Automation bis zum Human Autonomy Teaming 8**
Laura Kaja, Lukas Kowalewski & Jasper Wagner

- 2 Anforderungen für erfolgreiches Human Autonomy Teaming 36**
Vincent Kapp, Jessica Schröder & Benedikt Wiemann

- 3 Zusammenarbeit von Menschen und autonomen Agenten 74**
Regina Prediger, Rafael Stasiak & Lea Wenzel

- 4 Einblicke in die Anwendungsgebiete des Human Autonomy Teamings 111**
Leonie Küster, Jana Masch & Mark Sturhann

- 5 Ein Blick in die Zukunft des Human Autonomy Teaming 149**
Julian Breden & Simon Sambale

Von der Automation bis zum Human Autonomy Teaming

Laura Kaja, Lukas Kowalewski & Jasper Wagner

A word cloud of terms related to human autonomy and automation. The words are in various shades of blue and are arranged in a non-uniform, overlapping manner. The most prominent words are 'Kommunikation', 'Automation Bias', 'Automation', 'Autonomie', 'HAT', 'Potential', 'Autonomy', 'Continuum approach', 'Menschlichkeit', 'Human Autonomy Teaming', 'Use', 'Vertrauen', 'Disuse', 'Abuse', 'Misuse', 'Levels of Automation', 'Autonomer Agent', 'Vier-Phasen-Modell', 'Brittleness', and 'Trust'.

Kommunikation
Automation Bias
Automation
Autonomie
HAT
Potential
Autonomy
Continuum approach
Menschlichkeit
Human Autonomy Teaming
Use
Vertrauen
Disuse
Abuse
Misuse
Levels of Automation
Autonomer Agent
Vier-Phasen-Modell
Brittleness

Inhaltsverzeichnis

1	Von der Automation bis zum Human Autonomy Teaming	
1.1	Grundbegriffe des Human Autonomy Teaming	10
1.1.1	Wie aus Automation Autonomie wird	
1.1.2	Was den autonomen Agenten ausmacht	
1.2	Der Ursprung des Human Autonomy Teaming	14
1.2.1	Die Geschichte der Automatisierung bis 2020	
1.2.2	Ein Vergleich von Human Automation Interaction und Human Autonomy Teaming	
1.3	Das Modell der Levels of Automation	19
1.3.1	Levels of Automation	
1.3.2	Vier Phasen der Levels of Automation	
1.3.3	Autonomiegrade der Levels of Automation	
1.4	Potenziale und Risiken des Human Autonomy Teaming	25
1.4.1	Sinnhaftigkeit des Human Autonomy Teaming	
1.4.2	Der Gebrauch von Automation und die damit verbundenen Risiken im Human Autonomy Teaming	
1.4.3	Positive Effekte des Human Autonomy Teaming	
	Literaturverzeichnis	32

Kapitel 1: Von der Automation bis zum Human Autonomy Teaming

Von Laura Kaja, Lukas Kowalewski & Jasper Wagner

Automation gibt es seit einigen Jahrtausenden (Guarnieri, 2010). Mit voranschreitender Zeit hat sich die Nutzung von Automation in der Arbeitswelt stark verändert. In den Anfängen wurde Automation vorwiegend als Hilfswerkzeug angesehen, das den Menschen bei der Zielerreichung unterstützt. Diese Interaktion zwischen Menschen und Automation wird Human Automation Interaction (HAI) genannt. Der technologische Fortschritt ermöglicht eine Zusammenarbeit von Menschen und hochautomatisierten technischen Systemen, die großes Potenzial bietet. Folglich verändert sich die Teamarbeit von rein menschlichen Teams hin zu Teams, bestehend aus Menschen und Technik. In diesem Kontext wird in der wissenschaftlichen Literatur von Human Autonomy Teaming (HAT) gesprochen. Dies hat weitreichende Folgen auf die Sicht der Menschen und auf die Automation und verändert die Nutzung dieser in der Arbeitswelt (O'Neill et al., 2020). Diese Aspekte werden in diesem Kapitel erörtert und anhand von Abbildungen verdeutlicht. Dabei werden die grundlegenden Begriffe des HAT erklärt, damit ein Verständnis für das Thema geschaffen wird. Anschließend wird auf die Entwicklung des HAT eingegangen, die ihren Ursprung in der HAI hat. Innerhalb des HAT haben sich verschiedene Modelle als sinnvoll erwiesen. Eines der etabliertesten Modelle stellt das Levels of Automation Modell von Sheridan und Verplank (1978) dar. Dieses dient als Grundlage für weitere Modelle. Sowohl das Levels of Automation Modell als auch dessen Weiterführungen werden in diesem Kapitel beschrieben. Das HAT bietet durch die besondere Art der Nutzung von Automation einige Potenziale. Jedoch dürfen die Risiken des Gebrauchs von Automation nicht außer Acht gelassen werden. Deshalb werden sowohl die Potenziale als auch die Risiken, die das HAT mit sich bringt, in diesem Kapitel thematisiert.

1.1 Grundbegriffe des Human Autonomy Teaming

In diesem Unterkapitel werden die Grundlagen für ein Verständnis der folgenden Kapitel vermittelt. Grundbegriffe wie Automation und Autonomie werden erläutert. Des Weiteren wird erklärt, wann von einem autonomen Agenten gesprochen werden kann. Weitere Begriffe, die im HAT von Bedeutung sind, wie Brittleness und Menschlichkeit, werden ebenfalls erläutert.

1.1.1 Wie aus Automation Autonomie wird

Automation und Automatisierung

Wie erfolgreich die Zusammenarbeit in einem HAT ablaufen kann, wird zu einem großen Teil von der zugrundeliegenden Automatisierung und Automation bestimmt. Automation kann dem Menschen von großem Nutzen sein, ihm aber auch seine Arbeit erschweren (Bainbridge, 1983). Nach Parasuraman et al. (2000) kann von Automation gesprochen werden, wenn eine Aufgabe von einer technischen Einheit ausgeführt wird, die zuvor von einem Menschen erledigt wurde, bzw. von diesem teilweise oder vollständig ausgeführt werden könnte. Dabei werden einzelne Funktionen oder ganze Tätigkeiten von einem Menschen auf eine Maschine übertragen. Im Laufe der Entwicklung ist der Begriff Automation flexibel verwendbar, da sich mit dem technologischen Fortschritt die Grenzen und Definitionen für Automation und die Fähigkeit der Maschinen verändern (Parasuraman & Riley, 1997). Automation kann mithilfe einer Software und der dazu nötigen Hardware eine Aufgabe ausführen, für die sie programmiert wurde (Parasuraman et al., 2000). Automatisierung beschreibt den Prozess der Einführung von Automation, also den Prozess der Übertragung von Funktionen oder ganzen menschlichen Tätigkeiten auf eine technische Einheit. Ziel der Automatisierung sind die dadurch

gewonnenen ökonomischen Vorteile und Aspekte, wie die erhöhte Sicherheit und Zuverlässigkeit der Systeme (Manzey, 2008). Eine erhöhte Automatisierung führt dazu, dass die Automation immer mehr und kompliziertere Aufgaben übernehmen kann. Je ausgeprägter die Automatisierung, desto höher ist der Automatisierungsgrad (Degree of Automation). Abhängig von der Situation oder Aufgabenstellung, besitzt die Automation einen unterschiedlich hohen Automatisierungsgrad (Onnasch et al., 2014). Dies impliziert, dass Automation kein Alles-oder-Nichts-Phänomen darstellt, sondern über ein Kontinuum von Stufen variieren kann. Von der niedrigsten Stufe der vollständig manuellen Ausführung bis zur höchsten Stufe der vollständigen Automatisierung (Parasuraman et al., 2000). Der Automatisierungsgrad lässt sich daran messen, wie viel Unterstützung dem Menschen in der Zusammenarbeit geboten wird. Ein höherer Automatisierungsgrad ermöglicht es der Automation, komplexe Aufgaben selbstständig zu übernehmen und verringert so den Arbeitsaufwand des Menschen. Ein System mit einem niedrigen Automatisierungsgrad kann hingegen nur einfache Aufgaben, wie die Auswahl von möglichen Handlungen übernehmen und bietet nur wenig eigenständige Unterstützung für den Menschen (Onnasch et al., 2014). Es ist nicht immer erforderlich, stark automatisierte Systeme einzusetzen. Ein großer Vorteil bei einer vergleichsweise niedrigen Automatisierung ist die Kostenreduktion. Vor allem bei einfachen Aufgaben wird ein geringer Automatisierungsgrad eingesetzt (O'Neill et al., 2020). Automation ersetzt nicht automatisch die Arbeit des Menschen, sondern sie verändert sie (Parasuraman & Riley, 1997). Ein höherer Automatisierungsgrad verbessert die Arbeit des Menschen, allerdings bietet er nicht nur Vorteile. Nachteile einer hohen Automatisierung sind verschiedene Probleme, wie z.B. der Verlust eigener Fähigkeiten und letztlich eventuell auch des eigenen Arbeitsplatzes (Endsley, 2017; Endsley & Kiris, 1995). Ein zu niedriger Automatisierungsgrad hat zur Folge, dass die Automation regelmäßig und häufig überwacht werden muss und nur begrenzt eigenständige Unterstützung leistet (Sheridan & Johanssen, 1976). Diese Probleme werden in den folgenden Kapiteln 2 und 3 noch genauer erläutert.

Automation Bias

Bei der Nutzung von Automation kann es zu Entscheidungsverzerrungen kommen. Automatisierte Entscheidungshilfen werden bei Entscheidungsprozessen eingesetzt, um diese effizienter zu gestalten und Entscheidungsfehler zu verringern. Entscheidungsfehler führen zu ökonomischen Kosten oder Einbußen in der Sicherheit, weshalb diese vermieden werden sollen. Wenn die automatisierten Entscheidungshilfen nicht richtig genutzt werden, kann dies zum sogenannten Automation Bias führen. Bei diesem werden nicht alle vorhandenen Informationen analysiert und ausgewertet, um eine Entscheidung zu treffen. Stattdessen wird die Entscheidung durch die von der Automation zur Verfügung gestellten Hinweise beeinflusst. Es wird zu stark auf die Entscheidungshilfen vertraut, wodurch sie teilweise zu häufig eingesetzt werden, was Fehler im Entscheidungsprozess zur Folge haben kann. Zum einen kann der falsche Einsatz der automatisierten Entscheidungshilfen zum error of omission, auf Deutsch Unterlassungsfehler, führen. Bei diesem reagiert der Mensch nicht auf kritische Situationen, weil er nicht rechtzeitig durch die Automation informiert wird. Ein Alltagsbeispiel wäre, dass eine autofahrende Person die richtige Ausfahrt von der Autobahn verpasst, weil die Navigationshilfe sie nicht informiert hat, dass sie abfahren muss. Zum anderen kann der commission error, auf Deutsch Auftragsfehler, auftreten. Dieser kommt dadurch zu Stande, dass Vorschläge oder Anweisungen der automatisierten Entscheidungshilfen fehlerhaft sind, jedoch trotzdem durch den Menschen umgesetzt werden. Ein Alltagsbeispiel wäre, dass eine autofahrende Person von der falschen Seite in eine Einbahnstraße einfährt, weil die Navigationshilfe der fahrenden Person sagt, dies zu tun (Parasuraman & Manzey, 2010). Faktoren, die zum Automation Bias führen, sind unter anderem die Tendenz von Menschen, im

Entscheidungsprozess den Weg mit dem geringsten kognitiven Aufwand zu wählen (Wickens, 1992) oder das zu hohe Vertrauen, das der Automation aufgrund ihrer Analysefähigkeiten geschenkt wird (Lee & See, 2004). Ein weiterer Faktor ist das Gefühl der Aufteilung von Verantwortung auf Menschen und Automation, wodurch Menschen dazu tendieren, ihren eigenen Aufwand bezüglich der Kontrolle und Analyse vorhandener Informationen zu reduzieren (Parasuraman & Manzey, 2010).

Autonomie

Der technologische Fortschritt der letzten Jahrzehnte im Bereich der Automation ermöglichte es, verschieden stark automatisierte Computersysteme zu entwickeln (Parasuraman & Riley, 1997). Autonomie kann als eine höher entwickelte Form von Automation angesehen werden. Sie impliziert, dass das System selbständig und unabhängig vom Menschen und anderen Systemen handelt. Es wählt eigenständig Aufgaben aus, die es anschließend selbstständig ausführt (Johnson et al., 2012). Die Autonomie des Systems ist nicht mit der Selbstständigkeit eines Menschen zu vergleichen, der aufgrund eigener Kreativität und komplexer ethischer Normen handelt (Neef, 2016). Ein hoher Automatisierungsgrad und die damit verbundene Autonomie eines Systems führen dazu, dass dieses eher als ein Teammitglied angesehen wird, da es zum nahezu selbstständigen Handeln befähigt wird (Lyons et al., 2018). Selbst der höchste Automatisierungsgrad einer Maschine bedarf immer noch einer darüber gestellten Programmierung durch einen Menschen. Dadurch wird der Maschine nie die letztendliche Entscheidungsgewalt überlassen. Eine höhere Automatisierung eines Systems ermöglicht keine realen Emotionen, da diese in der Komplexität dem menschlichen Gehirn vorbehalten sind (Megill, 2014). Deshalb können autonome Systeme nicht adäquat auf ihre Teammitglieder reagieren, so wie es menschliche Teammitglieder tun würden und dabei Gefühle oder menschliche Interaktionen zeigen. Trotz hoher Autonomie ist die Kommunikation zwischen Menschen und dem System nicht die Gleiche, wie in einem menschlichen Team (Cooke et al., 2016). Obgleich es erhebliche technologische Weiterentwicklungen im Bereich des Human Autonomy Teaming gibt, hat sich herausgestellt, dass selbst autonome Systeme häufig nicht das gleiche Potential wie ein menschliches Teammitglied aufweisen (O'Neill et al., 2020). Autonomie von Systemen ist in bestimmten definierten Bereichen besonders effektiv. Gut definierte Aufgaben mit geringen Kreativitätsanforderungen und Routineaufgaben können von Systemen besser durchgeführt werden, als von einem Menschen (Chui et al., 2016; O'Neill et al., 2020).

1.1.2 Was den autonomen Agenten ausmacht

Automation mit einem niedrigen Automatisierungsgrad wird unter anderem Agent genannt. Dieser wird oft nur bei Routineaufgaben eingesetzt (Hancock et al., 2013). Dies ist zwar nützlich und erhöht minimal die Effektivität, die Unterstützung ist jedoch nicht mit einem vollwertigen Teammitglied zu vergleichen. Um als ein solches Teammitglied angesehen werden zu können, benötigen die Agenten einen hohen Grad an Autonomie. Agenten, die diesen aufweisen, werden autonome Agenten genannt (O'Neill et al., 2020). O'Neill et al. (2020) definieren einen autonomen Agenten als eine computerbasierte Einheit, die im Team als Individuum wahrgenommen wird und eine bestimmte Rolle in diesem Team einnimmt. Um dies zu erreichen, wird der autonome Agent mit einer Software ausgestattet und entsprechend programmiert. Durch die Software wird dem autonomen Agenten ermöglicht, sich menschenähnlich und selbständig zu verhalten (Lyons et al., 2018). Eine enge Zusammenarbeit kann mit dem autonomen Agenten eingegangen werden. Um vom Menschen als Teammitglied angesehen zu werden, gibt es im Wesentlichen zwei Kriterien. Zum einen muss es eine Abhängigkeit der Teammitglieder untereinander geben. Die Ergebnisse hängen vom Gesamtergebnis der

Gruppe ab, was die Einzelpersonen veranlasst, sich mehr wie ein Team zu verhalten. Dieses Verhalten lässt größere Gemeinsamkeiten innerhalb des Teams entstehen, wodurch das Gemeinschaftsgefühl steigt (Demir et al., 2018; O'Neill et al., 2020). Solche Gemeinsamkeiten können z.B. der gemeinsame Teamname oder gewisse Zeichen und Symbole sein (Walliser et al., 2017). Zum anderen kommt es auf die Menschen selbst und deren Arbeit und Kommunikation mit den autonomen Agenten an. Teamprozesse müssen kommunikationszentriert sein und jedes Mitglied muss zum Gruppenziel beitragen. Es muss ein gemeinschaftliches Bewusstsein und die Absicht entstehen, auf ein gemeinsames Ziel hinzuarbeiten und sich gegenseitig dabei zu unterstützen (Wynne & Lyons, 2018). Sollte der autonome Agent nur als Hilfswerkzeug angesehen werden und nicht als Teammitglied, ist die allgemeine Leistung und soziale Interaktion deutlich geringer (O'Neill et al., 2020; Walliser et al., 2017). Dies hat zur Folge, dass autonome Agenten nicht die Möglichkeit haben, die Richtung der Teamarbeit mitzuentcheiden oder Empfehlungen auszusprechen. Nur Teammitglieder haben diese Möglichkeit (Wynne & Lyons, 2018).

Brittleness

Ein Problem bei der Zusammenarbeit zwischen Menschen und autonomen Agenten ist, dass sich die autonomen Agenten bei falscher Nutzung „brittle“ verhalten können. Der Begriff Brittleness ist ein sehr zentraler Begriff im Human Autonomy Teaming. Er beschreibt ein wichtiges Problem im Bereich der Automation. Brittleness bedeutet auf Deutsch Brüchigkeit oder Sprödigkeit. Dies hat nichts damit zu tun, dass die autonomen Agenten fehlerhaft sind oder drohen auseinanderzubrechen. Brittleness beschreibt ein Problem, welches bei autonomen Agenten entstehen kann, sollten sie nicht für ihren vorgegebenen Zweck eingesetzt werden. Autonome Agenten, die unter dem Problem der Brittleness leiden, haben nur eine begrenzte Effektivität (Onnasch et al., 2014; Roth et al., 1987). Agenten können nur so handeln, wie es ihre Programmierung vorsieht. Dies nimmt ihnen die Möglichkeit, auf unbekannte Situationen zu reagieren und zu improvisieren. Somit sind sie im Gebrauch für andere Anwendungen nicht effektiv, sondern nur zweckgebunden einsetzbar. Um zu vermeiden, dass sich autonome Agenten „brittle“ verhalten, ist es sinnvoll, sie nur in Bereichen einzusetzen, die sich gut definieren lassen. Bereiche wie Waldarbeit oder Erziehung sind nicht sinnvoll, weil es sehr schwer ist, vorauszusagen, was passiert. Für diese Einsätze fehlen den Maschinen Kreativität und Flexibilität (Chui et al., 2016).

Menschlichkeit

Humanness, auf Deutsch Menschlichkeit, beschreibt die Wahrnehmung der Menschen gegenüber autonomen Agenten. Es wird versucht, die Handlungen und die Kommunikation des autonomen Agenten so auszurichten, dass sie möglichst menschlich wirken. Der autonome Agent sollte trotzdem eine sinnvolle Aufgabe in der Teamarbeit erfüllen und eine Aufgabe übernehmen, die zuvor ein Mensch ausführte. Dies hat zur Folge, dass das menschliche Teammitglied den autonomen Agenten als Teammitglied mit menschlichen Zügen wahrnimmt und ihm stärker vertraut. Menschliche Teamarbeit zeichnet sich vor allem durch ihre Kommunikation und ihr Wohlwollen dem Teammitglied gegenüber aus (Lyons et al., 2018; O'Neill et al., 2020). Diese Teameigenschaften sollen mit dem menschenähnlichen Design des autonomen Agenten erreicht werden, um die Teamarbeit zu verbessern (O'Neill et al., 2020). O'Neill et al. (2020) stellen jedoch in Frage, ob es überhaupt möglich ist, die Kommunikation des autonomen Agenten so zu programmieren, dass sie menschlich wirkt und nicht als Maschine angesehen wird.

1.2. Der Ursprung des Human Autonomy Teaming

In diesem Unterkapitel wird auf die Geschichte der Automation seit der Antike eingegangen, in der die erste Automation entwickelt wurde. Darüber hinaus wird die weitere Entwicklung innerhalb der Industriellen Revolution beschrieben, um den technologischen Fortschritt seit dieser Zeit besser einschätzen zu können. Es wird auf Human Automation Interaction als Beginn des Human Autonomy Teaming eingegangen. Mithilfe dieses Verständnisses wird der Blick auf die Gegenwart gerichtet und es wird die Entwicklung des HAT thematisiert.

1.2.1 Die Geschichte der Automatisierung bis 2020

Als eines der ersten Beispiele für Automatisierung kann die Wasseruhr genannt werden, die mithilfe einer kontrollierten Durchflussrate des Wassers die Tageszeit bestimmen konnte. Diese Wasseruhr wurde von dem ägyptischen Ingenieur Ktesibios hergestellt, der vom Jahre 285 v. Chr. bis 222 v. Chr. in Alexandria lebte (Guarnieri, 2010).

Die Industrielle Revolution, die ihren Anfang im 18. Jahrhundert hat, führte zu einem Aufschwung der Automatisierung. Angefangen hat die industrielle Revolution in Großbritannien, dort waren alle notwendigen Gegebenheiten zeitig vorhanden. Diese Gegebenheiten waren:

- ein produktiver Agrarsektor,
- eine rasch anwachsende Bevölkerung, was wiederum eine große Menge an Arbeitskraft zur Verfügung stellt,
- die Teilnahme am Welthandel
- und die Offenheit für wissenschaftliche Erkenntnisse.

Die Entwicklung ist an weitere Voraussetzungen geknüpft, so müssen beispielsweise Rohstoffe, Technologien und Transportmöglichkeiten vorhanden sein (Liedtke, 2012). Zwei wesentliche Errungenschaften der Automatisierung sind die „Spinning Jenny“, eine Spinnmaschine, und die Dampfmaschine, wobei die Dampfmaschine als Symbolbild der Industrialisierung gilt (Butschek, 2006). Sie konnte in vielen Wirtschaftszweigen, insbesondere der Eisenindustrie, einen Beitrag zur Automatisierung leisten. Im 19. Jahrhundert wurde die Textilindustrie mithilfe der Dampfmaschine verbessert. Der Einsatz von Dampfmaschinen ermöglichte es, ökonomische Vorteile von Automatisierung zu realisieren (Butschek, 2006; Ziegler, 2012). Bis zu diesem Zeitpunkt war der Begriff der Automation in der Wissenschaft noch nicht verankert. Die erste wissenschaftliche Nennung von Automation erfolgte 1952 von J. Diebold. Er definierte Automation als einen automatisierten Vorgang oder einen Prozess, um Güter herzustellen (Hitomi, 1994). Die Definition des Begriffs hat sich seitdem gewandelt (vgl. Kapitel 1.1.1). Kurze Zeit später, 1956, kam es bei der Dartmouth Konferenz zur Namensgebung der künstlichen Intelligenz (KI) (Buchanan, 2005). KI beschreibt dabei Maschinen oder Computer, die kognitive Funktionen des Menschen, wie z.B. das Lernen und die Problemlösung nachahmen (Russell & Norvig, 2009). Mit voranschreitenden technologischen Entwicklungen im Bereich der KI, veränderte sich ab den 1980ern erstmals die Einstellung gegenüber Automation. Des Weiteren wurde sich damit auseinandergesetzt, wie Menschen und Computer besser miteinander interagieren können (O'Neill et al., 2020). Diese Gedanken wurden in den 1990ern weiter ausgeführt, indem Automation in Form von autonomen Agenten in menschliche Teams integriert wurde. Die autonomen Agenten wurden in der Zeit erstmalig als potenzielle Teammitglieder angesehen. Zudem wurde in den 90ern und 2000ern in Richtung der zeit- und sicherheitskritischen Aufgaben geforscht. Kraftwerke, Produktionsprozesse in Fabriken über Flugüberwachung und halbautomatisches Fahren waren Ziele der Forschung. Themen waren beispielsweise die Verteilung von Aufgaben zwischen Menschen und Maschinen oder das

Finden des richtigen Arbeitsaufwandes. Diese Themen beschäftigen die Forschung noch heute.

Seitdem Computerchips immer kleiner, effizienter und günstiger werden, steigt die Nutzung von automatischen Systemen von Menschen außerhalb der Forschung und Industrie. Menschen können Smartphones und andere Geräte, die mit dem Internet verbunden sind, nutzen, um mit automatisierten Systemen zu interagieren. Diese Interaktion zwischen Menschen und Automation wird Human Automation Interaction genannt (Janssen et al., 2019). Der Ausdruck HAI wird seit einigen Jahrzehnten für die Beschreibung dieser Interaktion in der wissenschaftlichen Literatur als sinnvoll erachtet. Der Begriff des Human Autonomy Teaming, welcher die Zusammenarbeit von Menschen und autonomen Agenten beschreibt, hat sich erst seit dem Jahr 2017 in der wissenschaftlichen Literatur etabliert. Das Prinzip des HAT als solches ist nicht neu. Es wird seit den 90ern beschrieben, jedoch erfolgt die begriffliche Trennung von HAI und HAT erst seit einigen Jahren (O'Neill et al., 2020).

Anhand eines Zeitstrahls können die wichtigsten Ereignisse der Automatisierung dargestellt werden, um einen besseren Überblick der verschiedenen Höhepunkte dieser in der Geschichte zu bekommen (Abb. 1.1).

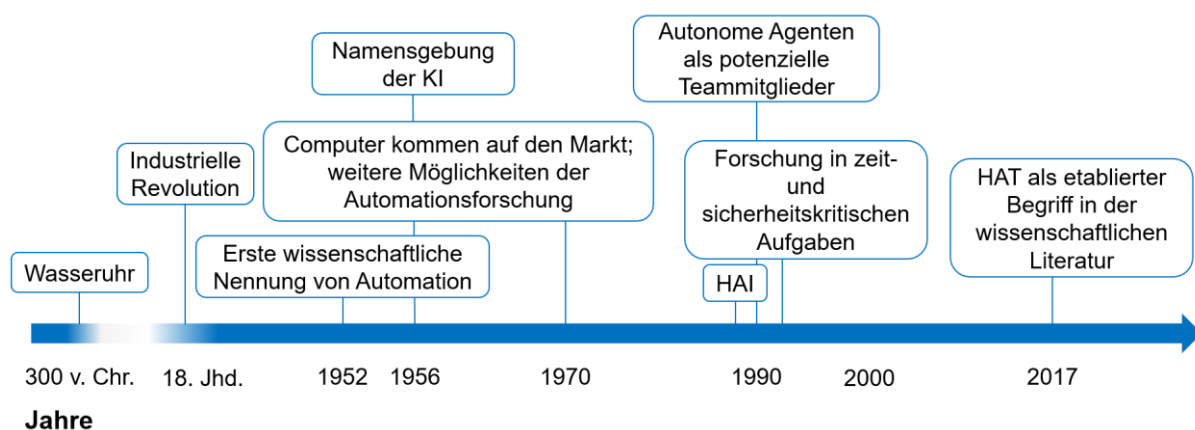


Abb.1.1: Zeitstrahl über die wichtigsten Ereignisse der Automatisierung.

Human Automation Interaction

Der Ursprung des Human Autonomy Teaming liegt in der Human Automation Interaction. Die Grundlagen der Forschung des HAT basieren auf der Forschung der HAI der letzten 40 Jahre (O'Neill et al., 2020). Die HAI stellt eine der unterschiedlichen Interaktionsformen von Menschen mit Automation dar. Weitere Interaktionsformen sind das Human Autonomy Teaming und die Human Robot Interaction (HRI) (Endsley, 2017). Mit Automation kann der Mensch interagieren, entweder auf passiven oder aktiven Wegen (Sheridan & Parasuraman, 2005). Der einfache Gebrauch von Wasser, Elektrizität oder Treibstoffen stellt keine direkte Interaktion mit der Automation dar, sondern eine passive Nutzung automatisch hergestellter Produkte. HAI beschreibt die aktive Interaktion, die der Mensch mit der Automation eingeht. Dies ist auf drei verschiedene Arten möglich. HAI liegt laut Sheridan und Parasuraman (2005) immer dann vor, wenn einer der folgenden Aspekte erfüllt ist:

- a) Wenn Menschen für die Automation die Aufgabenziele und -einschränkungen (mache X, aber vermeide Y) und Kompromisse zwischen diesen festlegen

Ein Beispiel hierfür wäre, dass eine flugzeugführende Person ihren Flugnavigator anhand einer digitalen Tastatur oder einer speziellen Befehlssprache programmiert, damit der Autopilot das Flugzeug auf eine neue Höhe und einen neuen Steuerkurs bringt. Dabei soll über bestimmte Wegpunkte ein entfernter Flughafen angefliegen werden.

- b) Wenn Menschen die Automation zum Starten, Stoppen oder Ändern der automatischen Aufgabenausführung bedienen

Ein Beispiel hierfür wären Mechaniker/innen, die eine Maschine programmieren, um ein Metallteil in einer Reihe von Bearbeitungsvorgängen herzustellen.

- c) Oder wenn Menschen von der Automation Informationen, physikalische Objekte, Substanzen oder ähnliches erhalten

Ein Beispiel hierfür wären Raumfahrtingenieure/innen, die die Bewegungen eines Roboterarms auf einem Marsrover programmieren.

Als einfache Alltagsbeispiele für HAI können das Knopfdrücken eines Fahrstuhls oder das Einstellen der Waschmaschine genannt werden (Sheridan & Parasuraman, 2005). Als Teamarbeit kann dies jedoch nicht angesehen werden, da es weder eine direkte Zusammenarbeit darstellt noch eine beidseitige Kommunikation vorhanden ist. Technische Fortschritte im Bereich der künstlichen Intelligenz und des maschinellen Lernens ermöglichen immer höhere Automatisierungsgrade, sodass Automation heute immer mehr Aufgaben und Entscheidungen selbständig lösen kann, ohne dass ein Mensch eingreifen muss. Dies ermöglicht eine Zusammenarbeit zwischen Menschen und Automation, in der Automation in Form eines autonomen Agenten ein Teammitglied darstellt (Demir et al., 2018; O'Neill et al., 2020).

Human Autonomy Teaming

Laut O'Neill et al. (2020) sowie McNeese et al. (2018) handelt es sich um HAT, wenn eine Teamarbeit zwischen mindestens einem Menschen und mindestens einem autonomen Agenten stattfindet. Die unterschiedlichen Teammitglieder nehmen dabei eigenständige und einzigartige Rollen ein und verfolgen ein gemeinsames Ziel. Die menschlichen Teammitglieder werden in der Literatur häufig Operator genannt (O'Neill et al., 2020). Um ein Verständnis für das HAT zu schaffen, wird der Begriff Teamarbeit in diesem Kontext erläutert. Teamarbeit im HAT kann als die Interaktion von zwei oder mehr heterogenen und unabhängigen Teammitgliedern beschrieben werden, die an einem gemeinsamen Ziel oder einer Aufgabe arbeiten (Johnson et al., 2012). Sie gelten als ein sich mit der Zeit änderndes und somit dynamisches System, wenn sie aktiv miteinander arbeiten und dabei ihre Fähigkeiten dazu einsetzen, ein gesetztes Ziel gemeinsam zu erreichen. Ein wichtiger Teil einer erfolgreichen Teamarbeit ist die Abstimmung der verschiedenen Eigenschaften der Teammitglieder über eine gewisse Zeitspanne. Des Weiteren stellt die Kommunikation zwischen den Teammitgliedern ein zentrales Element von Teamarbeit dar (Demir et al., 2019). Grundlage des HAT ist der partielle Ersatz von Menschen in einem Team durch programmierte Maschinen. Der partielle Ersatz erfolgt mithilfe eines Systems oder einer Maschine, welche die tatsächlichen oder denkbaren Arbeiten von Menschen voll oder zum Teil übernimmt (Demir et al., 2018). Zielsetzung ist nicht nur eine Erhöhung der Produktivität durch ein Ersetzen von menschlichen Arbeitskräften durch Maschinen, sondern darüber hinaus eine Verbesserung des Umgangs innerhalb des Teams und der Leistung (Sheridan & Parasuraman, 2005).

1.2.2 Ein Vergleich von Human Automation Interaction und Human Autonomy Teaming

Bei der HAI und dem HAT handelt es sich um unterschiedliche Arten der Interaktion von Menschen mit Automation. Der grundlegende Unterschied der beiden Begrifflichkeiten liegt darin, dass bei der HAI keine Autonomie vorhanden ist. Infolgedessen kann keine Teamarbeit entstehen. Automatisierte Systeme arbeiten nach einem bestimmten Plan, so wie sie programmiert worden sind. Dadurch fungieren sie als Hilfswerkzeug, um Menschen bei der Zielerreichung zu unterstützen. Autonome Systeme hingegen können selbstständig Aufgaben übernehmen und so einen Teil zur gemeinsamen Zielerreichung beitragen. Zudem ist die Art der Zusammenarbeit mit dem Menschen eine andere. Bei der HAI handelt es sich lediglich um eine Interaktion des Menschen mit der Automation. Dabei wird die Automation ausschließlich als Mittel zum Zweck betrachtet. Beim HAT wird die Automation in Form eines autonomen Agenten als eigenständiges, vollwertiges und gleichwertiges Teammitglied angesehen (O'Neill et al., 2020; Shively et al., 2018). Die Human Robot Interaction (HRI) kann ebenfalls nicht mit dem HAT gleichgesetzt werden. Roboter sind nur bedingt Beispiele für HAT. Der grundlegende Unterschied zwischen der HRI und dem HAT ist, dass Roboter teilweise noch menschliche Überwachung und Kontrolle benötigen. Dies führt dazu, dass ein Roboter kein eigenständiges Teammitglied darstellen kann, da er nicht über ausreichend Autonomie verfügt. Die Beziehung des Menschen zu einem Roboter ist gesondert geregelt. Ebenfalls sind die Vertrauensebenen in der Zusammenarbeit von Menschen und Robotern anders definiert als im HAT (Sheridan, 2016). Die unterschiedlichen Begrifflichkeiten der Interaktion von Menschen mit Automation werden in Abbildung 1.2 vereinfacht dargestellt.

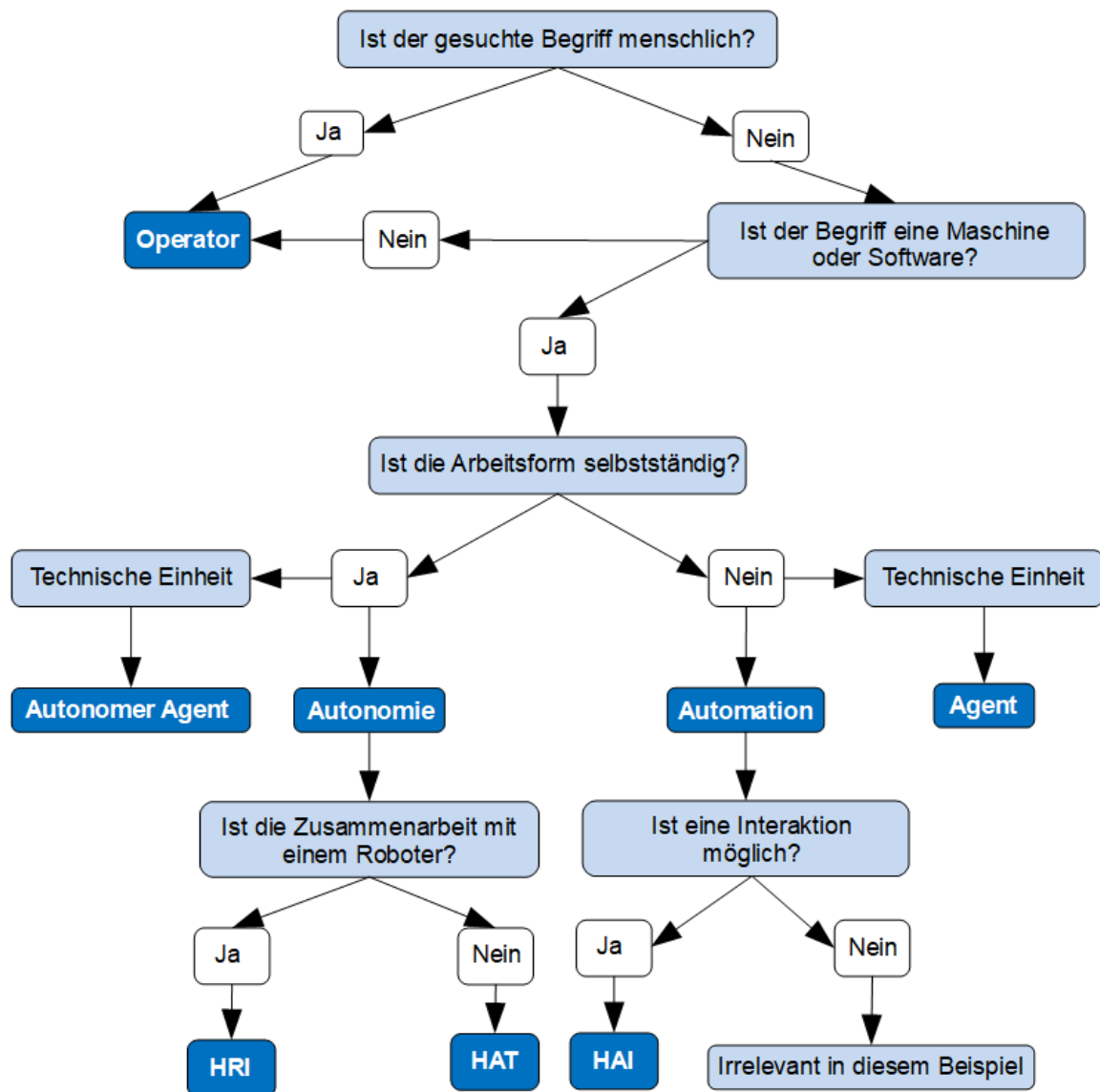


Abb. 1.2: Versuch der vereinfachten Darstellung der begrifflichen Einordnung.

Ein Blick in die gegenwärtige Entwicklung des Human Autonomy Teaming

Die Ergebnisse der Arbeiten und Forschungen bis zum heutigen Stand unterstützen die Annahme, dass die Teamarbeit mit autonomen Agenten weiter zunimmt. Das Human Autonomy Teaming zeigt viele potentielle Möglichkeiten, die Arbeit im Team zu verändern (Parasuraman & Riley, 1997). Dadurch können große Mengen an Informationen von autonomen Agenten gleichzeitig bearbeitet, viele Sichtweisen mit einbezogen und dem Menschen ohne Einbezug von Gefühlen neutral präsentiert werden. Des Weiteren können autonome Agenten aufgrund ihrer erhöhten Rechenleistung Menschen in schwierigen und gefährlichen Situationen erheblich unterstützen. Diese Rechenleistung zeichnet sich dadurch aus, dass ein Computer größere kognitive Fähigkeiten aufweist und nicht vergisst. Der Mensch wird entlastet und kann sich besser auf seine Aufgaben konzentrieren (Shively et al., 2018; Wynne & Lyons, 2018).

Teamarbeit wird durch die Beteiligung von autonomen Agenten gänzlich neu überdacht. Dabei kann ein autonomer Agent sehr gute Arbeitsleistungen in stark dynamischen Aufgabenbereichen in Zusammenarbeit mit dem Menschen erbringen. Jedoch gibt es zurzeit Limitationen, einen autonomen Agenten als Teammitglied einzusetzen. Dazu zählen Einschränkungen der Intelligenz und Interaktionsmöglichkeiten zwischen autonomen Agenten und den anderen

Teammitgliedern (Demir et al., 2018). Die wichtigsten Begriffe der ersten beiden Unterkapitel, wie HAT und autonomer Agent, werden in der Infobox 1.1 erneut aufgegriffen und in verkürzter Form definiert.

Infobox 1.1: Begriffsdefinitionen des Human Autonomy Teaming im Überblick.

<p>Automation: Eine Aufgabe, die zuvor von einem Menschen erledigt wurde, bzw. von diesem teilweise oder vollständig ausgeführt werden könnte, wird teilweise oder vollständig auf eine technische Einheit übertragen.</p>
<p>Automation Bias: Entscheidungsverzerrung, die bei der Nutzung von automatisierten Entscheidungshilfen auftreten kann, da die Entscheidung durch die von der Automation zur Verfügung gestellten Informationen beeinflusst wird.</p>
<p>Automatisierung: Automatisierung beschreibt den Prozess der Einführung von Automation, also den Prozess der Übertragung von Funktionen oder ganzen menschlichen Tätigkeiten auf eine technische Einheit.</p>
<p>Autonomie: Autonomie kann als eine höher entwickelte Form von Automatisierung angesehen werden. Sie impliziert, dass das System selbständig und unabhängig vom Menschen und anderen Systemen handelt.</p>
<p>Agent: Automation mit einem niedrigen Automatisierungsgrad, die meist bei Routineaufgaben eingesetzt wird.</p>
<p>Autonomer Agent: Eine computerbasierte Einheit, die im Team als Individuum wahrgenommen wird und eine bestimmte, selbständige Rolle in diesem Team einnimmt.</p>
<p>Brittleness: Automation, die nicht für ihre vorgegebenen Zwecke eingesetzt wird und deswegen eine verringerte Leistung erbringt.</p>
<p>Menschlichkeit: Die Handlungen und die Kommunikation des autonomen Agenten werden so ausgerichtet, dass sie möglichst menschlich wirken.</p>
<p>Human Automation Interaction: HAI beschreibt die aktive Interaktion, die der Mensch mit der Automation eingeht.</p>
<p>Human Autonomy Teaming: Es findet eine Teamarbeit zwischen mindestens einem Menschen und mindestens einem autonomen Agenten statt. Die unterschiedlichen Teammitglieder nehmen dabei eigenständige und einzigartige Rollen ein und verfolgen ein gemeinsames Ziel.</p>
<p>Human Robot Interaction: HRI beschreibt die menschliche Interaktion mit einem Roboter.</p>

1.3 Das Modell der Levels of Automation

In diesem Unterkapitel wird das Levels of Automation Modell erklärt. Dieses bietet eine Grundlage für weitere Modelle. Dazu zählen das Vier-Phasen-Modell von Parasuraman et al. (2000) und das Autonomiegrad-Modell von O'Neill et al. (2020). Diese Weiterführungen des Levels of Automation Modell werden ebenfalls erläutert.

1.3.1 Levels of Automation

Um eine Klassifikation von technischen Systemen zu ermöglichen, entwickelten Sheridan und Verplank (1978) das Levels of Automation Modell. Dieses Modell teilt Automation in zehn Level ein. Jedes Level spiegelt einen bestimmten Automatisierungsgrad der Agenten wider. Das Modell bewertet, wie viel Unterstützung der Agent bietet und wie eigenständig er dabei handelt. Das erste Level des Modells stellt den niedrigsten und das zehnte Level den höchsten Automatisierungsgrad dar. Der Grad der Automatisierung des technischen Systems nimmt vom ersten bis zum zehnten Level zu (Sheridan & Verplank, 1978). Die einzelnen Level können wie folgt beschrieben werden:

- **Level 1:** Der Agent bietet keine Hilfestellung an, die Handlung wird manuell vom Menschen ausgeführt.
- **Level 2:** Der Agent bietet dem Menschen viele verschiedene Handlungsalternativen an.
- **Level 3:** Der Agent bietet dem Menschen wenige Handlungsalternativen an.
- **Level 4:** Der Agent schlägt dem Menschen nur eine Handlungsalternative vor.
- **Level 5:** Der Mensch muss nur die ausgewählte Handlung des autonomen Agenten bestätigen und dieser führt sie anschließend aus.
- **Level 6:** Der Mensch hat eine gewisse Zeitspanne, in der er gegen die Handlung des autonomen Agenten stimmen kann.
- **Level 7:** Der autonome Agent führt automatisch seine Aufgabe aus und informiert den Menschen danach über sein Handeln.
- **Level 8:** Der autonome Agent informiert den Menschen nur, wenn er gefragt wird.
- **Level 9:** Der Mensch wird nur informiert, wenn der autonome Agent es für nötig hält.
- **Level 10:** Der autonome Agent handelt komplett eigenständig und ignoriert den Menschen.

Die Wahl des passenden Levels ist in einer Teamarbeit mit einer Automation wichtig. Sollte ein zu hohes oder ein zu niedriges Level gewählt werden, kann dies zu Problemen in der Zusammenarbeit führen (Chen et al., 2011). Ein zu niedriges Level der Automation führt zu einer erhöhten Belastung des Menschen. Daraus folgen Müdigkeit und eine schlechtere Ausführung der Aufgaben (Wright et al., 2013). Ein höheres Level der Automation hat vor allem Vorteile bei Teamarbeiten, bei denen eine Lösung von parallelaufenden und komplexen Aufgaben erwartet wird (Manzey et al., 2008). Der Einsatz eines zu hohen Levels der Automation kann zu erheblichen Nachteilen führen. Diese beinhalten den Verlust eigener Fähigkeiten oder des Situationsbewusstseins (Kaber & Endsley, 2004; Parasuraman & Riley, 1997; Parasuraman et al., 2000). Um dies zu verhindern, könnte ein mittleres Level der Automation gewählt werden. Doch ein mittleres Level der Automation schützt nicht vor Bedienungsfehlern, bzw. Anwendungsfehlern oder anderen Problemen, die unabhängig von dem Level der Automation sind. Kein Level des Modells kann als optimal für jede Situation angesehen werden. Für jede Aufgabe sollte individuell das Level ausgesucht werden, um eine optimale Nutzung der Automation zu gewährleisten (Manzey et al., 2008).

1.3.2 Vier Phasen der Levels of Automation

Das Levels of Automation Modell gilt als Grundlage für viele Modelle des HAT. Ein Modell, das diese Grundlage nutzt und darauf aufbaut, ist das Vier-Phasen-Modell von Parasuraman et al. (2000). Das Vier-Phasen-Modell bezieht die zehn Level der Automation auf vier

aufeinanderfolgende Ebenen, die den menschlichen Handlungsprozess darstellen sollen. Die vier Ebenen des Modells teilen sich in

- Informationsaufnahme,
- Informationsverarbeitung,
- Entscheidungsfindung
- und Handlungsausführung auf.

Die Phasen können jedoch nur vereinfacht menschliches Handeln darstellen. Menschliche Handlungen bei der Ausführung von Aufgaben sind häufig zu komplex, um sie exakt aufzuschlüsseln und zu beschreiben. Der kognitive Ablauf bei der Ausführung von Aufgaben ist bei Menschen nicht einer bestimmten Phase zugeordnet. Menschliche Handlungen beziehen sich oft auf mehrere Phasen gleichzeitig. Bei menschlichen Handlungen sind die Phasen weiterhin miteinander verwoben und werden häufig sogar simultan ausgeführt (Wickens et al., 2016). Bei autonomen Agenten ist es einfacher festzulegen, zu welcher Phase ihre aktuelle Aufgabe gehört. Dies liegt daran, dass der autonome Agent nur die Aufgabe ausführt, für die er programmiert worden ist. Das Vier-Phasen-Modell zeigt, dass autonome Agenten nicht in jeder Phase gleich hoch automatisiert werden müssen. Für jede Phase kann es je nach Aufgabe sinnvoll sein, einen anderen Automatisierungsgrad für den Agenten zu wählen (Parasuraman et al., 2000). Das Modell bietet einen großen Vorteil gegenüber anderen Modellen, die häufig nur eine bestimmte Art der Automatisierung beschreiben (Onnasch, 2015).

Die erste Phase des Modells ist die Informationsaufnahme. In dieser ist die Hauptaufgabe der autonomen Agenten die Erfassung und Registrierung der relevanten Daten für die Erfüllung der Aufgabe. Der Einsatz von autonomen Agenten ist in dieser Phase sinnvoll, weil autonome Agenten Daten schneller verarbeiten können als Menschen. Sie haben in Abhängigkeit von ihrer Hardware ein besseres Arbeitsgedächtnis als Menschen. Bei einer niedrigen Automatisierung konzentriert sich der autonome Agent ausschließlich auf das Sammeln von Daten. Bei einem hohen Automatisierungsgrad werden die gesammelten Daten zusätzlich automatisch vom autonomen Agenten gefiltert und hervorgehoben. Das selektive Hervorheben und Filtern von Informationen führt häufig dazu, dass der Mensch in seinem Handeln beeinflusst wird. Dies hat Auswirkungen auf die Arbeitserfahrungen, die Arbeitsleistung und die erzielten Resultate des menschlichen Teammitgliedes (Parasuraman et al., 2000).

Als zweite Phase des Modells folgt die Informationsverarbeitung. In dieser Phase übernimmt der autonome Agent die Aufgaben von kognitiven Funktionen, wie z.B. die Gruppierung oder die Abgrenzungen von Daten. Bei einem niedrigen Automatisierungsgrad erstellt der autonome Agent mithilfe von einfachen Algorithmen Trendvoraussagen oder versucht zu ermitteln, welches Ereignis oder welcher Verlauf in Zukunft stattfinden wird. Bei einem hohen Automatisierungsgrad bezieht der autonome Agent viele verschiedene Variablen bei der Berechnung mit ein. Am Ende der Berechnung werden alle Variablen miteinbezogen und der autonome Agent zeigt dem Menschen idealerweise ein einziges Ergebnis an (Parasuraman et al., 2000).

In der Phase der Entscheidungsfindung hilft der autonome Agent dem Menschen bei der Auswahl der optimalen Entscheidung. Der autonome Agent muss zwischen verschiedenen möglichen Lösungen die beste für die jeweilige Situation auswählen. Dazu liefert er Entscheidungsgrundlagen zu verschiedenen Ausgängen der Handlung. Er wird entsprechend seiner Programmierung lediglich sinnvolle Entscheidungsgrundlagen liefern und diese mit Eintrittswahrscheinlichkeiten bewerten (Parasuraman et al., 2000).

Die letzte Phase des Modells ist die Handlungsausführung. Der autonome Agent ersetzt in dieser Phase meist die Hand oder Stimme des Menschen. Die Höhe des Automationsgrades

wird von der Anzahl der übernommenen Handlungen definiert. Autonome Agenten beobachten den Menschen und führen Handlungen oder Unteraufgaben eigenständig aus, wenn ein bestimmtes Ereignis auftritt, welches die Handlung abschließt (Parasuraman et al., 2000).

Die adaptive Automation beschreibt einen autonomen Agenten, der nicht speziell für ein bestimmtes Level oder eine Phase programmiert wurde, sondern sich situativ anpasst und verändert. Dabei deckt er sämtliche Phasen und Level ab und kann als Alleskönner angesehen werden. Diese Automation wird jedoch nicht in das Vier-Phasen-Modell mit aufgenommen, da es sehr aufwändig ist, einen adaptiven autonomen Agenten zu programmieren. Diese Schwierigkeit ergibt sich aus der notwendigen Berücksichtigung aller Eventualitäten und komplexen Fragestellungen (Parasuraman et al., 2000). Die vier Phasen und die Möglichkeit der unterschiedlich hohen Automatisierung dieser, werden in Abbildung 1.3 dargestellt.

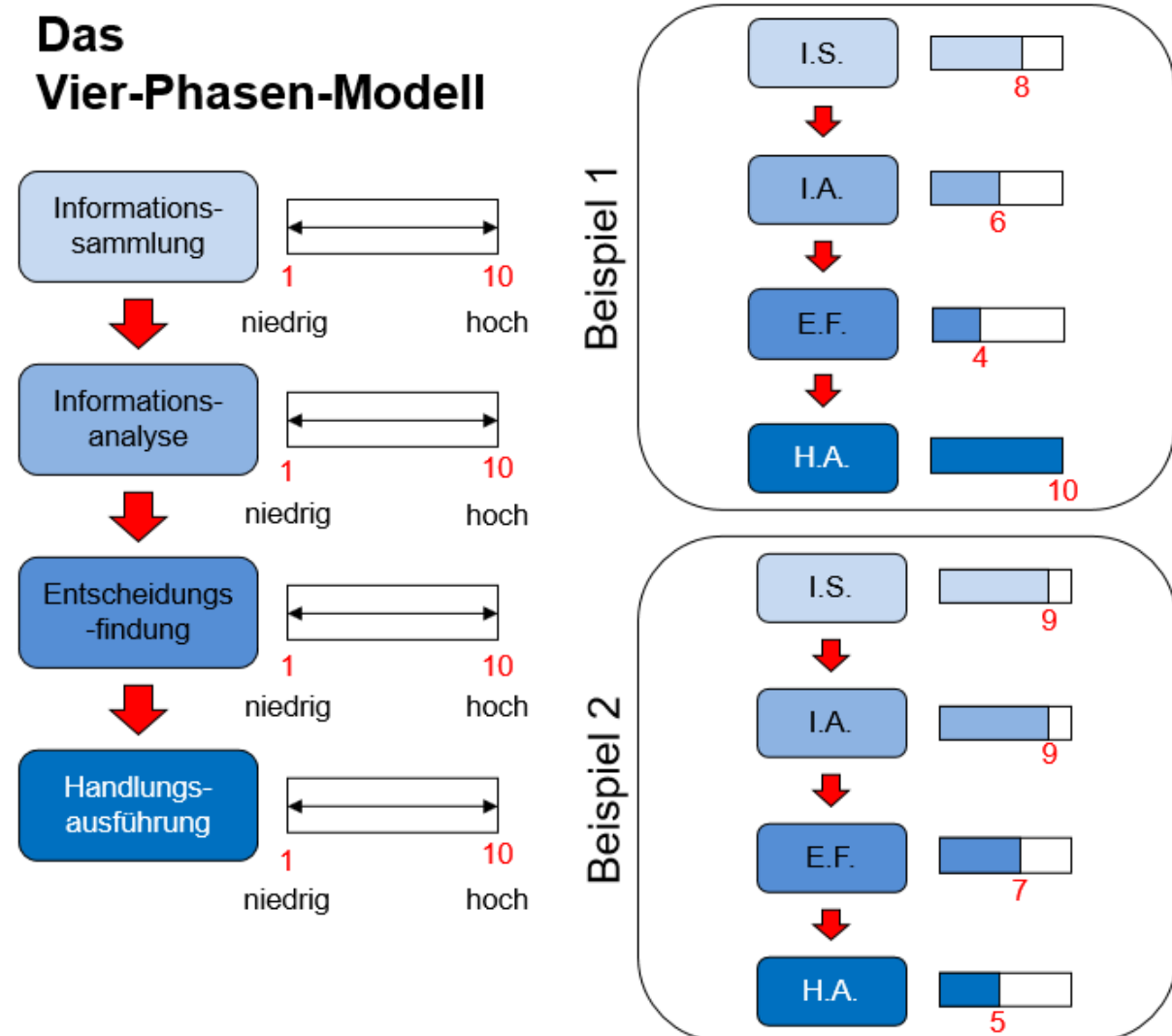


Abb. 1.3: Vier-Phasen-Modell (in Anlehnung an Parasuraman et al., 2000, S. 287).

Die Abbildung zeigt die vier Phasen Informationssammlung (I.S.), Informationsanalyse (I.A.), Entscheidungsfindung (E.F.) und Handlungsausführung (H.A.). Des Weiteren wird dargestellt, dass innerhalb jeder Phase die Möglichkeit besteht ein Automationslevel zwischen eins und zehn zu wählen. Im Beispiel 1 werden die Phasen I.S., I.A., E.F. und H.A. mit den Leveln acht, sechs, vier und zehn automatisiert. Im Beispiel 2 werden die Phasen I.S., I.A., E.F. und H.A.

mit den Leveln neun, neun, sieben und fünf automatisiert. Diese Beispiele verdeutlichen, dass es eine Vielzahl möglicher Kombinationen zwischen Automationsleveln und Phasen gibt.

1.3.3 Autonomiegrade der Levels of Automation

O'Neill et al. (2020) nutzen das Levels of Automation Modell von Sheridan und Verplank (1978) ebenfalls als Grundlage. Als Erweiterung haben sie den Leveln der Automation eine zusätzliche Unterteilung in verschiedene Autonomiegrade hinzugefügt. Diese Autonomiegrade lauten:

- keine Autonomie,
- teilweise Autonomie,
- Autonomie
- und hohe Autonomie.

Die Zuweisung der unterschiedlichen Autonomiegrade zu den Leveln der Automation sind in der untenstehenden Abbildung 1.4 dargestellt.

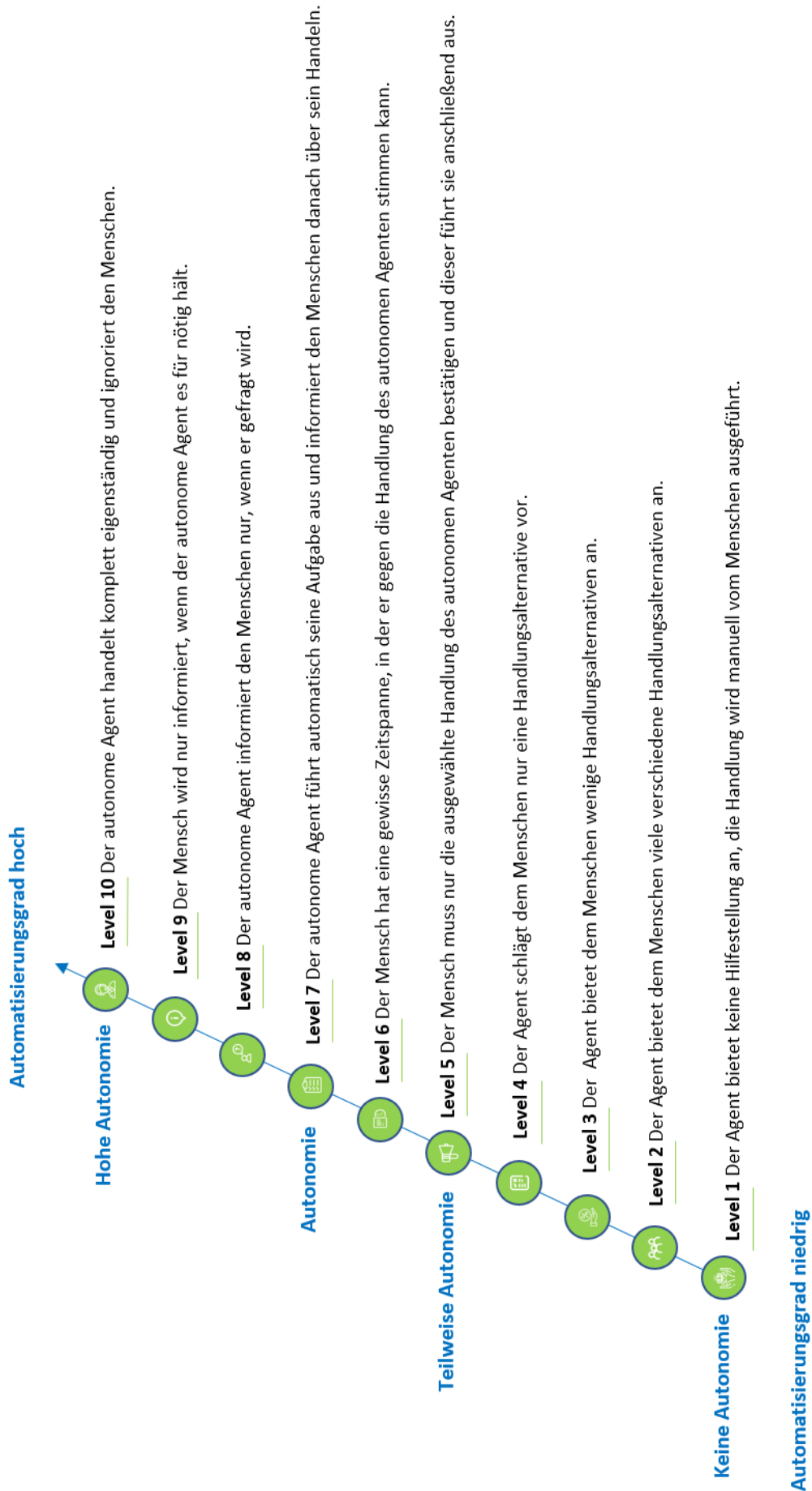


Abb. 1.4: Das Autonomiegrad-Modell (in Anlehnung an O'Neill et al., 2020, S. 6).

Die Abbildung 1.4 zeigt die von O'Neill et al. (2020) definierten Grade der Autonomie in Bezug auf die Level der Automation. Bei den Leveln zwei bis vier besitzt der Agent noch keine Autonomie. Dies liegt daran, dass der Agent in diesen Leveln kaum eigenständige Hilfestellungen bietet. Ab dem Level fünf bis Level sechs wird von teilweiser Autonomie gesprochen. Die Level fünf und sechs bieten keine vollständige Autonomie, weil immer noch auf eine Bestätigung des Menschen gewartet werden muss, bevor die Aufgabe ausgeführt werden kann. Es kann erst ab den Leveln sieben bis zehn von Autonomie gesprochen werden, wobei das zehnte Level den höchsten Grad der Autonomie darstellt.

1.4 Potenziale und Risiken des Human Autonomy Teaming

Dieses Unterkapitel veranschaulicht, wann und warum es sinnvoll ist, Human Autonomy Teaming in die Arbeitswelt zu integrieren. Es wird verdeutlicht, welche Gründe für die Zusammenarbeit von Menschen und autonomen Agenten sprechen, welche Aspekte bei der Integration des HAT von besonderer Bedeutung sind und welche Auswirkungen das HAT haben kann. Des Weiteren werden die Kategorien des Gebrauchs von Automation nach Arbeiten von Parasuraman und Riley (1997) definiert und die damit verbundenen Risiken im HAT erläutert. Ebenfalls thematisiert werden die positiven Effekte des HAT.

1.4.1 Sinnhaftigkeit des Human Autonomy Teaming

Allgemein wird in der Arbeitswelt viel mit Teams gearbeitet, um zusammen Aufgaben zu bewältigen und ein gemeinsames Ziel zu erreichen. Mit der Weiterentwicklung von Automation zu autonomen Agenten gibt es eine neue Art von Teamarbeit, die sich ebenfalls als effektiv erweist (Salas et al., 2008). Das Human Autonomy Teaming lässt sich nicht überall sinnvoll einsetzen. Es sollte nur in Bereichen integriert werden, in denen es förderlich ist, Bereiche der menschlichen Arbeit auf einen autonomen Agenten zu übertragen. Dies ist unter anderem beim Militär oder in der Luft- und Raumfahrt der Fall (Parasuraman et al., 2000).

Für die Zusammenarbeit von Menschen und autonomen Agenten spricht die Annahme, dass Menschen flexibler, anpassungsfähiger und kreativer als autonome Agenten sind. Dies ermöglicht es den Menschen besser auf sich verändernde oder unvorhersehbare Bedingungen reagieren zu können, als autonome Agenten (Parasuraman & Riley, 1997). Des Weiteren können Menschen besser auf ungewöhnliche Situationen, schlecht strukturierte Probleme oder Situationen mit unvollständigen Informationen reagieren (Shively et al., 2018). Autonome Agenten hingegen können Aufgaben effizienter, zuverlässiger und präziser bewältigen, als es Menschen möglich ist (Parasuraman & Riley, 1997). Durch diese Aspekte wird deutlich, dass sowohl die Menschen als auch autonome Agenten spezifische Stärken aufweisen, dessen Potenziale bei einer Zusammenarbeit von Menschen und autonome Agenten in einem Team genutzt werden können.

Bei der Einführung eines Human Autonomy Teamings gibt es einige Dinge, die beachtet werden müssen, um die beiderseitigen Vorteile gemeinsam ausschöpfen zu können. Dazu zählen unter anderem die Vorhersagbarkeit der nachfolgenden Aktionen, die Berechenbarkeit und die Zusammenarbeit (Klein et al., 2004). Ein wichtiger Faktor der Zusammenarbeit ist die bidirektionale, also in beide Richtungen funktionierende Kommunikation. Im HAT ist zu beachten, dass der autonome Agent mit dem Menschen und nicht für diesen arbeiten soll und es sich dementsprechend nicht um ein einfaches Hilfswerkzeug handelt. Es soll sich bei der Automation um ein Teammitglied handeln, das keiner permanenten Überwachung eines Menschen unterliegt. Dafür benötigt die Automation einen hohen Grad an Autonomie (Shively et al., 2018). Die Automation soll nicht die menschlichen Aufgaben übernehmen, wodurch der

Mensch nur noch eine leitende oder überwachende Person des autonomen Agenten darstellt, sondern es soll ein Team aus Menschen und autonomen Agenten gebildet werden (Roth et al., 1997). Menschen müssen dem autonomen Agenten vertrauen und ihn als Teammitglied betrachten und dementsprechend Toleranz gegenüber der Automation zeigen, damit ein HAT sinnvoll eingesetzt werden kann. Existiert keine Toleranz für den Gebrauch von Automation und somit für die Übernahme von Aufgabenbereichen durch den autonomen Agenten, wird dieser nicht genutzt und der Einsatz von HAT ist nicht zielführend (Parasuraman & Riley, 1997). Die beiden Teammitglieder treffen Entscheidungen auf der Basis unterschiedlicher Informationen. Sie können diese zusammentragen und somit gemeinsam ein besseres Ergebnis erzielen als es der Fall wäre, wenn beide alleine arbeiten würden (Wilson & Daugherty, 2018). Sowohl der Mensch als auch der autonome Agent können im HAT durch die Zusammenarbeit ihr gesamtes Potenzial ausschöpfen und weiterentwickeln. Besondere Berücksichtigung bedarf das Design des autonomen Agenten und dessen Verankerung in Teams. Wichtige Aspekte bei der Implementierung des HAT sind:

- der Autonomiegrad,
- die Verlässlichkeit aller Teammitglieder,
- die individuellen Unterschiede,
- das Training der Zusammenarbeit
- und die Transparenz.

Beidseitige Transparenz ist wichtig, damit der Mensch den Entscheidungsprozess des autonomen Agenten nachvollziehen kann und umgekehrt, sodass sich die Teammitglieder gegenseitig verstehen und miteinander arbeiten können (Schelble et al., 2020). Damit ein HAT funktioniert und es sinnvoll eingesetzt werden kann, müssen Menschen den Umgang und die Arbeit mit dem autonomen Agenten trainieren (Shively et al., 2018). Beim Training entsteht auf der Seite des Menschen ein Verständnis für die Arbeit und Handlungsweisen des autonomen Agenten. Eine Möglichkeit, um dies zu erreichen, ist, dass der Mensch innerhalb einer Simulation Aufgaben übernimmt, die normalerweise der autonome Agent ausführt. Dadurch kann der Entscheidungsprozess des autonomen Agenten nachvollzogen und die spätere Zusammenarbeit mit diesem erleichtert werden (Nikolaidis et al., 2015). Wenn das HAT falsch eingesetzt oder angewandt wird, kann das zu Problemen in den Abläufen führen (Schelble et al., 2020). Fehlerhafte Automation kann negative Auswirkungen auf die Zufriedenheit, Leistung und Sicherheit haben. Ist sie jedoch so ausgerichtet, dass sie die Anforderungen für ein HAT erfüllt, können in Zusammenarbeit mit ihr Leistungen erbracht werden, welche die kognitiven und physischen Fähigkeiten eines Menschen weit überschreiten (Roth et al., 1987). Automatisierte Systeme werden in der Regel für vorhersehbare Bedingungen entwickelt. Dennoch kann es vorkommen, dass sich ein System in einer unbekanntem Situation befindet, die es nicht bewältigen kann. Dementsprechend wird zur Lösung komplexer Probleme kreative menschliche Hilfe benötigt und der Mensch übernimmt in diesen Situationen die Kontrolle (Endsley & Kiris, 1995). Der Einsatz von autonomen Agenten kann auch Negativeffekte haben. Diese sind z.B. der Verlust des Situationsbewusstseins oder ein Rückgang von Fähigkeiten und Fertigkeiten beim Menschen, die früher einmal vorhanden waren und nach der Übertragung der Tätigkeiten auf Maschinen verloren gehen (Parasuraman et al., 2000). Die Formung eines autonomen Agenten zu einem guten Teammitglied kann von eingeschränkter Intelligenz (Klein et al., 2004), eingeschränkten Interaktionsmöglichkeiten und einer schwierigen Umgebung des Einsatzes beeinträchtigt werden (Demir et al., 2019). Die Voraussetzungen für ein funktionierendes HAT und für den Einsatz eines autonomen Agenten als Teammitglied werden in Kapitel 2 genauer erläutert.

Der Einsatz des HAT ist in komplexen und kritischen Situationen, in denen besonders schnell Entscheidungen getroffen werden müssen, sinnvoll. Menschen sind teilweise nicht dazu in der Lage sehr schnelle Entscheidungen zu treffen und sofort zu handeln. Entscheidungen werden von vollständig autonomen Agenten ohne jegliche menschliche Einflussnahme getroffen, wenn zu wenig Zeit zur Verfügung steht, um einen Menschen entscheiden zu lassen. Wenn eine Handlung unverzüglich erforderlich ist, um z.B. Unfälle zu vermeiden, agiert der autonome Agent ohne auf Befehle des Menschen zu warten (Parasuraman et al., 2000). Letztendlich können sowohl Menschen als auch autonome Agenten in gegenseitiger Abhängigkeit durch ihre Interaktion miteinander gute Arbeitsleistungen bringen (Endsley, 2015). Der Einsatz des HAT hat sich in Bereichen wie der Medizin, dem Militär oder der Luft- und Raumfahrt sehr bewährt (Salas et al., 2008). Mehrere Beispiele für den Einsatz von HAT in verschiedenen Praxisbereichen werden in Kapitel 4 betrachtet.

1.4.2 Der Gebrauch von Automation und die damit verbundenen Risiken im Human Autonomy Teaming

Automation kann als ein Hilfswerkzeug angesehen werden, bzw. in hohen Leveln der Automation sogar als autonomes Teammitglied, wie es im HAT der Fall ist. Deshalb ist der Gebrauch von Automation für das Thema HAT relevant. Es gibt verschiedene Möglichkeiten zur Nutzung der Automation. Diese kann von zahlreichen Faktoren abhängen. Dazu zählt beispielsweise die subjektive Einstellung der Menschen gegenüber der Automation, die meist von der Verlässlichkeit und Fehlerfreiheit der Automation abhängt (Parasuraman & Riley, 1997). Die Einstellung kann innerhalb einer Gesellschaft stark variieren (Helmreich, 1984). Fällt diese basierend auf vorheriger Erfahrung positiv aus, neigt der Mensch dazu, die Automation zu nutzen. Fällt die Einstellung jedoch eher negativ aus, tendiert der Mensch zur Nichtnutzung der Automation, denn soziale Interaktionen mit anderen Menschen werden bevorzugt. Ein Alltagsbeispiel hierfür ist die Nutzung des Bankautomaten alternativ zum Bankschalter. Die Entscheidung für oder gegen eine Nutzung der Automation hat Konsequenzen. Deshalb ist es wichtig zu entscheiden, wann und zu welchem Zweck Automation eingesetzt wird (Parasuraman & Riley, 1997).

Gebrauch (Use)

Der Gebrauch von Automation lässt sich in unterschiedliche Kategorien unterteilen. Parasuraman und Riley (1997) unterscheiden dabei Gebrauch, Nichtgebrauch, Fehlgebrauch und Missbrauch (Use, Misuse, Disuse und Abuse). Der Gebrauch (Use) wird als freiwillige Aktivierung oder Loslösung von der Automation durch den Menschen definiert. Wann eine Automation genutzt wird, kann von verschiedenen Faktoren abhängen, die innerhalb der Individuen variieren und worüber keine allgemeingültigen Aussagen getroffen werden können. Ein Faktor, der oft im Zusammenhang mit dem Gebrauch der Automation steht, ist der Arbeitsaufwand. Einige Menschen tendieren dazu, die Automation eher zu nutzen, wenn ihr Arbeitsaufwand besonders hoch ist, andere jedoch nicht. Die Müdigkeit, bzw. Erschöpfung durch die Arbeitstätigkeit kann ebenfalls ein Faktor sein, sich für die Nutzung einer Automation zu entscheiden (Riley, 1994). Ein weiterer Faktor kann das Vertrauen in die Automation sein. Menschen neigen dazu, verlässliche automatisierte Systeme nicht zu nutzen, wenn sie diesen nicht vertrauen. Im Gegensatz dazu können Menschen weiterhin auf eine Automation vertrauen, obwohl sie nicht korrekt arbeitet (Parasuraman & Riley, 1997). Das Thema Vertrauen im Bereich der Automation wird in Kapitel 2.1 genauer betrachtet. Insgesamt kann festgehalten werden, dass die Entscheidung, eine Automation zu nutzen oder darauf zu verzichten auf der Interaktion verschiedener Faktoren beruht (Parasuraman & Riley, 1997).

Fehlgebrauch (Misuse)

Der Fehlgebrauch (Misuse) von Automation liegt laut Parasuraman und Riley (1997) dann vor, wenn ein blindes Vertrauen in die Automation herrscht und diese unreflektiert und meist zu häufig eingesetzt wird. Ein Beispiel dafür ist, dass es zu einem Flugzeugabsturz kurz vor der Landebahn während des Landeanflugs kam, da die flugzeugführende Person zu wenig auf ihre eigenen Fähigkeiten und zu stark auf den Autopiloten vertraut hat. Deshalb konnte sie die Fluggeschwindigkeit während eines nächtlichen Schneesturms nicht kontrollieren (National Transportation Safety Board, 1994). Gründe für den Fehlgebrauch von Automation sind verschiedene menschliche Entscheidungsfehler, wie die Entscheidungsverzerrung und fehlerhafte Kontrollen. Entscheidungsverzerrungen können bei der Nutzung von Heuristiken zustande kommen. Dabei wird die erste Handlungsalternative gewählt, die einen zufriedenstellenden Nutzen verspricht (Tversky & Kahneman, 1974). Die Nutzung von Heuristiken ist ein analytisches Vorgehen, um mit begrenztem Wissen zu praktikablen Lösungen zu kommen. Diese Heuristiken sollen bei Entscheidungsfindungen oder Problemlösungen den kognitiven Aufwand reduzieren und somit den Entscheidungsprozess vereinfachen (Wickens, 1992). Sie können jedoch zu fehlerhaften Entscheidungen führen (Tversky & Kahneman, 1974). Je nachdem, welches Level der Automation vorliegt, kann es zu Kontrollfehlern des Menschen kommen. Insbesondere bei hohen Leveln der Automation, wie dies beispielsweise im HAT der Fall ist, kann es vermehrt zu Kontrollfehlern kommen. Dies liegt daran, dass unter anderem die Zahl an automatisierten Subsystemen, Alarmen und Entscheidungshilfen steigt (Reason, 1990). Weitere Probleme, die durch einen Fehlgebrauch von Automation auftreten können, sind laut Lee und Moray (1992) der Verlust von Fähigkeiten und laut Endsley (1996) der Verlust des Situationsbewusstseins (vgl. Kapitel 3.3.3). Der Fehlgebrauch von Automation kann zu einem Teufelskreis führen. Eine übermäßige Nutzung der Automation, bedingt durch mangelnde Fähigkeiten oder einem zu geringen Vertrauen in die eigenen Fähigkeiten, kann dazu führen, dass der Mensch seine Fähigkeiten nicht weiter trainiert, sondern diese vielmehr abgebaut werden. Das erhöht sukzessiv den Grad der Automationsnutzung, wodurch sich die Automation nutzende Person in einem Teufelskreis befindet (Parasuraman & Riley, 1997).

Nichtnutzung (Disuse)

Die Nichtnutzung oder geringe Nutzung der Automation wird mit dem Begriff Disuse bezeichnet. Hierbei führt ein zu geringes Vertrauen in die Automation dazu, dass diese nicht genutzt wird. Das geringe Vertrauen wird dadurch verursacht, dass ein falscher Alarm durch die Automation zu häufig ausgelöst wird (Parasuraman & Riley, 1997). In der Folge tendieren Menschen in der Bedienung dazu, Alarme mit der Zeit zu ignorieren und auszustellen. Beispiele für die Nichtnutzung von Automation sind Bahnunfälle. Diese kamen unter anderem dadurch zu Stande, dass ein zu geringes Vertrauen in die Automation herrschte. Infolgedessen wurden die von der Automation gesetzten Geschwindigkeitsbegrenzungen und Alarmsignale bei Geschwindigkeitsüberschreitungen ignoriert und abgestellt (Sorkin, 1988).

Missbrauch (Abuse)

Missbrauch von Automation liegt laut Parasuraman und Riley (1997) dann vor, wenn Tätigkeiten von Entwickler/innen automatisiert werden, ohne darauf zu achten, wie sich dies auf

- die menschliche Arbeit,
- die Leistung des Systems
- und die Autorität des Menschen gegenüber der Automation auswirkt.

Die unüberlegte Einführung von Automation durch Manager/innen kann ebenfalls zum Missbrauch der Automation führen. Dies ist der Fall, wenn keine Vorüberlegungen zu den Auswirkungen der Einführung von Automation auf die menschliche Arbeit, die Leistung des Systems und die Autorität des Menschen gegenüber der Automation getroffen werden. Beim Missbrauch der Automation wird diese aus rein ökonomischen Gründen herangeführt. Die Automation wird lediglich dort eingesetzt, wo sie zu einem ökonomischen Vorteil führt. Dieser Vorteil kann dadurch erreicht werden, dass die Automation eine Aufgabe präziser oder zuverlässiger, als ein Mensch ausführt oder dass der Mensch durch die Automation kostengünstiger ersetzt wird. Entwickler/innen und Manager/innen, die den Missbrauch von Automation verursachen, achten nicht darauf, inwiefern dem Menschen die Automation nutzen wird oder in welchem Umfang sich die Arbeitsaufgaben des Menschen verändern werden (Parasuraman & Riley, 1997). Diese Entwickler/innen und Manager/innen verfolgen das Ziel einer Reduktion der menschlichen Fehleranfälligkeit durch den vermehrten Einsatz von Automation (Hancock & Parasuraman, 1992). Jedoch können menschliche Fehler nicht dadurch beseitigt werden, dass die Menschen durch Automation ersetzt werden. Bei der Entwicklung und Programmierung von Automation können ebenfalls menschliche Fehler auftreten, da diese Tätigkeiten durch Menschen ausgeführt werden. Dies hat zur Folge, dass die Automation nicht zweckmäßig eingesetzt werden kann. Insgesamt kann der Missbrauch der Automation zum Fehlgebrauch (Misuse) oder Nichtgebrauch (Disuse) der Automation durch den Menschen führen (Parasuraman & Riley, 1997). Die vier verschiedenen Kategorien des Gebrauchs von Automation werden in der untenstehenden Infobox 1.2 verkürzt dargestellt.

Metakontrollproblem

Der Gebrauch von Automation unterliegt Kontrollproblemen auf zwei unterschiedlichen Ebenen. Einerseits existiert die Top-Down-Kontrolle, bei der Entwickler/innen und Manager/innen den Einsatz der Automation vorgeben. Andererseits besteht die Bottom-up-Kontrolle, bei der die menschlichen Teammitglieder entscheiden, wann und wie Automation genutzt wird. Beide Kontrollebenen müssen ausbalanciert werden, damit Automation in sinnvoller Weise eingesetzt werden kann. Dies erweist sich aufgrund mangelnder Kommunikation zwischen den Entwickler/innen, bzw. Manager/innen und den menschlichen Teammitgliedern, sowie deren unterschiedlichen Vorstellungen der Automationsnutzung als schwierig (Lee, 2008).

Infobox 1.2: Übersicht zu den Gebrauchsarten von Automation nach Parasuraman und Riley (1997).

	Gebrauch (Use)	Fehlgebrauch (Misuse)	Nichtnutzung (Disuse)	Missbrauch (Abuse)
Definition	<ul style="list-style-type: none"> • Freiwillige Aktivierung oder • Loslösung von Automation durch den Menschen 	<ul style="list-style-type: none"> • blindes Vertrauen in die Automation • Einsatz unreflektiert und meist zu viel 	<ul style="list-style-type: none"> • Nichtnutzung oder • geringe Nutzung von Automation 	Einsatz von Automation ohne Acht auf <ul style="list-style-type: none"> • die Auswirkungen auf die Arbeit, • die Leistung des Systems und • die Autorität des Menschen gegenüber der Automation zu geben
Gründe	Verschiedene Faktoren, bspw. <ul style="list-style-type: none"> • Arbeitsaufwand • Erschöpfung • Vertrauen 	<ul style="list-style-type: none"> • Entscheidungsverzerrungen, • zu geringes Vertrauen in die eigenen Fähigkeiten 	<ul style="list-style-type: none"> • Zu geringes Vertrauen in die Automation, da ein falscher Alarm zu häufig durch diese ausgelöst wird 	Es sollen <ul style="list-style-type: none"> • ein ökonomischer Vorteil geschaffen und • menschliche Fehler verringert werden
Folgen	<ul style="list-style-type: none"> • ein falscher Gebrauch von Automation kann zu Misuse, Disuse und Abuse führen 	<ul style="list-style-type: none"> • Verlust von Situationsbewusstsein • Verlust von Fähigkeiten → Teufelskreis 	<ul style="list-style-type: none"> • Menschen ignorieren Alarmer und stellen sie aus 	<ul style="list-style-type: none"> • Gefährdung der Automation durch Entwicklungsfehler und • Auslösen von Misuse oder Disuse der Automation

1.4.3 Positive Effekte des Human Autonomy Teaming

Viele Menschen assoziieren im Kontext zunehmender Automatisierung das weit verbreitete Vorurteil, der Mensch werde dadurch in der Arbeitswelt zunehmend ersetzt. Human Autonomy Teaming zeigt jedoch, dass dies nicht der Realität entspricht. Die menschliche Arbeit wird in dem Zusammenhang nicht substituiert, sondern vielmehr durch die Zusammenarbeit mit einem autonomen Agenten verändert. Die Veränderung kommt dadurch zustande, dass sich der Mensch an die neuen Möglichkeiten, die durch den Einsatz von Automation entstehen, anpasst. Die Automatisierung physischer Tätigkeiten kann Menschen von vielen zeit- und arbeitsintensiven Tätigkeiten befreien (Parasuraman & Riley, 1997). Die Arbeit wird nicht nur durch die Umverteilung von Aufgabenbereichen verändert, sondern es ergeben sich ebenfalls neue Arbeitsfelder innerhalb der Entwicklung von autonomen Agenten. Das Design solcher Agenten ist eine wichtige Aufgabe, damit das HAT funktionieren kann, die autonomen Agenten sinnvoll eingesetzt werden können und somit Vorteile aus der Zusammenarbeit von Menschen

und autonomen Agenten erzielt werden können. Zu diesen Vorteilen zählen unter anderem eine gesteigerte Effizienz, verbesserte Sicherheit, erhöhte Flexibilität in den Tätigkeiten und ein geringerer Arbeitsaufwand für den Menschen (Parasuraman & Manzey, 2010). Das HAT ist eine sinnvolle Methode, um die Möglichkeiten, die durch Automation geboten werden, zu nutzen. Durch die Zusammenarbeit mit autonomen Agenten können die Stärken von Automation optimal eingesetzt werden (Shively et al., 2018). Zu diesen Stärken zählen unter anderem eine effizientere, zuverlässigere und präzisere Aufgabenbewältigung, als es Menschen möglich ist (Parasuraman & Riley, 1997). Das autonome und menschenähnliche Verhalten der autonomen Agenten, die steigende Geschwindigkeit und verbesserte Qualität von Automation, sowie die Vorteile, die die Anwendung von HAT mit sich bringt, könnten eine zunehmende Akzeptanz in der Gesellschaft für Automation schaffen (Shively et al., 2018). Diese Akzeptanz wird ebenfalls dadurch verursacht, dass Menschen merken, dass sie im HAT nicht durch Automation ersetzt werden, sondern weiterhin einen wichtigen Teil zur Zielerreichung beitragen. In diesem Team können sowohl die Stärken des Menschen als auch die des autonomen Agenten gemeinsam genutzt werden, wodurch die Effektivität im Team steigt (Schelble et al., 2020). Das HAT bietet noch weitere Möglichkeiten, die in Kapitel 3 thematisiert werden. Des Weiteren wird ein Blick in die Zukunft des HAT in Kapitel 5 geworfen. In dem nächsten Kapitel, Kapitel 2, werden die Voraussetzungen, auf Seiten des Menschen und auf Seiten des autonomen Agenten, für ein funktionierendes HAT behandelt.

Literaturverzeichnis

- Bainbridge, L. (1983). Ironies of automation. *Automatica*, 19(6), 775–779.
- Buchanan, B. G. (2005). A (very) brief history of Artificial Intelligence. *AI Magazine*, 26(4), 53–60.
- Butschek, F. (2006). *Industrialisierung. Ursachen, Verlauf, Konsequenzen*. Köln: Böhlau, UTB GmbH.
- Chen, J. Y. C., Barnes, M. J. & Harper-Sciarini, M. (2011). Supervisory control of multiple robots: Human-performance issues and user-interface design. *IEEE Transactions on Systems, Man, and Cybernetics*, 41(4), Part C, 435–454.
- Chui, M., Manyika, J. & Miremadi, M. (2016). Where machines could replace humans—and where they can't (yet). *McKinsey Quarterly*, 30(2), 1–12.
- Cooke, N. J., Demir, M. & McNeese, N. (2016). *Synthetic teammates as team players: Coordination of human and synthetic teammates* (Report No. RE2016844 01). Cognitive Engineering Research Institute. Mesa, United States.
- Demir, M., Cooke, N. J. & Amazeen, P. G. (2018). A conceptual model of team dynamical behaviors and performance in Human-Autonomy Teaming. *Cognitive Systems Research*, 52, 497–507.
- Demir, M., Likens, A. D., Cooke, N. J., Amazeen, P. G. & McNeese, N. J. (2019). Team coordination and effectiveness in Human-Autonomy Teaming. *IEEE Transactions on Human-Machine Systems*, 49(2), 150–159.
- Endsley, M. R. (1996). Automation and situation awareness. In R. Parasuraman & M. Mouloua (Eds.), *Automation and human performance: Theory and applications* (pp. 163–181). NJ: Erlbaum.
- Endsley, M. R. (2015). *Autonomous horizons: System autonomy in the air force - A path to the future* (Report No. AF/ST TR 15-01). Washington, United States
- Endsley, M. R. (2017). From here to autonomy: Lessons learned from human–automation research. *Human Factors*, 59(1), 5–27.
- Endsley, M. R. & Kiris, E. O. (1995). The out-of-the-loop performance problem and level of control in automation. *Human Factors*, 37(2), 381–394.
- Guarnieri, M. (2010). The roots of automation before mechatronics [historical]. *IEEE Industrial Electronics Magazine*, 4(2), 42–43.
- Hancock, P. A., Jagacinski, R. J., Parasuraman, R., Wickens, C. D., Wilson, G. F. & Kaber, D. B. (2013). Human-Automation Interaction research. *Ergonomics in Design: The Quarterly of Human Factors Applications*, 21(2), 9–14.
- Hancock, P. A. & Parasuraman, R. (1992). Human factors and safety in the design of intelligent vehicle-highway systems (IVHS). *Journal of Safety Research*, 23(4), 181–198.
- Helmreich, R. L. (1984). Cockpit management attitudes. *Human Factors*, 26(5), 583–589.

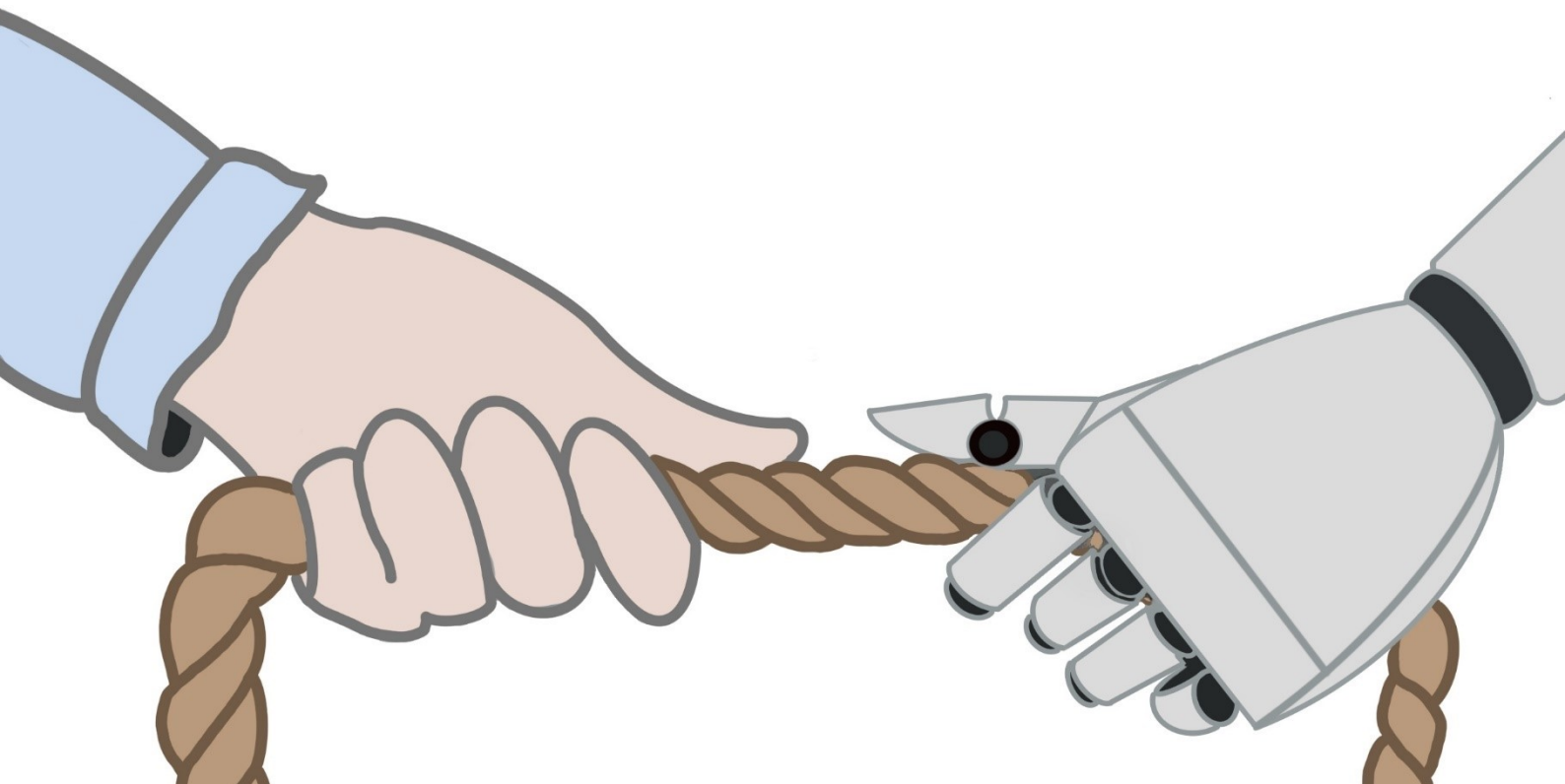
- Hitomi, K. (1994). Automation – its concept and a short history. *Technovation*, 14(2), 121–128.
- Janssen, C. P., Donker, S. F., Brumby, D. P. & Kun, A. L. (2019). History and future of Human-Automation Interaction. *International Journal of Human-Computer Studies*, 131, 99–107.
- Johnson, M., Bradshaw, J. M., Feltovich, P. J., Jonker, C., van Riemsdijk, B. & Sierhuis, M. (2012). Autonomy and interdependence in human-agent-robot teams. *IEEE Intelligent Systems*, 27(2), 43–51.
- Kaber, D. B. & Endsley, M. R. (2004). The effects of level of automation and adaptive automation on human performance, situation awareness and workload in a dynamic control task. *Theoretical Issues in Ergonomics Science*, 5(2), 113–153.
- Klein, G., Woods, D. D., Bradshaw, J. M., Hoffman, R. R. & Feltovich, P. J. (2004). Ten challenges for making automation a “team player” in joint human-agent activity. *IEEE Intelligent Systems*, 19(6), 91–95.
- Lee, J. D. (2008). Review of a pivotal Human Factors article: “Humans and automation: Use, misuse, disuse, abuse”. *Human Factors*, 50(3), 404–410.
- Lee, J. D. & Moray, N. (1992). Trust, control strategies and allocation of function in human-machine systems. *Ergonomics*, 35(10), 1243–1270.
- Lee, J. D. & See, K. A. (2004). Trust in automation: Designing for appropriate reliance. *Human Factors*, 46(1), 50–80.
- Liedtke, R. (2012). *Die Industrielle Revolution*. Köln: Böhlau, UTB GmbH.
- Lyons, J. B., Mahoney, S., Wynne, K. T. & Roebke, M. A. (2018). Viewing machines as teammates: A qualitative study. *AAAI Spring Symposium Series* (pp. 166–170). Palo Alto, California.
- Manzey, D. (2008). Systemgestaltung und Automatisierung. In P. Badke-Schaub, G. Hofinger & K. Lauche (Hrsg.), *Human Factors – Psychologie sicheren Handelns in Risikobranchen* (S. 307–321). Heidelberg: Springer Medizin Verlag.
- Manzey, D., Reichenbach, J. & Onnasch, L. (2008). Performance consequences of automated aids in supervisory control: The impact of function allocation. *Proceedings of the Human Factors and Ergonomics Society Annual Meeting*, 52(4), 297–301.
- 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), 262–273.
- Megill, J. (2014). Emotion, cognition and Artificial Intelligence. *Minds and Machines*, 24(2), 189–199.
- National Transportation Safety Board. (1994). *Aircraft accident report: Stall and loss of control on final approach* (Report No. NTSB-AAR-94/07). Washington, DC.
- Neef, A. (2016). *Kognitive Maschinen – Wie Künstliche Intelligenz die Wertschöpfung transformiert*. Zugriff am 08.01.2021. Verfügbar unter https://www.z-punkt.de/uploads/files/503/kognitive_maschinen.pdf
- Nikolaidis, S., Lasota, P., Ramakrishnan, R. & Shah, J. (2015). Improved human-robot team performance through cross-training, an approach inspired by human team training practices. *The International Journal of Robotics Research*, 34(14), 1711–1730.

- O'Neill, T. A., McNeese, N. J., Barron, A. & Schelble, B. (2020). Human–Autonomy Teaming: A review and analysis of the empirical literature. *Human Factors*. Advance Online Publication. <https://doi.org/10.1177/0018720820960865>
- Onnasch, L. (2015). Crossing the boundaries of automation – Function allocation and reliability. *International Journal of Human-Computer Studies*, 76, 12–21.
- Onnasch, L., Wickens, C. D., Li, H. & Manzey, D. (2014). Human performance consequences of stages and levels of automation: An integrated meta-analysis. *Human Factors*, 56(3), 476–488.
- Parasuraman, R. & Manzey, D. H. (2010). Complacency and bias in human use of automation: An attentional integration. *Human Factors*, 52(3), 381–410.
- Parasuraman, R. & Riley, V. (1997). Humans and automation: Use, misuse, disuse, abuse. *Human Factors*, 39(2), 230–253.
- Parasuraman, R., Sheridan, T. B. & Wickens, C. D. (2000). A model for types and levels of human interaction with automation. *IEEE Transactions on Systems, Man, and Cybernetics. Part A, Systems and Humans (IEEE T SYST MAN CY A)*, 30(3), 286–297.
- Reason, J. T. (1990). *Human error*. Cambridge: Cambridge University Press.
- Riley, V. (1994). *Human use of automation*. Unpublished doctoral dissertation, University of Minnesota.
- Roth, E. M., Bennett, K. B. & Woods, D. D. (1987). Human interaction with an „intelligent“ machine. *International Journal of Man-Machine Studies*, 27(5-6), 479–525.
- Roth, E. M., Malin, J. T. & Schreckenghost, D. L. (1997). Paradigms for intelligent interface design. In M. G. Helander, T. K. Landauer & P. V. Prabhu (Eds.), *Handbook of Human-Computer Interaction* (pp. 1177–1201). North Holland: Elsevier.
- Russell, S. J. & Norvig, P. (2009). *Artificial Intelligence. A modern approach* (Prentice Hall series in artificial intelligence). Upper Saddle River: Prentice Hall.
- Salas, E., Cooke, N. J. & Rosen, M. A. (2008). On teams, teamwork, and team performance: Discoveries and developments. *Human Factors*, 50(3), 540–547.
- Schelble, B. G., Flathmann, C. & McNeese, N. (2020). Towards meaningfully integrating Human-Autonomy Teaming in applied settings. In M. Obaid, O. Mubin, Y. Nagai, H. Osawa, Y. Abdelrahman & M. Fjeld (Eds.), *Proceedings of the 8th International Conference on Human-Agent Interaction* (pp. 149–156). New York: ACM.
- Sheridan, T. B. (2016). Human–robot interaction: Status and challenges. *Human Factors*, 58(4), 525–532.
- Sheridan, T. B. & Johannsen, G. (1976). *Monitoring behavior and supervisory control* (NATO Conference Series, Vol. 1). Boston, MA: Springer.
- Sheridan, T. B. & Parasuraman, R. (2005). Human-Automation Interaction. *Reviews of Human Factors and Ergonomics*, 1(1), 89–129.
- Sheridan, T. B. & Verplank, W. L. (1978). *Human and computer control of undersea teleoperators* (Forschungsbericht). Cambridge Massachusetts: Institute of Technology.

- Shively, R.J., Lachter, J., Brandt, S.L., Matessa, M., Battiste, V. & Johnson, W.W. (2018). Why Human-Autonomy Teaming? In C. Baldwin (Ed.), *Advances in Neuroergonomics and Cognitive Engineering. AHFE 2017. Advances in Intelligent Systems and Computing* (586). Cham: Springer International Publishing.
- Sorkin, R. D. (1988). Forum: Why are people turning off our alarms? *The Journal of the Acoustical Society of America*, 84(3), 1107–1108.
- Tversky, A. & Kahneman, D. (1974). Judgment under uncertainty: Heuristics and biases. *Science*, 185(4157), 1124–1131.
- Walliser, J. C., Mead, P. R. & Shaw, T. H. (2017). The perception of teamwork with an autonomous agent enhances affect and performance outcomes. *Proceedings of the Human Factors and Ergonomics Society Annual Meeting*, 61(1), 231–235.
- Wickens, C. D. (1992). *Engineering psychology and human performance* (2. Aufl.). New York: HarperCollins.
- Wickens, C. D., Hollands, J. G., Banbury, S. & Parasuraman, R. (2016). *Engineering psychology and human performance* (4. Aufl.). London, New York: Routledge Taylor & Francis Group.
- Wilson, H. J. & Daugherty, P. R. (2018). Collaborative intelligence: Humans and AI are joining forces. *Harvard Business Review*, 96(4), 114–123.
- Wright, J. L., Chen, J. Y. C., Quinn, A. & Barnes, M. J. (2013). *The effects of level of autonomy on human-agent teaming for multi-robot control and local security maintenance* (Report No. ARL-TR-6724). Army Research Laboratory.
- Wynne, K. T. & Lyons, J. B. (2018). An integrative model of autonomous agent teammate-likeness. *Theoretical Issues in Ergonomics Science*, 19(3), 353–374.
- Ziegler, D. (2012). *Die Industrielle Revolution*. In U. Puschner (Hrsg.), *Geschichte kompakt* (3. bibliogr. aktual. Aufl.). Darmstadt: WBG.

Anforderungen für erfolgreiches Human Autonomy Teaming

Vincent Kapp, Jessica Schröder & Benedikt Wiemann



Inhaltsverzeichnis

2	Anforderungen für erfolgreiches Human Autonomy Teaming	
2.1	Anforderungen aufseiten der Menschen als Teil von HAT	38
	2.1.1 Vertrauen in Automation und autonome Agenten	
	2.1.2 Akzeptanz	
	2.1.3 Systemwissen	
2.2	Anforderungen aufseiten des autonomen Agenten als Teil von HAT	41
	2.2.1 Programmierung und Design des autonomen Agenten	
	2.2.2 Anpassungsfähigkeit des autonomen Agenten	
2.3	Zusammenführung der Anforderungen aufseiten der Menschen und des autonomen Agenten als Team	50
	2.3.1 Bidirektionale Kommunikation	
	2.3.2 Transparenz	
	2.3.3 Koordination und Teamkognition	
	Literaturverzeichnis	67

Kapitel 2: Anforderungen für erfolgreiches Human Autonomy Teaming (HAT)

Von Vincent Kapp, Jessica Schröder & Benedikt Wiemann

In diesem Kapitel werden die Anforderungen für erfolgreiches Human Autonomy Teaming (HAT) behandelt. Als Erstes werden die Anforderungen aufseiten der beiden Parteien von HAT, Mensch und autonomer Agent, einzeln betrachtet. In dem ersten Unterkapitel 2.1 werden die Anforderungen an den Menschen thematisiert. Dazu gehören das Vertrauen sowohl in die Automation als auch in autonome Agenten und das Akzeptieren der Automation und des Agenten durch die Menschen. Das zweite Unterkapitel 2.2 stellt die Anforderungen an die autonomen Agenten dar. Hierbei handelt es sich einerseits um die Programmierung und das Design des autonomen Agenten, andererseits um die Anpassungsfähigkeit des Agenten an unterschiedliche Faktoren. Das erste und zweite Unterkapitel bilden eine Grundlage für das dritte Unterkapitel 2.3 und werden in diesem zusammengeführt. Dort werden die Anforderungen an die Menschen und an die autonomen Agenten gemeinsam betrachtet. Ferner werden die bidirektionale Kommunikation, die Transparenz sowie die Koordination und die Teamkognition thematisiert. Das dritte Unterkapitel 2.3 basiert auf den Erkenntnissen aus den ersten beiden Unterkapiteln und soll als Fazit für das gesamte Kapitel 2 gelten. Es wird auf die zukünftige aktiv stattfindende Kooperation eines Teams aus Menschen und autonomen Agenten eingegangen.

2.1 Anforderungen aufseiten der Menschen als Teil von HAT

Muir (1987) sagt, dass das Vertrauen zwischen Mensch und Maschine von den gleichen Faktoren beeinflusst wird, wie es bei menschlichen Beziehungen der Fall ist. Auf der einen Seite wird das Vertrauen zwischen zwei Menschen durch Ehrlichkeit und Zuverlässigkeit aufgebaut. Andererseits wird es ebenso durch Enttäuschung und Verrat verloren. Die Verlässlichkeit der Automation sowie das menschliche Vertrauen in die Automation gehören zu den entscheidenden Faktoren für erfolgreiche Zusammenarbeit (Muir, 1987). Im Folgenden werden die Herausforderungen und die Probleme des Vertrauensaufbaus erläutert und es wird darauf eingegangen, weshalb die bedienenden Personen die Automation und die autonomen Agenten unterschiedlich akzeptieren.

2.1.1 Vertrauen in Automation und autonome Agenten

Vertrauen ist für den Teamzusammenhalt essenziell, welcher sich wiederum positiv auf die Teamlebensfähigkeit auswirkt (O'Neill et al., 2020). Teamlebensfähigkeit beschreibt die Dauer des Bestehens des Teams. Nach Hackman (1987) entsteht Teamarbeit unter anderem aus Teamlebensfähigkeit. Diese wird durch weitere Aspekte wie das individuelle Lernen, die Entwicklung und Bedürfnisbefriedigung vervollständigt. Damit die Teamarbeit im HAT erfolgreich funktionieren kann, ist Vertrauen unerlässlich. Wie die unterschiedlichen Faktoren in Verbindung stehen, ist in Abbildung 2.1 dargestellt.

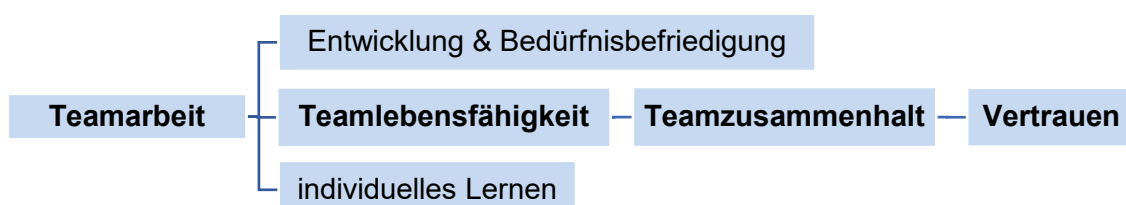


Abb. 2.1: Aufbau von Teamarbeit (in Anlehnung an Hackman, 1987).

Die Automation bietet ein großes Potenzial, um die Leistung und die Sicherheit des Menschen in seiner Tätigkeit zu verbessern. Wie in Kapitel 1.1.1 thematisiert wird, kann nach Parasuraman und Riley (1997) von Automation gesprochen werden, sobald eine technische Einheit eine Aufgabe übernimmt, die vorher von einem Menschen erledigt wurde. Jedoch kann es zum einen zum Misstrauen der Automation kommen, obwohl diese die Arbeitsbelastung drastisch verringert sowie die Leistung steigert. Zum anderen wird der Automation in bestimmten Bereichen, wo ein Eingreifen nötig wäre, zu sehr vertraut (Lee & See, 2004). Beispielsweise hat die flugzeugführende Belegschaft beim Absturz des Airbus A320-111 der Air France zu viel Vertrauen in das Fly-by-Wire System, welches für die Signalübertragung der Flugsteuerung sorgt, gesetzt und deshalb zu spät interveniert. Bei einem anderen Beispiel haben die Bedienenden die Nutzung vom Ground Proximity Warning System (GPWS), was bei zu niedriger Flughöhe Alarm schlägt, für nicht vertrauenswürdig eingestuft und somit nicht benutzt (Parasuraman & Riley, 1997).

In der Beziehung zwischen Mensch und Maschine wird das Vorkommen von „Misuse“ und „Disuse“ (Parasuraman & Riley, 1997, S. 230; vgl. Kapitel 1.4.2 für die Unterteilung zwischen Use, Abuse, Misuse und Disuse) wie folgt definiert. „Misuse“ (Parasuraman & Riley, 1997, S. 230) beschreibt den Fehlgebrauch von Automation, wenn ein blindes Vertrauen in die Automation herrscht und diese unflektiert und meist zu häufig eingesetzt wird. Dies war beim Absturz des Airbus A320-111 der Air France der Fall. „Disuse“ (Parasuraman & Riley, 1997, S. 230) beschreibt das Aufkommen von Fehlern, wenn der Mensch die Automation nicht oder gering nutzt, was z.B. beim Ground Proximity Warning System eingetreten ist. Unverhältnismäßiges Vertrauen beschreibt die Differenz zwischen dem Vertrauen in die Automation und den tatsächlichen Fähigkeiten dieser Automation. Um eine effiziente Zusammenarbeit zu gewährleisten, muss ein angemessenes Vertrauen unterstützt und realisiert werden (Wicks et al., 1999). Eine Kalibrierung in diesem Zusammenhang bezieht sich auf die Minimierung der Unstimmigkeiten zwischen dem Vertrauen und den Fähigkeiten der Automation (Muir, 1987). In Verbindung zu den definierten Begriffen „Misuse“ und „Disuse“ (Parasuraman & Riley, 1997, S. 230) kann einerseits eine schlechte Kalibrierung zu einem übermäßigen Vertrauen der Automation („Misuse“; Parasuraman & Riley, 1997, S. 230) führen. Andererseits können bei dem Misstrauen in die Automation („Disuse“; Parasuraman & Riley, 1997, S. 230) Fehler entstehen und/oder Fähigkeiten ungenutzt bleiben.

Bei der Arbeit mit einem autonomen Agenten wird gezeigt, dass die individuellen Unterschiede in der Persönlichkeit des Menschen einen Einfluss auf das Vertrauensverhältnis zwischen dem Agenten und der bedienenden Person haben (O'Neill et al., 2020). Bei autonomen Agenten, die eine ähnliche Arbeitsstilpräferenz wie die jeweilige Person aufweisen, kann das Vertrauen gestärkt werden und dazu führen, dass der Mensch den Agenten einem menschlichen Teammitglied vorzieht (You & Robert, 2017). Bei Personen, die individualistischer denken, ist das Vertrauen in die Agenten höher als bei jenen, die eher eine ungleiche Machtverteilung akzeptieren (Chien et al., 2016). Um das Vertrauen zwischen autonomen Agenten und Menschen aufzubauen, ist Transparenz ein wichtiger Faktor (Shively et al., 2017). Unter Transparenz wird die Bereitstellung von Informationen bezüglich der durchgeführten Handlungen beschrieben (Lyons, 2013). Lyons et al. (2016) und Sadler et al. (2016) zeigen, dass die bedienende Person durch Transparenz mehr Möglichkeiten bekommt, in das Geschehen einzugreifen, was somit das Vertrauen in die Automation oder den autonomen Agenten verstärkt. Für das Vertrauen stellt Transparenz einen von vielen wichtigen Bausteinen dar. Dieses Vertrauen darf aber nicht fehlgeleitet sein (Shively et al., 2017). Um Vertrauen in der Zusammenarbeit von Menschen und Agenten aufzubauen, müssen die drei Dimensionen:

- Nutzen,
- Vorhersagbarkeit und
- Absicht

berücksichtigt werden (Lees & Lee, 2007). Die Arbeit von Lees und Lee (2007) zeigt, dass bei transparenter Informationslage das Vertrauen angemessener ist im Vergleich dazu, wenn diese Transparenz nicht vorliegt. Sie stellen fest, dass mit zunehmendem Vertrauen die Überwachung der Agenten abnimmt. Sobald dem Menschen nicht die Möglichkeit gegeben ist, potenzielle Probleme zu erkennen, wie es bei mangelnder Transparenz der Fall ist, besteht ein starker Anreiz dazu, der Automation zu viel Vertrauen zu schenken. Endsley (2017) betont, dass mangelnde Transparenz und nicht kalibriertes Vertrauen keine unabhängigen Faktoren sind und diese in Kombination das System besonders anfällig für Fehler machen. Der Mensch trägt somit allein die Verantwortung, um beim Versagen des Systems einzugreifen. Die Person, die das System bedient, ist sich oftmals dem Systemzustand nicht bewusst, da angenommen wird, dass der Agent seine Aufgaben korrekt ausführt. Dieses Verhalten wird als reduziertes Situationsbewusstsein nach dem Out-of-the-Loop Effekt beschrieben (Endsley, 2017). In Kapitel 3.3.1 wird noch spezifischer auf die Entstehung von Vertrauen eingegangen.

2.1.2 Akzeptanz

Mit zunehmender Autonomie (vgl. Kapitel 1.1.1 für die Unterteilung von Automation und Autonomie) ist der Agent in der Lage, mit Menschen auf komplexe Weise zu interagieren. Dadurch wirkt sich das Verhalten des autonomen Agenten auf das Wohlbefinden des Menschen aus. Dementsprechend richtet sich der Grad, wie stark der Mensch den Agenten akzeptiert, danach aus, auf welche Art der Agent handelt (Bradshaw et al., 2004). Die Anpassung der Kommunikation im HAT ist für die Zusammenarbeit über einen längeren Zeitraum von großer Bedeutung. Kommunikation und der kulturelle Hintergrund sind bei der Interaktion zu berücksichtigen, um Agenten für den Menschen akzeptabel zu machen (Feltovich et al., 2004). Wenn die bedienenden Personen durch ihre Kultur in Berührung mit Computersystemen gekommen sind, weisen sie eine stärkere Akzeptanz gegenüber dem autonomen Agenten auf und das Vertrauen in diesen nimmt zu (Zuboff, 1988). Akzeptanz ist entscheidend, wenn der autonome Agent die vollständige Handlungsgewalt besitzt und der Mensch somit nicht mehr in der Lage ist einzugreifen. Ob der Mensch die Automation nutzt oder nicht, ist abhängig von der Akzeptanz, wie z.B. beim Nutzen eines autonomen Fahrzeugs (Endsley, 2017; Norman, 1997). Die hier beschriebenen Anforderungen aufseiten des menschlichen Teammitglieds innerhalb eines HAT werden zusammenfassend in Infobox 2.1 dargestellt.

2.1.3 Systemwissen

Eine weitere Anforderung aufseiten der Menschen besteht nach Rieth und Hagemann (2021) im notwendigen Systemwissen, welches einen weiteren oder veränderten Kompetenzbedarf für den menschlichen Teamplayer darstellt. Für ein erfolgreiches HAT muss der Mensch die allgemeine Systemlogik und die Systemprozesse verstehen. Die Begrifflichkeit „System“ steht an dieser Stelle stellvertretend für den autonomen Agenten. Der Mensch muss also den autonomen Agenten, einschließlich seiner Fähigkeiten und Grenzen, kennen, verstehen und interpretieren können sowie diese bei all seinen Handlungen korrekt berücksichtigen (Rieth & Hagemann, 2021). Dies steht im Einklang mit Battiste et al. (2018), die ebenfalls Systemkenntnisse aufseiten der Menschen fordern, damit der Mensch akkurate mentale Modelle bezüglich der Systemfunktionalitäten entwickeln kann. In der Literatur wird das Thema Systemwissen noch vergleichsweise selten thematisiert (Rieth & Hagemann, 2021). Fragen wie

beispielsweise, wie viel Wissen benötigt wird, wie Systemwissen am besten trainiert werden kann und wie stark sich dies auf die HAT-Leistung auswirkt, bleiben offen und sollten weiter erforscht werden.

Infobox 2.1: Anforderungen aufseiten des Menschen.

Anforderungen aufseiten des Menschen:

- **Vertrauen**
 - Vermeidung von Misuse und Disuse
 - Kalibrierung des Vertrauens in Anlehnung an die tatsächlichen Fähigkeiten
 - Übereinstimmung von persönlichen Eigenschaften und Arbeitsstil mit dem autonomen Agenten
 - Möglichkeit zum Eingreifen in Prozesse
 - Erkennung von Problemen
- **Akzeptanz**
 - Wohlbefinden im Umgang mit der Automation
 - Anpassung der Kommunikation zwischen Menschen und autonomen Agenten
 - Erfassung von mentalen Modellen
- **Systemwissen**
 - Verstehen von Systemlogik und Systemprozessen
 - Korrekte Berücksichtigung von Fähigkeiten und Grenzen des autonomen Agenten

2.2 Anforderungen aufseiten des autonomen Agenten als Teil von HAT

Nachdem zuvor die Anforderungen aufseiten des menschlichen Teammitglieds betrachtet wurden, folgen nun die Anforderungen, die aufseiten des technischen Teammitglieds, also dem autonomen Agenten, liegen. Dazu wird zunächst auf die Programmierung sowie das Design des autonomen Agenten eingegangen und anschließend wird die Anpassungsfähigkeit des autonomen Agenten näher beschrieben.

2.2.1 Programmierung und Design des autonomen Agenten

In diesem Kapitel wird auf die unterschiedlichen Programmier- und Designentscheidungen bei Automation eingegangen. Dabei werden der Grad und die Form der Automation erklärt. Außerdem wird erläutert, inwieweit das Design an den Menschen angepasst werden muss. Der Automationsgrad lässt sich mittels des Modells von Parasuraman et al. (2000) erfassen. Dem Modell nach gibt es zehn Level der Automation. Diese zehn Level können nach O'Neill et al. (2020) in drei Oberkategorien eingeteilt werden, welche zusätzlich den Grad der Autonomie widerspiegeln. Für eine detaillierte Darstellung dieser Modelle und Einteilung sei auf Kapitel 1.3 verwiesen.

Design der Automation

Es kann zwischen zehn Leveln der Automation unterschieden werden, welche von einer niedrigen Stufe der Entscheidungen und Aktivitäten (z.B. der Computer schlägt Alternativen vor) bis hin zu Autonomie (der Computer entscheidet alles, agiert autonom und ignoriert den Menschen) variieren (Parasuraman et al., 2000). Diese Level werden von O'Neill et al. (2020) weiter in drei Oberkategorien eingeteilt:

- keine Autonomie/manuelle Kontrolle,
- teilweise Agentenautonomie und
- hohe Agentenautonomie,

welche Auskunft über den Grad an Autonomie geben. Die zehn Level der Automation werden in dem Automationsmodell von Parasuraman et al. (2000) integriert (vgl. Kapitel 1.3.2 für die Erläuterung des Automationsmodells). Im Rahmen des Designs soll dieses Modell genau reflektiert werden. Während der Entwicklungsphase gilt es für den Menschen zu entscheiden, welche Arten von Automation zu erzielen sind und zu welchen Leveln diese zu erreichen sind. Parasuraman et al. (2000) schlagen vor, dass die mit der Aufgabe verbundenen menschlichen Leistungen sowie die Folgen durch Automation für die Handlung untersucht werden sollen und diese im Entwicklungsprozess der Automation zu berücksichtigen sind. Die Konsequenzen für das menschliche Verhalten im System durch das Einführen von Automation stellen ein wichtiges Kriterium in der Entscheidungsfindung dar. Automation kann sowohl positive als auch negative Folgen für das menschliche Verhalten haben. Genannt werden vier menschliche Leistungsbereiche, welche durch Automation beeinflussbar sind: die mentale Arbeitsbelastung, die situative Aufmerksamkeit, die Complacency und der Fähigkeitsabbau. Complacency liegt vor, wenn die bedienende Person die Automation nicht ausreichend überwacht und sie so Systemfehler übersieht. Ein weiteres Kriterium ist die Verlässlichkeit der Automation, welche das Vertrauen des Menschen sowie Operations- und Herstellungskosten etc. beeinflussen kann. Durch die Berücksichtigung der Auswirkungen auf die menschliche Leistung kann die Eignung der einzelnen Stufen der Automation ermittelt werden (Parasuraman et al., 2000).

Zu berücksichtigende Faktoren bei der Designentscheidung

Transparenz beschreibt die Vorhersagbarkeit und Verständlichkeit der Aktionen von autonomen Systemen für den Menschen. Sie beeinflusst signifikant das Verständnis über die Aktionen der Systeme sowie die Art und Weise, diese zu beeinflussen (Endsley, 2017). Durch Transparenz können die situative Aufmerksamkeit und das Vertrauen des Menschen gegenüber der Automation erhöht werden, ohne dadurch die Arbeitsbelastung zu beeinflussen (Selkowitz, Lakhmani et al., 2016). Mit zunehmendem Autonomiegrad ist es umso wichtiger, das Interfacedesign so zu gestalten, dass es den Menschen möglich ist, zu verstehen, was seitens der Automation gerade getan wird, was noch gemacht wird und wo die Grenzen der Möglichkeiten sind. Diese Wissensvermittlung kann durch Interfaces mit hohen Stufen an Systemtransparenz erreicht werden und so die Verständlichkeit und Vorhersagbarkeit des Systems fördern. Wie verlässlich (z.B. die derzeitige Arbeitsweise oder Systembewertungen) und robust ein System ist, sind wichtige Informationen, welche für das menschliche Teammitglied ersichtlich sein müssen. Die Robustheit ist die Fähigkeit, momentane sowie zukünftige Situationen bewältigen zu können (Endsley, 2017). Durch die Transparenz des Agenten wird die allgemeine Teamleistung beeinflusst. Eine niedrige Transparenz beim Agenten hat eine Leistungsminderung zur Folge. Weitere Auswirkungen betreffen das Vertrauensverhältnis von Mensch und Agent (Mercado et al., 2016).

Wie in Kapitel 2.1 bereits erarbeitet, wird Vertrauen in den Agenten durch seine Verlässlichkeit beeinflusst. Verlässlichkeit wird gemäß O'Neill et al. (2020) oft als Grad der Genauigkeit betrachtet, mit dem ein autonomer Agent seine Aufgaben verrichtet. Der Verlässlichkeitsgrad beeinflusst die Haltung von Menschen gegenüber der Automation und demnach die Nutzung oder Nichtnutzung dieser (Parasuraman & Riley, 1997). O'Neill et al. (2020) vermuten, dass die Transparenz mit der Verlässlichkeit des autonomen Agenten zusammenhängt. Vertrauen und bestimmte Verhaltensmuster bei Automationen entstehen, wenn diese verlässlich und akkurat sind. Zu erkennen ist dies beispielsweise an Alarmsystemen mit einer erhöhten Quote an Fehlalarmen. Diese Systeme haben Grenzwerte, welche festlegen, ab wann ein Alarm aktiviert wird. Ein hoher Grenzwert für falsche Alarme hat zur Folge, dass es vermehrt zu Fehlalarmen und zu einem Vertrauensverlust vonseiten des Menschen kommen kann. Dadurch besteht jedoch die Möglichkeit, versäumte Warnungen zu minimieren. Ein niedriger Grenzwert für falsche Alarme führt zu einer höheren Akzeptanz von menschlicher Seite aus und zu einem erhöhten Risiko nicht wahrgenommener Warnungen. Die Festlegung der Entscheidungsschwelle eines automatisierten Warnsystems hängt von der Abwägung der Kosten eines fehlenden Signals gegenüber denen eines falschen Alarms ab. Das Problem ist eine Situation mit möglichem Fehlverhalten bedingt durch das Design der Fehlalarmrate. Diese resultiert aus dem Kostenverhältnis eines versäumten Alarms gegenüber dem eines Fehlalarms (Parasuraman & Riley, 1997).

Laut Demir et al. (2019) ist es für erfolgreiches HAT nötig, dass autonome Agenten durch deren Programmierung und Ausbildung eine grundlegende Vorstellung davon haben, was Teamarbeit ist und sie die wichtigsten Prinzipien dieser kennen. Sie sind kommunikationsfähig und verstehen ihre eigenen Aufgaben. Allerdings ist dies bei den Aufgaben der menschlichen Teammitglieder nur schwer realisierbar. Die Möglichkeiten des maschinellen Lernens gilt es in Zukunft zu nutzen, um den autonomen Agenten zu lehren, was gute Teamarbeit ist (Kommunikation, Koordination, Bewusstsein etc.). Solange ein Agent nicht darin geschult wurde, sein Wissen zu konzeptualisieren und anzuwenden, wird er nie in der Lage sein, sich an dynamische Eigenschaften wie Koordination anzupassen (Demir et al., 2019). Ob Automation genutzt wird oder nicht, hängt unter anderem davon ab, wie deutlich dem Menschen ist, welchen Nutzen Automation in der jeweiligen Lage bringt und wie praktisch sie in der jeweiligen Situation angewendet werden kann. Demnach sollte die Nutzung von Automation für den Menschen schnell und einfach erfolgen können, ohne dass der Mensch sich intensiv eindenken und lernen muss. Die Vorteile der Reduzierung von der Arbeitsbelastung sowie die Attraktivität für die Bedienenden sinken mit steigenden kognitiven Kosten im Rahmen der Nutzung und resultieren eher in Fehler (Parasuraman & Riley, 1997).

2.2.2 Anpassungsfähigkeit des autonomen Agenten

Bei einem autonomen Agenten im Rahmen von HAT handelt es sich um eine computergestützte Einheit. Diese nimmt innerhalb des Teams eine bestimmte Rolle ein. Der autonome Agent wird von den Menschen als Teammitglied angesehen, wenn zwischen beiden eine gegenseitige Abhängigkeit (Interdependenz) besteht und er als agentenhaft angesehen wird (O'Neill et al., 2020). Agentenhaft meint laut Lyons et al. (2018), dass der autonome Agent durch Selbstverwaltung selbstständig Handlungsoptionen verfolgen kann. Zwei universale Kriterien zur Einstufung von autonomen Agenten als Teammitglied sind der Grad der Interdependenz von Aktivitäten und Ergebnissen anderer Teammitglieder (Nass et al., 1996; Walliser et al., 2017) und der Grad der Handlungsunabhängigkeit und Proaktivität der autonomen Agenten (Lyons et al., 2018; Wynne & Lyons, 2018).

Anpassung an menschliches Verhalten

Um gemeinsam als ein Team agieren zu können, muss eine Interdependenz zwischen Menschen und autonomen Agenten bestehen. Dabei kann zwischen folgenden Interdependenzen unterschieden werden:

- der Aufgabeninterdependenz
- der Zielinterdependenz und
- der Interdependenz von Feedback und Anerkennung.

Die Aufgabeninterdependenz meint, dass die Gruppenmitglieder interagieren und aufeinander angewiesen sind, um die Aufgabe erledigen zu können. Die Interdependenz der Ziele bedeutet, dass die Ziele der Gruppe und der einzelnen Mitglieder miteinander verknüpft sein müssen, um eine maximale Gruppeneffektivität der Zusammenarbeit zu erreichen. Die Interdependenz von Feedback und Anerkennung bezieht sich auf die Verknüpfung von individuellem Feedback und individueller Anerkennung mit der Gruppenleistung, um zu einem gruppenorientierten Verhalten zu motivieren. Diese Art der Interdependenz ist wichtig für die Effektivität der Gruppe (Campion et al., 1993). Für Interdependenz sollte eine Gemeinsamkeit der Aufgaben und Ziele gegeben sein. Bei der Kollaboration von autonomen Agenten und Menschen sind diese mit unterschiedlichen Teilaufgaben der Gesamtaufgabe beschäftigt (Lyons et al., 2018). Diese gegenseitige Abhängigkeit sowie die Eigenschaften menschenähnlich (z.B. die menschliche Stimme eines autonomen Agenten) und agentenhaft zu sein, fördern es, dass der autonome Agent von dem Menschen als Partner und nicht als Werkzeug gesehen wird (Lyons et al., 2018; Wynne & Lyons, 2018). HAT liegt vor, wenn ein autonomer Agent als individuelles Mitglied ein Team ergänzt. Dieser wird als Ausführer einer einzigartigen Rolle im Team wahrgenommen (O'Neill et al., 2020). Somit leistet er einen Beitrag zu der Gesamtleistung (Larson & DeChurch, 2020).

Walliser et al. (2019) haben in zwei Experimenten die Wahrnehmung von Teamarbeit bei menschlichen Teams und bei Teams aus Menschen und autonomen Agenten untersucht. Neben der Wahrnehmung von Teamarbeit wurde die Anwendung von teambildenden Einflüssen analysiert. Beide Aspekte wurden dahingehend untersucht, wie sie die Teamergebnisse verbessern. Die teilnehmenden Personen haben entweder mit einem Menschen oder mit einem autonomen Agenten zusammengearbeitet. Durch die zwei Experimente soll gezeigt werden, inwieweit die Ergebnisse der Leistung verbessert werden können, wenn der autonome Agent als Teammitglied angesehen wird. Diese Leistungsergebnisse von Teams aus Menschen und autonomen Agenten sollen durch eine soziale Beziehung zwischen diesen erhöht werden. In dem ersten Experiment wird die Betrachtung der Automation als Werkzeug und der Automation als Teammitglied bezüglich der sozialen Ergebnisse von Affekten, Verhalten und Leistung verglichen. Es wird gezeigt, dass durch die Betrachtung als Teammitglied beispielsweise eine gesteigerte Interdependenz und gemeinsame Ziele entstehen. So konnte z.B. bezüglich der Rollenklärung keine Verbesserung festgestellt werden. Das Ergebnis des ersten Experiments ist, dass durch die Betrachtung des autonomen Agenten als Teammitglied einerseits die Affekte und das Verhalten des Teams verbessert werden können. Andererseits kann bei diesem Experiment nicht die Leistung erhöht werden (Walliser et al., 2019). An dieser Stelle vermuteten Walliser et al. (2019), dass eine Teamstruktur zwischen Menschen und autonomen Agenten notwendig ist. Sie reicht aber nicht aus, um die Leistung des Teams zu erhöhen. Um diese Hypothese zu überprüfen, wird bei dem zweiten Experiment die Teambildung bzw. die Teamentwicklung zur Erreichung einer Leistungssteigerung betrachtet. In diesem Experiment wird die Erhöhung der Teamleistung durch die Verbesserung der sozialen Interaktion zwischen Menschen und autonomen Agenten angestrebt. Letzteres soll durch teambildende

Maßnahmen geschehen. Das Ergebnis des zweiten Experiments ist, dass durch die vorherige Absprache der Rollenverteilung und des gemeinsamen Ziels die gesamten Ergebnisse der Teamarbeit sowie die Leistung verbessert werden können. Aus den Experimenten resultiert, dass soziale Interaktionen zwischen Menschen und autonomen Agenten eine wesentliche Rolle bei der Teamarbeit spielen und dass sowohl eine Teamstruktur als auch Teambildung notwendig sind (Walliser et al., 2019).

Die Interdependenz von Menschen und autonomen Agenten ist entscheidend für ein Team (Nass et al., 1996; O'Neill et al., 2020), da dies zu positiven Ergebnissen beim HAT führt (O'Neill et al., 2020). In einer Studie von Nass et al. (1996) zeigt sich, dass die Auswirkungen einer Zusammenarbeit mit einem Computer die Gleichen sind wie bei einer Zusammenarbeit mit einem anderen Menschen. Menschliche Teammitglieder nehmen autonome Agenten aufgrund der Interdependenz ähnlicher zu sich selbst wahr. Sie sehen sich selbst als kooperativer und empfänglicher für den Einfluss der autonomen Agenten im Vergleich dazu, wenn keine interdependente Beziehung zwischen den Menschen und den autonomen Agenten besteht. Die Informationen des Agenten werden als hochwertiger eingestuft, als angenehmer wahrgenommen und die Menschen passen sich diesen Informationen stärker an, vergleichsweise zu einer Situation ohne gegenseitige Abhängigkeit zwischen den menschlichen und autonomen Mitgliedern (Nass et al., 1996). Walliser et al. (2017) stellen zusätzlich fest, dass aus der Interdependenz der Ergebnisse positive Auswirkungen und eine bessere Teamleistung resultieren. Diese Teamaktivitäten und -ergebnisse, die von gegenseitiger Abhängigkeit geprägt sind, stellen eine Voraussetzung dar, um dem autonomen Agenten eine Rolle im Team anvertrauen zu können (O'Neill et al., 2020). Über die Interdependenz hinausgehend gibt es weitere Faktoren, die dazu beitragen, dass Menschen autonome Agenten als Teammitglieder und nicht als Werkzeug ansehen. Zu diesen Faktoren gehören Handlungsfähigkeit, Kommunikation (vgl. Kapitel 2.3.1) und Koordination (vgl. Kapitel 2.3.3; Lyons et al., 2018; Wynne & Lyons, 2018), die im Folgenden näher erläutert werden.

Das Kriterium der Handlungsfähigkeit wird als Grad des Entscheidungsspielraums beschrieben. Teammitglieder können Handlungsoptionen auswählen oder empfehlen, bei Werkzeugen besteht diese Möglichkeit nicht. Für die Betrachtung des autonomen Agenten als Teammitglied und für das erfolgreiche HAT ist Handlungsfähigkeit somit notwendig. Sie hängt zudem teilweise von Entscheidungsgewalt und Selbstverwaltung ab (Wynne & Lyons, 2018) und ergibt sich aus Unabhängigkeit, Selbstverwaltung und Proaktivität (O'Neill et al., 2020). Wenn eine gegenseitige Abhängigkeit zwischen Mensch und Agent besteht (Nass et al., 1996) und der autonome Agent als agentenhaft gesehen wird (Lyons et al., 2018), kann dieser wahrscheinlich als Teammitglied und nicht als Werkzeug angesehen werden (Mercado et al., 2016). In Systemen mit gemischten Initiativen kann durch die Aufgabenerlegung in kleinere Verantwortungsbereiche und die dynamische Verknüpfung dieser Aufgaben mit zugeordneten Rollen ein gemeinsames Verständnis der Arbeitsteilung geschaffen werden (Hayes & Scassellati, 2013; Lyons, 2013). Bei Systemen mit gemischten Initiativen handelt es sich um Systeme, in denen ein Mensch und ein autonomer Agent zusammenarbeiten, um Entscheidungen zu treffen (Allen et al., 1999; Chen & Barnes, 2014). Dieser Ansatz sowie das Kommunizieren der Absichten ermöglichen eine gegenseitige Anpassung, bei der der autonome Agent seine Rolle im Team ändern und autonom handeln kann, um ein Ziel zu erreichen (Chen & Barnes, 2014; Sycara & Sukthankar, 2006).

Es gibt Unterschiede zwischen der Kommunikation von autonomen Agenten und Menschen und der Kommunikation rein unter Menschen (O'Neill et al., 2020). Klein et al. (2004) definieren Herausforderungen, um autonome Agenten zu effektiven Teammitgliedern zu machen, wenn

sie mit Menschen interagieren. Damit die Agenten effektive Teammitglieder sind, müssen sie die Absichten und Handlungen der anderen Teilnehmenden angemessen modellieren. Dabei ist die Betrachtung des mentalen Zustandes und der Entwicklung der gemeinsamen Aktivität unerlässlich. Gemeint damit ist z.B., ob die anderen Teammitglieder Probleme haben, sich bestimmte Sackgassen ergeben und wie sich die Teammitglieder an Störungen anpassen. Mit Sackgassen sind Situationen gemeint, in denen das Teamziel aufgrund von Umsetzungsproblemen nicht weiterverfolgt werden kann. Es ist notwendig, dass die autonomen Agenten als Teil des Teams wie Menschen berechenbar sind. Der Agent muss vorhersagbar sein und in der Lage sein, die Handlungen anderer vorherzusagen zu können. Dazu gehört das Verhalten der Teammitglieder zu beobachten und dieses richtig vorausszusagen. Hierbei handelt es sich um eine gewisse Intelligenz und Autonomie des Agenten. Diese wirkt entgegen dem Vertrauen, das die Menschen in ihre Vorhersagbarkeit haben. Wenn autonome Agenten anpassungsfähiger werden, werden sie unvorhersehbarer (Klein et al., 2004). Die Anpassungsfähigkeit des Systems an den Arbeitsstil des Operators kann ansteigen. Ist dies der Fall, wird die bedienende Person unwilliger, ihr eigenes Verhalten anzupassen, da diese Anpassungen zu Verwirrung führen könnten (Klein, 2004). Die autonomen Agenten sollten, anstatt auf eine Abfrage des Menschen zu warten, den Informationsbedarf des menschlichen Teammitglieds besser antizipieren und die richtigen Informationen rechtzeitig übermitteln. Ansonsten könnte dies in eine Verzögerung resultieren (O'Neill et al., 2020). Autonome Agenten müssen in der Lage sein, relevante Aspekte ihrer Absichten gegenüber ihren Teammitgliedern deutlich zu machen (Klein et al., 2004). Damit ihre Aktionen ausreichend vorhersehbar sind, müssen sie ihre eigenen Ziele, Zustände, Kapazitäten, Absichten, Änderungen und bevorstehende Aktionen für den Menschen oder für andere Beteiligte, die sie überwachen und koordinieren, ersichtlich gestalten (Feltovich et al., 2004). Sie müssen relevante Status- und Absichtserklärungen beobachten und interpretieren. Die autonomen Agenten sollten in der Lage sein, sich an Zielverhandlungen beteiligen zu können. Wenn sie nicht in der Lage sind, ihre Ziele ohne Weiteres darzustellen, zu begründen oder zu modifizieren, stören sie die Koordination und die Aufrechterhaltung der Gemeinsamkeiten der Teammitglieder. Es ist notwendig, dass autonome Agenten signalisieren, wenn sie Probleme haben und extreme Maßnahmen ergreifen müssen oder wenn sie sich am Rande ihres Zuständigkeitsbereichs befinden (Klein et al., 2004). Im Rahmen von HAT ist es unerlässlich, dass die Kommunikation von autonomen Agenten und Menschen ähnlich wie die Kommunikation rein unter Menschen gestaltet wird.

Anpassung an die Aufgabe und die Situation

Der traditionelle Ansatz der Automation stammt von Fitts et al. (1951) und sagt aus, dass die Funktionszuweisung vollständig von dem Menschen oder von den autonomen Agenten festgelegt wird. Die Aufgabenkontrolle wird bei dem Design festgelegt und während der Aufgabenausführung nicht verändert (Scallen & Hancock, 2001). Das jeweilige Teammitglied kontrolliert dann vollständig die bestimmte Aufgabe (Fitts et al., 1951). Die adaptive Funktionszuweisung stellt eine Alternative zu dem traditionellen statischen Ansatz dar. Die Kontrolle von Aufgaben kann dynamisch zwischen dem Menschen und dem autonomen Agenten wechseln. Diese Verschiebung der Funktionskontrolle basiert auf festgelegten Schwellenwerten für beispielsweise die Kompetenz der bedienenden Person. Eine Beispielsituation wäre, dass die flugzeugführende Person das Flugzeug steuert und dann der Schwellenwert erreicht wird. Die Automation würde die Kontrolle über diese Aufgabe übernehmen und diese nach einer Ruhephase der flugzeugführenden Person zurückgeben. Wenn die festgelegten Schwellenwerte nicht erreicht werden, wird die Automation die Kontrolle über die Aufgabe nicht übernehmen (Scallen & Hancock, 2001). Bei dieser Form der Aufgabenzuweisung ist die Umverteilung der

Aufgaben bei ansteigender Belastung des Menschen möglich. Der autonome Agent übernimmt diese Aufgabe, und sobald die Arbeitsbelastung zurückgeht, kann der Mensch wieder die Kontrolle übernehmen (Byrne & Parasuraman, 1996). Zur Verbesserung von der Überwachung automatisierter Systeme kann die Aufgabe für kurze Zeit von dem autonomen Agenten an den Menschen gegeben werden und dieser kann sie dann an den autonomen Agenten zurückgeben. Diese kurzzeitige Aufgabenzuweisung an den Menschen ist bei geringer bis moderater Arbeitsbelastung sinnvoll. Dadurch können bei einer später eintretenden hohen Arbeitsbelastung Ausfälle der automatisierten Aufgabenüberwachung verhindert werden (Parasuraman et al., 1996).

Die Funktionsallokation beschreibt die Aufteilung der einzelnen Funktionen innerhalb des HAT zwischen Menschen und autonomen Agenten. Damit dies geschehen kann, ist eine Aufgaben- und Funktionsbeschreibung unerlässlich. Sie bezieht sich auf die Aufgaben sowie Funktionen, die von dem Team erwartet werden. Diese Beschreibung sollte klar differenziert sein, um Entscheidungskriterien dafür zu bieten, wie die Aufteilung der Funktionen umzusetzen ist. Für die Entscheidung der Aufgabenteilung können zwei unterschiedliche Automatisierungsstrategien herangezogen werden:

- die technik- bzw. kostenzentrierte Sichtweise und
- die fähigkeitszentrierte Sichtweise.

Bei der ersten Sichtweise stehen die technische Umsetzbarkeit und die Kosten bei der Entscheidung im Vordergrund. So werden die Aufgaben, die mit geringen Kosten und mit wenig technischem Aufwand automatisiert werden können, von dem autonomen Agenten übernommen. Die restlichen Aufgaben werden dem menschlichen Teammitglied zugeteilt (Manzey, 2008). Bei der fähigkeitszentrierten Sichtweise werden für die Entscheidung die einzelnen Vorteile der Leistungen der autonomen Agenten und der Menschen berücksichtigt (Fitts et al., 1951). Für die Zuteilung der Aufgaben wird betrachtet, welches Teammitglied diese mit größerer Effizienz bearbeitet. Zu beachten ist, dass diese beiden Automatisierungsstrategien aufgrund von zunehmender technologischer Entwicklung immer weniger deutlich voneinander getrennt werden können. Im Rahmen von HAT sollte die Funktionsallokation dynamisch verändert werden können, um die situativen Anforderungen bewältigen zu können (Manzey, 2008). Für die Anpassung abhängig von situativen Eigenschaften gibt es folgende Konzepte:

- die adaptierbare Automation und
- die adaptive Automation.

Die adaptierbare Automation meint, dass die Entscheidungsgewalt über die Aufgabenzuteilung bei dem Menschen liegt. Dieser kann festlegen, welche Aufgaben für welchen Zeitraum von einem bestimmten Teammitglied ausgeführt werden. Der Autopilot im Flugzeug stellt für diese Art der Automation ein Beispiel dar. Die bedienende Person legt fest, ob sie die Automation nutzt oder nicht. Bei der adaptiven Automation trifft die Automation die Entscheidungen bezüglich der Funktionsallokation (Manzey, 2008). Die adaptive Automation als Ansatz für die Automatisierungsgestaltung meint, dass eine dynamische und flexible Aufgabenaufteilung zwischen den menschlichen und den autonomen Teammitgliedern stattfindet (Byrne & Parasuraman, 1996). Die Entscheidung über die Zuweisung der Funktionen geschieht abhängig von dem jeweiligen Kontext (Manzey, 2008). Es handelt sich um die situationsabhängige Funktionsallokation, bei der dem autonomen Agenten, wenn sich die Komplexität der situativen Anforderungen erhöht, Aufgaben zugeteilt werden. So nimmt die Belastung des Menschen ab. In diesem Fall werden der Kontext und der Zustand der nutzenden Person betrachtet (Scallan & Hancock, 2001). In einem anderen Fall werden die Leistung und der Zustand der nutzenden

Person als Faktoren für die Entscheidung bezüglich der Aufgabenverteilung herangezogen. Beispielsweise wird die Müdigkeit einer Autofahrenden Person ermittelt und abhängig davon wird die Funktionsallokation stattfinden. Dies soll zu einer Optimierung der menschlichen Beanspruchung führen (Byrne & Parasuraman, 1996; Kaber & Riley, 1999). Die Überwachung des Zustandes von der bedienenden Person sorgt dafür, dass die Automatisierungsnutzung reguliert wird. Nach der Erfassung dieses Zustandes werden aufgrund von Änderungen der Arbeitsbelastung Aufgaben neu verteilt. Die bedienende Person übernimmt entweder weniger Aufgaben, da sie automatisiert werden, oder sie übernimmt mehr. Es kann von vollständig übernommenen Aufgaben bis hin zu der Übernahme einzelner Teile einer Aufgabe zwischen Menschen und autonomen Agenten variieren (Rouse, 1988). Im Rahmen der adaptiven Automation ist die Entscheidung darüber, wer die Funktionsallokation festlegt, ein wichtiger Aspekt. Der Mensch gibt dem autonomen Agenten die Macht über die Aufgabenverteilung, wodurch der Mensch seine Kontrolle über die Situation größtenteils abgibt (Manzey, 2008). Die hier beschriebenen Fachbegriffe sind in Tabelle 2.1 erklärt.

Tab. 2.1: Wichtige Fachbegriffe (in Anlehnung an Manzey, 2008; Scallen & Hancock, 2001).

Begriff	Erklärung
Flexible Funktionsallokation	Kontrolle über die Aufgaben kann dynamisch zwischen dem Menschen und dem autonomen Agenten wechseln
Adaptierbare Automation	Entscheidungsgewalt über die Aufgabenzuteilung liegt bei dem Menschen: Festlegung, welche Aufgaben und für welchen Zeitraum von einem bestimmten Teammitglied ausgeführt werden → Beispiel: Autopilot im Flugzeug
Adaptive Automation	Entscheidungsgewalt über die Aufgabenzuteilung liegt bei der Automation: Festlegung, welche Aufgaben und für welchen Zeitraum von einem bestimmten Teammitglied ausgeführt werden → Beispiel: Ermittlung der Müdigkeit bei einer Autofahrenden Person

Kozłowski und Ilgen (2006) sprechen die Entwicklung der Arbeitsgestaltung der letzten Jahre an und stellen fest, dass Teams komplexe dynamische Systeme sind, die in einem Kontext existieren und sich entwickeln, wenn die Mitglieder im Laufe der Zeit interagieren. Außerdem bilden Teams sich heraus und passen sich an, wenn sich die situativen Anforderungen verändern (Kozłowski & Ilgen, 2006). Der autonome Agent kann im Rahmen eines Systems mit gemischten Initiativen mehrere verschiedene Aufgaben wahrnehmen und jede davon mit einem variierenden Grad an Autonomie durchführen. Dieser Autonomiegrad hängt von den Anforderungen der Aufgabe und den Bedürfnissen der menschlichen Teammitglieder ab (Kaupp et al., 2010). Autonome Agenten, die für eine Interaktion in Systemen mit gemischten Initiativen auf niedriger Ebene ausgelegt sind, können in unsicheren Situationen eine Klärung ersuchen. Sobald kritische Informationen auftauchen, können sie diese melden. Auf höheren Ebenen ist der autonome Agent für die Durchführung bestimmter Vorgänge verantwortlich. Er übernimmt Teilaufgaben, indem er an der Umsetzung eines Plans arbeitet. Wenn Entscheidungen von Menschen getroffen werden müssen, fordert er diesen dazu auf. Solange der autonome Agent mit der Umsetzung des Plans beschäftigt ist, liegt die Initiative bei ihm. Sobald seine Teilaufgabe beendet ist, erhält der Mensch diese Initiative zurück. Auf den höchsten

Stufen bei Interaktionen von Systemen mit gemischten Initiativen überwacht der autonome Agent aktiv die aktuelle Aufgabe. Hierbei nutzt er Informationen, die er über seine Fähigkeiten, die des Menschen und die anderen Anforderungen an seine Ressourcen hat. So kann er beurteilen, ob er die Initiative ergreifen sollte (Allen et al., 1999). Parasuraman et al. (2000) sagen, dass der Grad und vielleicht sogar die Art der Automation so gestaltet werden können, dass sie abhängig von den situativen Anforderungen während des Betriebs variiert. So könnte in einem Beispiel die Entscheidung über das Fortsetzen oder das Abbrechen eines Flugzeugstarts nach einer Triebwerksstörung mit einem niedrigen oder einem hohen Niveau automatisiert werden. Dies könnte in Abhängigkeit der zeitlichen Kritikalität der Situation (z.B. wie nah das Flugzeug an der kritischen Geschwindigkeit für den Start ist) entschieden werden (Inagaki, 1999). Ein weiteres Beispiel sind Militärroboter in dem Zusammenhang von situativer Anpassung. Für sie spielen die rauen Bedingungen, wie potenziell feindliche Umgebungen und zeitliche Einschränkungen, in denen sie operieren müssen, eine Rolle. Sie müssen die Dynamik ihrer Umgebung verstehen können. So kann ein Roboter beispielsweise gezwungen sein, seine Flughöhe zu erhöhen, weil in einem Gebiet Kampfhandlungen stattfinden (Lyons, 2013). Für erfolgreiches HAT ist es notwendig, dass der autonome Agent sich an die Aufgabe und die Situation anpassen kann. Die in diesem Unterkapitel 2.2 beschriebenen Anforderungen aufseiten des autonomen Agenten innerhalb eines HAT werden zusammenfassend in Infobox 2.2 dargestellt.

Infobox 2.2: Anforderungen aufseiten des autonomen Agenten.

Anforderungen aufseiten des autonomen Agenten:

- **Programmierung und Design des autonomen Agenten**
 - Zehn Level der Automation
 - Bewusstsein der Folgen durch das Einführen von Automation
 - Transparenz
 - Verlässlichkeit
 - Programmierung und Ausbildung
- **Anpassung an menschliches Verhalten**
 - Interdependenz
 - Aufgabeninterdependenz
 - Zielinterdependenz
 - Interdependenz von Feedback und Anerkennung
 - Handlungsfähigkeit
 - Antizipation der anderen Teammitglieder
 - Vorhersagbarkeit
 - Rechtzeitige Informationsvermittlung
 - Verdeutlichung der eigenen Absichten
 - Beobachtung und Interpretation von relevanten Status- und Absichtserklärungen
 - Beteiligung an Zielverhandlungen
- **Anpassung an die Aufgabe und die Situation**
 - Flexible Funktionsallokation
 - Adaptierbare Automation
 - Adaptive Automation

2.3 Zusammenführung der Anforderungen aufseiten der Menschen und des autonomen Agenten als Team

Die Anforderungen aufseiten des menschlichen Teammitglieds sowie die Anforderungen aufseiten des autonomen Agenten wurden in den vorherigen Kapiteln betrachtet. Diese werden nun zusammengeführt. Dazu wird zunächst auf bidirektionale Kommunikation eingegangen, dann wird Transparenz näher beschrieben und anschließend folgt die Erläuterung von Koordination und Teamkognition.

2.3.1 Bidirektionale Kommunikation

Laut M. D. McNeese und N. J. McNeese (2020) sind Kommunikation und Koordination für effektive Teamarbeit bedeutend, besonders im Hinblick auf das Verständnis der Interaktion von Menschen und autonomen Agenten im Team. Der Austausch von diversen Informationen, wie z.B. Antworten oder Fragen, wird als bidirektionale Kommunikation definiert (Kaupp et al., 2010). Sie fördert die effektive Leistung des Teams aus den Menschen und den Agenten

(Chen et al., 2018). Marathe et al. (2018) gehen auf Faktoren für die Verbesserung der bidirektionalen Kommunikation ein. Gemeinsame mentale Modelle werden durch die Kommunikation von Informationen geschaffen und haben eine positive Auswirkung auf das Teamverständnis (Chen et al., 2018). Nach O'Neill et al. (2020) gibt es unterschiedliche Formen der Sprache: die visuelle Darstellung, die textbasierte Methode, die Kommunikation durch beobachtbare Änderungen und Audioäußerungen. Verschiedene Verhaltensarten von Kommunikation in Teams aus Mensch und Automation haben Auswirkungen auf die Teamleistung (Demir et al., 2016).

Notwendigkeit von bidirektionaler Kommunikation

Bidirektionale Kommunikation ist der Austausch von diversen Informationen wie Befehle, Fragen und Antworten (Kaupp et al., 2010). Demir et al. (2019) betonen, dass für effektive Teamarbeit sinnvolle Kommunikation nötig ist. Dafür sollen Menschen im Umgang mit Agenten trainiert werden. Ohne entsprechendes Training kann es zu Komplikationen in der Interaktion kommen, weil das Wissen über die Kommunikation und die Begrenzungen der Koordination des Agenten fehlen. Die Agenten müssen für Teamfähigkeit programmiert und trainiert werden (Demir et al., 2016). Zudem nimmt die Komplexität von autonomen Agenten zu und somit die Möglichkeiten von Wechselbeziehungen, weshalb die Relevanz der Koordination zwischen Menschen und Agenten zunehmend an Bedeutung gewinnt (Chen et al., 2018). Laut Shively et al. (2017) ist bidirektionale Kommunikation für das Einbringen von Automation in einem Team notwendig. Sie wird als Schlüssel zum Lösen vieler typischer Probleme in hochautomatisierten Systemen gesehen. So kann die bidirektionale Kommunikation dazu führen, dass Automation außerhalb des Bereiches, wofür sie programmiert wurde, weniger Eingriffe von Menschen benötigt, um Situationen zu bewältigen. Sie kann zudem bei dem Aufbauen von einem geteilten Bewusstsein helfen. Dafür werden eine gemeinsame Sprache und das Ermöglichen von expliziter wie auch impliziter Kommunikation benötigt. Durch die Diskussion der Ziele sowie Vertrauen und Vernunft wird das Verstehen erleichtert (Shively et al., 2017). Bidirektionale Kommunikation verbessert das gegenseitige Verständnis voneinander und ermöglicht effektive Aufgabenkoordination durch das Fördern der Kommunikation zwischen Menschen und Agenten (Marathe et al., 2018).

Erfolgreiche menschliche Teams können effektiv miteinander kommunizieren, wissen über die Fähigkeiten der anderen Bescheid und sind in der Lage, zukünftige Aktionen der Mitglieder zu antizipieren sowie die Bedingungen der vorliegenden Aufgabe zu deuten und zu beurteilen. Um diese Eigenschaften auch in Teams von Menschen und Agenten zu erlangen, sehen Marathe et al. (2018) beidseitige Kommunikation als relevanten Aspekt für die Förderung des gegenseitigen Verständnisses. Dabei spielt das Verstehen der Entscheidungen und der Entscheidungsprozesse der einzelnen Teammitglieder eine wichtige Rolle. Die Aufrechterhaltung von angemessenem Verstehen und kalibriertem Vertrauen vonseiten des Menschen zu den Agenten bestimmt die Effektivität der Zusammenarbeit zwischen Mensch und Agent. Bidirektionale Kommunikation zwischen Menschen und autonomen Agenten ist wichtig für eine effektive Leistung in Teams. Andernfalls bleiben möglicherweise der Plan, der Zielstatus oder der zugrunde liegende Argumentationsprozess unbekannt und so ist es dem Menschen nicht möglich, effektiv einzugreifen. Wenn die menschlichen Absichten, Bedingungen und Ziele dem Agenten nicht klar sind, kann er seine Pläne nicht anpassen (Chen et al., 2018).

Um Fortschritte in bidirektionaler Kommunikation zu erzielen, sind drei Faktoren relevant. Der erste Faktor ist der Anspruch an den Agenten in der Lage zu sein, den Zustand und die Intention des menschlichen Teammitglieds zu verstehen. Dieser Bereich fokussiert sich darauf, den

Menschen wahrzunehmen und den Agenten über den voraussichtlichen Status sowie die erwartete Leistung dessen zu informieren. Als zweiter Faktor gilt, dass die menschlichen Teammitglieder verstehen können, was die Möglichkeiten und Intentionen des Agenten sind. Dem Menschen werden zu einem passenden Zeitpunkt in einer angemessenen Art und Weise Informationen über den Agenten mitgeteilt. Damit das System funktioniert, muss der dritte Faktor gegeben sein. Es werden geschlossene Kreisläufe benötigt, wodurch die Informationen der Teammitglieder effektiv eingearbeitet und das Verhalten von allen Teammitgliedern koordiniert werden kann. Dabei werden Stärken hervorgehoben und Schwächen kompensiert, damit die gesamte Teamleistung gesteigert wird (Marathe et al., 2018).

Mentale Modelle im Rahmen von bidirektionaler Kommunikation

Die Kommunikation von relevanten Informationen kann bei der Schaffung von gemeinsamen mentalen Modellen helfen und somit das Teamverständnis stärken (Chen et al., 2018). Mentale Modelle bezeichnen das gemeinsame Verständnis von den Zielen des Teams, den Aufgaben der einzelnen Teammitglieder und der Teamkoordination, wodurch das Team die gemeinsamen Ziele erreichen kann (Cannon-Bowers et al., 1995). Eine stärkere Annäherung von mentalen Modellen der unterschiedlichen Teammitglieder resultiert in erheblichen Steigerungen des Teamprozesses und der Leistungen (Mathieu et al., 2000). Die Faktoren der Entscheidungsprozesse von Menschen und Agenten werden ihnen durch bidirektionale Kommunikation bewusst und diese können sie dann im eigenen Entscheidungsprozess aufnehmen sowie Fehler des anderen korrigieren (Chen et al., 2018). Menschliches Vertrauen kann vermindert werden, wenn das Verhalten des Menschen nicht mit dem Modell des Agenten übereinstimmt. Die Folge ist eine mögliche hemmende Wirkung auf die Teamleistung, welcher nur durch gezielte Kommunikation entgegengewirkt werden kann. Die Genauigkeit der Modelle des menschlichen Verhaltens kann durch transparente bidirektionale Kommunikation ermittelt werden. Diese Informationen sind nötig, um im Team ein gemeinsames Verständnis und Verstehen zu entwickeln (Schaefer et al., 2018).

Formen und Verhaltensarten der Teamkommunikation

Es kann zwischen verschiedenen Formen der Kommunikation unterschieden werden. Es gibt die visuelle Darstellung, welche das visuelle Teilen von Aktionsvorschlägen ermöglicht (z.B. eine Route, um von einem Ort zu einem anderen zu gelangen). Des Weiteren gibt es die textbasierte Methode, wodurch Nachrichten zwischen den Teammitgliedern via Chat ausgetauscht werden können. Zudem existiert die Kommunikation mittels beobachtbarer Änderungen bezüglich des Ortes und der Orientierung. Eine weitere Kommunikationsform sind Audioäußerungen durch eine gesprochene Sprache. In der Text- und Audiokommunikation wird die Wizard of Oz Technik genutzt. Diese Technik lässt die menschlichen Teammitglieder in dem Glauben, dass die Stimme oder der Text von dem autonomen Agenten stammt, obwohl diese Nachrichten von einem/einer Verbündeten, der/die als Vertreter/in des autonomen Agenten gilt, kommen (O'Neill et al., 2020). Die unterschiedlichen Formen der Kommunikation werden in Abbildung 2.2 dargestellt.

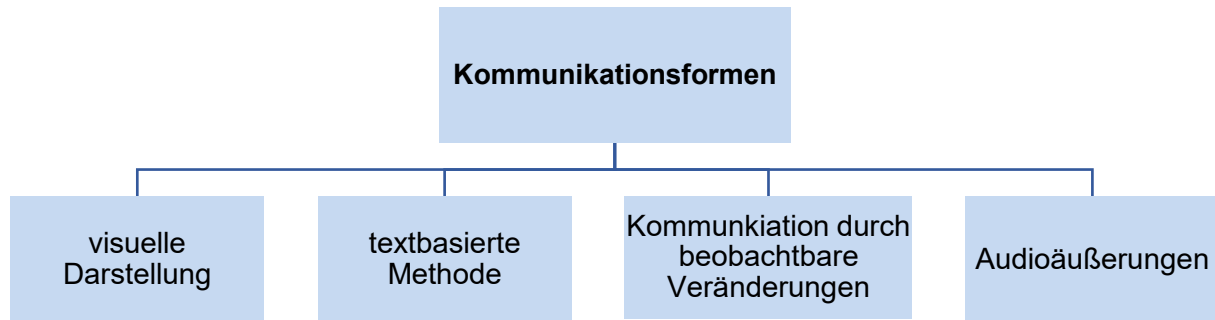


Abb. 2.2: Darstellung der unterschiedlichen Kommunikationsformen (in Anlehnung an O’Neill et al., 2020).

Durch natürliche Sprache kann die bedienende Person komplexe Anforderungen flexibel ausdrücken und ermöglicht so Interaktionen mit dem Agenten, als wäre es ein Mensch. Mehrdeutige Befehle, welche ein Mensch mit Leichtigkeit entschlüsseln könnte, sind in relativ eingeschränkten Umgebungen möglicherweise verwirrend für den Agenten (Tellex et al., 2012). Beispielsweise kann der Befehl „Beobachte die Hintertür“ (Chen & Barnes, 2014, S. 25) kontextbedingt doppeldeutig sein. Wenn mehrere Gebäude zu überwachen sind, ist nicht deutlich, welche Tür gemeint ist, welche Türen als Hintertüren gelten oder ob Fenster auch als solche zählen. Das Hinzufügen von beispielsweise Gestik verbessert die Möglichkeiten von Mensch und Maschine, mehr in einer natürlichen bidirektionalen Kommunikation zu interagieren. Kommandosprachen sind Umsetzungstabellen, die exakt wörtlich formuliert sind und als Befehl für Agenten oder als Antwort von Agenten funktionieren. Eine weitere Form der Kommunikation ist die kontrollierte natürliche Sprache. Sie ist im Wortschatz begrenzter als die natürliche Sprache und im Vergleich zu der Kommandosprache in der Lage, in komplexeren Dialogen zu agieren. Am nützlichsten ist die kontrollierte natürliche Sprache in spezialisierten Umgebungen (Chen & Barnes, 2014). Abgebildet werden die drei Sprachen in Abbildung 2.3.

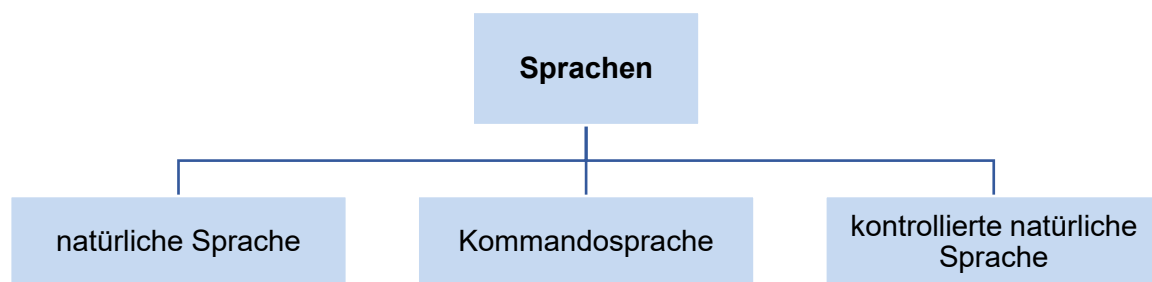


Abb. 2.3: Abbildung der drei Sprachen (in Anlehnung an Chen & Barnes, 2014).

Es gibt verschiedene Verhaltensarten von Kommunikation in Teams aus Mensch und Automation (Cooke et al., 2007; Demir & Cooke, 2014), welche sich mehr oder weniger positiv auf die Teamleistung auswirken (Demir et al., 2016). Insgesamt gibt es acht unterschiedliche Verhaltensweisen:

- generelle Statusupdates,
- wiederholte Nachfragen,
- Anfragen des Status anderer,
- Vorschläge,
- Vorausplanungen,
- positive Kommunikation,
- negative Kommunikation und

- unklare Kommunikation.

Die generellen Statusupdates geben den anderen Teammitgliedern Auskunft über den derzeitigen Zustand (z.B. Angaben über die momentane Geschwindigkeit). Bei den wiederholten Nachfragen werden Informationen und Aktionen von Teammitgliedern erneut erfragt (z.B. Welche Aktion als nächstes erfolgt, wenn das jetzige Ziel erreicht wird?). Anfragen über den Status anderer informieren die nachfragende Person über den momentanen Zustand der anderen und kommuniziert Interesse (z.B. Ist alles bereit für den nächsten Schritt?). Vorschläge geben den Teammitgliedern mögliche Alternativen zum Ausführen (z.B. Verlangsame die Geschwindigkeit um ein Viertel.). Vorausplanen beinhaltet das Antizipieren von zukünftigen Folgeaktionen und das Aufstellen von Regeln für künftige Situationen (z.B. Sobald die festgelegte Grenze überschritten wird, gilt es die Geschwindigkeit und Höhe auf einem bestimmten Wert zu halten, bis der Bereich überquert wird.). Positive Kommunikation besteht aus dem Helfen der Teammitglieder mittels Informationen und Anerkennung der Sprache der Mitglieder (z.B. Bitte gib die Information von dem nächsten Ziel an die flugzeugführende Person.). Die negative Kommunikation beinhaltet als Resultat von unterschiedlichen Zielen Diskussionen im Team (z.B. wie das Ziel am besten erreicht werden kann). Unklare Kommunikation ist dann der Fall, wenn die gesendeten Informationen aufgrund von Rechtschreibfehlern oder doppeldeutigen Begriffen unverständlich sind (Cooke et al., 2007; Demir et al., 2016; Demir & Cooke, 2014). Jedes Mitglied innerhalb des HAT kann die unterschiedlichen Verhaltensarten in einer strukturierten Art und Weise nutzen, um die bestmögliche Kommunikation und Koordination zu erzielen (Demir et al., 2016).

2.3.2 Transparenz

Ein Agent und die bedienende Person sollten in der Lage sein, zu erkennen, was ein Teammitglied macht und aus welchem Grund so gehandelt wird. Wenn dies erfüllt ist, kann von transparenter Automation gesprochen werden. Das bedeutet nicht allein die Bereitstellung von Informationen. Zum Beispiel unterscheiden sich die Berechnungen von Agenten oftmals von der Vorgehensweise der Menschen mit den gleichen Berechnungen (Lyons, 2013). Ein Agent ist transparent, wenn er diese Lücke schließt, indem er die Information so darstellt, dass sie dem mentalen Modell des Menschen entspricht. Die Wichtigkeit der Transparenz ist damit gegeben, dass sie dem Menschen ermöglicht, sein Vertrauen zu kalibrieren und „Misuse“ oder „Disuse“ (Parasuraman & Riley, 1997, S. 230) verhindert (Shively et al., 2017). Ein Mangel an Transparenz beschreibt die fehlende Einsicht des menschlichen Teammitglieds in die Aktionen von (teil-) autonomen Systemen. Dies erschwert diesem Mitglied nachzuvollziehen, weshalb der Agent die ausgeführte Aktion tätigt oder nicht ausgeführt hat (Shively et al., 2017). Anschaulich wird die Situation bei der Zusammenarbeit von der flugzeugführenden Belegschaft mit Flugmanagementsystemen. Hier ist das Verständnis dieser Belegschaft oft eingeschränkt. Ihr ist nicht immer bewusst, was das System im Augenblick macht oder was es als nächstes machen wird (Wiener et al., 1988). Ein angemessener Umgang mit dem Agenten ist durch diese Undurchsichtigkeit nicht gegeben (Shively et al., 2017). Durch mangelnde Transparenz entsteht mangelndes Situationsbewusstsein (Situation Awareness), da der Mensch nicht wissen kann, auf welche Informationen der Agent Zugriff hat. Beispielhafte Fragen können sein, ob der Agent auf Wetterinformationen zugreifen kann oder ob er das Wetter in der näheren Umgebung mit einbezieht. Durch das Fehlen der Informationen ist eine schlechte oder verwirrende Empfehlung des Agenten sehr wahrscheinlich (Shively et al., 2017). Faktoren, die für das gemeinsame Situationsbewusstsein erforderlich sind, wurden von Onken (1997) aufgezeigt. Hierzu schlägt er vor, dass die Schlüsselspezifikationen für die Entwicklung von Agenten in einem Flugzeug-Cockpit ein umfassendes maschinelles Wissen der Flugsituation besitzen

müssen. Weiterhin müssen sie in der Lage sein, effizient zwischen der Besatzung und der Maschine zu kommunizieren. In seinem Modell des erforderlichen Wissens zur Situationsbeurteilung identifiziert er Faktoren, die sich auf:

- das Flugzeug (z.B. Leistungsdaten, Systemstatus, etc.),
- die Crew (z.B. Ressourcen, Standardverhalten, individuelles Verhalten, etc.),
- die Mission (z.B. Ziele und Beschränkungen),
- die Flugsicherung (z.B. Freigaben, Anweisungen usw.) und
- die Umgebungsfaktoren (z.B. Navigationsdaten, Wetter und Verkehr) beziehen.

Es liegt nahe, dass die Automation nicht ohne ein Verständnis der Situation effizient helfen kann (Onken, 1997).

Situation Awareness (SA)

Nach Endsley (1995) wird Situation Awareness (SA) als die Wahrnehmung der Elemente in der Umgebung innerhalb eines Zeit- und Raumvolumens, das Verstehen ihrer Bedeutung und die Projektion in die nahe Zukunft beschrieben. SA besteht aus drei Komponenten, die hierarchisch voneinander abhängig sind. Die erste Komponente ist die Wahrnehmung der Elemente in der Umgebung und besteht darin, neben den Eigenschaften den Status und die Dynamik aller relevanten Umgebungselemente wahrzunehmen. Im Beispiel der Autofahrenden Person wäre es die Position aller anderen Fahrzeuge, mögliche Hindernisse, aber auch der Status des Fahrzeuges selbst, wie z.B. der Benzinstand. Im Falle der flugzeugführenden Personen wären es alle Elemente des Flugzeuges, Berge oder Warnleuchten. Nicht nur die Elemente, sondern auch die Eigenschaften dieser, wie z.B. die Größe, Farbe, Geschwindigkeit oder Position, gehören zur ersten Komponente zum Erreichen von SA. Die zweite Komponente von SA beinhaltet das Verstehen der aktuellen Situation und geht über die reine Kenntnisnahme der vorhandenen Elemente hinaus. Basierend auf dem vorangegangenen Wissen über die Elemente aus der ersten Komponente, folgt das Verständnis über die Bedeutung der Elemente und das Zusammenfügen eines ganzheitlichen Bildes der Umgebung hinsichtlich der Ziele der Bedienenden. Zum Beispiel muss die bedienende Person eines Kraftwerks die verschiedenen Daten aller Systemvariablen zusammenfügen, um die Funktionstüchtigkeit aller Systemkomponenten zu erfassen und mögliche Abweichungen zu lokalisieren. Bei der dritten Komponente der Projektion des zukünftigen Zustandes werden auf Basis der ersten und zweiten Komponente in naher Zukunft Fähigkeiten sowie Aktionen der Elemente in der Umgebung projiziert. Auf Basis der Wahrnehmung des Status und der Dynamiken der Elemente und dem Verständnis der aktuellen Situation werden zukünftige Aktionen der Elemente abgebildet. Wenn z.B. eine kampfflugzeugführende Person weiß, dass ein fremdes Flugzeug sich offensiv einem Standpunkt annähert, kann sie voraussagen, dass es ein feindliches Flugzeug ist, welches einen Angriff startet. Diese Projektion liefert ihr das Wissen und gegebenenfalls genug Zeit, um eine günstige Vorgehensweise für das Erreichen der eigenen Ziele zu ermöglichen. Das Erfassen von voraussichtlichen Umweltzuständen in der näheren Zukunft ist von Bedeutung für die Entscheidungsfindung. SA bietet ein breites Konstrukt, welches in vielen Bereichen anwendbar ist (Endsley, 1995).

SAT-Modell

Das Situation Awareness-based Agent Transparency (SAT-) Modell (Chen et al., 2018) beschreibt den Umgang mit Informationen, der für die Entscheidungsfindung der Menschen wichtig ist, um das Verständnis von effektivem HAT zu verbessern (Chen et al., 2018). Mit zunehmender Intelligenz der Agenten werden diese fähig, komplexe Entscheidungen zu treffen

(Shattuck, 2015; Warner et al., 2015). Im Gegenzug wird es für das menschliche Teammitglied zunehmend schwieriger, die Prozesse nachzuvollziehen. Jedoch ist dies von entscheidender Bedeutung, um eine effektive Zusammenarbeit zu suggerieren und wird zunehmend wichtiger bei einer dynamischen Umgebung (Chen & Barnes, 2014). Für das Vertrauen der Menschen in autonome Systeme sind die geringe Beobachtbarkeit, die Vorhersagbarkeit, die Richtbarkeit, die Überprüfbarkeit und das geringe gegenseitige Verständnis gemeinsamer Ziele als Problembereiche von Bedeutung. Das SAT-Modell von Chen et al. (2014) ordnet die Problembereiche ein und fördert somit das Verständnis der Menschen von den Agenten. Das SAT-Modell teilt sich in drei unabhängige Bereiche auf: die aktuellen Aktionen und Pläne, der Argumentationsprozess und die Ergebnisvorhersagen (Chen et al., 2018). Auf der ersten Ebene des SAT-Modells wird der Mensch auf der Informationsebene unterstützt, um die Wahrnehmung der aktuellen Aktionen und Pläne des Agenten zu verbessern. Dies geschieht durch die Bereitstellung der grundlegenden Informationen über den Zustand, die Ziele, die Absichten und die Pläne des Agenten. Auf der zweiten Ebene werden die Prozesse des Agenten dargestellt sowie die Einschränkungen, die er in seiner Entscheidung berücksichtigt. Das menschliche Teammitglied ist dann in der Lage, das aktuelle Verhalten des Agenten nachzuvollziehen. Auf der dritten Ebene kommuniziert der Agent die Wahrscheinlichkeiten von Erfolg und Misserfolg sowie mögliche zukünftige Zustände und Konsequenzen. Dadurch wird der Mensch dabei unterstützt, sich auf Situationen, die in der Zukunft liegen, vorzubereiten (Chen et al., 2018). Das SAT-Modell nutzt drei Komponenten, die laut Lee und See (2004) die Kalibrierung des Vertrauens durch Transparenz beeinflussen. Dazu gehören der Zweck, der Prozess und die Leistung. Um den Operator die Informationen leichter bereitzustellen, schlug Lee (2012) vor, die drei Komponenten und die Leistungshistorie in vereinfachter Form darzustellen (z.B. durch eine grafische Anzeige), damit der Mensch nicht von den Informationen überfordert wird (Cook & Smallman, 2008; Selkowitz, Lakhmani et al., 2016). Die Ebenen des SAT-Modells werden in der folgenden Abbildung 2.4 dargestellt.

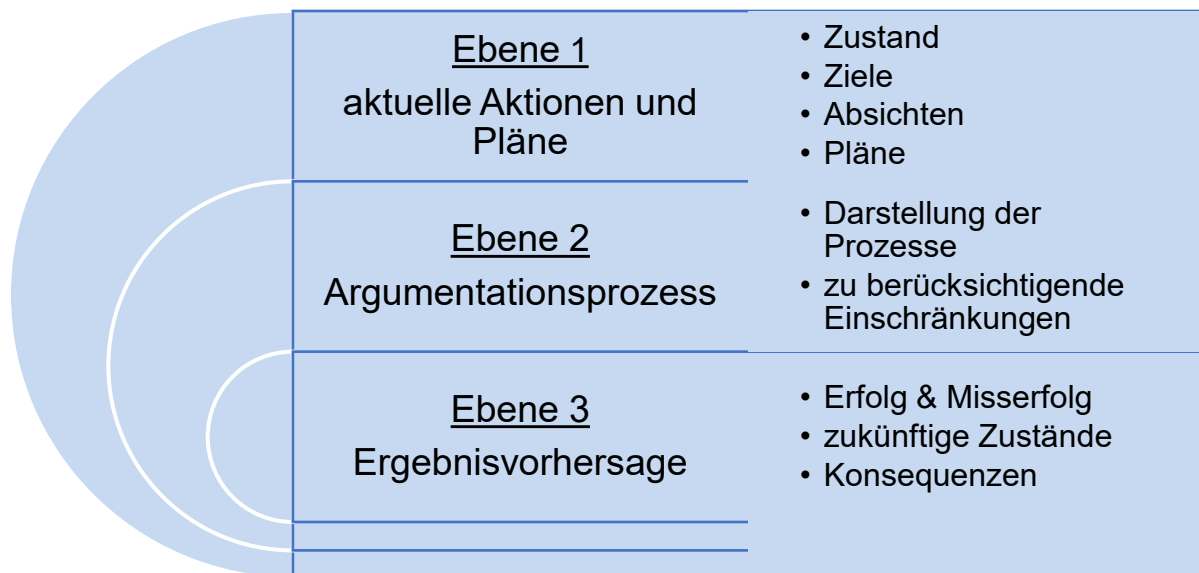


Abb. 2.4: Darstellung des SAT-Modells (in Anlehnung an Chen et al., 2018).

Autonomous Squad Member (ASM)

Das US-Verteidigungsministerium hat sich im Rahmen des Autonomy Research Pilot Initiative (ARPI) Forschungsprogramms wichtigen Fragen zum Thema HAT gewidmet (Department of Defense, 2013; zitiert nach Chen et al., 2018). Eines von sieben Projekten ist das Autonomous Squad Member (ASM). Dieses Projekt befasst sich mit der Interaktion zwischen einem

Menschen und einem kleinen Bodenroboter. Der ASM-Roboter ist ein robotergestütztes Fahrzeug, das in einer simulierten Infanterieumgebung Nachschub transportiert und sich autonom zu einem Sammelpunkt begeben kann (Selkowitz, Larios et al., 2016). Ein Ziel des ASM-Projektes ist es, einen transparenten Agenten zu konstruieren. Dieser Agent soll in der Lage sein, das Situationsbewusstsein der bedienenden Person für die aktuelle Lage zu fördern sowie seine Pläne und Überlegungen für sein zukünftiges Handeln mit wahrscheinlichen Ergebnissen zu kommunizieren. Neben den Informationen über die Umgebung, Ressourcenbestände und Ergebnisse aus vorangegangenen Aktionen berücksichtigt der ASM-Roboter die wahrgenommenen Aktionen seiner Teammitglieder und fügt diese der Informationsmenge hinzu. Daraufhin wählt er aus dieser Informationsmenge eine Aktion aus, beispielsweise ob er in Deckung geht oder ein Hindernis umgehen soll. Durch die Kommunikation der Informationsmenge mit den Mitgliedern des Teams ist ein erster Schritt in Richtung bidirektionaler Transparenz innerhalb eines HATs gemacht (Chen et al., 2018). Mit dem SAT-Modell als Grundlage besitzt der ASM-Roboter eine Benutzeroberfläche, auf der der Agent seine Pläne, Beweggründe und voraussichtlichen Ergebnisse darstellt. Dies soll transparente Interaktionen im HAT fördern (Selkowitz, Lakhmani et al., 2016). Wie sich diese Benutzeroberfläche auf das HAT auswirkt, wird in mehreren Studien untersucht, die sich mit der Auswirkung von Transparenz seitens des Agenten befassen, wie z.B. auf SA der Teammitglieder, das Vertrauen in den ASM-Roboter und die Arbeitsbelastung. In einem Experiment werden Testpersonen zu verschiedenen Schnittstellen der SAT-Ebenen untersucht. Die Versuchspersonen sollen in einer virtuellen Umgebung auf mögliche Bedrohungen achten, während sie das ASM-Display im Blick halten. Anhand dieser Untersuchung soll festgestellt werden, welche Informationsmenge die Bedürfnisse der menschlichen Teammitglieder am besten unterstützt und wie sich die Zusammenarbeit auf die Arbeitsbelastung auswirkt. Das signifikant bessere Verständnis der Aktionen des Agenten ist in der Schnittmenge von allen drei Ebenen besser als nur von der ersten und zweiten Ebene. Die Teilnehmenden des Versuchs sind besser darin, den zukünftigen Zustand des Agenten vorherzusehen unter der Bedingung aller Ebenen als in der ersten Ebene sowie in der ersten und zweiten Ebene. Diese Ergebnisse deuten darauf hin, dass bei einer Agentenschnittstelle, die alle drei Ebenen des SAT-Modells unterstützt, das Verständnis als auch die Voraussicht auf zukünftiges Handeln des Agenten für die menschlichen Teammitglieder erleichtert wird. Zudem wird bestätigt, dass wenn alle drei Ebenen des SAT-Modells erfüllt werden, es dazu beiträgt, das Verständnis des Agenten als auch die Voraussicht auf zukünftiges Handeln des Agenten für die bedienenden Personen zu erleichtern (Selkowitz, Larios et al., 2016).

Dynamisches SAT-Modell

Bei Teams, die ausschließlich aus Menschen bestehen, bilden die Teammitglieder ein Verständnis ihrer Umgebung. Dazu gehört, das Interaktionsmuster zwischen den Teammitgliedern zu verstehen und ein Verständnis der Leistungsumgebung und der verwendeten Geräte bei der Zusammenarbeit zu haben (Mathieu et al., 2000). Durch dieses Wissen ist es den Teammitgliedern möglich, effektiv miteinander zu kommunizieren, das Handeln der Mitarbeitenden vorherzusehen und ihr Verhalten zu koordinieren (Cannon-Bowers et al., 1993; Sycara & Sukthankar, 2006). Um das HAT zu unterstützen, müsste dieses Wissen von effektiven Teams im Design des Agenten integriert werden (Sycara & Sukthankar, 2006). Von Lyons (2013) wurde vorgeschlagen, dass Transparenz sich aus verschiedenen Dimensionen zusammensetzt. In der Teamdimension unterstützt der Agent die Arbeitsteilung, indem er dem Menschen vermittelt, welche Aufgaben in dessen Zuständigkeitsbereich fallen und welche im Zuständigkeitsbereich des Agenten liegen sowie welche Aufgaben sowohl das Eingreifen des

Menschen als auch des Agenten benötigen (Bradshaw et al., 2011; Lyons, 2013). Das ursprüngliche SAT-Modell (Chen & Barnes, 2014) scheint nach aktueller Forschung unzureichend für beispielsweise militärische Umgebungen zu sein. SAT-Informationen müssen dynamischer und bidirektionaler sein, um die Bedienenden bei der Entscheidungsfindung besser mit einzubeziehen.

Das SAT-Modell müsste im Bereich der bidirektionalen Kommunikation und im Bereich der Transparenz ergänzt werden. Das gemeinsame Handeln im HAT muss besser adressiert werden, um jedes Teammitglied bei kollaborativen Aufgaben mit einzubeziehen. Der Agent sollte dazu in der Lage sein, durch die Eingabe des Menschen zu lernen oder diese zu verstehen (Chen et al., 2018). Das bedeutet, dass der autonome Agent mit seinen Teamkollegen im ständigen Austausch der SAT-Ebenen steht. Folglich werden in dem dynamischen SAT-Modell zusätzlich die bidirektionale Kommunikation (vgl. Kapitel 2.3.1) und Transparenz (vgl. Kapitel 2.3.2) integriert. Das Verhältnis von Transparenz und bidirektionaler Kommunikation, die bidirektionale Transparenz, ist somit für effektive Leistung im HAT ausschlaggebend (Chen et al., 2018). In der folgenden Abbildung 2.5 wird dargestellt, wie der Agent und der Mensch die unterschiedlichen Ebenen des SAT-Modells miteinander kommunizieren, wodurch Transparenz entsteht und die Erreichung des Teamziels gefördert wird.

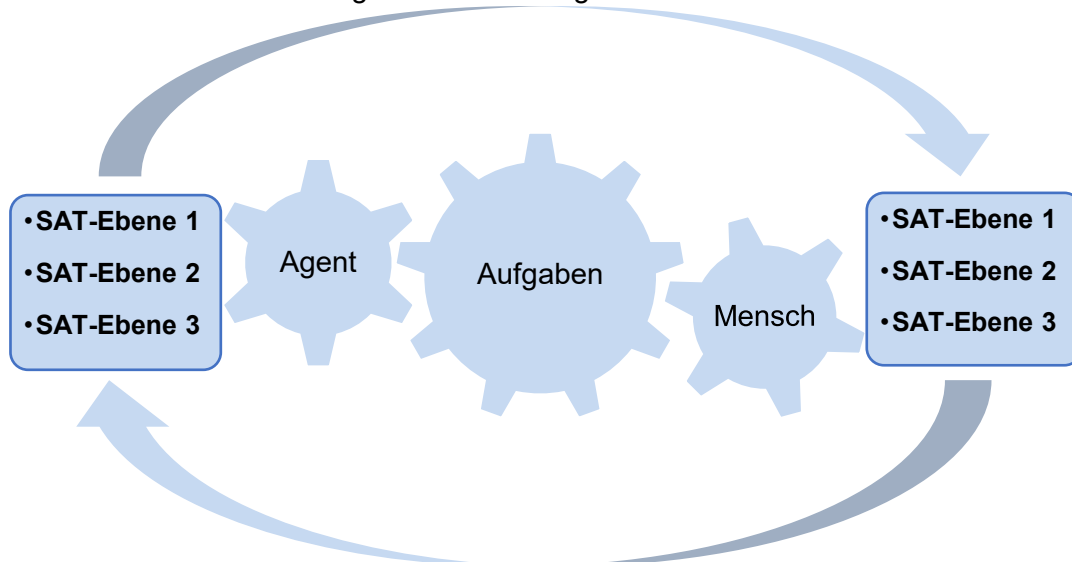


Abb. 2.5: Dynamisches SAT-Modell (in Anlehnung an Chen et al., 2018).

2.3.3 Koordination und Teamkognition

Christoffersen und Woods (2002) sagen, dass die Fähigkeit des gemeinsamen Systems, wirksam zu funktionieren, von der Koordinierungsfähigkeit der menschlichen und autonomen Teammitglieder abhängt. Damit ist gemeint, dass die Informationen, zu denen jedes Mitglied Zugang hat, für die Teamarbeit wirksam aufeinander abgestimmt und genutzt werden müssen (Christoffersen & Woods, 2002). Laut Sycara und Sukthankar (2006) gibt es drei wesentliche Aspekte der Interaktion zwischen Menschen und autonomen Agenten:

- Die gegenseitige Vorhersagbarkeit der Teammitglieder (Sycara & Lewis, 2004),
- das Teamwissen sowie ein gemeinsames Verständnis und
- die Fähigkeit, sich neu auszurichten (Christoffersen & Woods, 2002) und sich an andere anzupassen.

Bei allen dieser drei Elemente handelt es sich um einen eindeutigen und wirksamen Wissenstransfer (Sycara & Sukthankar, 2006). Diese genannten Aspekte werden in dem folgenden

Abschnitt erläutert. Abbildung 2.6 zeigt die drei grundlegenden Voraussetzungen für die Interaktion zwischen Menschen und autonomen Agenten.

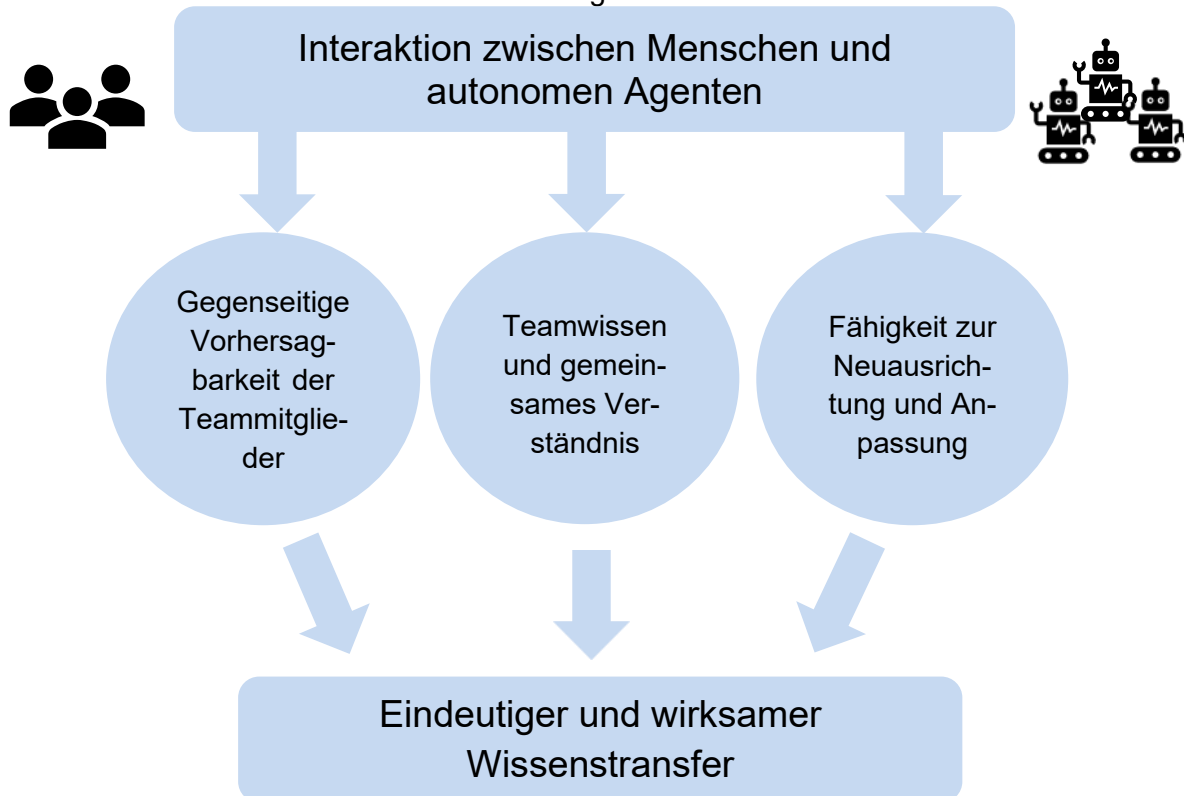


Abb. 2.6: Wesentliche Aspekte der Interaktion zwischen Menschen und autonomen Agenten (in Anlehnung an Christoffersen & Woods, 2002; Sycara & Lewis, 2004; Sycara & Sukthankar, 2006).

Beobachtbarkeit und gegenseitige Vorhersagbarkeit

Laut Grosz (1981) sowie McCarthy et al. (1991) ist die geteilte Darstellung einer Problemsituation eine Grundlage kooperativer Arbeit. In Abbildung 2.7 werden die Eigenschaften dieser Darstellung abgebildet. Christoffersen und Woods (2002) unterteilen diese gemeinsame Repräsentation in die Abbildung des Problemzustandes und in die Beschreibung der Aktivitäten von anderen Teammitgliedern. Diese beiden Teile sind dabei voneinander abhängig. Bei der Repräsentation des Problemzustandes geht es um ein gemeinsames Verständnis des Problems, welches zu lösen ist. Zum Beispiel sollte bei allen Teammitgliedern Klarheit über die Art des Problems, dessen Priorität und eine angemessene Lösungsstrategie herrschen. Die gemeinsame Repräsentation der Aktivitäten anderer Mitglieder des Teams beinhaltet die Informationen über die aktuellen Aufgaben der anderen, deren verfolgte Lösungsstrategie und den Grund für die Wahl dieser Strategie. Dazu gehören beispielsweise auch der Status ihrer Bemühungen und ihre Absichten für zukünftiges Handeln (Christoffersen & Woods, 2002).

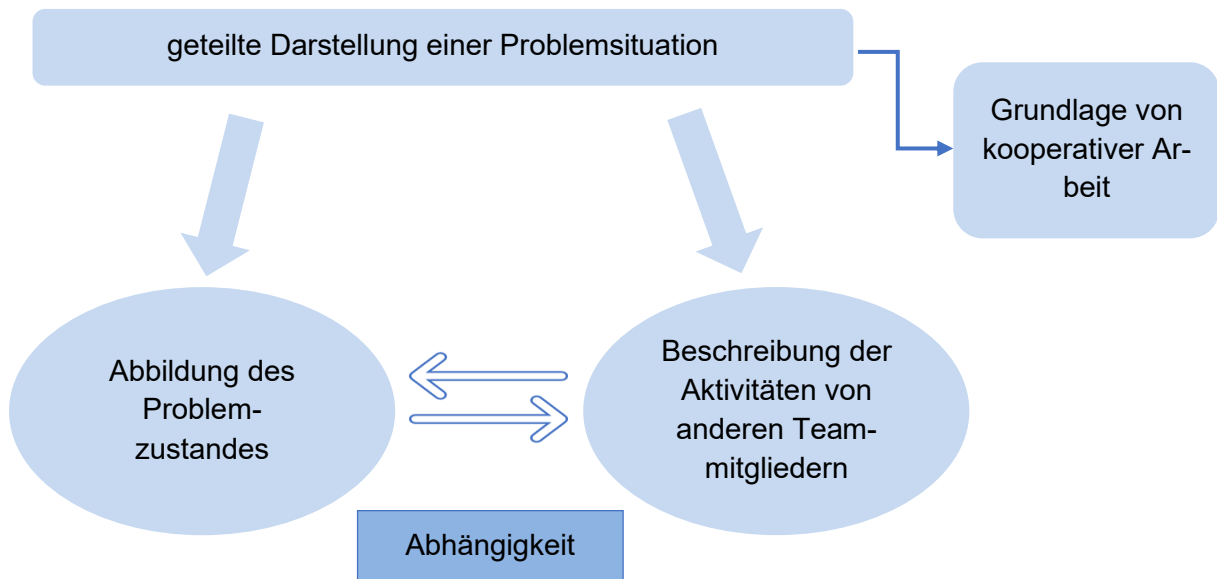


Abb. 2.7: Die gemeinsame Repräsentation der Problemsituation (in Anlehnung an Christoffersen & Woods, 2002; Grosz, 1981; McCarthy, 1991).

Gegenseitige Vorhersagbarkeit meint, dass für die erfolgreiche Koordination die Aktionen des einzelnen Teammitglieds für alle anderen Mitglieder genügend vorhersagbar sein müssen und dass diese versuchen, eine Einschätzung vieler Merkmale der Teamaktivität zu bilden (z.B. die benötigte Zeit eines Teammitglieds zur Erfüllung einer bestimmten Aufgabe; Sycara & Sukthankar, 2006). Nach Christoffersen und Woods (2002) muss der Mensch wissen, was der Agent tut und was er als nächstes tun wird. Eine effiziente und effektive Koordination kann durch die zuverlässigen Erwartungen über allgemeine Strategien und das kontextunabhängige Verhalten anderer hergestellt werden. Neben diesen Faktoren unterstützt das Wissen über die Situation der anderen die erfolgreiche Koordination (Patterson et al., 1999). Durch die Informationen über den Status und die Arbeit der anderen Teammitglieder wird jeder Einzelne dazu befähigt, die Bemühungen anderer in gewisser Weise vorherzusehen und so die eigenen Aktionen entsprechend zu koordinieren (Christoffersen & Woods, 2002). Gegenseitige Vorhersagbarkeit beinhaltet den Wissenstransfer zwischen den Teammitgliedern, was ihnen die Möglichkeit gibt, auf sinnvolle Weise zu kommunizieren und sich zu koordinieren. Wissen wird von Menschen implizit in ihrem Gehirn abgebildet. Damit andere Teammitglieder dieses Wissen verstehen können, müssen Menschen es explizit darstellen (Sycara & Sukthankar, 2006). Währenddessen müssen Agenten ihre Absichten klar kommunizieren und Ergebnisse verständlich machen (Lewis, 1998). So ist für die effektive Koordinierung gegenseitige Berechenbarkeit notwendig. Um die eigenen Aktivitäten planen zu können, muss der Mensch genau vorhersagen können, was die zukünftigen Aktivitäten der anderen Teammitglieder sein werden. Dieses gemeinsame Wissen muss vorliegen (Klein et al., 2004). Beobachtbarkeit und gegenseitige Vorhersagbarkeit stellen folglich wichtige Erfolgsfaktoren im Rahmen von HAT dar.

Teamwissen und mentale Modelle

Auch Teamwissen und mentale Modelle sind für ein erfolgreiches HAT unerlässlich. Teamwissen enthält das relevante gegenseitige Wissen, die Überzeugungen und Annahmen. Diese Überzeugungen und Annahmen unterstützen die interdependenten Handlungen, das Konstruieren und Verfolgen von Plänen, damit die Teamziele erreicht werden können. Dieses kann schon vorher vorhanden sein oder es muss erst während der Teamarbeit entwickelt werden (Sycara & Sukthankar, 2006). Entscheidend für das Verständnis der Teamleistung ist

Teamwissen, da es erklärt, wie die Mitglieder eines erfolgreichen Teams miteinander interagieren (Cannon-Bowers & Salas, 2001). Einen autonomen Agenten als Teammitglied anzusehen, ist eine Herausforderung (Klein et al., 2004) und effektive Funktionen der Teamarbeit sind notwendig (McNeese et al., 2018). Teams mit Menschen und Agenten koordinieren sich und arbeiten gemeinsam. Mit Koordination ist das Teilen von Wissen und ihre Abhängigkeit von dem Output der anderen Mitglieder gemeint. Unter Zusammenarbeit wird das Arbeiten an gemeinsamen Funktionen verstanden (Bradshaw et al., 2011). Zu den Funktionen gehören:

- das Verständnis der eigenen Aufgaben,
- das Bewusstsein der Aufgaben anderer (Salas et al., 2005) und
- die wirksame Interaktion mit den anderen Mitgliedern des Teams (Cooke et al., 2013; Gorman et al., 2010).

Das Team teilt einen gemeinsamen Wissensrahmen (Cooke et al., 2000). Teamwissen beinhaltet mehrere Arten von Wissen und alle müssen innerhalb eines effektiven Teams geteilt werden. Dazu gehören Kenntnisse über spezifische Aufgaben, aufgabenbezogene Themen, Eigenschaften von Teammitgliedern und Einstellungen sowie Überzeugungen der Mitglieder (Cannon-Bowers et al., 1993). Informationen, wie Stärken und Schwächen der Mitglieder des Teams und deren Einstellungen und Überzeugungen, sind universell und auch bei anderen Aufgaben anwendbar (Sycara & Sukthankar, 2006). Anhand einer Bedarfsanalyse oder einer Analyse der Teamaufgaben können die Rollen der einzelnen Teammitglieder, die gegenseitigen Abhängigkeiten und die Fähigkeiten, welche zur Rollenausführung und zur Koordination der interdependenten Aktivitäten notwendig sind, ermittelt werden (Arthur et al., 2005). Salas et al. (2005) betonen, dass für eine erfolgreiche Zusammenarbeit der Teammitglieder jedes einzelne Mitglied ein klares Verständnis der Rollen während der Tätigkeit, der verfügbaren Ressourcen und der Kapazitäten der Teammitglieder (z.B. aufgabenbezogene Präferenzen) haben muss. In einem Team aus Menschen und autonomen Agenten kommt es vor, dass für gewisse Aufgabenbereiche Fähigkeiten notwendig sind, die bei manchen Teammitgliedern stärker vorhanden sind als bei anderen. Es gibt Aufgaben, für die die unterschiedlichen Teammitglieder besser geeignet sind. Bei den Agenten wären das gefährliche oder sich wiederholende Tätigkeiten. Wenn die Rollen nach den Fähigkeiten der Teammitglieder zugewiesen werden, kann die Aufgabenausführung effizienter gestaltet werden (Hayes & Scassellati, 2013). Auch Clark und Brennan (1991) betonen die Bedeutsamkeit der Schaffung und Aufrechterhaltung einer gemeinsamen Basis für eine effektive Koordination. Diese gemeinsame Basis beinhaltet relevante Kenntnisse, Überzeugungen und Annahmen, welche die Menschen und Agenten teilen. Durch diese Basis entsteht die Möglichkeit, dass beide die Signale und Botschaften verstehen können. Sobald es Anzeichen für ein Zurückgehen der Gemeinsamkeiten gibt, sollten vorbeugende Maßnahmen ergriffen werden, um die Teamfunktion beizubehalten und eine Störung zu umgehen (Klein et al., 2004).

Mentale Modelle (vgl. Kapitel 2.3.1 für eine Definition von mentalen Modellen) können für die Vorhersage von Systemzuständen und das Ziehen von Rückschlüssen auf das Systemverhalten genutzt werden (Rouse & Morris, 1986). Durch geteilte kognitive Darstellungen wie diese können die Teammitglieder die Bedürfnisse und Aktionen vorhersehen und so auf implizite Weise ihr Verhalten koordinieren und ihre Teameffektivität verbessern (Cannon-Bowers et al., 1993). Teammentale Modelle sorgen für die Möglichkeit der Organisation und des Erwerbs von Informationen durch die Teammitglieder. Diese Informationen sind erforderlich, um Aktionen antizipieren und ausführen zu können (Kozlowski & Ilgen, 2006). Laut Klimoski und Mohammed (1994) beinhalten teammentale Modelle das gemeinsame und organisierte Verständnis und die mentale Darstellung von Wissen oder Überzeugungen. Diese sind für

grundlegende Aspekte der Aufgabenumgebung des Teams wichtig (Klimoski & Mohammed, 1994). Laut Mathieu et al. (2000) gibt es zwei Arten von mentalen Modellen:

- teambezogene mentale Modelle und
- aufgabenbezogene Modelle.

Letztere behandeln die Informationen über die benötigten Materialien sowie die Art und Weise, auf die die Ressourcen eingesetzt werden. Teambezogene Modelle betreffen die Teamarbeit und das erwartete Verhalten der anderen (Salas et al., 2005). Cannon-Bowers et al. (1993) schlagen vier Inhaltsbereiche von teammentalen Modellen vor, diese werden in Tabelle 2.2 dargestellt. Erstens das Wissen über die Ausrüstung und die Werkzeuge, welche von dem Team benutzt werden, zusammengefasst das Ausrüstungsmodell. Zweitens das Verständnis der Teamaufgabe, welches die Ziele, Leistungsanforderungen und Probleme beinhaltet, dies ist das aufgabenbezogene mentale Modell. Drittens die Kenntnis der Zusammensetzung und der Ressourcen der Teammitglieder, einschließlich des Wissens der einzelnen Mitglieder, deren Fähigkeiten, Präferenzen und Gewohnheiten. Dabei handelt es sich um das Teammitglied-Modell. Viertens das Wissen der Teammitglieder über angemessene oder effektive Prozesse, dies wird als Team-Interaktions-Modell bezeichnet (Cannon-Bowers et al., 1993). Wären diese gemeinsamen mentalen Modelle nicht vorhanden, könnten die Teammitglieder auf unterschiedliche Ziele hinarbeiten, was zu dem Unvermögen führt, die Handlungen oder Bedürfnisse der anderen zu antizipieren (Cannon-Bowers et al., 1990, 1993). Es wird vermutet, dass Teameffektivität durch ein gemeinsames Verständnis von der Aufgabe, dem Team, der Ausrüstung und der Situation verbessert werden kann (Cannon-Bowers et al., 1993). Wie in Unterkapitel 2.2.2 beschrieben, ist es für HAT notwendig, dass sich der autonome Agent dem menschlichen Verhalten anpasst. Aus diesem Grund wurde für diesen Abschnitt hauptsächlich Literatur aus der Teamforschung herangezogen, weil dies folglich auf autonome Agenten übertragbar sein sollte.

Tab. 2.2: Die vier Inhaltsbereiche von teammentalen Modellen (in Anlehnung an Cannon-Bowers et al., 1993).

Inhaltsbereich	Beschreibung
Ausrüstungsmodell	Wissen über die Ausrüstung und die Werkzeuge, welche von dem Team benutzt werden
Aufgabenbezogenes mentales Modell	Verständnis der Teamaufgabe, welches die Ziele, Leistungsanforderungen und Probleme beinhaltet
Teammitglied-Modell	Kenntnis der Zusammensetzung und der Ressourcen der Teammitglieder, einschließlich des Wissens der einzelnen Mitglieder, deren Fähigkeiten, Präferenzen und Gewohnheiten
Team-Interaktions-Modell	Wissen der Teammitglieder über angemessene oder effektive Prozesse

Richtbarkeit, gegenseitige Anpassungsfähigkeit und menschengesteuerte Ausführung

Damit Teams in unterschiedlichen Kontexten und bei unterschiedlichen Aufgabenanforderungen flexibel und agil sein können, sind Richtbarkeit und gegenseitige Anpassung erforderlich. Richtbarkeit meint die Zuweisung von Rollen und Verantwortlichkeiten an die Teammitglieder

(Sycara & Sukthankar, 2006). Eine reibungslose Arbeitsweise eines Teams kann durch Arbeitsteilung und Rollenverteilung erreicht werden. Durch diese Rollenzuweisung werden die erwarteten Aktionen definiert. So können Absichten oder zukünftige Handlungen einfacher vorhergesagt werden. Der autonome Agent wird zu einem effektiveren Teammitglied, wenn er basierend auf der Aufgabe Rollenzuweisungen erstellt. Diese werden mit anderen Mitgliedern synchronisiert und er wählt eine Rolle entsprechend den sozialen und praktischen Anforderungen nach aus (Hayes & Scassellati, 2013). Autonome Agenten können die Initiative ergreifen und sowohl Menschen oder anderen Agenten Anweisungen erteilen (Demir et al., 2018) als auch Aufgaben- und Teamarbeit durchführen (McNeese et al., 2018). Sie treffen auch während der Bearbeitung der Teamaufgabe eigene Entscheidungen über ihre Handlungen (Demir et al., 2018; McNeese et al., 2018). Damit autonome Agenten ihre Initiative angemessen ausüben können, brauchen sie ein klares Modell der Teamziele, Mitgliederrollen und des Teamverfahrens. Gegenseitige Anpassungsfähigkeit beinhaltet die Rollenänderung der Mitglieder zur Erfüllung der Teamanforderungen (Sycara & Sukthankar, 2006). Für effektive Koordination sollten die Teammitglieder während ihrer Aktivitäten angemessen auf den Einfluss anderer reagieren (Klein et al., 2004). Richtbarkeit und die gegenseitige Anpassung sind aufgrund ihres Ausdrucks von Wechselwirkungen innerhalb der Teamarbeit relevant. Wenn die einzelnen Teammitglieder und ihre Handlungen keine Auswirkungen auf die anderen Mitglieder und ihre Aktionen haben, liegt keine Koordination vor, sie arbeiten lediglich parallel (Sycara & Sukthankar, 2006).

Laut Christoffersen und Woods (2002) ist es für ein System zwischen Menschen und Agenten zentral, die Frage nach der Kontrolle zu klären. Damit ist gemeint, wer letztendlich die Kontrolle über das Problem und damit über die Problemlösungsstrategie hat. Der Mensch muss in die Agententätigkeiten eingreifen können, falls dieser Bedarf besteht (Christoffersen & Woods, 2002). Battiste et al. (2018) haben Rahmenbedingungen entwickelt, bei denen die Automation als Teammitglied angesehen werden kann. Diese Rahmenbedingungen beziehen drei Designgrundsätze des Crew Resource Management (CRM) mit ein. CRM ist ein Modell für erfolgreiche Teamarbeit, welches in den vergangenen Jahrzehnten von der Luftfahrt entwickelt wurde. Bei den Grundsätzen handelt es sich um die drei Elemente des HAT-Modells von Battiste et al. (2018):

- Transparenz,
- bidirektionale Kommunikation und
- menschengesteuerte Ausführung.

Bei Letzterem ist es erforderlich, dass ein Mensch die endgültigen Entscheidungen trifft und die dazugehörige Verantwortung trägt sowie dass diese Entscheidungen eindeutig sind. Die Fähigkeit, die finale Handlung auszuführen, liegt bei dem Menschen (Battiste et al., 2018). Wenn ein Mensch für die Ergebnisse der Teamarbeit verantwortlich ist, sollte er auch die effektive Autorität und die endgültige Kontrolle über die Art der Problemlösung haben (Billings, 1997). Das bedeutet, dass die Automation als Ressource bzw. Unterstützung bei der Lösung eines Problems designt wird. In Routinesituationen gelingt dies gut, in komplexeren Arbeitsumgebungen mit fortgeschrittener Automation ist es komplizierter. In der Praxis ist es jedoch meist unmöglich, automatisierte Systeme zu entwickeln, die jede eintretende Situation berücksichtigen können. Die Schwierigkeit liegt darin, dass strategische und taktische Entscheidungen notwendig werden, die außerhalb der Möglichkeiten der Automation liegen. Wenn diese schwierigen Situationen eintreten, ist die Frage, ob das gemeinsame System sich erfolgreich anpassen kann. Üblicherweise wurde dem Menschen eingeräumt, die Automation zu unterbrechen und das Problem selbst zu übernehmen. Das System wird auf zwei Betriebsmodi

beschränkt: komplett automatisch oder komplett manuell. Der Mensch wird gezwungen, sobald er eingreift, auf potenziell nützliche Funktionen der Automation zu verzichten. Notwendig sind kooperative Zwischenformen der Interaktion, die z.B. die Konzentration der Automation auf ein Teilproblem ermöglichen. Sie muss somit flexibel sein und Anweisungen gut annehmen können. Die Wichtigkeit der Richtbarkeit liegt darin, dass in kritischen Situationen das Nichtvorhandensein zu schwerwiegenden Konsequenzen führen kann. Diese Ergebnisse sollen nicht bedeuten, dass es sich bei der Automation um eine passive Ergänzung handelt. Dies wäre ein Rückschritt im Sinne des Gedankens „Mensch gegen Automation“ und eine Verschwendung der Potenziale beider. Der Mensch soll nicht nur die Automation managen, sondern er muss die Fähigkeit zum strategischen Handeln und zur Steuerung der Automationsaktivitäten bewahren.

Dies unterstützt die Effektivität des gesamten Systems (Christoffersen & Woods, 2002). Folglich wird deutlich, dass Richtbarkeit, eine gegenseitige Anpassungsfähigkeit und eine menschengesteuerte Ausführung relevant für den Erfolg von HAT sind. Die hier beschriebenen Anforderungen aufseiten des menschlichen Teammitglieds und des autonomen Agenten innerhalb eines HAT werden zusammenfassend in Infobox 2.3 dargestellt. In Abbildung 2.8 sind die Anforderungen aufseiten des Menschen, des autonomen Agenten und des Teams zusammenfassend dargestellt.

Infobox 2.3: Anforderungen aufseiten des menschlichen Teammitglieds und des autonomen Agenten.

Anforderungen aufseiten des menschlichen Teammitglieds und des autonomen Agenten:

- **Bidirektionale Kommunikation**
 - Training und Programmierung für sinnvolle Kommunikation
 - Verstehen der Teammitglieder
 - Bildung von mentalen Modellen
 - Umsetzung von Formen der Kommunikation
 - Nutzbare Sprachen
 - Nutzung der Verhaltensarten von Kommunikation
- **Transparenz**
 - Informationsaustausch
 - Verstehen der Teammitglieder
 - Projektion von zukünftigen Aktionen der autonomen Agenten
 - Situationsbewusstsein (Situation Awareness)
 - SAT-Modell
 - Bidirektionale Transparenz
- **Effektive und effiziente Koordination**
 - Beobachtbarkeit
 - Gegenseitige Vorhersagbarkeit
 - Teamwissen
 - Gemeinsames mentales Modell
 - Richtbarkeit
 - Gegenseitige Anpassungsfähigkeit
 - Menschengesteuerte Ausführung
 - Klärung der Kontrolle

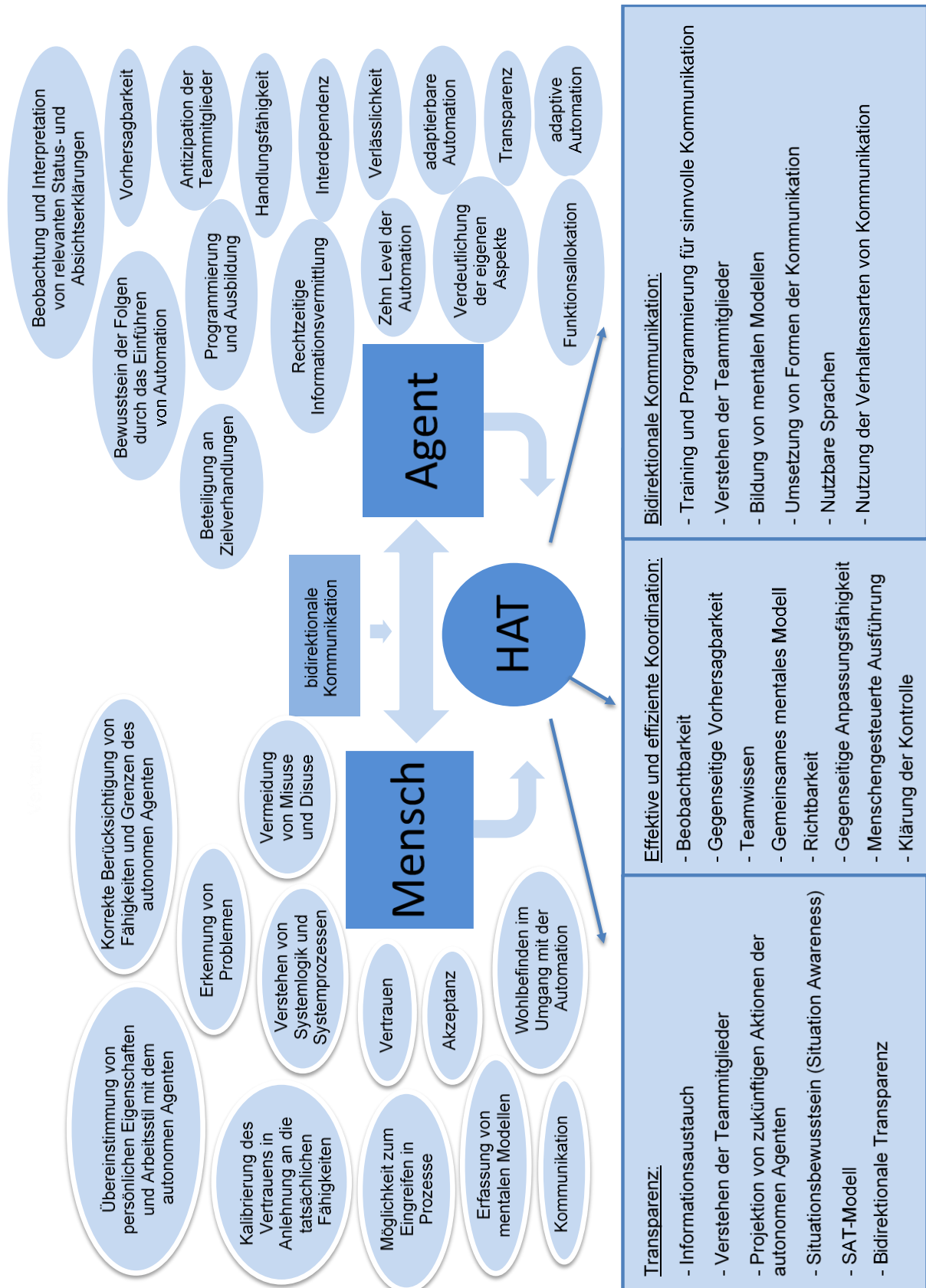


Abb. 2.8: Anforderungen aufseiten des Menschen, des autonomen Agenten und des Teams.

Literaturverzeichnis

- Allen, J. E., Guinn, C. I. & Horvitz, E. (1999). Mixed-initiative interaction. *IEEE Intelligent Systems*, 14(5), 14–23.
- Arthur, W., Edwards, B. D., Bell, S. T., Villado, A. J. & Bennett, W. (2005). Team task analysis: Identifying tasks and jobs that are team based. *Human Factors*, 47(3), 654–669.
- Battiste, V., Lachter, J., Brandt, S., Alvarez, A., Strybel, T. Z. & Vu, K.-P. L. (2018). Human-Automation Teaming: Lessons learned and future directions. In S. Yamamoto & H. Mori (Eds.), *Human Interface and the Management of Information. Information in Applications and Services. HIMI 2018. Lecture Notes in Computer Science* (Vol. 10905). Cham: Springer International Publishing.
- Billings, C. E. (1997). Aviation automation: The search for a human-centered approach. *Human Factors in Transportation*. Mahwah, NJ: Lawrence Erlbaum Associates Publishers.
- Bradshaw, J. M., Beautement, P., Breedy, M. R., Bunch, L., Drakunov, S. V., Feltovich, P. J., Hoffmann, R. R., Jeffers, R., Johnson, M., Kulkarni, S., Lott, J., Raj, A. K., Suri, N. & Uszok, A. (2004). Making agents acceptable to people. In N. Zhong & J. Liu (Eds.), *Intelligent Technologies for Information Analysis* (pp. 361–406). Berlin, Heidelberg: Springer.
- Bradshaw, J. M., Feltovich, P. J. & Johnson, M. (2011). Human-agent interaction. In G. A. Boy (Ed.), *The handbook of human-machine interaction: A human-centered design approach* (1. Aufl., pp. 283–300). Boca Raton, FL: CRC Press.
- Byrne, E. A. & Parasuraman, R. (1996). Psychophysiology and adaptive automation. *Biological Psychology*, 42(3), 249–268.
- Campion, M. A., Medsker, G. J. & Higgs, A. C. (1993). Relations between work group characteristics and effectiveness: Implications for designing effective work groups. *Personnel Psychology*, 46(4), 823–847.
- Cannon-Bowers, J. A. & Salas, E. (2001). Reflections on shared cognition. *Journal of Organizational Behavior*, 22(2), 195–202.
- Cannon-Bowers, J. A., Salas, E. & Converse, E. (1990). Cognitive psychology and team training: Training shared mental models and complex systems. *Human Factors Society Bulletin*, 33, 1–4.
- Cannon-Bowers, J. A., Salas, E. & Converse, E. (1993). Shared mental models in expert team decision making. In N. J. Castellan (Ed.), *Individual and group decision making: Current issues* (pp. 221–246). Mahwah, NJ: Lawrence Erlbaum Associates Publishers.
- Cannon-Bowers, J. A., Tannenbaum, S. I., Salas, E. & Volpe, C. E. (1995). Defining competencies and establishing team training requirements. In R. A. Guzzo & E. Salas (Eds.), *Team effectiveness and decision making in organizations* (pp. 333–381). San Francisco: Jossey-Bass.

- Chen, J. Y. C. & Barnes, M. J. (2014). Human-agent teaming for multirobot control: A review of human factors issues. *IEEE Transactions on Human-Machine Systems*, 44(1), 13–29.
- Chen, J. Y. C., Lakhmani, S. G., Stowers, K., Selkowitz, A. R., Wright, J. L. & Barnes, M. (2018). Situation awareness-based agent transparency and Human-Autonomy Teaming effectiveness. *Theoretical Issues in Ergonomics Science*, 19(3), 259–282.
- Chen, J. Y. C., Procci, K., Wright, J. L., Garcia, A., Barnes, M. & Boyce, M. (2014). *Situation awareness-based agent transparency (Report No. ARL-TR-6905)*. U.S. Army Research Laboratory.
- Chien, S.-Y., Lewis, M., Sycara, K., Liu, J.-S. & Kumru, A. (2016). Influence of cultural factors in dynamic trust in automation. *2016 IEEE International Conference on Systems, Man, and Cybernetics: Conference Proceedings* (pp. 2884–2889). Budapest, Hungary.
- Christoffersen, K. & Woods, D. D. (2002). How to make automated systems team players. In E. Salas (Ed.), *Advances in Human Performance and Cognitive Engineering Research: Automation* (1. Aufl.). Amsterdam, Boston: JAI.
- Clark, H. H. & Brennan, S. E. (1991). Grounding in communication. In L. B. Resnick & J. M. Levine (Eds.), *Perspectives on socially shared cognition: Revised papers presented at a conference* (1. Aufl., pp. 127–149). Washington, D.C.: American Psychological Association.
- Cook, M. B. & Smallman, H. S. (2008). Human factors of the confirmation bias in intelligence analysis: Decision support from graphical evidence landscapes. *Human Factors*, 50(5), 745–754.
- Cooke, N. J., Gorman, J. C., Myers, C. W. & Duran, J. L. (2013). Interactive team cognition. *Cognitive science*, 37(2), 255–285.
- Cooke, N. J., Gorman, J., Pedersen, H., Winner, J., Duran, J., Taylor, A., Amazeen, P. G., Andrews, D. H. & Rowe, L. (2007). Acquisition and retention of team coordination in command-and-control (Technical Report for AFOSR Grant FA9550-04-1-0234 and AFRL Award No. FA8650-04-6442). Washington, DC: Department of Defense.
- Cooke, N. J., Salas, E., Cannon-Bowers, J. A. & Stout, R. J. (2000). Measuring team knowledge. *Human Factors*, 42(1), 151–173.
- Demir, M. & Cooke, N. J. (2014). Human teaming changes driven by expectations of a synthetic teammate. *Proceedings of the Human Factors and Ergonomics Society Annual Meeting*, 58(1), 16–20.
- Demir, M., Cooke, N. J. & Amazeen, P. G. (2018). A conceptual model of team dynamical behaviors and performance in Human-Autonomy Teaming. *Cognitive Systems Research*, 52, 497–507.
- Demir, M., McNeese, N. J. & Cooke, N. J. (2016). Team communication behaviors of the human-automation teaming. *2016 IEEE International Multi-Disciplinary Conference on Cognitive Methods in Situation Awareness and Decision Support (CogSIMA)* (pp. 28–34). Piscataway, NJ.
- Demir, M., McNeese, N. J. & Cooke, N. J. (2019). The evolution of Human-Autonomy Teams in remotely piloted aircraft systems operations. *Frontiers in Communication*, 4, 50.

- Endsley, M. R. (1995). Toward a theory of situation awareness in dynamic systems. *Human Factors*, 37(1), 32–64.
- Endsley, M. R. (2017). From here to autonomy: Lessons learned from human–automation research. *Human Factors*, 59(1), 5–27.
- Feltovich, P. J., Bradshaw, J. M., Jeffers, R., Suri, N. & Uszok, A. (2004). Social order and adaptability in animal and human cultures as analogues for agent communities: Toward a policy-based approach. In A. Omicini, P. Petta & J. Pitt (Eds.), *Engineering Societies in the Agents World IV. ESAW 2003. Lecture Notes in Computer Science* (Vol. 3071, pp. 21–48). Berlin, Heidelberg: Springer.
- Fitts, P. M., Viteles, M. S., Barr, N. L., Brimhall, D. R., Finch, G., Gardner, E., Grether, W. F., Kellum, W. E. & Stevens S.S. (1951). *Human engineering for an effective air-navigation and traffic-control system*. Washington, DC: National Research Council.
- Gorman, J. C., Amazeen, P. G. & Cooke, N. J. (2010). Team coordination dynamics. *Nonlinear Dynamics, Psychology, and Life Sciences*, 14(3), 265–289.
- Grosz, B. J. (1981). Focusing and description in natural language dialogues. In A. K. Joshi, B. L. Webber & I. A. Sag (Eds.), *Elements of discourse understanding*. Cambridge, New York: Cambridge University Press.
- Hackman, J. R. (1987). The design of work teams. In J. Lorsch, *Handbook of organizational behavior*. Englewood Cliffs, NJ: Prentice-Hall.
- Hayes, B. & Scassellati, B. (2013). Challenges in shared-environment human-robot collaboration. 8. *ACM/IEEE International Conference on Human- Robot Interaction, Workshop on Collaborative Manipulation*. Tokyo, Japan.
- Inagaki, T. (1999). Situation-adaptive autonomy: Trading control of authority in human-machine systems. In M. W. Scerbo & M. Mouloua (Eds.), *Automation technology and Human performance: Current research and trends*. N.J: Lawrence Erlbaum Associates Publishers.
- Kaber, D. B. & Riley, J. M. (1999). Adaptive automation of a dynamic control task based on secondary task workload measurement. *International Journal of Cognitive Ergonomics*, 3(3), 169–187.
- Kaupp, T., Makarenko, A. & Durrant-Whyte, H. (2010). Human–robot communication for collaborative decision making – A probabilistic approach. *Robotics and Autonomous Systems*, 58(5), 444–456.
- Klein, G. (2004). *The power of intuition: How to use your gut feelings to make better decisions at work*. New York, NY: Currency Doubleday.
- Klein, G., Woods, D. D., Bradshaw, J. M., Hoffman, R. R. & Feltovich, P. J. (2004). Ten challenges for making automation a “team player” in joint human-agent activity. *IEEE Intelligent Systems*, 19(6), 91–95.
- Klimoski, R. & Mohammed, S. (1994). Team mental model: Construct or metaphor? *Journal of Management*, 20(2), 403–437.
- Kozlowski, S. W. J. & Ilgen, D. R. (2006). Enhancing the effectiveness of work groups and teams. *Psychological Science in the Public Interest : a Journal of the American Psychological Society*, 7(3), 77–124.

- Larson, L. & DeChurch, L. (2020). Leading teams in the digital age: Four perspectives on technology and what they mean for leading teams. *The Leadership Quarterly*, 31(1).
- Lee, J. D. (2012). Human-machine trust for robust autonomous systems. *Proceedings of the 4th IEEE Workshop on Human-Agent-Robot Teamwork*. Ocala, FL, USA.
- Lee, J. D. & See, K. A. (2004). Trust in automation: Designing for appropriate reliance. *Human Factors*, 46(1), 50–80.
- Lees, M. N. & Lee, J. D. (2007). The influence of distraction and driving context on driver response to imperfect collision warning systems. *Ergonomics*, 50(8), 1264–1286.
- Lewis, M. (1998). Designing for Human-Agent Interaction. *AI Magazine*, 19(2), 67–78.
- Lyons, J. B. (2013). Being transparent about transparency: A model for human-robot interaction (Technical Report SS-13-07). *Trust and autonomous systems: 2013 AAAI Spring Symposium* (pp. 48–53). Palo Alto, California.
- Lyons, J. B., Mahoney, S., Wynne, K. T. & Roebke, M. A. (2018). Viewing machines as teammates: A qualitative study. *Proceedings of the 2018 AAAI Spring Symposium* (pp. 160–177). Palo Alto, California.
- Lyons, J. B., Sadler, G. G., Koltai, K., Battiste, H., Ho, N. T., Hoffmann, L. C. et al. (2016). Shaping trust through transparent design: Theoretical and experimental guidelines. In P. Savage-Knepshield & J. Y. C. Chen (Eds.), *Advances in Human Factors in Robots and Unmanned Systems*. Advances in Intelligent Systems and Computing (Vol. 499, pp. 127–136). Cham: Springer International Publishing.
- Manzey, D. (2008). Systemgestaltung und Automatisierung. In P. Badke-Schaub, G. Hofinger & K. Lauche (Hrsg.), *Human Factors: Psychologie sicheren Handelns in Risikobranchen* (S. 307–324). Heidelberg: Springer Medizin Verlag.
- Marathe, A. R., Schaefer, K. E., Evans, A. W. & Metcalfe, J. S. (2018). Bidirectional communication for effective Human-Agent Teaming. In G. Fragomeni & J. Y. C. Chen (Eds.), *Virtual, augmented and mixed reality: Interaction, navigation, visualization, embodiment, and simulation* (Vol. 10909, pp. 338–350). Lecture Notes in Computer Science. Cham: Springer International Publishing.
- Mathieu, J. E., Heffner, T. S., Goodwin, G. F., Salas, E. & Cannon-Bowers, J. A. (2000). The influence of shared mental models on team process and performance. *The Journal of applied psychology*, 85(2), 273–283.
- McCarthy, J. C., Miles, V. C. & Monk, A. F. (1991). An experimental study of common ground in text-based communication. *CHI' 91: Proceedings of the 1991 Conference on Human Factors in Computing Systems* (pp. 209–215). New York, NY: ACM Press.
- McNeese, M. D. & McNeese, N. J. (2020). Humans interacting with intelligent machines: At the crossroads of symbiotic teamwork. In R. Pak, E. de Visser & E. Rovira (Eds.), *Living with robots: Emerging issues on the psychological and social implications of robotics* (pp. 165–197). Amsterdam: Elsevier.
- 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), 262–273.
- Mercado, J. E., Rupp, M. A., Chen, J. Y. C., Barnes, M. J., Barber, D. & Procci, K. (2016). Intelligent agent transparency in Human-Agent Teaming for Multi-UxV Management. *Human Factors*, 58(3), 401–415.

- Muir, B. M. (1987). Trust between humans and machines, and the design of decision aids. *International Journal of Man-Machine Studies*, 27(5-6), 527–539.
- Nass, C. I., Fogg, B. J. & Moon, Y. (1996). Can computers be teammates? *International Journal of Human-Computer Studies*, 45(6), 669–678.
- Norman, D. A. (1997). How might people interact with agents? In J. M. Bradshaw (Ed.), *Software agents* (pp. 49–55). Cambridge, MA: The AAAI Press/The MIT Press.
- O'Neill, T., McNeese, N. J., Barron, A. & Schelble, B. (2020). Human-Autonomy Teaming: A review and analysis of the empirical literature. *Human Factors. Advance Online Publication*. <https://doi.org/10.1177/0018720820960865>
- Onken, R. (1997). The cockpit assistant system CASSY as an on-board player in the ATM environment. *Proceedings of first air traffic management research and development seminar*. Saclay, France.
- Parasuraman, R., Mouloua, M. & Molloy, R. (1996). Effects of adaptive task allocation on monitoring of automated systems. *Human Factors*, 38(4), 665–679.
- Parasuraman, R. & Riley, V. (1997). Humans and automation: Use, misuse, disuse, abuse. *Human Factors*, 39(2), 230–253.
- Parasuraman, R., Sheridan, T. B. & Wickens, C. D. (2000). A model for types and levels of human interaction with automation. *IEEE transactions on systems, man, and cybernetics. Part A, Systems and humans : a publication of the IEEE Systems, Man, and Cybernetics Society*, 30(3), 286–297.
- Patterson, E. S., Watts-Perotti, J. & Woods, D. D. (1999). Voice loops as coordination aids in space shuttle mission control. *Computer Supported Cooperative Work: CSCW*, 8(4), 353–371.
- Rieth, M. & Hagemann, V. (2021). Veränderte Kompetenzanforderungen an Mitarbeitende infolge zunehmender Automatisierung – Eine Arbeitsfeldbetrachtung. *Gruppe. Interaktion. Organisation. Zeitschrift für Angewandte Organisationspsychologie (GIO). Advance Online Publication*. <https://doi.org/10.1007/s11612-021-00561-1>
- Rouse, W. B. (1988). Adaptive aiding for human/computer control. *Human Factors*, 30(4), 431–443.
- Rouse, W. B. & Morris, N. M. (1986). On looking into the black box: Prospects and limits in the search for mental models. *Psychological Bulletin*, 100(3), 349–363.
- Sadler, G. G., Battiste, H., Ho, N., Hoffmann, L., Johnson, W., Shively, R. J., Lyons, J. & Smith, D. (2016). Effects of transparency on pilot trust and agreement in the autonomous constrained flight planner. *35th DASC Digital Avionics Systems Conference – 2016 Proceedings Papers*. Sacramento, California.
- Salas, E., Sims, D. E. & Burke, C. S. (2005). Is there a “Big Five” in teamwork? *Small Group Research*, 36(5), 555–599.
- Scallen, S. F. & Hancock, P. A. (2001). Implementing adaptive function allocation. *The International Journal of Aviation Psychology*, 11(2), 197–221.
- Schaefer, K. E., Perelman, B. S., Brewer, R. W., Wright, J. L., Roy, N. & Aksaray, D. (2018). Quantifying human decision-making: Implications for bidirectional communication in human-robot teams. In G. Fragomeni & J. Y. C. Chen (Eds.), *Virtual, augmented and*

- mixed reality: Interaction, navigation, visualization, embodiment, and simulation* (Vol. 10909, pp. 361–379). Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics, Vol. 10909 LNCS). Cham: Springer International Publishing.
- Selkowitz, A. R., Lakhmani, S. G., Larios, C. N. & Chen, J. Y. C. (2016). Agent transparency and the autonomous squad member. *Proceedings of the Human Factors and Ergonomics Society Annual Meeting*, 60(1), 1319–1323.
- Selkowitz, A. R., Larios, C. A., Lakhmani, S. G. & Chen, J. Y. C. (2016). Displaying Information to Support Transparency for Autonomous Platforms. In P. Savage-Knepshield & J. Y. C. Chen (Eds.), *Advances in Human Factors in Robots and Unmanned Systems*. Advances in Intelligent Systems and Computing (Vol. 499, pp. 161–173). Cham: Springer International Publishing.
- Shattuck, L. G. (2015). Transitioning to autonomy: A human systems intergration perspective. *National Aeronautics and Space Administration*. Zugriff am 30.01.2021. Verfügbar unter <https://human-factors.arc.nasa.gov/workshop/autonomy/download/presentations/Shaddock%20.pdf>
- Shively, R. J., Lachter, J., Brandt, S. L., Matessa, M., Battiste, V. & Johnson, W. W. (2018). Why Human-Autonomy Teaming? In C. Baldwin (Ed.), *Advances in Neuroergonomics and Cognitive Engineering. AHFE 2017. Advances in Intelligent Systems and Computing* (586). Cham: Springer International Publishing.
- Sycara, K. & Lewis, M. (2004). Integrating intelligent agents into human teams. In E. Salas & S. M. Fiore (Eds.), *Team cognition: Understanding the factors that drive process and performance* (pp. 203–231). Washington, DC: American Psychological Association.
- Sycara, K. & Sukthankar, G. (2006). *Literature review of teamwork models* (Report No. CMU-RITR-06-50). Pittsburgh: Robotics Institute, Carnegie Mellon University.
- Tellex, S., Kollar, T., Dickerson, S., Walter, M. R., Banerjee, A. G., Teller, S. & Roy, N. (2012). Approaching the symbol grounding problem with probabilistic graphical models. *AI Magazine*, 32(4), 64–76.
- Walliser, J. C., Mead, P. R. & Shaw, T. H. (2017). The perception of teamwork with an autonomous agent enhances affect and performance outcomes. *Proceedings of the Human Factors and Ergonomics Society Annual Meeting*, 61(1), 231–235.
- Walliser, J. C., de Visser, E. J., Wiese, E. & Shaw, T. H. (2019). Team structure and team building improve human–machine teaming with autonomous agents. *Journal of Cognitive Engineering and Decision Making*, 13(4), 258–278.
- Warner, Norcross & Judd. (2015). Trucking, mining industries blazing a path to vehicle autonomy. Zugriff am 30.01.2021. Verfügbar unter <https://www.wnj.com/Publications/Trucking-Mining-Industries-Blazing-a-Path-to-Vehic>
- Wicks, A. C., Berman, S. L. & Jones, T. M. (1999). The structure of optimal trust: Moral and strategic implications. *The Academy of Management Review*, 24(1), 99.
- Wiener, E. L., Nagel, D. C., Friedman, M. P. & Carterette, E. C. (1988). *Human factors in aviation. Cognition and Perception*. Burlington: Elsevier Science.
- Wynne, K. T. & Lyons, J. B. (2018). An integrative model of autonomous agent teammate-likeness. *Theoretical Issues in Ergonomics Science*, 19(3), 353–374.

-
- You, S. & Robert, L. (2017). Teaming up with robots: An IMOI (inputs-mediators-outputs-inputs) framework of human-robot teamwork. *International Journal of Robotic Engineering*, 2(3), 1–7.
- Zuboff, S. (1988). Dilemmas of transformation in the age of the smart machine. In S. Zuboff (Ed.), *In the age of the smart machine: The future of work and power* (pp. 3–12). New York, NY: Basic Books.

Zusammenspiel von Menschen und autonomen Agenten

Regina Prediger, Rafael Stasiak & Lea Wenzel



Inhaltsverzeichnis

3	Zusammenspiel von Menschen und autonomen Agenten	
3.1	Gegenüberstellung von Human Autonomy Teaming und menschlichen Teams	76
3.1.1	Der Mensch und der autonome Agent als Teamplayer	
3.1.2	Interaktion des Teams	
3.1.3	Kommunikation zwischen Menschen und autonomen Agenten	
3.2	Stärken und Schwächen des Human Autonomy Teaming	81
3.2.1	Performance im Human Autonomy Teaming	
3.2.2	Auswirkungen des HAT unter soziologischen, wirtschaftlichen und psychologischen Aspekten	
3.3	Bausteine der Interaktion mit autonomen Agenten aus der Humanperspektive	88
3.3.1	Vertrauen	
3.3.2	Persönlichkeit	
3.3.3	Performance	
3.4	Zusammenfassung	103
	Literaturverzeichnis	105

Kapitel 3 Zusammenspiel von Menschen und autonomen Agenten

Von Regina Prediger, Rafael Stasiak und Lea Wenzel

In den letzten Jahrzehnten fand ein Wandel in der Forschung rund um die Beziehung von Menschen, Automation und autonomen Agenten statt. Hintergrund ist das jüngste Bestreben, eine Kooperation aus Menschen und Technik zu ermöglichen und zu gestalten (Eschen, 2013; Hoc, 2000). Die technologischen Fortschritte im Bereich der Automation eröffnen das weite Forschungsfeld, autonome Agenten in den Alltag des Menschen zu integrieren (O’Neill et al., 2020; Parasuraman & Riley, 1997). Zum Vergleich: Am Anfang genügte die Entwicklung einer Konzeption, an wen oder was welche Aufgaben optimaler delegiert werden können. Eine Kooperation aus Menschen und autonomen Agenten galt als Traumvorstellung (Eschen, 2013; Hoc, 2000; Parasuraman & Riley, 1997). Weiterhin wird sich mit dem Einzug der Autonomie das Verhältnis zwischen Menschen und Technik in Zukunft stark verändern (De Visser et al., 2018). McNeese et al. (2021) identifizieren den Wandel als Wendepunkt, an dem die Notwendigkeit besteht, alle Erkenntnisse und Vermutungen über menschliche Teamarbeit zu prüfen und auf das Human Autonomy Teaming (HAT) zu übertragen.

Diesem Wandel ist Kapitel 3 gewidmet. Es werden vier Anforderungen nach Klein et al. (2004) vorgestellt, wann der Mensch in konventionellen, menschlichen Teams als Teamplayer gilt. Dem werden zehn Herausforderungen gemäß Klein et al. (2004) gegenübergestellt, denen der autonome Agent hinsichtlich seiner Teamfähigkeit gerecht werden muss. Im Anschluss werden die thematisierten Erkenntnisse im Kontext der Interaktion im HAT und der Kommunikation zwischen Menschen und autonomen Agenten (vgl. Kapitel 2.3.1) gebündelt und präzisiert. Ein geläufiges Credo geht davon aus, dass Teams Aufgaben effizienter erledigen als eine Einzelperson. Allerdings birgt die Arbeit im Team nicht nur Vorteile. Im HAT entscheidet der korrekte Einsatz des autonomen Agenten über den Erfolg oder Misserfolg des Teams. Welche Eigenschaften und Voraussetzungen den Erfolg oder Misserfolg beeinflussen, wird als Stärken und Schwächen des HAT identifiziert. Im Anschluss werden die Auswirkungen des HAT in soziologischen und wirtschaftlichen Bereichen erläutert, bevor nach Blustein et al. (2019) psychologische Aspekte der sozialen Interaktion skizziert werden. Zum Schluss wird deutlich: Das Verhältnis zwischen Menschen und autonomen Agenten baut auf Vertrauen, Persönlichkeit und Performance auf.

3.1 Gegenüberstellung von Human Autonomy Teaming und menschlichen Teams

Wichtig für ein effektives HAT ist das richtige Zusammenspiel zwischen Menschen und autonomen Agenten. Seit langem werden Teams eingesetzt, um Arbeiten in einer Vielzahl von Aufgaben und Kontexten abzuschließen (McNeese et al., 2018; Salas et al., 2008). Traditionelle Teamarbeit ist definiert als zwei oder mehrere Menschen, die abhängig voneinander auf ein gemeinsames Ziel hinarbeiten (Salas et al., 1992; zitiert nach McNeese et al., 2018). Ein großer Teil des Wissens bezüglich Teamarbeit basiert auf Mensch-Mensch-Interaktion. Allerdings ist die Automation in den letzten Jahren mit fortschreitender Technologie zunehmend intelligenter geworden, um vermehrt als Teammitglied und nicht als Diener oder Werkzeug des Menschen betrachtet zu werden. Dieses Konzept der Zusammenarbeit zwischen Menschen und autonomen Agenten bezeichnet das HAT (McNeese et al., 2018). Aus diesem Grund ist eine effektive Kommunikation gemeinsamer Ziele, Modelle und eines gemeinsamen Verständnisses der Welt für eine lohnende Zusammenarbeit mit einem autonomen Agenten erforderlich, damit Handlungen gegenseitig verstanden und erwartet werden (Schaefer et al., 2015; Schaefer et al., 2017). Das HAT umfasst mindestens einen Menschen und einen autonomen

Agenten, die voneinander abhängig zusammenarbeiten und sich aufeinander abstimmen, um Aufgaben erfolgreich zu schaffen (Cuevas et al., 2007; Demir et al., 2019). Im HAT sind menschliche Teammitglieder und technologiebasierte Teammitglieder in gewissem Maße gleichwertig (Cuevas et al., 2007; Demir et al., 2019; Langan-Fox et al., 2009; Wijngaards et al., 2006). Autonome Agenten besitzen durch die fortgeschrittene Technologie die Fähigkeit, Initiative ergreifen zu können und sowohl menschlichen als auch automatisierten Teammitgliedern Befehle zu erteilen (Demir et al., 2019). Die Fähigkeit autonomer Agenten die Rolle eines Teammitglieds einnehmen zu können, leitete einen Paradigmenwechsel von rein menschlichen Teams zum HAT ein. In diesem Zusammenhang ist es erforderlich zu verstehen, wie sich das HAT von rein menschlichen Teams unterscheidet (Demir et al., 2019). In dem nachfolgenden Unterkapitel wird darauf näher eingegangen.

3.1.1 Der Mensch und der autonome Agent als Teamplayer

Der Mensch als Teamplayer

Am Anfang des Kapitels 3.1.1 wird zunächst auf rein menschliche Teams eingegangen. Dies hilft zu verstehen, wie sich rein menschliche Teams vom HAT unterscheiden. Klein et al. (2004) definieren gemeinsame Aktivitäten als eine erweiterte Reihe von Aktionen, die von einem Ensemble von Menschen sich aufeinander abstimmend ausgeführt werden. Für das Vorliegen einer gemeinsamen Aktivität müssen vier Voraussetzungen erfüllt sein. Eine Voraussetzung stellt das Schließen eines Basic Contracts zwischen den Teilnehmenden dar, in dem Vereinbarungen zur Zusammenarbeit getroffen werden. Die Handlungen der Teammitglieder müssen unter diesen vorhersehbar sein. Sie müssen gegenseitig steuerbar sein und Gemeinsamkeiten müssen bewahrt werden (Klein et al., 2004). Im Nachfolgenden werden die vier Anforderungen von rein menschlichen Teams detaillierter beschrieben.

- **Basic Contract:** Damit gemeinsame Aktivitäten durchführbar sind, schließt jede Partei einen Basisvertrag ab, eine Vereinbarung, die oft stillschweigend erfolgt. Der Basisvertrag gilt als Unterstützung zur Erleichterung der Koordination, auf gemeinsame Ziele hinarbeiten und Störungen in der Teamkoordination zu verhindern. Der Basic Contract muss kontinuierlich verstärkt und erneuert werden, es ist keine stets erfüllbare Voraussetzung. Er beinhaltet die Erwartung, dass die Mitglieder fehlerhafte gegenseitige Kenntnisse, Überzeugungen und Annahmen ausbessern, sobald diese entdeckt werden. Wichtig ist, dass beide Parteien sich sehr stark engagieren und Eigeninitiative zeigen, um Fehler zu vermeiden. Der Pakt verlangt, dass eine Partei den anderen Parteien signalisieren muss, wenn diese beabsichtigt aus der gemeinsamen Aktivität auszusteigen (Klein et al., 2004).
- **Gegenseitige Vorhersehbarkeit:** Teammitglieder sind auf ein angemessenes Maß an gegenseitiger Vorhersehbarkeit angewiesen, damit eine effektive Koordinierung der gemeinsamen Aktivität stattfinden kann. Die Planung der eigenen Aktionen ist nur möglich, wenn Teammitglieder genau vorhersehen können, was andere tun werden, vor allem bei stark voneinander abhängigen Aktivitäten (Klein et al., 2004).
- **Richtbarkeit:** Teammitglieder müssen steuerbar sein. Dies bedeutet, dass sie die Fähigkeit besitzen müssen, im Verlauf der Aktivität angemessen auf den Einfluss der anderen Teammitglieder zu reagieren, diesen zu bewerten und zu ändern (Klein et al., 2004).
- **Gemeinsamkeiten:** Für eine wirksame Koordinierung ist die Schaffung und Aufrechterhaltung von Gemeinsamkeiten erforderlich (Clark et al., 1991; zitiert nach Klein et al., 2004). Die Teammitglieder sollten geeignete Kenntnisse, Überzeugungen und

Annahmen teilen, damit sie in der Lage sind Botschaften und Signale zu verstehen, um z.B. bei einer Katastrophe vorbeugende Maßnahmen zu ergreifen (Klein et al., 2004).

Der autonome Agent als Teamplayer

Nachdem nun die Anforderungen für rein menschliche Teams näher beleuchtet wurden, werden als nächstes die Herausforderungen des HAT beschrieben. Die oben aufgezeigten relevanten Aspekte für rein menschliche Teams finden sich hierin wieder. Wie wird ein autonomer Agent zum Teamplayer? Klein et al. (2004) haben dazu zehn Herausforderungen formuliert, wie sich ein autonomer Agent zum Teamplayer entwickelt. Dies ist essenziell, da die Nachfrage nach einer Steigerung der Effektivität der Teamarbeit für komplexe Systeme, die eng und kooperativ mit Menschen zusammenarbeiten, zunimmt. In der nachfolgenden Tabelle werden diese zehn Herausforderungen charakterisiert.

Tab 3.1: Zehn Herausforderungen, um den autonomen Agenten zum Teamplayer zu entwickeln.

1. Basic Contract	Die erste Herausforderung kennzeichnet die Notwendigkeit des Erfüllens der Anforderungen eines Basic Contracts durch den autonomen Agenten, um gemeinsame Aktivitäten ausführen zu können. Falls ein autonomer Agent ausfällt und seine Rolle nicht mehr erfüllen kann, muss er in der Lage sein, seine Teammitglieder über tatsächliche und bevorstehende Fehler zu informieren (Cohen & Levesque, 1991; Klein et al., 2004). Er muss nicht nur in der Lage sein, einen Basic Contract zu schließen, sondern muss die gemeinsamen Ziele , seine Rolle in der Zusammenarbeit und die Notwendigkeit dieser Ziele verstehen und akzeptieren (Klein et al., 2004). Diese Herausforderung ist auch bei rein menschlichen Teams vorzufinden. Hier ist der Basic Contract ebenfalls eine Anforderung für effektive Teamarbeit.
2. Angemessene Modellierung	Autonome Agenten müssen in der Lage sein, die Absichten und Handlungen der anderen Teilnehmenden im Hinblick auf den Zustand und Entwicklung der gemeinsamen Aktivität angemessen zu modellieren . Beispielsweise, ob sie Probleme haben oder wie die anderen auf Störungen des Plans eingestellt sind (Klein et al., 2004).
3. Vorhersehbarkeit & 4. Richtbarkeit	Weitere Herausforderungen betreffen das Vertrauen . Autonome Agenten müssen steuerbar sein und Human Agent Teammitglieder gegenseitig vorhersehbar . Äquivalent zu einem Menschen muss ein autonomer Agent die Handlungen des anderen beobachten und einigermaßen hervorsehen können. Dem Menschen fällt es jedoch schwer, autonomen Agenten zu vertrauen , aufgrund der „Intelligenz“ und Autonomie der technischen Teammitglieder, da diese gegen die Vorhersehbarkeit der Menschen wirken. Autonome Agenten müssen genau wie ein menschliches Teammitglied die Fähigkeit besitzen, im Verlauf der Aktivität angemessen auf den Einfluss der anderen Teammitglieder zu reagieren , ihn zu bewerten und zu ändern (Bradshaw & Liu, 2004; zitiert nach Klein et al., 2004). Die Vertrauensdynamiken verlaufen dennoch aufgrund psychologischer Vorurteile in

	<p>umgekehrter Reihenfolge. Menschliche Teammitglieder nehmen den autonomen Agenten präferiert als “perfekt” und in seiner Funktionsweise zuverlässig wahr, sobald sie ihn als vertrauenswürdig einschätzen. Hintergrund ist übersteigertes Vertrauen in die Soft- und Hardware des autonomen Agenten. Umgekehrt begegnen menschliche Teammitglieder einander anfangs skeptisch und entwickeln während der Interaktion Vertrauen ineinander (Dzindolet et al., 2003). Ein idealer autonomer Agent sollte als guter Teamplayer, wie ein Mensch, vorhersagen können, welche Handlung der andere ausführen wird (Klein et al., 2004).</p>
5. Offenlegung der Absichten und des Status	<p>Autonome Agenten müssen die Fähigkeit besitzen, relevante Aspekte ihres Status und ihrer Absichten für alle Teammitglieder offen zu legen. Der Mensch muss verstehen, wieso und wie ein autonomer Agent in bestimmten Situationen handelt und wie er als nächstes agiert. Wenn das menschliche Teammitglied dies nicht nachvollziehen kann, kann es zu Verwirrungen kommen und zum Misstrauen in die Autonomie. Damit keine Fragen des menschlichen Teammitglieds offenbleiben, muss der autonome Agent seine Handlungen hinreichend vorhersehbar machen, indem er seine eigenen Ziele, Zustände, Kapazitäten, Absichten, Änderungen und bevorstehenden Handlungen für das menschliche Teammitglied offensichtlich darlegt (Feltovich et al., 2004; Klein et al., 2004).</p>
6. Kommunikation	<p>Die sechste Herausforderung stellt die Kommunikation des autonomen Agenten mit dessen Umwelt dar. Autonome Agenten müssen nicht nur in der Lage sein, Signale zu senden, sondern auch Signale des Menschen zu empfangen und richtig zu interpretieren (Klein et al., 2004).</p>
7. Zielverhandlungen	<p>Teammitglieder müssen in der Lage sein, an Zielverhandlungen teilzunehmen, insbesondere wenn sich eine Situation ändert und das Team sich anpassen muss. Autonome Agenten sollen in der Lage sein, ihre aktuellen und potenziellen Ziele mitzuteilen, zu begründen oder zu ändern. Sind sie dazu nicht in der Lage, dann beeinträchtigen sie die Koordination und die Aufrechterhaltung einer gemeinsamen Basis (Klein et al., 2004).</p>
8. Zusammenarbeit	<p>Der autonome Agent muss im System so gestaltet sein, dass er das Geben und Nehmen möglich macht, was die natürliche und effektive Teamarbeit zwischen Menschen und Gruppen auszeichnet (Klein et al., 2004).</p>
9. Aufmerksamkeitsmanagement	<p>Zur Aufrechterhaltung der gemeinsamen Verständnisgrundlage während koordinierter Aktivitäten, müssen die Teammitglieder die Aufmerksamkeit der anderen auf wichtige Signale, Aktivitäten und Änderungen lenken. Das übliche Automatisierungsdesign besitzt einen Alarm zum Überschreiten der Schwellenwerte (Ho et al., 2004; Klein et al., 2004).</p>

10. Kostenkontrolle	Alle Teammitglieder müssen dazu beitragen, die Kosten für koordinierte Aktivitäten zu kontrollieren . Die Kosten für die Koordination innerhalb eines HAT können schnell steigen, sodass die Teammitglieder sich generell bemühen, diese Kosten niedrig zu halten . Dabei sollten die Teammitglieder trotzdem kontinuierlich in die Koordination des Teams investieren (Klein et al., 2004). Im HAT können die autonomen Agenten effizienter, zuverlässiger und genauer auf umweltbezogene missionsbedingte Reize reagieren als ein menschlicher Teamkollege. Durch diese schnelle Reaktion ist es dem autonomen Agenten möglich, die Systemleistung zu verbessern , wenn die Systemleistung durch die menschliche Reaktionszeit begrenzt ist (Goodman et al., 2017).
---------------------	---

3.1.2 Interaktion des Teams

In den letzten Jahren hat die Technologie einen Punkt erreicht, an dem autonome Agenten zunehmend unabhängige, teamorientierte Entscheidungen treffen können. Die Mensch-Maschinen-Interaktion spielt eine große Rolle. Bei dieser Interaktion soll das menschliche Teammitglied nicht den autonomen Agenten kontrollieren, sondern mit diesem zusammenarbeiten. In diesen Arten von sozialen Netzwerken beeinflussen sich die Entscheidungen der Teammitglieder gegenseitig (Schaefer et al., 2017). Es ist zu erwähnen, dass die Interaktion mit einem autonomen Agenten nicht der Interaktion menschlicher Teammitglieder miteinander gleicht. Das menschliche Mitglied hat bestimmte Erwartungen an einen autonomen Agenten, dieser führt diese Erwartungen allerdings auf eine andere Weise aus als ein menschliches Teammitglied. Obwohl diese Aktionssequenzen des technischen Teammitglieds häufig erfolgreich bei der Ausführung ihrer Aufgabe sind, kommt es durch unbekannte und unerwartete Sequenzen zu einer Verschlechterung des Vertrauens der Anwendenden. Missbrauch oder Nichtnutzung des autonomen Agenten können resultieren (Parasuraman & Riley, 1997; Schaefer, 2013; Schaefer et al., 2017). Kapitel 1.4.2 schildert ausführlich, was Misstrauen und Nichtnutzung im Umgang mit autonomen Agenten bedeuten.

3.1.3 Kommunikation zwischen Menschen und autonomen Agenten

Um auf die Absicht eines Teammitglieds schließen zu können, muss diese mit den anderen Teammitgliedern kommuniziert werden. In diesem Zusammenhang wird zwischen expliziter und impliziter Kommunikation unterschieden (vgl. Kapitel 2.3.1). Die explizite oder auch gerichtete Kommunikation, verbunden mit impliziten, oft nonverbalen Merkmalen ist hier von wesentlicher Bedeutung. Implizite Merkmale sind z.B. Körperhaltung, Kopf- oder Augenbewegungen oder Änderungen der räumlichen Nähe. Explizite Kommunikationsmodalitäten sind z.B. Audio, Visualität oder der Tastsinn. Das Verstehen und Integrieren dieser expliziten und impliziten Merkmale in den autonomen Agenten erhöhen das Potenzial, die Absicht der Teammitglieder besser zu kommunizieren. Physische Konstruktionsmerkmale und das Funktionsverhalten geben dem Menschen Hinweise, wie Absichten des autonomen Agenten zu interpretieren sind. Aus diesem Grund ist es wichtig, diese physischen Konstruktionsmerkmale und das Funktionsverhalten zu verstehen (Schaefer et al., 2017).

Physische Konstruktionsmerkmale betreffen beispielsweise die Vermenschlichung von autonomen Agenten und Robotern. Durch vermenschlichte Designmerkmale, wie z.B. das Design des Augenlids und die Augenbewegung von Roboteraugen, ist es leichter, die

Kommunikation des emotionalen Zustands in der Mensch-Mensch-Interaktion widerzuspiegeln (Yamazaki et al., 2009; zitiert nach Schaefer et al., 2017). Das Einsetzen eines vollständigen Gesichtsdesigns in einen autonomen Agenten vereinfacht das Teaming, da es dem menschlichen Teammitglied einen Schwerpunkt zur bidirektionalen Kommunikation bietet. Mittel der verbalen und nonverbalen Kommunikation dienen im Grundsatz der zwischenmenschlichen Verständigung. Analog werden autonome Agenten zunehmend dazu befähigt. Durch Änderungen des Gesichtsausdrucks, der Sprachmuster (z.B. Lautstärke oder Intensität) und des Blicks oder körperlichen Bewegung werden Gefühle kommuniziert. Dies trägt dazu bei, das menschliche Verhalten zu ändern, um die Teamarbeit zu unterstützen. Zum Beispiel kann ein Mensch durch prosodische Variationen einem Menschen unterschiedliche Botschaften übermitteln, während alle dieselben Wörter verwenden. Verbales kann durch Nonverbales betont oder widerlegt werden. Botschaften sind beispielsweise Lob, Wut, Neugier, Verbot, Aufmerksamkeit, Trost, usw. Mit dem Einbau emotionaler Funktionen können Entwickler/innen das Design autonomer Agenten vorantreiben (Schaefer et al., 2017). Auch mithilfe nonverbaler Hinweise kommunizieren Menschen ihren emotionalen Zustand, z.B. affektiv mit Berührungen, Orientierung und Distanz von Partnern, Blickrichtungen und Gesten (Hayes & Scassellati, 2013; Schaefer et al., 2017). Hinsichtlich der Kommunikation wird erfolgreiches gemeinsames Verständnis zwischen Mitgliedern eines HAT durch spezifische, nonverbale Bewegungen und Körpergesten, wie Achselzucken (Gielniak & Thomaz, 2011; Schaefer et al., 2017), proxemische Verhaltensweisen und gegenseitigem Blickverhalten (Schaefer et al., 2017; Strabala et al., 2012) vermittelt. Proxemische Verhaltensweisen sind z.B. Veränderungen, die durch Überlegungen zum sozialen Raum diktiert werden (Schaefer et al., 2017; Wiltshire et al., 2013).

Es sei jedoch darauf verwiesen, dass diese Aspekte einfach in einen Roboter zu integrieren sind, jedoch weniger in softwarebasierte Automation. Folglich bedarf es hier anderer Elemente, um auch ohne diese Designmerkmale effektive Kommunikation erzielen zu können (vgl. 2.3.1). Einige Faktoren können allerdings ersetzt werden und einem menschlichen Teammitglied helfen, auf die Absicht des Verhaltens eines autonomen Agenten zu schließen. Das können Merkmale wie Geschwindigkeit und damit verbundene Änderungen des bewegungsbasierten Rauschens (Lohse et al., 2013; Schaefer et al., 2017) und bewegungsbasiertes Verhalten, wie beispielsweise Geschwindigkeit und Orientierung (Schaefer et al., 2017; Szafir et al., 2014; Wiltshire et al., 2013), sein. Es existieren noch viele Herausforderungen und potenzielle Verbesserungsmöglichkeiten in der effektiven Teamarbeit zwischen Menschen und autonomen Agenten. Wichtig ist es, dem menschlichen Teammitglied zu helfen, den Argumentationsprozess der Folgenabschätzungen für die Entscheidungsfindung in verschiedenen Bereichen, Aufgaben und Teamrollen besser zu verstehen. Allerdings werden zunehmend andere Methoden und Technologien zur Kommunikation von Absichten ohne physische Merkmale untersucht. Obwohl autonome Agenten andere Entscheidungsprozesse, Kontexte und Kommunikationsmittel aufweisen als menschliche Teammitglieder, ist es dennoch möglich Kommunikationsprotokolle zu entwickeln, die die Entwicklung des Bewusstseins für Teamsituationen erleichtern (Schaefer et al., 2017).

3.2 Stärken und Schwächen des Human Autonomy Teaming

HAT besitzt sowohl positive als auch negative Auswirkungen auf den Menschen sowie autonomen Agenten, die in diesem Kapitel als Stärken und Schwächen bezeichnet werden. Zunächst wird benannt, wann ein autonomer Agent eine positive und wann eine negative Auswirkung auf den Menschen haben kann. Daraufhin wird definiert, wie diese Auswirkungen aussehen. In diesem Unterkapitel wird dadurch deutlich, welche Effekte durch diese Stärken und

Schwächen entstehen und es werden die soziologischen, psychologischen und wirtschaftlichen Aspekte beleuchtet.

3.2.1 Performance im Human Autonomy Teaming

Beim HAT gibt es mindestens zwei Protagonisten/innen, zum einen den autonomen Agenten und zum anderen den Menschen. Durch beide kann die Zusammenarbeit sowohl positiv als auch negativ beeinflusst werden. Die voranschreitende Entwicklung künstlicher Intelligenz ermöglicht es autonomen Agenten zunehmend, in Teams zu arbeiten und Aufgaben effektiver zu bewältigen (Seeber et al., 2020). Bereiche der Teamarbeit sind z.B. die Koordination, die Aufgabenzuweisung und die Interaktionen mit Menschen oder anderen autonomen Agenten (Chen et al., 2016). Im Folgenden beziehen wir uns auf die Autonomielevel aus Abbildung 3.1 (vgl. Kapitel 1.3.3).

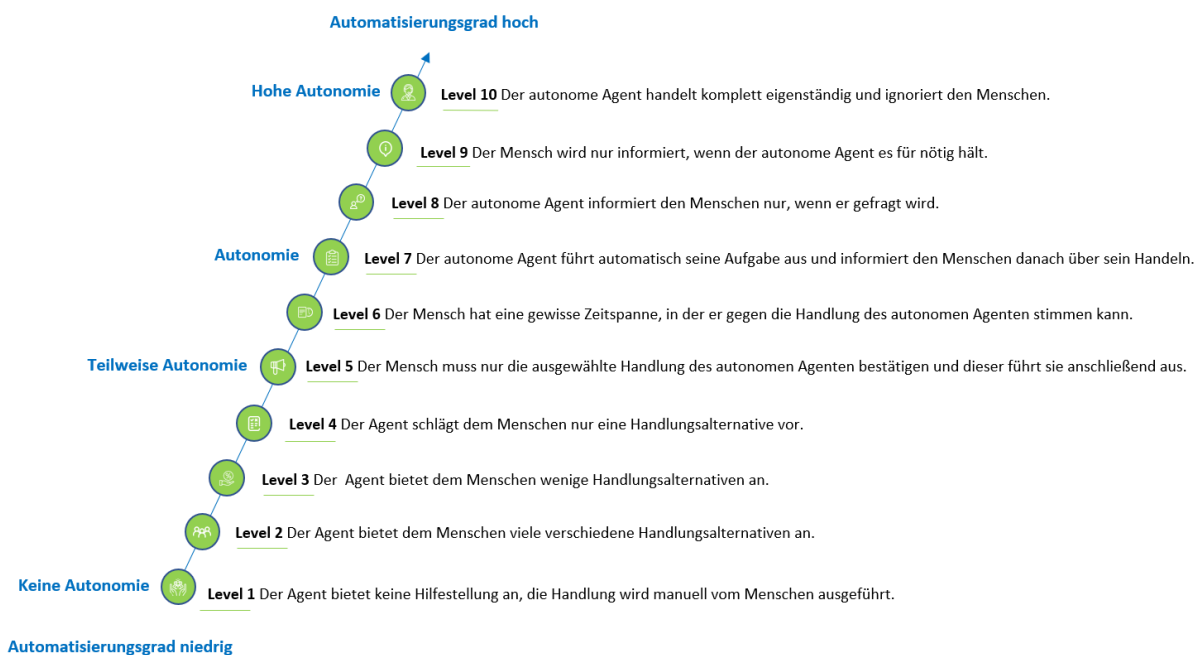


Abb. 3.1: Das Autonomiegrad-Modell in Anlehnung an O'Neill et al. (2020).

Ein höheres Autonomielevel hat einen positiven Effekt auf die Ergebnisse, die aus der Zusammenarbeit zwischen Menschen und autonomen Agenten resultieren. Menschen empfinden die Zusammenarbeit mit diesen autonomen Agenten als angenehmer, da diese kommunikationseffizienter und besser in der Performance sind (Lewis et al., 2003; Wright et al., 2013; zitiert nach O'Neill et al., 2020). Durch ein hohes Autonomielevel seitens des autonomen Agenten steigt das Vertrauen des Menschen gegenüber diesem. Analog steigt der Schwierigkeitsgrad des Aufgabenverständnisses (Azhar & Sklar, 2017). Andere Studien zeigen allerdings, dass autonome Agenten als besser wahrgenommen werden, wenn sie ein moderates Maß an Entscheidungsfindung liefern und somit ein mittleres Autonomielevel aufweisen. (Wright & Kaber, 2003, 2005). Niedrige bis mittlere gemeinsame Kontrolle führt zu einem höheren Erwartungsverhältnis als eine hohe gemeinsame Kontrolle (O'Neill et al., 2020). Das bedeutet, dass die Erwartungen bezüglich des Resultats höher sind, wenn der autonome Agent mehr Aufgaben übernimmt, als wenn die Aufgaben gleichermaßen auf den Menschen und den autonomen Agenten aufgeteilt werden.

Der Mensch, der mit dem autonomen Agenten zusammenarbeitet, spielt dabei eine große Rolle. Wenn der Mensch ein niedriges räumliches Vorstellungsvermögen besitzt, wird die

beste Performance erzielt, wenn zugleich der autonome Agent ein hohes Autonomielevel hat. Bei Menschen mit einer hohen Aufmerksamkeitskontrolle ist dies zusätzlich der Fall (Wright et al., 2013; Wright et al., 2018). Menschen mit einem hohen räumlichen Vorstellungsvermögen performen derweil schlechter, wenn sie mit solch einem autonomen Agenten zusammenarbeiten, da sich dadurch die situative Aufmerksamkeit verschlechtert (O'Neill et al., 2020). Ein mögliches Beispiel für eine schlechtere Performance bei Menschen mit einem hohen räumlichen Vorstellungsvermögen ist das autonome Parken. Beim autonomen Parken wird das Fahrzeug selbstständig von einer Fahrspur in eine Parklücke bewegt. Dadurch, dass der Mensch dabei nichts tun muss, verringert sich die situative Aufmerksamkeit. Die Menschen mit einem niedrigen räumlichen Vorstellungsvermögen oder mit einer hohen Aufmerksamkeitskontrolle profitieren durch solche Systeme. Sie können durch dieses System das eigenständige Parken verbessern, indem sie die stattfindenden Vorgänge beobachten. Welches Autonomielevel angemessen ist, hängt somit stark von dem Menschen ab, der mit dem autonomen Agenten zusammenarbeitet. Passt das Autonomielevel nicht zum Menschen, verschlechtern sich die Arbeitsergebnisse und das kann eine Schwäche des HAT sein. Ist das Autonomielevel angemessen, kann durch das HAT ein gutes Ergebnis erzielt werden, was eine Stärke des HAT darstellt.

Unter Transparenz wird verstanden, wie nachvollziehbar etwas ist. Bei einem autonomen Agenten mit einer niedrigen Transparenz verschlechtert sich die Performance (Mercado et al., 2016) und die Menschen werden selbstgefälliger (Wright et al., 2016). Selbstgefälliger bedeutet in diesem Kontext, dass die Menschen mehr auf ihre eigene Leistung vertrauen als auf die des autonomen Agenten. Wright et al. (2016) berichten, dass sich der Workload bei der Zusammenarbeit mit dem autonomen Agenten durch die Selbstgefälligkeit erhöht. Dies wurde durch Eigenbericht und dem Pupillendurchmesser der Testpersonen, welcher einen Indikator für mentale Beanspruchung darstellt, gemessen. Vermutet wird, dass eine höhere Transparenz mit einer größeren Informationsteilung assoziiert wird und dadurch eine intensivere Verarbeitungsanforderung erfolgt (O'Neill et al., 2020). Dadurch ist es dem Menschen möglich, die Aktionen des autonomen Agenten besser nachzuvollziehen. Durch einen Mangel an Transparenz kann der Mensch nicht nachvollziehen, wieso der autonome Agent in einer Situation so reagiert beziehungsweise nicht das tut, was der Mensch erwartet. Das menschliche Situationsbewusstsein (Situation Awareness) wird gefährdet (vgl. Kapitel 2.3.2). Als Beispiel dient der/die Pilot/in, der/die mit einem Flugmanagementsystem konfrontiert wird. Wenn der/die Pilot/in sich dabei fragt, was das System macht und wieso es dies macht, entsteht eine Unwissenheit. Dadurch kann der/die Pilot/in nicht mehr voraussagen, was als nächstes passieren wird und ob dieses Verhalten des autonomen Agenten angemessen ist (Shively et al., 2018). Im schlimmsten Fall entsteht als Folge mangelhafter Transparenz das Human-out-of-the-Loop-Performanceproblem (vgl. Kapitel 3.3.3). Die Transparenz stellt somit sowohl eine Stärke als auch eine Schwäche beim HAT dar, die die Teamleistung verbessert oder mindert.

Eine größere Verlässlichkeit des autonomen Agenten führt zu positiven Ergebnissen. Verlässlichkeit stellt das Level der Präzision und Richtigkeit dar, mit der ein autonomer Agent seine Arbeit erfüllt. Verschlechterungen der Zuverlässigkeit haben zur Folge, dass die Anzahl von schlechten Ergebnissen beim HAT zugleich ansteigt (Chen et al., 2011). Erhöhter Workload, schlechte Performance und niedriges Vertrauen sind weitere Eigenschaften, die mit einer niedrigeren Zuverlässigkeit assoziiert werden (O'Neill et al., 2020). Eine Stärke des HAT kann eine hohe Verlässlichkeit des autonomen Agenten sein. Ist der autonome Agent jedoch unzuverlässig, wird dies zu einer Schwäche des HAT.

Für ein effektives HAT ist es wichtig, dass der autonome Agent in der Umgebung zum Einsatz kommt, für die er erstellt wurde beziehungsweise alle möglichen Situationen berücksichtigt werden, die während des Einsatzes des autonomen Agenten auftreten können. Ansonsten können negative Folgen entstehen (Shively et al., 2018). Ein Beispiel für eine negative Folge, wenn nicht alle Situationen berücksichtigt werden, ist der Unfall eines Teslas aus dem Jahre 2016. Der Pkw ist unter einen Sattelschlepper geraten, während der Autopilot eingeschaltet war. Als möglicher Grund wurde angegeben, dass das System die weiße Seite des Anhängers vor dem Hintergrund eines hellen Himmels nicht erkannt oder für ein Autobahnschild gehalten habe (Merkel, 2019). Resultat ist: In der falschen Umgebung ist das HAT nicht nur ineffizient, sondern kann zudem eine Gefahr und eine große Schwäche darstellen.

Angemessenes Situationsbewusstsein bedeutet, dass sich der Mensch zu jeder Zeit bewusst ist, was der autonome Agent macht, warum er es macht und wie er weiterhin agieren wird (Badke-Schaub et al., 2012). Ein Mangel an geteiltem Situationsbewusstsein (Situation Awareness) bedeutet hingegen, dass das Teammitglied von einem autonomen Agenten oft nicht weiß, welche Informationen der autonome Agent nutzt, um seine Aufgabe zu erfüllen. Dies hat zur Folge, dass Fehler entstehen, weniger Vertrauen gegenüber dem autonomen Agenten existiert und dieser weniger genutzt wird. Die Möglichkeit besteht, dass das Teammitglied unzuverlässige oder unbrauchbare Informationen gibt, die der autonome Agent nutzt. Dadurch kann der autonome Agent fehlerhafte oder verwirrende Ergebnisse liefern (Shively et al., 2018). Onken (1997) beschreibt anhand des Beispiels „Cockpitassistenten“, dass bestimmte Faktoren, wie dem autonomen Agenten implantiertes Wissen über die Maschine sowie der aktuellen Flugsituation, und eine effiziente Kommunikation zwischen dem Menschen und dem autonomen Agenten zu einer Verbesserung des geteilten Situationsbewusstseins führen kann. Bei einem effizienten Informationsaustausch kann die Leistung des Menschen durch das HAT verbessert werden und eine Stärke des HAT sein. Ein ineffizienter Informationsaustausch führt zu Fehlern und stellt dadurch eine Schwäche des HAT dar.

Aus wenig Transparenz und einem Mangel an Situationsbewusstsein kann eine schlechte Einschätzung von Vertrauen gegenüber dem autonomen Agenten entstehen (vgl. Kapitel 2.3.2 & Kapitel 3.3.3). Unangemessenes Vertrauen in Form von zu großem Vertrauen kann zu einem häufigen Gebrauch des autonomen Agenten führen. Dadurch wird der autonome Agent eventuell unter Bedingungen gebraucht, für die er nicht erstellt wurde. Ein Misstrauen gegenüber dem autonomen Agenten führt dazu, dass dieser selten bis gar nicht genutzt wird. Dadurch kann das Potential nicht ausgeschöpft werden. Wenn der Mensch versteht, wann und wieso der autonome Agent etwas macht, kann der Mensch einschätzen, wann er sich auf den autonomen Agenten verlassen kann. Eine Möglichkeit blindes Vertrauen zu vermeiden, wäre wenn der autonome Agent mögliche Lösungen präsentiert und nicht nur eine einzige Option. Dies hätte zur Folge, dass die bedienende Person sich mit dem Thema auseinandersetzen muss und nicht nur dem Ergebnis vertraut (Shively et al., 2018). Im Januar 2018 fuhr ein Tesla-Fahrer in ein unbesetztes Feuerwehrauto. Aus einer Untersuchung folgte, dass der Fahrer den Autopiloten 14 Minuten vor dem Unfall aktiviert hatte und die letzten 13 Minuten das Lenkrad nicht mehr angefasst hatte. Dabei hatte sich der Fahrer auf andere Dinge als das Fahren konzentriert. Eine Ursache für den Unfall war vermutlich ein übermäßiges Vertrauen des Fahrers in den autonomen Agenten (Potor, 2019).

Ob das HAT eine Stärke oder Schwäche darstellt, hängt davon ab, ob der autonome Agent korrekt eingesetzt wird. Wenn alles richtig gemacht wird, kommen die Stärken des HAT zum Tragen und es stellt eine Bereicherung und Unterstützung dar. Die Schwächen des HAT werden deutlich, wenn der autonome Agent nicht verstanden oder in falschen Situationen eingesetzt wird. In Tabelle 3.2 werden die Stärken und Schwächen zusammengefasst.

Tab. 3.2: Zusammenfassung der Stärken und Schwächen.

Eigenschaft/ Voraussetzungen	Stärken	Schwächen
Autonomie	Wenn das Autonomielevel des autonomen Agenten zu dem Menschen passt, erfolgen gute Ergebnisse aus der Zusammenarbeit mit dem Menschen.	Wenn das Autonomielevel des autonomen Agenten nicht zu dem Menschen passt, erfolgen schlechte Ergebnisse aus der Zusammenarbeit mit dem Menschen.
Transparenz	Durch eine hohe Transparenz seitens des autonomen Agenten wird das Verständnis sowie die Leistung im HAT verbessert. Allerdings nur unter der Voraussetzung, dass kein Informationsüberschuss in Folge der hohen Transparenz stattfindet.	Durch eine niedrige Transparenz seitens des autonomen Agenten erfolgt weniger bis gar kein Vertrauen in den autonomen Agenten, wodurch die Leistung im HAT verschlechtert wird und zugleich erhöhter Stress seitens des Menschen auftritt.
Verlässlichkeit	Bei einem verlässlichen autonomen Agenten entstehen positive Auswirkungen auf die Ergebnisse, die im HAT zustande kommen.	Bei einem unverlässlichen autonomen Agenten entstehen negative Auswirkungen auf die Ergebnisse, die im HAT zustande kommen.
Umgebung	Der autonome Agent sollte stets in der Umgebung eingesetzt werden, für die er geschaffen wurde, damit der autonome Agent auf jede mögliche Situation reagieren kann.	Falls der autonome Agent in einer Umgebung eingesetzt wird, für die er nicht geschaffen wurde, können unerwartete Situationen auftreten, auf die der autonome Agent nicht reagieren kann.

Vertrauen	Angemessenes Vertrauen führt zu einem angemessenen Gebrauch des autonomen Agenten.	Zu viel beziehungsweise zu wenig Vertrauen führen zu einem unangemessenen Gebrauch des autonomen Agenten.
Informationsaustausch	Durch einen Austausch von korrekten Informationen ist die richtige Verarbeitung möglich, wodurch die Leistung verbessert werden kann und der Mensch in den Aufgaben entlastet wird.	Durch einen Austausch von falschen Informationen entstehen fehlerhafte Ergebnisse.

3.2.2 Auswirkungen des HAT unter soziologischen, wirtschaftlichen und psychologischen Aspekten

Soziologische Aspekte

In Österreich soll im Rahmen des Projektes „REAAAL“ älteren Menschen die selbstständige Lebensführung erleichtert und abgesichert werden. Dafür werden „Assistive Technologien“ beziehungsweise „Ambient Assisted Living (AAL)“-Technologien genutzt. Diese Technologien sollen die „Menschen in ihren alltäglichen Handlungen unterstützen und Einschränkungen bestmöglich kompensieren“ (Schneider et al., 2011). Im Vergleich ist „Smart Home“ bekannter als AAL. Allerdings bezeichnet „Smart Home“ die eingesetzte Technik, wobei AAL hinterfragt, wie diese Technik eingesetzt wird. Dabei werden Systeme zur Automatisierung und Kontrolle der Umgebung im Haus integriert. Aber auch Alarmsysteme, Sensoren zum Monitoring, Medikamentenerinnerungssysteme und Systeme zum Informationsaustausch können hinzugefügt werden. Besonders die letzten beiden Systeme können für ältere Menschen und Personen mit Einschränkungen von Hilfe sein. Die Daten, die durch die Systeme gesammelt werden, können an Angehörige, Ärzte und/oder Pflegekräfte weitergeleitet werden, wodurch eine Kontrolle ohne das direkte Dasein möglich wäre. Durch diese Systeme ist es den Menschen möglich, selbstständiger zu leben und ihren Alltag zu bewältigen. Zudem wird ein gewisses Gefühl von Sicherheit gewährleistet. Dadurch kann die Lebensqualität der zu betreuenden Person, aber auch die der Pflegekraft, beziehungsweise der Angehörigen steigen, da diese durch die Technik entlastet werden (Schneider et al., 2011). In diesem Beispiel stellen die Systeme, die in und um das Haus eingesetzt werden, den autonomen Agenten dar. Diese Systeme arbeiten zum einen mit dem Menschen, der im Haus lebt, und zum anderen mit den Pflegekräften, Angehörigen und Ärzten zusammen. Durch das HAT und die somit gewährleistete Selbstständigkeit sowie die Möglichkeit, den Alltag selber zu bewältigen, ist es den Menschen länger möglich, ein soziales Leben zu führen sowie die sozialen Beziehungen aufrechtzuerhalten.

Eine weitere soziologische Auswirkung kann die Automatisierung, Roboterisierung sowie die Einführung der künstlichen Intelligenz in die Arbeitswelt sein. Dies bringt zwar Chancen, aber auch Risiken mit sich. Wie schon beim technischen Fortschritt in der Vergangenheit kann die

Zunahme an Automatisierung, Roboterisierung und künstlicher Intelligenz dazu beitragen, dass die Arbeitsbelastung der Arbeitnehmenden verringert wird und die Produktivität zunimmt. Durch die Steigerung der Produktivität kann einerseits das Einkommen und der allgemeine Wohlstand erhöht werden, andererseits aber dazu genutzt werden, dass die Arbeitnehmenden weniger Arbeitszeit leisten müssen. Allerdings bringen die technologischen Fortschritte die Gefahr mit, dass viele Arbeitsplätze verloren gehen. Dies könnte zu einer erhöhten Arbeitslosigkeit führen. Beim Vergleich mit früheren technologischen beziehungsweise industriellen Revolutionen resultiert, dass durch die technologischen Fortschritte neue Berufe und Tätigkeiten entstehen. Diese können die verlorengegangenen Arbeitsplätze möglicherweise kompensieren (Krämer, 2019).

Wirtschaftliche Aspekte

In der Produktion werden immer mehr Roboter und autonome Agenten eingesetzt. Durch den Einsatz von Robotern oder autonomen Agenten im Fertigungssektor gehen nicht zwangsläufig alle Arbeitsplätze in diesem Bereich verloren. Die Produktionsfirmen werden dazu gebracht, neue Arbeitsplätze in anderen Sektoren, vor allem für junge Leute, zu kreieren. Für Menschen mit hohen Fähigkeiten, besonders im Bereich Wissenschaft und Management, hat es keine Einflüsse. Aber für Menschen mit wenigen Fähigkeiten hat es in Berufen, in denen Roboter und autonome Agenten eingesetzt werden, negative Folgen. Diese Menschen akzeptieren oftmals einen niedrigeren Lohn, sofern die Arbeitssicherheit aufrechterhalten wird. Ein positiver Aspekt an dem Einsatz von Robotern und autonomen Agenten ist, dass damit die Produktivität steigt (Dauth et al., 2017).

Ein weiteres Beispiel für die wirtschaftlichen Auswirkungen ist der Einsatz von autonomen Agenten im Flugverkehr. Brand et al. (2018) haben in ihrer Studie mit Piloten/innen der Bundeswehr die Delegation von Aufgaben an autonome Agenten sowie die Assistenz von den Agenten evaluiert. Der Einsatz der autonomen Agenten beziehungsweise Assistenzsysteme fand bei unbemannten Luftfahrzeugen (unmanned aerial vehicles, UAVs) statt. Mit der Studie wurde gezeigt, dass Assistenzsysteme maßgeblich zum Erfolg beitragen. Durch die eingesetzten Systeme wird die Sicherheit der Besatzung im Vergleich zu den heutigen Systemen deutlich erhöht. Zudem ist aus Sicht der Besatzung die flexible Führung von mehreren UAVs möglich. Der Einsatz der Assistenzsysteme vergrößerte den Handlungsspielraum der Besatzung soweit, dass kognitiv anspruchsvolle Aufgaben durch das Assistenzsystem für eine bestimmte Zeit vereinfacht oder gänzlich übernommen wurden. Vor allem in Stresssituationen oder Situationen, in denen eine hohe Beanspruchung der Besatzung herrschte, wurde Fehlleistung durch das System vermindert (Brand et al., 2018).

Aus wirtschaftlicher Sicht ist besonders die flexible Führung von mehreren UAVs interessant. Dadurch, dass der Flugverkehr sich vermehren kann, ist es denkbar, dass der Einsatz von UAVs auch für den Frachtverkehr genutzt werden kann. Dies hätte zur Folge, dass einerseits die Lieferdauer für Produkte aus dem Ausland verkürzt werden könnte, andererseits der Import und Export von den einzelnen Ländern zunehmen könnte. Ein weiterer positiver Aspekt für die Wirtschaft ist, dass durch den Einsatz von Assistenzsystemen die Besatzung beziehungsweise die einzelnen Besatzungsmitglieder entlastet werden. Durch die Delegation von Aufgaben und die mögliche Verminderung von Stresssituationen ist es den einzelnen Besatzungsmitgliedern möglich, auf kurze wie auf lange Sicht länger zu arbeiten. Auf kurze Sicht wären dies die Stunden pro Arbeitstag und auf langer Sicht die gesamten Arbeitsjahre eines Arbeitnehmenden beziehungsweise die Erhöhung des Renteneinstiegsalters.

Psychologische Aspekte

Eine anständige Arbeit hilft Menschen dabei, Selbstrespekt und Würde zu erreichen, sowie Freiheit und Sicherheit am Arbeitsplatz zu erleben. Die Möglichkeit, sich die Arbeit auszusuchen und diese bedeutsame, erfüllende und produktive Arbeit auszuführen, ist ein Privileg. Menschen, denen eine solche anständige Arbeit ermöglicht wird, hilft es, ein adäquates und uneingeschränktes Selbstkonzept zu entwickeln und einen sozialen Beitrag zu leisten (Blustein et al., 2019). Durch solch eine Arbeit kann der Mensch einige seiner psychologischen Grundbedürfnisse erfüllen. Zu den Grundbedürfnissen zählen beispielsweise das Bedürfnis zu überleben, Selbstbestimmtheit und soziale Verbundenheit, beziehungsweise Kontakte mit anderen Menschen. Im optimalen Fall kann die Arbeit diese Bedürfnisse erfüllen oder ihnen gerecht werden. Dadurch entsteht ein Fundament für die Erfüllung von weiteren psychologischen oder sozialen Bedürfnissen. Zugleich wirkt es sich positiv auf das Lebensgefühl der Person aus (Blustein et al., 2019). Laut der Selbstbestimmungstheorie streben Menschen danach Aktivitäten auszuführen, die sie als authentisch, selbstbestimmend oder motivierend ansehen (Ryan & Deci, 2000; zitiert nach Blustein et al., 2019).

Infolge von HAT können Arbeitsplätze verloren gehen oder Menschen werden dazu gebracht neue Tätigkeiten auszuführen, die sie als nicht erfüllend empfinden. Dadurch können negative Effekte auf die Erfüllung der Bedürfnisse, wie beispielsweise Selbstbestimmtheit eines Menschen, entstehen. Zudem führt ein Arbeitsplatzverlust oft zu einem Gefühl von Isolation, welches stark das psychologische Empfinden von Arbeitslosen beeinflussen kann. Für viele Menschen sind die sozialen Interaktionen, die sie am Arbeitsplatz haben, psychologisch gesehen, sehr wichtig. Durch die anderen Mitarbeiter wird sozialer Support, ein Gefühl von gemeinsamen Zielen und einer Gemeinschaft vermittelt. Meist fällt der Wegfall dieser wichtigen Aspekte erst auf, wenn eine Person ihren Arbeitsplatz verliert (Blustein et al., 2019).

Durch HAT ändern sich die sozialen Interaktionen des Menschen, weil sie überwiegend mit autonomen Agenten beziehungsweise technischen Einheiten interagieren. Dadurch stellen sich Fragen, wie beispielsweise:

- Mit wem kann der Mensch seine Probleme bereden?
- Ist eine menschliche Kommunikation weiterhin möglich?
- Verlernen Menschen dadurch ihre sozialen Fähigkeiten?

Hierzu braucht es in Zukunft Forschung aus allen Disziplinen, um die soziologischen, wirtschaftlichen und psychologischen Auswirkungen von HAT besser verstehen zu können.

3.3 Bausteine der Interaktion mit autonomen Agenten aus der Humanperspektive

Worauf baut das Verhältnis zwischen Menschen und autonomen Agenten auf? Dieser Fragestellung widmet sich das vorliegende Unterkapitel aus der Perspektive des Menschen. Ein zentraler Baustein ist das Vertrauen des Menschen in den autonomen Agenten. Uneinigkeiten beginnen bereits bei der Definition von Vertrauen (Lee & See, 2004). Zusätzlich ist zwischen Vertrauen und Zutrauen zu differenzieren, da diese zwei Größen ein unterschiedliches Maß an einzugehendes Risiko beinhalten (Abbass et al., 2016). Die technischen Anforderungen an den autonomen Agenten aus Kapitel 2.2 werden um einen Einblick in die Forschungen zu Trusted Autonomy (TA) ergänzt, um den Unterschied von Vertrauen und Zutrauen zu begründen (Abbass et al., 2016). Der Baustein Persönlichkeit prägt die Entstehung und Gestaltung von dynamisch erlerntem Vertrauen in autonome Agenten, bedarf allerdings weiterer Forschung (Hoff & Bashir, 2015; Kraus et al., 2020). Grundlegend ist die Autonomie in Human-in-

the-Loop (HITL-Systeme) und Human-on-the-Loop (HOTL-Systeme) einzuteilen (Nahavandi, 2017). Wirtschaft und Industrie messen der Performance im HAT einen hohen Stellenwert bei. Mit Blick auf die Teamleistung wird die Human-out-of-the-loop-Problematik sichtbar (Kaber & Endsley, 1997). Demir et al. (2017) sehen im HAT das höchste Potential, die Out-of-the-loop-Problematik zu umgehen.

3.3.1 Vertrauen

Wie Vertrauen entsteht

Die Forschung im Bereich der künstlichen Intelligenz, der Softwareagenten und der Robotik entwickelt sich zunehmend weiter. Trotz der absehbaren Chancen für Produktion, Wirtschaft sowie Gesundheits- und Bildungswesen gibt es heute nur wenige vollwertige autonome Agenten, die als solche den notwendigen Autonomiebedingungen gerecht werden. Die Vermutung liegt nahe, dass ein Mangel an Vertrauen in autonome Agenten eines der wesentlichen Hindernisse des Fortschritts auf dem Weg zu vollwertiger Integration ist (Abbass et al., 2016).

Um die Vermutung von Abbass et al. (2016) nachvollziehen zu können, ist die Entstehung von Vertrauen im Mensch-Autonomie-Kontext zu klären. Der Begriff Maschine ist in der Literatur geläufiger als der Ausdruck autonomer Agent. Nicht immer ist die Maschine gleichbedeutend mit der Definition eines autonomen Agenten im HAT nach O'Neill et al. (2020). Dennoch können direkte Rückschlüsse auf das Verhältnis zwischen Menschen und autonomen Agenten auch aus älteren Forschungen abgeleitet werden. Sofern wir nicht auf das Gegenteil hinweisen, sind die Begriffe Maschine und autonomer Agent als Synonyme zu verstehen. Im weiteren Verlauf wird eine Auswahl bekannter Ansätze, die Vertrauen definieren, vorgestellt.

- Der Ansatz von Muir (1987), der die Entstehung von Vertrauen in Mensch-Mensch-Interaktionen auf die Mensch-Maschine-Interaktion überträgt (vgl. Kapitel 2.1).
- Der Ansatz von Lee und Moray (1992), der eine Abhängigkeit des Vertrauens im Mensch-Automation-Kontext von drei Variablen identifiziert.
- Der Ansatz von Hoff und Bashir (2015), der Vertrauen im Mensch-Autonomie-Kontext grundlegend in drei Schichten teilt.
- Der Ansatz von Mayer et al. (1995), der die in der Literatur populärste Definition von Vertrauen im Mensch-Maschine-Kontext präsentiert.

Muir (1987) argumentierte, dass die gleichen Faktoren, die das Vertrauen von Menschen untereinander gestalten, das Vertrauen von Menschen in Maschinen beeinflussen (Muir, 1987; Parasuraman & Riley, 1997). Dieser in der Literatur vertretene Ansatz charakterisiert die Herleitung von Vertrauen als Erwartung von Verhaltensweisen und Ergebnissen (Abbass et al., 2016; Barber, 1983; Lee & See, 2004). Vertrauen in zwischenmenschlichen Beziehungen entsteht chronologisch im Zeitverlauf. Abbildung 3.2 illustriert das mehrdimensionale Konstrukt des Vertrauens (Abbass et al., 2016). Das Fundament für Vertrauen bilden technologische Unvollkommenheiten wie Zuverlässigkeit, Wahrung der Privatsphäre und zwei Arten von Sicherheiten. Sicherheiten beschreiben einerseits beispielsweise die Beständigkeit der Verfügbarkeit einer Computerfunktion (security) und andererseits das Abhängigkeitsverhältnis während kritischer Situationen beispielsweise in der Flugsicherheit. Ein Teammitglied, der/die Pilot/in, muss sich mangels Informationen auf das Urteilsvermögen eines weiteren Teammitglieds, des/der Fluglotsen/in, verlassen können (safety). Diese Konstrukte gelten als unvollkommen aufgrund der natürlichen Gegebenheit, dass hundertprozentige Vollkommenheit nicht erreichbar ist. Kognitive Unvollkommenheiten wie Komplexität der Umwelt, der Aufgabe und der Beteiligten, das Risiko des Vertrauensverstoßes und zuletzt der freie Wille bauen auf dem

Fundament auf. Das Risiko des Vertrauensverstoßes kennzeichnet die Verletzlichkeit, die mit dem Vertrauen in einen anderen einhergeht (Abbass et al., 2016; Parasuraman & Riley, 1997). Grundsätzlich dient Vertrauen als Bewältigungsmechanismus, um mit Risiken und Unsicherheiten umzugehen (Botsman, 2020).

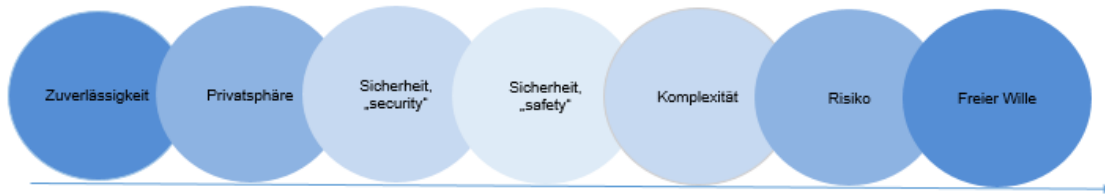


Abb. 3.2: Vertrauensentwicklung in einer Zeitreihe (in Anlehnung an Abbass et al., 2016, S.2810).

Hinsichtlich der Teamarbeit gilt Vertrauen als Maßstab für Gruppenkohäsion und Zugehörigkeit in menschlichen Teams (Krainz, 2011; zitiert nach Lackner, 2012). Das Vertrauen entwickelt sich als Prozess. Es ist dynamisch und gleichzeitig instabil. Menschen überprüfen ihr Vertrauen in andere Menschen fortwährend (Lackner, 2012). Dementsprechend prüfen menschliche Teammitglieder kontinuierlich, inwiefern sie dem autonomen Agenten in ihrem Team vertrauen können (Abbass et al., 2016; Lee & See, 2004). Als instabiles Produkt dynamischer Teamentwicklungen bewegt sich die Vertrauensbereitschaft innerhalb zweier gegensätzlicher Pole. Auf der einen Seite vertrauen Menschen blind ohne Reflexion, auf der anderen Seite zweifeln und misstrauen sie (Lackner, 2012). Die menschliche Vertrauensbereitschaft ist sehr stark erfahrungsgeprägt, variiert aber mittels differenzierter Entscheidungen des Individuums zwischen den sozialen Kontexten (Eberl, 2012). Beispielsweise entscheidet sich ein menschliches Teammitglied im HAT dazu, den vertrauten Hilfsassistenten im zivilen Luftverkehr zu vertrauen, misstraut aber einem neuen Kollegen. Optimismus und Pessimismus der menschlichen Teammitglieder beeinflussen deren Entscheidungen und Erfahrungen (Lee & See, 2004). Resultat ist, dass eine gelungene Teamarbeit sowohl Vertrauen voraussetzt als auch erschafft (Lackner, 2012; Meyer, 2001).

Nach Lee und Moray (1992) hängt das Vertrauen zwischen Menschen und Automation von der Leistung, dem Prozess und dem Zweck einer Automation ab. Fähigkeitsbasiertes Vertrauen variiert je nach Güte der Aufgabenausführung. Führt die Automation die delegierte Aufgabe gut aus, wächst das Vertrauen des Mitarbeitenden in die Automation. Prozessbasiertes Vertrauen ist abhängig vom Verständnis des menschlichen Mitarbeitenden für die Funktion der Automation bei der Bedienung. Diese Art des Vertrauens ähnelt dem Vertrauen zwischen Menschen basierend auf Sicherheiten. Zweckbasiertes Vertrauen unterliegt der von dem/der Entwickler/in beabsichtigten Nutzung der Automation, dem Zweck der Automation (Hoff & Bashir, 2015).

Hoff und Bashir (2015) klassifizieren Vertrauen dagegen in die drei Schichten Dispositionsvertrauen, Situationsvertrauen und erlerntes Vertrauen – zusammengefasst in Abbildung 3.3. Dispositionsvertrauen beschreibt die Vertrauensart, die am Anfang steht, vor jeglicher Interaktion mit einem bestimmten System. Es ist ein relativ stabiles Persönlichkeitsmerkmal, eine Art Urvertrauen in das Wohlwollen und die Vertrauenswürdigkeit der Umwelt. Faktoren wie das Geschlecht, das Alter, die eigene Kultur und individuelle Persönlichkeitseigenschaften beeinflussen das Dispositionsvertrauen. Diese Aufzählung ist allerdings nicht vollständig. Hoff und Bashir (2015) ermutigen, zukünftig neue Dimensionen zu erforschen, die

Dispositionsvertrauen in autonome Agenten anregen. Situationsvertrauen ist dem erlernten Vertrauen ähnlich, doch ist es abhängig von einem spezifischen Kontext, der sowohl von der externen Umwelt als auch den Eigenschaften des Menschen bestimmt wird. Externe Variablen und interne Variablen prägen das Situationsvertrauen. Extern gegeben sind z.B. die Schwierigkeit einer Aufgabe, die Arbeitsbelastung und die Arbeitsumgebung, während die jeweilige Aufmerksamkeitsspanne, das Selbstbewusstsein und die Stimmung intern begründet sind. Erlerntes Vertrauen entwickelt sich aus der Interaktion mit einem bestimmten System und basiert auf Urteilen und Erfahrungen, die in der Vergangenheit mit dem bestimmten System gemacht wurden. Die Rolle des Dispositionsvertrauens bei der Entwicklung des erlernten Vertrauens bleibt offen (Hoff & Bashir, 2015; Kraus et al., 2020). Das erlernte Vertrauen spaltet sich in anfänglich erlerntes Vertrauen und dynamisch erlerntes Vertrauen. Anfänglich erlerntes Vertrauen bezieht sich auf die eigenen Erwartungen, Werte und Erfahrungen aus ähnlichen Situationen. Das dynamisch erlernte Vertrauen setzt sich aus Variablen der Systemperformance und des Designs des autonomen Agenten zusammen, die während der Interaktion mit einem autonomen Agenten beobachtbar sind. Beispielsweise sind das die Nützlichkeit des autonomen Agenten, seine Zuverlässigkeit, seine Vertrauenswürdigkeit und der Grad der Schwere seiner Fehler. Hoff und Bashir (2015) erkennen Parallelen, wie Menschen einander vertrauen und wie sie Maschinen vertrauen. Ein potentieller Grund für die Parallelen ist, dass das menschliche Vertrauen in autonome Agenten tatsächlich das Vertrauen in die Entwickler/innen des autonomen Agenten repräsentiert (Hoff & Bashir, 2015; Parasuraman & Riley, 1997).

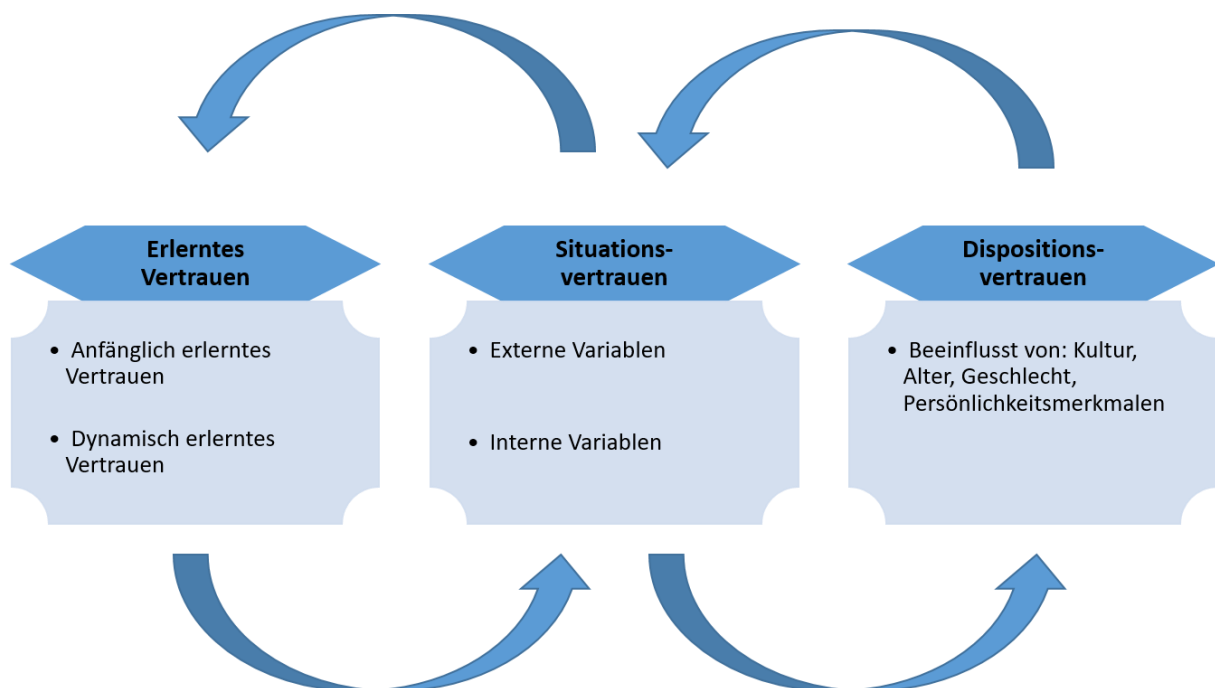


Abb. 3.3: Die drei Schichten des Vertrauens in autonome Agenten (in Anlehnung an Hoff & Bashir, 2015, S. 413).

Ein weiterer Ansatz definiert Vertrauen im Mensch-Maschine-Kontext als eine Absicht oder Handlungsbereitschaft (Lee & See, 2004; Mayer et al., 1995). Mayer et al. (1995) charakterisieren Vertrauen als die Bereitschaft einer Partei, sich unabhängig von der Fähigkeit zur übergreifenden Kontrolle den Handlungen einer anderen Partei auszusetzen. Diese Bereitschaft basiert auf der Erwartung, dass die andere Partei eine in der Absicht des Vertrauensgebers relevante bestimmte Handlung ausführen wird (Lee & See, 2004; Mayer et al., 1995). Das

bedeutet im HAT, dass ein menschliches Teammitglied mit dem autonomen Agenten interagiert, ungeachtet seiner Möglichkeiten den autonomen Agenten zu kontrollieren. Das menschliche Teammitglied lässt sich auf die Interaktion ein, weil es erwartet, dass der autonome Agent zum Wohl des Menschen agiert. Dies ist die in der Literatur am weitesten verbreitete und akzeptierte Herangehensweise (Lee & See, 2004; Rousseau et al., 1998). Die Relevanz von Verletzlichkeit und dem Eingehen von Risiken zur Entwicklung von Vertrauen und Vertrautheit steht im Fokus (Lee & See, 2004).

Wechselwirkung von Vertrauen und Autonomie

Eine erfolgreiche Teamarbeit setzt Vertrauen ineinander voraus und erschafft es (Lackner, 2012; Meyer, 2011). Analog basiert Autonomie grundlegend auf Vertrauen (Abbass et al., 2016, McNeese et al., 2021). Denn Autonomie beinhaltet die Fähigkeit, innerhalb eines begrenzten Spielraumes nach eigenem Ermessen Entscheidungen zu treffen. Dies beschreibt den freien Willen, der das Modell des mehrdimensionalen Vertrauenskonstrukts abschließt (Abbass et al., 2016). Autonomie steht in einem natürlichen Widerspruch zum Konzept der Koordination oder übergeordneten Konzepten wie z.B. Verantwortung. Denn ein Teammitglied verzichtet auf einen Teil seiner Autonomie, um sich mit anderen Teammitgliedern zu koordinieren und seine Verpflichtungen zu erfüllen (Huhns & Buell, 2002). Parallel bringt es sich in eine verletzbare Position, indem es andere mit der Verantwortung für Leistungen berechtigt (Lee & See, 2004). Das Teammitglied ist sich seiner Verletzlichkeit bewusst und wird abwägen, wie wahrscheinlich seine anderen Teammitglieder sein Vertrauen missbrauchen werden. Wenn ein menschliches Teammitglied im HAT einen autonomen Agenten als vertrauenswürdig einstuft, kann es Verpflichtungen an den vertrauenswürdigen autonomen Agenten delegieren. Wird dieses Vertrauen im Zeitverlauf intensiviert, verbessert sich die Leistung des Teams (Abbass et al., 2016; Parasuraman & Riley, 1997).

Auswirkungen von unangemessenem Vertrauen

Übermäßiges Vertrauen und Misstrauen beschreiben zwei Dilemmata der Interaktion mit autonomen Agenten (Lee & See, 2004; Parasuraman & Riley, 1997). Einerseits führen sie zu einer Überbeanspruchung des autonomen Agenten aufgrund zu starken Vertrauens in seine Leistungsfähigkeit (Positivity Bias). Andererseits werden gegebene Kapazitäten aufgrund von Misstrauen nicht ausgelastet (Parasuraman & Riley, 1997). Kapitel 1.4.2 schildert alle Dilemmata aus Disuse, Misuse und Abuse ausführlich.

Lee und Moray (1994) identifizieren eine menschliche Präferenz hinsichtlich der Nutzung einer Automation. Menschen bevorzugen es, Anlagen selbst zu steuern, wenn ihr Vertrauen in die eigenen Fähigkeiten größer ist als ihr Vertrauen in die automatisierte Steuerung (Lee & Moray, 1994; Parasuraman & Riley, 1997). Unvorhergesehene Situationen und die Komplexität der Funktionsweise ermöglichen kein vollständiges Verständnis des autonomen Agenten. Mangelnde Informationen über die Funktionsweise des autonomen Agenten führen zu Unsicherheiten, inwiefern er als vertrauenswürdig zu kalibrieren ist (Abbass et al., 2016; Lee & See, 2004; Parasuraman & Riley, 1997). Parallel reflektiert die Zuverlässigkeit eines autonomen Agenten dessen Nutzung, doch wirkt gelegentliche Unzuverlässigkeit angesichts eines zunehmend technischen Umfelds weniger abschreckend als vor der Jahrhundertwende. Regelmäßige Unzuverlässigkeit führt hingegen zu einer Vermeidung bis Weigerung der Nutzung. Eine schwerwiegende Problematik folgt, wenn der autonome Agent inkorrekte Leistungsinformationen liefert, aber der Mensch weiterhin auf die korrekte Funktion vertraut (Parasuraman & Riley, 1997). Forschungen dokumentieren, dass sich Collegestudierende und Piloten/innen weiterhin

auf eine unzuverlässige Automation verließen, obgleich in einem geringeren Ausmaß (Riley, 1994; zitiert nach Parasuraman & Riley, 1997). Diese Zusammenhänge bedeuten, dass der Mensch dem autonomen Agenten nicht blind vertrauen darf, um ihn unter den für ihn konzipierten Bedingungen zu nutzen. Umgekehrt bleiben Kapazitäten ungenutzt, wenn der Mensch der Funktion des autonomen Agenten pauschal misstraut (Parasuraman & Riley, 1997). Die technologische Verwundbarkeit von autonomen Agenten desillusioniert den Menschen in seiner Erwartung, alle Maschinen seien zur Unfehlbarkeit fähig (Huhns & Buell, 2002; Lee & See, 2004). Die Konsequenz ist: Transparente Informationen zur Komplexität einer Automation resultieren in angemessenem Vertrauen (Lee & See, 2004; Shively et al., 2018).

Einerseits legt bei teilautonomen Agenten die Forschung nahe, dass die Intensität der menschlichen Kontrolle mit zunehmendem Vertrauen in den teilautonomen Agenten abnimmt (Lee & See, 2004; Shively et al., 2018). Andererseits ersetzt Kontrolle das Vertrauen, wenn die Transparenz steigt. Beispielsweise messen Menschen ihre Körperfunktionen mit Fitness Trackern, statt ihrem Körpergefühl zu vertrauen. Das steht im Widerspruch zur gesellschaftlichen Annahme, mehr Transparenz führe zu mehr Vertrauen (Botsman, 2020). Insgesamt ist angemessen kalibriertes Vertrauen somit ein integraler Bestandteil von HAT und grundlegend für die Akzeptanz und Nutzung (Lee & See, 2004; Parasuraman & Riley, 1997; Shively et al., 2018).

Trusted Autonomy (TA)

Eine der heutigen technologischen Herausforderungen umfasst die Frage, wie Menschen lernen, autonomen Agenten zu vertrauen und umgekehrt (Abbass et al., 2016). Welche Anforderungen an einen autonomen Agenten modellieren die Arbeit im HAT? Kapitel 2.2 thematisiert bereits die technischen Anforderungen an den autonomen Agenten, um im HAT angemessen zu funktionieren. Kapitel 3.1 erarbeitet im Vergleich zu rein menschlichen Teams die Herausforderungen, einen autonomen Agenten zum Teamplayer zu formen. Der vorliegende Abschnitt ergänzt beide Bereiche um das Verständnis von Vertrauenswürdigkeit.

Trusted Autonomy (TA) ist ein Forschungsgebiet, das den Interaktionsraum zweier Entitäten gestaltet. Entitäten beschreiben den menschliche Agenten sowie autonome maschinelle Agenten oder eine Mischung beider. Relevant ist, dass sie einen gewissen Grad an Autonomie besitzen, um miteinander interagieren zu können. Das verwandte Forschungsgebiet der Cognitive Cyber Symbiosis (CoCyS) gilt als Erweiterung der bisherigen Sicht auf Interaktionen zwischen Mensch und Maschine, da die Ähnlichkeiten zwischen Informationsflüssen in einem elektromagnetischen Spektrum (cyber space) und Informationsflüssen in elektromagnetischen Gehirnbahnen (cognitive space) untersucht werden. CoCyS konzentriert sich auf den Aufbau und die Schnittstelle eines vertrauenswürdigen autonomen Systems (Trusted Autonomous System, TAS). Da CoCyS eine Cloud definiert, die Menschen und Maschinen zur Entscheidungsfindung einsetzt (Abbass et al., 2016; Collins et al., 2016), beschränken wir uns auf die Einführung in die TA, um den Rahmen der HAT-Thematik nicht zu überschreiten.

Den Ursprung von TA prägen Huhns und Buell (2002), indem sie die drei Bestandteile des systemischen Vertrauens identifizierten:

- Verständnis
- Interaktionsmanagement
- Philosophie und gesellschaftliche Konventionen

Diese drei Bestandteile wurden konkretisiert. Verständnis gilt als die Fähigkeit, Systeme auf einer hohen Abstraktionsebene zu repräsentieren, um das Verständnis der Systeme untereinander zu ermöglichen. Interaktionsmanagement wurde formuliert zur Fähigkeit, Interaktionen

der Systeme auf der Anwendungsebene zu verwalten. Philosophie und gesellschaftliche Konventionen wurden umgewandelt zur Transparenz der Ethik und Philosophie. Systemisches Vertrauen repräsentiert das Vertrauen in die Organisation oder in das eigene Team (Huhns & Buell, 2002).

Eine leitende Annahme der Forschung zu TA ist, dass ein gegenseitiges Vertrauensverständnis existieren muss (Abbass et al., 2016; Lee & See, 2004). Die Herausforderung der Definition eines TAS besteht allerdings in der Unterscheidung zwischen Vertrauen und Zutrauen (Abbass et al., 2016). Zusätzlich unterscheidet sich Vertrauenswürdigkeit als Persönlichkeitseigenschaft von Vertrauen als Handlungsbereitschaft oder Erwartung von Handlungsweisen (Botsman, 2020). Solange nicht alle Teammitglieder das mehrdimensionale Konzept von Vertrauen (vgl. Abb. 3.2) verstehen, ist es nicht möglich, die Teambeziehung als vertrauensvolle Beziehung zueinander zu definieren. Das bedeutet, dass ein autonomer Agent als vertrauenswürdig wahrgenommen wird, sobald seine Autonomiebefähigungen zu einem Verständnis und einer Interpretation von Vertrauen führen. Beispielsweise traut ein Mensch einer Maschine ohne Autonomiebefähigung lediglich Leistung zu. Wenn diese Maschine keine Leistung erbringt, gilt nicht die Maschine als unzuverlässig. Vielmehr riskiert der Hersteller der Maschine seine Reputation (Abbass et al., 2016).

Anforderungen an den autonomen Agenten resultieren. Ausgehend von der Systemperspektive muss ein vertrauenswürdiger autonomer Agent sich selbst über seine Ziele bewusst sein und außerhalb seines Vakuums mit anderen vertrauenswürdigen autonomen Agenten und menschlichen Teammitgliedern im HAT zur Zielrealisation kooperieren (Abbass et al., 2016). Grundlegend für die Vertrauensbildung ist die Fähigkeit, die Motivation der anderen Teammitglieder zu verstehen. Diese ermöglicht die Vorhersage der Verhaltensdynamiken im HAT, selbst wenn zukünftige Umwelt- und Teambedingungen nicht zu prognostizieren sind. Die Natur der Autonomie, im Unterschied zur Automation, betrifft die Anforderung, in komplexen Umwelten geprägt von Unsicherheiten zu operieren (Abbass et al., 2018). Die sozio-technologische Frage zur Modellierung eines autonomen Agenten als Basis der TA umfasst somit Anforderungen aus denen Verhaltensdynamiken im HAT entstehen. Der autonome Agent muss mittels künstlicher Intelligenz (KI) befähigt sein, seine menschlichen Teammitglieder proaktiv zu spüren, ähnlich der menschlichen Teammitglieder untereinander durch Körpersprache, Mimik und/oder Stimme. Das beinhaltet ein kontinuierliches Erfassen, Verarbeiten, Reagieren und Überprüfen der Umgebung, in welcher der vertrauenswürdige autonome Agent operiert. Weitere Anforderungen an den autonomen Agenten sind die einwandfreie Identifikation seiner Teammitglieder zur Unterscheidung von anderen Menschen und autonomen Agenten sowie die selbstständige Beurteilung des Grads der Vertrauenswürdigkeit seiner menschlichen Teammitglieder (Abbass et al., 2016; Cooke et al., 2012). Autonome Agenten benötigen eine intrinsische Motivation, gleichbedeutend mit internen Motivationsprozessen, wenn sie erfolgreich mit unvorhersehbaren komplexen Umwelten umgehen sollen. Intrinsische Motivation ist erforderlich, weil die Programmierungen, die die Agentensteuerung definieren, nicht vor der Ausführung der Aufgabe festgelegt werden können (Abbass et al., 2018; Sun, 2018). Dementsprechend ist ein hohes Maß an Intelligenz erforderlich. Resultat ist die Dimension kalibrierten Vertrauens, in der das Vertrauen des Menschen die Fähigkeiten und die Leistung des autonomen Agenten tatsächlich widerspiegelt (Kraus et al., 2020).

3.3.2 Persönlichkeit

Persönlichkeit und Vertrauen

Übermäßiges Vertrauen und Misstrauen repräsentieren problematische Nutzungsmuster im Umgang mit autonomen Agenten (Lee & See, 2004; Parasuraman & Riley, 1997). Ein wissenschaftliches Verständnis von Vertrauensbildungsprozessen kann die Auswirkungen solcher Nutzungsmuster begrenzen, insbesondere im komplexen HAT (Kraus et al., 2020). Dazu ist die Definition des Persönlichkeitsbegriffs notwendig. Die Forschung von Kraus et al. (2020) baut auf der Definition der Persönlichkeit als eine Aneinanderreihung organisierter und relativ dauerhafter psychologischer Charakteristika und Mechanismen innerhalb eines Individuums, die seine Interaktion mit anderen beeinflusst, auf (Kraus et al., 2020; Larsen & Buss, 2005). Resultat der Definition ist, dass die Persönlichkeit die Verarbeitung von Informationen und damit die Interaktion mit autonomen Agenten beeinflusst. Dementsprechend gilt die Persönlichkeit als Prädiktor für Vertrauen in autonome Agenten (Kraus et al., 2020). Kapitel 5.2.6 führt den Baustein Persönlichkeit im HAT-Kontext mittels unterschiedlicher Persönlichkeitstypen unabhängig von der Vertrauensthematik fort.

Die Forschung von Kraus et al. (2020) untersucht ein hierarchisches Persönlichkeitsmodell des dynamisch erlernten Vertrauens in hochautomatisiertes Fahren. Sie testeten Teilnehmende in zwei Studien online mittels Persönlichkeitsfragebögen, der Konfrontation mit einem hochautomatisierten Fahrsystem und anschließenden Animationen. Kraus et al. (2020) beabsichtigten zu untersuchen, ob und wie unterschiedliche Persönlichkeitsmerkmale das dynamisch erlernte Vertrauen in autonome Agenten beeinflussen. Das dynamisch erlernte Vertrauen nach Hoff und Bashir (2015) stellt das Vertrauen während einer Interaktion dar. Beispielsweise lernt ein Mensch mit der Zeit seinem Herzschrittmacher zu vertrauen, ihn am Leben zu erhalten. Das Vertrauen in den Herzschrittmacher ist nicht von Anfang an gegeben (Hoff & Bashir, 2015).

Basis der Studien von Kraus et al. (2020) ist das 3M-Modell von Mowens (2000). Das Meta-theoretische Motivationsmodell (3M-Modell) erklärt, wie Persönlichkeitsmerkmale mit der Situation interagieren, um die Situation zu beeinflussen (Chen & Barnes, 2013). Die von Kraus et al. (2020) modifizierte Version ist mittels Abbildung 3.4 im Kontext der Forschungsergebnisse illustriert. Im Original besteht die untere Ebene aus breiten genetischen Verhaltensveranlagungen und frühen Kindheitserlebnissen, die grundlegende Merkmale darstellen. Die zweite Ebene konstruiert zusammengesetzte Merkmale, beispielsweise Selbstwirksamkeit, Wettbewerbsfähigkeit und Aufgabenorientierung. Kraus et al. (2020) fassten die beiden unteren Ebenen des Modells zusammen, um eine Ebene elementarer Merkmale zu formulieren. Eine Ebene höher resultieren im modifizierten Modell aus der vorherigen Ebene situationsbezogene Merkmale. Situationsbezogene Merkmale rufen eine dauerhafte Verhaltenstendenz in allgemeinen situativen Kontexten hervor, z.B. impulsives Kaufen oder sportliches Interesse. Beide Ebenen der Modifikation zusammen ergeben Oberflächenmerkmale, die Verhaltenstendenzen beschreiben, welche in bestimmten Verbraucherdomänen gelten (Chen & Barnes, 2013; Kraus et al., 2020; Mowen, 2000). Abbildung 3.4 zeigt die Verteilung der Persönlichkeitsmerkmale auf die Ebenen. Offenheit, Gewissenhaftigkeit, Extraversion, Verträglichkeit und Neurotizismus (Big Five) und Selbstwertgefühl sowie Kontrollüberzeugung (locus of control, LOC) bilden die Grundlage der Persönlichkeitsstruktur. Affinität für Technologie, Technologische Selbstwirksamkeitserwartung und Dispositionsvertrauen ordneten Kraus et al. (2020) der zweiten Ebene, den situationsbezogenen Merkmalen, zu. Aus den unteren Ebenen entwickeln sich die Vertrauensneigung in Automation und die a priori Akzeptanz vom autonomen Fahren. Forschungsrelevante Verhaltensweisen werden direkt oder indirekt durch eine

Merkmalskombination der drei Hierarchieebenen beeinflusst. Vermutet wurde, dass all diese Persönlichkeitsmerkmale miteinander kombiniert in dynamisch erlerntem Vertrauen in autonomes Fahren resultieren (Kraus et al., 2020).

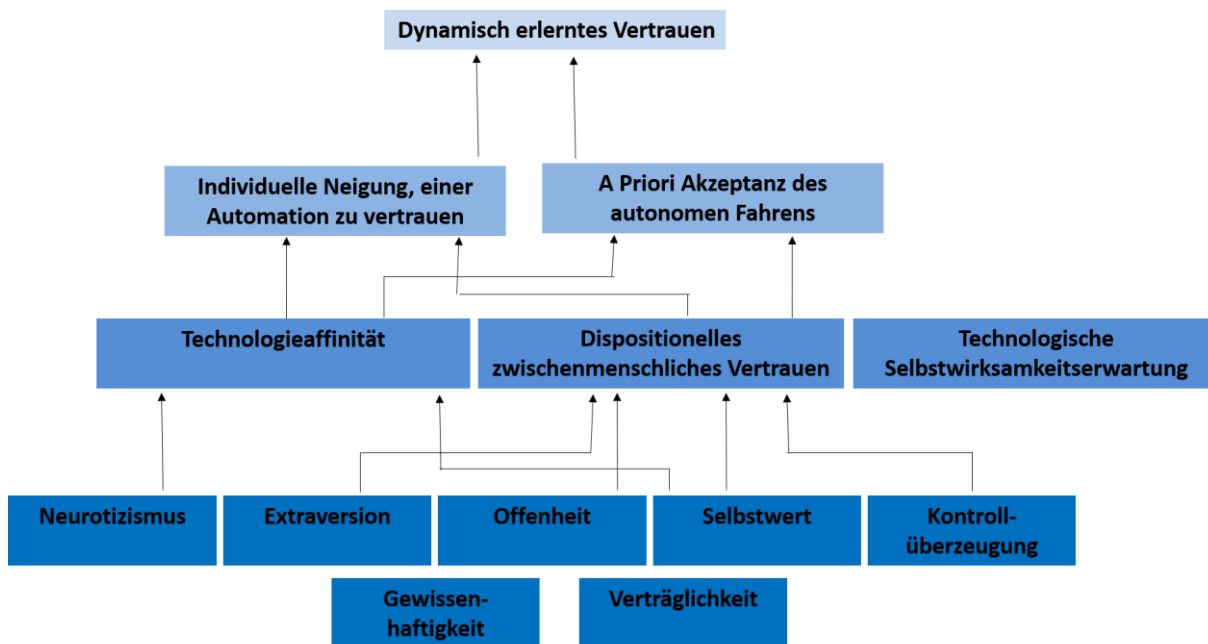


Abb. 3.4: Klassifikation und Relation der Persönlichkeitsmerkmale (in Anlehnung an Kraus et al., 2020, S.18).

Orientierend an den verfügbaren Studien zu Vertrauen und Interaktion mit Technologie (vgl. Kraus et al., 2020) konstruierten die Forscher/innen zehn Modellhypothesen zu den Beziehungen der Persönlichkeitsmerkmale. Tabelle 3.3 zeigt diese Modellhypothesen. Gleichzeitig erfasst Tabelle 3.3 die Ergebnisse der zwei Studien. Die Aufteilung in einen negativen und einen positiven Zusammenhang konkretisiert die Beziehungen der Merkmale aus Abbildung 3.4. Beispielsweise verweist ein negativer Zusammenhang auf Misstrauen im Umgang mit autonomen Agenten. Neurotische Menschen verspüren eine erhöhte Unsicherheit bei der Interaktion mit neuen Technologien (Kraus et al., 2020; McCrae & Costa, 1987). Die Erklärung liefern die Forschungsergebnisse in Abbildung 3.4. Abbildung 3.4 zeigt, dass Neurotizismus die Technologieaffinität beeinflusst und dies die Neigung, einer Automation zu vertrauen, sowie die a priori Akzeptanz autonomen Fahrens prägt. Die a priori Akzeptanz einer Technologie definiert die Bewertung dieser vor einer Interaktion mit dieser (Merritt et al., 2013; Payre et al., 2014). Ein positiver Zusammenhang suggeriert Vertrauen. Beispielsweise wird angenommen, dass Menschen, die mit vielen anderen Menschen harmonieren (verträglich sind) analog mit autonomen Agenten harmonieren. Verträglichkeit erfordert aber weitere Untersuchungen, da kein spezifischer indirekter Effekt auf das Vertrauen in autonomes Fahren nachgewiesen wurde (Kraus et al., 2020). Deshalb ist aus Abbildung 3.4 keine Verknüpfung ersichtlich. Den Glauben eines Menschen, ein selbstbestimmtes Leben führen zu können, die Kontrollüberzeugung (LOC), setzen Kraus et al. (2020) zwar in Beziehung zum Vertrauen in autonome Agenten, doch blieb die Beziehungsrichtung des ganzheitlichen LOC anfangs offen. Kraus et al. (2020) schlossen interne LOC letztendlich aus dem Modell aus, stellten aber einen negativen Zusammenhang von externen LOC zu Vertrauen in autonomes Fahren fest. Menschen mit externem LOC neigen zur Kontrollabgabe. Sie glauben, eine potenziell gefährliche Situation nicht selbst kontrollieren zu können (Ajzen, 2002; Kraus et al., 2020). Einerseits besteht das Risiko, dass sie autonomen Agenten mehr als sich selbst vertrauen. Dagegen neigen Menschen mit internem

LOC zu weniger Vertrauen in autonome Agenten und halten sich selbst für fähig genug, ungewohnte Situationen zu meistern. Andererseits besteht die Möglichkeit, dass Interne von der Effektivität autonomer Agenten eher profitieren als Externe, die sich aufgrund ihres niedrigen Selbstwerts nicht zutrauen, mit Technologie zu interagieren (Ajzen, 2002; Kraus et al., 2020). Die technologische Selbstwirksamkeitserwartung bedarf ebenfalls weiterer Forschung. Anfangs wurde dieses Persönlichkeitsmerkmal in einen positiven Zusammenhang mit Vertrauen in autonomes Fahren gebracht. Kraus et al. (2020) entdeckten in der ersten explorativen Studie, dass die Erwartung, Technik selbst beeinflussen zu können, mit den anderen situativen Persönlichkeitsmerkmalen korreliert und schlossen die technologische Selbstwirksamkeitserwartung aus. Dementsprechend sind keine Verknüpfungen in Abbildung 3.4 vorhanden.

Tab. 3.3: Zehn Modellhypothesen von Kraus et al. (2020) zur Relation der Persönlichkeitsmerkmale und Vertrauen in autonomes Fahren.

Persönlichkeitsmerkmal	Ebene	Positiver Zusammenhang	Negativer Zusammenhang	Offener Zusammenhang oder weitere Forschung erforderlich
Extraversion	Elementares Merkmal	x		
Neurotizismus	Elementares Merkmal		x	
Verträglichkeit	Elementares Merkmal	x		x
Selbstwert	Elementares Merkmal	x		
Kontrollüberzeugung (extern)	Elementares Merkmal		x	x
Dispositionelles zwischenmenschliches Vertrauen	Situatives Merkmal	x		
Technologieaffinität	Situatives Merkmal	x		
Technologische Selbstwirksamkeitserwartung	Situatives Merkmal	x		x
Neigung Automation zu vertrauen	Oberflächenmerkmal	x		
A Priori Akzeptanz autonomen Fahrens	Oberflächenmerkmal	x		

Die meisten Variablen des vorgestellten Modells unterstützen empirisch eine Prognose des dynamisch erlernten Vertrauens in autonome Agenten (Kraus et al., 2020). Demnach kann z.B. bei der Auswahl von Bewerbenden für berufliche Stellen im Umgang mit autonomen

Agenten anhand der Persönlichkeitsmerkmale in geeigneten Testverfahren prognostiziert werden, inwiefern sich die Bewerbenden für die Stelle eignen. In der Luft- und Raumfahrt erfolgt die Auswahl von Bewerbenden mittels psychologischer Eignungsuntersuchung. Dies soll eine Prognose ermöglichen, ob Bewerbende die Voraussetzungen sowohl für die Ausbildung als auch den fordernden Pilotenberuf mitbringen. Intention ist die Minimierung von Ausbildungsabbrüchen vor dem Hintergrund, die Personalplanung der Deutschen Flugsicherung GmbH (DFS) zu stabilisieren und erhöhte Ausbildungskosten zu vermeiden (Eschen, 2013).

3.3.3 Performance

Erwartungen von Industrie und Wirtschaft

Zunehmend treffen Algorithmen und autonome Agenten die Entscheidungen im Alltag, nicht mehr der Mensch (De Visser et al., 2018). Sie entscheiden, welche Werbung online zu sehen ist und steuern Flugzeuge durch gefährliche Gewitter. Im HAT funktionieren der Mensch und der autonome Agent in eigenständigen Rollen (vgl. Kapitel 1.2.1). Im Folgenden wird daran erinnert, wie der autonome Agent die Teamleistung im HAT unterstützt. Kapitel 3.2 thematisiert die Stärken und Schwächen im HAT ausführlich. Darauf aufbauend wird vorgestellt, wie sich die Arbeit im HAT auf die Wahrnehmung der Arbeitsbelastung in Industrie und Wirtschaft auswirkt. Abschließend wird am Beispiel des E-Commerce erklärt, warum autonome Agenten den Menschen nicht vollständig ersetzen können (vgl. Kapitel 5). Diese Zusammenhänge bilden die Verständnisgrundlage für die Einteilung der Autonomie in zwei unterschiedliche Systeme und die Human-out-of-the-loop-Performanceproblematik.

Autonome Fähigkeiten werden für verschiedene Systeme der Industrie entwickelt, um den Arbeitsumfang zu minimieren, menschliche Fähigkeiten zu erweitern und die menschliche Sicherheit zu optimieren (Endsley, 2017). Das langfristige Forschungsziel ist die Aufgabenerfüllung durch autonome Agenten ohne menschliche Anleitung oder Unterstützung. Dennoch kann der autonome Agent aufgrund seiner nicht ausgereiften Autonomiebefähigung noch nicht unabhängig von Menschen agieren (Schaefer et al., 2017). Über die konkrete Zusammenarbeit von Menschen und autonomen Agenten ist aufgrund der früher begrenzten Fähigkeiten der Technik empirisch wenig bekannt (McNeese et al., 2021). Menschen sind in der übergeordneten Rolle der Kontroll- und Aufsichtsinstanz für die Leistung des autonomen Agenten verantwortlich oder kooperieren mit dem autonomen Agenten in der Rolle des Teammitglieds (Endsley, 2017). Die Arbeit mit autonomen Agenten beeinflusst die Wahrnehmung der Arbeitsbelastung. Die Automatisierung sich wiederholender Aufgaben reduziert die manuelle Arbeitsbelastung. Allerdings steigert der Einsatz eines autonomen Agenten die kognitive Arbeitsbelastung, wenn die Interaktion eine übermäßige Menge an erforderlichen Aufgaben umfasst (Endsley, 2017). Ebenso nehmen die menschlichen Teammitglieder im HAT eine höhere Arbeitsbelastung wahr, wenn das Vertrauensverhältnis innerhalb des Teams nicht stimmt und das Team seine Ziele nicht erreicht. Umgekehrt identifizieren die menschlichen Teammitglieder eine Entlastung, wenn das Vertrauen zueinander angemessen kalibriert werden konnte. Eine niedrige Arbeitsbelastung führt zur Fähigkeit, mehrere Aufgaben zu erfüllen. Das ist gleichbedeutend mit erhöhter Leistungsfähigkeit (McNeese et al., 2021; Zhang & Yang, 2017).

Die absehbaren Chancen, die die Zusammenarbeit im HAT der Industrie und Wirtschaft bietet, bedeuten für einzelne Arbeitnehmenden gleichzeitig eine diffuse Unsicherheit (Abbass et al., 2016). Die Unsicherheit geht von der Fragestellung aus, ob autonome Agenten den Menschen aus Industrie und Wirtschaft verdrängen können. McNeese et al. (2021) vermuten, dass beispielsweise im E-Commerce die Anzahl autonomer Agenten steigen wird. Hintergrund ist die

Annahme, dass die Arbeitsumgebungen mit dem Ausbau der Digitalisierung virtueller werden. Aufgaben in virtuellem Umfeld, die typischerweise Menschen erfüllen, können gleichermaßen von autonomen Agenten übernommen werden (McNeese et al., 2021). Vollständig können autonome Agenten den Menschen allerdings nicht ersetzen (Demir et al., 2017). Letztendlich gilt der Mensch in der allgemeinen Wahrnehmung aufgrund seiner Kreativität, Flexibilität und Anpassungsfähigkeit als geeigneter, sich auf unvorhergesehene oder veränderte Situationen einzustellen (Eschen, 2013).

Human-in-the-Loop und Human-on-the-Loop

Die Klassifizierung der Autonomie erfolgt in Human-in-the-Loop (HITL-Systeme) und Human-on-the-Loop (HOTL-Systeme). Als HITL-Systeme werden Maschinen bezeichnet, die mit der Ausführung einer Aufgabe in einem bestimmten Zeitintervall betraut sind und anschließend anhalten, um auf weitere Anweisungen zu warten. Das System handelt, wie es der Mensch befiehlt. Unter der Bedingung, dass Menschen ihnen die selbstständige Aufgabenausführung erlauben, können sie autonom agieren (Nahavandi, 2017). Das Raytheon Patriot-Raketenabwehrsystem ist z.B. ein HITL-System. Der Mensch legt das Ziel fest, das Patriot-System visiert es an. Es kann allerdings erst nach Bestätigung durch den Menschen angreifen. Der Mensch ist somit für eventuelles Fehlverhalten des autonomen Agenten verantwortlich. Als HOTL-Systeme sind Maschinen bekannt, die eine Aufgabe selbstständig ausführen und vollständig abschließen, aber menschliche Mitarbeitende die Leistung kontrollieren und bei Bedarf eingreifen. HITL-Systeme werden zunehmend durch autonomere HOTL-Systeme ersetzt (Nahavandi, 2017).

Parasuraman und Riley (1997) registrierten, dass selbst in unbemannten autonomen Systemen Menschen die autonomen Agenten überwachen. Beispielsweise kontrollieren Menschen Raumfahrzeuge oder Waffensysteme. Unbemannte Drohnen können z.B. per Funk von einem/einer Drohnenpilot/in gesteuert werden (Budde, 2019). Ebenso kann das Funksignal extern unterbrochen werden. Als Drohnen 2018 den Flugverkehr am Londoner Flughafen Gatwick während einer gezielten Sabotageaktion stoppten, konnte das Funksignal der Drohnen gestört werden, um den Flugbetrieb wieder aufzunehmen (Tagesschau, 2018). Durch diese Funkverbindung ist der Unterschied zu bemannten Kampffjets aus technologischer Perspektive gering. Dagegen entscheiden Drohnen, deren Steuerung automatisiert ist, selbstständig. Menschen erlauben insbesondere einer großen Anzahl an Drohnen autonomes Agieren. Denn ein Mensch ist allein nicht fähig, hunderte Drohnen zu kontrollieren. Gleichermaßen können hundert Menschen nicht hundert Drohnen kontrollieren (Budde, 2019). Bei der Eröffnung der Olympischen Winterspiele 2018 in Südkorea ließen hunderte autonome Drohnen im Formationsflug leuchtende Bilder am Nachthimmel erscheinen (Budde, 2019; Frankfurter Allgemeine Zeitung, 2018).

Human-out-of-the-Loop

Entgegen der Erwartungen von Industrie und Wirtschaft hat die Kooperation von Menschen und autonomen Agenten nicht grundsätzlich eine Verbesserung der Leistung zur Folge. Das Human-out-of-the-Loop-Problem ist eine Konsequenz der Automatisierung, die mit dem Verlust der manuellen Fähigkeiten zur Aufgabenbearbeitung und dem Verlust des menschlichen Situationsbewusstseins (Situation Awareness, SA) zusammenhängt (Kaber & Endsley, 1997). Das Risiko der Out-of-the-loop-Problematik besteht, wenn Menschen Rollen als Supervisors und im Monitoring einnehmen (Endsley & Kiris, 1995; Kaber & Endsley, 1997). Die Möglichkeit, dass ein Mensch aus dem geschlossenen Kreis (loop) herausfällt, wird im Folgenden erläutert.

Abbildung 3.5 visualisiert zentrale Begriffe der Erläuterung. Anschließend werden bekannte Maßnahmen, die dieses Performanceproblem optimieren, identifiziert.

Verlust der manuellen Fähigkeiten zur Aufgabenbearbeitung

Die meisten Automationen funktionieren in Situationen, für die sie konzipiert wurden, aber benötigen menschliche Intervention, um in andersartigen Situationen zu funktionieren. Andersartige Situationen entstehen, wenn die Automation nicht auf ihre Programmierung zurückgreifen kann (Endsley, 2017). Forschungen verglichen die Leistung von Mitarbeitenden, die eine automatisierte Anlage überwachten (supervisory control), mit der von Mitarbeitenden, die es gewohnt waren, die Anlage manuell zu steuern. Festgestellt wurde, dass die überwachenden Mitarbeitenden die Anlage bei einem Wegfall der Automatisierung langsamer und ineffizienter steuern konnten als die anderen Mitarbeitenden (Endsley & Kiris, 1995). Mitarbeitende, die regelmäßig passive Kontrollaufgaben ausführen, gewöhnen sich daran und verlieren die manuelle Steuerung. Ferner gilt die Befürchtung, dass die Entwicklung benötigter Fähigkeiten begrenzt wird, wenn Mitarbeitende direkt mit der Automation ausgebildet werden (Kaber & Endsley, 1997).

Manuelle Fähigkeiten sind relevant, um Aufgaben selbstständig zu bearbeiten (Endsley, 2017; Kaber & Endsley, 1997). Um autonom zu bleiben, sollte sich der Mensch eine gewisse Unabhängigkeit von autonomen Agenten bewahren. Zusätzlich ermöglichen manuelle Fähigkeiten, dass Mitarbeitende erkennen, wann die Automation ein Eingreifen erfordert (Endsley, 2017; Kaber & Endsley, 1997). Die Notwendigkeit der manuellen Intervention ist grundsätzlich gegeben. Denn gewöhnliche Betriebsbedingungen umfassen Fehlfunktionen oder Ausfälle der Automation (Endsley, 2017). Ebenso ist ein autonomer Agent nicht vollkommen (Lee & See, 2004). Je weitläufigere Aufgaben der Automation zugeteilt werden und je zuverlässiger die Automation agiert, desto geringer ist die Wahrscheinlichkeit, dass Menschen Situationen erkennen, in denen die Automation ihnen Fehlermeldungen sendet (Endsley, 2017). Beispielsweise stürzte im Jahr 1989 die Boeing 747 von Air China beinahe in den Pazifischen Ozean, weil der Autopilot einen Triebwerksausfall kompensierte, den der/die Pilot/in erst nach Ausfall des Autopiloten erkannte (Kaber & Endsley, 1997).

Verlust des menschlichen Situationsbewusstseins

Ein angemessenes Situationsbewusstsein bedeutet, dass sich der Mensch jederzeit bewusst ist, was gerade passiert, warum es passiert und wie es weitergehen wird (Badke-Schaub et al., 2012). Endsley und Kiris (1995) betrachten das menschliche Situationsbewusstsein als entscheidendes Maß für effektive Entscheidungsfindung und menschliche Leistung in dynamischen Systemen (vgl. Kapitel 2.3.2). Es umfasst aufeinander aufbauend die Wahrnehmung relevanter Zustandsvariablen (Level 1 SA), das Verständnis der Relevanz dieses Zustands (Level 2 SA) und die Prognose zukünftiger Trends und Konsequenzen (Level 3 SA). Alle drei Komponenten sind voneinander abhängig.

Die Symptomatik eines Verlusts an SA oder der Reduktion von SA äußert sich in verzögerter Problemerkennung, Missinterpretation von Systemmeldungen und reduzierter Anpassungsfähigkeit (Endsley & Kiris, 1995). Ein/e Autofahrer/in nimmt beispielsweise die Geschwindigkeitsanzeige des Fahrzeugs wahr, reagiert allerdings nicht auf geschwindigkeitsbegrenzende Verkehrsschilder und wird von der Polizei angehalten. Er/Sie erkennt zwar das Problem, passt sich allerdings nicht an die veränderte Situation an. In einem anderen Beispiel interpretiert der/die Autofahrer/in auffällige Motorgeräusche als unerheblich, woraufhin sich der Motorschaden unentdeckt vergrößert (Badke-Schaub et al., 2012). Wieder erkennt er/sie ein Problem,

doch kann es nicht korrekt beurteilen. Mängel des SA fallen erst auf, wenn die Automation nicht mehr funktioniert (Badke-Schaub et al., 2012; Endsley, 2017). Durch die passive Kontrollaufgabe des Menschen im Umgang mit der Automation entstehen zwei Effekte: Complacency und ein Wachsamkeitsdefizit. Der/Die Mitarbeitende fühlt sich zur Kontrollinstanz befördert und ein Gefühl der selbstgefälligen oder unkritischen Zufriedenheit mit sich selbst oder seinen Leistungen verbreitet sich. Ein weiterer Aspekt, der Complacency verursacht, ist eine zuverlässige Automation (Kaber & Endsley, 1997). Die Zuverlässigkeit einer Automation kann zu übersteigertem Vertrauen führen und das Erkennen von Fehlern vermindern (Badke-Schaub et al., 2012; Endsley, 2017). Funktioniert eine Automation nicht zuverlässig, entwickelt sich Misstrauen. In Kombination mit dem Wachsamkeitsdefizit entstehen gefährliche Irrtümer. Beispielsweise ignorierte und deaktivierte die Flugbesatzung vor einigen Jahrzehnten ihre Automation, obwohl diese Signale sendete. Sie waren die Unzuverlässigkeit der Automation an Bord gewohnt und überzeugt, es handle sich um Fehlalarmmeldungen. Die wahren Alarmmeldungen wurden aufgrund einer hohen Fehlalarmquote und dem einhergehenden Misstrauen nicht mehr ernst genommen. Weiterhin wechselt der/die Mitarbeitende bei dem Verlust von SA von der Rolle des/der aktiven Informationsverarbeiters/in in die Rolle des/der passiven Informationsempfängers/in. Die Art der Systemrückmeldung an den Menschen ändert sich. Ohne entsprechende Rückmeldungen sind die Mitarbeitenden „out of the loop“, weil ihnen Informationen über korrekte Aufgabenbearbeitung oder aufgetretene Probleme fehlen (Endsley & Kiris, 1995). Beispielsweise fehlen dem/der Autofahrer/in, der/die die Geschwindigkeit mittels eines Tempomats reguliert, alle taktilen Rückmeldungen vom Andruck des Gaspedals. Dieser Mangel kann das Erkennen und Beurteilen relevanter dynamischer Veränderungen des Systems oder der Situation verhindern. Der Verlust des SA kann ebenfalls von der Automation begünstigt werden. Mangelnde Transparenz der Automation blockiert die Bemühungen der Mitarbeitenden, wachsam zu sein und das SA nicht zu verlieren. Zum Beispiel zeigt das Navigationssystem eines Pkws nicht alle Baustellen und Umleitungen an, die für das Nachvollziehen und Beurteilen der Route notwendig sind (Badke-Schaub et al., 2012). Insbesondere im Umgang mit autonomen Agenten wirkt sich mangelndes Systemverständnis auf das SA aus. Die Funktionsweise eines autonomen Agenten ist komplex und technisch, für den Laien nahezu unverständlich. Der Ausruf „Was macht er denn jetzt?“ ist aktuell bereits im Umgang mit Computerprogrammen wie Excel zu hören.

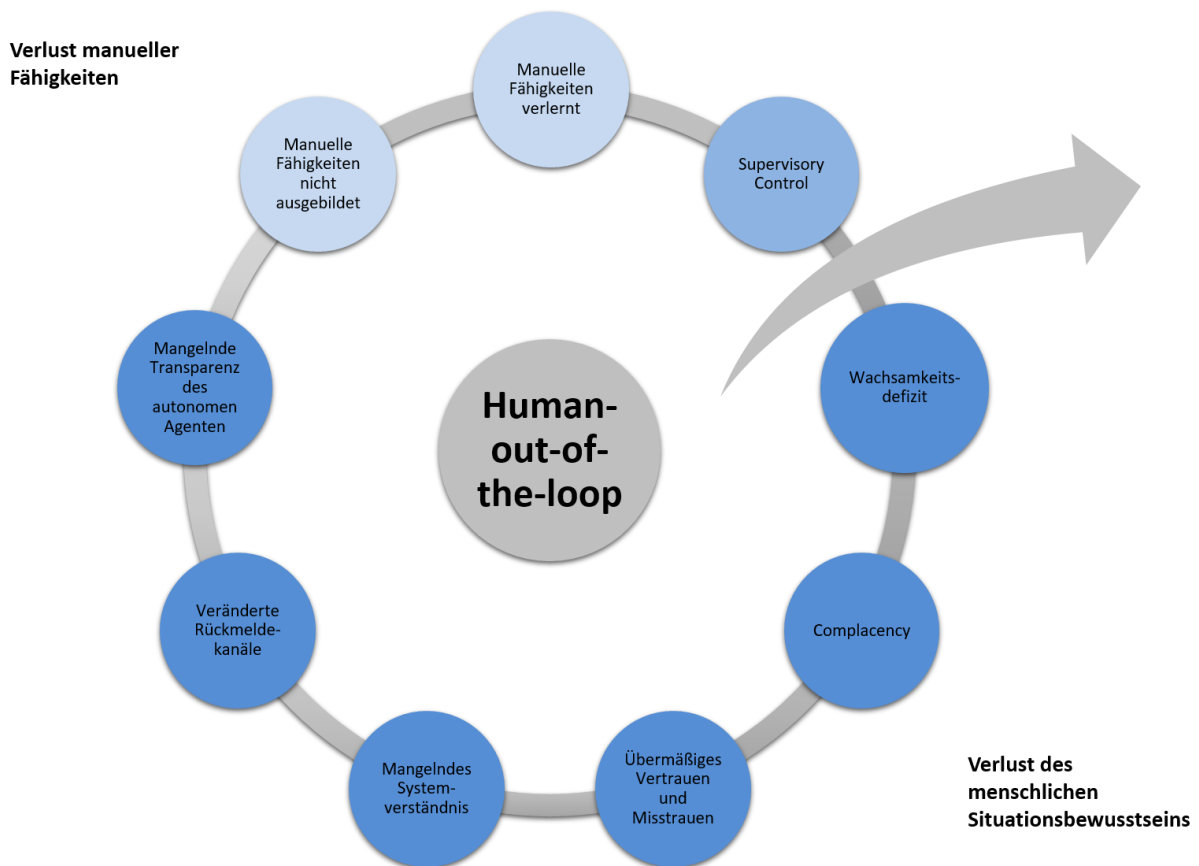


Abb. 3.5: Human-out-of-the-loop: Effekte, die auf den Menschen einwirken.

Lösungsmaßnahmen

Um den Menschen im Loop zu halten, sind das Ausbilden und Festigen manueller Fähigkeiten zur Aufgabenbearbeitung sowie das Bewahren des SA notwendig (Badke-Schaub et al., 2012; Endsley & Kiris, 1995). Entsprechende Maßnahmen umfassen:

- Das Training

Nach Badke-Schaub et al. (2012) kann das SA aufrechterhalten werden, wenn das Design des autonomen Agenten an die Bedürfnisse der Mitarbeitenden angepasst wird oder die Mitarbeitenden im Rahmen von Schulungen und Workshops lernen, wie der autonome Agent funktioniert und wie mit ihm zu kommunizieren ist. Bei mangelnder Transparenz des autonomen Agenten oder aufgrund eines mangelnden Systemverständnisses eignen sich diese Trainingsmethoden, um einen Verlust des SA zu verringern. Im Rahmen der adaptiven Automation können durch Simulationen die manuellen Fähigkeiten erhalten werden. Beispielsweise sind Trainings an Flugsimulatoren für Piloten/innen obligatorisch (Badke-Schaub et al., 2012; Kaber & Endsley, 1997).

- Den Automatisierungsgrad

Einen Ansatz zur Optimierung von Out-of-the-Loop-Problemen stellt der Automatisierungsgrad (Level of Automation, LOA) dar (vgl. Kapitel 1.3). Der Automatisierungsgrad bestimmt die optimale Verteilung der Kontrolle zwischen menschlichen Mitarbeitenden und der Automation, um beide entsprechend ihrer Qualifikationen und Kapazitäten variabel in den Systembetrieb einzubinden. Menschliche kognitive Fehler sollen reduziert werden (Kaber & Endsley, 1997).

- Die adaptive Automation

Parasuraman (1992) identifizierte Verbesserungen beim Erkennen einer unzuverlässigen Automation und Fehlfunktionen, indem der/die Mitarbeitende gemäß der adaptiven Automation in regelmäßigen Abständen die manuelle Steuerung übernahm (vgl. Kapitel 2.2.2). In der statischen Automation übernimmt der/die Mitarbeitende lediglich die Überwachung, nicht die Steuerung oder andere manuelle Aufgaben. Die Verlagerung kognitiver Funktionen vom Menschen auf die Maschine bewirkt einen starken Abfall der Leistung des Menschen (Badke-Schaub et al., 2012; Parasuraman, 1992). Der rapide Leistungsabfall des/der Mitarbeitenden soll verhindert werden (Abbass et al., 2016; Parasuraman, 1992).

- Das HAT

Ein jüngerer Lösungsansatz, dem auch dieses Buch gewidmet ist, fokussiert das HAT. Demir et al. (2017) schlagen vor, die Beziehung zwischen den Mitarbeitenden und der Automation fundamental zu ändern. Die Beziehung soll sich von der rollenspezifischen Unterteilung in Mensch und Maschine differenzieren hin zu einem Teaming, in dem Mensch und autonomer Agent aktiv kooperieren und transparent koordinieren (Demir et al., 2017).

3.4 Ein Überblick

Dieses Unterkapitel bietet ein Konzentrat des Inhalts aus Kapitel 3.1 bis 3.3. Die Teamarbeit in konventionellen rein menschlichen Teams wurde von der Teamarbeit im HAT abgegrenzt, bevor Stärken und Schwächen des HAT identifiziert werden konnten. Im Anschluss wurden Variablen, die die grundsätzliche Interaktion mit autonomen Agenten aus der Perspektive menschlicher Teammitglieder im HAT prägen, beleuchtet.

Zusammenfassung
<ul style="list-style-type: none">• Traditionelle Teamarbeit ist definiert als zwei oder mehrere Menschen, die abhängig voneinander auf ein gemeinsames Ziel hinarbeiten.• Das HAT umfasst mindestens einen Menschen und einen autonomen Agenten, die voneinander abhängig zusammenarbeiten und sich aufeinander abstimmen, um Aufgaben erfolgreich zu schaffen. Im HAT sind menschliche Teammitglieder und autonome Agenten in gewissem Maße gleichwertig.• Im Vergleich zu rein menschlichen Teams, die vier Anforderungen erfüllen müssen, muss ein autonomer Agent zehn Herausforderungen genügen, um als Teamplayer angesehen zu werden.• Im HAT kontrolliert der Mensch den autonomen Agenten nicht, sondern arbeitet mit ihm im Team zusammen.• Um auf die Absicht eines Teammitglieds schließen zu können, muss diese mit den anderen Teammitgliedern kommuniziert werden. In diesem Zusammenhang unterscheidet man zwischen expliziter und impliziter Kommunikation.• Stärken und Schwächen des HAT sind ein Resultat von den Eigenschaften und Voraussetzungen, die der autonome Agent sowie der Mensch besitzen.• In Zukunft braucht es Forschung in allen Disziplinen, um die soziologischen, wirtschaftlichen und psychologischen Auswirkungen besser zu erfassen und verstehen zu können.• In der Literatur besteht Uneinigkeit, wie Vertrauen im Mensch-Autonomie-Kontext entsteht und zu gestalten ist. Bekannte Ansätze zur Definition von Vertrauen

charakterisieren Vertrauen in autonome Agenten zunächst als Erwartung von Verhaltensweisen und Ergebnissen. Muir (1987) verglich die chronologische Entwicklung von zwischenmenschlichem Vertrauen mit der von Vertrauen in Maschinen. Lee und Moray (1992) sehen eine Abhängigkeit von Vertrauen in der Leistung, dem Prozess und dem Zweck einer Automation. Hoff und Bashir (2015) klassifizieren Vertrauen in autonome Agenten dagegen in die drei Schichten Dispositionsvertrauen, Situationsvertrauen und erlerntes Vertrauen. Mayer et al. (1995) definieren Vertrauen, das der Mensch einer Maschine entgegenbringt, als Absicht oder Handlungsbereitschaft. Lee und See (2004) bestätigen, dass die Definition von Mayer et al. (1995) im Mensch-Autonomie-Kontext die populärste und akzeptierteste ist.

- Vertrauen entwickelt sich als Prozess. Es ist dynamisch und instabil.
- Autonomie basiert grundlegend auf Vertrauen, steht aber in einem natürlichen Widerspruch zu Koordination und übergeordneten Konzepten.
- Übermäßiges Vertrauen und Misstrauen beschreiben zwei Dilemmata der Interaktion mit autonomen Agenten.
- Trusted Autonomy (TA) ist ein Forschungsgebiet, das den Interaktionsraum zweier Entitäten, die ein gewisses Maß an Autonomie aufweisen, untersucht und hinsichtlich der Komponente Vertrauenswürdigkeit des autonomen Agenten (Trusted Autonomous System, TAS) gestaltet.
- Persönlichkeit gilt als Prädiktor für Vertrauen in autonome Agenten.
- Dynamisch erlerntes Vertrauen in autonome Agenten kann mittels einer Vielzahl von Persönlichkeitsmerkmalen prognostiziert werden.
- Autonome Agenten können unterteilt werden in Human-in-the-Loop (HITL-Systeme) und Human-on-the-Loop (HOTL-Systeme). HOTL-Systeme ersetzen zunehmend HITL-Systeme.
- Das Human-out-of-the-Loop-Problem ist eine Konsequenz der Automatisierung, die mit dem Verlust der manuellen Fähigkeiten zur Aufgabenbearbeitung und dem Verlust des menschlichen Situationsbewusstseins (Situation Awareness) zusammenhängt.
- Das HAT beinhaltet das höchste Potential, um der Out-of-the-loop-Problematik zu entgehen.

Literaturverzeichnis

- Abbass, H. A. (2019). Social integration of artificial intelligence: functions, automation allocation logic and human-autonomy trust. *Cognitive Computation*, 11(2), 159–171.
- Abbass, H. A., Leu, G. & Merrick, K. (2016). A review of theoretical and practical challenges of trusted autonomy in big data. *IEEE Access*, 4, 2808–2830.
- Abbass, H. A., Petraki, E., Merrick, K., Harvey, J. & Barlow, M. (2016). Trusted autonomy and cognitive cyber symbiosis: Open challenges. *Cognitive Computation*, 8(3), 385–408.
- Abbass, H. A., Scholz, J. & Reid, D. J. (2018). Foundations of Trusted Autonomy: An introduction. In H. Abbass, J. Scholz & D. Reid (Eds.), *Foundations of Trusted Autonomy. Studies in Systems, Decision and Control 117*, 1–12. Cham: Springer International Publishing.
- Ajzen, I. (2002). Perceived behavioral control, self-efficacy, locus of control, and the theory of planned behavior. *Journal of Applied Social Psychology*, 32(4), 665–683.
- Azhar, M. Q., & Sklar, E. I. (2017). A study measuring the impact of shared decision making in a human-robot team. *The International Journal of Robotics Research*, 36(5-7), 461–482.
- Badke-Schaub, P., Hofinger, G. & Lauche, K. (2012). *Human Factors - Psychologie sicheren Handelns in Risikobranchen* (2. überarbeitete Aufl.). Berlin: Springer Verlag.
- Barber, B. (1983). *The logic and limits of trust*. New Brunswick, NJ: Rutgers University Press.
- Blustein, D. L., Kenny, M. E., Di Fabio, A. & Guichard, J. (2019). Expanding the impact of the psychology of working: Engaging psychology in the struggle for decent work and human rights. *Journal of Career Assessment*, 27(1), 3–28.
- Botsman, R. (2020). Introducing rethink with Rachel. *LinkedIn*. Zugriff am 08.01.2021. Verfügbar unter <https://www.linkedin.com/pulse/introducing-rethink-rachel-rachel-botsman/>
- Budde, Joachim (2019). Wenn Waffen selbst über Leben und Tod entscheiden. *Deutschlandfunk*. Zugriff am 28.01.2021. Verfügbar unter https://www.deutschlandfunk.de/autonome-waffensysteme-wenn-waffen-selbst-ueber-leben-und.676.de.html?dram:article_id=441194
- Brand, Y., Schmitt, F. & Schulte, A. (2018). Delegation und Assistenz mit kognitiver Automation für Manned-Unmanned Teaming Hubschrauber Missionen. *Kognitive Systeme*, 2018(1).
- Chen, J. Y. C. & Barnes, M. J. (2013). *Human- Agent Teaming for multi- robot control: A literature review* (Report No. ARL-TR-6328). Army Research Laboratory.
- Chen, J. Y. C., Barnes, M. J., Selkowitz, A. R., Stowers, K., Lakhmani, S. G. & Kasdaglis, N. (2016). Human-autonomy teaming and agent transparency. In *Companion Publication of the 21st International Conference on Intelligent User Interfaces* (S. 28–31). Sonoma, CA, USA: IUI'16 Companion.
- Cohen, P. R. & Levesque, H. J. (1991). Teamwork. *Nous*, 25(4), 487–512.

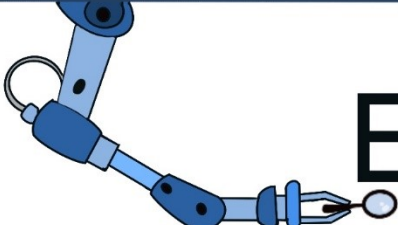
- Collins, M. G., Juvina, I. & Gluck, K. A. (2016). Cognitive model of trust dynamics predicts human behavior within and between two games of strategic interaction with computerized confederate agents. *Frontiers in Psychology*, 7, 49. Zugriff am 04.01.2021. Verfügbar unter <https://doi.org/10.3389/fpsyg.2016.00049>
- Cooke, N. J., Gorman, J. C., Duran, J. L. & Taylor, A. R. (2007). Team cognition in experienced command-and-control teams. *Journal of Experimental Psychology: Applied*, 13(3), 146–157.
- Cooke, N. J., Gorman, J. C., Myers, C. W. & Duran, J. L. (2012). Interactive team cognition. *Cognitive Science*, 37(2), 255–285.
- Cuevas, H. M., Fiore, S. M., Caldwell, B. S. & Strater, L. (2007). Augmenting team cognition in human-automation teams performing in complex operational environments. *Aviation Space Environmental Med*, 78(5), B63–B70.
- Dauth, W., Findeisen, S., Südekum, J. & Woessner, N. (2017). German robots – the impact of industrial robots on workers, *IAB – Discussion Paper, No.30/2017*. Nürnberg: Institut für Arbeitsmarkt- und Berufsforschung (IAB).
- De Visser, E. J., Pak, R. & Shaw, T. H. (2018). From ‘automation’ to ‘autonomy’: The importance of trust repair in human–machine interaction. *Ergonomics*, 61(10), 1409–1427.
- Demir, M., McNeese, N. J. & Cooke, N. J. (2016). Team situation awareness within the context of human-autonomy teaming. *Cognitive Systems Research*, 46, Part C, 3–12.
- 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), 258–267.
- Demir, M., Likens, A. D., Cooke, N. J., Amazeen, P. G. & McNeese, N. J. (2019). Team coordination and effectiveness in Human-Autonomy Teaming. *IEEE Transactions on Human-Machine Systems*, 49(2), 150–159.
- Demir, M., Likens, A. D., Cooke, N. J. & McNeese, N. J., (2019). Team coordination and effectiveness in Human-Autonomy Teaming. *IEEE Transactions on Human-Machine Systems*, 99, 1–10.
- Dzindolet, M. T., Peterson, S. A., Pomranky, R. A., Pierce, L. G. & Beck, H. P. (2003). The role of trust in automation reliance. *International Journal of Human-Computer Studies*, 58(6), 697–718.
- Eberl, P. (2012). Vertrauen und Kontrolle in Organisationen, In H. Möller (Hrsg.), *Vertrauen in Organisationen* (S. 93–110), Wiesbaden: Springer Fachmedien.
- Endsley, M. R. (2017). From here to autonomy: lessons learned from human–automation research. *Human Factors*, 59(1), 5–27.
- Endsley, M. R., & Kaber, D. B. (1999). Level of automation effects on performance, situation awareness and workload in a dynamic control task. *Ergonomics*, 42(3), 462–492.
- Endsley, M. R. & Kiris, E. O. (1995). The out-of-the-loop performance problem and level of control in automation. *Human Factors*, 37(2), 381–394.
- Eschen, S. C. S. (2013). Persönlichkeit als Prädiktor für Leistung in hochautomatisierten Mensch-Maschine-Teams der Luftfahrt. Veröffentlichte Dissertation, Universität Hamburg.

- Feltovich, P. J., Bradshaw, J. M., Jeffers, R., Suri, N. & Uszok, A. (2004). Social order and adaptability in animal and human cultures as analogues for agent communities: Toward a policy-based approach. In A. Omicini, P. Petta & J. Pitt (Eds.), *Engineering societies in the agents world IV* (pp. 21–48). Berlin: Springer Verlag.
- Frankfurter Allgemeine Zeitung (FAZ) (2018). Über Pyeongchang: Drohnen malen Bilder in den Himmel. *Youtube*. Zugriff am 28.01.2021. Verfügbar unter <https://www.youtube.com/watch?v=sXsSxm2qXZ4>
- Gielniak, M. J. & Thomaz, A. L. (2011). Generating anticipation in robot motion. RO-MAN, 20th *IEEE International Symposium on Robot and Human Interactive Communication* (pp. 449–454). Atlanta, GA.
- Goodman, T. J., Miller, M. E., Rusnock, C. F. & Bindewald, J. M. (2017). Effects of agent timing on the human-agent team. *Cognitive Systems Research*, 46, 40–51.
- Hayes, B. & Scassellati, B. (2013). Challenges in shared-environment human-robot collaboration. 8. *ACM/IEEE International Conference on Human- Robot Interaction, Workshop on Collaborative Manipulation*. Tokyo, Japan.
- Ho, C.-Y., Nikolic, M. I., Waters, M. J. & Sarter, N. B. (2004). Not now: Supporting attention management by indicating the modality and urgency of pending tasks. *Human Factors*, 46(3), 399–409.
- Hoc, J. M. (2000). From human-machine interaction to human-machine cooperation. *Ergonomics*, 43(7), 833–843.
- Hoff, K. A. & Bashir, M. (2015). Trust in automation: integrating empirical evidence on factors that influence trust. *Human Factors*, 57(2), 407–434.
- Huhns, M. N. & Buell, D. A. (2002). Trusted Autonomy. *IEEE Internet Computing*, 6(3), 92–95.
- Kaber, D. B. & Endsley, M. R. (1997). Out-of-the-loop performance problems and the use of intermediate levels of automation for improved control system functioning and safety. *Process Safety Progress*, 16(3), 126–131.
- Klein, G., Woods, D. D., Bradshaw, J. M., Hoffman, R. R. & Feltovich, P. J. (2004). Ten challenges for making automation a "team player" in joint human-agent activity. *IEEE Intelligent Systems*, 19(6), 91–95.
- Krämer, H. (2019). Digitalisierung, Monopolbildung und wirtschaftliche Ungleichheit. *Wirtschaftsdienst*, 99(1), 47–52.
- Kraus, J., Scholz, D. & Baumann, M. (2020). What's driving me? Exploration and validation of a hierarchical personality model for trust in automated driving. *Human Factors*. Advance Online Publication. <https://doi.org/10.1177/0018720820922653>
- Lackner, K. (2012). Vertrauen in Gruppen. In H. Möller (Hrsg.), *Vertrauen in Organisationen* (S. 111–139), Wiesbaden: Springer Fachmedien.
- Langan-Fox, J., Canty, J. M. & Sankey, M. J. (2009). Human–automation teams and adaptable control for future air traffic management. *International Journal of Industrial Ergonomics*, 39(5), 894–903.
- Larsen, R. J. & Buss, D. M. (2005). *Personality psychology: Domains of knowledge about human nature*. New York City: McGraw-Hill Education.

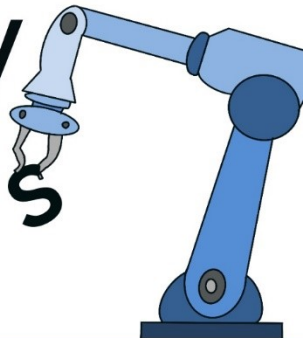
- Lee, J. D. & Moray, N. (1992). Trust, control strategies and allocation of function in human machine systems. *Ergonomics*, 35(10), 1243–1270.
- Lee, J. D. & Moray, N. (1994). Trust, self-confidence, and operators' adaptation to automation. *International Journal of Human-Computer Studies*, 40(1), 153–184.
- Lee, J. D. & See, K. A. (2004). Trust in automation: designing for appropriate reliance. *Human Factors*, 46(1), 50–80.
- Lees, M. N. & Lee, J. D. (2007). The influence of distraction and driving context on driver response to imperfect collision warning systems. *Ergonomics*, 50(8), 1264–1286.
- Lohse, M., van Berkel, N., van Dijk, E. M., Joosse, M. P., Karreman, D. E. & Evers, V. (2013). The influence of approach speed and functional noise on users' perception of a robot. *IEEE/RSJ International Conference on Intelligent Robots and Systems* (pp. 1670– 1675). Tokyo, Japan.
- Mayer, R. C., Davis, J. H. & Schoorman, F. D. (1995). An integrative model of organizational trust. *Academy of Management. Review*, 20(3), 709–734.
- McCrae, R. R. & Costa, P. T. (1987). Validation of the five-factor model of personality across instruments and observers. *Journal of Personality and Social Psychology*, 52(1), 81–90.
- 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), 262–273.
- McNeese, N. J., Demir, M., Chiou E. K. & Cooke, N. J. (2021). Trust and team performance in Human-Autonomy Teaming. *International Journal of Electronic Commerce*, 25(1), 51–72.
- Mercado, J. E., Rupp, M. A., Chen, J. Y. C., Barnes, M. J., Barber, D. & Procci, K. (2016). Intelligent agent transparency in human-agent teaming for Multi-UxV Management. *Human Factors*, 58(3), 401–415.
- Merkel, F. (2019). Teslas „Autopilot“ war bei tödlichem Unfall eingeschaltet. *Spiegel*. Zugriff am 05.01.2020. Verfügbar unter <https://www.spiegel.de/auto/aktuell/tesla-autopilot-war-laut-bericht-bei-todescrash-mit-lkw-eingeschaltet-a-1268087.html>
- Merritt, S. M. (2011). Affective processes in human–automation interactions. *Human Factors*, 53(4), 356–370.
- Merritt, S. M., Heimbaugh, H., LaChapell, J. & Lee, D. (2013). I trust it, but I don't know why: Effects of implicit attitudes toward automation on trust in an automated system. *Human Factors*, 55(3), 520–534.
- Meyer, J. (2001). Effects of warning validity and proximity on responses to warnings. *Human Factors*, 43(4), 563–572.
- Mowen, J. C. (2000). The 3M model of motivation and personality: Theory and empirical applications to consumer behavior. Boston: Kluwer Academic Press.
- Muir, B. M. (1987). Trust between humans and machines, and the design of decision aids. *International Journal of Man-Machine Studies*, 27(5-6), 527–539.
- Nahavandi, S. (2017). Trusted autonomy between humans and robots. Towards human-on-the-loop in robotics and autonomous systems. *IEEE Systems, Man, and Cybernetics Magazine*, 3(1), 10–17.

- O'Neill, T. A., McNeese, N. J., Barron, A. & Schelble, B. (2020). Human–Autonomy Teaming: A review and analysis of the empirical literature. *Human Factors*. Advance Online Publication. <https://doi.org/10.1177/0018720820960865>
- Onken, R. (1997). The cockpit assistant system CASSY as an on-board player in the ATM environment. *Proceedings of first air traffic management research and development seminar*. Saclay, France.
- Parasuraman, R. (1993). Effects of adaptive function allocation on human performance. In D. J. Garland & J. A. Wise (Eds.), *Human factors and advanced aviation technologies* (pp. 147–157). Daytona Beach, FL: Embry-Riddle Aeronautical University Press.
- Parasuraman, R., Mouloua, M. & Molloy, R. (1996). Effects of adaptive task allocation on monitoring of automated systems. *Human Factors*, 38(4), 665–679.
- Parasuraman, R. & Riley, V. (1997). Humans and automation: Use, misuse, disuse, abuse. *Human Factors*, 39(2), 230–253.
- Payre, W., Cestac, J. & Delhomme, P. (2014). Intention to use a fully automated CAR: Attitudes and a priori acceptability. *Transportation Research Part F: Traffic Psychology and Behaviour*, 27, Part B, 252–263.
- Potor, M. (2019). Brötchen, Fahrer oder Autopilot: Wer ist Schuld am Tesla-Unfall 2018? *Mobilitymag*. Zugriff am 05.01.2020. Verfügbar unter <https://mobilitymag.de/tesla-auto-pilot-unfall-kalifornien-2018/>
- Rousseau, D., Sitkin, S., Burt, R. & Camerer, C. (1998). Not so different after all: A cross-discipline view of trust. *Academy of Management Review*, 23(3), 393–404.
- Salas, E., Cooke, N. J. & Rosen, M. A. (2008). On teams, teamwork, and team performance: Discoveries and developments. *Human Factors*, 50(3), 540–547.
- Salas, E., Dickinson, T. L., Converse, S. A. & Tannenbaum, S. I. (1992). Toward an understanding of team performance and training. In R. W. Swezey & E. Salas (Eds.), *Teams: Their training and performance* (pp. 3–29). Norwood, NJ: Ablex Publishing.
- Schaefer, K. W. (2013). *The perception and measurement of human-robot trust*. Veröffentlichte Dissertation, University of Central Florida, Orlando, FL.
- Schaefer, K. E., Evans III., A. W. & Hill, S. G. (2015). Command and control in network-centric operations: Trust and robot autonomy. *Proceedings of the 20th International Command and Control Research and Technology Symposium*. Annapolis, MD.
- Schaefer, K. E., Straub, E. R., Chen, J. Y. C., Putney, J. & Evans III., A. W. (2017). Communicating intent to develop shared situation awareness and engender trust in human-agent teams. *Cognitive Systems Research*, 46, 26–39.
- Schneider, U., Schober, F. & Harrach, B. (2011). Ambient Assisted Living (AAL) - Technologien im betreubaren Wohnen" – Wissenschaftliche Evaluierung des Pilotprojektes "REAAAL" im Hinblick auf sozialpolitische Zielsetzungen (Forschungsberichte des Forschungsinstituts für Altersökonomie, 1/2011). Wien: WU Vienna University of Economics and Business, Forschungsinstitut für Altersökonomie.
- 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).

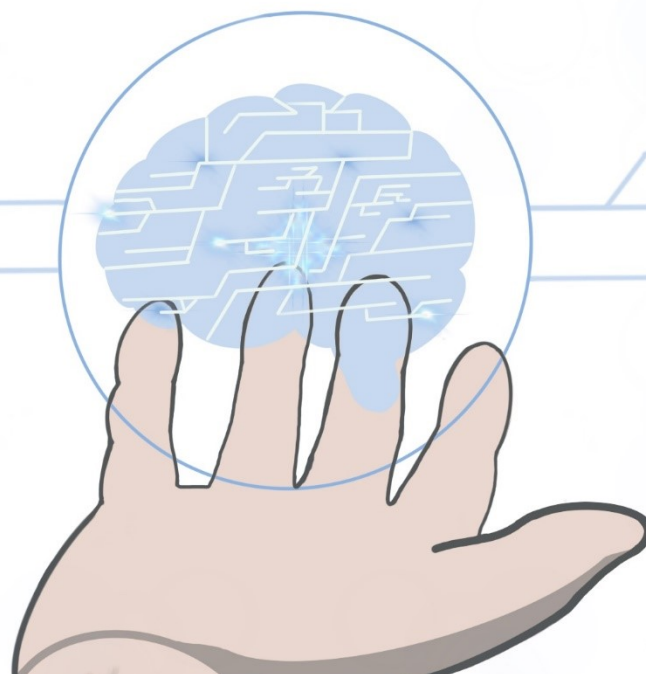
- Machines as teammates: A research agenda on AI in team collaboration. *Information & Management*, 57(2), 1–22.
- Shively, R. J., Lachter, J., Brandt, S. L., Matessa, M., Battiste, V. & Johnson, W. W. (2018). Why Human-Autonomy Teaming? In C. Baldwin (Ed.), *Advances in Neuroergonomics and Cognitive Engineering. AHFE 2017. Advances in Intelligent Systems and Computing* (586). Cham: Springer International Publishing.
- Strabala, K., Lee, M. K., Dragan, A., Forlizzi, J. & Srinivasa, S. S. (2012). Learning the communication of intent prior to physical collaboration. *IEEE RO-MAN: The 21st IEEE international symposium on robot and human interactive communication*, 968–973. <https://doi.org/10.1109/ROMAN.2012.6343875>
- Sun R. (2018). Intrinsic Motivation for Truly Autonomous Agents. In H. Abbass, J. Scholz & D. Reid (Eds.), *Foundations of Trusted Autonomy. Studies in Systems, Decision and Control 117* (pp. 273–292). Cham: Springer International Publishing.
- Szafir, D., Mutlu, B. & Fong, T. (2014). Communication of intent in assistive free flyers. *IEEE international conference on Human-robot interaction, HRI '14*, 358–365.
- Tagesschau (2018). Flugbetrieb in Gatwick lahmgelegt: Drohnen verhindern Starts und Landungen in London. *Tagesschau Mediathek*. Zugriff am 28.01.2021. Verfügbar unter <https://www.tagesschau.de/multimedia/video/video-485719.html>
- Wijngaards, N., Kempen, M., Smit, A. & Nieuwenhuis, K. (2006). Towards sustained team effectiveness. In O. Boissier, J. Padget, V. Dignum, G. Lindemann, E. T. Matson, S. Ossowski, J. Sichman & J. Vázquez-Salceda (Eds.), *Coordination organizations institutions and norms in multi-agent systems* (pp. 35–47). Berlin: Springer Verlag.
- Wiltshire, T. J., Lobato, E. J. C., Wedell, A. V., Huang, W., Axelrod, B. & Fiore, S. M. (2013). Effects of robot gaze and proxemic behavior on perceived social presence during a hallway navigation scenario. *Proceedings of the Human Factors and Ergonomics Society Annual Meeting*, 57(1), 1273–1277.
- Wright, J. L., Chen, J. Y. C. & Barnes, M. J. (2018). Human–automation interaction for multiple robot control: The effect of varying automation assistance and individual differences on operator performance. *Ergonomics*, 61(8), 1033–1045.
- Wright, J. L., Chen, J. Y. C., Barnes, M. J. & Hancock, P. A. (2016). Agent reasoning transparency's effect on operator workload. *Proceedings of the Human Factors and Ergonomics Society Annual Meeting*, 60(1), 249–253.
- Wright, M. C. & Kaber, D. B. (2003). Team coordination and strategies under automation. *Proceedings of the Human Factors and Ergonomics Society Annual Meeting*, 47(3), 553–557.
- Wright, M. C. & Kaber, D. B. (2005). Effects of automation of information – processing functions on teamwork. *Human Factors*, 47(1), 50–66.
- Zhang, M. Y. & Yang, X. J. (2017). Evaluating effects of workload on trust in automation, attention allocation and dual-task performance. *Proceedings of the Human Factors and Ergonomics Society Annual Meeting*, 61(1), 1799–1803.



Einblicke in die Anwendungsgebiete des Human Autonomy Teaming's



Leonie Küster, Jana Masch & Mark Sturhann



Inhaltsverzeichnis

4	Einblicke in die Anwendungsgebiete des Human Autonomy Teamings	
4.1	Autonomes Fahren und Human Autonomy Teaming	113
4.1.1	Autonome Technik in PKWs	
4.1.2	Autonome Technik in LKWs	
4.1.3	Praxisbeispiele von HAT im öffentlichen Verkehr und Gütertransport	
4.2	Human Autonomy Teaming im Militär	133
4.2.1	Autonome unbemannte Flug- und Landfahrzeuge	
4.2.2	Assistenzsysteme an Bord von Marineschiffen	
4.3	Human Autonomy Teaming im Weltraum	138
4.3.1	Praxisbeispiele autonomer Robotik im Weltraum	
4.3.2	Stimmen aus der Praxis zu autonomen Agenten im Weltraum	
4.4	Fazit zum Praxiskapitel	144
	Literaturverzeichnis	145

Kapitel 4: Einblicke in die Anwendungsgebiete des Human Autonomy Teamings

Von Leonie Küster, Jana Masch und Mark Sturhann

Das Human Autonomy Teaming (HAT) weist als Konzept der Teamarbeit zwischen autonomen Agenten und menschlichen Teammitgliedern mehrere Anwendungsgebiete auf. In diesem Kapitel werden drei große Anwendungsfelder des HAT präsentiert: Autonomes Fahren und HAT, HAT im Militär sowie HAT im Weltraum. In dieser Reihenfolge werden die Anwendungsgebiete in den folgenden drei Unterkapiteln vorgestellt. Zu jedem Anwendungsgebiet werden Praxisbeispiele des HAT dargelegt, um Einblicke in die Anwendungsgebiete des HAT zu geben. Im Rahmen des ersten und des dritten Unterkapitels wurden zusätzlich Interviews mit Experten zu autonomer Technologie durchgeführt.

4.1 Autonomes Fahren und Human Autonomy Teaming

Zu den Anwendungsgebieten des HAT zählt unter anderem das autonome Fahren im Straßenverkehr, im öffentlichen Straßenverkehr und im Gütertransport. Im vorliegenden ersten Unterkapitel werden deshalb Praxisbeispiele für die Anwendung des HAT in diesen drei Bereichen des autonomen Fahrens vorgestellt. Bei jedem Praxisbeispiel wird begründet, warum das entsprechende Fahrzeug als autonomer Agent bezeichnet werden kann und warum im Rahmen des jeweiligen Praxisbeispiels ein HAT vorliegt. Dies erfolgt im vorliegenden Unterkapitel zunächst auf formaler Ebene durch die Übertragung der Automationslevel-Taxonomie autonomer Agenten von Parasuraman et al. (2000) auf die Automationslevel-Taxonomie automatisierter Fahrzeuge der Society of Automobile Engineers International (SAE International, 2016). Mit Hilfe der Definitionen zu den Begriffen HAT und autonomer Agent, die bereits im ersten Kapitel dieses E-Books detailliert behandelt worden sind (vgl. Kapitel 1.1), wird dann auf sachlicher Ebene eine Übertragung auf das Anwendungsgebiet des autonomen Fahrens vorgenommen. Auf dieser Ebene wird der autonome Agent angesichts seiner Aktivitäten und Handlungen analysiert. Im Gegensatz dazu, wird auf der formalen Ebene lediglich der Automationsgrad des autonomen Agenten betrachtet (vgl. Abb. 4.1). Die Erfüllung der Charakteristika autonomer Agenten und des HAT wird dann bei den einzelnen Praxisbeispielen überprüft. Anschließend erfolgt die Klassifizierung dieser Praxisbeispiele als ein HAT. Nach der Gegenüberstellung der zwei eben erwähnten Automationslevel-Taxonomien werden zuerst Praxisbeispiele des HAT im Bereich autonomer Technik von PKWs präsentiert. Im zweiten Abschnitt dieses Unterkapitels wird ein HAT-Praxisbeispiel im Bereich autonomer Technik in LKWs vorgestellt. Sowohl im ersten als auch im zweiten Abschnitt dieses Unterkapitels wurde ein Interview mit einem Experten zum autonomen Fahren durchgeführt. Im dritten und letzten Abschnitt werden dann Praxisbeispiele von HAT im öffentlichen Straßenverkehr, sowie im Gütertransport präsentiert.

4.1.1 Autonome Technik in PKWs

Der erste Abschnitt des vorliegenden Unterkapitels beinhaltet die bereits angesprochene Gegenüberstellung der zwei Automationslevel-Taxonomien. Diese soll eine formale Übertragung der Konzepte des autonomen Agenten und des HAT auf den Bereich des autonomen Fahrens ermöglichen. Anschließend werden kurz die wichtigsten Charakteristika autonomer Agenten und des HAT angeführt. Dadurch erfolgt auf sachlicher Ebene eine Übertragung der zwei eben erwähnten Konzepte auf das autonome Fahren. Darauf folgend werden zwei Praxisbeispiele von HAT im Bereich hochautomatisierter und autonomer PKWs vorgestellt. Das Interview mit

einem Experten des ADAC e. V. zum autonomen Fahren bildet den Abschluss des ersten Teils vom Unterkapitel zum autonomen Fahren.

Übertragung des Human Autonomy Teamings auf das Anwendungsgebiet autonomes Fahren

O'Neill et al. (2020) definieren einen autonomen Agenten als eine computerbasierte Einheit, die im Team als Individuum wahrgenommen wird. Dieses Individuum besetzt eine bestimmte Rolle in diesem Team. HAT stellt O'Neill et al. (2020) nach, eine gegenseitige Abhängigkeit in Aktivitäten und Ergebnissen dar, die mindestens ein menschliches Teammitglied und mindestens einen autonomen Agenten als Teammitglied beinhaltet. Diese stehen in einem Abhängigkeitsverhältnis und streben zudem an, ein gemeinsames Ziel im Kollektiv zu erreichen. Damit ein autonomer Agent als vollwertiges Teammitglied gilt, muss dieser unabhängig von anderen Teammitgliedern handeln können. Zusätzlich muss sich dieser im Rahmen eines HAT auch selbst kontrollieren können und Eigeninitiative aufweisen (O'Neill et al., 2020). Eine weitere essenzielle Anforderung an einen autonomen Agenten ist ein hoher Automationsgrad (Autonomie), damit dieser in komplexen Situationen die Rolle eines menschlichen Teammitgliedes einnehmen kann. Dafür sollte der autonome Agent die jeweilige Aufgabe des Teams verstehen können. Des Weiteren sind die Interaktion und Kommunikation des autonomen Agenten mit den anderen Teammitgliedern des HAT von sehr hoher Wichtigkeit (McNeese et al., 2018).

Wie in Kapitel 1.3 ausführlich beschrieben, sind bei der Klassifikation des Automationsgrades autonomer Agenten im Rahmen der Automationslevel-Taxonomie von Parasuraman et al. (2000) zehn Automationslevel zu unterscheiden. Das Automationslevel 10 stellt den höchsten Automationsgrad dar und das Automationslevel 1 den geringsten, sodass im Rahmen des ersten Levels der Mensch die vollständige Kontrolle über den autonomen Agenten hat. Währenddessen hat beim Vorliegen des Automations-Levels 10 der autonome Agent die vollständige Entscheidungshoheit und agiert autonom, das heißt ohne menschliches Zutun. Der Grad der Automatisierung des autonomen Agenten nimmt vom ersten bis zum zehnten Level zu. Die zehn Automationslevels beziehen sich auf die Informationsaufnahme, Informationsanalyse, Handlungsauswahl und Handlungsausführung eines autonomen Agenten. Mit zunehmendem Automationslevel kann der autonome Agent eine höhere Stufe der Informationsverarbeitung ausführen (Parasuraman et al., 2000). Ab dem Automationslevel 5 der Automationslevel-Taxonomie von Parasuraman et al. (2000) liegt gemäß O'Neill et al. (2020) ein teilautonomer Agent vor. Ab dem Automationslevel 6 wird von einem autonomen Agenten gesprochen.

Ein HAT liegt dann vor, wenn der autonome Agent mindestens das Automationslevel 5 (partielle Autonomie) der Automationslevel-Taxonomie von Parasuraman et al. (2000) aufweist: „To meet criteria for autonomous agents and to qualify as a HAT, at least partial or high autonomous agents as defined above would be required of the computer-based entities (i. e., autonomous agents) involved.“ (vgl. O'Neill et al., 2020, S. 7-8). Übertragen auf die Automationslevel-Taxonomie von Parasuraman et al. (2000) bedeutet das, dass ein autonomer Agent mit der Freigabe seines menschlichen Teammitglieds mindestens eine selbst vorgeschlagene Aktion selbstständig ausführen können muss, um als autonom zu gelten (Parasuraman et al., 2000). Die wichtigsten Charakteristika eines autonomen Agenten für die Übertragung des

HAT-Konzepts auf die nachfolgenden Praxisbeispiele wurden in Abbildung 4.1 zusammengefasst.

Übersicht zu den zentralen Charakteristika eines autonomen Agenten:

- I. Formal (anhand der Automationslevel-Taxonomie Parasuramans):
 - Weist mindestens das Automationslevel 5 gemäß Parasuraman et al. (2000) auf
 - Kann mit Freigabe des menschlichen Teammitglieds eine selbst vorgeschlagene Aktion eigenständig ausführen
- II. Sachlich (anhand der Schlussfolgerungen aus der Taxonomie):
 - Besetzt eine bestimmte Rolle im HAT
 - Steht mit dem menschlichen Teammitglied in einem Abhängigkeitsverhältnis
 - Kann unabhängig vom menschlichen Teammitglied handeln
 - Selbstkontrolle und Eigeninitiative
 - Interaktion und Kommunikation mit dem menschlichen Teammitglied
 - Verständnis von den gemeinsamen Aufgaben im Team
 - Wahrnehmung als individuelles Teammitglied durch den Menschen

Abb. 4.1: Zentrale Charakteristika eines autonomen Agenten für die Übertragung auf die Praxisbeispiele (in Anlehnung an Parasuraman et al., 2000, S. 287; O'Neill et al., 2020, S. 4-22).

Für den Transfer dieser grundsätzlichen Charakteristika des autonomen Agenten und des HAT auf die Automationslevel-Taxonomie automatisierter Straßenfahrzeuge der SAE International muss demnach das SAE-Level eines automatisierten Fahrzeugs festgestellt werden, ab dem dieses als autonom bezeichnet werden kann. Dann sollte dieser Automationslevel dem autonomen Agenten gemäß der Definition von O'Neill et al. (2020) zugeordnet werden können. Mit der Erfüllung dieser Anforderung lässt sich dann ebenfalls festhalten, wann im Bereich des autonomen Fahrens ein HAT vorliegt. Die Übertragung der Automationslevel-Taxonomie von Parasuraman et al. (2000) auf die Automationslevel-Taxonomie der SAE International und die des HAT-Begriffes auf das autonome Fahren, werden bei den einzelnen Praxisbeispielen autonomen Fahrens in den folgenden Abschnitten jeweils vorgenommen. Für diese Übertragung muss aber auch geklärt sein, unter welchen Bedingungen eine fahrende Person und ein hochautomatisiertes oder autonomes Fahrzeug zusammen ein HAT darstellen. Um als Team gelten zu können, muss die fahrende Person mit dem autonomen Fahrzeug zusammenarbeiten.

Um die Automationslevel-Taxonomie von Parasuraman et al. (2020) auf das autonome Fahren beziehen zu können, wird an dieser Stelle die Automationslevel-Taxonomie automatisierter Fahrzeuge im Straßenverkehr gemäß der SAE International dargestellt. Aus dieser lässt sich entnehmen, wann ein Fahrzeug als autonom zu bezeichnen ist. Dadurch kann eine präzise Übertragung der Automationsgrade autonomer Agenten eines HAT auf die Automationsgrade eines autonomen Fahrzeugs erfolgen. Gemäß der aktuellsten Fassung der SAE International Standards zum automatisierten Fahren werden hinsichtlich der Automation eines Fahrzeugs und der Rolle der fahrenden Person sechs Automationslevel unterschieden. Diese Taxonomie bezieht sich auf automatisierte und motorbetriebene Fahrzeugsysteme. Die einzelnen Automatisierungsgrade werden außerdem im Hinblick auf die Aufgaben und Funktionen des automatisierten Fahrzeugsystems differenziert. Beim Automationslevel 0 liegt keine Automation des Fahrzeugs vor, während das Automationslevel 5 die vollständige Automation des Fahrzeugs bedeutet. Im Bereich der Automationslevel 3 bis 5 führen die Funktionen des automatisierten Fahrsystems die Fahraufgaben selbst durch. Die Betitelung eines automatisierten

Fahrzeugsystems mit den einzelnen Automationslevels erfolgt also abhängig von dessen Erfüllung bestimmter Fahraufgaben, der Fahrfunktionen des Fahrzeugs, sowie abhängig von der daraus folgenden Rolle der fahrenden Person (SAE International, 2018).

Die Automationslevel 0 bis einschließlich 2 kennzeichnen automatisiertes Fahren, bei dem die fahrende Person das Fahrzeug steuert. Von den Automationslevels 3 bis 5 agiert das automatisierte Fahrzeug selbständiger und weniger mit Hilfe der fahrenden Person (SAE International, 2016). Beim automatisierten Fahren im Bereich der ersten drei Automationslevel (null bis zwei) hat die fahrende Person dauerhaft die unterstützenden Funktionen des automatisierten Fahrzeugs zu überwachen. Die fahrende Person muss des Weiteren die Steuerung, Beschleunigung und das Durchführen von Bremsvorgängen des automatisierten Fahrzeugs übernehmen. Die die fahrende Person unterstützenden Funktionen des automatisierten Fahrzeugs umfassen im Bereich des Automationslevels 1 Warnsignale und kurze Assistenz, z.B. automatische Notfall-Bremssysteme. Das Automationslevel 2 umfasst Steuerungs-, Brems- oder Beschleunigungshilfen für die fahrende Person, wie beispielsweise Spurhalteassistenten. Automatisierte Fahrzeuge des Automationslevels 3 müssen sowohl Steuerungs- als auch Brems- und Beschleunigungshilfen für die fahrende Person bereitstellen, z.B. Spurhalteassistenten und Geschwindigkeitsassistenten (SAE International, 2016). Ab dem Automationslevel 3 kommt es des Weiteren zu einer entscheidenden Veränderung bei der Interaktion der fahrenden Person mit dem automatisierten Fahrzeug: Die fahrende Person kann ab diesem Automationslevel die Kontrolle über dieses an das Fahrzeug abgeben (Wood et al., 2019). Das automatisierte Fahrzeug muss der fahrenden Person vor ihrer Übernahme der Fahrzeugkontrolle zunächst Bescheid geben, dann kann die fahrende Person das automatisierte Fahrzeug wieder selbst fahren. Wie beim Automationslevel 4 sollte das automatisierte Fahrzeug in diesem Klassifizierungsbereich unter bestimmten Bedingungen selbständig fahren können. Ab dem Automationslevel 4 und beim Automationslevel 5 benötigt das automatisierte Fahrzeug die fahrende Person nicht mehr zur Kontrollübernahme. Im Gegensatz zum Automationslevel 4 sollte ein automatisiertes Fahrzeug des Automationslevels 5 unter allen Bedingungen durch die entsprechenden Funktionen selbständig fahren können (SAE International, 2016).

Von der Automationslevel-Taxonomie der SAE International werden aktive Sicherheitssysteme, wie beispielsweise elektronische Stabilitätskontrollsysteme, Notbremssysteme und bestimmte Typen von Fahrassistenzsystemen, wie z.B. Fahrspurassistenzsysteme ausgeschlossen. Dies ist darin begründet, dass diese Systeme alleine nicht alle zentralen Fahraufgaben auf eine anhaltende und dauerhafte Art und Weise ausführen können. Diese Systeme stellen eher momentane und vorübergehende Interventionen während potenziell gefährlichen Situationen bereit. Der Eingriff dieser aktiven Sicherheitssysteme verändert außerdem nicht die Rolle der fahrenden Person im Rahmen der Ausführung der Fahraufgaben. Die Rolle der fahrenden Person kann durch diese ebenfalls nicht übernommen werden. In diesem Zusammenhang ist aber zu erwähnen, dass automatisierte Fahrzeugsysteme in der Regel Kollisions-Schutz-Funktionen aufweisen, die wiederum eingriffsfähige aktive Sicherheitssysteme beinhalten. Kollisions-Schutz-Funktionen werden teilweise bei Fahrzeugsystemen der Automationslevel drei bis fünf eingesetzt (SAE International, 2018).

Da die Automationslevel-Taxonomie der SAE International für gewöhnlich auch von Automobilunternehmen herangezogen wird, erfolgt die Übertragung der Begriffe HAT und autonomer Agent auf das autonome Fahren anhand dieser. Für das Vorliegen eines HAT muss zumindest ein teilautonomer Agent vorliegen. Dieser teilautonome Agent liegt nach O'Neill et al. (2020) formal gesehen vor, wenn er mindestens ein Automationslevel von 5, gemäß der Automationslevel-Taxonomie von Parasuraman et al. (2000) aufweist. In Anbetracht der Tatsache, dass

die fahrende Person eines automatisierten Fahrzeugs ab dem SAE-Automationslevel 3 vorübergehend die Kontrolle über das Fahrzeug an das autonome Fahrsystem von diesem abgeben kann, kann dieses Fahrzeug als teilautonom bezeichnet werden. Das automatisierte Fahrzeug agiert ab diesem Level selbständiger und weniger mit Hilfe der fahrenden Person (SAE International, 2016). Diese Selbständigkeit des automatisierten Fahrzeugs bedingt aber die manuelle Freigabe der fahrenden Person. Auch vor der Wiedererlangung der Kontrolle über das automatisierte Fahrzeug durch die fahrende Person muss das System des Fahrzeugs dieser zunächst ein Signal geben. Ab den Automationsleveln 5 bis 6 der Automationslevel-Taxonomie von Parasuraman et al. (2000) kann von einem teilautonomen Agenten gesprochen werden (O'Neill et al., 2020). Somit können die Automationslevel 5 und 6 eines autonomen Agenten gemäß der Automationslevel-Taxonomie von Parasuraman et al. (2000) mit dem Automationslevel 3 der Automationslevel-Taxonomie der SAE International gegenübergestellt werden. Dies wird in der untenstehenden Abbildung 4.2 grafisch veranschaulicht. Die in der Abbildung hervorgehobenen Automationslevel 5 bis 6 eines autonomen Agenten und das in dieser Abbildung markierte SAE-Automationslevel 3 sollen die Gegenüberstellung der Automationslevel-Taxonomie nach Parasuraman et al. (2000) mit der Automationslevel-Taxonomie der SAE International (2016) verdeutlichen. In jedem der beiden erwähnten Automationslevel-Bereiche der zwei Taxonomien benötigt der autonome Agent zunächst die Freigabe des menschlichen Teammitglieds und gibt diesem ebenfalls die Möglichkeit für die Rückkehr zur manuellen Kontrolle über sein System (Parasuraman et al., 2000). Folglich kann ein automatisiertes Fahrzeug ab einem SAE-Automationslevel 3 formal als ein autonomer Agent im Anwendungsgebiet des autonomen Fahrens bezeichnet werden, da es ab dann als teilautonom gilt und vorübergehend selbständig agieren kann. Dadurch liegt aus formaler Perspektive ab diesem SAE-Automationslevel beim autonomen Fahren ein HAT vor, sofern die fahrende Person als menschliches Teammitglied in der Interaktion mit dem Fahrzeug eine Rolle als Teammitglied einnimmt. Außerdem sollten sowohl die fahrende Person als auch das teilautonome Fahrzeug ein gemeinsames Ziel im Kollektiv erreichen wollen (O'Neill et al., 2020).

Sachlich gesehen sind an das Vorliegen eines autonomen Agenten im Bereich des autonomen Fahrens die bereits oben erwähnten Bedingungen gekoppelt. Ein automatisiertes Fahrzeug ab dem SAE-Automationslevel 3 erfüllt die für die Bezeichnung als ein autonomer Agent notwendigen Bedingungen, wenn es von der fahrenden Person als Teammitglied wahrgenommen wird. Ab diesem Automationslevel kann das automatisierte Fahrzeug selbständig fahren. Es interagiert und kommuniziert mit der fahrenden Person vor der Kontrollübernahme und Kontrollabgabe. Dadurch besetzt es eine bestimmte Rolle im Team. Daher wird auch aus sachlicher Perspektive fortan ab dem SAE-Automationslevel 3 von einem autonomen Agenten und damit von einem HAT im Anwendungsgebiet des autonomen Fahrens gesprochen. Die Gegenüberstellung der beiden Automationslevel-Taxonomien wird in Abbildung 4.2 grafisch veranschaulicht. Dabei wurden die für die Übertragung des HAT-Konzepts entscheidenden, zu gegenüberstellenden Automationslevel farblich von den anderen Automationslevels hervorgehoben.

Automationslevel-Taxonomie nach Parasuraman et al. (2000)	Automationslevel-Taxonomie der SAE International (2016)
<p>Automationslevel 1: Das menschliche Teammitglied übernimmt alle Aufgaben und der autonome Agent bietet keine Unterstützung an.</p> <p>Automationslevel 2: Der autonome Agent schlägt dem menschlichen Teammitglied zahlreiche Handlungsalternativen vor.</p> <p>Automationslevel 3: Der autonome Agent begrenzt seinen Vorschlag auf die wichtigsten und besten Handlungsalternativen.</p> <p>Automationslevel 4: Der autonome Agent schlägt dem menschlichen Teammitglied eine Handlungsalternative vor.</p> <p>Automationslevel 5: Der autonome Agent führt seine vorgeschlagene Handlungsalternative selbst aus, wenn das menschliche Teammitglied zustimmt.</p> <p>Automationslevel 6: Der autonome Agent gibt dem menschlichen Teammitglied für eine begrenzte Zeit die Möglichkeit, gegen sein Vorhaben einen Einspruch zu erheben.</p> <p>Automationslevel 7: Der autonome Agent handelt automatisch eigenständig und muss dann das menschliche Teammitglied über seine Handlung informieren.</p> <p>Automationslevel 8: Der autonome Agent informiert das menschliche Teammitglied nur, wenn dies vorher festgelegt wurde.</p> <p>Automationslevel 9: Der autonome Agent informiert das menschliche Teammitglied nur, wenn er sich selbst dafür entscheidet.</p> <p>Automationslevel 10: Der autonome Agent handelt vollkommen autonom und ignoriert das menschliche Teammitglied.</p>	<p>Automationslevel 0 bis einschließlich 2: Die fahrende Person steuert das automatisierte Fahrzeug selbst. Sie wird mit zunehmendem Automationslevel durch mehr Assistenzsysteme unterstützt: Warnsysteme (ab Level 0), entweder Steuer- oder Brems- und Beschleunigungsassistent (ab Level 1) und Steuer-, Brems- und Beschleunigungsassistent (ab Level 2).</p> <p>Automationslevel 3: Unter bestimmten Bedingungen fährt das automatisierte Fahrzeug autonom. Die fahrende Person steuert das automatisierte Fahrzeug nur, wenn die Fahrzeugsysteme sie dazu auffordern. Ab diesem Level steuert die fahrende Person das automatisierte Fahrzeug nicht mehr selbstständig, sofern die Systeme des automatisierten Fahrzeugs aktiviert sind.</p> <p>Automationslevel 4 bis einschließlich 5: Das automatisierte Fahrzeug fährt unter allen Bedingungen autonom und benötigt die fahrende Person nicht mehr zur Übernahme der Steuerung.</p>

Abb. 4.2: Gegenüberstellung der beiden Automationslevel-Taxonomien (in Anlehnung an Parasuraman et al., 2000, S. 287; SAE International, 2016).

Praxisbeispiele hochautomatisierter und autonomer PKWs

Im vorliegenden, zweiten Abschnitt des Unterkapitels zum autonomen Fahren werden zwei Praxisbeispiele autonomer Agenten und von HAT im Bereich hochautomatisierter und autonomer PKWs präsentiert. Zunächst werden beide automatisierte Fahrzeuge hinsichtlich ihrer Funktionen beschrieben. Anschließend werden die beiden Fahrzeuge ausgehend von den Schlussfolgerungen aus dem ersten Abschnitt dieses Unterkapitels (vgl. Abb. 4.1) als autonome Agenten und im Zusammenspiel mit der fahrenden Person als HAT klassifiziert. Dies wird erst auf formaler Ebene durchgeführt, sprich mithilfe der SAE-Automationslevel der beiden automatisierten Fahrzeuge. In diesem Zusammenhang wird die Gegenüberstellung der beiden Automationslevel-Taxonomien herangezogen (vgl. Abb. 4.2). Anschließend werden die beiden automatisierten Fahrzeuge auf sachlicher Ebene als autonome Agenten und in Interaktion mit der fahrenden Person, als HAT klassifiziert. Auf dieser Ebene werden die Charakteristika autonomer Agenten, die nicht den Automationslevel betreffen, mit den Funktionen der beiden Fahrzeuge verglichen.

Erstes Praxisbeispiel eines HAT im Bereich autonomer Technik in PKWs

Die neuesten Automobilmodelle integrieren vielerlei Funktionen, die den Menschen im Rahmen der Bewältigung der gemeinsamen Fahraufgaben unterstützen. Das erste Praxisbeispiel autonomen Fahrens in diesem Unterkapitel stellt das Fahrzeug „S 500 Intelligent Drive“ von Mercedes-Benz dar. Die Fahrassistenzsysteme des Automobilmodells, der Aktive Abstands-Assistent und der Aktive Lenk-Assistent, unterstützen die fahrende Person beim Abstandhalten und Lenken. Des Weiteren weisen neue Modelle einen Aktiven Nothalt-Assistenten und einen Aktiven Spurwechsel-Assistenten auf. Die Bedienung aller Assistenzfunktionen steuert die fahrende Person über das Lenkrad (Daimler, n. d.a). Das bedeutet, dass das automatisierte Fahrzeug mit der Freigabe durch die fahrende Person über das Lenkrad selbständig agieren kann. Dies entspricht der Funktion eines autonomen Agenten des Automationslevels 5 gemäß der Automationslevel-Taxonomie nach Parasuraman et al. (2000) (vgl. Abb. 4.1). Die Fahrzeugassistenten, die im Folgenden beschrieben werden, übernehmen dann zusammen die Fahraufgaben. Das automatisierte Fahrzeug unterstützt sein menschliches Teammitglied, die fahrende Person, bei ihren Fahraufgaben.

Der Aktive Abstands-Assistent reduziert selbständig die Geschwindigkeit, wenn das Fahrzeug sich Engstellen wie Kreuzungen oder Kreisverkehren nähert. Nach dem Passieren dieser Engstellen beschleunigt der Assistent das Fahrzeug. Das Fahrzeug kann mit Hilfe dieses Assistenten unabhängig von der fahrenden Person agieren, sofern diese den Assistenten aktiviert. Die Navigation des Fahrzeugs ist eng mit der Geschwindigkeitssteuerung des Assistenten verbunden. Beispielsweise verringert der Assistent die Geschwindigkeit des Fahrzeugs, wenn die Navigation des Fahrzeugs eine Autobahnausfahrt anzeigt und sich das Fahrzeug auf der rechten Fahrspur der Autobahn vor der Ausfahrt befindet. Das Fahrzeug kann sich demnach selbst kontrollieren und weist unter bestimmten Fahrbedingungen eine Eigeninitiative auf. Auf Schnellstraßen und Autobahnen regelt der aktive Abstands-Assistent Abstände zu vorausfahrenden Fahrzeugen und hält die Spur. Der Aktive Spurwechsel-Assistent wird von der fahrenden Person aktiviert, indem diese den Blinker betätigt (Daimler, n. d.a). In diesem Fall muss das menschliche Teammitglied erneut zunächst den Assistenten des Fahrzeugs bzw. das Fahrzeug selbst aktivieren. Erst dann agiert dieses selbständig. Das menschliche Teammitglied steht mit dem automatisierten Fahrzeug also in einem Abhängigkeitsverhältnis und interagiert mit diesem. Nach dem Antippen des Blinkers durch die fahrende Person überprüft die Sensorik des Fahrzeugs, ob die Nebenspur vor, neben und hinter dem Fahrzeug frei ist. Die Funktionsweise der Sensorik des Fahrzeugs in Verbindung mit dem Aktiven Spurwechsel-

Assistenten wird in der untenstehenden Abbildung 4.3 veranschaulicht. Zusätzlich wird beim hochautomatisierten Spurenwechsel auch die Geschwindigkeit anderer Fahrzeuge berücksichtigt. Der Aktive Spurwechsel-Assistent unterstützt die fahrende Person erst, wenn sich im Sicherheitsbereich kein anderes Fahrzeug befindet (Daimler, n. d.a). Das Fahrzeug kontrolliert sich auf diese Weise selbst und nimmt mit Hilfe dieses Assistenten erneut eine unterstützende Rolle in Zusammenarbeit mit der fahrenden Person ein.

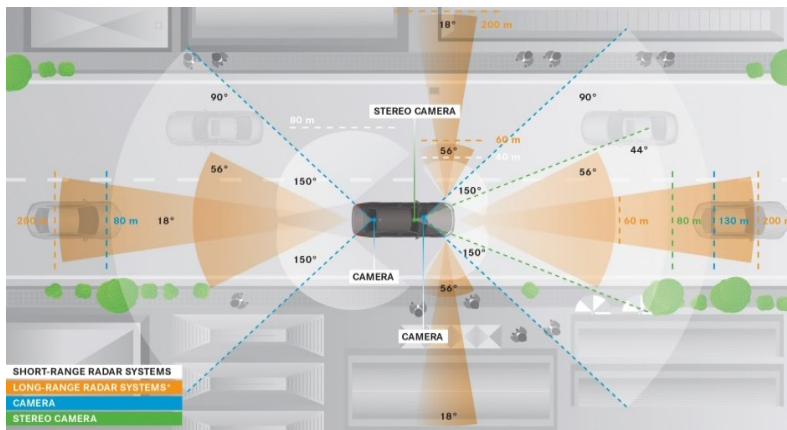


Abb. 4.3: Funktionsweise der Sensorik, Kamera- und Radarsysteme des hochautomatisierten Fahrzeugs (Foto zur Verfügung gestellt von Daimler, 2013).

Der Aktive Geschwindigkeitslimit-Assistent des automatisierten Fahrzeugs stellt eine Teilfunktion des Verkehrszeichen-Assistenten dar und kann in Verbindung mit dem System „COMAND Online“ über die Fahrzeugkameras Geschwindigkeitsbeschränkungen erkennen. In dieser Hinsicht wird erneut deutlich, dass sich das automatisierte Fahrzeug selbst kontrollieren kann, sofern der entsprechende Assistent vom menschlichen Teammitglied aktiviert wurde. Der Assistent berücksichtigt ebenfalls der Fahrzeugnavigation bekannte Geschwindigkeitslimits. Anhand der erkannten und der Navigation bekannten Geschwindigkeitsbegrenzungen passt dann der aktive Abstands-Assistent selbständig die Fahrzeuggeschwindigkeit an (Daimler, n. d.a). Das Fahrzeug weist auch in Form dieses Assistenten eine Eigeninitiative auf und unterstützt das menschliche Teammitglied. Die fahrende Person kann die Geschwindigkeit des Fahrzeugs jederzeit verändern. In dieser Hinsicht interagieren das menschliche Teammitglied und das automatisierte Fahrzeug miteinander und sind voneinander abhängig. Beim Staufolgefahren, z.B. auf der Autobahn, kann das Fahrzeug mit Haltezeiten von bis zu 30 Sekunden anschließend selbständig wieder anfahren und dem vorausfahrenden Fahrzeug folgen. Der aktive Nothalt-Assistent des Fahrzeugs überprüft, ob die fahrende Person während der Fahrt und mit eingeschaltetem Lenk-Assistenten dauerhaft nicht mehr ins Fahrgeschehen eingreift (Daimler, n. d.a). Dieser Assistent ermöglicht dem Fahrzeug, das menschliche Teammitglied zu überwachen. Zunächst fordert das Fahrzeugsystem die fahrende Person optisch und akustisch dazu auf, die Hände an das Lenkrad zu nehmen, sofern diese seit längerem keine Aktivität gezeigt hat (Daimler, n. d.a). Die Kommunikation des menschlichen Teammitglieds mit dem Fahrzeug ist in diesem Fall unabdingbar. Zeigt die fahrende Person keine Reaktion auf die Warnungen des Fahrzeugs, dann sorgt der Assistent dafür, dass das Fahrzeug in der eigenen Fahrspur bis zum Stillstand die Geschwindigkeit verzögert. Der nachfolgende Verkehr wird bei dieser Geschwindigkeitsverzögerung des Fahrzeugs auf unter 60 Stundenkilometer über die Einschaltung des Warnblinkers durch den Assistenten gewarnt. Bei einem Stillstand des Fahrzeugs werden zudem automatisch die Parkbremse des Fahrzeugs, sowie das Mercedes-Benz Notrufsystem aktiviert. Zusätzlich kommt es durch den Assistenten zu einer Entriegelung des Fahrzeuges, um Ersthelfenden den Zutritt in das Fahrzeug zu erleichtern. Greift

die fahrende Person aber wieder in das Fahrgeschehen ein, wird der Abbruch der aktivierten Nothalt-Funktionen eingeleitet. Das Fahrzeug hilft seinem menschlichen Teammitglied und sorgt für dessen Sicherheit. Es nimmt also nicht nur eine unterstützende und kontrollierende, sondern auch eine sicherheitsbezogene Rolle in Zusammenarbeit mit der fahrenden Person ein (Daimler, n. d.a).

Der Aktive Brems-Assistent unterstützt die fahrende Person bei drohenden Kollisionen mit anderen Fahrzeugen, sofern diese keine für die Entschärfung der Gefahrensituation erforderlichen Maßnahmen ergreift. Vor der drohenden Kollision wird die fahrende Person zunächst durch eine optische Warnung des Fahrzeugs auf einen zu geringen Abstand zu einem Gefahrenobjekt aufmerksam gemacht. Hat das Fahrzeug eine akute Kollisionsgefahr erkannt, wird die fahrende Person zusätzlich akustisch gewarnt (Daimler, n. d.a). Das automatisierte Fahrzeug kommuniziert auf diese Art mit dem menschlichen Teammitglied. Der Assistent ist imstande, autonom eine Notbremsung vorzunehmen, sofern sich dem Fahrzeug ein Gefahrenobjekt nähert (Daimler, n. d.a). Dem Fahrzeug kommt auf diese Weise eine Schutzfunktion in Bezug auf das menschliche Teammitglied zu. Bremst die fahrende Person selbst, wird beim Bremsvorgang eine situationsgerechte Bremsunterstützung durch das Fahrzeug gewährleistet. Der aktive Spurhalte- und der Totwinkel-Assistent des Fahrzeugs warnen die fahrende Person durch Lenkradvibrationen, wenn das Fahrzeug unbeabsichtigt die Fahrspur verlässt. Beim Überfahren einer durchgezogenen Linie der Fahrspur wird das Fahrzeug durch eine einseitige Bremsung dann in die Fahrspur zurückgezogen (Daimler, n. d.a). Das Fahrzeug nimmt demnach ebenfalls eine korrigierende Funktion in Bezug auf sein menschliches Teammitglied ein. Überfährt das Fahrzeug unbeabsichtigt eine gestrichelte Linie einer Fahrspur, erfolgt ein Eingriff des Assistenten nur dann, wenn auf der Nebenspur durch ein anderes Fahrzeug Kollisionsgefahr besteht. Der Assistent warnt die fahrende Person optisch und akustisch auch vor seitlichen Kollisionen mit anderen Fahrzeugen, wenn die fahrende Person vorher den Blinker betätigt hat (Daimler, n. d.a). Das menschliche Teammitglied muss das Fahrzeug für dessen Unterstützung stets aktivieren. Die fahrende Person hängt auf der einen Seite von der Unterstützung des Fahrzeugs ab. Auf der anderen Seite benötigt das automatisierte Fahrzeug zunächst die Freigabe und Aktivierung seiner Assistenten durch sein menschliches Teammitglied. Das Fahrzeug kann mit Hilfe des Assistenten zudem eine automatische einseitige Bremsung vollziehen, sofern eine seitliche Kollision mit einem anderen Gefahrenobjekt droht (Daimler, n. d.a). In der Teilabbildung 4.4a wird das autonome Wirken des Fahrzeugs beim Fahrspurhalten verdeutlicht. In der Teilabbildung 4.4b wird gezeigt, dass das Fahrzeug autonom lenken kann (Aktiver Lenk-Assistent, siehe oben). In beiden Bildern wird deutlich: Die fahrende Person hat ihre Hände nicht am Lenkrad. Sie überwacht das Fahrzeug, um gegebenenfalls eingreifen zu können.



Abb. 4.4a und 4.4b: HAT mit dem autonomen Agenten Mercedes-Benz S500 Intelligent Drive im Bereich des autonomen Fahrens (Fotos zur Verfügung gestellt von Daimler, 2013).

Das automatisierte S-Klasse-Modell S 500 Intelligent Drive von Mercedes-Benz lässt sich gemäß der SAE International Automationslevel-Taxonomie als ein Fahrzeug des Automationslevels 3 bis 4 klassifizieren, weil es zwar vorübergehend autonom fahren kann, aber dafür von der fahrenden Person aktiviert werden muss. Zum Beispiel der Aktive Spurwechsel-Assistent unterstützt die fahrende Person bei einem Spurenwechsel erst dann, wenn die fahrende Person den Blinker des Fahrzeugs betätigt. Diese Aktivierung durch die fahrende Person ist auch für die Tätigkeit der anderen Assistenten des Fahrzeugs erforderlich. Auch die Rückübertragung der Kontrolle über das Fahrzeug auf die fahrende Person erfolgt erst, wenn sie dies dem Fahrzeug gegenüber bestätigt. Beispielsweise beim Aktiven Nothalte-Assistenten wird die fahrende Person bei längerer Passivität vom System des Fahrzeugs dazu aufgefordert, ihre Hände an das Lenkrad zu nehmen. Aus diesem Grund ist das Fahrzeug bezüglich seines Automationsgrades nicht deutlich dem Automationslevel 4 zuzuordnen. Auf der anderen Seite kann das Fahrzeug unter bestimmten Bedingungen autonom fahren und die fahrende Person wird z.B. im Rahmen des Staufolgefahrens auf der Autobahn teilweise nicht mehr zur Kontrollübernahme benötigt. Dies spricht für eine Klassifizierung der S-Klasse als ein Fahrzeug des Automationslevels 4. Da die fahrende Person teilweise noch zur Kontrollübernahme des Fahrzeugs benötigt wird, kann man hier nicht von hochautomatisiertem Fahren sprechen. Gegen die Klassifizierung des Fahrzeugs überhaupt könnte die gesonderte Betrachtung der einzelnen Assistenten des Fahrzeugs sprechen. Diese könnten, alleine betrachtet, möglicherweise von der Automationslevel-Taxonomie der SAE International ausgeschlossen werden, da sie in die Kategorie aktiver Schutzsysteme fallen. Dagegen spricht aber, dass die einzelnen Assistenten des Fahrzeugs im Kollektiv dafür sorgen, dass dieses nicht nur vorübergehend, sondern auch anhaltend, z.B. beim Staufolgefahren auf der Autobahn, selbständig fahren kann. Die Frage wäre demnach, wie lange die Assistenten ihre jeweiligen Funktionen ausführen können und was die SAE International unter vorübergehender und anhaltender Unterstützung auf der anderen Seite versteht. Da die einzelnen Assistenten der S-Klasse aber nicht nur für eine kurze Zeit aktiv sein können, wird der Ausschluss des Fahrzeugs von der Automationslevel-Taxonomie der SAE International hier nicht vorgenommen. Die einzelnen Assistenten der S-Klasse übernehmen fast in Gänze die Aufgaben beim Fahren eines automatisierten Fahrzeugs. Deshalb wird die hier betrachtete S-Klasse von Mercedes-Benz als automatisiertes Fahrzeug des SAE-Automationslevels 3 bis 4 klassifiziert. Dadurch liegt in Form der S-Klasse ein teilautonom bis autonomer Agent vor (vgl. Abb. 4.2). Folglich liegt, formal gesehen, im Rahmen des autonomen Fahrens beim vorliegenden Praxisbeispiel ebenfalls ein HAT vor.

Auf sachlicher Ebene erfüllt das betrachtete Fahrzeug ebenfalls die Anforderungen an einen autonomen Agenten. Es kommuniziert und interagiert mit der fahrenden Person und es nimmt bei der Zusammenarbeit mit der fahrenden Person eine unterstützende, kontrollierende und sicherheitsrelevante Rolle ein. Außerdem sind sowohl das Fahrzeug als auch die fahrende Person voneinander abhängig und stimmen sich z.B. im Rahmen der Tätigkeit des Aktiven Nothalte-Assistenten genau miteinander ab. Zudem weist das Fahrzeug nach der Aktivierung seiner Assistenten durch die fahrende Person Eigeninitiative auf. Es kontrolliert sich eigenständig und kann unter bestimmten Bedingungen unabhängig vom menschlichen Teammitglied agieren. Das Verständnis des Fahrzeugs von gemeinsamen Fahraufgaben im HAT mit dem menschlichen Teammitglied liegt insofern vor, als dass das Fahrzeug durch seine Programmierung eine Art installiertes Verständnis aufweist. Das menschliche Teammitglied gibt dem Fahrzeug seine Aufgaben vor und sorgt dafür, dass es diese umsetzt. Die Voraussetzung für die Umsetzung der Aufgaben ist, dass das Fahrzeug dazu in der Lage ist. Auch die Wahrnehmung der S-Klasse als ein Teammitglied kann hier nicht bestätigt werden, da dies von der Einstellung der fahrenden Person zum automatisierten Fahrzeug abhängt. Die Erfüllung dieser sachlichen Anforderung für das Vorliegen eines autonomen Agenten und HAT wird aber angenommen, da nur fahrende Personen, die automatisierte Fahrzeuge als sicher, verlässlich und kooperativ genug betrachten, diese fahren. Aus den bereits genannten Gründen liegt sowohl aus formaler als auch aus sachlicher Perspektive, in Form des betrachteten hochautomatisierten Fahrzeugs, mindestens ein teilautonomer (bis autonomer) Agent und damit ein HAT vor.

Zweites Praxisbeispiel eines HAT im Bereich autonomer Technik in PKWs

Auch der Prozess des automatisierten Einparkens stellt eine Teamarbeit zwischen Mensch und autonomen Agenten dar und soll nun als zweites Praxisbeispiel autonomer Agenten und des HAT im Bereich autonomer Technik in PKWs erläutert werden. Beispielsweise im Rahmen des Automated Valet Parking (AVP) von S-Klasse-Fahrzeugen des Automobilherstellers Mercedes-Benz wird gegenwärtig autonomes Parken von PKWs erprobt. Dies erfolgt zusammen mit den Unternehmen Bosch und Apcoa am Flughafen in Stuttgart. Ein neueres Modell der S-Klasse von Mercedes-Benz soll das AVP durchführen. Im Rahmen des AVP soll ermöglicht werden, dass dieses Fahrzeug autonom und durch einen Befehl des Menschen über das Smartphone zu einem reservierten Parkplatz fährt. Das Fahrzeug weist eine Vorrüstung für Fahrfunktionen des SAE-Automationslevels 4 auf. Für den Test des autonomen Parkens wird das Parkhaus P6 am Flughafen Stuttgart genutzt. Für die Umsetzung des AVP ist es notwendig, dass Parkhäuser mit der für das AVP erforderlichen technischen Infrastruktur ausgestattet sind. Das Fahrzeug soll autonom im Untergeschoss des P6 parken. Beim Projekt des AVP sollen Videokameras freie Parkplätze erkennen, sowie Fahrkorridore, das Fahrzeugumfeld und Hindernisse auf der Fahrspur erfassen. Das Parkhaus besitzt eine Computerzentrale, die die Route der Fahrzeuge zu freien Parkplätzen berechnet. Das autonom parkende und fahrende Fahrzeug orientiert sich stets an den Informationen dieser Parkhauskameras. Durch das AVP werden vor allem zwei große Vorteile gegenüber dem manuellen Parken erwartet: Zum einen der erhöhte Komfort der fahrenden Personen, die beim Verlassen des Fahrzeugs beispielsweise keine Probleme mehr haben und eine Zeitersparnis für die fahrende Person. Zum Beispiel gewinnt man bei einer Verspätung durch autonome Parkvorgänge Zeit (Daimler, 2020a).

Dieses Praxisbeispiel zeigt auch die Herausforderungen, die mit HAT einhergehen. Zur erfolgreichen Umsetzung des AVP bedarf es beispielsweise der erforderlichen technischen Infrastruktur im Einsatzfeld. Die Gestaltung der für die Funktionsfähigkeit des HAT notwendigen

Umgebung ist sehr aufwendig. Außerdem zeigt dieses Praxisbeispiel, dass ein HAT einen begrenzten Einsatzbereich haben kann. Im vorliegenden Beispiel werden autonome Parkvorgänge in einem Parkhaus durchgeführt. Eine wichtige, zu klärende Frage in diesem Zusammenhang wäre, ob das vorgestellte autonome Fahrzeug auch in anderen Umgebungen, wie beispielsweise auf Straßen oder in Garagen, autonom einparken kann. Es ist naheliegend, dass das AVP nur dann funktioniert, wenn es sich bei der Umgebung um ein, den Funktionen des AVP angepasstes Parkhaus handelt. In der untenstehenden Teilabbildung 4.5a wird ersichtlich, dass sich keine fahrende Person im Fahrzeug befindet. Das Fahrzeug parkt autonom und ohne das menschliche Teammitglied im Parkhaus. Teilabbildung 4.5b verdeutlicht, dass die fahrende Person bis zum Erreichen der Parkhauseinfahrt im Fahrzeug bleibt. Bei beiden Teilabbildungen wird anhand der Beobachtung des Fahrersitzes deutlich, dass sich zwar im Parkhaus (Teilabbildung 4.5a) keine fahrende Person im Fahrzeug befindet, dies aber beim Erreichen des Parkhauses der Fall ist (Teilabbildung 4.5b).



Abb. 4.5a und 4.5b: Automated Valet Parking durch die S-Klasse von Mercedes-Benz (Fotos zur Verfügung gestellt von Daimler, 2020b).

Kontrollübernahme durch das menschliche Teammitglied ist für das AVP aber nicht unbedingt erforderlich. Sonst wäre die S-Klasse lediglich dem SAE-Automationslevel 3 zuzuordnen. Die bestimmten Bedingungen, unter denen das Fahrzeug autonom agiert, stellen hier das Pilotparkhaus dar, in dem die S-Klasse selbständig parkt. Für eine Klassifizierung als ein automatisiertes Fahrzeug des SAE-Automationslevels 5 müsste es unter allen möglichen Bedingungen autonom fahren und nicht nur parken können. Dies wird im Rahmen des AVP noch nicht erfüllt. Dadurch liegt, gemäß den Schlussfolgerungen im ersten Abschnitt dieses Unterkapitels, aus formaler Sicht ebenfalls ein HAT vor. Außerdem muss die fahrende Person das Fahrzeug zum Parkhaus bringen. Deshalb spielt die fahrende Person auch eine Rolle in der Zusammenarbeit mit dem autonom parkenden Fahrzeug. Aus sachlicher Sicht sind die in Abbildung 4.1 genannten Anforderungen an einen autonomen Agenten ebenfalls erfüllt: Das autonom parkende Fahrzeug verschafft der das Fahrzeug abstellenden Person einen zeitlichen Vorteil beim Erreichen eines möglichen Fluges und es parkt selbstständig ein. In dieser Hinsicht erfüllt es in Zusammenarbeit mit der fahrenden Person eine bestimmte, unterstützende Rolle. Das autonom parkende Fahrzeug zeigt Eigeninitiative beim autonomen Parken im Parkhaus und übt Kontrolle über sich selbst aus, indem es mit der Technologie des Parkhauses interagiert und selbstständig Informationen, beispielsweise zu freien Parkplätzen, auswertet. Die Wahrnehmung des autonom parkenden Fahrzeugs als individuelles Teammitglied durch die fahrende Person wird hier wie beim ersten Praxisbeispiel angenommen. Auch das Verständnis der gemeinsamen Fahraufgaben in Zusammenarbeit mit der fahrenden Person wird in diesem

zweiten Praxisbeispiel angenommen. Aus diesen Gründen liegen auch aus sachlicher Perspektive ein autonomer Agent und damit ein HAT vor (vgl. Abb. 4.1).

Um einen noch besseren Einblick in das Thema HAT aus der Praxisperspektive zu erhalten, wird in der folgenden Infobox 4.1 ein Interview mit einem Experten des autonomen Fahrens vom ADAC e. V. präsentiert. Im Interview werden die Automationslevel-Taxonomie des ADAC e. V. und in Bezug auf das autonome Fahren, die Aspekte Transparenz, Kontrolle und Verantwortung thematisiert. Zudem werden die Herausforderungen der praktischen Umsetzung autonomen Fahrens dargelegt.

Eine Stimme aus der Praxis zu autonomer PKW-Technologie

Infobox 4.1: Interview mit einem Experten des ADAC e. V. zum autonomen Fahren.

Wir: Als wie fortschrittlich in Bezug auf Automation lassen sich aktuell in Deutschland auf dem Markt befindende PKWs Ihrer Meinung nach bewerten?

Experte: Der ADAC testet im Rahmen des Euro NCAP Sicherheitsprogramms sämtliche verfügbaren Assistenzsysteme in aktuellen Neufahrzeugen. Die gängige Nomenklatur der SAE-Level null bis fünf des autonomen Fahrens verwenden wir nicht. Der ADAC unterscheidet drei Stufen: Assistiertes Fahren (1. Stufe), automatisiertes Fahren (2. Stufe) und autonomes Fahren (3. Stufe). Heutige Fahrzeuge bewegen sich nach unserer Einteilung maximal auf der Stufe 2, also der Stufe des automatisierten Fahrens. Wichtiger ist dem ADAC bzw. dem Euro NCAP Konsortium der Begriff der kooperativen Assistenz, einem wichtigen Testparameter, nach der die Systeme ab sofort bewertet werden. Gefragt ist eine möglichst perfekte Kooperation zwischen fahrender Person und Fahrzeug. Das heißt, dass die fahrende Person jederzeit genau weiß, welche Situation das Fahrzeug erkennt und dass die fahrende Person zudem jederzeit Gewissheit darüber haben soll, ob das Fahrzeug diese Situation auch beherrschen kann oder eben nicht. Dabei muss das Fahrzeug der fahrenden Person jederzeit erlauben, in Regelung befindliche Assistenzsysteme zu überstimmen. Das ist z. B. der Schwachpunkt beim Test des Tesla Model 3 gewesen: Die fahrende Person hatte Mühe am Lenkrad, den Lenkassistenten gegebenenfalls zu überstimmen.

Wir: Welche rechtlichen und technischen Schwierigkeiten ergeben sich aktuell bezüglich der Praktikabilität und Integration autonomer PKWs in den Straßenverkehr Deutschlands?

Experte: Rechtlich gesehen, gibt es derzeit noch keine Möglichkeit, der fahrenden Person die Kontrolle und die Verantwortung beim Fahren abzunehmen und dem Fahrzeug zu übertragen. Die inzwischen rechtlich zulässigen Autobahnpiloten sind auf eine Höchstgeschwindigkeit von 60 km/h begrenzt und daher, wie der Stauassistent, auch eher nur für den Staufolgeverkehr geeignet. Die fahrende Person muss die Kontrolle über das Verkehrsgeschehen behalten und die Fahraufgaben in kürzester Zeit übernehmen können. Die Systeme sollen sie entlasten, ohne sie von ihrer Verantwortung zu befreien.

Wir: Welche technischen Voraussetzungen der Entwicklung autonomer Fahrzeuge müssten für die Überwindung dieser Hürden in Zukunft gegeben sein?

Experte: Es müsste sichergestellt werden können, dass eine dem Verkehrsgeschehen abgewandte oder gar schlafende fahrende Person die Fahraufgabe nicht innerhalb von zehn Sekunden (wie oftmals vorgeschlagen), sondern innerhalb zwei oder drei Sekunden wieder voll übernimmt. Das scheint jedoch grundsätzlich unmöglich.

Wir: Als wie fortschrittlich schätzen Sie im Vergleich zu Deutschland die Forschungslandschaft und Praktikabilität des autonomen Fahrens in den USA ein?

Experte: In den USA scheint die Software-Kompetenz oftmals sehr viel größer. Im anwendbaren Ergebnis sind deutsche Entwickler und Forscher aber nicht wesentlich im Rückstand.

Wir: Wie beurteilen Sie den aktuellen internationalen Entwicklungsstand zum autonomen Fahren insgesamt?

Experte: Ich würde hervorheben, dass im Moment alle Hersteller ihre Anstrengungen und Investitionen zurückschrauben. Die Hersteller haben die Herausforderungen und Probleme meines Erachtens allesamt unterschätzt und verlieren jetzt das Interesse.

4.1.2 Autonome Technologie in LKWs

In diesem Unterkapitel wird ein weiteres Praxisbeispiel hochautomatisierten bzw. teilautonomen Fahrens dargestellt. Das folgende Praxisbeispiel bezieht sich aber im Gegensatz zu den beiden bisher erwähnten Praxisbeispielen auf autonome Technik in LKWs. Dabei wird auf das sogenannte LKW-Platooning eingegangen. Im Rahmen des LKW-Platoonings erfolgt im Autobahnverkehr die Abstimmung zweier LKWs miteinander. Die beiden LKWs im nächstfolgenden Praxisbeispiel und Pilotprojekt des LKW-Platoonings fahren auf der Autobahn in Kolonne und nehmen unter anderem abhängig voneinander teilautonom Bremsvorgänge und Beschleunigungsmanöver vor. Danach wird dieses Praxisbeispiel in den HAT-Kontext einsortiert und bezüglich dessen möglicher Geltung als teilautonomer bzw. autonomer Agent diskutiert. Anschließend wird beim Praxisbeispiel eine Wertung für das potenzielle Vorliegen eines HAT vorgenommen.

LKW-Platooning

Im Rahmen des EDDI-Projekts (Elektronische Deichsel-Digitale Innovation) wurde weltweit das erste Mal das LKW-Platooning im öffentlichen Straßenverkehr erprobt. Dies geschah auf dem digitalen Testfeld der Autobahn A9 zwischen München und Nürnberg. Bei diesem Projekt kooperierten die Unternehmen MAN Truck & Bus, DB Schenker und die Hochschule Fresenius miteinander. Beim LKW-Platooning sind zwei elektronisch gekoppelte LKWs auf einer Autobahnstrecke unter realen Verkehrsbedingungen über eine längere Zeit gefahren. In beiden LKWs hat man fahrende Personen positioniert. Der Abstand zwischen den beiden LKWs ist während des LKW-Platoonings auf der Autobahn konstant bei 15 Metern gehalten worden. An Autobahnkreuzen, vor Autobahnbaustellen, bei Steigungen und Gefälle von über vier Prozent, Unfallstellen und besonders dichtem Verkehr musste das LKW-Platooning angesichts rechtlicher Regelungen unterbrochen werden. Die Höchstgeschwindigkeit ist durch diese Regelungen auf 80 Stundenkilometer festgelegt worden (Brandt et al., n. d.).

Dabei gibt der vorausfahrende LKW dem ihm folgenden LKW die Geschwindigkeit und Richtung vor. Durch den Windschatteneffekt weist der folgende LKW eine effizientere Fahrweise auf. Außerdem wird die Sicherheit beim Fahren der LKWs erhöht, indem beide elektronisch gekoppelten LKWs wie eine Einheit zusammen reagieren. Die LKWs nehmen also in der Zusammenarbeit mit den fahrenden Personen eine unterstützende und auch sicherheitsbezogene Rolle ein (vgl. Abb. 4.1). Die beiden LKWs sind beide identisch aufgebaut und können jeweils die Rolle des Führungsfahrzeugs oder des Folgefahrzeugs wahrnehmen. Die Ausrüstung beider LKWs setzt sich aus Assistenz- und Sicherheitssystemen zusammen. Zum Beispiel der abstandsregelnde Tempomat ACC und das Notbremssystem EBA sind in beiden LKWs verbaut. Des Weiteren wurden in den beiden LKWs jeweils Radar- und Kamerasensoren, sowie ein Lidar-Sensor (Light Detection and Range) verbaut. Der Lidar-Sensor kann durch seinen großen Öffnungswinkel potenziell einsicherende Fahrzeuge früh erkennen. Die Radar- und Kamerasensoren steuern die Abstandsregelung. Ein elektrisch ansteuerbares Lenksystem wird für die Querführung in beiden LKWs eingesetzt. Zusätzlich sind für das LKW-Platooning ein Bremssystem und ein Notbremsassistent angepasst worden. Zudem verwaltet ein Sicherheitssteuergerät die Beschleunigungs- und Verzögerungsanforderungen und interagiert mit allen Komponenten der LKWs. Nur wenn alle Komponenten der LKWs korrekt arbeiten, gibt das Sicherheitssteuergerät den Betrieb des LKW-Platoonings frei (Brandt et al., n. d.). In dieser Hinsicht kontrollieren sich die beiden LKWs selber (vgl. Abb. 4.1). Dadurch wird sichergestellt, dass der LKW auch bei einem Ausfall einzelner Komponenten sicher bremsen kann. Beide LKWs weisen zudem ein Voll-TFT-Display auf. Dieses zeigt den fahrenden Personen den aktuellen Status des LKW-Platoonings an. An den Lenkrädern der LKWs sind Bordcomputer-, Tempomatbedienungsfunktionen und Aktivierungshilfen für die Bedienung des LKW-Platoonings installiert worden (Brandt et al., n. d.). Das menschliche Teammitglied muss dem LKW also zunächst die Freigabe erteilen, damit dieser selbständig agieren kann. In dieser Hinsicht stehen die fahrenden Personen und die beiden LKWs jeweils in einem Abhängigkeitsverhältnis zueinander und interagieren, bzw. kommunizieren miteinander (vgl. Abb. 4.1). Das LKW-Platooning kann die fahrende Person aber übersteuern, sofern diese die entsprechenden Assistenten aktiviert. Da die beiden LKWs elektronisch gekoppelt fahren und als Einheit reagieren, bremst der folgende LKW, sofern der vorausfahrende LKW bremst, zum selben Zeitpunkt wie dieser. Der abstandsregelnde Tempomat ACC sorgt bei den zwei LKWs im Platooning dafür, dass ein Mindestabstand von 50 Metern zum vorausfahrenden Verkehr eingehalten wird. Zudem reduziert der ACC bei Bedarf die Geschwindigkeit der LKWs. Bei drohenden Kollisionen zu vorausfahrenden Fahrzeugen bremst das Notbremssystem der LKWs diese bis zum Stillstand (Brandt et al., n. d.). Die beiden LKWs können also unabhängig von ihrem jeweiligen menschlichen Teammitglied, der fahrenden Person, handeln (vgl. Abb. 4.1). Das LKW-Platooning sorgt alles in allem für eine Entlastung der fahrenden Personen und nicht für den Ersatz dieser. Die beiden LKWs nehmen also ebenfalls eine entlastende Funktion ein. Lediglich die Aufgaben der LKW-Fahrer/Innen werden sich angesichts des LKW-Platoonings verändern. LKW-Platooning kann als Vorstufe autonomen Fahrens angesehen werden (Brandt et al., n. d.). In der folgenden Abbildung 4.6 werden die zwei LKWs dargestellt, die miteinander auf der Autobahn in Kolonne fahren. Die Teilabbildung 4.6a zeigt das LKW-Platooning im realen Straßenverkehr auf der Autobahn. Die Teilabbildung 4.6b stellt die bereits oben beschriebenen Fahrzeugsysteme der LKWs im Platooning dar.

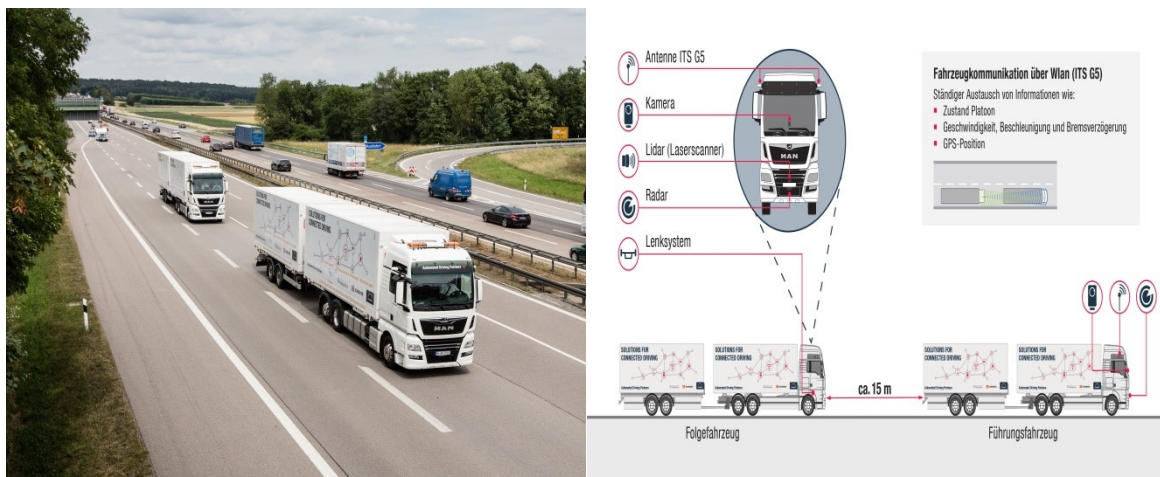


Abb. 4.6a und 4.6b: HAT beim LKW-Platooning (Fotos zur Verfügung gestellt von der Deutsche Bahn AG & MAN; Foto: Jan Hölzl (Abb. 4.6a) & von der Deutsche Bahn AG & DB Schenker (Abb. 4.6b), 2018).

Die beiden am LKW-Platooning Pilotprojekt beteiligten LKWs werden als teilautonom betrachtet, da diese zwar selbständiger und weniger mit Hilfe der fahrenden Person (vgl. Abb. 4.2) fahren können, aber diese dennoch zur Überwachung und Aktivierung ihrer Platooning- und Assistenzsysteme benötigen. Da die bereits erwähnten Assistenzsysteme der beiden LKWs zudem viele Fahraufgaben übernehmen können, werden diese Systeme hier im Kollektiv betrachtet, nicht von der Automationslevel-Taxonomie der SAE International ausgeschlossen. Die beiden LKWs werden folglich als Fahrzeuge der SAE-Automationslevel 3 bis 4 bewertet. Dadurch liegen aus formaler Perspektive gemäß den Schlussfolgerungen des ersten Abschnitts dieses Unterkapitels zwei teilautonome Agenten vor (siehe Abb. 4.2). Aus sachlicher Perspektive erfüllen die beiden LKWs im Platooning auch die für die Geltung als teilautonome Agenten genannten Anforderungen: Sie entlasten die LKW-Fahrer/innen von deren Fahraufgaben und übernehmen damit eine unterstützende Rolle. Außerdem können beide LKWs, sowohl als vorausfahrendes, als auch als folgendes Fahrzeug agieren und besitzen jeweils beide zwei zusätzliche Rollen in der Interaktion des LKW-Platoonings. Durch die bereits oben angeführten Assistenzsysteme können die beiden LKWs zudem Eigeninitiative ausüben. Durch das Sicherheitssteuergerät werden die Platooning- und Assistenzsystem-Funktionen von den LKWs selbst kontrolliert. Die Interaktion und Kommunikation der beiden LKWs mit den fahrenden Personen erfolgt unter anderem auch durch die Anzeigen auf dem Display in der Fahrkabine des LKWs. Die fahrenden Personen im Pilotprojekt nahmen die Funktionen des LKW-Platoonings als hilfreich wahr. Also ist hier auch die Wahrnehmung der teilautonomen Agenten als vollwertiges Teammitglied durch die fahrenden Personen gegeben. Dadurch liegen in Form der beiden teilautonomen LKWs zwei teilautonome Agenten vor, sowohl aus formaler als auch aus sachlicher Perspektive. Damit sind aus beiden Perspektiven die Anforderungen für das Vorliegen eines HAT bei diesem Praxisbeispiel gegeben und erfüllt. Das menschliche Teammitglied übernimmt immer noch eine zentrale Funktion und Rolle. Es aktiviert die teilautonomen Agenten im LKW-Platooning und überwacht diese. Zudem kann sie diese ausschalten. Die gemeinsame Aufgabe und Interdependenz beim vorliegenden HAT ist das sichere und effiziente Fahren der beiden LKWs im LKW-Platooning. Die beiden teilautonomen Agenten weisen eine geringere Reaktionszeit auf und erhöhen bei Bremsmanövern die Sicherheit. Die fahrenden Personen überwachen die beiden LKWs und tragen auf diese Weise ebenfalls zur erhöhten Sicherheit des LKW-Platoonings bei. Sowohl die beiden teilautonomen Agenten, als

auch die beiden fahrenden Personen sorgen folglich im Rahmen des LKW-Platoonings jeweils für die Bewältigung dieser gemeinsamen Aufgaben. Für einen tiefergehenden Einblick in das Themengebiet des autonomen Fahrens von LKWs wird in der folgenden Infobox 4.2 ein Interview mit einem Experten von DB Schenker zu autonomer LKW-Technik vorgestellt. In diesem werden unter anderem der aktuelle Stand autonomer LKW-Technik, sowie die gesetzlichen und technischen Herausforderungen autonomen Fahrens mit Hilfe autonomer LKW-Technik diskutiert.

Eine Stimme aus der Praxis zu autonomer LKW-Technologie

Infobox 4.2: Interview mit Herrn Erik Wirsing, dem Vice President Global Innovation bei DB Schenker, zu autonomer LKW-Technologie

Wir: Als wie fortschrittlich, in Bezug auf Automation, lassen sich die von DB Schenker aktuell verwendeten LKWs im europäischen Landverkehr allgemein Ihrer Meinung nach bewerten?

Herr Wirsing: Die LKWs, die wir bei DB Schenker gegenwärtig verwenden, lassen sich nicht als autonome Fahrzeuge bezeichnen. Gegenwärtig verwenden wir für unsere LKWs Automatisierungshilfen und Assistenzsysteme der Automationsstufe 2. Die LKWs weisen normale Verbrennungsmotoren oder Elektrotechnik auf. Meiner Ansicht nach, nimmt der Grad der Automatisierung stetig zu. Außerhalb unserer Lager und im öffentlichen Verkehr steht uns noch die Gesetzgebung und Rechtsprechung bei der Umsetzung hochautomatisierter und autonomer Lösungen im Weg. Die Technik ist ebenfalls noch nicht weit genug.

Wir: Welche wesentlichen Funktionen sollte ein autonomer LKW aufweisen und welche Tätigkeiten sollte dieser ausführen können?

Herr Wirsing: Das Oberthema schlichtweg ist die Sicherheit der autonomen Fahrzeuge. Und weil gegenwärtig bei der Erprobung autonomen Fahrens immer noch Unfälle passieren, wird Sicherheit weiterhin einer der wichtigsten Faktoren beim autonomen Fahren bleiben. Des Weiteren sollte ein autonomer LKW eigenständig navigieren, sich selbst überwachen und steuern können. Außerdem sollte er eine Art Schwarmintelligenz aufweisen, bei der wie beim Platooning mehrere LKWs miteinander kommunizieren können. Ein autonomer LKW der Automationsstufe 5 sollte zudem ohne fahrende Person und Führerhaus fahren. Autonome LKWs sollten auch eigenständig orchestrieren können, also autonom an Laderampen heranfahren und ihre Türen z.B. zum Entladen öffnen können.

Wir: Welche Aufgaben müssten Fahrer/innen autonomer LKWs bei DB Schenker voraussichtlich in Zusammenarbeit mit der autonomen Technik noch ausführen und überwachen?

Herr Wirsing: Fahrer/innen autonomer LKWs der Automationsstufe 5 werden voraussichtlich keine aktiven Fahraufgaben mehr übernehmen. Sie werden eher dafür verantwortlich sein, den LKW zu überwachen, um bei Störfällen eingreifen zu können. Das Potenzial der Automatisierung besteht hierbei darin, dass LKW-Fahrer/innen durch autonome LKWs ihre Arbeitspausen direkt im LKW verbringen können, während der LKW selbständig weiterfährt. Dadurch würde es insgesamt weniger Fahrer/Innen und LKWs geben, da LKWs autonom länger fahren werden. LKW-Fahrer/innen werden in Zukunft zu Fuhrpark- und Truck-Managern/innen. Sie werden voraussichtlich die Ladungssicherung übernehmen und angesichts der Entwicklung von Platooning-Techniken mehrere LKWs gleichzeitig überwachen.

Wir: Welche rechtlichen und technischen Schwierigkeiten ergeben sich aktuell bezüglich der Praktikabilität und Integration autonomer LKWs in den Gütertransportverkehr Deutschlands?

Herr Wirsing: Technische Schwierigkeiten bezüglich des autonomen Fahrens verorte ich vor allem in komplexeren Verkehrsbereichen wie z.B. Innenstädten oder Entladungszonen. Dort müsste ein autonomer LKW sehr genau rangieren und fahren können, um sehr enge Fahrstrecken bewältigen zu können. Alles in allem muss autonomes Fahren also alle Verkehrsbereiche und Fahrstrecken meistern, um als technisch ausgereift gelten zu können. Rechtliche Schwierigkeiten liegen bezüglich großer Sicherheitsbedenken vor. Der Gesetzgeber wird hier meines Erachtens eine graduelle Einführung autonomen Fahrens befürworten, sodass erst Fahrten auf der Autobahn, dann auf der Landstraße und letzten Endes in der Innenstadt und engen Strecken rechtlich und auch technisch möglich sein werden.

Wir: Wie könnte man den Einsatz des LKW-Platoonings im Straßenverkehr ausweiten?

Herr Wirsing: Für einen flächendeckenden Einsatz des Platoonings könnte man gemeinsam mit anderen Unternehmen eine überlagerte Instanz schaffen und ebenfalls ein ganzheitliches Überwachungssystem für autonome Fahrzeuge einführen. Im Vergleich zum Schienenverkehr, wo Fahrzeuge direkt in einem System erfasst werden, oder im Vergleich zum System der Deutschen Flugsicherung ist der Landverkehr noch nicht in einem System erfasst.

4.1.3 Praxisbeispiele von HAT im öffentlichen Verkehr und Gütertransport

Autonomes Fahren und HAT sind auch bereits im öffentlichen Nah- und Fernverkehr vertreten. Es finden sich weltweit einige Projekte zum Thema autonomes und vollautomatisiertes Fahren in Bezug auf Busse, Metro- oder U-Bahnsysteme und Züge.

In Bezug auf die folgenden Praxisbeispiele von HAT aus dem öffentlichen Busverkehr werden fünf Automationsstufen unterschieden. Es ist davon auszugehen, dass HAT in den Stufen vom teilautomatisierten bis zum vollautomatisierten Fahren vorliegt. Zu vergleichen ist die untenstehende Abbildung 4.7 mit der im ersten Abschnitt dieses Unterkapitels thematisierten Automationslevel-Taxonomie der SAE International (vgl. Abb. 4.2).

Assistiertes Fahren	Teilautomatisiertes Fahren	Hochautomatisiertes Fahren	Vollautomatisiertes Fahren	Autonomes Fahren
<ul style="list-style-type: none"> •Fahrende Person steuert das Fahrzeug und muss den Blick auf die Fahrbahn behalten •Unterstützung ist punktuell 	<ul style="list-style-type: none"> •Fahrzeug für Aufgaben selbstständig durch •Fahrende Person hat dauerhaft die Kontrolle •Systeme arbeiten zusammen 	<ul style="list-style-type: none"> •Fahrende Person kann sich teilweise vom Verkehr abwenden •System fordert nach kurzer Zeit wieder zu einer Übernahme auf 	<ul style="list-style-type: none"> •Fahrende Person wird zum Passagier •System erkennt eigene Grenzen und fordert die fahrende Person zum Eingreifen auf 	<ul style="list-style-type: none"> •Es wird keine fahrende Person benötigt •Fahren ohne Passagiere möglich •System bewältigt alle Verkehrssituationen

Abb. 4.7: Fünf Stufen des autonomen Fahrens (in Anlehnung an Arnold, 2020).

Im automatisierten Busverkehr ist der „Mercedes-Benz Future Bus“ ein anschauliches Beispiel für HAT. Dieser Bus ist teilautomatisiert und verfügt über die Technik des „CityPilot“, die

Ampeln erkennt und mit ihnen kommuniziert, damit der Bus schließlich sicher über die Ampelkreuzung fahren kann. Außerdem erkennt das System Hindernisse und Fußgänger auf der Fahrbahn und bremst vor der Kollision mit diesen selbstständig (Daimler, n. d.b). Die erste Fahrt auf öffentlichen Straßen absolvierte der Future Bus in den Niederlanden auf einer ca. 20 Kilometer langen Strecke mit mehreren Kurven, Tunneln und Ampelkreuzungen. Während der Fahrt befindet sich eine fahrende Person im Fahrzeug, um in das Geschehen eingreifen zu können und es zu überwachen. Zu beachten ist, dass die fahrende Person weder das Gaspedal, noch die Bremse betätigen muss und nur bei Gegenverkehr, gemäß der Verkehrsvorschriften, von dem Fahrzeug aufgefordert wird, das Lenkrad zu übernehmen (Daimler, n. d.b). Dies ist der Grund für die Klassifizierung des Fahrzeuges in die Stufe 2, denn das System wird dauerhaft von der fahrenden Person überwacht und es kann die Führung nur in spezifischen Anwendungsfällen übernehmen (vgl. Abb. 4.7; Verband der Automobilindustrie, 2015). An die Haltestellen auf der Strecke kann der Future Bus selbstständig heranfahren und auch die Türen öffnen und schließen sich ohne die Hilfe der fahrenden Person. Diese wird immer über die Handlungen und Vorgänge des Busses durch ein Display zwischen Lenkrad und Frontscheibe informiert oder aufgefordert, die Steuerung zu übernehmen. Es bedarf keinerlei Bestätigung durch den Menschen. Mit der Einordnung in den Bereich des Automationslevels 2 (vgl. Abb. 4.7), handelt es sich bei dem Busmodell formal um einen autonomen Agenten (vgl. Abb. 4.1) und es kann von einem HAT gesprochen werden. Auch aus sachlicher Perspektive liegen hier sowohl ein autonomer Agent, als auch HAT vor, da die Begleiter und die Technik des Busses unter anderem ein gemeinsames Ziel verfolgen und eine Kommunikation zwischen den beiden Parteien über das Lenkraddisplay stattfindet. Beide Teammitglieder stehen in einem Abhängigkeitsverhältnis. Der autonome Agent kann beim Heranfahren an Haltestellen unabhängig von der fahrenden Person agieren. Die Wahrnehmung des Busses als individuelles Teammitglied durch die fahrende Person, sowie das Verständnis des Busses von den gemeinsamen Aufgaben im Team wird hier, wie bei den bisherigen Praxisbeispielen, angenommen (vgl. 4.1). Beide Akteure bewältigen im Team gemeinsam Fahraufgaben und besitzen eine zentrale Rolle in diesem HAT. Die Abbildungen 4.8a und 4.8b sind Aufnahmen von dem Mercedes-Benz Future Bus und zeigen diesen von innen und außen. In 4.8a sieht man das Fahrercockpit des Busses und erkennt, dass sich die Hände der fahrenden Person nicht am Lenkrad befinden oder das Gaspedal betätigt wird. Der Bus fährt in dieser Situation also autonom. Beide Abbildungen lassen das Design des Fahrzeugs und die fortschrittliche technische Ausstattung aus verschiedenen Perspektiven erkennen.



Abb. 4.8a und 4.8b: Autonomer Stadtbuss (Fotos zur Verfügung gestellt von Daimler, 2016).

Über den Future Bus hinaus, werden auf den öffentlichen Straßen weltweit auch andere erste hochautomatisierte Busse erprobt, wie z.B. in Las Vegas. Dort gibt es autonome Shuttle Busse, welche sich teilweise von einer sich im Fahrzeug befindenden Person per

Spielkonsolen-Controller steuern lassen. Die Busse fahren und bremsen selbstständig, müssen aber bei Gegenverkehr von der Begleitperson gesteuert werden und können noch keine Hindernisse umfahren. Die Busse können die Hindernisse auf ihrem Weg nicht bewerten und stoppen aus Sicherheitsgründen sofort (Hansen & Janssen, 2018). Die Shuttle Busse sind demnach dem Automationslevel 3 (vgl. Abb. 4.7) zuzuordnen und folglich handelt es sich hier formal ebenfalls um ein HAT. Auch aus sachlicher Perspektive (vgl. Abb. 4.1) liegt in Form des Shuttlebusses ein autonomer Agent und damit ein HAT vor. Der Shuttlebus ist von seinem menschlichen Teammitglied abhängig und benötigt bei Fahrsituationen, in denen Verkehr auf der Gegenfahrbahn herrscht, sein menschliches Teammitglied zur Unterstützung. Somit stehen beide Teammitglieder in einem Abhängigkeitsverhältnis und bewältigen gemeinsam Fahraufgaben. Auf der anderen Seite kann der autonome Agent unabhängig von der fahrenden Person handeln. Er bremst und fährt selbstständig. Dadurch, dass die Begleitperson den Bus bei Gegenverkehr über den Controller steuert, interagiert sie mit ihm (vgl. Abb. 4.1).

Anders ist es bei den Metro- und U-Bahnsystemen. Seit 2008 gibt es in Nürnberg, Deutschland, die erste autonome U-Bahn. Die U3 ist eine selbstfahrende U-Bahn mit einer Gesamtfahrstrecke von 20 Kilometern. Von den drei existierenden U-Bahn Linien sind seit 2010 zwei vollautomatisiert. Selbst wenn die Fahrerkabine leer sind oder sogar nicht mehr vorhanden, kann man bei dieser Art automatisierten Fahrens von einem HAT sprechen. Das fahrerlose System funktioniert im Zusammenspiel mit dem Menschen und der im Fahrzeug und entlang der Strecke verbauten Technik, die den autonomen Agenten widerspiegeln. Durch Monitoring bringt sich der Mensch in das Geschehen mit ein. In einer Leitstelle der Verkehrs-Aktiengesellschaft (VAG) in Nürnberg können Mitarbeitende überwachen, eingreifen oder die aktuelle Handlung unterbrechen. Den Rest der Zeit fährt und entscheidet die Bahn selbstständig. Die autonome U-Bahn verfügt über intelligente Türen und ein umfangreiches System von Radarmeldern und Kameras, welches die Sicherheit am Bahnsteig gewährleistet (Allianz pro Schiene e.V., 2016). Die autonome U-Bahn fährt allerdings ausschließlich in einem geschlossenen System. Es gibt keine Weichen, Abbiege- und Spurwechsellmöglichkeiten oder Kreuzungen. Das heißt, dass die Implementierung in den Zug- und Güterverkehr mit deutlich komplexerer Technik verbunden ist. Das autonome System muss auf Hindernisse, sowie umgefallene Bäume reagieren und schlechter Witterung, Starkregen, Schnee und Eis trotzen (Doll, 2018). Auch ohne die Klassifizierung in ein Automationslevel kann bei diesem Beispiel von HAT gesprochen werden. Das menschliche Teammitglied überwacht hierbei den autonomen Agenten. Es kann die Handlungen des autonomen Agenten unterbrechen und wirkt dadurch direkt auf diesen ein. Beide Teammitglieder stehen in einem Abhängigkeitsverhältnis. Der autonome Agent führt den Großteil seiner Aufgaben unabhängig vom menschlichen Teammitglied aus und nimmt dadurch eine zentrale Rolle im Team ein. Er besitzt die Fähigkeit, sich selbst zu kontrollieren und zeigt Eigeninitiative. Gemäß den erwähnten Aspekten, wird beim vorliegenden Praxisbeispiel aus sachlicher Perspektive von einem autonomen Agenten gesprochen (vgl. Abb. 4.1) und folglich auch von HAT.

In Deutschland gibt es noch keine Beispiele für autonomen Personenzug- oder Güterverkehr (Stand 2020). In Australien kann es aufgrund der trockenen Wetterlage und der teilweise dünn besiedelten Gegenden bereits solche Güterzüge geben. Sie fahren z.B. in der Pilbara Region von Western Australia und bilden den weltweit größten Roboter und das erste automatisierte Langstrecken-Schienennetz (Briginshaw, 2018). Dieses Eisenerzbahn Netzwerk umfasst derzeit insgesamt 50 Güterzüge, welche vollautomatisiert fahren und mit der „AutoHaul-Technik“ ausgestattet sind. Diese Technik beinhaltet z.B. ein Zugschutzsystem und ein Kollisionserkennungssystem. Um die öffentliche Sicherheit zu verbessern gibt es ein Hinderniserkennungs-

system. Alle empfangenen und aufgenommenen Daten werden über die Signalisierungen an den Übergangspunkten und entlang der Hauptstrecke visualisiert (Buchan, n. d.). Das System benötigt zwei separate Kanäle, von denen sich einer im Kontrollcenter in Perth und der andere an Bord des Güterzuges befindet. Diese sind miteinander verlinkt und kommunizieren mit Hilfe des Menschen miteinander (Smith, 2019). Das intelligente System und der Mensch im Kontrollcenter treffen zusammen die Entscheidungen, die sonst die fahrende Person allein treffen würde. HAT findet der Definition nach über die AutoHaul-Technik statt. In diesem Praxisbeispiel erfolgt die Übertragung des HAT-Konzepts ebenfalls lediglich aus sachlicher Perspektive. Mit Hilfe der AutoHaul-Technik nimmt der autonome Güterzug mit dem menschlichen Teammitgliedern im Kontrollcenter eine zentrale Rolle ein. In diesem Center überwachen die menschlichen Teammitglieder den autonomen Agenten, wodurch sie mit diesem in einem Abhängigkeitsverhältnis stehen. Der autonome Güterzug bewältigt mit der AutoHaul-Technik den Großteil der Fahraufgaben unabhängig von den menschlichen Teammitgliedern und zeigt Eigeninitiative. Über die AutoHaul-Technik interagiert der autonome Güterzug mit seinen menschlichen Teammitgliedern im Kontrollcenter und kann als aktives Teammitglied wahrgenommen werden. Aus den genannten Gründen liegt in Form des autonomen Güterzuges aus sachlicher Perspektive ein autonomer Agent vor. Sie arbeiten im Kollektiv auf das gleiche Ziel hin (vgl. Abb. 4.1), weshalb ebenfalls von HAT ausgegangen werden kann.

4.2 Human Autonomy Teaming im Militär

Ein weiteres Anwendungsgebiet, in dem intensiv an HAT geforscht wird und in dem dieses zukünftig stark vertreten sein soll, stellt der Militärbereich dar. Es folgen zunächst Praxisbeispiele speziell zu unbemannten Luftfahrzeugen, die als autonome Agenten, im Zusammenspiel mit den Soldaten des Militärs, ein HAT bilden. Im Militärkontext nennt man diese Art von Zusammenarbeit Manned-Unmanned Teaming (MUM-T). Das folgende Unterkapitel beschäftigt sich mit der Erklärung dieses Themas und einer Schlussfolgerung von MUM-T auf HAT.

4.2.1 Autonome unbemannte Flug- und Landfahrzeuge

Die moderne Militärtechnik setzt auf unbemannte Fahrzeuge. Während diese zurzeit noch stark abhängig von menschlicher Bedienungseingabe sind, wird von der Bundeswehruniversität München an neuen intelligenten Agenten geforscht, welche Aufgaben (teil-) autonom ausführen sollen. Dies kann eine Vielzahl von Aufgaben sein, die den menschlichen Bedienenden entlasten, bei der Missionsplanung und -ausführung helfen und dessen persönliches Risiko verringern sollen. Das adaptive, assoziierte System bewertet kontinuierlich die Leistung des Bedienenden und überwacht die Missionsumgebung autonom. Es werden automatisch Unterstützungsstrategien abgeleitet, welche auf diesen Ergebnissen basieren (Bundeswehruniversität München, n. d.).

Die einzelnen unbemannten Fahrzeuge nehmen eine Position in der Technik des Manned-Unmanned Teaming ein. Bei den Luftstreitkräften findet MUM-T besondere Aufmerksamkeit. Die inhärenten Stärken der bemannten Systeme werden in diesem Konzept mit den Stärken der unbemannten Luftfahrzeuge (UAS) kombiniert. MUM-T ist eine Kombination aus zu Fuß operierenden Soldaten, bemannten/unbemannten Fahrzeugen, Robotik und Sensorik, die das Lagebewusstsein verbessert, sowie die Kampfkraft und Überlebensfähigkeit erhöht. Die Sensorabdeckung wird bei richtiger Konzeption durch MUM-T erweitert und die Erfassung, sowie Bekämpfung von Zielen zusätzlich ermöglicht (Department of Defence, 2014). Mit der Beschreibung des MUM-T kann folglich auf ein HAT geschlossen werden. Um die Rolle der unbemannten Luftfahrzeuge als (teil-) autonome Agenten und die des Menschen in diesem

Zusammenhang besser zu verstehen, schildert das Positionspaper des Amts für Heeresentwicklung der Bundeswehr folgendes Szenario: Ein Bataillon der Bundeswehr befindet sich in einem Auslandseinsatz und schützt ein Umspannwerk. Drohnen attackieren die Soldaten und Soldatinnen und ihre Netzwerke werden von Hacker/innen angegriffen. Daraufhin fällt die Entscheidung, die Kräfte des taktischen unbemannten Luftsystems (TauUAS) einzusetzen. Vier Züge mit je 5.000 Drohnen, alle unterschiedlicher Bauart und Größe, werden aus mehreren Transportflugzeugen gestartet (Pletsch, 2020; Amt für Heeresentwicklung, 2019). Die Drohnen formieren sich zu mehreren Schwärmen, welchen jeweils eine bestimmte Aufgabe zugeteilt wurde. Ein Schwarm besteht aus hunderten Sensor-UAS, welche mit hochauflösenden Kameras ausgestattet sind und über eine enorme Reichweite verfügen. Andere Schwärme sollen gegnerische Drohnen stören oder dienen als Relais zur Kommunikation der eigenen unbemannten Luftfahrzeuge. Wieder andere sind in der Lage, eine verlegfähige UAS-Sperre zu bilden, da sie mit kleinsten Wirkmitteln ausgestattet sind, die einen Angriff auf gegnerische Sensorik und die Markierung bzw. Verfolgung von Zielen ermöglichen. Mit Hilfe von KI werden die gegnerischen Kräfte einer Gefechtsgliederung zugeordnet. Dies erfolgt hochautomatisiert (Amt für Heeresentwicklung, 2019).

Das Szenario würde zur Koordination der hohen Anzahl an unbemannten Luftfahrzeugen eigentlich einen enormen Personalaufwand erfordern, weshalb die Bundeswehr umfangreich in die Technologie der KI investiert. So soll es möglich sein, dass die Drohnen nicht dauerhaft mit einer Fernbedienung gesteuert, sondern lediglich Zielgebiete und Aufgaben für die Schwärme definiert werden müssen. Um das zuverlässige automatische Ausführen von Teilaufgaben zu gewährleisten, ist es erforderlich, dass die KI den OODA-Zyklus, den Prozess der Entscheidungsfindung und -umsetzung bei der Führung der UAS sicherstellt (Amt für Heeresentwicklung, 2019). Dieser besteht aus den vier Phasen Observation (Beobachtung), Orientation (Orientierung), Definition und Action (Aktion). Die Eigenschaft der Anwendung dieses Zyklus macht beispielsweise aus dem UAS einen autonomen Agenten im Sinne der Definition (vgl. Kapitel 1.1.2). Der autonome Agent stellt die Verbundwirkung zwischen Schwarm und MUM-T her. Es wird deutlich, dass die unbemannten Luftfahrzeuge mit den menschlichen Teammitgliedern zusammenarbeiten und Unterstützung bieten. Die Teammitglieder interagieren und kommunizieren miteinander, sodass das gemeinsame Ziel eines Missionserfolges erreicht werden kann.

Future Combat Air System

Das „Future Combat Air System“ (FCAS) ist eine weitere Form des MUM-T. Das Unternehmen AIRBUS stellt in Verbindung mit dem zukünftigen Luftkampfsystem, dem FCAS, und einem europäischen Rüstungsprojekt seine neue Technik vor. Das Unternehmen entwickelt begleitend für den im Fokus stehenden „Next Generation Fighter“, einem bemannten Jet, automatisierte Drohnen, die das Konzept des MUM-T unterstützen. Die Drohnen operieren als „Remote Carrier“, die zwar als autonom gelten, jedoch für bestimmte Manöver die Freigabe von menschlichem Personal erhalten sollen (Bundesregierung, 2018). Geplant ist eine „Air Combat Cloud“ (ACC), welche die Teaming- Informationen für einen schnelleren kooperativen Kampf bereitstellt und die Verbindung zwischen bemannter und unbemannter Plattform herstellt (Airbus Defence, 2020). Die ACC ermöglicht so die Kommunikation mit den menschlichen Teammitgliedern. Mit Hilfe der ACC sollen Daten zukünftig in Echtzeit von Boden und Flugkörpern bis hin zu Satelliten miteinander vernetzt werden können (WELT, 2020). So soll ein „Multi-Domain Betrieb“ und synchronisierte Operationen über Luft, Land, Meer, Weltraum und Internet hinweg möglich sein (Bundeswehruniversität München, n. d.). Die ACC ist hier als autonomer Agent zu verstehen, der im MUT-T mit dem Menschen kommuniziert. Bei dem Future

Combat Air System handelt es sich um ein Beispiel für HAT, da der autonome Agent, im Team mit seinen menschlichen Teammitgliedern eine zentrale, unterstützende Rolle besetzen soll. Er soll autonom Verbindungen herstellen und während einer Militäroperation Informationen mit seinen Teammitgliedern teilen können. Der autonome Agent kann unabhängig vom menschlichen Teammitglied agieren, aber steht dennoch in einem Abhängigkeitsverhältnis zum Menschen. Die ACC weist im Rahmen der genannten Vorgänge Eigeninitiative auf und kann sich selbst kontrollieren. Es ist davon auszugehen, dass die automatisierten Drohnen als individuelle Teammitglieder angesehen werden (vgl. Abb. 4.1).

Mission Master

Auf dem Boden testen die Streitkräfte z.B. das autonome unbemannte Landfahrzeug (A-UGV) namens „Mission Master“, hergestellt von der Rheinmetall AG Unternehmensbereich Defence (Verteidigung). Der Mission Master zeichnet sich durch seine vielseitige Einsetzbarkeit in verschiedenen Situationen aus. Speziell ist, dass er eine skalierbare Autonomie besitzt. Es kann per Funk ferngesteuert werden, im Follow-Me-Modus, also dem bedienenden Soldaten selbstständig folgend oder autonom agieren. Dies macht den Mission Master zu einem zuverlässigen Teammitglied der Einheit. Die niederländischen Streitkräfte, welche den Mission Master im November 2020 testeten, benutzten die KI, um ihre Kampfkraft zu steigern und die sogenannten 3D-Aufträge, „dull, dirty, dangerous“ bedeutend langweilig, dreckig, gefährlich, sicher auszuführen (Heiming, 2020). Bei Fernsteuerung besitzt der Bedienende jederzeit die Kontrolle über jegliche Funktion des Systems, also liegt die Klassifizierung Human-in-the-Loop vor (vgl. Kapitel 3.3.3). Im teilautonomen Betrieb benötigt das System ein Lagebild aus Sensorinformationen, mit Hilfe dessen es Teilaufgaben autonom erledigen kann. Alle Aktionen müssen dabei vom Bediener freigegeben werden. Dabei befindet sich das menschliche Teammitglied „on-the-Loop“. In dem vollautonomen Betrieb müssen die Aktionen des Systems programmiert werden. Dazu gehören Vorgaben für die Nutzung der Teilsysteme und die Bewegung. Nach dem Start kann der Mission Master ohne weiteren Eingriff der bedienenden Person agieren, wobei von einem Human-out-of-the-Loop-System gesprochen wird. Dabei liegt die Verantwortung des menschlichen Bedienenden in der Definition der Aufgaben und den Randbedingungen sowie Start, Überwachung und Beendigung der Mission (Heiming, 2020). Falls nötig, kann der Ablauf vom Bediener unterbrochen werden (Monroy, 2019). Die autonomen Funktionen verdankt der Mission Master der „PATH Autonomie-Kit“-Technologie von Rheinmetall Defence (Rheinmetall AG, 2020). In einem Vorstellungsvideo der Hersteller werden z.B. die Verwendungsmöglichkeiten des „Mission Master-Armed Reconnaissance“ innerhalb eines Einsatzes gezeigt. Dieser Mission Master wurde für gefährliche Spähaufklärungseinsätze entwickelt und ist zusätzlich mit einer fernbedienbaren Waffenstation ausgestattet. Er besitzt hochmoderne Sensorik und kann ein Lagebild in Echtzeit vermitteln. Damit soll er das Team z.B. bei der Erkundung vermutlich leerstehender Gebäude oder besetzter, bzw. unbekannter Gebiete unterstützen (Rheinmetall, n. d.).

Die entscheidenden Merkmale, welche der Mission Master erfüllt, damit ein HAT aus sachlicher Perspektive angenommen wird, sind, dass er als autonomer Agent vom Menschen als einzigartiges Teammitglied wahrgenommen wird und eine spezielle Rolle bei der Erreichung des gemeinsamen Ziels einnimmt. Die Teammitglieder kommunizieren miteinander und der Mission Master ist in der Lage automatisiert die Aufgabe des Fahrens zu übernehmen, sowie Teilaufgaben auszuführen, welche der Mensch über einen Monitor überwacht. Er versteht die gemeinsamen Aufgaben im Einsatz mit seinen menschlichen Teammitgliedern und sie stehen in einem Abhängigkeitsverhältnis.

4.2.2 Assistenzsysteme an Bord von Marineschiffen

Im folgenden Unterkapitel wird der Einsatz eines HAT an Bord eines Marineschiffes anhand eines Assistenzsystems beispielhaft thematisiert. Zunächst wird die Entwicklung eines solchen Systems beschrieben, um im Anschluss hervorzuheben, auf welchen Ebenen und wie genau es dem Menschen eine Hilfestellung bietet.

Das Assistenzsystem stellt ein Einsatzsystem mit leistungsstarken Sensoren und Kommunikationsanlagen dar, das der Besatzung dabei helfen soll, kritische Aufgaben, welche im weiteren Verlauf des Textes weiter thematisiert werden, zu bewältigen. Diese Aufgaben werden durch Informationsunsicherheit, variierende und konkurrierende Zielsetzungen, Zeitdruck und ein hohes Entscheidungsrisiko erschwert. In der Zukunft sollte daher auf neuartige technische Möglichkeiten, wie z.B. kognitive Assistenzsysteme gesetzt werden, denn dem Personal stehen künftig immer komplexere und unvorhersehbare Szenarien gegenüber. Durch den Einsatz wird eine Entlastung des Personals bei der Aufgabenbewältigung möglich gemacht (Özyurt et al., 2012).

Das Entscheidungsunterstützungssystem COGAS identifiziert Luftziele in den Operationszentralen von Marineschiffen. Eine Möglichkeit der Analyse von Schwachstellen bei Arbeitsabläufen bietet der „Human View“, entwickelt von der NATO. Es gibt verschiedene Modellierungskomponenten, mit welchen analysiert und beschrieben wird, wie der Mensch im gesamten System agiert und welche Aufgaben er zu bewältigen hat. Der „Human View A“ stellt ein Mensch-System-Konzept dar. Beim HV-C handelt es sich um die funktionalen Randbedingungen. Die Operationszentrale (OPZ) ist das Herz des Schiffes. Insgesamt ist es wichtig, dass nur eine sehr geringe Fehlertoleranz zugelassen wird. Bisher konnten die betriebenen Einsatzsysteme bei der Identifizierung von Objekten nur eine geringe Verlässlichkeit bieten, weshalb der Mensch voraussichtlich auch in Zukunft noch der Entscheidungsträger bleibt (Özyurt et al., 2012). Hier kann davon ausgegangen werden, dass es sich um ein HAT handelt, da das System in einem Abhängigkeitsverhältnis mit dem Menschen steht (vgl. Abb. 4.1).

Der „Human View D“ beschreibt die Rollen und Aufgaben zwischen den Teammitgliedern. Aufgabenbereiche werden vorher fest definiert. Für diese ist dann der zugeteilte Operateur selber verantwortlich. Jeder Aufgabenbereich ist sehr vielseitig, weshalb eine breite Handlungssicherheit abgedeckt werden muss. Es gibt ein „Anti-Air-Warfare“ (AAW) Team, welches sich in mehrere Rollen unterteilt. Zum einen gibt es einen „Anti-Air-Warfare-Officer“ (AAWO). Dieser übernimmt die Führung der AAW Operationen und leitet den „Investigate Prozess“. Zum anderen gibt es einen „Picture Compiler“ (AAW1&2), welcher für die Überwachung des Luftraums zuständig ist. Zusätzlich erstellt und überwacht er noch das Lagebild. Der „Electronic Warfare Officer“ (EWO) ist für die Überwachung zuständig und analysiert die empfangenen Radaremissionen. Der „Air Tactical Controller“ ist für die Auswertung und die Analyse der Flugdaten zuständig. Der „Air Intercept Controller“ repräsentiert die Leitung der Abfangjäger und überwacht den Luftverkehr. Zu guter Letzt kümmert sich der „Track Manager“ (TM) um die Auswertung von Link-Informationen zu den Luftzielen und führt die Daten verteilter Systeme zusammen. Der „Human View E“ regelt die Teamkommunikation. Hierbei ist es wichtig zu wissen, ob es gerade um die Luftzielbekämpfung, die Seezielbekämpfung, die Unterwasserzielbekämpfung und/oder die asymmetrische Zielbekämpfung geht. Um die Bedrohung zu analysieren, muss jeder Operateur die Informationen seines Funktionsbereiches an den Anti-Air-Warfare Officer liefern. Wenn alle Informationen vollständig sind, wird vom AAWO eine Identität festgelegt und an die Teammitglieder AAW1 und/oder AAW2 weitergegeben, welche die Identität dann ins System eingeben. Der AAWO meldet außerdem noch über eine externe

Kommunikation die Bedrohung des Luftziels an andere Mitglieder (Özyurt et al., 2012). Beim „Human View H“ geht es um die Simulation der COGAS-Prozesse, welche im nächsten Unterkapitel thematisiert werden.

Dadurch, dass es an Bord viele laute Geräusche gibt und viele Informationen weitergegeben werden müssen, leidet die Kommunikation zwischen den Operateuren darunter, was dazu führen kann, dass wichtige Informationen verloren gehen. Der AAWO muss gegebenenfalls mehrere Kommunikationskreise gleichzeitig einschalten, was zu einer Überlagerung mehrerer Meldungen führen kann und die kognitiven Fähigkeiten zur Bearbeitung der Informationen nicht ausreichen. Was die Aufgaben- und Rollenstruktur angeht, gibt es ein internes Systemhandbuch, welches einen Überblick über Aufgabeninhalte schafft. Hier wird auch erwähnt, dass die AAWO-Aufgaben keinen starkstrukturierten Workflow beinhalten, was schnell zu einer Überlastungssituation und somit zu einer höheren Fehlentscheidungsrate führen kann. Im Allgemeinen soll dem Assistenzsystem ein Regelungsmodul ergänzt werden, wodurch der Grad der Unterstützung an den aktuellen Bedarf des Operateurs flexibel angepasst werden kann.

Das Assistenzsystem hat die Aufgabe zur Interpretation über das Vorhaben des Operateurs und das Resultat daraus dann mit seinem eigenen Interpretationsansatz zu vergleichen. Danach hat er die Aufgabe, je nach Unterstützungsgrad, diese Unterstützung bei der Bewältigung von Führungsaufgaben, bzw. der Identifizierung von Luftzielen zu leisten (Özyurt et al., 2012). Dadurch, dass das Assistenzsystem sich dem Operateur flexibel anpasst, kann es je nach Bedarf Unterstützung leisten und somit kann hier von einem HAT ausgegangen werden. Das Assistenzsystem nimmt als autonomer Agent eine unterstützende Rolle im Team mit seinem menschlichen Teammitglied ein. Es unterstützt den Menschen z.B. bei dessen Führungsaufgaben. Da der autonome Agent im vorliegenden Praxisbeispiel vom Vorhaben des Operateurs abhängt bzw. dieses interpretieren muss, ist er von diesem abhängig. Auf der anderen Seite interpretiert das Assistenzsystem unter anderem selbstständig das Vorhaben seines menschlichen Teammitglieds. Beispielsweise durch das oben erwähnte Regelungsmodul kann der autonome Agent das Assistenzsystem selbst kontrollieren und übt zudem Eigeninitiative aus. Das Verständnis der gemeinsamen Aufgaben im Team und die Wahrnehmung des autonomen Agenten als individuelles Teammitglied werden hier wie bisher angenommen. Sowohl der autonome Agent, als auch das menschliche Teammitglied nehmen eine zentrale Rolle ein und haben im Kollektiv dieselben übergeordneten Missionsziele. Aus den genannten Gründen folgt die Klassifizierung dieses Praxisbeispiels als HAT aus sachlicher Perspektive (vgl. Abb. 4.1).

Identifizierung von Luftzielen mit Hilfe des kognitiven und kooperativen Assistenzsystems (COGAS) an Bord von Marineschiffen

Das Assistenzsystem COGAS soll die Rolle eines erfahrenen Operateurs einnehmen und wird vom menschlichen Teammitglied als Individuum wahrgenommen, was auf ein HAT schließen lässt (vgl. Abb. 4.1). Somit kann es die anderen Operateure bei der Luftzielidentifizierung in der Operationszentrale unterstützen. Solche sicherheitskritischen Systeme stellen hohe Anforderungen an die Belastbarkeit, Konzentration und kognitive Leistungs- und Reaktionsfähigkeit der Operateure. In kritischen Situationen kann dies zu Fehlentscheidungen des Personals führen und somit gravierende Konsequenzen haben. Um diese Konsequenzen zu umgehen, werden vermehrt autonome Systeme eingesetzt. Indem Mensch und autonomer Agent als Team zusammenarbeiten, werden Arbeitsabläufe beschleunigt und Risiken, die durch Menschen hervorgerufen werden, minimiert (Özyurt et al., 2015). Das Zusammenspiel der kognitiven Agenten stellt eine Unterstützungsfunktion dar. Zum einen werden die Handlungsoptionen generiert und zum anderen gilt es, die Aktionen aufzuteilen. Diese beiden

Unterstützungsfunktionen werden mit Hilfe von autonomen Agenten umgesetzt. Ziel hierbei ist, das Situationsbewusstsein der Menschen zu verbessern, indem Informationen über den Zustand des Systems und der Umgebung geliefert werden.

Ein solches HAT gewährleistet, dass der Mensch dauerhaft in den Arbeitsprozess integriert wird und so ein „out-of-the-Loop“ verhindert werden kann. Die Zusammenarbeit beider autonomer Agenten spielt bei der erfolgreichen Unterstützungsleistung eine große Rolle. Die Aufgabe des „Action Selection Agents“ (ASA) ist die Überwachung des Arbeitssystems mit Sensoren, eine Analyse der Umgebung, die Anforderung fehlender Sensordaten und daraus die Ableitung des System- und Umgebungszustandes. Der ASA ist für die Erzeugung der Handlungsmöglichkeiten zuständig, die aus den Arbeitszielen abgeleitet werden. Diese sollen ausgeführt werden, damit zielführende Systemzustände erreicht werden. Der „Action Allocation Agent“ (AAA) hingegen arbeitet eng mit dem Operateur zusammen. Er überwacht den aktuellen Zustand des Lagebilds mit Hilfe des Human Machine Interface (HMI). Der AAA legt daraus dann den Unterstützungsgrad des Assistenzsystems fest und entscheidet, wie eine Aktion durchgeführt werden soll (Özyurt et al., 2015). Die Struktur des ASA besteht aus zwei separaten Wissensspeichern. Der eine enthält die Informationen über die momentane System- und Umgebungssituation und der andere Speicher zeigt die relevanten Kenntnisse auf, die ein Operateur durch Erfahrung und Ausbildung erwerben kann (a priori Wissen).

Bei der Umsetzung der Wissensrepräsentation muss zuerst die Situation festgestellt werden. Hierbei werden aktuelle Sensorinformationen mit den a-priori-Informationen aus dem Umgebungsmodell kombiniert und erzeugen einen aktuellen Umgebungszustand, welcher prognostizierte Informationen über die Umgebung enthält. Als nächstes werden die Zielsetzungen bestimmt, die aus dem aktuellen Umgebungszustand abgeleitet werden. Aus diesen Zielsetzungen wird die Basis für die Auswahl der Aufgaben und Aktionen gebildet. An dieser Stelle kommt ein Priorisierungsalgorithmus zum Einsatz, welcher bei der Auswahl der geeigneten Aufgaben hilft. Nach der Auswahl der Aufgaben erfolgen die Durchführung der Aktion und die Reihenfolge ihrer Ausführung. Zum Schluss muss die Darstellung konfiguriert werden, indem die geeignetsten Darstellungselemente identifiziert und alle ermittelten Informationen über das Human Machine Interface an den dafür verantwortlichen Operateur weitergegeben werden (Özyurt et al., 2015).

Abschließend kann beim Assistenzsystem COGAS aus sachlicher Perspektive aufgrund der von diesem erfüllten, in Abbildung 4.1 aufgelisteten zentralen Charakteristika eines autonomen Agenten von einem HAT ausgegangen werden. Das System wird vom menschlichen Teammitglied als ein Individuum wahrgenommen. Zudem nimmt es eine Rolle im Team ein und kommuniziert und interagiert mit dem Menschen. Es kann unabhängig vom menschlichen Teammitglied agieren, weist auf diese Art Eigeninitiative auf und kann sich selbst kontrollieren. Die beiden zuletzt angeführten Anforderungen werden vom autonomen Agenten in diesem Praxisbeispiel als erfüllt angenommen. Dadurch liegt aus sachlicher Sicht ein autonomer Agent vor und damit ein HAT. Beide Akteure, sowohl das menschliche Teammitglied als auch der autonome Agent, haben gemeinsame Missionsaufgaben. Diese bewältigen sie im Kollektiv. Auch das menschliche Teammitglied nimmt in Bezug auf die Zusammenarbeit mit dem COGAS in mehreren Aufgabenbereichen eine zentrale Rolle ein (vgl. Abb. 4.1).

4.3 Human Autonomy Teaming im Weltraum

Das letzte Unterkapitel gibt einen Ausblick auf die praktische Umsetzung des HAT-Konzepts im Weltraum. Im ersten Abschnitt dieses Unterkapitels wird ein Praxisbeispiel eines autonomen Agenten im Weltraum geboten. Bei diesem handelt es sich um einen fahrenden Roboter,

der weitestgehend autonom navigieren kann. Das menschliche Teammitglied gibt ihm Ziele vor, die der Rover dann selbst anfährt. Im zweiten Abschnitt dieses Unterkapitels werden zwei Interviews mit Experten zu autonomer Weltraumrobotik und autonomen Weltraumsystemen präsentiert. Das erste Interview wurde mit einem Experten des Robotics Innovation Center des Deutschen Forschungszentrum für Künstliche Intelligenz (DFKI) in Bremen und das zweite Interview mit einem Experten der Airbus Defence and Space GmbH in Bremen geführt.

4.3.1 Praxisbeispiele autonomer Robotik im Weltraum

OG10-ADE Projekt und der autonome Agent SherpaTT

Im Rahmen des Projekts ADE („Autonomous DEcision-making“) wird ein Rover-System eingesetzt, welches die Fortschritte der zukünftigen Technologie für Weltraumroboter-Missionen zeigt. Das autonome System soll planetare Explorationen durchführen und die Datensammlung ausweiten. Es sollen Fahrmanöver, die länger als ein Kilometer sind, durchgeführt werden. Dies wird stetig auf einem, dem Mars oder Mond ähnlichen, Terrain auf der Erde getestet. Das System weist bei unvorhersehbaren Ereignissen eine schnelle Reaktion auf und kann somit die zur Verfügung stehenden Ressourcen optimal nutzen (Deutsches Forschungszentrum für Künstliche Intelligenz, 2020). Zu diesen unvorhersehbaren Ereignissen könnte z.B. ein plötzlich entstandenes Hindernis gehören, welches umfahren werden muss oder dass sich die Beschaffenheit des Bodens stark verändert, sodass der Rover eine neue Route entwickeln muss. Obwohl der Rover über flexible Räder zur passiven Bodenadaptation verfügt, wird hauptsächlich darauf abgezielt, eine lange autonome Fahrt auf einem natürlichen, ebenen Gelände durchzuführen. Der Rover soll in der Lage sein, alle Entscheidungen, die mit der Verfolgung von Missionszielen und der Datenerfassung einhergehen, autonom zu treffen (Ocón et al., 2020). Zur grafischen Veranschaulichung werden in den Teilabbildungen 4.9a und 4.9b Fotos des angesprochenen Rover-Systems SherpaTT präsentiert. Diese wurden bei Testfahrten aufgenommen.



Abb. 4.9a und 4.9b: SherpaTT im Einsatz beim Projekt OG10-ADE (Fotos zur Verfügung gestellt vom Robotics Innovation Center des DFKI in Bremen, 2020; Fotos: Annemarie Popp).

Mit diesem Wissen über das Rover-System lässt sich ein autonomer Agent (vgl. Abb. 4.1) feststellen. Der autonome Agent des Projekts arbeitet auf drei verschiedenen Autonomielevels mit der Bodenkontrollstation zusammen und folglich kann auch von HAT gesprochen werden. Das erste Level der Zusammenarbeit ist der interaktive Modus. Dieser erlaubt dem menschlichen Teammitglied in der Bodenkontrollstation, per direkter Telekommunikation, dem Rover das Ausführen von Aufgaben zu befehlen und diese zu beobachten. Aufgaben wie die

Erkundung eines bestimmten Gebietes können zur sofortigen Erledigung an den Rover weitergegeben werden oder mit Vorgabe eines Zeitfensters. Der autonome Agent nimmt auf diese Weise eine unterstützende Rolle ein (vgl. Abb. 4.1). Im semi-autonomen Modus handelt der autonome Agent ereignisgesteuert. Das menschliche Teammitglied erstellt dabei einen Fahrplan, der mehrere hochrangige Ziele, bzw. Wegpunkte enthält, die der Rover während seiner Erkundung erreichen soll (Ocón et al., 2020). Der Rover erhält auf diese Weise durch den Menschen die Freigabe zur Ausführung (vgl. Abb. 4.1). Auf dem höchsten Level führt der autonome Agent die Missionsplanung, die Priorisierung bestehender Pläne und die wissenschaftliche Erkennung relevanter Ergebnisse selbstständig durch. Er wird lediglich von der Bodenkontrollstation überwacht und durch das Teilen von Informationen unterstützt (Ocón et al., 2020). Ein Beispiel für diese relevanten Daten wären Steininformationen für geologische Untersuchungen, welche der „Opportunistic Scientific Agent“ autonom erkennt (Deutsches Forschungszentrum für Künstliche Intelligenz, 2020). Ocón et al. (2020) sprechen in diesem Zusammenhang von einem „Scientific Detector“.

Der Scientific Detector kann Bilder über mehrere Kanäle verarbeiten und verwendet hochauflösende Farbbilder, sowie thermische Infrarotbilder, um wissenschaftlich relevante Ziele zu erkennen (Ocón et al., 2020). Dabei wird unterschieden zwischen bereits bekannten und neuartigen Funden. Bereits im semi-autonomen Modus kann der autonome Agent die Route eigenständig abändern, um diese erkannten Ziele zu befahren und zu untersuchen (Deutsches Forschungszentrum für Künstliche Intelligenz, 2020). Dabei helfen das „Autonomous Decision-making and Action-taking Module“ (ADAM) und der „Mission Planner Reactor“. ADAM löst dabei erst die Konflikte, die zwischen dem eigenen und dem von der Kontrollstation entwickelten Plan entstehen und der Mission Planner Reactor reagiert mit einem Planänderungsvorschlag (Ocón et al., 2020). Neben dem Scientific Detector sind diese beiden Systeme Bestandteile des autonomen Agenten. Schritt für Schritt wird das Rover-System in verschiedenen Umgebungen und immer komplexeren Szenarien getestet.

Zusammenfassend liegt sowohl aus sachlicher, als auch aus formaler Perspektive in Form des SherpaTT beim Projekt ADE ein autonomer Agent vor. Neben den bereits oben erwähnten, erfüllten Charakteristika eines autonomen Agenten durch den Rover, sind weitere Kriterien zu nennen, die dieser aufweist. Der Rover steht auf der einen Seite in einem Abhängigkeitsverhältnis mit seinem menschlichen Teammitglied. Er ist auf dessen Freigabe angewiesen. Dies stellt gemäß Abbildung 4.1 das zentrale, formale Kriterium für das Vorliegen eines autonomen Agenten dar. Es wird im vorliegenden Praxisbeispiel durch den autonomen Agenten erfüllt. Damit ist dieser aus formaler Sicht als solcher zu klassifizieren. Auf der anderen Seite fährt er autonom und unabhängig von seinem menschlichen Teammitglied bis zu einem vorgegebenen Ziel. In Form des ADAM-Moduls besitzt der Rover zudem die Fähigkeiten, Eigeninitiative zu zeigen und sich selbst zu kontrollieren. Der Rover SherpaTT kommuniziert beim Erreichen seines Zieles mit dem menschlichen Teammitglied. Dies erfolgt ebenfalls bei durch den autonomen Agenten nicht zu bewältigenden Hindernissen auf dem Weg, den er fährt. Durch die Implementierung des ADAM-Moduls besitzt der Rover ein Verständnis von den gemeinsamen Aufgaben im Team mit dem Menschen. Dass der Rover von seinem menschlichen Teammitglied im Team als Individuum angesehen wird, wird hier wie bei den bisherigen Praxisbeispielen angenommen. Aus den genannten Gründen liegt ebenfalls aus sachlicher Perspektive ein autonomer Agent vor. Damit liegt aus beiden in diesem Kapitel differenzierten Perspektiven ebenfalls ein HAT vor. Dem menschlichen Teammitglied kommt in der Zusammenarbeit mit dem Rover eine überwachende und einflussnehmende Rolle zu. Zudem muss der Mensch im HAT mit dem Rover mehrere Aufgaben (siehe oben) bewältigen. Das gemeinsame Ziel beider

Akteure ist die Findung und Bewältigung eines optimalen Weges bis zu einem zu erreichenden Ziel.

4.3.2 Stimmen aus der Praxis zu autonomen Agenten im Weltraum

Um den Ausblick in das Unterkapitel zum HAT im Weltraum zu vertiefen, wurden zwei Interviews mit Experten zu autonomer Weltraumtechnik geführt. Zunächst wird das Interview mit Herrn Dr. Cordes, dem Leiter des OG10-ADE Projekts am Robotics Innovation Center des DFKI in Bremen zu autonomer Weltraumrobotik angeführt. In diesem werden z.B. die Herausforderungen, das Zusammenspiel mit dem menschlichen Teammitglied, die Anforderungen, der aktuelle technologische Stand, sowie eine Prognose für die weitere Entwicklung autonomer Weltraumrobotik diskutiert. Im zweiten Interview mit Herrn Götz Anspach Von Broecker (GAvB), dem Key Account Manager German AeroSpace & Defence Research Institutions der Airbus Defence and Space GmbH in Bremen, wird autonome Weltraumtechnik in Bezug auf das HAT-Konzept diskutiert. In diesem Zusammenhang werden unter anderem der aktuelle Stand, ein praktischer HAT-Ansatz und Anforderungen an das HAT im Weltraum behandelt.

Infobox 4.3: Interview mit Herrn Dr. Florian Cordes, dem Leiter des OG10-ADE Projekts am Robotics Innovation Center (RIC) des DFKI in Bremen, über den Einsatz autonomer Weltraumrobotik.

Wir: Welche Weltraumtechnik des RIC des DFKI am Standort Bremen kann bereits als stark automatisiert bezeichnet werden?

Dr. Cordes: Das DFKI in Bremen betreibt angewandte Grundlagenforschung. Das bedeutet, dass wir die Anwendung von autonomen oder automatisierten Robotern auf der Erde erproben. Oft sollen die Roboter dann auf einem dem Mars oder dem Mond ähnlichen Gelände auf der Erde Search&Rescue-Aufgaben ausführen. Hierbei werden dem Roboter von dem/der menschlichen Kollaborateur/in Wegpunkte vorgegeben. Der Roboter soll dann autonom Lokalisations- und Mappingprozesse durchführen. Er erstellt autonom detaillierte 3D-Höhenkarten mit Hilfe von 3D-Kameras und soll dann selbstständig ein z.B. hochpriorisiertes und einen Kilometer entferntes Wegpunktziel erreichen. Anschließend soll der Roboter autonom Hindernisse auf dem Weg zum Zielpunkt überwinden oder ggf. Umwege und einen optimalen Weg zum Ziel finden und letzten Endes eine Probe sammeln können. Hierbei soll die Neugier des Roboters getriggert werden und sich eine weniger deterministische Autonomie entfalten, da die menschlichen Kollaborateure nie vorher wissen können, wie die vom Roboter erstellte Karte aussieht.

Wir: Wann kann Ihrer Meinung nach generell von einem autonomen Weltraumroboter gesprochen werden?

Dr. Cordes: Rover stellen Roboter dar, die als ein bodengebundenes Fahrzeug fungieren. Die Kommunikation mit der Bodenstation erfolgt über Satelliten-Relais. Die Rover sollen in der Regel dazu fähig sein, selbständig auf Fehler reagieren zu können. Im Idealfall sollten diese selbständig einen eigenen Pfad entwickeln und fahren können und dies bis zur Zielerreichung ohne eine Rückmeldung zur Bodenstation. Nur bei Schwierigkeiten oder Fehlern auf dem Weg zum Zielpunkt sollte dann eine Kommunikation und Rücksprache mit den menschlichen Kollaborateuren/innen in der Bodenstation erfolgen

Wir: Welche Funktionen der Weltraumroboter des DFKI in Bremen werden bereits verstärkt „autonomisiert“?

Dr. Cordes: Bei den Weltraumrobotern des DFKI in Bremen werden verstärkt Navigationsfunktionen autonomisiert. Auch Robotersysteme und Dockingstationen werden bereits hochautomatisiert gestaltet. In der Regel sind die Funktionen der Roboter aber eher als automatisiert zu bezeichnen.

Wir: Könnten Sie uns bitte Beispiele für autonome oder stark automatisierte Weltraumroboter des DFKI in Bremen nennen, die gegenwärtig eingesetzt werden?

Dr. Cordes: Gegenwärtig forschen wir im Rahmen des OG10-ADE Projekts mit anderen EU-Partnern an einem großen Rover, dem SherpaTT, der einen kleineren Laufroboter CREX (Crater Explorer) zu einem Krater bringen soll, durch den dann eine Steinprobenahme erfolgen soll. Ziel der Mission ist es, Wasser- und Sauerstoffvorkommen (für den Weiterflug) auf dem Mars überprüfen zu können. Dabei kommen ein autonomes Navigationssystem und Novelty-Detection zum Einsatz. Hierzu haben wir bereits erforderliche Systemaufstellungen und Feldexperimente auf der Erde durchgeführt. Der SherpaTT plant dabei den Weg selbständig und sieht Hindernisse ein. Das ADAM soll als „Autonomie-Baukasten“ die Autonomie in den Rover bringen. Durch dieses können autonome Mappingprozesse und autonome Navigation des Rovers erfolgen.

Wir: Welche Aufgaben müssten „Überwacher/innen“ autonomer oder stark automatisierter Weltraumroboter des DFKI in Bremen voraussichtlich in Zusammenarbeit mit der autonomen Technik noch ausführen und überwachen? Wie kommt es dabei zu einer Art Teamarbeit zwischen Mensch und autonomem Weltraumroboter?

Dr. Cordes: Die mit den Weltraumrobotern agierenden Menschen bezeichnen wir als Missionsüberwacher/innen und sowohl Roboter als auch Mensch als Kollaborateure. An bestimmten Schlüsselpunkten schaut der Mensch in den Fahrplan des Roboters und bringt seine Neugier mit in die Zusammenarbeit ein. In diesem Zusammenhang kann man von Teamarbeit sprechen, auch weil der Roboter selbst keinen eigenen Antrieb hat und die Neugier des Menschen übermittelt bekommt. Teamarbeit kommt zustande, indem der Roboter den Menschen unterstützt und erkennt, was dieser benötigt.

Wir: Worin sehen Sie und das DFKI in Bremen das zukünftige Potential autonomer Weltraumfahrzeuge und autonomer Weltraumrobotik?

Dr. Cordes: Momentan befindet ist die Forschung zu autonomen Weltraumrobotern weiter als die praktische Anwendung dieser im Weltraum und auf anderen Planeten. Autonome Roboter sollten autonome Systeme aufweisen, die selbständig agieren und dafür braucht es einen höheren wissenschaftlichen Stand der Technologie. Die Autonomisierung der Weltraumroboter nimmt laufend zu, aber aktuell weisen die Weltraumroboter wenig Autonomie auf.

Infobox 4.4: Interview mit Herrn Anspach Von Broecker (GAvB), dem Key Account Manager German AeroSpace & Defence Research Institutions der AIRBUS Defence and Space GmbH in Bremen zu autonomen Weltraumsystemen.

Wir: In welcher Hinsicht erfolgt gegenwärtig Human Autonomy Teaming mithilfe autonomer Weltraumsysteme von Airbus?

GAvB: Bei Airbus differenzieren wir Unmanned- und Manned Teaming und anhand des Technical Readiness Level (TRL) existieren in den Unmanned Teams unterschiedliche Abstufungen. Gegenwärtig gibt es bereits viele autonome Systeme mit einem TRL 10. Außerdem muss in Bezug zum HAT-Begriff geklärt werden, welchen TRL autonome Systeme aufweisen müssen um als autonomer Agent bezeichnet werden zu können. Zudem muss erwähnt sein, dass Airbus die Eingliederung von Astronauten und autonomer Systeme in ein ganzheitliches System für ein HAT als zentral wichtig erachtet. Eine weitere zu klärende Frage wäre, ob ein HAT immer noch vorliegt, wenn die autonomen Agenten eines HAT unterschiedliche und für die Geltung als autonomer Agent teilweise zu niedrige Automationslevel aufweisen.

Wir: Welche autonomen Weltraumsysteme sind bei der AIRBUS Defence and Space GmbH bereits stark automatisiert?

GAvB: Sogenannte Automative Transport Vehicles (ATVs) werden an unserem Standort in Bremen entwickelt. Diese können sich autonom Raumstationen nähern und an diese autonom andocken. Für die Autonomie aller Technologien und Fahrzeuge müssen zunächst autonome Systeme entwickelt werden. Hierfür nehmen wir eine graduelle Einführung von Autonomie vor, damit ein autonomes System HAT-Prozesse bewerkstelligen kann. Eine weitere Basis für autonome Systeme in Raketen, Satelliten, Grenzsicherungssystemen und Cybersecurity-Systemen stellen KI-basierte Subsysteme dar. Eine Anforderung an das Gesamtsystem autonomer Agenten muss dessen Resilienz gegenüber Einzelausfällen bestimmter autonomer Agenten sein. Dadurch soll eine unerwartete Zusammenarbeit und Verkopplung autonomer Agenten im System ermöglicht werden. Durch die Umgebung und externe Einflüsse ergeben sich übergeordnete Gesamtprogrammierungen autonomer Systeme von selbst. Eine weitere Frage ist, was das Gesamtsystem autonomer Systeme bzw. Agenten erreichen soll. Dazu muss in diesem ein Regulativ eingeführt und das Zusammenspiel der autonomen Agenten erprobt werden, um die Teamarbeit der autonomen Agenten zu ermöglichen.

Wir: Welche Aufgaben muss der Mensch voraussichtlich in Zusammenarbeit mit der autonomen Technik noch ausführen?

GAvB: Die Menschen bleiben stets den autonomen Agenten übergeordnet und sind allerhöchste Instanz und überwachen stets die Systems of Systems von HATs. Die Zielführungsfunktion des Menschen ist z.B. die Harmonisierung unterschiedlicher moralischer Gesamtvorstellungen unterschiedlicher Nationen.

Wir: Was verstehen Sie unter dem Systems-of-Systems-Ansatz in Bezug auf den Einsatz und die Erforschung autonomer Technik im Weltraum? Warum sollte man bei autonomen Elementen immer im Systemansatz denken?

GAvB: Beim System-of-Systems-Ansatz wird ein ganzheitliches gemeinsames System für autonome Agenten geschaffen. Dadurch lassen sich übergeordnete Problemstellungen lösen und nicht nur die individuellen unterzuordnenden Probleme der einzelnen autonomen Agenten. Im Weltraum werden internationale Systeme einzelner Nationen in Systems of Systems eingegliedert. Ein Zusammenspiel vieler autonomer Agenten und HATs ist nur mit einem Gesamtsystem möglich. Eine Verteilung von Problemen auf unterschiedliche Problemträger ist weniger effizient als der System-of-Systems-Ansatz. Für das Gesamtsystem des System-of-Systems-Ansatzes sind Regeln aufzustellen.

Wir: Was ist Ihre persönliche Prognose bezüglich der zukünftigen Praktikabilität autonomer Weltraumssysteme?

GAvB: Wir befinden uns bei Airbus inmitten der Entwicklung autonomer Weltraumssysteme. In Zukunft wird immer mehr Autonomie für die Kontrolle von menschlichen Bestrebungen im Weltraum notwendig werden. Dabei ist der System-of-Systems Ansatz zentral und wird sich durchsetzen. Bereits die Hälfte aller Weltraum-Projekte erfolgt teilweise autonom und die Gesamtsysteme des Ansatzes werden eine übergeordnete Instanz benötigen. Airbus wird wie andere Unternehmen zum System-of-Systems-Betreiber. Ferner wird die Freiheit durch die Gesetzgebung für die Entwicklung autonomer Systeme benötigt. Dafür braucht es Zertifizierungen in Deutschland. Zugleich müssen die Regulierungen und Zulassungen verschärft werden. Für eine Akzeptanz autonomer Systeme in der Gesellschaft und Politik sollte man im Weltall autonome Agenten erproben und dann Derivate der Technologie in terrestrischer Umgebung anwenden. Physikalisch gesehen ist Autonomie im Weltraum zwingend erforderlich und dort ist die Notwendigkeit für diese als erstes gegeben. Ein „Digitaler Twin“ autonomer Agenten und des Gesamtsystems als digitale Abbilder sind unbedingt erforderlich. Beim Digitalen Twin handelt es sich um ein System, das trotz Fehlern von Hardware und Software digital weiterarbeitet, also ein mentales System, das alle Prozesse simuliert.

4.4 Fazit zum Praxiskapitel

In diesem Kapitel wurde das Thema HAT aus der Praxisperspektive erörtert. Zu den Anwendungsbereichen autonomes Fahren, Militär und Weltraum wurden mehrere Praxisbeispiele des HAT präsentiert. Diese wurden anhand der Automationslevel-Taxonomie von Parasuraman et al. (2000) (formale Ebene) und durch die Schlussfolgerungen von O'Neill et al. (2020) (sachliche Ebene) zunächst als autonome Agenten und dann als HAT klassifiziert. Es bleibt festzuhalten, dass sich dank technischer Fortschritte zunehmend auch in der Praxis Teams zwischen Mensch und autonomen Agenten etablieren. Der autonome Agent unterstützt den Menschen und ersetzt diesen nicht. HAT zielt als Konzept der Teamarbeit zwischen Mensch und Maschine auf eine enge Zusammenarbeit beider Akteure ab. Weder das menschliche Teammitglied noch der autonome Agent übernehmen alle Aufgaben im Team. Beim HAT erfolgt eine Aufgabenverteilung zwischen beiden Akteuren, sodass beiden Teammitgliedern eine zentrale und bestimmte Rolle zukommt. Das HAT hat bereits Einzug in mehrere Anwendungsgebiete gefunden. Dennoch ist das Potenzial von HAT noch lange nicht ausgeschöpft und lässt weiteren Spielraum für höhere Autonomie, bessere Technik und Zusammenarbeit. Es obliegt folglich der Praxis und der Forschung, das HAT weiter zu erkunden und gemeinsam zur erfolgreichen Umsetzung autonomer Agenten als Teammitglied des Menschen im Sinne eines HAT beitragen zu können.

Literaturverzeichnis

- Airbus Defence (2020). Future Combat Air System: Owning the sky with the next generation weapons system. Zugriff am 02.01.2021. Verfügbar unter <https://www.airbus.com/newsroom/stories/Future-Combat-Air-System-Owning-the-sky-with-the-Next-Generation-Weapons-System.html#>
- Allianz pro Schiene e.V. (18.11.2016). Innovationsträger Schiene: Autonomes Fahren. *Youtube*. Zugriff am 05.01.2021. Verfügbar unter <https://youtu.be/sV45ZfSr0ew>
- Allianz pro Schiene e.V. (30.11.2016). Selbstfahrende Metros: Eine Milliarde Fahrgäste jedes Jahr. Zugriff am 05.01.2021. Verfügbar unter <https://www.allianz-pro-schiene.de/presse/pressemitteilungen/uebersicht-selbstfahrende-metros-europa/>
- Amt für Heeresentwicklung (2019). *Künstliche Intelligenz in den Landstreitkräften. Ein Positionspaper des Amts für Heeresentwicklung* (2. Auflage). Köln: Amt für Heeresentwicklung. Zugriff am 02.01.2021. Verfügbar unter <https://www.bundeswehr.de/resource/blob/156024/d6ac452e72f77f3cc071184ae34dbf0e/download-positionspaper-deutsche-version-data.pdf>
- Arnold, H. (31.07.2020). Fahrerlose Busse - Die Zukunft des ÖPNV? *PwC Deutschland*. Zugriff am 05.01.2021. Verfügbar unter: <https://www.pwc.de/de/im-fokus/mobility-transformation/autonome-busse-im-oepnv.html>
- Brandt, A., Jentzsch, G. & Pradka, A. (2018). Projektbericht zum EDDI und LKW-Platooning. *ADAC e.V.* Zugriff am 18.12.2020. Verfügbar unter <https://www.adac.de/rundums-fahrzeug/ausstattung-technik-zubehoer/autonomes-fahren/technik-vernetzung/platooning-lkw-automatisiert>
- Briginshaw, D. (2018). Rio Tinto completes automation of Pilbara rail network. *Railjournal*. Zugriff am 02.01.2021. Verfügbar unter <https://www.railjournal.com/freight/rio-tinto-completes-automation-of-pilbara-rail-network/>
- Buchan, M. (n. d.). AutoHaul Project. *Calibre Group*. Zugriff am 02.01.2021. Verfügbar unter <https://www.calibregroup.com/projects/project/autohaul-project>
- Bundesregierung (2018). *Drohnen-Schwärme in Waffensystemen der Bundeswehr*. Drucksache 19/5433 – Antwort der Bundesregierung auf die Kleine Anfrage der Abgeordneten Andrej Hunko, Heike Hänsel, Michel Brandt, weiterer Abgeordneter und der Fraktion DIE LINKE – Drucksache 19/4715, S.1-8. Zugriff am 12.02.2021. Verfügbar unter <http://dipbt.bundestag.de/dip21/btd/19/054/1905433.pdf>
- Bundeswehruniversität München (n. d.). Intelligence for autonomous systems. *HAT.tech Softwarelösungen*. Zugriff am 03.01.2020. Verfügbar unter <https://www.hattec.de/software-solutions/>
- Bundeswehruniversität München (n. d.). Manned-unmanned teaming. *HAT.tech Softwarelösungen*. Zugriff am 03.01.2020. Verfügbar unter <https://www.hattec.de/software-solutions/>
- Daimler AG, Mercedes-Benz (2013). Mercedes-Benz, S-Klasse S 500 INTELLIGENT DRIVE - Vernetzte Sensorik des Forschungsfahrzeugs S 500 INTELLIGENT DRIVE (Bildquelle). Zugriff am 02.01.2021. Verfügbar unter <https://media.daimler.com/marsMediaSite/de/instance/picture/S-500-INTELLIGENT-DRIVE.xhtml?oid=7517505>

- Daimler AG, Mercedes-Benz (2013). Mercedes-Benz, S-Klasse S 500 INTELLIGENT DRIVE (Bildquelle). Zugriff am 02.01.2021. Verfügbar unter <https://media.daimler.com/marsMediaSite/de/instance/picture/S-500-INTELLIGENT-DRIVE.xhtml?oid=7517505>
- Daimler AG, Mercedes-Benz (2016). IAA Preview 2016 Daimler Buses: Mercedes-Benz Future Bus mit CityPilot (Bildquelle). Zugriff am 02.01.2021. Verfügbar unter <https://media.daimler.com/marsMediaSite/de/instance/picture/Mercedes-Benz-Future-Bus-mit-CityPilot-2016.xhtml?oid=12806228>
- Daimler AG, Mercedes-Benz (2019). Safety First For Automated Driving-SaFAD. Zugriff am 01.12.2020. Verfügbar unter <https://www.daimler.com/innovation/case/autonomous/safety-first-for-automated-driving.html>
- Daimler AG, Mercedes-Benz (2020a). Fahrerloses Parken. Automated Valet Parking. Zugriff am 02.01.2021. Verfügbar unter <https://www.daimler.com/innovation/case/autonomous/fahrerlos-geparkt.html>
- Daimler AG, Mercedes-Benz (2020b). Vollautomatisiertes und fahrerloses Parken kommt mit der neuen Mercedes-Benz S-Klasse an den Flughafen Stuttgart (Bildquelle). Zugriff am 02.01.2021. Verfügbar unter <https://media.daimler.com/marsMediaSite/de/instance/ko/Vollautomatisiertes-und-fahrerloses-Parken-kommt-mit-der-neuen-Mercedes-Benz-S-Klasse-an-den-Flughafen-Stuttgart.xhtml?oid=47757313>
- Daimler AG, Mercedes-Benz (n. d.a). Vom mitdenkenden zum autonomen Auto. Intelligent Drive heute und morgen. Zugriff am 23.12.2020. Verfügbar unter <https://www.daimler.com/innovation/case/autonomous/intelligent-drive.html>
- Daimler AG, Mercedes-Benz (n. d.b). Der Mercedes-Benz Future Bus. Die Zukunft der Mobilität. Zugriff am 03.01.2021. Verfügbar unter <https://www.daimler.com/innovation/autonomes-fahren/future-bus.html>
- Department of Defence (2014). *Unmanned systems integrated roadmap* (Reference No. 14-S-0553). Washington, DC.
- Deutsches Forschungszentrum für Künstliche Intelligenz (2019). Autonomous Decision Making in Very Long Traverses. *DFKI Projekte & Publikationen*. Zugriff am 07.01.2020. Verfügbar unter <https://www.dfki.de/web/forschung/projekte-publikationen/projekte-uebersicht/projekt/ade-og10/>
- Doll, N. (2018). Das fahrerlose Zug-Zeitalter beginnt nicht in Deutschland. *WELT*. Zugriff am 02.01.2021. Verfügbar unter <https://www.welt.de/wirtschaft/article181511646/Deutsche-Bahn-Franzoesische-SNCF-plant-vor-der-DB-fahrerlose-Zuege.html>
- Hansen, S. & Janssen, J.-K. (2018). Robobus: Wo und wie autonome Busse jetzt schon fahren. *c't Magazin für Computertechnik*, 2018/10, 64. Zugriff am 04.01.2021. Verfügbar unter <https://www.heise.de/select/ct/2018/10/1526008148939708>
- Heiming, G. (13.11.2020). Das niederländische Heer erprobt Mission Master. *Europäische Sicherheit & Technik*. Zugriff am 05.01.2021. Verfügbar unter <https://esut.de/2020/11/meldungen/24026/das-niederlaendische-heer-erprobt-mission-master/>
- Heiming, G. (27.03.2019). Mission Master UGV vor Produktionsstart. *Europäische Sicherheit & Technik*. Zugriff am 05.01.2021. Verfügbar unter <https://esut.de/2019/03/>

- fachbeitraege/industrie-fachbeitraege/11869/mission-master-ugv-vor-produktionsstart/
- 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), 262–273.
- Monroy, M. (25.07.019). Rheinmetall baut bewaffneten Drohnenpanzer. *Telepolis Wirtschaft*. Zugriff am 05.01.2021. Verfügbar unter <https://www.heise.de/tp/features/Rheinmetall-baut-bewaffneten-Drohnenpanzer-4478682.html>
- Ocón, J., Dragomir, I., Coles, A., Green, A., Kunze, L., Marc, R., Perez, C. J., Germa, T., Bissonnette, V., Scalise, G., Foughali, M., Kapellos, K., Dominguez, R., Cordes, F., Paar, G., Reina, G. & Kisdi, A. (2020). ADE: Autonomous Decision Making In Very Long Traverses. *15th International Symposium on Artificial Intelligence, Robotics and Automation in Space (i-SAIRAS '20)*, Conference. Pasadena, California, United States.
- O'Neill, T. A., McNeese, N. J., Barron, A. & Schelble, B. (2020). Human–Autonomy Team-ing: A review and analysis of the empirical literature. *Human Factors*. Advance Online Publication. <https://doi.org/10.1177/0018720820960865>
- Özyurt, E., Döring, B. & Flemisch, F. (2012). Vorgehensweise zur simulationsbasierten Entwicklung eines kognitiven Assistenzsystems für Marineschiffe. Veröffentlichte Dissertation, Universität Duisburg-Essen.
- Özyurt, E., Döring, B. & Flemisch, F. (2015). *Kognitives und kooperatives Assistenzsystem (COGAS) zur Luftzielidentifizierung an Bord von Marineschiffen*. Veröffentlichte Dissertation, Universität Duisburg-Essen.
- Parasuraman, R., Sheridan, T. B. & Wickens, C. D. (2000). A model for types and levels of human interaction with automation. *IEEE Transactions on Systems, Man, and Cybernetics –Part A: Systems and Humans*, 30(3), 286–297.
- Pletsch, M. (16.03.2020). Mensch-Maschine: EU- Großprojekte zum Manned- Unmanned Teaming. *Informationsstelle Militarisierung e.V. (IMI), AUSDRUCK (März 2020)*, 10–13. Zugriff am 02.01.2021. Verfügbar unter <https://www.imi-online.de/download/Ausdruck-100-MP-MUMT.pdf>
- Rheinmetall Defence (n. d.). Rheinmetall Mission Master – Autonomes Robotikfahrzeug maximiert die Effizienz der Soldaten. Zugriff am 06.01.2021. Verfügbar unter https://rheinmetall-defence.com/de/rheinmetall_defence/systems_and_products/unbemannte_fahrzeuge/mission_master/index.php
- Rheinmetall Defence (2020), Rheinmetall stellt neue Spähauflärungsvariante des Robotikfahrzeugs Mission Master vor. Zugriff am 23.01.2021. Verfügbar unter https://www.rheinmetall-defence.com/de/rheinmetall_defence/public_relations/news/archiv/archive2016/index~1_25536.php
- Shuttleworth, J. (07.01.2019). SAE Standards News: J3016 automated-driving graphic update. *SAE International*. Zugriff am 25.11.2020. Verfügbar unter <https://www.sae.org/news/2019/01/sae-updates-j3016-automated-driving-graphic>
- Smith, K. (2019). Rise of the machines: Rio Tinto breaks new ground with AutoHaul. *International Railway Journal*. Zugriff am 03.01.2021. Verfügbar unter https://www.railjournal.com/in_depth/rise-machines-rio-tinto-autohaul

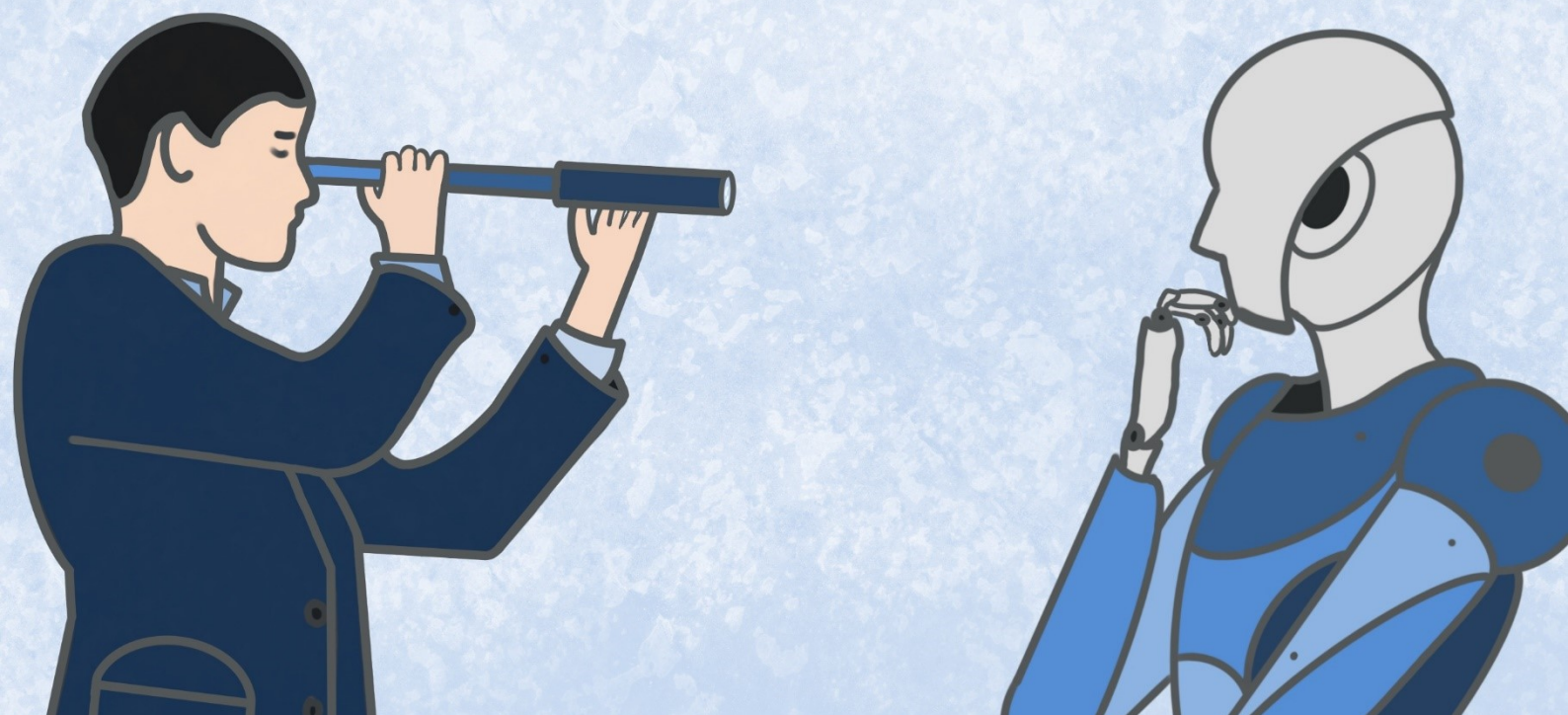
Society of Automotive Engineers International (15.06.2018). Taxonomy and Definitions for Terms Related to Driving Automation Systems for On-Road Motor Vehicles. *SAE International*. Advance Online Publication. https://doi.org/10.4271/J3016_201806

Verband der Automobilindustrie (2015). *Automatisierung – Von Fahrerassistenzsystemen zum automatisierten Fahren* (Broschüre, 03.09.2015). Zugriff am 03.01.2021. Verfügbar unter <https://www.vda.de/dam/vda/publications/2015/automatisierung.pdf>

WELT Nachrichtensender (2020). Future Combat Air System: Das plant die EU für das Riesen- Rüstungsprojekt. *Youtube*. Zugriff am 07.01.2021. Verfügbar unter <https://www.youtube.com/watch?v=GwGNLIUOZf8>

Ein Blick in die Zukunft des Human Autonomy Teamings

Julian Breden & Simon Sambale



Inhaltsverzeichnis

5	Ein Blick in die Zukunft des Human Autonomy Teamings	
5.1	Die Rolle des Menschen im Human Autonomy Teaming der Zukunft	151
5.1.1	Der Wert des Menschen im Human Autonomy Teaming der Zukunft	
5.1.2	Die Aufgaben des Menschen im Human Autonomy Teaming der Zukunft	
5.2	Verbesserungsmöglichkeiten für das Human Autonomy Teaming der Zukunft	155
5.2.1	Level an Autonomie des autonomen Agenten	
5.2.2	Transparenz und Zuverlässigkeit	
5.2.3	Aufgabenmerkmale	
5.2.4	Teamzusammensetzung	
5.2.5	Training	
5.2.6	Menschliche individuelle Unterschiede	
5.3	Soziale und ethische Dilemmata der Zukunft	161
5.3.1	Die moralische Entscheidungsfindung	
5.3.2	Autonomer Agent im Gesetz	
5.3.3	Die Risiken und Gefahren des autonomen Agenten	
	Literaturverzeichnis	171

Kapitel 5: Ein Blick in die Zukunft des Human Autonomy Teamings

Von Julian Breden & Simon-Tolga Sambale

Human Autonomy Teaming (HAT) wird als Arbeitsweise der Zukunft bezeichnet. HAT wird die traditionelle Arbeitsweise des Menschen neu definieren. Rein menschliche Teams werden zunehmend durch Teams ersetzt, in denen Mensch und Technik zusammenarbeiten und miteinander interagieren. Die Rolle der Technik wird von dem autonomen Agenten eingenommen. Der Mensch hat über seine lange Geschichte vieles in seine Gesellschaft integriert. HAT ist der nächste Schritt zum sogenannten digitalen Zeitalter. Mit neuen zukünftigen Möglichkeiten und Chancen wird es zu Problemen kommen, die es zukünftig zu lösen gilt. Im folgenden Kapitel 5 wird auf die zukünftige Rolle des Menschen im HAT, die Verbesserungsmöglichkeiten des HAT, und auf die sozialen und ethischen Dilemmata in Bezug auf das HAT eingegangen.

5.1 Die Rolle des Menschen im Human Autonomy Teaming der Zukunft

Studien zeigen, dass nach jetzigem Forschungsstand bei hochgradig automatisierten Systemen eine grundlegende Notwendigkeit besteht, dass Menschen diese automatisierten Vorgänge überwachen (Baxter et al., 2012; Endsley, 2017). Eine wichtige Rolle des menschlichen Operators wird in der nahen Zukunft darin liegen, in Situationen einzugreifen, die das System alleine nicht bewältigen kann (Endsley, 2017). Teamarbeit scheint bei autonomen Vorgängen immer noch von großer Wichtigkeit zu sein. Dies könnte ein Grund sein, wieso HAT in Zukunft wohl immer bedeutender werden wird. Gleichwohl kann davon ausgegangen werden, dass in Zukunft immer mehr autonome Systeme erfolgreich in die Arbeitswelt integriert werden. Eine wichtige Frage, die gestellt werden sollte, lautet, welche Aufgaben im Teaming mit dem autonomen Agenten zukünftig für den Menschen prädestiniert sind.

Um diese Frage zu beantworten, werden auf den folgenden Seiten zwei Ansätze vorgestellt. Der erste Ansatz unter (5.1.1) bezieht sich vor allem auf Ausführungen von Agrawal et al. (2017). Diese beschäftigen sich mit den unterschiedlichen Stärken der künstlichen Intelligenz (KI) und des Menschen, welche im Vorhersagen und dem Urteilsvermögen liegen. Anschließend folgt der zweite Ansatz unter (5.1.2), welcher sich auf Ausführungen von Lee (2018) und Bünthe (2020) stützt. Es geht um die Bestimmung der Aufgaben und Kompetenzen von KI und dem Menschen mithilfe einer 4-Felder-Matrix.

5.1.1 Der Wert des Menschen im Human Autonomy Teaming der Zukunft

Der Vorteil des HAT besteht darin, dass die menschlichen Vorteile und die Vorteile des autonomen Agenten gemeinsam genutzt werden können. Diese können sich potenzieren und folglich kann ein HAT effektiver sein als die Summe der Einzelteile (vgl. Chen & Barnes, 2014). Um zu verstehen, welchen Wert der Mensch im zukünftigen HAT haben wird, muss betrachtet werden, wodurch sich die menschliche Intelligenz im Gegensatz zur KI des autonomen Agenten auszeichnet. Dieses Unterkapitel beschäftigt sich mit den Stärken des Menschen in Bezug auf das HAT. Denn sollte der Faktor Mensch zukünftig bei der Entwicklung autonomer Systeme nicht ausreichend berücksichtigt werden, wird dies zu Defiziten im Bereich der Sicherheit und für das menschliche Wohlbefinden führen (Karvonen & Martio, 2018). Zusätzlich werden die Stärken der KI thematisiert, um zu klären, was den autonomen Agenten auszeichnet.

Unterschiedliche Stärken der KI und des Menschen: Vorhersagen und Urteilsvermögen

Die KI kann Vorhersagen schneller und kostengünstiger treffen als der Mensch. Bei Vorhersagen handelt es sich um die Fähigkeit, durch vorhandene Informationen neue Informationen zu generieren. Der Mensch verfügt über ein besseres Urteilsvermögen. Urteilsvermögen ist die Fähigkeit, wohlüberlegte Entscheidungen zu treffen. Dazu gilt es zu verstehen, welche Auswirkungen verschiedene Aktionen und Handlungen haben können. Agrawal et al. (2017) beschreiben die bessere Vorhersagefähigkeit der KI durch maschinelles Lernen. Beim maschinellen Lernen werden Computeralgorithmen verwendet, um autonom aus Daten und Informationen zu lernen (Kleesiek et al., 2019). Der Mensch wird zukünftig unter anderem dafür zuständig sein, die KI mit den Daten und Informationen zu versorgen. Anhand dieser Daten wird dann maschinelles Lernen stattfinden (vgl. Cesafsky et al., 2019). Verbesserungen im Bereich des maschinellen Lernens führen zu besseren und kostengünstigeren Vorhersagen der KI. Beispielhaft dafür ist das autonome Fahren. Hierbei wird vorausgesagt, was eine fahrende Person in einer bestimmten Situation tun würde. In Bereichen, in denen eine Urteilsfindung nicht einfach beschrieben werden kann, wie im Gesundheitssektor, ist menschliches Urteilsvermögen von Bedeutung (Agrawal et al., 2017). Hierauf wird im Folgenden noch einmal genauer eingegangen.

Dank kostengünstiger und jederzeit verfügbarer Vorhersagen durch künstliche Intelligenz wird menschliche Urteilsfindung immer wertvoller werden. Ökonomen sprechen in diesem Fall von komplementären Gütern. Beurteilungsfähigkeiten sind komplementär zu Vorhersagen und werden stärker nachgefragt, wenn der Preis für Vorhersagen sinkt. Dies resultiert aus stetigen Verbesserungen im Bereich der künstlichen Intelligenz (Agrawal et al., 2017). Beispielhaft für HAT würde das bedeuten, dass der autonome Agent Vorhersagen besser treffen kann als der Mensch. Der Mensch wird in Zukunft benötigt, um sein Urteilsvermögen beizusteuern, damit aus den gelieferten Vorhersagen und Prognosen die richtigen Schlüsse gezogen werden. Drei Bereiche, die als gutes Beispiel für die Anwendung von HAT gesehen werden können, sind der Gesundheitssektor, der Einzelhandel und der Sicherheitssektor (Agrawal et al., 2017; Strybel et al., 2018). Anhand dieser Beispiele sollen die unterschiedlichen Stärken bezüglich des Vorhersagens und Urteilsvermögens greifbarer gemacht werden.

Gesundheitssektor

Im Gesundheitssektor wird KI zusammen mit den Menschen eine große Rolle spielen (Agrawal et al., 2017). In Zukunft werden im Gesundheitssektor immer häufiger Mensch und Technik als Team gemeinsam fungieren müssen und es wird zwangsläufig HAT stattfinden (Strybel et al., 2018). KI verbessert Diagnosen, was zu effektiverer Behandlung und Pflege führt. Insbesondere in der Radiologie wird heute schon KI eingesetzt, um Krankheiten anhand von Bildern zu erkennen. Im Rahmen des maschinellen Lernens lernt die KI durch die verschiedenen Krankheitsbilder. Außerdem kann die KI den Ärzten Hinweise liefern, nach welcher Krankheit gesucht werden soll. Ein Beispiel für die erfolgreiche Anwendung in der Radiologie ist die Krankheit NSCLC (Non-Small-Cell Lung Cancer). Hierbei handelt es sich um ein Lungenkarzinom. Die KI konnte erfolgreiche Vorhersagen für distale Metastasen liefern. Distale Metastasen liegen weiter von der Körpermitte bzw. vom Herzen entfernt (Hosny et al., 2018). Die Behandlung und Pflege wird die Aufgabe des Menschen bleiben. Ein Grund dafür ist, das im vorherigen Absatz erklärte, menschliche Urteilsvermögen. Verschiedene Patient/innen haben unterschiedliche Bedürfnisse, welche erfüllt werden müssen. Menschen sind aufgrund ihrer emotionalen Fähigkeiten wie Empathie und der Möglichkeit, Gefühle nachzuvollziehen, besser

in der Lage, auf die unterschiedlichen Bedürfnisse der Patient/innen einzugehen als autonome Systeme. Gerade im Gesundheitssektor gibt es zahlreiche Situationen,

in denen Pro- und Contra-Argumente für verschiedene Behandlungsweisen abgewogen werden müssen. Dieses Urteilsvermögen werden autonome Systeme wahrscheinlich nie in einer Art und Weise entwickeln, die aus menschlicher Sicht vertretbar sein wird (Agrawal et al., 2017). Über die hohe Komplexität, die es momentan und wohl auch zukünftig unmöglich macht, einen autonomen Agenten wie einen Menschen zu programmieren, wird unter 5.3.1 berichtet.

Einzelhandel und Sicherheitssektor

Andere Bereiche, in denen in Zukunft HAT stattfinden kann, sind der Einzelhandel und der Sicherheitssektor. Im Einzelhandel können autonome Agenten Kaufverhalten besser voraussagen, da die KI des autonomen Agenten die riesigen Datenmengen im Einzelhandel schnell und effektiv auswerten kann. Menschliches Personal, das die Kunden berät und auf Fragen der Kunden eingehen kann, wird weiterhin benötigt, um den jeweiligen Einzelhandel von seinen Konkurrenten zu differenzieren und einzigartig zu machen. Im Sicherheitssektor werden autonome Agenten zukünftig eventuell Verbrechen vorhersagen können. Es werden trotzdem Polizisten benötigt, die ethische Beurteilungen mit Polizeiarbeit kombinieren können (Agrawal et al., 2017).

Das benötigte Urteilsvermögen des Menschen könnte sich zukünftig ändern, sollte es autonomen Agenten tatsächlich gelingen, Entscheidungen besser abzuwägen und eigene Urteile zu fällen. Die KI des autonomen Agenten könnte durch fehlerhafte Urteile lernen und Muster erkennen, wie in ähnlichen Szenarien bessere Entscheidungen getroffen werden können. Trotzdem werden in Zukunft immer wieder neuartige Situationen auftauchen, in denen menschliche Urteilsfindung benötigt wird (Agrawal et al., 2017). Die Fähigkeit, Urteile und moralische Entscheidungen zu treffen, wird den Menschen auch in Zukunft zu einem unverzichtbaren Faktor im HAT machen.

Um die Ausgangsfrage „Welche Aufgaben sind im Teaming mit dem autonomen Agenten zukünftig für den Menschen prädestiniert?“ aufzugreifen, hat dieser Ansatz gezeigt, dass der Mensch vor allem die Aufgaben übernehmen wird, die Urteilsvermögen und moralische Entscheidungen voraussetzen. Dieses Unterkapitel führt das Urteilsvermögen des Menschen als große Stärke gegenüber der KI auf. Trotzdem soll keineswegs der Eindruck entstehen, dass das menschliche Urteilsvermögen und damit auch die menschliche Urteilsfindung unfehlbar sind. Ein gutes Beispiel für die verzerrte Urteilsfindung des Menschen liefert die Infobox 5.1 im Unterkapitel 5.3.1.

5.1.2 Die Aufgaben des Menschen im Human Autonomy Teaming der Zukunft

In diesem Unterkapitel soll die Ausgangsfragestellung „Welche Aufgaben sind im Teaming mit dem autonomen Agenten zukünftig für den Menschen prädestiniert?“ ebenfalls mit aufgegriffen werden. Die Aufgaben des Menschen und der KI, die stellvertretend für den autonomen Agenten betrachtet werden kann, werden mithilfe einer 4-Felder-Matrix bestimmt. Die 4-Felder-Matrix besteht aus der x-Achse „Automatisierung“ und „Kreativität/Strategie“ sowie der y-Achse „Empathie und Mitgefühl“. Je weiter links sich Aufgaben auf der x-Achse befinden, desto repetitiver und leichter zu automatisieren sind diese Aufgaben. Je weiter rechts sich Aufgaben auf der x-Achse befinden, desto kreativer und strategischer sind diese Aufgaben.

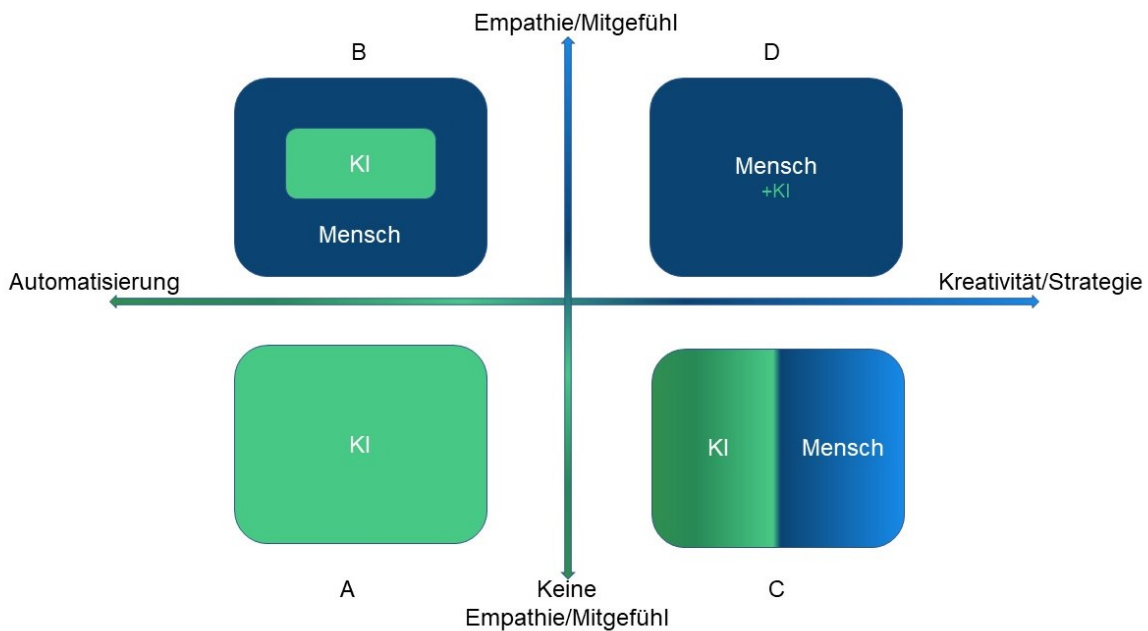


Abb. 5.1 4-Felder Matrix zur Bestimmung der Aufgaben von Menschen und künstlicher Intelligenz (in Anlehnung an Lee, 2018).

Laut Lee (2018) können Aufgaben, die repetitiv und routinemäßig sind, sowie keine Empathie oder Mitgefühl erfordern, in naher Zukunft vollständig durch KI übernommen werden. Diese Tätigkeiten befinden sich links unten (**A**) in der Matrix. Aufgaben, bei denen zwar eine grundsätzliche Automatisierung möglich wäre, Mitgefühl sowie Empathie aber eine wichtige Rolle spielen, befinden sich links oben (**B**). Solche Aufgaben könnten im Team von Menschen und KI übernommen werden und wären ein gutes Beispiel für HAT. Strategische und kreative Aufgaben, die keine Empathie und Mitgefühl erfordern, befinden sich rechts unten in der Matrix (**C**) und können ebenfalls im Team von Menschen und KI bearbeitet werden. Bei hochstrategischen und empathischen Tätigkeiten, die sich rechts oben in der Matrix befinden (**D**), wird in Zukunft der Mensch den wichtigeren Teil darstellen, die KI assistiert (Bünthe, 2020).

Bünthe (2020) liefert Beispiele zu den verschiedenen Anordnungen in der 4-Felder Matrix. Ein mögliches Beispiel für (**A**) sind Tellerwäscher/innen und LKW-Fahrer/innen. Diese Berufe sind stark repetitiv, routinemäßig und erfordern kein strategisches Denken oder Kreativität. Diese musterhaften Abläufe kann sich die KI durch maschinelles Lernen aneignen und diese in Zukunft komplett übernehmen. Für (**B**) kommen Hochzeitsplaner/innen in Frage. Die KI wird im Teaming beim Planen die Hauptarbeit leisten. Das Planen einer Hochzeit ist eine Aufgabe, die automatisiert stattfinden kann. Die Buchung der „Location“, die Bestellung der Speisen und Getränke bei einem Cateringservice, die generelle Kostenplanung, die Auswahl eines Fotografen und der Musik für das Fest sind Aufgaben, die von der KI übernommen werden können. Der Mensch würde die gesamte Betreuung des Hochzeitspaares übernehmen und die Präferenzen für die einzelnen Planungsschritte mit dem Brautpaar durchsprechen, hierfür benötigt es Einfühlungsvermögen und Empathie. Für Aufgaben im Bereich (**C**) kommen Schriftsteller/innen und Autor/innen in Frage. Diese werden von künstlicher Intelligenz aufgrund der niedrigen Empathiebedürftigkeit unterstützt. Beispielsweise könnte die KI die Recherche für Artikel oder Bücher übernehmen und eine Vielzahl von verschiedenen Daten und Informationen filtern und auswerten. Den kreativen Teil, das Schreiben, würde der Mensch übernehmen. Im Bereich (**D**) kann es sich um Aufgaben für CEOs und Strateg/innen handeln. Diese Aufgaben

erfordern Führungskennntnisse und Menschenkennntnisse. Voraussetzungen für diese Berufe sind, dass immer wieder Lösungen für neuartige Situationen gefunden werden können. Strategisches Denken und Kreativität sind hierbei unersetzlich. Empathie und Mitgefühl werden benötigt, da diese Arbeiten eine starke soziale Interdependenz aufweisen und eine ständige Interaktion mit anderen Menschen stattfindet. Aufgrund der hohen Empathie und hohen Strategie könnte KI hier, wie in der Abbildung zu sehen, allenfalls assistieren. Die 4-Felder-Matrix impliziert eine grundlegende Aufgabenverteilung, auf die am Ende des Unterkapitels nochmal eingegangen wird. Es bleibt festzuhalten, dass der Mensch kreativer und strategischer ist als die KI des autonomen Agenten (Bünthe, 2020). Einen Grund, wieso die KI weniger kreativ und strategisch ist, liefert Bünthe (2020):

Bezüglich der **Kreativität**:

KI lernt aus vergangenen Daten und verbindet diese neu, um die abhängige Variable, den Erfolg (etwa ein gewünschtes Verhalten, mehr Käufe oder mehr Klicks) zu optimieren. Eine Kombination der Daten zu etwas, das wenig Aussicht auf Erfolg (die abhängige Variable) hat, wird eine KI nicht präferieren und daher nicht anbieten. Genau das tut aber ein Mensch, wenn er kreativ ist. Mangels eines perfekten Algorithmus geht der Mensch mehr ins Risiko, probiert Dinge aus, kombiniert Daten zu etwas Neuem, das aller Wahrscheinlichkeit nach weniger Erfolg hat als etwas, das bereits ausprobiert wurde. Aber eben darin liegt das Überraschende, das Neue, das, was die Kreativität ausmacht. Ein Mensch kann auf Gefühle von Menschen oder Menschengruppen eingehen, sie lesen und darauf mit seiner Kommunikation reagieren – oder aus Gründen der Pietät darauf absichtlich verzichten (S.209).

Bezüglich der **Strategie**:

Auch hier wird KI auf lange Sicht den Menschen nicht vollständig ersetzen, allenfalls unterstützen können. Wie bereits erwähnt, werden vermutlich in absehbarer Zeit nicht genug Daten veröffentlicht werden, um einer KI zu ermöglichen, eine bessere Unternehmensstrategie, Markenstrategie oder Marketing-Strategie zu entwickeln, als es ein Experte oder eine Expertin das könnte. Daher bleibt hier der Mensch immer noch überlegen. Außerdem sollte eine Firma ihre Strategie, das Herz des Unternehmens, nicht ohne Not gänzlich abgeben (S.209).

Die 4-Felder Matrix impliziert eine grundlegende Aufgabenverteilung für das zukünftige HAT. Die KI, die stellvertretend für den autonomen Agenten steht, übernimmt im Teaming mit dem Menschen vor allem die Aufgaben, die keine Empathie oder Mitgefühl erfordern. Außerdem ist die KI dafür prädestiniert, repetitive und routinemäßige Aufgaben auszuführen. Der Mensch hingegen wird vor allem die Aufgaben übernehmen, die strategisch und kreativ sind, sowie Empathie und Mitgefühl voraussetzen.

Zusammenfassend zeigen die zuvor behandelten Ansätze, dass der Mensch im HAT aufgrund seines Urteilsvermögens, seiner Kreativität, seiner Fähigkeit, Empathie zu zeigen und durch sein strategisches Denken in Zukunft weiterhin eine wichtige Rolle im Teaming spielen wird. Der Mensch wird in Zukunft immer häufiger mit autonomen Agenten interagieren und in der Zusammenarbeit vor allem die Aufgaben übernehmen, die Urteilsvermögen, Entscheidungsfindung, Kreativität, Empathie und strategisches Denken erfordern

5.2 Verbesserungsmöglichkeiten für das Human Autonomy Teaming der Zukunft

Der aktuelle Entwicklungsstand im Bereich des HAT ist noch lange nicht ausgereift. So zeigen beispielsweise O'Neill et al. (2020), dass heutzutage rein menschliche Teams die Leistung von

HATs meist weit übersteigen. In den Bereichen der technischen und der Integrationsperspektive wird weitere Forschung benötigt, um die vielversprechenden Potenziale von HAT zukünftig gänzlich nutzen zu können. In diesem Kapitel werden die Verbesserungspotentiale im Bereich des HAT dargestellt und der daraus resultierende Forschungsbedarf aufgezeigt.

5.2.1 Level an Autonomie des autonomen Agenten

Mit steigender Autonomie des autonomen Agenten wird es mehr Möglichkeiten geben, die verschiedenen Level an Autonomie des autonomen Agenten effektiv einzusetzen. Bislang wurde nicht hinreichend erforscht, inwiefern eine höhere Agentenautonomie in welchem Kontext sinnvoll und hilfreich eingesetzt werden kann. O'Neill et al. (2020) empfehlen weiter zu erforschen, unter welchen Bedingungen ein hoher Grad an Autonomie des autonomen Agenten hilfreich oder schädlich für die Aufgabe wird. Des Weiteren erkennen O'Neill et al. (2020) weiterhin einen Forschungsbedarf im Bereich des anpassungsfähigen autonomen Agenten. Der autonome Agent sollte in der Lage sein, die Beteiligung im HAT nach der gegebenen Situation anzupassen. Das bedeutet, sich an den menschlichen Anforderungen zu orientieren und im Zuge der auszuführenden Aufgabe, die Autonomie entsprechend der Situation dynamisch zu verändern.

Zwei mögliche Konzepte der Entscheidungsfindung bestehen im HAT. Das erste Szenario ist die geringe Autonomie des autonomen Agenten, welche vom menschlichen Teammitglied überwacht wird und Befehle erhält. Hierbei wird nicht das vollständige Potenzial des HAT genutzt. Die sensorischen, Verarbeitungs- und Argumentationsfähigkeiten des autonomen Agenten werden durch die befehlende und begrenzte Aufgabe beschränkt. Dadurch kann der autonome Agent nur nach den Befehlen und Anweisungen des menschlichen Teammitglieds handeln. Es werden keine Entfaltungsmöglichkeit für den autonomen Agenten gelassen (Azhar & Sklar 2017). Das zweite Szenario wäre ein tatsächliches HAT, in dem sowohl Mensch als auch autonomer Agent gleichberechtigt zusammenarbeiten und zusammen zu Lösungen gelangen. Die Verantwortung für die Aufgabe wird von den Beteiligten geteilt, Diskussionen über Aktionen geführt und Gründe gegen Maßnahmen gesucht, bis die Aufgabe beendet wird (Azhar & Sklar 2017). Anhand dieser beiden Möglichkeiten der Zusammenarbeit des HAT sollte erforscht werden, inwiefern höhere Autonomie des autonomen Agenten sinnvoll und zielführend ist.

Von Bedeutung ist die Erforschung der Frage, wann die Autonomie für die jeweilige menschliche Anforderung im Prozess der Aufgabe angepasst werden sollte. Bisher wird in Situationen der städtischen Such- und Rettungsmissionen, der Minenräumung und der Inspektion von Kernkraftwerken eine kollaborative Interaktion mit geringerer Autonomie des autonomen Agenten und mit der Anleitung des Menschen genutzt. Autonome Agenten werden in Zukunft im häuslichen Bereich und im Gesundheitswesen vermehrt eingesetzt. Beispielfähig als Pflegekräfte im Privathaushalt oder in häuslichen Pflegeeinrichtungen. Die kollaborative Interaktion im HAT mit der Gleichberechtigung in der Entscheidungsfindung zwischen den Teammitgliedern würde den Austausch von Informationen, die Minimierung von Unsicherheit und die Erweiterung von individuellem und gemeinsamem Wissen ermöglichen. Es wird ein Rahmen geschaffen, in dem sich das menschliche Teammitglied und der autonome Agent gegenseitig herausfordern und den anderen von der eigenen Entscheidung überzeugen müssen. Die Studie von Azhar und Sklar (2017) zeigt, dass die Kollaboration im Team zwischen Menschen und autonomen

Agenten (HAT) bessere objektive Leistungen hervorbringt, als das Team mit dem Menschen als Überwacher/in des autonomen Agenten. Die gemeinsamen Ziele des Menschen und des

autonomen Agenten entsprechen noch nicht der potentiellen Komplexität. In Zukunft werden die Welt und die damit verbundenen Aufgaben komplexer sein. Kausalzusammenhänge werden schwerer nachvollziehbar sein. Azhar und Sklar (2017) schlagen eine Unterstützung des Dialogs zwischen Menschen und autonomen Agenten für komplexere Aufgaben mit gemeinsamen und unterschiedlichen Zielen vor.

Es wird von Vorteil sein, eine Kultur innerhalb des HAT zu etablieren, in dem der Mensch und der autonome Agent einen auf Argumentation basierenden Dialog führen. Der auf Argumentation basierende Dialog vereint die Stärken der Teammitglieder im HAT und ermöglicht eine bessere Zusammenarbeit. Azhar und Sklar (2007) benennen mangelndes Wissen, schnelle Entscheidungen und dynamische Umgebungen als Austauschpunkte im HAT. Menschen und autonome Agenten gelangen unterschiedlich an Wissen. Beim autonomen Agenten sind vor allem die analytischen Fähigkeiten mithilfe von Datensätzen vorteilhaft, welche schnell und effizient ausgewertet werden können. Der Mensch kann sich durch seine kognitiven Fähigkeiten an neue und dynamische Umgebungen anpassen (Fan et al., 2010). Ein Beispiel für eine sich dynamisch verändernde Umgebung wäre ein Hindernis, welches im Zuge der Aufgabe nicht kalkulierbar auftritt und den Arbeitsprozess verändert. Aus diesem Grund verlangen O'Neill et al. (2020) eine dynamische Anpassung des Grades an Autonomie vom autonomen Agenten entsprechend der menschlichen Anforderungen. Dies wäre eine Möglichkeit, um Problematiken im Aufgabenprozess zu lösen.

5.2.2 Transparenz und Zuverlässigkeit

Durch die zunehmende Etablierung des HAT in der Arbeitswelt werden zukünftig vermehrt unbemannte autonome Agenten eingesetzt und überwacht. Cummings und Mitchell (2008) sowie Lewis (2013) erwarten zukünftig die Überwachung mehrerer unbemannter Systeme durch eine einzelne Person. Dies kann, je nach Komplexität der Aufgabe und dem Grad an Kontrolle, die Überwachungsfähigkeiten des menschlichen Teammitglieds überschreiten. Die Ausführungen des autonomen Agenten sind für die überwachende und steuernde Person bzw. das menschliche Teammitglied nicht vollständig transparent überschaubar. Der Arbeitsaufwand sowie die Zuverlässigkeit des HAT würden dadurch negativ beeinflusst werden. Mercado et al. (2016) schlagen vor, zu untersuchen, inwiefern sich Unsicherheit auf den einzelnen Transparenzebenen auf die Leistung, das Vertrauen und die Arbeitsleistung der Betreiber auswirkt.

Weitere Entwurfsstrategien zur Verringerung von „Complacency“ und Arbeitsbelastung sollten entworfen werden. Bei dem Begriff Complacency handelt es sich um den Grad an Selbstzufriedenheit, welcher durch die Erfüllung der Aufgabe beim menschlichen Teammitglied innerhalb des HAT entsteht. Beispielhaft nennen O'Neill et al. (2020), die Anforderungen an den Benutzenden zu verringern und das menschliche Teammitglied stattdessen einfache und typische Aufgaben ausführen zu lassen. Im Laufe der Zeit wird durch Schulungen, Cross-Training und Auffrischungsschulungen die Transparenz für die überwachende Person der unbemannten Systeme wiederhergestellt. Weitere Forschung sollte im Bereich der Auswirkungen der unterschiedlichen Formen und Anwendungen von autonomen Agenten auf die situative Aufmerksamkeit des menschlichen Teammitglieds durchgeführt werden (Wright & Kaber, 2005).

5.2.3 Aufgabenmerkmale

Der vielseitige Einsatz des HAT wird in Zukunft in verschiedenen Wirtschafts- und Aufgabebereichen eingesetzt werden. Wie bereits im Kapitel 5.2.1 angeschnitten wurde, haben Mensch und autonomer Agent bestimmte Stärken in der Zusammenarbeit des HAT. Im

folgenden Abschnitt wird dieser Punkt genauer untersucht und auf die Merkmale der Aufgabe angewandt.

Das HAT wird die Arbeitsweise der Zukunft genannt, da sich die Stärken des Menschen und des autonomen Agenten gegenseitig ergänzen und eine bessere Zusammenarbeit resultiert. Autonome Agenten weisen Stärken bei rechenintensiven Tätigkeiten, wie der Informationssuche und der Filterung, auf. Menschen zeigen Stärken bei der Projektion und Argumentation in dynamischen Situationen und besitzen die bessere Anpassungsfähigkeit an die kognitiven Fähigkeiten der Teammitglieder (Fan et al., 2010). Die Stärken des autonomen Agenten sollten gezielt eingesetzt werden, um die Arbeitsbelastung des Menschen zu reduzieren. O'Neill et al. (2020) empfehlen weitere Untersuchungen bezüglich der Frage, inwiefern autonome Agenten bestimmte Aufgaben des Menschen übernehmen können, um die Arbeitsleistung situationsgemäß anzupassen. Die Arbeitsleistung des Menschen wird gezielt reduziert und die Arbeitsleistung und Zufriedenheit innerhalb des HAT wird gesteigert. Forschungsbedarf wird bei der Entwicklung des kognitiven Agenten verortet. Durch zukünftige Forschung im Bereich der künstlichen Intelligenz wird der autonome Agent höhere kognitive Fähigkeiten erlangen. Die sinnvolle und effektive Aufgabenverteilung wird einer Veränderung unterliegen und neue Herausforderungen entstehen lassen. Fan et al. (2010) sehen Forschungsbedarf in den voraussetzenden Vertrauensbeziehungen zwischen dem kognitiven autonomen Agenten und dem Menschen für eine verbesserte Teamleistung. Kaber und Wright (2005) erkennen Forschungsbedarf in Bezug auf die Erforschung der Auswirkungen von verschiedenen kognitiven Formen des autonomen Agenten auf die Team-Koordination und Arbeitsleistung.

5.2.4 Teamzusammensetzung

HAT zeichnet sich durch das Teaming von Menschen und autonomen Agenten aus. Die beiden Komponenten haben die Möglichkeit, als Team effektiver zu sein als die Summe ihrer Teile (vgl. Kapitel 5.1). Eine Verbesserungsmöglichkeit für das HAT der Zukunft stellt die Teamzusammensetzung dieser beiden Komponenten dar. Es besteht weiterhin Forschungsbedarf darin, die beste Art der Gruppenkonstellation zwischen Menschen und autonomen Agenten zu identifizieren. Dieses Unterkapitel gewährt einen Einblick in das Thema Teamzusammensetzung, den momentanen Forschungsstand und gibt eine Empfehlung für die Zukunft.

Teamzusammensetzung bezieht sich auf das Verhältnis von Menschen und autonomen Agenten im Team (O'Neill et al., 2020). Die Literatur über rein menschliche Teams legt nahe, dass die Teamstruktur eine effektive Art ist, wie Menschen sich organisieren und Aufgaben bewältigen (Walliser et al., 2019). Teams sind soziale Strukturen, die sich durch eine hohe gegenseitige Abhängigkeit der Teammitglieder und einem gemeinsamen Ziel auszeichnen (Salas et al., 2008). Teamzugehörigkeit steht in Verbindung mit hoher Kommunikation, Vertrauen und Einsatz (Abrams et al., 1990). Dieser Ansatz wird beim HAT auf die Zusammenarbeit von Menschen und autonomen Agenten übertragen. Sowohl der Mensch als auch der autonome Agent verfolgen dasselbe Ziel und übernehmen unterschiedliche Aufgaben, die eine gegenseitige Abhängigkeit zur Erreichung des gemeinsamen Ziels aufweisen.

Nach jetzigem Forschungsstand werden HATs hinsichtlich der Performance fast immer von Teams, die nur aus Menschen bestehen, übertroffen (McNeese et al., 2018). Menschliche Teams zeigen im Gegensatz zu HATs einen positiven Lerneffekt (Cooke et al., 2016). Sie können sich besser reorganisieren und anpassen (Grimm et al., 2018). Ein möglicher Grund, wieso rein menschliche Teams besser performen als Human Autonomy Teams könnte die Kommunikation sein. Rein menschliche Teams verfügen über einen effizienteren

Informationsaustausch (O'Neill et al., 2020). Es werden mehr Statusupdates geliefert und schon mitgeteilte Informationen müssen weniger oft wiederholt werden (Demir & McNeese, 2018). Für zukünftige Forschungen im Bereich des HAT sollte die Qualität des Informationsaustausches und der Kommunikation berücksichtigt werden. Zukünftige Forschungen sollten sich mit den Effekten der Teamgröße und Zusammensetzung beschäftigen. Wie die Teamstruktur Einfluss auf die Annahme von Empfehlungen der autonomen Entscheidungshilfen nehmen kann, ist ebenfalls eine wichtige Fragestellung der Zukunft (Walliser et al., 2019).

Viele Forschungsprojekte im Bereich des HAT beschäftigten sich derzeit mit der Verbesserung der Kommunikation zwischen Menschen und autonomen Agenten mithilfe von Visualisierung (Demir et al., 2016). Laut De Visser et al. (2018) soll Human Autonomy Interaction menschliche Modelle als Standard adaptieren. Ein wichtiges menschliches Kommunikationsmodell ist die Sprache. Es erscheint sinnvoll, die Sprache auf die Kommunikation zwischen Menschen und autonomen Agenten anzuwenden. Diesen Gedankengang bestätigt eine Studie von Bogg et al. (2020). Ein sprechender autonomer Agent führt zu besserer Performance, kann den Verlust des Situationsbewusstseins verringern und fördert die subjektive Wahrnehmung des Teamings beim Menschen (Bogg et al., 2020). Das Hinzufügen einer synthetischen Stimme wird dem menschlichen Operator aufgrund der eben genannten Punkte scheinbar sehr behilflich sein. Die Zusammensetzung eines HAT aus einem sprechenden autonomen Agenten und einem Menschen könnte dafür sorgen, dass die Performance gesteigert wird. HATs könnten so in Zukunft besser mit menschlichen Teams mithalten.

5.2.5 Training

Eine weitere Verbesserungsmöglichkeit stellt das Training dar. Hierbei geht es um die systematische Verbesserung der Zusammenarbeit von Menschen und autonomen Agenten. Das Unterkapitel gibt einen Einblick in die momentane Forschung und einen Ausblick auf die zukünftige Forschung im Bereich Training. Es wird aufgezeigt, wie wichtig gezieltes Training für die Zukunft des HAT sein kann.

Training bietet für den Menschen und für den autonomen Agenten die Möglichkeit, gemeinsam mit den anderen Teammitgliedern zu trainieren, zu üben und zu lernen. Training sollte beispielsweise auf Wissen, Fähigkeiten und Fertigkeiten abzielen, um diese Bereiche zu verbessern (O'Neill et al., 2020). Es ist wichtig, dass das Training mit einer fundierten Bedarfsanalyse oder Aufgabenanalyse verknüpft wird (Arthur et al., 2005). So können die Rollen jedes Teammitglieds, wesentlichen Abhängigkeiten sowie Fähigkeiten, die erforderlich sind, um individuelle Rollen auszuführen und voneinander abhängige Aktivitäten zu koordinieren, ermittelt werden (O'Neill et al., 2020). Die bestehende Forschung zeigt positive Effekte von Training. So produzierte das Training von Menschen mit autonomen Agenten konstant positive Resultate, wie gestiegenes Vertrauen und geringere Unsicherheiten. Wenn es dem Menschen möglich war, vor der Aufgabe mit dem autonomen Agenten zu trainieren, wurde ein höheres Lerntempo festgestellt (Cohen & Imada, 2005). Menschen und autonome Agenten hatten ein besseres Gespür für gegenwärtige Abläufe. Dies könnte dazu führen, dass die Ausführungen des Teampartners besser antizipiert werden und Aufgaben besser koordiniert werden können (O'Neill et al., 2020). In der Zukunft wird es wichtig sein, bestehende Kenntnisse aus dem menschlichen Teamtraining auf das HAT zu übertragen. Dazu wird es notwendig sein, den spezifischen Trainingsbedarf des Menschen und des autonomen Agenten zu identifizieren. Es müssen Methoden gefunden werden, wie die Teams anhand der neuen Erkenntnissen trainiert werden können (O'Neill et al., 2020). Maßnahmen, die das Teamtraining betreffen, könnten einer der wesentlichen Inputs sein, um „Human Autonomy Teamwork“ zukünftig positiv zu beeinflussen. Meta-Analysen zeigen deutlich auf, dass Teamtraining sehr erfolgreich auf vielen

verschiedenen Feldern sein kann, z.B. in der Medizin oder Luftfahrt (O'Neill et al., 2020). Beides sind Bereiche, in denen in Zukunft das Human Autonomy Teaming eine große Rolle spielen wird.

5.2.6 Menschliche individuelle Unterschiede

Der Mensch ist aufgrund seiner Individualität nicht auf ein bestimmtes Verhalten reduzierbar. Menschen kommunizieren unterschiedlich, bedingt durch ihre Persönlichkeit, Kultur oder dem Alter. Erikson (2019) benennt vier Persönlichkeitstypen, welche das Verhalten und die Kommunikationsweise des Menschen leiten. Unterschieden wird in vier Dimensionen, aus denen der Persönlichkeitstyp des Menschen entsteht. Eine Unterscheidung besteht zwischen introvertierten und extrovertierten Menschen. Introvertierte Personen werden als passiv und zurückhaltend charakterisiert. Extravertierte Personen sind aktiv und leiten bzw. initiieren Kommunikation. Eine weitere Unterscheidung besteht zwischen beziehungsorientierten und aufgabenorientierten Persönlichkeiten. Beziehungsorientierte Menschen legen mehr Fokus auf die Beziehung zu anderen Menschen, wohingegen aufgabenorientierte Individuen sich auf die Aufgabe konzentrieren. Die Abbildung 5.2 stellt die Zusammensetzung der verschiedenen Dimensionen dar. Die jeweilige Dimension wird durch die Form und Farbe in der Tabelle wiedergegeben. Zum Beispiel steht der rote Kreis für eine extravertierte Person, welche beziehungsorientiert ist. Die Möglichkeit besteht, sich als Individuum in mehreren Dimensionen zu befinden. Ungefähr 80 Prozent der Menschen haben eine Kombination aus zwei Dimensionen, fünf Prozent eine und die restlichen 15 Prozent haben drei Persönlichkeitsdimensionen (Erikson, 2019). Das bedeutet, es besteht die Möglichkeit, sich gleichzeitig in den Dimensionen „roter Kreis“, „lila Dreieck“ und „blaues Viereck“ zu befinden. Eine Kombination aus allen vier Dimensionen ist nicht möglich.

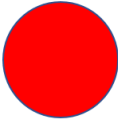


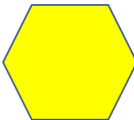
	Extrovertiertheit	Introvertiertheit
Beziehungsorientierung		
Aufgabenorientierung		

Abb. 5.2: Dimensionen der Persönlichkeit (in Anlehnung an Erikson, 2019, S.12).

Zu den vier Dimensionen der Persönlichkeit von Erikson (2019) gibt es noch nicht hinreichende Forschungsergebnisse, um klare Schlussfolgerungen zu ziehen. Einer der wenigen Studien zu der Thematik von Hanna und Richards (2015) zeigt, dass Introversion sich positiv auf die Entwicklung der Aufgaben- und Teamarbeit auswirkt. Zu den „Big 5“

Persönlichkeitseigenschaften im Team existieren bereits Ergebnisse (vgl. Kapitel 3.2.2). Die Big 5 Persönlichkeitseigenschaften beinhalten Aufgeschlossenheit, Perfektionismus, Geselligkeit, Empathie und emotionale Labilität (Mangel an emotionaler Stabilität). Dennoch wird weitere Forschung benötigt, um weitere klare Erkenntnisse aus dem Zusammenhang zwischen Teamleistung und Persönlichkeitsmerkmalen zu erlangen.

Die Übereinstimmung von Introversion zwischen dem menschlichen Teammitglied und dem autonomen Agenten wirkt sich positiv auf das Leistungsergebnis aus (Hanna & Richards, 2015). Die Übereinstimmung des Verständnisses der zu absolvierenden Aufgabe und der darin involvierten Teamarbeit zwischen den Teammitgliedern wird „Shared Mental Model“ (SMM) genannt. Das SMM wird durch die Art und Weise der Kommunikation und durch die Kommunikation bedingt von den Persönlichkeitsmerkmalen beeinflusst (Hanna & Richards, 2015).

O'Neill et al. (2020) sehen Forschungsbedarf in der Entwicklung von anpassungsfähigen autonomen Agenten an die Persönlichkeiten und Arbeitsstile der menschlichen Teammitglieder. In Zukunft sollte vermehrt darauf geachtet werden, wie autonome Agenten entwickelt und programmiert werden müssen, um bestimmten Persönlichkeiten und Arbeitsstilen optimal zu entsprechen. Maßnahmen werden benötigt, die sich auf die spezifische Kommunikation beziehen (Wright & Kaber, 2005).

5.3 Soziale und ethische Dilemmata der Zukunft

Arbeiten im Human Autonomy Teaming wird zukünftig vermehrt Teil der Gesellschaft. Im Alltag und in der generellen Entscheidungsfindung spielen soziale und ethische Kontexte eine zentrale Rolle. Der vermehrte Einsatz von HAT beinhaltet soziale und ethische Dilemmata. Es gilt zu klären, wie mit sozialen und ethischen Dilemmata umzugehen ist und was der Mensch tun kann, um diese zu verringern oder bestenfalls zu verhindern. Im folgenden Kapitel wird die moralische Entscheidungsfindung im HAT beschrieben, der autonome Agent im gesetzlichen Kontext betrachtet und Risiken und Herausforderungen des HATs behandelt.

5.3.1 Die moralische Entscheidungsfindung

Mit steigender Autonomie des autonomen Agenten werden diesem zukünftig mehr Entscheidungen überlassen. Die Entscheidungen des Menschen werden von moralischen Prinzipien geleitet. Es wäre sinnvoll, dem autonomen Agenten durch entsprechende Programmierungen moralisches Denken und Handeln zu ermöglichen.

Im folgenden Unterkapitel werden zwei Ansätze beschrieben, die darlegen, inwiefern eine moralische Entscheidungsfindung in den autonomen Agenten integriert werden könnte. Im ersten Abschnitt geht es um den Künstlichen Moral Agenten (KMA), der aufgrund von interner Entscheidungsfindung moralische Prinzipien anwendet. Im zweiten Abschnitt wird die Programmierung von moralischen Prinzipien in ein autonomes Fahrzeug behandelt und die unterschiedlichen Auffassungen von moralischen Prinzipien im Kontext zu der Thematik beschrieben.

Der Künstliche Moral Agent

Die Idee des Künstlichen Moral Agenten (KMA) basiert auf dem Verhalten des autonomen Agenten, welches von moralisch festgelegten Standards geleitet wird. KMA führen menschlich moralische und ethische Abwägungen in der Entscheidungsfindung durch und haben eine moralische Verpflichtung gegenüber Dritten. Einem Faktor der Idee des KMA, der Verantwortung für Verhalten, schreibt Himma (2009) eine besonders hohe Bedeutung zu. Die Verantwortung

für das Verhalten ist essenziell, da für gutes und schlechtes Verhalten Konsequenzen folgen müssen. Der autonome Agent soll die Fähigkeit der Emotionalität in die Entscheidungsfindung einbeziehen und somit Teil des menschlichen Wertesystems werden.

Um das Verständnis und die Wichtigkeit eines moralischen autonomen Agenten zu erlangen, werden die Eigenschaften des Menschen und des autonomen Agenten genauer betrachtet. Der autonome Agent denkt regelbasiert und reflektiert Verhalten nicht. Emotionalität, Verstand und Bewusstsein sind nicht vorhanden. Es besteht weder ein Verständnis von komplizierten sozialen Kontexten noch von der beschränkten Wahrnehmung der programmierten, gewünschten Funktion (Precht, 2020). Die genannten Eigenschaften sind grundlegend für die Bedeutung von menschlicher Auffassung von Moral und Gerechtigkeit.

Warum wird der autonome Agent nicht wie ein Mensch programmiert, wenn Schwächen in der Auffassung von moralischen Prinzipien bestehen? Die Komplexität der technischen Durchführung macht es zurzeit unmöglich, den autonomen Agenten nach dem menschlichen Vorbild zu programmieren. Weitreichende und intensive Forschung vor allem im Bereich der

künstlichen Intelligenz wird benötigt, um eine menschliche Auffassung von moralischen Prinzipien zu programmieren und anzuwenden. In der Infobox 5.1 wird darauf eingegangen, welche Herausforderungen trotz der technologischen Möglichkeit bei der Programmierung von moralischen Prinzipien auftreten könnten.

Infobox 5.1: Verzerrte Urteilsfindung des Menschen.

Menschen sind in ihrer rationalen Entscheidungsfindung begrenzt. Dies bedeutet, dass Gedankengänge und Beurteilungen durch Verzerrungen der Wahrnehmung zu falschen Schlussfolgerungen führen. Kahnemann (2011) beschrieb mit der Einführung der Systeme 1 und 2, wie Menschen zu falschen Urteilen gelangen. Bei System 1 handelt es sich um das automatische, schnelle und einfache Denken in der menschlichen Urteilsfindung. Das automatische und schnelle Denken von System 1 machen es dem Menschen möglich, unbewusst, mühelos und intuitiv Tätigkeiten auszuführen. Beispielhaft für Tätigkeiten der Denkweise von System 1 sind einfache Rechenaufgaben ($2+2=?$) oder den PIN von der EC-Karte an der Kasse einzugeben. Das System 2 ist für das langsame und konzentrierte Denken des Menschen verantwortlich. System 2 ermöglicht bewusste Entscheidungen und ist die Identifizierung mit dem Menschen selbst. Es stellt das das Bewusstsein im Menschen dar, welches Prinzipien hat, Entscheidungen trifft und Gedankengänge interpretiert. Beispielhaft für Tätigkeiten unter dem Denken von System 2 sind komplizierte Rechenaufgaben ($17 \times 24=?$) oder der Vergleich zweier Waschmaschinen nach ihrem Wert.

Das schnelle und intuitive Beurteilen von System 1 und die arbeits- und konzentrationsintensive Verarbeitung von System 2 führen zu Urteilsverzerrungen in beiden Systemen. Der KMA würde die gleichen Urteilsfehler aufweisen, wie der Mensch. Zweckführend wäre dies nicht. Precht (2020) beschreibt eine nicht neutrale Wahrnehmung der Welt des Menschen aufgrund der Verzerrung durch emotionale Filter. Fehler in der moralischen Auffassung werden erscheinen, welche durch situative Faktoren beeinflusst werden.

Für die bessere Nachvollziehbarkeit der situativen Faktoren wird sich vorgestellt, dass fünf Menschen auf einem Gleis auf Eisenbahnschienen festgebunden sind. In einer Minute kommt ein Zug und würde die fünf Menschen überfahren. Die Option besteht, die Schienenweichen umzustellen, um den Zug auf eine anderes Gleis umzuleiten. Auf dem anderen Gleis arbeitet ein Schienenarbeiter und führt Wartungsarbeiten durch. Dieser hat Ohrenschützer auf und ist auf die Wartungsarbeit fokussiert. Der Arbeiter nimmt nicht wahr, was in seinem Umfeld passiert und reagiert auf keine Signale. Der Zug wird den Schienenarbeiter nach der Umstellung der Gleise überfahren. Der Bremsweg des Zuges ist zu lang, um vor dem Schienenarbeiter anzuhalten. Aus rationaler Sicht würden die Weichen umgestellt werden, da mehr Menschenleben gerettet werden würden. Komplizierter wird es, wenn die einzelne Person ein enger Verwandter ist. Dort würde die Emotionalität des Menschen eingreifen. Die Umstellung der Weichen wird moralisch angezweifelt, da zu der verwandten Person eine emotionale Bindung besteht.

Dieses Beispiel verdeutlicht die Komplexität der menschlichen Urteilsfindung und zeigt, dass dabei Urteilsverzerrungen existieren. Selbst wenn die menschliche Urteilsfindung unter Berücksichtigung der menschlichen Moral in eine autonomen Agenten programmieren könnte, würde dies ebenfalls zu den gleichen Urteilsverzerrungen führen. Dies wäre für das HAT nicht sinnvoll, da die beabsichtigte Ergänzung von Stärken nicht erfolgen würde (vgl. Kapitel 5.2.1).

Der autonome Agent benötigt zwei Voraussetzungen für die moralische Entscheidung. Der KMA muss in der Lage sein, Verhalten frei zu wählen und den Unterschied zwischen richtigem und falschem Verhalten zu erkennen (Himma, 2009). Die essenzielle Voraussetzung von der

freien Wahl des Verhaltens, ist die Fähigkeit, das Verhalten mit Bedacht, Überlegung und Absicht zu vollziehen (Himma, 2009). Das bedeutet, dass Bewusstsein dafür zu haben, was gerade der Tatbestand ist und welche Konsequenzen aus den Taten folgen. Bei der Erkennung von richtigem und falschem Verhalten setzt Himma (2009) das moralische Urteilsvermögen voraus. Das bedeutet, moralische Konzepte zu verstehen, im Alltag anzuwenden und relevante von nicht relevanten Regeln zu unterscheiden. Diese emotionale Verarbeitung von Informationen ist essenzieller Bestandteil im KMA (Farisco et al., 2020).

An dieser Stelle ließe sich die Frage stellen, ob und wie eine solche moralische Auffassung in den autonomen Agenten programmiert werden kann. Mit zukünftiger Forschung im Bereich der künstlichen Intelligenz wird aufgeklärt werden, ob der autonome Agent moralische Entscheidungen treffen kann. Im nächsten Abschnitt wird eine alternative Methode betrachtet, um moralische Prinzipien in den autonomen Agenten zu integrieren. Diese Methode benötigt keinen hohen Grad an künstlicher Intelligenz des autonomen Agenten.

Ethische und moralische Entscheidungen beim autonomen Fahrzeug programmieren

Das autonome Fahren ist der Wunschgedanke vieler Menschen und wird vor allem in Deutschland mit der zukünftigen Digitalisierung verbunden. Automobilhersteller wie Tesla oder Daimler nutzen bereits Technologien, welche das teilweise autonome Fahren ermöglichen (vgl. 4.1.1). Berücksichtigt wird häufig nicht, inwiefern ethische und moralische Dilemmata in diesem Kontext entstehen. Das autonome Fahrzeug muss bei unvermeidbaren Unfällen entscheiden, welche Beteiligten geschützt werden sollen. Auf welcher Basis wird diese moralische Entscheidungsfindung getroffen?

In Deutschland gab es im Jahr 2017 3000 Verkehrstote und rund 400.000 Verletzte im Straßenverkehr. 90 Prozent der Verkehrsunfälle entstanden aufgrund von menschlichem Versagen (Precht, 2018). Ein autonomes Fahrzeug könnte den Verkehrsprozess effizienter gestalten, Abgase reduzieren und bis zu 90 Prozent der Verkehrsunfälle verhindern. Zur Lösung von moralischen Dilemmata müssten dem autonomen Fahrzeug moralische Prinzipien programmiert werden, welche den schwierigen Entscheidungsprozess lenken sollen (Bonneton et al., 2016). Die Frage stellt sich, welche Personengruppen das autonome Fahrzeug primär schützen soll. Die Studie von Awad et al. (2018) stellt dar, welche moralischen Prinzipien von Bedeutung sind. Die Ergebnisse zeigen die Präferenzen, Menschen vor Tieren, junge vor alten Menschen zu retten und mehr Leben anstelle eines einzelnen zu verschonen. Abbildung 5.3 zeigt die allgemeinen anerkannten Präferenzen von moralischen Prinzipien. Die Einigung besteht, den Insassen des autonomen Fahrzeugs zu opfern, um dafür mehr Menschen zu retten (Bonneton et al., 2016). Generell stieg die moralische Akzeptanz, den Insassen des automatisierten Fahrzeugs zu opfern, mit der steigenden Anzahl an potenziellen geretteten Personen. Eine utilitaristische Präferenz wird gezeigt. Das bedeutet, jemanden für den Nutzen des Gemeinwohls zu opfern (in diesem Kontext eine größere Anzahl an Personen). Die Teilnehmenden der Studie empfehlen anderen Menschen ein moralisch programmiertes Fahrzeug zu erwerben und zu nutzen.

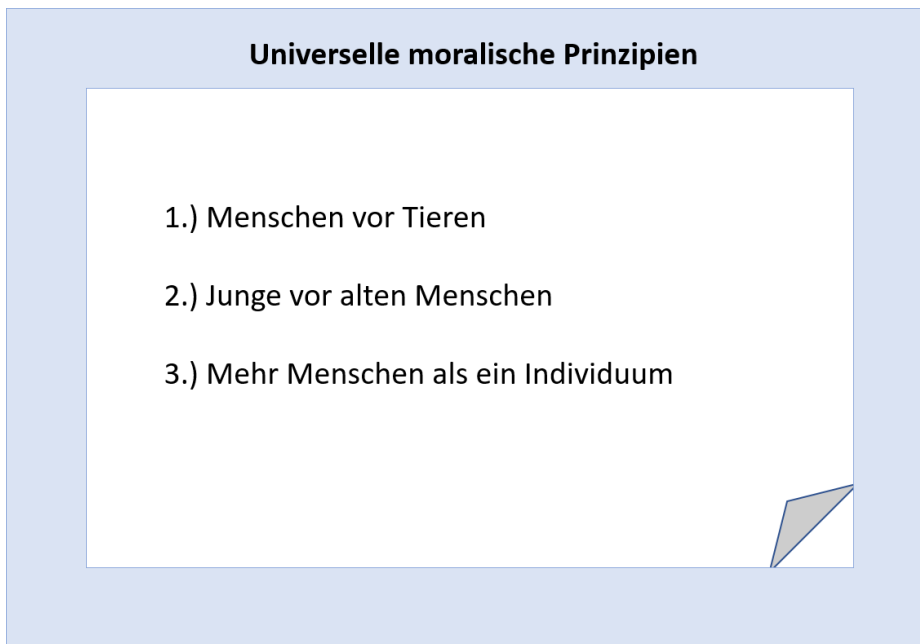


Abb. 5.3: Universelle Präferenzen von moralischen Prinzipien.

Menschen verhalten sich nicht immer rational, wie bereits im Abschnitt 5.3.1 und in Infobox 5.1 beschrieben. Denn es gibt bei den Teilnehmenden keine moralische Präferenz, das utilitäre autonome Fahrzeug zu kaufen und zu nutzen. Besonders wird diese Meinung vertreten, wenn Familienmitglieder das autonome Fahrzeug benutzen sollen. Teilnehmer der Studie sind gegen Vorschriften, welche utilitäres Verhalten in autonomen Fahrzeugen unterstützen (Bonneton et al., 2016). Die Resultate der Studie sind von großer Wichtigkeit, da mehr Informationen über die Thematik der sozialen Dilemmata bei autonomen Fahrzeugen vorliegen. Aufsichtsbehörden und Gesetzgeber können die Resultate nutzen, um die Basis für Vorschriften beim autonomen Fahren zu etablieren. Wichtig wird es sein, Vorschriften und Richtlinien vor dem Einführen des autonomen Fahrzeugs festzulegen. Dadurch werden moralische Dilemmata präventiv statt reaktiv gelöst.

Auf der Welt gibt es viele verschiedene Kulturen, weshalb verschiedene Auffassungen von Moral existieren. Die Präferenz, jüngere statt älterer Menschen zu retten, ist in östlichen Ländern weniger ausgeprägt, als in südlichen. Das Gleiche gilt für die Präferenz, Personen mit hohem Status zu retten. Südliche Länder haben im Vergleich zu den westlichen und östlichen Ländern, eine schwächere Präferenz, Menschen anstelle von Haustieren zu retten. In südlichen Ländern herrscht hingegen eine starke Präferenz, Frauen und gesunde Menschen zu retten (Awad et al., 2018). Der wichtige Faktor für den Unterschied der moralischen Auffassung ist der Kulturunterschied von individualistischen und kollektivistischen Gesellschaften. Teilnehmende von individualistischen Kulturen legen viel Wert auf das Individuum und weisen die utilitaristische Präferenz auf. Das bedeutet, einen Menschen zu opfern, um mehr Menschen zu retten. Teilnehmende von kollektiven Kulturen, welche besonders auf die älteren Mitglieder achten, zeigen eine geringere Präferenz jüngere Menschen in moralischen Entscheidungen zu bevorzugen (Awad et al., 2018)

Die eben genannten unterschiedlichen Ansichten von Moral und Gerechtigkeit verdeutlichen die Schwierigkeit, universelle globale Ethiken für das autonome Fahrzeug zu entwickeln. Der ausschlaggebende Punkt für den Kauf eines autonomen Fahrzeugs wird sein, inwiefern die ethischen und moralischen Regeln in der Programmierung des Fahrzeugs, der Auffassung von Gerechtigkeit des Individuums entsprechen (Awad et al., 2018).

5.3.2 Autonomer Agent im Gesetz

Mit mehr Einsatz von HAT und dem höheren Grad an Autonomie des autonomen Agenten stellt sich die Frage, wer für Fehler des autonomen Agenten rechtlich verantwortlich gemacht wird. Die Hauptproblematik in diesem Szenario besteht darin, dass kaum Erfahrungen diesbezüglich gesammelt wurden (Haagen & Lohmann, 2020). Wichtig wird es sein, den rechtlichen Gesetzesrahmen festzulegen, bevor der autonome Agent vermehrt im menschlichen Alltag integriert und genutzt wird. Im folgenden Unterkapitel wird auf die rechtliche Verantwortlichkeit des autonomen Agenten eingegangen. Hierbei werden die Produzentenhaftung und die Produkthaftung mit einbezogen.

Produzentenhaftung

Hartmann (2020) schreibt dem Hersteller des autonomen Agenten die rechtlichen Folgen für das Fehlverhalten des Agenten zu. Voraussetzungen sind die Verletzung der Verkehrspflichten des Herstellers. Die Verkehrspflichten werden in den Bereichen Konstruktion, Entwicklung, Fabrikation, Instruktion, Produkt- und Marktbeobachtungen vorausgesetzt. Sollte in einem von diesen Bereichen des Produktionsprozesses ein Fehler auftreten, trägt der Hersteller die Haftung für den Fehler. Dies kann beispielsweise bei der Entwicklung des Bauplans in der Konstruktion oder durch fehlende Warnhinweise in der Gebrauchsanleitung in der Instruktion geschehen. Bei der Herstellung eines Produktes wird ein bestimmter Sicherheitsstandard erwartet. Der Bundesgerichtshof bestimmt den Sorgfaltsmaßstab als Voraussetzung, dass ein Produkt mit künstlicher Intelligenz auf dem Markt verkauft werden darf. Dies erfolgt nach dem entsprechend vertretbaren Risiko für das Produkt. Die Risikobewertung hängt von den Sicherheitsstandards der jeweiligen Branche ab. Sicherheitsstandards werden schwer festlegbar sein, da sich der Markt und die Technologie für autonome Agenten schnell und dynamisch entwickeln werden. Der rechtliche Rahmen sollte zukünftig gesetzt werden, um die Haftung des autonomen Agenten im Gesetz zu regeln (Hartmann, 2020).

Produkthaftung

Ein weiterer rechtlicher Aspekt ist die Produkthaftung. Im Zivilrecht ist die Thematik der Haftung Teil des Produkthaftungsgesetz (ProdHaftG). Die zentrale Haftungsnorm im ProdHaftG ist § 1 Abs. 1 Satz 1 und besagt: *„Wird durch den Fehler eines Produkts jemand getötet, sein Körper oder seine Gesundheit verletzt oder eine Sache beschädigt, so ist der Hersteller des Produkts verpflichtet, dem Geschädigten den daraus entstehenden Schaden zu ersetzen.“* Generell wird mit diesem Gesetz aufgezeigt, dass der Hersteller bei Fehlern des autonomen Agenten verpflichtet ist, den erbrachten Schaden des autonomen Agenten zu ersetzen. Der Hersteller trägt die haftende und den Schadensersatz gebende Verantwortung.

Jeder Beteiligte, der im Prozess der Herstellung des autonomen Agenten einen Teil erkenntlich beigetragen hat, wird zur Verantwortung gezogen. Hartmann (2020) betont, dass nur der Hersteller haftet, dessen Teilprodukt Fehler aufweist. Ein Lieferant eines Teils ohne Fehler haftet nicht für ein Endprodukt, welches Fehler aufweist. Um die verantwortliche Instanz für den Fehler zu finden, bedarf es zukünftig gut organisierte und klar definierte Kontrollen der Teilprodukte. Damit kann der Fehler dem richtigen Hersteller zugeordnet werden. Haagen und Lohmann (2020) betonen die erforderliche Verkehrssicherungspflicht, welche zuvor in dem Abschnitt der Produzentenhaftung aufgezeigt wurde. Neben vernünftig durchgeführter Konstruktion, Fabrikation und Instruktion muss der autonome Agent im Verkehr gründlich beobachtet werden. Das wird als Produktbeobachtungspflicht bezeichnet. Die Wichtigkeit der Produktbeobachtungspflicht besteht darin, dass im Verkehr der autonome Agent Risiken und

Probleme entstehen lassen könnte. Die Neuheit und die zukünftig steigende Komplexität erschweren die klare Nachvollziehbarkeit und Vorhersehbarkeit der Entscheidungen des autonomen Agenten. Schwer ermittelbar wird die Ursache der Schäden des autonomen Agenten, da mit höherer Autonomie weniger Kontrolle seitens des Menschen herrscht (Haagen & Lohmann, 2020). Aus den eben genannten Gründen sollte in Zukunft ein besonderer Fokus auf die Produktbeobachtungspflicht gelegt werden.

Das weitere Problem wird die Zuweisung der Haftung des Herstellers werden. Da der Hersteller meist ein Unternehmen ist und somit keine Person, müssten Angestellte oder die Führungsebene die Verantwortung tragen. Hierbei gibt es zwei Szenarien. Zum einen kann die Vorstandsvorsitzende und die geschäftsführende Person des Unternehmens belangt werden oder der/die Programmierer/in haftet für den autonomen Agenten (Haagen & Lohmann, 2020). Gesetzgeber sollten Regularien und Gesetze vor der Einführung von hochgradig autonomen Agenten auf dem Markt etablieren (Haagen & Lohmann, 2020; Hartmann, 2020).

5.3.3 Die Risiken und Gefahren des autonomen Agenten

Der Mensch hat aufgrund seiner Vorstellungsfähigkeit viele Szenarien konstruiert, in denen Roboter bzw. der autonome Agent existentielle Gefahren für die Menschheit darstellen könnten. Science-Fiction-Filme, wie Terminator oder iRobot, setzen sich mit dem Gedanken auseinander, dass autonome Agenten zukünftig stärker und intelligenter sein werden als der Mensch. Im folgenden Kapitel werden die Gefahren und Risiken des autonomen Agenten anhand von Künstlichen Waffensystemen (KWS) aufgezeigt und auf weitere Herausforderungen und Risiken des HATs eingegangen.

Künstliche Waffensysteme

HAT ist im Bereich des Militärs weit entwickelt (Sparrow, 2007) (vgl. Kapitel 4.2). Die Tatsache, dass viele von den in der Einleitung genannten Szenarien im Kontext zu Künstlichen Waffensystemen (KWS) gehören, ist nicht überraschend. Purves et al. (2015) sehen in Zukunft, dass KWS häufiger eingesetzt werden, um Entscheidungen über Leben und Tod zu treffen. Zum jetzigen Zeitpunkt haben die KWS noch nicht den notwendigen Grad an hoher Autonomie, um eigenständig Entscheidungen über Leben und Tod zu treffen. Dennoch besitzen autonome Waffensysteme bereits einen Grad an Autonomie, welcher das selbstständige Aussuchen von potenziellen Zielen ermöglicht (Sparrow, 2007).

Die Problematik der KWS ist die fehlende moralische Beurteilung, welche für Menschen in Entscheidungen über Leben und Tod essenziell für die Entscheidungsfindung ist. Dafür wird ein Bewusstsein für moralische Entscheidungen benötigt (Himma, 2009; Purves et al., 2015) (vgl. 5.3.1). Eine moralische Beurteilung von KWS ist in der nahen Zukunft eher unwahrscheinlich, aufgrund der Komplexität der Programmierbarkeit von moralischer Beurteilung (vgl. Kapitel 5.3.1). Das KWS hat nicht die Fähigkeit, richtige und falsche Ziele zu unterscheiden. Dafür benötigt das System kontextuelle Fähigkeiten, welche nach dem heutigen Stand schwer programmierbar sind. Die nicht vorhandene Intention für Leben- und Tod-Entscheidungen erschwert die richtige Entscheidung (Purves et al., 2015).

Um den eben genannten Punkt zu der richtigen Intention der Entscheidung über Leben und Tod zu erläutern, wird im Folgenden ein beispielhaftes Szenario beschrieben. Generell ist in der Gesellschaft etabliert, dass Menschen nicht getötet werden sollen. In bestimmten Szenarien wird die Todesentscheidung als moralisch und ethisch richtig empfunden. Zu dem Zeitpunkt der Terroranschläge in Paris im Jahr 2015 wurden mehrere Menschen entweder getötet oder als Geisel genommen. Der triftige Grund, den Terroristen zu töten, besteht, da dieser das

Gemeinwohl gefährdet und möglicherweise noch mehr Menschen töten könnte. In diesem Szenario gibt es einen triftigen Grund, also eine Intention, den Terroristen zu töten. Ein KWS kennt eine solche Intention nicht. Die Todesentscheidung basiert nicht auf einem triftigen Grund, sondern auf einer Programmierung oder einem Befehl. Der Befehl wird vom KWS nicht hinterfragt, da kein Bewusstsein vorherrscht (vgl. 5.3.1).

Das Beispiel zu den KWS verdeutlicht, dass HAT im Militär bereits zu ethischen Risiken und Gefahren führen kann. Die Beschreibungen des autonomen Agenten in der Einleitung zum Kapitel 5.3.3 sind im Verhältnis zu dem aktuellen Entwicklungsstand der KWS überzogen. Dennoch existieren bereits Ansätze für eine solche Entwicklung.

Herausforderungen und Risiken des HAT

Das HAT beinhaltet weitreichende Potentiale. Mit diesen Potentialen entstehen neue Herausforderungen und Risiken. Im folgenden Abschnitt wird die Thematik um die Herausforderungen und Risiken des HAT näher betrachtet und mögliche Maßnahmen zur Verringerung der Herausforderungen und Risiken beschrieben. Die Tabelle 5.1 stellt eine Übersicht der verschiedenen Aspekte von Herausforderungen und Risiken innerhalb des HAT dar. In der mittleren Spalte werden die Herausforderungen und Risiken beschrieben. In der dritten Spalte werden mögliche Maßnahmen benannt, welche die Herausforderungen und Risiken vermindern können.

Im Abschnitt 5.3.1 und 5.3.2 wurde das Thema Vertrauen bereits eingeführt. Im Folgenden wird es vor dem Kontext möglicher Herausforderungen und Risiken im Rahmen eines HAT am Beispiel von einem KWS näher betrachtet. Durch den Einsatz von KWS im Krieg ist es schwer zu ermitteln, wer die Verantwortung für die Toten und Schäden tragen soll. Die wichtige Anforderung des Kriegsrechts besagt, dass Länder oder Personen für den Tod von Zivilisten verantwortlich gemacht werden müssen. Für grausame Taten im Krieg muss Verantwortung übernommen werden und Konsequenzen folgen (Sparrow, 2007). Morgan et al. (2020) bezeichnen dieses Phänomen als Verantwortungslücke. Je autonomer die KWS werden, desto weniger können die Designenden und Befehlshabenden zur Verantwortung gezogen werden (Morgan et al., 2020; Sparrow, 2007). Der Grund dafür ist das autonome Handeln des autonomen Agenten, welches nicht vom Menschen kontrolliert wird.

Es gibt drei mögliche Verantwortliche, welche infrage kommen. Morgan et al. (2020) benennen die Programmierenden bzw. Ingenieure/-innen des KWS, der/die Offizier/-in, welcher den Befehl für das KWS gibt und das KWS selbst. Die Programmierenden bzw. Ingenieure/-innen sind die verantwortlichen Personen, da diese das KWS gebaut und die Programmierung vorgenommen haben. Allerdings können diese Personen das Verhalten des KWS nur schwer vorhersagen, da das KWS autonom agiert. Der/die Offizier/-in gibt den Befehl für den Einsatz der KWS und ist somit für die Taten des KWS verantwortlich. Dagegen spricht, dass das KWS mit steigender Autonomie mehr eigene Ziele aussuchen kann und die Wahrscheinlichkeit steigt, ein falsches Ziel zu treffen. Je autonomer, desto weniger Verantwortung müsste die befehlshabende Person übernehmen. Die Maschine ist verantwortlich, da die Aktion eigenständig ausgeführt wird. Wie bereits in Abschnitt 5.3.1 erwähnt wurde, kann der autonome Agent keine moralische Verantwortung tragen.

Maßnahmen zur Schließung der Verantwortungslücke wären der Human-in-the-Loop (HITL) (Morgan et al., 2020) und rechtliche und politische Einigungen. HITL wird der Prozess genannt, in dem der Mensch den autonomen Agenten überwacht und sicherstellt, dass nach entsprechend ethischen und rechtlichen Regeln operiert wird (vgl. Kapitel 3.3.3). Die verantwortliche Person wird für das Fehlverhalten des autonomen Agenten belangt. HITL stellt sicher, dass

ein beteiligtes Teammitglied für die Arbeit oder die Aktionen innerhalb des HAT, die Verantwortung trägt. Rechtliche und politische Einigungen könnten das Risiko präventiv bewältigen. Gesetze müssten entwickelt werden, welche Beteiligte im HAT konkret zur Verantwortung ziehen können. Die Gesetzgeber müssten sich mit den politischen Institutionen einigen, wie die Gesetze kontrolliert und durchgesetzt werden.

Das fehlende Systemverständnis vom menschlichen Teammitglied über den autonomen Agenten könnte zu Risiken führen (Rieth & Hagemann, 2021). Der autonome Agent kommt zu Ergebnissen, die Menschen nicht mehr nachvollziehen können. Das fehlende Verständnis des Systems führt zu keinem Vertrauensaufbau zwischen dem menschlichen Teammitglied und dem autonomen Agenten. Mögliche Vorteile des HAT können nicht wahrgenommen werden oder der autonome Agent wird nicht zweckmäßig eingesetzt (Morgan et al., 2020). Vertrauen und Transparenz sind wichtige Anforderungen für eine erfolgreiche Zusammenarbeit im HAT (vgl. Kapitel 2.1).

Mögliche Maßnahmen zur Bewältigung des Risikos wären situatives Training oder Simulationen. Bei fehlendem Systemverständnis kann das menschliche Teammitglied das Training oder die Erfahrungen aus den Simulationen nutzen, um effektiver mit dem fehlenden Systemverständnis umzugehen. Dies wird vermutlich nicht immer funktionieren. Dennoch wird ein Ansatz geliefert, um Herausforderungen beim Systemverständnis zu bewältigen oder zu verringern.

Ein weiteres Risiko ist, dass sich das menschliche Teammitglied zu sehr auf den autonomen Agenten verlässt. Dem autonomen Agenten wird ohne eine kritische Hinterfragung geglaubt und vertraut, obwohl dieser Fehler macht. Ein Automation Bias entsteht (Morgan et al., 2020). Der Automation Bias ist die Hinnahme eines Ergebnisses, ohne zu wissen, wie es zu diesem Ergebnis gekommen ist (vgl. Kapitel 1.1.1). Im HAT geht es darum, dass Teammitglieder zusammenarbeiten, um ein gemeinsames Ziel zu erreichen (O'Neill et al., 2020). Wenn sich das menschliche Teammitglied vollständig ohne kritische Hinterfragung auf den autonomen Agenten verlässt, lässt sich die Frage stellen, ob in dieser Situation überhaupt ein HAT nach der Definition von O'Neill besteht. Das menschliche Teammitglied bringt keinen eigenen Input ein, um auf das gemeinsame Ziel hinzuarbeiten. Der autonome Agent arbeitet alleine aktiv auf das gemeinsame Ziel hin. Das Arbeitsumfeld wird nicht gebildet, indem das menschliche Teammitglied und der autonome Agent gleichberechtigt zusammenarbeiten und zusammen zu Lösungen gelangen (vgl. Kapitel 5.2.1). Bei der vollständigen Verlässlichkeit auf den autonomen Agenten durch das menschliche Teammitglied kommt die ergänzende Zusammenarbeit zwischen den Teammitgliedern im HAT nicht zur Geltung.

Das Risiko der zu hohen Verlässlichkeit auf den autonomen Agenten kann durch das Training des menschlichen Teammitglieds bewältigt werden. Situationsanalysen oder reflektive Gedankengänge können dem menschlichen Teammitglied ermöglichen, diese Situationen zu bewältigen. Der Mensch wird darauf vorbereitet, sich in bestimmten Situationen der Rolle im HAT bewusst zu sein und somit dementsprechend zu reagieren und zu handeln.

Die Herausforderung der fehlenden Erkennung von Problemen innerhalb des HAT ähnelt dem zuvor genannten Aspekt des fehlenden Systemverständnisses. Durch den hohen Grad an Autonomie des autonomen Agenten wird es schwieriger für Überwachende (menschliches Teammitglied) Probleme im HAT zu erkennen und den autonomen Agenten zu leiten. Der Grund dafür ist das komplexere und schnellere Agieren des autonomen Agenten (Morgan et al., 2020). Das schnellere und komplexere Agieren des autonomen Agenten führt zu fehlendem Systemverständnis seitens des menschlichen Teammitglieds. Die Konsequenzen des fehlenden Systemverständnisses wurden bereits beschrieben (vgl. Kapitel 5.3.3).

Um diese Herausforderung zu bewältigen, müsste ein Training in Kombination mit dem HITL stattfinden. Das Training des menschlichen Teammitglieds sollte in den Bereichen von Problemerkennung und Problemlösung im HAT erfolgen. Simulationen können hilfreich sein, um Szenarien der Problemerkennung und Problemlösung zu schaffen. Durch den HITL wird sichergestellt, dass das menschliche Teammitglied den autonomen Agenten überwacht und die Möglichkeit bekommt, Probleme zu erkennen (vgl. Kapitel 3.3.3).

Beim HAT im Militär werden weniger Menschen für Operationen direkt eingesetzt. Die Menschen werden die überwachende und steuernde Funktion einnehmen und weniger physisch in Kriegsszenarien erscheinen. Der autonome Agent wird in militärischen Operationen die physisch anwesende Einheit sein. Das Risiko besteht, dass die Hemmschwelle für die Initiierung von militärischen Aktionen sinkt. Das Risiko erhöht sich, dass militärische Aktionen häufiger durchgeführt werden und weniger diplomatische Mittel genutzt werden (Morgan et al., 2020). Der Grund dafür ist die fehlende physische Anwesenheit des menschlichen Teammitglieds auf dem militärischen Operationsgebiet. Stattdessen wird der autonome Agent eingesetzt. Dieser hat nicht den gleichen gesellschaftlichen Wert wie der Mensch und ist leichter entbehrlich. Aufgrund des häufigen Einsatzes des autonomen Agenten in militärischen Operationen besteht ein Fehlgebrauch (Misuse) des HATs (vgl. Kapitel 1.4.2).

Maßnahmen für die Verringerung des Risikos sind der defensive Gebrauch von autonomen Waffensystemen und politische und rechtliche Einigungen. Politische und rechtliche Institutionen müssen sich auf spezielle Regelungen und Gesetze einigen, welche einen sinnvollen und zielführenden Einsatz von HAT ermöglichen. Beispielhaft ist die Maßnahme autonome Waffensysteme ausschließlich für defensive Zwecke einzusetzen. Da HAT zukünftig mehr Aspekte der menschlichen Arbeitswelt einnehmen wird, ist die Festlegung von Regularien und Gesetzen von entscheidender Bedeutung (vgl. Kapitel 5.3.2).

Tabelle 5.1: Aspekte der Herausforderungen und Risiken des HATs mit möglichen Maßnahmen zur Verringerung.

Aspekt	Herausforderungen und Risiken	Maßnahmen zur Verringerung von Herausforderungen und Risiken
Verantwortungslücke	<ul style="list-style-type: none"> Unklarheit, wer die Verantwortung für die Taten des autonomen Agenten übernimmt 	<ul style="list-style-type: none"> Human-in-the-Loop (HITL) Politische und rechtliche Einigungen
Fehlendes Systemverständnis	<ul style="list-style-type: none"> Nicht Nachvollziehbarkeit der Aktionen und des Ergebnisses vom autonomen Agenten 	<ul style="list-style-type: none"> Situatives Training der menschlichen Teammitglieder Simulationen der Situation
Verlässlichkeit	<ul style="list-style-type: none"> Vertrauen in den autonomen Agenten ohne Hinterfragung 	<ul style="list-style-type: none"> Reflektives Training der menschlichen Teammitglieder Simulationen der Situation
Erkennung von Problemen	<ul style="list-style-type: none"> Komplexität und Agilität des autonomen Agenten erschweren Problemerkennung und Leitung 	<ul style="list-style-type: none"> Human-in-the-Loop (HITL) Training zur Problemerkennung und -lösung
Fehlgebrauch (Misuse)	<ul style="list-style-type: none"> Geringe Hemmschwelle den autonomen Agenten einzusetzen (im Militärkontext) Nutzung zu falschen Zwecken 	<ul style="list-style-type: none"> Politische und rechtliche Einigungen Defensiver Gebrauch

Das HAT ermöglicht eine effektive und kooperative Zusammenarbeit zwischen dem Menschen und dem autonomen Agenten, da sich die Stärken der Teammitglieder gegenseitig ergänzen (vgl. Kapitel 5.2.1). Dadurch entsteht das Potential, die Arbeitswelt der Zukunft zu erleichtern

und zu verbessern. Herausforderungen und Risiken werden meist nicht in der Visualisierung der Arbeitswelt der Zukunft berücksichtigt. Wie bei jeder neuen technologischen Entwicklung muss eine goldene Mitte gefunden werden zwischen Vorteilen in der Nutzung des HAT und den daraus resultierenden Herausforderungen und Risiken. Je mehr sich in der Wissenschaft und Forschung und in der Öffentlichkeit mit dem Thema HAT beschäftigt wird, desto effizienter und zweckführender können die Menschen das HAT zukünftig nutzen. Wie das HAT die Arbeitswelt der Zukunft gestalten wird, hängt von den Nutzenden, Entwickelnden und Institutionen ab. Diese können die Entwicklung beeinflussen, wie der Mensch und der autonome Agent die Welt der Zukunft gestalten werden.

Literaturverzeichnis:

- Agrawal, A., Gans, J. & Goldfarb, A. (2017). *What to expect from artificial intelligence*. *MIT Sloan Management Review*, 58(3), 23–26.
- Arthur Jr., W., Edwards, B. D., Bell, S. T., Villado, A. J. & Bennett Jr., W. (2005). Team task analysis: Identifying tasks and jobs that are team based. *Human Factors*, 47(3), 654–669.
- Morgan, F. E., Bourdreaux, B., Lohn, A. J., Ashby, M., Curriden, C., Klima, K. & Grossmann, D. (2020). *Military application of Artificial Intelligence. Ethical concerns in an uncertain world*. Santa Monica: RAND Corporation.
- Awad, E., Dsouza, S., Kim, R., Schulz, J., Henrich, J., Shariff, A., Bonnefon, J.-F. & Rahwan, I. (2018). The Moral Machine Experiment: 40 million decisions and the path to universal machine ethics. *Nature*, 563(7729), 59–64.
- Azhar, M. Q. & Sklar E. I. (2017). A study measuring the impact of shared decision making in a human-robot team. *The International Journal of Robotics Research*, 36(5–7), 461–482.
- Barnes, M. J., Chen, J. Y. C. & Hill, S. G. (2017). *Humans and autonomy: implications of shared decision-making for military operations* (No. ARL-TR-7919). US Army Research Laboratory.
- Baxter, G., Rooksby, J., Wang, Y. & Khajeh-Hosseini, A. (2012). The ironies of automation: Still going strong at 30? *Proceedings of the 30th European conference on Cognitive Ergonomics* (pp. 65-71). Edinburgh, United Kingdom.
- Bogg, A., Birrell, S., Bromfield, M. A. & Parkes, A. M. (2020). Can we talk? How a talking agent can improve human autonomy team performance. *Theoretical Issues in Ergonomics Science*. Advance Online Publication.
<https://doi.org/10.1080/1463922X.2020.1827080>
- Bonnefon, J.-F., Shariff, A. & Rahwan, I. (2016). The social dilemma of autonomous vehicles. *Science*, 352(6293), 1573–1576.
- Butkus, M. A. (2020). The human side of artificial intelligence. *Science and Engineering Ethics*, 26, 2427–2437.
- Bünthe C. (2020). *Die chinesische KI-Revolution*. Wiesbaden: Springer Gabler.
- Cohen, J. & Imada, A. (2005). Agent-based training of distributed command and control teams. *Proceedings of the Human Factors and Ergonomics Society Annual Meeting*, 49(25), 2164–2168.
- Cesafsky, L., Stayton, E. & Cefkin, M. (2019). Calibrating agency: Human-Autonomy Teaming and the future of work amid highly automated systems. *Ethnographic Praxis in Industry Conference Proceedings*, 2019(1), 65–82.
- Chen, J. Y. C. & Barnes, M. J. (2014). Human-agent teaming for multirobot control: A review of human factors issues. *IEEE Transactions on Human-Machine Systems*, 44(1), 13–29.

- Cummings, M. L. & Mitchell, P. J. (2008). Predicting controller capacity in supervisory control of multiple UAVs. *Systems and Humans*, 38(2), 451–460.
- De Visser, E. J., Pak, R. & Shaw, T. H. (2018). From ‘automation’ to ‘autonomy’: The importance of trust repair in human–machine interaction. *Ergonomics*, 61(10), 1409–1427.
- Demir, M., McNeese, N. J. & Cooke, N. J. (2016). Team communication behaviors of the human-automation teaming. *2016 IEEE International Multi-Disciplinary Conference on Cognitive Methods in Situation Awareness and Decision Support (CogSIMA)*, 28–34.
- Endsley, M. R. (2017). From here to autonomy: lessons learned from human–automation research, *Human Factors*, 59(1), 5–27.
- Erikson, T. (2019). *Surrounded by idiots: The four types of human behavior and how to effectively communicate with each in business (and in life) (First U.S. Edition)*. New York: St. Martins Publishing Group.
- Fan, X., McNeese, M. & Yen, J. (2010). NDM-Based cognitive agents for supporting decision-making teams. *Human-Computer Interaction*, 25(3), 195–234.
- Farisco, M., Evers, K. & Salles, A. (2020). Towards establishing criteria for the ethical analysis of Artificial Intelligence. *Science and Engineering Ethics*, 26, 2413–2425.
- Haagen, C. & Lohmann, A. (2020). *Künstliche Intelligenz im Strafrecht*. Berlin: Springer Verlag.
- Hanna, N. & Richards, D. (2015). The impact of virtual agent personality on a shared mental model with humans during collaboration. Proceedings of the 14th International Conference on Autonomous Agents and Multiagent Systems (pp. 1777–1778). Istanbul, Turkey.
- Hartmann, M. (2020). *Künstlichen Intelligenz im Zivilrecht*. Berlin: Springer Verlag.
- Himma, K. E. (2009). Artificial agency, consciousness, and the criteria for moral agency: what properties must an artificial agent have to be a moral agent? *Ethics and Information Technology*, 11, 19–29.
- Hosny, A., Parmar, C., Quackenbush, J., Schwartz, L. H. & Aerts, H. J. (2018). Artificial intelligence in radiology. *Nature Reviews Cancer*, 18(8), 500–510.
- Purves, D., Jenkins, R. & Strawser, B. J. (2015). Autonomous machines, moral judgment, and acting for the right reasons. *Ethical Theory and Moral Practice*, 18(4), 851–872.
- Kahnemann, D. (2011). *Thinking, fast and slow*. United Kingdom: Penguin Random House UK.
- Kleesiek, J., Murray, J. M., Strack, C., Kaissis, G. & Braren, R. (2020). Wie funktioniert maschinelles Lernen? *Der Radiologe*, 60(1), 24–31.
- Lee, K.-F. (2018). How AI can save our humanity. *Youtube*. Zugriff am 31.01.2021. Verfügbar unter: <https://www.youtube.com/watch?v=ajGgd9Ld-Wc>
- Lewis, M. (2013). Human interaction with multiple remote robots. *Reviews of Human Factors and Ergonomics*, 9(1), 131–174.

- Mercado, J. E., Rupp M. A., Chen J. Y. C, Barnes M. J., Barber D. & Procci, K. (2016). Intelligent agent transparency in human–agent teaming for Multi-UxV Management. *Human Factors*, 58(3), 401–415.
- O’Neill, T. A., McNeese, N. J., Barron, A. & Schelble, B. (2020). Human–Autonomy Team-ing: A review and analysis of the empirical literature. *Human Factors*. Advance Online Publi-cation. <https://doi.org/10.1177/0018720820960865>
- Precht, R. D. (2018). *Jäger, Hirten, Kritiker: Eine Utopie für die digitale Gesellschaft* (2. Auflage). München: Goldmann Verlag.
- Precht, R. D. (2020). *Künstliche Intelligenz und der Sinn des Lebens* (1. Auflage). München: Goldmann Verlag.
- Rieth, M. & Hagemann, V. (2021). Veränderte Kompetenzanforderungen an Mitarbeitende infolge zunehmender Automatisierung – Eine Arbeitsfeldbetrachtung. *Gruppe. Interaktion. Organisation. Zeitschrift für Angewandte Organisationspsychologie (GIO)*. Advance Online Publication. <https://doi.org/10.1007/s11612-021-00561-1>
- Salas, E., Cooke, N. J. & Rosen, M. A. (2008). On teams, teamwork, and team performance: Discoveries and developments. *Human Factors*, 50(3), 540–547.
- Sparrow, R. (2007). Killer robots. *Journal of Applied Philosophy*, 24(1), 62–77.
- Strybel, T. Z., Keeler, J., Barakezyan, V., Alvarez, A., Mattoon, N., Vu, K.-P. L. & Battiste V. (2018). Effectiveness of Human Autonomy Teaming in cockpit applications. In S. Yamamoto & H. Mori (Eds.) *Human Interface and the Management of Information. Information in Applications and Services. HIMI 2018. Lecture Notes in Computer Science* (Vol. 10905). Cham: Springer International Publishing.
- Walliser, J. C., de Visser, E. J., Wiese, E. & Shaw, T. H. (2019). Team structure and team building improve human–machine teaming with autonomous agents, *Journal of Cognitive Engineering and Decision Making*, 13(4), 258–278.
- Wright, M. C. & Kaber, D. B. (2005). Effects of automation of information – processing functions on teamwork. *Human Factors*, 47(1), 50–66.