

Assistierter Schlaf.

Smarte Kommunikationsassistenten und ihre Nutzer*innen unter dem Blickwinkel von HAL 9000 in 2001: A Space Odyssey, und am Beispiel der Schlafassistenz.

Masterarbeit im Fach:

Medienkultur- und Kunsttheorien (MKKT)

Eingereicht von:

Julian Umhaller, BA

Zur Erlangung des akademischen Grades:

Master of Arts (MA)

Angefertigt am:

Institut für Medien

Kunstuniversität Linz

Linz, 2023

Betreuung:

Prof. Dr.phil. Gloria Meynen

Abstract

Wildgewordene Maschinen, wie HAL 9000 in *2001: A Space Odyssey* zeichnen ein dystopisches Bild der Zukunft. Sowohl durch die Gefahr, die diese fiktiven Maschinen und Algorithmen für ihre Nutzer*innen darstellen, als auch durch ihre vermeintliche Belebtheit. Dem gegenüber stehen gegenwärtige smarte Kommunikationsassistenten, die weder die Belebtheit, noch die Funktionalität, oder die Handlungsfähigkeit ihrer fiktiven Gegenüber besitzen. Diese Disparität zwischen Realität und Science Fiction bildet den konzeptionellen Ausgangspunkt für diese Arbeit und mündet in der Frage, wie sich die Beziehung zwischen smarten Kommunikationsassistenten und ihren Nutzer*innen gestaltet.

Dabei lassen sich, ausgehend von der frühen Computerforschung sowohl Chancen als auch Grenzen von smarten Assistenten ableiten. Diese Bedingungen diktieren wesentlich die Interaktion. Sei es durch die Notwendigkeit Maschinen, und ihren Zertifizierungen zu vertrauen, oder durch den Umstand, dass Funktionalität auch immer an die Gefahr des Scheiterns geknüpft ist.

So werden zuerst smarte Kommunikationsassistenten definiert und die Figur des HAL 9000 als Antithese zu gegenwärtigen smarten Assistenten beleuchtet. Infolgedessen werden die technologischen Eigenschaften und zugrundeliegenden Konzepte von smarten Assistenten analysiert, und die Möglichkeiten der Datenerhebung, Verarbeitung und Automation erfasst. Schließlich werden dann gegenwärtige Möglichkeiten der Schlafassistenz mit den Eigenschaften, und den Grenzen von smarten Assistenten verschränkt. Dabei zeigt sich, dass Science Fiction auch das Potenzial besitzt, technologische Implikationen aufzuzeigen und neue Perspektiven darzustellen. So können fiktive, wildgewordenen Maschinen, trotz ihrer Realitätsferne, einen Blick auf Chancen, Gefahren und neue Möglichkeiten von gegenwärtigen Technologien eröffnen.

Abstract Englisch

Machines gone wild, such as HAL 9000 in *2001: A Space Odyssey*, depict a dystopian view of our future. Both through the danger, that these fictitious machines and algorithms pose to their users, as well as through their supposed aliveness. In comparison, current smart personal assistants are neither alive, nor do they have the functionality (and ability to act), that their fictitious counterparts hold. This disparity between reality and science fiction forms the conceptual basis for the work, and results in the question of how the dynamics between smart personal assistants and their users are constituted.

Starting from early computer science, the opportunities and limits of smart assistants can be derived. These properties dictate the interaction quite substantially. – Both due to the need to trust these machines and their certifications, and because of the fact that functionality always involves the risk of failure.

Initially, smart personal assistants need to be defined, and HAL 9000 is explored as a possible antithesis to current smart personal assistants. Moreover, the technological properties and underlying concepts of smart assistants need to be analyzed. This includes the possibilities of data collection, processing and automation. Finally, within the context of the properties and limits of smart assistants, current sleep assistance tools are discussed. Additionally, it can be concluded, that science-fiction-narratives hold the potential to uncover technological implications and offer alternative perspectives. In this way, fictitious machines that have gone wild can, despite their lack of reality, open up new perspectives about the opportunities and dangers of current technologies.

Danksagung

Allem voraus, möchte ich mich bei meinen Eltern bedanken. Mein Interesse für Kunst und Kultur habt ihr mir bereits in die Wiege gelegt, und eure Begeisterung dafür hat mich schon bald angesteckt. So schätze ich die vielen Abende der bereichernden Diskussionen und Gespräche. Nicht zuletzt danke ich euch natürlich auch für die großartige Unterstützung bei Korrektur und Layout.

Weiters möchte ich mich bei meiner Partnerin bedanken. Du hast nicht nur meine Sprunghaftigkeit ertragen, sondern es auch geschafft, mich immer wieder zu motivieren. Und ohne dein Talent, komplexe Formulierungen zu vereinfachen, wäre die Arbeit vermutlich kein Spaß zu lesen. Tausend Dank dir.

Schließlich möchte ich mich vor allem bei Prof. Dr.phil Gloria Meynen von ganzem Herzen für die persönliche und konstruktive Betreuung, sowie für den theoretischen Input bedanken; und auch für den Ansporn, weiter, über erforschte Sphären hinaus zu denken. So konnte ich eine Masterarbeit schreiben, die meine Erwartungen übertroffen hat.

Inhaltsverzeichnis

Abstract	II
Abstract Englisch	III
Danksagung	IV
Einleitung	6
Was sind smarte Kommunikationsassistenten?	7
Versionen von 2001: A Space Odyssey	11
Science Fiction und Zukunftsvisionen	14
Das Uncanny Valley	18
Die Unheimliche Maschine	20
Intelligent Machinery	24
Die Grenzen von Künstlicher Intelligenz	28
Das Problem mit Emotional Intelligence	30
Ansätze des Affective Computing	34
Sensoren und Daten	39
Nutzer*innenprofile und Künstliche Neuronale Netzwerke	44
Künstliche Neuronale Netzwerke in der Praxis	49
Vertrauen und Überprüfbarkeit	50
Server und Clients	52
Assistierter Schlaf	55
Digitale Dienerschaft	62
Science Fiction und Realität	65
Konklusion	68
Ausblick	71
Literaturverzeichnis	73
Digitalquellen	77
Abkürzungsverzeichnis	79

Einleitung

Diese Arbeit befasst sich mit der Beziehung zwischen smarten Kommunikationsassistenten, und ihren Nutzer*innen, und vergleicht diese, unter anderem am Beispiel der Schlafsteuerung, mit der Dynamik zwischen HAL 9000 und den Astronauten in der Handlung von *2001: A Space Odyssey*. Dabei muss zuerst ein theoretischer Rahmen beschrieben werden, in dem die Science-Fiction-Narration (sowohl des Romans, als auch des Films) mit gegenwärtigen technologischen Entwicklungen verglichen werden kann. Neben möglichen Überschneidungen zwischen Science Fiction und Zukunftsforschung, wird dabei auch die Narration von *A Space Odyssey* thematisiert.

Ausgehend von dem theoretischen Zugang des Uncanny Valley, das die emotionale Abneigung gegenüber anthropomorph anmutenden Maschinen (etwa durch Unbehagen, Unheimlichkeit) beschreibt, gilt es weiters, den Aspekt menschlicher Imitation als wesentliche Devise von maschineller Intelligenz zu fassen. Diese Imitation von Menschen – sowohl strukturell in Form von Künstlichen Neuronalen Netzwerken, wie auch beim Interfacing in Form von Sprachsteuerung – ist dabei zentral für die Dynamik zwischen smarten Kommunikationsassistenten und ihren Nutzer*innen. So erlaubt es diese Imitation zwar, Interfaces für die Nutzer*innen effizienter zu gestalten, sie scheitert jedoch daran, mehr als nur quantitative Daten zu erheben. Zugänge des Affective Computing versuchen zwar, diese Problematik zu umgehen, schaffen das aber nur begrenzt, und nur durch vielschichtige Datenerhebungen.

Darüber hinaus erfüllen smarte Kommunikationsassistenten eine Doppelrolle: Sie sind zum einen Werkzeuge (Tools) um, auf vereinfachte Weise, Abläufe und Prozesse zu automatisieren, und besitzen zum zweiten eine Eigenmächtigkeit, fungieren als Dienstleister, deren Services auch blockieren können. Daher sollen smarte Assistenten unter dem Blickwinkel von Digitaler Dienerschaft untersucht werden. Jene beschreibt das Zusammenspiel aus Funktionalität und Risiko, sowie die Notwendigkeit der Maschine entweder Vertrauen entgegenzubringen, Verantwortung (inform der Haftung des Herstellers) herzustellen, oder mögliche Fehler zumindest zu tolerieren. Außerdem ist eine Erörterung des Zusammenspiels von Servern und Clients notwendig.

Am Beispiel der digitalen Schlafsteuerung und -überwachung zeigt sich dabei vor allem eines: Die Gefahr einer dystopischen Zukunft ist (zumindest vorläufig) gebannt. Gegenwärtige Applikationen zur Schlafvermessung und -steuerung schöpfen nämlich die technologischen Möglichkeiten kaum aus; so müssten etwa weit mehr Daten erhoben werden, um eine effizientere Messung und Interpretation des Schlafs durchführen zu können. Dafür wären jedoch Sensoren notwendig, die über die Möglichkeiten von Smartphone und Fitnesstracker hinausgehen. Schließlich besitzen gegenwärtige Applikationen zur Schlafvermessung und -steuerung auch nur ein geringes Maß an Handlungsfähigkeit. Zwar können die Nutzer*innen nach dem Aufwachen an den gemessenen Daten ablesen, wie erfolgreich ihr Schlaf war, doch weder smarte Assistenten, noch die Applikationen selbst erlauben es, mehr als nur den Aufweckzeitpunkt zu variieren. Alle anderen Änderungen müssen aktiv von den Nutzer*innen selbst getätigt werden, wohl um im Zweifelsfall nicht für Fehler haften zu müssen. Damit sind diese Technologien aber, zumindest vorläufig, reine Lifestyle Produkte.

Ähnliche Tendenzen zeigen sich auch in der Beziehung von smarten Kommunikationsassistenten und ihren Nutzer*innen. Während HAL 9000 in *2001: A Space Odyssey* direkt und aktiv den Alltag und die Schlaf- sowie Lebensqualität der Astronauten mitbestimmt, sind gegenwärtige smarte Assistenten hauptsächlich entbehrliche Interfaces. Dies umschifft damit zwar die Gefahr, dass smarte Assistenten ihren Nutzer*innen bewusst, oder durch fehlerhafte Software/Hardware schaden können. – Doch im Umkehrschluss bleiben smarte Kommunikationsassistenten damit redundante Werkzeuge zur Automation, mit Menschen-imitierenden Interfaces.

Was sind smarte Kommunikationsassistenten?

Um die Beziehung zwischen smarten Kommunikationsassistenten, und ihren Nutzer*innen erörtern zu können – und im Vergleich zu HAL 9000 in *2001: A Space Odyssey*, sowie am Beispiel des Assistierte Schlafs analysieren zu können – muss dabei zuerst der Begriff des smarten Assistenten abgegrenzt werden. Denn es gibt keine formalisierte Definition, wo Assistenten beginnen, und wo sie enden.

Allerdings lassen sich in einem ersten Schritt Eigenschaften beschreiben, die sich in gängigen smarten Kommunikationsassistenten wiederfinden lassen. So beschreiben etwa Knote et. al. (2021) verschiedene Funktionen die smarte Assistenten¹ besitzen können:

»[1] Communication mode refers to the primary way(s) a user communicates with an SPA and vice-versa. Communication is either primarily text-based [...], voice-based [...], visual-sensor-based [...], text-and-vision-based [...], voice-and-vision-based [...], or passively observational, i.e., the SPA assists by gathering context data without being consciously perceived by the user [...]. [2] Directionality comprises unidirectional interaction [...] and bidirectional interaction [...]. Unidirectional interaction means that either the user or the SPA provides information which is intentionally directed towards the other, but thereafter, the recipient does not respond to the sender's request. Bidirectional means that the SPA co-creates value in communicational exchange. [3] Integration refers to an SPA's outreach to other smart things in the network or to the user's digital life through external control, e.g., concerning an ecosystem integration. One can broadly distinguish between SPAs with the ability to, e.g., control smart household objects, post on social media, or shop on behalf of the user [...] and SPAs designed solely for question answering and information recall without external control [...]. It is also possible that an SPA has no external control because it operates in isolation from other systems [...].«²

Damit ergeben sich bereits diverse Anwendungsbereiche und Ausprägungen. smarte Kommunikationsassistenten sollen jedoch zumindest eine Möglichkeit zur Interaktion mit ihren Nutzer*innen bieten, also die Funktion eines Interfaces erfüllen. Außerdem sollen diese Interfaces die Möglichkeit bieten, Feedback zu erzeugen. Schließlich sollen smarte Kommunikationsassistenten als Interfaces auch die Möglichkeit besitzen, diverse Aufgaben und Prozesse zu automatisieren (also fernzusteuern). – Je besser und/oder je mehr dabei ferngesteuert werden kann, um so nützlicher sind smarte Kommunikationsassistenten dabei als Interfaces. Doch smarte Assistenten sind mehr als nur Interfaces:

¹ Smarte Kommunikationsassistenten werden unter anderem als Smart Personal Assistants (SPAs) und als Intelligent Personal Assistants (IPAs) bezeichnet. Siehe dazu etwa: Yorke-Smith, Neil et. al.; The Design of a Proactive Personal Agent for Task Management. In: International Journal on Artificial Intelligence Tools, Volume 21, Issue 1, 1250004; 2012; World Scientific Publishing Company;

² Knote, Robin et. al.; Value Co-Creation in Smart Services: A Functional Affordances Perspective on Smart Personal Assistants. In: Journal of the Association for Information Systems, Volume 22, Issue 2, 2021; S. 437 f.

»Apple's Siri [...], Google's Google Now [...], and Microsoft's Cortana [...] represent a class of emerging Web-service applications known as intelligent personal assistants (IPAs). An IPA is an application that uses inputs such as the user's voice, vision (images), and contextual information to provide assistance by answering questions in natural language, making recommendations, and performing actions.«³

Diese Definition von Hauswald et. al. (2016), beschreibt nicht nur die gängigsten Einsatzbereiche von smarten Kommunikationsassistenten,⁴ sondern eben auch, dass smarte Assistenten nicht nur (etwa sprachliche) Inputs in Computercode transformieren, sondern tatsächlich unterstützen indem sie etwa Empfehlungen abgeben oder an Aufgaben und Verabredungen erinnern. Diese Fähigkeiten der Software lassen sich dabei anhand von fünf Aspekten theoretisch fassen.

Erstens ist es für die Effektivität von smarten Assistenten entscheidend wie und zu welcher Art von Wissen sie Zugang haben. – Dies kann sowohl den Zugang zu öffentlich einsehbaren Web-Informationen,⁵ zu softwarespezifischen Datenbanken,⁶ oder auch zu etwaigen Nutzungsdaten⁷ umfassen. Zweitens ist es für die Funktionalität von smarten Assistenten entscheidend, bis zu welchem Detailgrad die Software genutzt werden kann. Dies betrifft zum einen die Automatisierbarkeit von Aufgaben, zum anderen hat etwa der Detailgrad der Spracherkennung, direkten Einfluss drauf, ob etwa eine Wissensabfrage erfolgreich ist oder nicht.⁸ Drittens ist das Anpassungsverhalten entscheidend. Dies umfasst zum einen etwaiges Lernverhalten – wenn die Maschine etwa von selbst beginnt,

³ Hauswald, Johann et. al.; Designing Future Warehouse-Scale Computers for Sirius, an End-to-End Voice and Vision Personal Assistant. In: ACM Transactions on Computer Systems, Volume 34; Issue 1; 2016; New York: Association for Computing Machinery; S. 2

⁴ Diese sind (Home-)Automation, Datenaufzeichnung (Tracking), Beantwortung von Fragestellungen mittels Google-Suche, und Erinnerungen sowie nutzer*innenspezifische Empfehlungen.

⁵ Etwa mittels einer Google-Suche.

⁶ Beispielsweise der Wortschatz für die Spracherkennung.

⁷ Etwa Datensätze die durch Sensoren, Mikrofone und/oder Kameras erhoben wurden.

⁸ Wenn die Aufgabenstellung für das jeweilige System zu komplex ist, oder etwa die Suchbegriffe nicht erkannt werden, scheitert selbst ein vergleichsweise simple Wissensabfrage.

den Wecker für Montagmorgen zu stellen – zum anderen bezieht sich das Anpassungsverhalten auch auf die Anpassung an variierende Nutzer*innen.⁹

Viertens wird die Effektivität von smarten Assistenten wesentlich durch kollektives Lernverhalten beeinflusst. Dies umfasst etwa die Möglichkeit, nutzer*innenspezifische Sprachwortschätze an zentrale Server zu senden und dort eine umfassende Datenbank aufzubauen, auf die im Umkehrschluss alle smarten Assistenten Zugriff haben. Fünftens werden smarte Kommunikationsassistenten trotz umfassender Serverkommunikation klar als lokal gespeicherte Software gefasst, die ohne notwendige Internetverbindung funktionieren muss.¹⁰ Diese fünf Aspekte kommen in gegenwärtigen smarten Kommunikationsassistenten vor¹¹ und unterscheiden schließlich Assistenten von Interfaces oder Chatrobotern – selbst wenn sie Funktionen von beiden erfüllen.

Darüber hinaus muss ein weiteren Aspekt von smarten Kommunikationsassistenten thematisiert werden. – Nämlich das Framework,¹² das der Datenerfassung, -verarbeitung und Automation zugrunde liegt. Obwohl diese Begleitsoftware zwar nicht direkt an das Interface gekoppelt ist, ist das Framework der Ort an dem lokal erhobene Daten zwischengespeichert werden, und die Schnittstelle, über die etwaige Zusatzgeräte¹³ in die Infrastruktur von smarten Assistenten integriert werden können.

⁹ So wird etwa für alle Nutzer*innen eines spezifischen smarten Assistenten ein eigenes Nutzer*innenprofil angelegt, das auf personenspezifische Eigenarten, wie etwa Aussprache, oder verwendete Begriffe Bezug nimmt.

¹⁰ Außerdem können smarte Kommunikationsassistenten durch Abschaltung der zugehörigen Maschinen im Regelfall deaktiviert werden.

¹¹ Siehe dazu: Knotte, Robin et. al.; Value Co-Creation in Smart Services: A Functional Affordances Perspective on Smart Personal Assistants. In: Journal of the Association for Information Systems, Volume 22, Issue 2, 2021; S. 438 f.

¹² Übersetzt als Grundgerüst beschreibt der Begriff eine Basissoftware die die Kommunikation zwischen verschiedenen Programmen ermöglicht. Teilweise kann auch auf diese Basissoftware direkt zugegriffen werden, wie etwa bei dem Apple Betriebssystem IOS, wo die Gesundheitsdaten mittels der Software Health abgerufen werden können. (<https://developer.apple.com/documentation/healthkit> Abgerufen am 1. Dezember 2022)

¹³ Etwa: Fitnesstracker.

Smarte Kommunikationsassistenten werden folglich im Rahmen dieser Arbeit als Programmierungen definiert, die als Interfaces, zur Datenerfassung (und -verarbeitung) sowie zur Automation verwendet werden können. Darüber hinaus besitzen smarte Assistenten – gegenüber anderen Programmen – erstens autonomen Zugang zu Informationen und Datenbanken, können zweitens mit (mehr oder weniger) komplexen Inputs umgehen, besitzen drittens ein individuelles, und viertens ein kollektives Anpassungs- sowie Lernverhalten. Fünftens handelt es sich bei smarten Assistenten schließlich auch immer um eine lokal installierte Software. Um die Beziehung zwischen diesen Programmen und ihren Nutzer*innen, etwa am Beispiel des assistierten Schlafs erschließen zu können, muss die Dynamik die durch diesen Anspruch an smarte Assistenten entsteht, analysiert werden. Zuerst muss jedoch der Supercomputer HAL 9000 und die Handlung von *2001: A Space Odyssey* theoretisch verortet, sowie auf die unterschiedlichen Versionen der Science-Fiction-Handlung Bezug genommen werden.

Versionen von *2001: A Space Odyssey*

Für die Analyse von HAL 9000 in *2001: A Space Odyssey* im Kontext der Beziehung von smarten Kommunikationsassistenten und ihren Nutzer*innen, muss zuerst die mediale Aufbereitung thematisiert werden. So wird zum einen auf die Handlung des 1968 erschienenen Films von Stanley Kubrick Bezug genommen. Zum zweiten wird der, ebenfalls 1968 erschienene, gleichnamige Roman des Science Fiction Autors Arthur C. Clarke referenziert. Dabei überschneiden sich die Handlung und zugrundeliegende Themen, vor allem wegen des Entstehungskontexts:

»The two started working together on the script, but soon realized that the script format was too limiting; they thus decided to try out their ideas in a novel instead. The script developed from the early drafts of the novel, though — as Clarke points out — ›toward the end, both novel and screenplay were being written simultaneously, with feedback in both directions‹ [...] Eventually, when the film was released in 1968 the credits attributed the script to both Kubrick and Clarke. The novel came out shortly after the film, [...] with Clarke as the only author.«¹⁴

¹⁴ Caracciolo, Marco; Bones in Outer Space: Narrative and the Cosmos in *2001: A Space Odyssey* and Its Remediations. In: *Image [&] Narrative*; Vol. 16, No. 3 (2015); Löwen: Katholieke Universiteit Leuven; S. 74 nach: Clarke, Arthur C.; 1972; S. 31

Trotz dieser verwobenen Produktion gibt es jedoch einige Unterscheidungen in der Handlung, wie etwa eine Verkürzung der Geschichte (durch das Überspringen des Kapitels zur Kulturgeschichte der Menschheit)¹⁵ und die Abänderung des Ziels der Raumfahrt von Saturn zu Jupiter.

»Why the change from Saturn to Jupiter? Well, it made a more straightforward story-line - and, more important, the special effects department couldn't produce a Saturn that Stanley found convincing. If it had done so, the movie would be now have been badly dated, since the Voyager missions showed Saturn's rings to be far more implausible than anyone had ever dreamed.«¹⁶

Die Bedingung der visuellen Darstellbarkeit ist dabei auch der entscheidende Grund für die Unterschiede zwischen Film und Roman. Unterschiede in der Narration ergeben sich dabei entweder aufgrund von Fragen der Ästhetik (etwa bei Saturn und Jupiter) oder wegen der spannenderen Vermittlung von Handlungssträngen – durch meist visuelles Storytelling. Für die Analyse von HAL 9000 und die Dynamik zwischen dem Supercomputer und den Astronauten, sind dabei zwei narrative Änderungen von Bedeutung:

Zum einen beratschlagen sich die Astronauten Frank Poole und Dave Bowman darüber, ob HAL 9000 fehlerhaft geworden ist. Im Roman findet diese Diskussion nur am Rande statt, und die Astronauten sind sich darüber bewusst, dass der Supercomputer mithört.¹⁷ Im Film hingegen, versuchen die Astronauten sich von HAL 9000 abzukapseln, indem sie ein Reperaturshuttle betreten. Da jener nicht mehr zuhören kann, verfolgt HAL schließlich über die Lippenbewegungen der Astronauten das Gespräch, und ihre Äußerungen des Misstrauens.¹⁸

¹⁵ Siehe Dazu: Caracciolo, Marco; Bones in Outer Space: Narrative and the Cosmos in 2001: A Space Odyssey and Its Remediations. In: Image [&] Narrative; Vol. 16, No. 3 (2015); Löwen: Katholieke Universiteit Leuven; S. 76

¹⁶ Clarke, Arthur C.; 2001: A Space Odyssey; 2018; London: Orbit S. xvi

¹⁷ Siehe dazu: Ebd. S. 143 f.

¹⁸ Kubrick, Stanley; 2001: A Space Odyssey; 1970; Los Angeles: Warner Brothers; Pos. 1:23:40-1:27:21

Zum zweiten versucht der Supercomputer im Roman, den Astronauten Dave Bowman zu töten, indem er die Luftschleusen öffnet. Der Mordversuch findet statt, als Bowman, nach dem Tod von Frank Poole, einen der sich im Weltraum schlaf befindenden, Astronauten aufwecken will.¹⁹ Daraufhin flüchtet sich Bowman in ein »emergency shelter,«²⁰ in dem sich auch ein Raumanzug befindet. Geschützt vor den äußeren Einflüssen, und nachdem er erkannt hat, dass die schlafenden Astronauten bereits tot sind, macht sich Bowman auf den Weg um HAL 9000 abzuschalten.²¹

Im Film hingegen, versucht Bowman, den ins All driftenden Frank Poole noch vor seinem Tod zu retten. Dafür springt er zügig in ein Reparaturshuttle, vergisst dabei jedoch dem Helm von seinem Raumanzug mitzunehmen. Bowman erreicht den leblosen Körper von Frank Poole, und will ihn durch die Luftschleuse in das Raumschiff zurückbringen. Doch HAL 9000 verweigert den Befehl zur Öffnung der Schleuse. Der Supercomputer hat während dieser Rettungsmission bereits die lebenserhaltenden Funktionen der schlafenden Astronauten deaktiviert, und will die Kontrolle übernehmen, indem er Bowman (ohne seinen Helm) aus dem Schiff sperrt. Doch Bowman rettet sich, indem er die Luftschleuse zerstört, einen Notfallhelm aufsetzt, und beginnt HAL 9000 zu deaktivieren.²²

Beide Handlungsstränge sind dabei verschiedene Ausprägungen derselben narrativen Funktion: HAL 9000 will die Kontrolle übernehmen, und die Astronauten beseitigen. Und der Supercomputer hört alles mit (selbst als die Astronauten sich abschotten wollen). Die medienspezifische Aufbereitung von *A Space Odyssey* wird davon diktiert, diese Bedrohung, die HAL für die Astronauten darstellt, adäquat vermitteln zu können.

Darüber hinaus stellten Kubrick und Clarke den Anspruch, dass die Handlung und Themen der Geschichte nicht durch technologische Entwicklungen zunichte gemacht

¹⁹ Siehe dazu: Clarke, Arthur C.; 2001: *A Space Odyssey*; 2018; London: Orbit S. 160

²⁰ Ebd. S. 165

²¹ Siehe dazu: Ebd. S. 168 f.

²² Kubrick, Stanley; 2001: *A Space Odyssey*; 1970; Los Angeles: Warner Brothers; Pos. 1:32:55-1:49:55

werden. Und es galt eine Handlung zu erzeugen, die ohne spezifisches Fachwissen verstanden werden konnte, und nicht zu abstrakt technologisch wurde.

»Our main problem, therefore, was creating a story which would not be made obsolete - or even worse, ridiculous - by the events of the next few years. We had to outguess the future; one way of doing that was to be so far ahead of the present that there was no danger of the facts overtaking us. On the other hand, if we got too far ahead, there was grave risk of losing contact with our audience... '«²³

Dieses Bedürfnis, dass die in der Geschichte erschaffene, fiktive Zukunftsvision möglichst lange glaubwürdig bleibt, betrifft dabei nicht nur die Handlung von 2001: A Space Odyssey, sondern kann als eine wesentliche Prämisse von Science Fiction verstanden werden.

Science Fiction und Zukunftsvisionen

Für das Verhältnis von Science Fiction zu Zukunftsforschung und -vorhersage²⁴ ist dabei die Art und Weise wie auf die Realität Bezug genommen wird, entscheidend. Obwohl Science-Fiction-Handlungen keineswegs realitätsnah, und wissenschaftlich (oder technologisch) nachvollziehbar sein müssen, gilt es doch eine logische und konsistente Welt aufzubauen. Wenzel Mehnert (2019) beschreibt die Eigenarten von Science Fiction dabei wie folgt:

»Das Genre unterteilt sich in eine Vielzahl von Subgenres, die sich vor allem durch die Ausprägung zweier Merkmale unterscheiden: das Novum und die Storyworld. Das Novum bildet den Kern jeder Science-Fiction-Geschichte und bezeichnet die fiktive(n) Neuheit(en) des Zukunftspostulats. [...] Quantitativ kann das Novum unterschiedliche Größenordnungen haben und von einzelnen (oder mehreren) Erfindungen bis hin zu ganzen Milieus und veränderten Umwelten reichen. [...] Ausgehend von dem Novum spannt sich spekulativ die Storyworld [...] auf. Sie beschreibt die räumliche und zeitliche Umwelt der Narration, den fiktiven Erfahrungsraum der Protagonisten und damit das Umfeld ihrer Handlungen. Gleichzeitig erweitert die Storyworld die Frage »Was wäre,

²³ Clarke, Arthur C.; 2001: A Space Odyssey; 2018; London: Orbit S. x

²⁴ Der Versuch wissenschaftlicher Zukunftsvorhersage lässt sich als »Foresight« bezeichnen. Siehe dazu: Steinmüller, Karlheinz; Science Fiction: eine Quelle von Leitbildern für Innovationsprozesse und ein Impulsgeber für Foresight. 2010; Bonn: iFQ S. 27

wenn...?« um kausale Implikationen: Wenn X passiert, passiert als Folge davon Y, das wiederum führt zu Z [...].«²⁵

In der Anwendung dieser beiden Begriffe, ist das Novum in der Handlung von *2001: A Space Odyssey*, die Existenz von Monolithen, die die Entwicklung menschlicher Existenz bedingen (oder zumindest begünstigen). Und weil dieses Novum bereits in der fernen Vergangenheit, nämlich in der Evolution der Menschenaffen ansetzt, erlaubt jenes die Handlung, als eine (in sich geschlossene) Parallelrealität zu begreifen. Explizit werden die Abweichungen und/oder technologische Sprünge in der Storyworld²⁶ jedoch erst mit der Entdeckung des Monolithen auf der Mondoberfläche beginnen. Diese Veränderungen der Storyworld, ergeben sich natürlich daraus, dass während der Konzeption von *A Space Odyssey* gesellschaftspolitische und technologische Entwicklungen der Zukunft, wie die Erfindung und Verbreitung von selbstdenkenden Maschinen, nur vermutet werden konnten. Trotzdem wirkt die Handlung von *A Space Odyssey* nicht altmodisch oder überholt. Die Aktualität erhält sich dabei aufgrund von zwei Aspekten: Einerseits werden die Überlegungen von Clarke und Kubrick nur selten durch nicht-fiktive technologische Entwicklungen zunichte gemacht, der Versuch des »outguessing the future« hat damit funktioniert.²⁷ – Wesentliche Abweichungen von Clarke's und Kubrick's Konzeption finden sich vor allem in der Handlungsfähigkeit,²⁸ und der (narrativ ungemein spannenden) Menschenähnlichkeit, die HAL 9000 besitzt.

Andererseits ist die logische Konsistenz der Handlung entscheidend, da etwa die Gefährlichkeit des HAL 9000 bereits in der narrativen Einführung des Supercomputers erörtert wird. So stellt sich bereits während der Beschreibung der Position und Funktionen die HAL 9000 einnimmt, eine unheimliche Vorahnung ein, dass jener die ihm übertragenen Funktionen nutzen wird, um den Astronauten zu schaden.

²⁵ Mehnert, Wenzel; *The Future is Near: Schnittstellen einer negativen Zukunft*. 2019; Hamburg: Textem S. 55 f.

²⁶ Etwa hin zur Existenz von selbstdenkenden Maschinen im Jahre 2001, oder dem noch immer existierenden Kalten Krieg zwischen UdSSR und USA. Siehe dazu: Clarke, Arthur C.; *2001: A Space Odyssey*; 2018; London: Orbit S. 186

²⁷ Clarke, Arthur C.; *2001: A Space Odyssey*; 2018; London: Orbit S. x

²⁸ Dem Umstand wie einflussreich (und damit auch fatal) die Handlungen des Supercomputers für die Astronauten sein können.

»His prime task was to monitor the life-support systems, continually checking oxygen pressure, temperature, hull leakage, radiation and all the other interlocking factors upon which the lives of the fragile human cargo depended. [...] And he could watch over the hibernators, making any necessary adjustments to their environment, and doling out the minute quantities of intravenous fluids that kept them alive. [...] The time might even come when Hal would take command of the ship. In an emergency, if no one answered his signals, [...] he would take what measures he deemed necessary to safeguard the ship and to continue the mission.«²⁹

Diese Unheimlichkeit erhält HAL 9000 dabei nur, indem die Vorahnung, der Supercomputer würde beginnen seine Fähigkeiten nicht für die Lebenserhaltung der Astronauten sondern für ihren Tod einzusetzen, sich auch bewahrheitet. Neben dieser Dystopie – fehlerhafte Technologien sind ein wesentliches Stilmittel der Science Fiction – ist es entscheidend, dass der potenzielle Schaden, den HAL 9000 anrichten kann, auch eintritt, und sich die durch Novum und Storyworld etablierten Regeln und Möglichkeiten nicht nachträglich ändern. Dies verleiht der Handlung eine nachvollziehbare Logik und dem Supercomputer eine greifbare, realistische Handlungsfähigkeit.

Schließlich kann diese, logisch konsistente und fiktive Zukunft nicht nur als eine dystopische Fortführung der Computerforschung des 20. Jahrhunderts, und vom Wettlauf ins All gesehen werden, sondern besitzt darüber hinaus auch einen prototypischen Blick auf die technologische Entwicklung hin zu gegenwärtigen smarten Kommunikationsassistenten und deren Einsatzgebieten. So werden in den folgenden Kapiteln sowohl auf konzeptioneller, als auch auf technologischer Ebene, Ähnlichkeiten wie auch Unterschiede zwischen HAL 9000 und gegenwärtigen smarten Assistenten beschrieben. Diese lassen sich zwar zum Teil durch den Rückgriff auf dieselben theoretischen Zugänge der frühen Computer- und Künstlichen Intelligenzforschung erklären, zum Teil aber auch nicht. Darüber hinaus ist unklar ob, (und falls, in welchem Ausmaß) sich Entwickler*innen Inspiration von Science-Fiction-Narrationen, und damit Figuren wie HAL 9000 geholt haben (und immer noch holen). So können manche Überschneidungen zwischen HAL 9000 und smarten Kommunikationsassistenten reiner Zufall sein, doch es liegt nahe, dass die Narration des Supercomputers, etwa als popkulturelles Negativbeispiel, die Ausformung der smarten Assistenten wesentlich beeinflusst haben kann.

²⁹ Clarke, Arthur C.; 2001: A Space Odyssey; 2018; London: Orbit S. 98 f.

Karlheinz Steinmüller beschreibt diese Vorbildwirkung, die Science-Fiction-Narrationen dabei einnehmen können, wie folgt:

»[Science-Fiction] trägt in diesem Falle aber zur Suche nach und zur Diskussion über neue Gesellschaftsmodelle bei [...] ist Einübung von Zukunftsdenken. Sie popularisiert neue Ideen wie den Raumflug und motiviert für (oder im Einzelfall auch gegen!) Veränderungen. Sie kann [...] in der Funktion eines ›Mind Openers‹ zur Überwindung von Denkbarrieren beitragen. SF kann als Indikator für kulturelle und soziale Trends, für virulente technische und nichttechnische Ideen, für Wertungen und Bewertungen und allgemeiner, als ein Zugang zu populären Zukunftserwartungen gesehen und genutzt werden. Sie dient dann als ein Frühwarnsystem, das Schwache Signale aufnimmt und transportiert, die auf Künftiges, auf Innovationen, einen Einstellungswandel in der Gesellschaft etc. hinweisen. Nicht zuletzt kann SF heuristisch als Ideen- und Inspirationsquelle eingesetzt werden, beispielsweise wenn es darum geht, mögliche Trendbrüche und Folgen technologischer Durchbrüche zu identifizieren oder Innovationsprozesse durch Anregungen von außen zu unterstützen. Insbesondere die heuristische Funktion der SF wird bisweilen im Rahmen von Innovationsprozessen genutzt.«³⁰

In dieser Weise gilt es schließlich auch das Verhältnis von HAL 9000 in *2001: A Space Odyssey* und smarten Kommunikationsassistenten zu betrachten. HAL wäre demnach nicht zwingend Prototyp für smarte Assistenten, sondern eine Antithese. Im Sinne einer in die Zukunft gedachten Parallelrealität, ist der Supercomputer ein Mahnmal, eine Gegenüberstellung. Dieser dystopische Extremfall konfrontiert reale technologische Chancen und Entwicklungen, mit ihren möglichen Konsequenzen – und trägt damit auch aktiv dazu bei, diese Zukunft zu vermeiden. Die Utopie hingegen imaginiert die positiven Effekte eines technologischen Novums.³¹ In beiden Fällen kann Science Fiction jedoch Empfehlungen abgeben, wohin technologische Entwicklungen führen können, und sollen.

»Wenn wir als Basishypothese voraussetzen, dass SF-Autoren jeweils derzeitig entstehende wissenschaftliche und technische Impulse frühzeitig aufgreifen und in ihren Gedankenexperimenten auf mögliche soziale, kulturelle, wirtschaftliche, ökologische, philosophische etc. Implikationen prüfen, dann liegt es nahe, SF auch für Foresight zu nutzen. Denn stets ist Science

³⁰ Steinmüller, Karlheinz; Science Fiction: eine Quelle von Leitbildern für Innovationsprozesse und ein Impulsgeber für Foresight. 2010; Bonn: iFQ S. 27

³¹ Ob dieses Novum dabei in der fernen oder nahen Zukunft tatsächlich erreicht werden könnte, sei dahingestellt.

Fiction zugleich ein Spiegel aktueller Trends und Entwicklungen und damit ein Medium, das insbesondere Technikvisionen aufgreift, umformuliert und in Diskurse neu einspeist.«³²

Im Umkehrschluss liegt es daher nahe, besonders Science Fiction (und damit HAL 9000 in *2001: A Space Odyssey*) zu einem Vergleich heranzuziehen. So ist es nicht nur das Ziel, dieser Gegenüberstellung zu analysieren, ob (und wenn, in welcher Weise) die Dystopie des allmächtigen Supercomputers eintreten kann (oder gar bereits eingetreten ist), und Überschneidungen zu dokumentieren; so müssen ebenfalls die (potenziellen) Handlungsfähigkeiten, von Nutzer*innen wie auch Assistenten analysiert, und die Interaktion und Dynamik zwischen Menschen und Maschinen beschrieben werden. Ein wesentlicher Grundtenor in der Konzeption, Entwicklung und auch Interaktion mit Maschinen (das heißt: Computern, Kommunikationsassistenten, Robotern) ist dabei der Aspekt der Imitation menschlichen Verhaltens. Dieser zeigt sich etwa am Auftreten einer bestimmten Form von Unheimlichkeit, im Zuge derer die Maschine weder als lebendig, noch als tot, weder als menschlich, noch als unmenschlich wahrgenommen wird.

Das Uncanny Valley

Ein theoretischer Zugang, der sich in der Handlung von *2001: A Space Odyssey* erkennen lässt, und in der Auseinandersetzung mit smarten Kommunikationsassistenten notwendig ist, ist das Phänomen des Uncanny Valley. Dieser Begriff des Unheimlichen Tals³³ wurde erstmals 1970 von Masahiro Mori im Zuge der Akzeptanz gegenüber menschlich wirkenden, also anthropomorphen Robotern benutzt. Das Tal bezeichnet dabei das Phänomen, dass Maschinen, ab einem gewissen Punkt der Ähnlichkeit beginnen, unheimlich zu werden. Und diese Unheimlichkeit äußert sich in einer Abnahme der Affinität (der Akzeptanz und Interaktion) mit menschenähnlichen Robotern.

»Mir ist aufgefallen, dass es in diesem Sinne eine [...] Beziehung zur Affinität gegenüber Robotern gibt, nämlich wenn Roboter äußerlich dem Menschen nähergebracht werden: Unsere Affinität steigt so lange an, bis wir in ein Tal kommen [...]. Ich nenne es das unheimliche Tal. [...] Es ist

³² Steinmüller, Karlheinz; Science Fiction: eine Quelle von Leitbildern für Innovationsprozesse und ein Impulsgeber für Foresight. 2010; Bonn: iFQ S. 23

³³ Die Begriffe Uncanny Valley und Unheimliches Tal bezeichnen dabei dasselbe Phänomen.

unnötig zu erwähnen, dass es das Ziel der Robotik ist, einen künstlichen Menschen zu erschaffen und verschiedene Einrichtungen arbeiten daran, das Erscheinungsbild von Robotern immer menschlicher zu gestalten.[...] Sie scheinen aber nur auf den ersten Blick echt zu sein. Sobald klar wird, dass sie künstlich sind, wecken sie auf einmal ein unheimliches Gefühl in uns. Ein Händedruck mit einer schlaffen, knochenlosen Hand jagt uns, huch, einen kalten Schauer über den Rücken! Durch solch ein Erlebnis verlieren wir unsere Affinität – die Hand wird unheimlich.«³⁴

Mori beschreibt dieses Phänomen dabei primär am visuellen Erscheinungsbild von Maschinen. Neben der Künstlichen Hand – egal ob einem Roboter oder einem Menschen (als Prothese) zugehörig – kann, so Mori, ein anthropomorphes Erscheinungsbild, ab einem gewissen Imitationsgrad, Unheimlichkeit hervorrufen. Zusätzlich zu dieser Unheimlichkeit durch visuelle Ähnlichkeit (die etwa auch eine Wachsfigur oder Schaufensterpuppe aufweisen kann) sind dabei aber auch die Bewegungsabläufe (und damit das Verhalten) von Robotern entscheidend.

»Beispielsweise ist für uns der ausgeschaltete Industrieroboter nur eine ölige Maschine. Doch wenn er so programmiert wird, dass er seine Greifarme wie menschliche Hände bewegt (natürlich muss dafür die Geschwindigkeit und Beschleunigung stimmen), weckt er in uns ein gewisses Maß an Affinität. Wenn jedoch eine Handprothese aus dem unheimlichen Tal anfangen würde, sich zu bewegen, wäre sie noch viel unheimlicher. [...] Ein Roboter hatte beispielsweise 29 Paare künstlicher Muskeln im Gesicht [...], damit er wie ein Mensch lächeln konnte. Laut dem Entwickler war sein Lächeln eine dynamische Sequenz von Gesichtskrümmungen, bei denen besonders Geschwindigkeit eine wichtige Rolle spielt. Falls der Roboter sich auch nur halb so schnell bewegen würde, um ihn langsamer lächeln zu lassen, könnte sein Grinsen schnell unheimlich wirken. Das Beispiel zeigt, dass etwas [...] durch eine abweichende Bewegung schnell in das unheimliche Tal hinabstürzen kann.«³⁵

Es geht also um den Aspekt der Imitation des Menschen – etwa durch ein anthropomorphes Erscheinungsbild und Verhaltensweisen. Die Nuancen dieser Imitation sind dabei für die Frage nach dem Uncanny Valley entscheidend. Denn die Unheimlichkeit des HAL 9000 entsteht nicht dadurch, dass der Supercomputer Kameras für die visuelle Datenerfassung, oder eine Sprachsteuerung für die effizienter Vermittlung von Information

³⁴ Masahiro, Mori; Das Unheimliche Tal. Übersetzung: MacDorman, Karl et. al.; 2019; Hamburg: Textem; S. 212-214.

³⁵ Masahiro, Mori; Das Unheimliche Tal. Übersetzung: MacDorman, Karl et. al.; 2019; Hamburg: Textem; S. 215 f.

nützt, sondern durch das Erleben eines sprechenden und sehenden Roboters – der Menschen zwar soweit imitiert, um anthropomorph aufzutreten – dessen Imitation eines Menschen jedoch in den Feinheiten scheitert. Dies ist, zusätzlich zur Kontrolle des Supercomputers über das Raumschiff,³⁶ wesentlich für die wahrgenommene Bedrohlichkeit des HAL 9000, der durch das irrationale Handeln schließlich zur realen Bedrohung für das Überleben der Astronauten wird. Die Handlung von *A Space Odyssey* treibt diese Imitation von menschlichem Verhalten (und menschlichen Emotionen) sogar so weit, dass der einzige überlebende Astronaut, Dave Bowman das Versagen des Supercomputers als vermeintlich menschlich interpretiert, und ihm deswegen verzeiht:

»If it could happen to a man, then it could happen to Hal; and with that knowledge the bitterness and the sense of betrayal he felt towards the computer began to fade.«³⁷

Narrativ erreicht HAL 9000 an diesem Punkt die intellektuelle Ebene der Astronauten, und hat, zumindest für Bowman, die menschliche Imitation vollendet. Doch die Handlung lässt (zumindest bis zu diesem Urteil) die Frage offen, ob HAL 9000 entweder nur ein nützliches Werkzeug mit viel Handlungsfähigkeit ist,³⁸ oder ob der Supercomputer menschliches Verhalten imitiert – sei es zur einfachen und direkten Kommunikation mit den Astronauten, oder um selbst zum Menschen zu werden. Diese Ambiguität zwischen Leblosigkeit und einer vermeintlichen Lebendigkeit mündet schließlich in einer unheimlichen Maschine.

Die Unheimliche Maschine

Der Uncanny-Valley-Effekt, den HAL 9000 ausübt, lässt sich dabei theoretisch durch zwei Einordnungsprobleme fassen. Erstens durch eine Uneindeutigkeit in der Kategorisierung (»category uncertainty«³⁹), und zweitens durch Unterschiedlichkeiten im Detailreichtum

³⁶ Etwa: Die Möglichkeit des HAL 9000 die lebenserhaltenden Systeme zu beeinflussen.

³⁷ Clarke, Arthur C.; 2001: *A Space Odyssey*; 2018; London: Orbit S. 187

³⁸ Und bei dem lediglich die Programmierung fehlerhaft ist.

³⁹ Chattopadhyay, D.; MacDorman, K. F.; Familiar faces rendered strange: Why inconsistent realism drives characters into the uncanny valley. 2019; *Journal of Vision*, 16(11):7; S. 2

und damit der Menschenähnlichkeit (»perceptual mismatch«⁴⁰), die sich auf die Unterscheidbarkeit zwischen real und fiktiv auswirken können (»realism inconsistency«⁴¹). Demnach entsteht, so MacDorman und Chattopadhyay (2016, 2019), die Unheimlichkeit vor allem aufgrund dieser drei Uneindeutigkeiten, indem menschen-imitierende Maschinen die Grenzen zwischen Werkzeug und Lebewesen, sowie zwischen realer Welt und Fiktion verwischen. Dafür sind vor allem Unterschiedlichkeiten im Detailreichtum entscheidend.

Beispielsweise erscheinen verschiedene Aspekte von HAL 9000 verschieden anthropomorph. Wo die Fähigkeit des Roboters zu sprechen (und Sprache zu verstehen), sich sehr an eine Kommunikation mit Menschen annähert,⁴² sind andere Aspekte der Imitation weniger ausgeprägt. So besitzt der Supercomputer neben Kameras (die als die *Augen* HALs fungieren) und einem Speichermedium, das das menschliche Gehirn (zumindest in der Funktion) imitiert, kaum eine Ausstattung, die als dezidierter *Körper* HALs fungiert. Zwar bezeichnet ihn der Astronaut Dave Bowman als Nervensystem des Raumschiffs,⁴³ doch hier lässt sich argumentieren, dass der Computer lediglich als zentrale Steuereinheit, also als Interface, fungiert. – Denn HAL 9000, bezeichnet sich schließlich auch selbst⁴⁴ als eine abgeschlossene (Berechnungs-)Einheit, die zwar auch als Bordcomputer fungiert, aber nicht das Raumschiff selbst, sondern lediglich die Steuerung der Schiffsfunktionen übernimmt. An diesen drei Aspekten des HAL 9000, lassen sich schließlich bereits Unterschiede im Detailreichtum erkennen.

Der Detailgrad der menschlichen Imitation ist, etwa bei den Kameralinsen viel weniger hoch, als beim Verhalten des Computers, oder bei der Fähigkeit mit HAL Dialoge zu

⁴⁰ Ebds. S. 1

⁴¹ Chattopadhyay, D.; MacDorman, K. F.; Reducing consistency in human realism increases the uncanny valley effect; increasing category uncertainty does not. 2016; Cognition, Volume 146, S. 191

⁴² Eine wesentliche Unterscheidung liegt etwa (noch) in der monotonen Betonung des Supercomputers.

⁴³ Siehe dazu: Clarke, Arthur C.; 2001: A Space Odyssey; 2018; London: Orbit S. 169

⁴⁴ Die Frage ob HAL 9000 sich, selbst erkennen kann und als Lebewesen erlebt, ist nicht klar zu beantworten. Allerdings führt der Supercomputer in Clarke's Roman mehrmals innere Monologe, weswegen von einem Selbstverständnis HALs als zumindest abgegrenzte Entität ausgegangen werden kann. Siehe Dazu: Ebd. S. 161 f.

führen. Darüber hinaus imitiert der zentrale Speicher HALs, auf funktionaler Ebene, so exakt das menschliche Gehirn, dass es für den Astronauten Dave Bowman möglich wird, den Supercomputer zu lobotomisieren.⁴⁵ Und da HAL 9000 auch Angst vor einem vermeintlichen Tod, oder einem Verlust der kognitiven Fähigkeiten verspürt, wird die Unheimlichkeit noch verstärkt. Diese verschiedenen Detailgrade – eines funktional menschenähnlichen Gehirns, einer weitgehend authentischen Sprachsteuerung (samt computergeneriertem Räuspern und monotoner Stimme), und von Kameralinsen als kruden Augen – sorgen dafür, dass HAL 9000 an der Grenze zwischen Maschine und Lebewesen erlebt wird.

Zum einen wird diese Grenze auch in der Handlung thematisiert, die letztendlich eine Entwicklung vom neutralen, Menschen dienenden Werkzeug, hin zu einem wildgewordenen, irrational handelnden Lebewesen, mit der Angst vor der eigenen Abschaltung beschreibt. Zum zweiten bedeutet diese Schwelle an der HAL 9000 sich befindet, dass der Computer zwischen Realitätsnähe und unmissverständlicher Fiktion oszilliert. Dadurch wird es unklar, ob wildgewordene Maschinen nur in der Fiktion stattfinden können, oder ob die Bedrohung des mordenden und täuschenden Supercomputers tatsächlich real sein kann. – Diese Unsicherheit zwischen Realität, und Fiktion ist, so MacDorman und Chattopadhyay (2016, 2019), das Fundament hinter der Unheimlichkeit in Mori's Uncanny Valley. Jene lässt sich dabei noch weiter denken: Denn das unheimliche Tal wird ebenso von der Bedingung diktiert, nicht (mehr) hinter die menschlich auftretende Maske blicken zu können.⁴⁶ Damit der Spuk der uneindeutigen, und damit die eigene Realität und Zurechnungsfähigkeit infrage stellenden Maschine gelingen kann, darf keine Möglichkeit bestehen sie auf technologischer Ebene nachzuvollziehen.

Ein Einblick in die Funktionsweise und das Verständnis der Berechnungen kann die Maschine nämlich als reine Imitation enttarnen, und ihr jedwede Unheimlichkeit entziehen. Das Phänomen des Uncanny Valley funktioniert somit ähnlich wie eine Geisterbahn: Um den Trick (der menschlich anmutenden und unheimlichen Maschine) funktionieren zu lassen, muss zum einen der Glaube an, oder zumindest die Unsicherheit über die

⁴⁵ Siehe Dazu: Ebd. S. 170

⁴⁶ Gepaart mit der Unsicherheit ob die Bedrohung nun real oder nur eingebildet ist.

Möglichkeit bestehen, dass Technologie vermeintlich lebendig werden könne. Zum zweiten, dürfen dann diese (vermeintlich lebendigen) Geister nicht als technologische Tricks enttarnt werden. Denn sobald die Mechaniken (oder verkleideten Menschen) hinter dem Spuk aufgedeckt werden, ist die Illusion verschwunden. Daher ist das Uncanny Valley wesentlich an die Nicht-Einsehbarkeit in die Maschine, die als unheimlich erlebt wird, beziehungsweise an das Unwissen über die Strukturalität von etwa HAL 9000 (oder Computersystemen im Allgemeinen) gebunden.

Für die Beziehung zwischen smarten Kommunikationsassistenten und ihren Nutzer*innen bedeutet dies, dass Konsistenz im Interface-, und Hardwaredesign entscheidend für die Akzeptanz von smarten Assistenten ist. Dies betrifft nicht nur eine etwaige intuitive Nutzung, sondern vor allem den Grad menschlicher Imitation. Zwar können Maschinen Abschreckung oder Abneigung bei den Nutzer*innen auslösen, doch es stellt sich die Frage ob (und falls unter welchen Umständen) andere emotionale Reaktionen möglich sind. Das Forschungsgebiet des Affective Computing versucht schließlich Maschinen mit dieser Fähigkeit auszustatten. So sollen Maschinen, wie etwa smarte Kommunikationsassistenten, mittels kluger und umfassender Datenerfassung und -interpretation die Fähigkeit erhalten, sowohl Emotionen wahrzunehmen und zu interpretieren, als auch, mittels audiovisuellen Symbolen und Signalen, emotionale Reaktionen bei Nutzer*innen auszulösen.

Um dabei die Grenzen von Affective Computing aufzeigen zu können, muss die Entwicklungsgeschichte, bis hin zum Versuch Maschinen mit Emotionen zu versehen, nachgezeichnet werden. Der entscheidende Aspekt, den dabei Theorien (und Praxen) des Affective Computing zu imitieren versuchen, ist die Emotional Intelligence⁴⁷ von Menschen, die es etwa erlaubt, Vermittlungs- und Informationskontexte von Sprache, oder generelle Stimmungen zu lesen, und etwa auf ein wütendes Gegenüber anders zu reagieren, als auf ein freundliches. Dort finden Theorien des Affective Computing auch bereits Anwendung, da gegenwärtige smarte Assistenten durch Affective Computing vermeintlich besseres (und feinfühligeres) Feedback geben können, und dies die Software dazu befähigt, detaillierte Computermodelle anzulegen, an denen dann etwaige

⁴⁷ Der Begriff wurde dabei 1990 von Peter Salovey und John D. Mayer geprägt. Siehe dazu: Salovey, Peter; Mayer, John D.; Emotional Intelligence. In: Imagination, Cognition and Personality, Volume 9, Issue 3; 1990; Sage; S. 185-211

Bedürfnisse von Nutzer*innen genauer vorhergesagt werden können. Um jedoch erörtern zu können, wie die Imitation von Emotional Intelligence in Theorie und Praxis funktioniert, und was Modelle des Affective Computing implizieren,⁴⁸ gilt es zuerst grundlegende Konzepte zur Intelligenz von Maschinen zu thematisieren.

Intelligent Machinery

Ob und wie eine Maschine intelligent sein oder werden kann (und soll), lässt sich mittels Alan Turings Konzept der Imitation fassen. Der in *Intelligent Machinery* (1948) erörterte Imitationsbegriff bedeutet dabei nicht zwingend, dass Maschinen menschliches Verhalten imitieren oder kopieren, sondern primär die Möglichkeit einer Maschine, die Aufgabe (und damit die Programmierung) einer anderen Maschine übernehmen zu können. Diese Eigenschaft, Arbeitsprozesse flexibel kopieren zu können, ist für Turing das Merkmal einer »Universal Machine«⁴⁹, im Gegensatz zur diskreten »Logical Computing Machine«⁵⁰ (L.C.M.):

»[...] it is possible to design a ›universal machine‹ which is an L.C.M. such that if the standard description of some other L.C.M. is imposed on [it] [...] it will carry out the operations of the particular machine whose description it was given. [...] We do not need to have an infinity of different machines doing different jobs. A single one will suffice.«⁵¹

Turings Universal Machine ist damit nicht nur das Fundament für die konzeptionelle Trennung zwischen Computer (der als Hardware fungiert) und Programmierung (die als Software den Computer auf eine Aufgabe hin ausrichtet). – Die Konzeption von Universal Machines umfasst zusätzlich das Statement, dass schlußendlich nur eine einzige Maschine benötigt werden soll, um alle (maschinellen) Aufgaben zu lösen.

⁴⁸ Und woran jene scheitern.

⁴⁹ Turing, Alan M.; *Intelligent Machinery*; 1948; In: Meltzer, Bernard Et. al. (Hsg.); *Machine Intelligence* 5; 1969; Edinburgh: Edinburgh University Press; S. 7

⁵⁰ Ebd. S. 6

⁵¹ Ebd. S. 7

Dieser Trend hin zur Singularität, ist der zentrale schöpferische Anspruch. Und dieser Anspruch lässt sich in beinahe allen Bereichen der Computerforschung wiederfinden.⁵²

Um, so Turing, intelligente Maschinen herstellen zu können, gilt es jedoch nicht nur eine Maschine so flexibel wie möglich zu gestalten, sondern ebenso, dass die Maschine sich bei Änderungen im Datensatz selbst anpassen kann: Zum einen durch die Möglichkeit zwischen verschiedenen Programmierungen (also Imitationen von Maschinen) hin und her zu springen, zum zweiten beschreibt Turing aber auch konzeptionelle Ansätze, wie Maschinen ihre eigene Programmierung selbst verändern (und damit verbessern) können. Hier beschreibt Turing, wie maschinelles Lernverhalten, den menschlichen Verstand imitiert.

»If we are [...] following the human model as closely as we can, we should begin with a machine with very little capacity to carry out elaborate operations, or to react in a disciplined manner to orders (taking the form of interference). Then by applying appropriate interference, mimicking education, we should hope to modify the machine until it could be relied on to produce definite reactions to certain commands. This would be the beginning of the process.«⁵³

Dieser Prozess, kann dabei, so Turing, intelligente Maschinen erschaffen. Schließlich hält Turing aber ebenso fest, dass Flexibilität, und Lernverhalten bei Maschinen, nicht dasselbe sind wie Intelligenz.

»The extent to which we regard something as behaving in an intelligent manner is determined as much by our own state of mind and training as by the properties of the object under consideration. If we are able to explain and predict its behaviour or if there seems to be little underlying plan, we have little temptation to imagine intelligence. With the same object therefore it is possible that one man would consider it as intelligent and another would not; the second man would have found out the rules of its behaviour.«⁵⁴

⁵² So auch in Masahiro Moris Ausführungen zum Uncanny Valley. Laut Mori ist es etwa das Ziel der Robotik einen künstlichen Menschen zu erschaffen, oder zumindest einen Roboter der in der Lage ist, jedwede menschliche Fähigkeit nachzubilden. Siehe dazu: Masahiro, Mori; Das Unheimliche Tal. Übersetzung: MacDorman, Karl et. al.; 2019; Hamburg: Textem; S. 212-214.

⁵³ Turing, Alan M.; Intelligent Machinery; 1948; In: Meltzer, Bernard Et. al. (Hsg.); Machine Intelligence 5; 1969; Edinburgh: Edinburgh University Press; S. 14

⁵⁴ Ebd. S. 23

Maschinelle Intelligenz ist also kein beweisbarer Zustand, oder Detailgrad, den eine Maschine erreichen muss, um als intelligent zu gelten, sondern ein Urteil. Zwar können Universal Machines mit einprogrammiertem Lernverhalten logisches Berechnungsvermögen und Flexibilität beweisen, und diese Möglichkeiten lassen sich mit logisch-intelligentem Verhalten von Menschen vergleichen. Doch schlussendlich ist eine Maschine nur dann intelligent, wenn sie es schafft, dass jene Menschen, die darüber urteilen, etwa von ihrem logischen Berechnungsvermögen getäuscht werden. Einerseits führt diese Erkenntnis bereits auf den Turing Test (der eigentlich den Titel »imitation game«⁵⁵ trägt) hin, mit dem Ziel Menschen über eine bestimmte Zeitdauer hinweg zu täuschen. Andererseits beschreibt Turing aber auch, dass intelligentes Verhalten von Maschinen immer auch die Imitation von Lebewesen (beziehungsweise Menschen) erfordert. In manchen Aspekten, wie der mathematischen Berechnung, oder der Speicherung von quantifizierbaren Informationen in Datenbanken, fällt diese Imitation Maschinen sehr leicht, und wird aufgrund ihrer Komplexität auch schnell für Menschen unverständlich. Somit wirkt (und wird) die Maschine intelligent.

Doch das Ziel⁵⁶ ist es eine intelligente Maschine zu entwickeln, die mehr als logische und statistische Aufgaben bearbeiten kann; indem technologische Entwicklungen es entweder erlauben, mehr als das logische Denkvermögen⁵⁷ zu imitieren, oder es schaffen die gesamte menschliche Existenz mittels Nullen und Einsen zu quantifizieren, und in Form eines Computerprogramms zu replizieren. Ersteres, und die unlösbaren Probleme hinter dem Versuch, thematisiert Joseph Weizenbaum in *Computer Power and Human Reason* (1976). Jener setzt damit auch den Rahmen fest, was wir überhaupt von (mehr oder weniger) intelligenten Maschinen, wie etwa smarten Kommunikationsassistenten, erwarten können.

⁵⁵ Der Begriff und das Konzept eines Tests, der die Fähigkeit von Maschinen, Menschen zu imitieren, messen soll, wird dabei erstmals 1950 von Turing formuliert: Turing, Alan M.; I. – Computing Machinery and Intelligence. In: *Mind*, Volume LIX, Issue 236; 1950; Oxford: Oxford University Press; S. 433

⁵⁶ Wie es zumindest Mori bezüglich Roboter, sowie Weizenbaum bezüglich früher Künstlicher Intelligenzforschung beschreibt, und wie sich durch das Ziel der singulären Maschine bei Turing ableiten lässt.

⁵⁷ Also beispielsweise, Emotionen

Letzteres ist der wesentliche Anspruch von Affective Computing, als Methode, um Emotionen (und in letzter Instanz menschliche Existenzen) zu messen, zu speichern, und erfolgreich zu vermitteln. Diese Konzepte kommen auch bei smarten Assistenten, im Rahmen der technologischen Möglichkeiten zum Einsatz.

In gewisser Weise referenzieren dabei auch Kubrick und Clarke mit HAL 9000 ganz bewusst die Ziele, Thesen und Wünsche der frühen Computerforschung. Zum einen beurteilen sie HAL 9000 bereits bevor jener in der Handlung auftritt als intelligent und verweisen auch explizit darauf, dass die Maschine es schafft, den Turing Test (auf unbestimmte Zeit) zu bestehen.⁵⁸ So zeigt HAL die dystopische Zukunft auf, die eintreten kann, wenn entweder Maschinen beginnen tatsächlich emotional zu handeln, oder wenn jene vorschnell als fähig beurteilt werden – obwohl sie etwa nicht die notwendige (emotionale) Standfestigkeit besitzen, um das wahre Ziel der Weltraummission geheimhalten zu können. Zum anderen bezieht sich ein Maschine die sich emotional verhält (und letztendlich die Nerven verliert) auch immer auf menschliches Verhalten:

»Very often the use of these devices serves as an analogy for the human condition, either how we as a species are becoming more machine-like and losing our own humanity or as an evolutionary extension of where we are headed and how a more powerful brain might function«⁵⁹

Schließlich, kann die Figur des HAL 9000 ebenso als eine Metapher zur Abhängigkeit von digitalen Systemen gesehen werden. Ein Beispiel, dafür (unabhängig von etwaigen Prophezeiungen), zeigt sich etwa bei der Technologie des Apollo Guidance Computers (AGC),⁶⁰ der zwar eine rudimentäre digitale Technologie darstellte, dessen Funktionieren aber über Leben und Tod der Astronauten entschied. Daher ist es naheliegend, dass Kubrick und Clarke mit HAL 9000 auf diese spezielle Computertypologie Bezug nahmen:

⁵⁸ Siehe dazu: Clarke, Arthur C.; 2001: A Space Odyssey; 2018; London: Orbit S. 99

⁵⁹ Frinzi, Joe R.; Kubrick's Monolith. The Art and Mystery of 2001: A Space Odyssey; 2017; Jefferson: McFarland & Company; S. 52

⁶⁰ Siehe dazu: The Apollo guidance computer: Hardware. In: Computers in Spaceflight: The NASA Experience; Chapter Two. Computers On Board The Apollo Spacecraft; (<https://history.nasa.gov/computers/Ch2-5.html> Abgerufen am 1. Dezember 2022)

Dass der erste AGC tatsächlich einwandfrei funktionieren würde,⁶¹ konnte nämlich nicht garantiert werden.⁶² Letztlich ist es die entscheidende Frage, ob HAL 9000 wegen seiner Handlungsmacht oder der vermeintlichen (auch emotionalen) Intelligenz als bedrohlich auftritt. Dies wird schließlich anhand des Assistierte Schlafs, und von Theorien zu digitaler Dienerschaft, erörtert. Davor müssen jedoch die Grenzen der Künstlichen Intelligenz thematisiert werden, und wie smarte Kommunikationsassistenten versuchen, diese mittels Affective Computing zu umgehen.

Die Grenzen von Künstlicher Intelligenz

In *Computer Power and Human Reason* (1976) analysiert Weizenbaum vordergründig die Frage wo die Fähigkeiten, Möglichkeiten aber auch Grenzen von Künstlicher Intelligenz liegen. Dabei stellt er fest, dass es mehrere Probleme gibt, die der Entwicklung einer Künstlichen Intelligenz, die mehr als nur eine intelligent arbeitende Maschine sein soll, im Weg stehen.

»First [...], the ability of even the most advanced of currently existing computer systems to acquire information by means other than [...] ›being spoon-fed‹ is still extremely limited. [...] Second, it is not obvious that all human knowledge is encodable in ›information structures,‹ however complex. A human may know, for example, just what kind of emotional impact touching another person's hand will have both on the other person and on himself. The acquisition of that knowledge is certainly not a function of the brain alone; [...] The knowledge involved is in part kinesthetic; its acquisition involves a hand, to say the very least. There are, in other words, some things humans know by the virtue of having a human body. No organism that does not have a human body can know these things in the same way humans know them. Every symbolic representation of them must lose some information that is essential for some human purposes. Third [...], there are some things people come to know only as a consequence of having been treated as human beings by other human beings. [...] Fourth, and finally, even the kinds of knowledge that appear superficially to be communicable from one human being to another in language alone are in fact not altogether

⁶¹ Sowohl in der fehlenden Gravitation und Atmosphäre im Weltraum, als auch im Bezug auf die Programmierung. Siehe dazu: Verification and Validation. In: Computers in Spaceflight: The NASA Experience; Chapter Two. Computers On Board The Apollo Spacecraft; (<https://history.nasa.gov/computers/Ch2-6.html> Abgerufen am 1. Dezember 2022)

⁶² Trotzdem führte die NASA etwa zwischen 1000 und 1200 Testläufe durch. Ebd. (<https://history.nasa.gov/computers/Ch2-6.html> Abgerufen am 1. Dezember 2022)

so communicable. [...] even in abstract information theory, the ›information content‹ of a message is not a function of the message alone but depends crucially on the state of knowledge, on the expectations, of the receiver.«⁶³

Diese Probleme lassen sich dabei zum Teil durch die Datengewinnung beheben (oder zumindest minimieren). Etwa indem mehr und mehr (Zusatz-) Informationen aufgezeichnet werden, um die maschinelle Reproduktion, also die symbolische Kopie, genauer an die Realität anzunähern. Wie in anderen Medien⁶⁴ auch, kann der auf dem Speichermedium gesicherte Inhalt natürlich niemals die Realität exakt abbilden. Doch solange der Detailgrad konsistent ist,⁶⁵ und die zugrundeliegenden Computermodelle akkurat arbeiten, können auch Emotionen, oder körperliches Empfinden (etwa mittels Affective Computing) relativ zielsicher gespeichert und vermittelt werden. Zum Teil sind die Unterschiede zwischen Menschen und Maschinen aber auch unüberwindbar. Weizenbaum macht das Grundsatzproblem wie folgt deutlich:

»Wether or not this program can be realized depends on whether man really is merely a species of the genus ›information-processing system‹ or whether he is more than that. Man is not a machine. [...] although man most certainly processes information, he does not necessarily process it in the way computers do. Computers and men are not species of the same genus.«⁶⁶

Das entscheidende Problem – der Aspekt, der die Imitation damit grundsätzlich unmöglich macht – ist somit, dass Menschen sich in ihrem Denken, Tun und Sein nicht immer nachvollziehbar verhalten. Ein Beispiel ist für Weizenbaum dabei die Kommunikation mittels Sprache. Verschiedene Äußerungen haben in verschiedenen Kontexten verschiedene Bedeutungen.

»[...] even in abstract information theory, the ›information content‹ of a message is not a function of the message alone but depends crucially on the state of knowledge, on the expectations, of the receiver. The Message ›Am arriving on 7 o'clock plane, love Bill‹ has different information

⁶³ Weizenbaum, Joseph, Computer Power and Human Reason. From Judgment to Calculation; 1976; San Francisco: W.H. Freeman S. 208 f.

⁶⁴ Dies umfasst genau genommen ebenso analoge Medien, wie Sprache und Schrift.

⁶⁵ Um den Uncanny-Valley-Effekt zu vermeiden.

⁶⁶ Weizenbaum, Joseph, Computer Power and Human Reason. From Judgment to Calculation; 1976; San Francisco: W.H. Freeman S. 203

content for Bill's wife, who knew he was coming home, but not on precisely what airplane, than for a girl who wasn't expecting Bill at all and who is surprised by his declaration of love. [...] The human use of language manifests human memory. And that is a quite different thing than the store of the computer, which has been anthropomorphized into ›memory.«⁶⁷

Mittlerweile sind diverse Programme zur Spracherkennung zwar in der Lage, einige Subtexte und auch kontextbezogene Formulierungen zu erfassen. Doch effektiv passiert nicht mehr, als dass bestimmte Schlagwörter erkannt und, mittels Wahrscheinlichkeitsrechnung, interpretiert werden. Wenn es das Ziel ist, dass smarte Kommunikationsassistenten, aus diesen Schlagwörtern eine Handlungsanweisung ableiten – wie die Wohnräume vorzuheizen, oder ein Haushaltsgerät zu aktivieren – dann ist eine intelligente Interaktion möglich: Einerseits gibt es nur eine begrenzte Anzahl an sinnvollen Argumentationen um die eigene Ankunftszeit mitzuteilen (und es müssen nur Zahlen sowie Orte erkannt/analysiert werden). Andererseits ist ein Scheitern des Assistenten, zwar eventuell frustrierend, aber weitgehend harmlos. Dass Menschen sich nicht nach mathematischen (oder logischen) Prinzipien verhalten, und Informationen die Menschen⁶⁸ kommunizieren nicht zwingend für bare Münze genommen werden dürfen, sind dabei zwei essentielle Schwierigkeiten, die mittels Affective Computing umgangen oder gelöst werden müssen, und sind maßgebliche Probleme für die digitalen Erfassung von Emotional Intelligence.

Das Problem mit Emotional Intelligence

Die Definition des Begriffs Emotional Intelligence ist dabei keineswegs eindeutig. So gibt es mehrere theoretische Annäherungen. Im Wesentlichen beschreibt die Emotionale Intelligenz die – bei Menschen beobachtbare – Qualität, Gefühle (und Gefühlslagen) aufnehmen, verarbeiten und produzieren zu können. Darüber hinaus stellen Theorien zur Emotional Intelligence fest, dass Emotionen eine wesentliche Rolle in intelligentem Verhalten spielen, und damit sowohl den Verstand, wie auch das logische Denken maßgeblich beeinflussen.

⁶⁷ Ebd. S. 200

⁶⁸ Untereinander, wie auch einer Maschine gegenüber.

Somit kritisieren Theorien der Emotional Intelligence vor allem Ansätze zur Intelligenzermittlung durch den Intelligenzquotienten (IQ), der menschliche Intelligenz auf das mathematisch, logische Denkvermögen reduziert. Peter Salovey und John D. Mayer (1993) definieren Emotional Intelligence dabei vor allem über die Fähigkeiten, die damit einhergehen. So definieren die beiden den Begriff wie folgt:

»Emotional intelligence is a type of social intelligence that involves the ability to monitor one's own and others' emotions, to discriminate among them, and to use the information to guide one's thinking and actions [...]. The scope of emotional intelligence includes the verbal and nonverbal appraisal and expression of emotion, the regulation of emotion in the self and others, and the utilization of emotional content in problem solving.«⁶⁹

Emotional Intelligence umfasst damit zum einen die soziale Kompetenz (bei etwa zwischenmenschlicher Kommunikation), Subtexte lesen zu können, aber auch kulturell spezifische Eigenarten bewusst oder unterbewusst aufzunehmen und zu vermitteln. Zum zweiten beschreibt Emotional Intelligence die Fähigkeit, die eigene Gefühlswelt zu regulieren, Emotionen einzuordnen, und Affekte zuzulassen oder überwinden zu können. Und schließlich beschreibt Emotional Intelligence den Umstand, dass Emotionen den Geisteszustand und damit das Denken unmittelbar beeinflussen.

»People who experience varying emotions will also experience varying thoughts along with them. People in good moods perceive positive events as more likely, and negative events as less likely to occur; [...] such changes may, in consequence, enhance their fluency in generating alternative outcomes;«⁷⁰

Emotionen, und die Fähigkeiten diese maßzuregeln, üben damit gewaltigen Einfluss auf die Interaktion mit digitalen Kommunikationsassistenten aus. – Vor allem weil die Art und Weise, sowie die Intensität in der Emotionen etwa logische Denkabläufe, Erwartungshaltungen, oder Bedürfnisse beeinflussen, sich von Person zu Person unterscheidet. Menschliches Verhalten funktioniert also nicht nach mathematisch logischen Prinzipien,

⁶⁹ Salovey, Peter; Mayer, John D.; The Intelligence of Emotional Intelligence. In: Intelligence, Volume 17, Issue 4; 1993; Elsevier; nach: Salovey, Peter; Mayer, John D.; Emotional Intelligence. In: Imagination, Cognition, and Personality, Volume 9, Issue 3; 1990; Sage; S. 433

⁷⁰ Siehe dazu: Salovey, Peter; Mayer, John D.; The Intelligence of Emotional Intelligence. In: Intelligence, Volume 17, Issue 4; 1993; Elsevier; S. 436

und variiert von Mensch zu Mensch. Wie sollen (und können) demzufolge Menschen imitiert werden? – Entweder durch eine bewusste Abstraktion (beziehungsweise Reduktion) oder durch extrem detaillierte (auf einzelne Nutzer*innen zugeschnittene) Datensätze.

Weizenbaum (1976) beschreibt dabei die Möglichkeit, nur die Performance zu imitieren. Dieser Performanceaspekt sieht vor, etwa die Funktionalität des menschlichen Sprachverständnisses nachzubilden, ohne dabei zwingend Gehör oder Gehirn nachbauen zu müssen.⁷¹ Diese Reduktion auf die Performance bedeutet ebenso, dass besonders der Eindruck (oder die Illusion) von menschen-imitierendem Verhalten zählt. Das zeigt sich etwa an Weizenbaums eigenem Chatprogramm, ELIZA, das als der erste Chatroboter gilt, bei dem die Nutzer*innen über die Intelligenz der Maschine getäuscht werden. Selbst wenn diese Täuschung nicht absichtlich passiert, so ist sie doch ein bewusst toleriertes Begleitphänomen bei dem Versuch, menschenähnlich zu wirken.

»ELIZA was a program consisting mainly of general methods for analyzing sentences and sentence fragments, locating so-called key words in texts, assembling sentences from fragments, and so on. In a sense ELIZA [...] commanded a set of techniques but [...] had nothing of her own to say. The script, in turn, was a set of rules which permitted [...] [the machine] to improvise on whatever resources provided. The first extensive script I prepared [...] was one that enabled it to parody the responses of a nondirective psychotherapist in an initial psychiatric interview. [...] I reasoned, a psychiatrist can reflect [...] »My mommy took my teddy bear away from me,« by saying »Tell me more about your parents,« without really having to know anything about teddy bears, [...] some of which, it might use in later responses.«⁷²

Das Psychotherapie-Skript von ELIZA versucht also eine*n Psychotherapeut*in performativ zu imitieren. Um dies zu erreichen, schreibt Weizenbaum ein Skript, laut dem ELIZA den Nutzer*innen offene Gegenfragen stellt, die auf vorher definierte Schlüsselbegriffe reagieren, ohne dabei irgendein Verständnis, außer der Transformation von Text, zu haben.

⁷¹ Siehe dazu: Weizenbaum, Joseph, *Computer Power and Human Reason. From Judgment to Calculation*; 1976; San Francisco: W.H. Freeman S. 164 f.

⁷² Weizenbaum, Joseph, *Computer Power and Human Reason. From Judgment to Calculation*; 1976; San Francisco: W.H. Freeman S. 188 f.

»Nevertheless, ELIZA created the most remarkable illusion [...]. People who knew very well that they were conversing with a machine soon forgot that fact [...]. This illusion was especially strong and most tenaciously clung to among people who knew little or nothing about computers. They would [...] converse with the system in private, and would after conversing with it for a time, insist, in spite of my explanations, that the machine really understood them.«⁷³

In gewisser Weise spiegelt sich diese Illusion von Intelligenz bereits in Turings Ausführungen⁷⁴ wider. Und auch die Zuschreibung von Emotional Intelligence (oder gar Lebendigkeit) ist ein Urteil, das Menschen über Maschinen treffen, wenn entweder die Prozesse der Maschine nicht mehr nachvollzogen werden können, oder aber auch wenn die Maschine es schafft, eine überzeugende Performance abzuliefern. Im besten Fall ist dabei jedoch eine intelligente Maschine eine nützliche Illusion, im schlimmsten Fall bewusste Täuschung.

Für die Beziehung zwischen smarten Kommunikationsassistenten und ihren Nutzer*innen bedeutet dies, dass jedwede Interaktionsdynamik nie über diesen Punkt hinaus wachsen kann. Entweder das vermeintlich intelligente Gegenüber ist eine Illusion, der sich die Nutzer*innen hingeben, im vollen Bewusstsein darüber, dass der Eindruck von Intelligenz (über etwaige mathematische Fähigkeiten hinaus) aus Algorithmen besteht, die emotionale Kompetenz zum Zweck der vermeintlich intuitiveren Interaktion vorgaukeln. Oder das Gegenüber ist technologisch so sehr entwickelt und/oder abgeschottet, dass kein intellektueller Hebel (seitens der Nutzer*innen) mehr verwendet werden kann, um die Intelligenz der Maschine zu entkräften. Doch dann ist (und bleibt) nicht erkennbar, ob die Maschine tatsächlich intelligent ist, oder nur eine kunstvolle Täuschung.⁷⁵

Der Zugang des Affective Computing will jedoch diese Grenze – an der nicht mehr zwischen Täuschung oder wahrhafter Intelligenz unterschieden werden kann – als vermeintliches Ideal erreichen, und zwar unter dem Aufhänger einer besseren Interaktion.

⁷³ Ebd. S. 189

⁷⁴ Siehe dazu: Turings Erörterungen zu Intelligenz als Zuschreibung, beziehungsweise Urteil. Turing, Alan M.; *Intelligent Machinery*; 1948. In: Meltzer, Bernard Et. al. (Hsg.); *Machine Intelligence 5*; 1969; Edinburgh: Edinburgh University Press; S. 14

⁷⁵ Dies spiegelt das Grundsatzproblem wider, dass Intelligenz bei Menschen zwar approximiert, aber niemals definiert werden kann.

Ansätze des Affective Computing

Affective Computing ist ein Begriff der jüngeren Computerwissenschaft. Der Begriff wurde von Rosalind Picard (2000) geprägt. Jene beschreibt Affective Computing dabei wie folgt:

»[...] computing that relates to, arises from, or deliberately influences emotions. [...] Affective computing includes implementing emotions and therefore can aid the development and testing of new and old emotion theories. However, affective computing also includes many other things, such as giving a computer the ability to recognize and express emotions, developing its ability to respond intelligently to human emotion, and enabling it to regulate and utilize its emotions.«⁷⁶

Das Ziel von Affective Computing ist es folglich, dass Emotionen gemessen, verarbeitet, gespeichert, und auf Knopfdruck wieder abgerufen werden können, und zwar durch digitale Computer. Picard widmet sich dabei einzeln den jeweils dafür notwendigen Komponenten. Wenn ein affektiver Computer etwa Emotionen erkennen und aufzeichnen soll, dann muss dafür das Verhalten von Menschen so detailliert wie möglich – bis hin zu den feinsten physiologischen Details – vermessen werden. Picard fasst dabei jedwede Abweichung vom vermeintlichen Standardwert (dem emotionalen Ausgangspunkt), egal ob bewusst oder unbewusst, absichtlich, oder unfreiwillig, unter dem Begriff der »sentic modulation«⁷⁷ zusammen.

»Sentic modulation, such as voice inflection, facial expression, and posture is the physical means by which an emotional state is typically expressed, [...]. In fact few people are good at articulating their emotional state, but expressing it through sentic modulation is natural, and usually subconscious. [...] There are many physiological responses that vary with time and that might potentially be combined to assist in recognition of emotional states. These include heart rate, diastolic and systolic blood pressure, pulse, pupillary dilation, respiration, skin conductance and color, and temperature. [...] Given that a person is experiencing an emotion, e.g., hate, then sentic modulation may result in a tense voice, glaring expression, or finger pressure strongly away from the body. Respiration rate and heart rate may also increase.«⁷⁸

⁷⁶ Picard, Rosalind W.; Affective Computing; 1997; Massachusetts: MIT Press; S. 3

⁷⁷ Ebd. S. 25

⁷⁸ Ebd. S. 25, S. 29

Nach der Datenerhebung werden die erhobenen Datensätze in einem zweiten Schritt miteinander verknüpft, und interpretiert. Um menschliche Emotionen korrekt einordnen zu können, ist es dabei notwendig, menschliches Verhalten mittels (fein aufgelösten) Algorithmen nachzubilden. Zwar hält Picard fest, dass keine gesamtheitliche Reproduktion des Menschen benötigt wird.⁷⁹ Doch ein akkurater Algorithmus, der die erfassten Daten – also im Idealfall jede menschliche Emotion – vermeintlich korrekt auswertet, stellt im Endeffekt den Anspruch, ein perfektes Computermodell von menschlicher Emotional Intelligence zu sein. Korrekt bedeutet dabei, dass in einem Test, Computer zu demselben Ergebnis kommen, wie Menschen.

»A digital video of a person [...] could be observed by both a group of humans and a group of computers. The humans and computers are asked what emotion the person in the video is expressing. When the group of computers and the group of humans respond with the same distribution of answers, then we could say that the computers are recognizing emotions as well as the humans.«⁸⁰

Abgesehen davon, dass solch ein Test unzählige Male wiederholt werden müsste, und sich nur auf ein Modell, also eine Programmierung/Software bezieht, die danach nie mehr modifiziert oder erweitert werden dürfte, zeigen sich an solchen Tests Probleme der Repräsentation. Denn egal wie oft dieser Test wiederholt werden würde, es könnten nur begrenzte Personen im Video auftreten, und ein begrenztes Urteil über Gefühlszustände getroffen werden. Dieses Beispiel, wie Picard vermeintlich korrekte Erkenntnisse verifizieren will, zeigt, dass es unmöglich ist mehr als nur bestimmte Aspekte wie sich Menschen verhalten, digital zu approximieren. Und es stellt sich dabei immer die Frage, welche Menschen die Vorbilder für die digitalen Modelle sind.

Im Umkehrschluss heißt das nicht nur, dass soziale und kulturelle Urteile über die Programmierer*innen ihren Weg in das Computermodell finden. Sondern es bedeutet, dass auch bei der Abstraktion oder Reduktion des Computermodells bestimmte Menschen oder bestimmte Vorannahmen zum Menschsein im Hintergrund stehen. Picard versucht dieses Problem zu umschiffen, indem sie auf individuelle Nutzerprofile

⁷⁹ Ebd. S. 49

⁸⁰ Ebd. S. 51

fokussiert.⁸¹ Wenn Algorithmen auf die jeweiligen Nutzer*innen zugeschnitten werden, umgeht dies vor allem das Problem, dass sentic modulation, pro Nutzer*in unterschiedlich ausgeprägt ist.

»In many applications it is only necessary that your personal computer be able to recognize your affect; it can then translate this information to others if you wish.«⁸²

Doch selbst wenn die Datengewinnung und -auswertung auf individuell zugeschnittene Standardwerte und Abweichungen hin geeicht ist, kann die zugrundeliegende Programmierung durch technologische Rahmenbedingungen Einfluss auf das Computermodell ausüben. Dies lässt sich schlicht und ergreifend als Bias bezeichnen, den die Programmierung von den Programmierer*innen übernimmt. Dies illustriert Toby Walsh (2022) etwa am Beispiel von Gesichtserkennungssoftware:

»A major ethical concern behind many of these calls is bias. [...] [facial recognition software] often works better on men than women, better on white people than Black people, and worst of all on Black women. And while some facial recognition software has been improved in response, significant biases remain. In June 2020, in the first known case of its type, a man in Detroit was arrested in front of his family for burglary because he was mistakenly identified by facial recognition software. It may come as no surprise the man was Black.«⁸³

Selbst wenn sich dabei der Bias, also bewusste, oder unbewusste Vorannahmen, nie ganz eliminieren lässt, kann in der Praxis natürlich ein plurales Menschsein in Computermodellen verstärkt werden.

»There is promising, if somewhat slow, progress on making facial recognition software less biased. This ranges from using more representative datasets (such as the 10K U.S. Adult Face database that consists of a large demographically balanced set of faces) to debiasing methods (such as selectively resampling biased data to make it less so).«⁸⁴

⁸¹ Ebd. S. 32

⁸² Ebd. S. 33

⁸³ Walsh, Toby; The troubling future for facial recognition software. In: Communications of the ACM, Volume 65, Issue 3; 2022; New York: Association for Computing Machinery; S. 35 f.

⁸⁴ Ebd. S. 36.

Dies ist aber ein aktiver Schritt. Zusätzlich sollte die Maschine im Idealfall nicht alleine Entscheidungen darüber treffen, ob etwa eine Person bei sich zuhause ist, oder gerade versucht einzubrechen. Die Entscheidungsgewalt von etwa Technologien zur Gesichtserkennung – die zur polizeilichen Überwachung eingesetzt werden – eröffnet dabei ebenso Fragen zur Handlungsmacht und Verantwortlichkeit von Maschinen. Diese gilt es schließlich, mittels Konzepten zu digitaler Dienerschaft zu thematisieren. Darüber hinaus widmet sich Rosalind W. Picard (2000) ebenso der Frage, wie Maschinen Emotionen akkurat vermitteln oder synthetisieren können. Für die etwaige Gefühlssynthese braucht das Computerprogramm dabei die Fähigkeit, mit den Nutzer*innen zu kommunizieren, und die Möglichkeit Symbole zu vermitteln, die Gefühlsreaktionen auslösen.

»The basic requirements for a computer to have the ability to express emotions is that the machine have channels of communication such as voice or image, and an ability to communicate affective information over those channels. For example, a computer or software agent that displays a face could use an expression such as a smile—like a Macintosh does, upon booting up.«⁸⁵

Affektive Maschinen sollen also etwa audiovisuelle Symbole nutzen um verschiedene Gefühlslagen zu übermitteln. Doch egal ob ein lächelndes Gesicht beim Einschalten des Computers oder ein veränderter Tonfall bei der Sprachsynthese⁸⁶ – die entscheidende Frage bleibt, warum Maschinen überhaupt Emotionen vermitteln sollen? Eine mögliche Antwort darauf ist, dass etwa ein Lächeln beim Einschalten eine Form von angenehmem Feedback, (etwa: der Computer funktioniert ordnungsgemäß) darstellt. Eine zweite Antwort ist, dass die Fähigkeit Emotionen mittels Symbolen auszudrücken, in der digitalen Kommunikation zwischen zwei Nutzer*innen (über zwei Computer) Anwendung findet, wie etwa anhand von Emojis. Schließlich stellt dieser emotional-aufgeladene Output jedoch eine Möglichkeit dar, wie eine intensivere Kommunikation zwischen Mensch und Maschine erreicht werden kann. Das Ziel hinter Affective Computing ist es nämlich die Affinität zu, und Akzeptanz von Maschinen zu verstärken:

⁸⁵ Picard, Rosalind W.; *Affective Computing*; 1997; Massachusetts: MIT Press; S. 56

⁸⁶ Siehe dazu etwa: Ebd. S. 58

»The priorities of your personal software agent need to shift with your affective state, and what it predicts would influence your feelings. [...] Suppose your news agent learned of a building that was bombed. It should predict that such a tragic event will cause you concern, especially if you know people in the town where it happened. If it learned by scanning your address book [...] that you had a dear friend there, it might choose to interrupt you right away with this news. It would also pre-fetch the phone number for that friend, so you could call immediately if you wanted too.«⁸⁷

In der praktischen Anwendung geht es also vor allem darum eine Performance abzuliefern, die die Interaktion mit Maschinen vereinfachen, intensivieren und besser auf die jeweilige Nutzer*in anpassen soll. Ob Illusion oder Täuschung ist dabei zweitrangig. Denn in der Methodik von Affective Computing ist es wichtiger ein, für bestimmte Bereiche funktionales Computermodell zu erzeugen, als Menschen zu imitieren. Allerdings lässt sich trotzdem ein schöpferischer Anspruch von künstlicher (emotionaler) Intelligenz, als – nach menschlichem Ebenbild geschaffen – wiederfinden. So kokettiert Picard sehr wohl mit etwaigen Ähnlichkeiten von Maschinen/Robotern und Menschen:

»Robot bodies, like human bodies, can encounter problems that could use some formst of sentic modulation. [...] I was shown a mobile robot [...] [that] stopped in its tracks. [...] It was [...] much later that they determined its buffers had overfilled, a very simple problem. In fact, an analogous problem happens to humans when overwhelmed with too much information. A human will appear confused, or her eyes may glaze over, and she may be unable to move mentally beyond the present point. Whether or not she is aware that her ›buffers are full,‹ this information is signaled to outsiders through subtle emotional expressions, indicating that it is useless to give her any more information. This robot could have helped its makers by having a spontaneous expression to indicate this internal state.«⁸⁸

Folglich muss festgehalten werden, dass Affective Computing zum einen sehr performance- und zweckorientiert gesehen werden kann, da Emotionen quantifiziert und durch Computermodelle interpretiert⁸⁹ werden. Außerdem soll durch Affective Computing auch vermeintlich besseres (weil emotional aufgeladenes) maschinelles Feedback erreicht werden, sowie die Kommunikation von zwei Menschen mittels Computerprogrammen

⁸⁷ Ebd. S. 102

⁸⁸ Ebd. S. 73

⁸⁹ Beziehungswiese rekonstruiert.

verbessert werden. Zum zweiten liegt dem Anspruch (und Wunsch) aber ebenso die Prämisse zugrunde, dass prinzipiell die gesamte menschliche Existenz durch einen Computerprozessor berechnet und auf einer Festplatte abgespeichert werden könne – die Technologie wäre nur noch nicht leistungsstark genug. Doch es ist (und bleibt) unmöglich selbst einzelne Nutzer*innen gesamtheitlich zu beschreiben. Zwar lassen sich Verhaltensweisen messen und approximieren, aber menschliches Denken, Verhalten, (und Sein) ist nicht statisch, sondern fluide. Maschinelle Berechnungen sind hingegen exakt, und zielorientiert.

Bei der Umsetzung von Methoden des Affective Computing in smarten Kommunikationsassistenten zeigen sich dabei ähnliche Potenziale, und Schwierigkeiten. Um eine funktionale Performance abzuliefern, müssen Assistenten daher zwei wesentliche Bedingungen erfüllen: Sie müssen sentic modulation über so viele (und diverse) Datenkanäle wie möglich aufzeichnen können. Und sie müssen die erhobenen Daten mit flexiblen Computermodellen funktional verarbeiten können.

Sensoren und Daten

Im Regelfall gelangen smarte Kommunikationsassistenten über diverse Messinstrumente, sogenannte Sensoren, an ihre Daten. Jeder Sensor lässt sich dabei als ein Datenkanal verstehen. In der Summe der (möglichst vielschichtigen) Aufzeichnung, kann dann nach Überschneidungen in den Daten gesucht, und damit ein Bild der Außenwelt erzeugt werden. Die Datenkanäle von smarten Kommunikationsassistenten lassen sich dabei in zwei Bereiche aufteilen – die internen, also die der Hardware von smarten Assistenten zugehörigen Sensoren – und die externen Sensoren, also zusätzliche Messobjekte, die ihre Daten an die Hardware von smarten Assistenten vermitteln. Sensoren die entweder nur zum Zweck der Datenerhebung, oder für zusätzliche Funktionen (wie etwa als Uhr) nah am Körper, oder direkt an der Haut getragen werden, werden als Wearables⁹⁰ bezeichnet.

»Wearables are miniature embedded computing systems worn by people. [...] Sensing is the primary functionality of wearables. [...] The data gathered is then processed by a [...]

⁹⁰ Die Kurzform für Wearable Technology.

microprocessor. Raw sensing data is typically discarded afterwards. The processed information, in much more compacted form, is stored and transferred to remote devices, e.g., a mobile phone, through low-power wireless interface [...].«⁹¹

Wie Williamson et. al. festhalten, sortieren die jeweiligen Wearables bereits die Daten, und senden (je nach Handlungsanweisung) bereits ausgewertete Datensätze an eine Datensammelstelle. Diese Zentralstelle ist der physische Ort an dem die unterschiedlichen Datensätze miteinander verknüpft werden. Doch nicht nur Wearables stellen Datenkanäle zu Verfügung, Smartphones sind ebenso mit einer Vielzahl an Sensoren ausgestattet, und auch in vielen alltäglichen Gegenständen⁹² werden mehr und mehr Sensoren verbaut. Ein wesentliches Schlagwort ist hier der Begriff des Internet of Things (IoT).⁹³ Das Internet der Dinge, bezeichnet dabei erstens die Verwendung von digitalen Technologien für die Datenübertragung, zweiten beschreibt der Begriff aber auch einen Meilenstein:

»IoT is simply the point in time when more “things or objects” were connected to the Internet than people.«⁹⁴

Drittens sind IoT-Technologien in ihrer Umsetzung besonders auf eine nahtlose Integration, wie etwa in alltägliche Gegenstände, fokussiert. Die Fähigkeit Daten zu sammeln und zu übermitteln, kann dabei auch zum wesentlichen Ziel des nunmehr smarten Objekts werden, doch die alltäglichen Gegenstände (etwa: eine Waschmaschine) müssen weiterhin ihre ursprünglichen Aufgaben erfüllen können. So ist es Bad Practice, wenn die smarte, digitale Nutzung erzwungen wird, und im Gegenzug die direkte (analoge) Interaktion zurückgenommen wird.

⁹¹ Williamson, James, et. al.; Data sensing and analysis: Challenges for wearables; In: 20th Asia and South Pacific Design Automation Conference, ASP-DAC; 2015; S. 136

⁹² Und auch in Haushaltsgeräten.

⁹³ Der Begriff wurde erstmalig 1985 von einem der Pioniere des US-amerikanischen Mobilfunks, Peter T. Lewis in einer Rede verwendet, und bezeichnete die Möglichkeit nicht nur menschliche Kommunikation über das Internet auszutauschen, sondern auch Datensätze hin und her zu senden. Siehe dazu: (<http://www.chetansharma.com/correcting-the-iot-history/> Abgerufen am 22. November 2022)

⁹⁴ Evans, Dave; The Internet of Things. How the Next Evolution of the Internet Is Changing Everything; 2011; Cisco IBSG; S. 2

In der Praxis bedeutet dies, dass mehr und mehr elektronische Alltagsgeräte, zusätzlich zur bereits vorhandenen Funktion, mit Sensoren, und der notwendigen digitalen Verarbeitungs-Infrastruktur⁹⁵ ausgestattet werden. Smarte Kommunikationsassistenten nutzen dabei die folgenden Anwendungsbereiche von IoT-Geräten zur Datenerfassung:

»Devices (wearables and ingestibles) to monitor and maintain human health and wellness; disease management, increased fitness, higher productivity [as well as] [...] Home controllers and security systems [...]«⁹⁶

Diverse Wearables (Tracker, Pulsmesser, smarte Uhren, etc.), Überwachungskameras, smarte Kühlschränke, Smarte Waschmaschinen/Trockner, Bewegungsmelder, Luftfeuchtigkeitmessgerätee und smarte Thermostate; All diese Geräte des Internet of Things sind mögliche Datenkanäle, auf die smarte Kommunikationsassistenten, nach der Aktivierung durch die Nutzer*in, zugreifen können. Die Kommunikation funktioniert dabei meist über etablierte Netzwerkprotokolle, wie Bluetooth oder drahtlose IP-Netzwerke. Wearables, wie auch smarte Alltagsobjekte, lassen sich somit als IoT-Geräte definieren, da sie über digitale Netzwerke mit anderen Maschinen kommunizieren. – Das Smartphone hingegen ist kein IoT-Gerät (mehr):

»All of the definitions describe scenarios in which network connectivity and computing capability extends to a constellation of objects, devices, sensors, and everyday items that are not ordinarily considered to be ›computers‹; this allows the devices to generate, exchange, and consume data, often with minimal human intervention.«⁹⁷

Denn gegenwärtige Smartphones sind (zumindest im Rahmen des Internet of Things) zu vollwertigen Computern avanciert;⁹⁸ so besitzen jene einige Besonderheiten, die sie von

⁹⁵ Etwa Mikroprozessoren zur Vor-Verarbeitung, Speicherung, und Übermittlung an einen Hub, sowie Technologien zur digitalen Fernsteuerung.

⁹⁶ Manyika, James et.al.; The Internet of Things: Mapping the Value beyond the Hype; 2015; McKinsey Global Institute: McKinsey & Company; S. 3

⁹⁷ Singhania, Vivek; The Internet of Things: An Overview. Understanding the Issues and Challenges of a More Connected World; 2015; Genf, Reston: The Internet Society (ISOC); S. 12

⁹⁸ Siehe dazu etwa: Perazzo, Pericle et al.; Performance Evaluation of Attribute-Based Encryption on Constrained IoT Devices. In: Computer Communications, Volume 170; 2021; Sciencedirect: Elsevier; S. 1

IoT-Peripheriegeräten abheben: Erstens besitzen gegenwärtige Smartphones mittlerweile genug Rechenleistung und Speicherkapazität um einen Softwareclient, eines smarten Assistenten, installieren und betreiben zu können. Zweitens erlaubt das Touchscreen Interface (samt digitaler Tastatur) etwaige Softwareeinstellungen einfach anzupassen. Und drittens können durch eine stetige Internetverbindung (auch im Mobilfunknetz) die smarten Assistenten laufend synchronisiert und aktualisiert werden.⁹⁹

Zusätzlich können smarte Kommunikationsassistenten aber auch auf sogenannten Smart Home Hubs installiert werden. Diese Technologie besteht im Wesentlichen aus Mikroprozessoren mit diversen Netzwerkanbindungen, verbaut in einer möglichst unauffälligen Form (wie etwa einer grauen Dose). Hier kann meist nur über Sprachsteuerung direkt mit dem Hub interagiert werden – für Softwareeinstellungen, muss mit Smartphone oder Computer zugegriffen werden. Außerdem können Hubs ebenso IoT-Peripheriegeräte sein, wenn sie beispielsweise ein Geflecht (Mesh-Netzwerk) bilden, bei dem ein Hub das Rechenzentrum darstellt, und die anderen, als Peripheriegeräte, Daten zwischenspeichern und weiterleiten. Smart Home Hubs finden schließlich vor allem im Kontext von Home-Automation (der Automatisierung von Aufgaben im eigenen Zuhause) ihre Anwendung.¹⁰⁰ Während Smart Home Hubs dabei maßgeblich von den, im Netzwerk eingegliederten IoT-Peripheriegeräten abhängig sind, die ihnen Datensätze übermitteln müssen, können Smartphones selbst (ohne Unterstützung von Peripheriegeräten) eine Vielzahl an Daten erheben.

So besitzen jene im Regelfall eine Vielzahl an Sensoren, die bereits im Gerät verbaut sind, und konstant Daten erheben. Eine Methode (die auch Smart Hubs beherrschen¹⁰¹) ist die Audioaufzeichnung über das Mikrofon.

⁹⁹ Darüber hinaus werden in der Forschung und Entwicklung andere Ansprüche an Smartphone, als an IoT-Geräte gestellt. So sollen IoT-Geräte etwa meist möglichst kompakt sein, keine Wartung benötigen, und möglichst lange unter diversen Bedingungen funktionieren. Siehe dazu etwa: Williamson, James, et. al.; Data sensing and analysis: Challenges for wearables; In: 20th Asia and South Pacific Design Automation Conference, ASP-DAC; 2015; S. 136

¹⁰⁰ Siehe dazu: (<https://uk.pcmag.com/home-automation-hubs/35777/what-is-a-smart-home-hub-and-do-you-need-one> Abgerufen 23. November 2022)

¹⁰¹ Für Sprachsteuerung müssen Mikrofone in Smart Home Hubs verbaut werden.

Smartphones verfügen darüber hinaus meist über Sensoren zur Bestimmung des Standorts, der Beschleunigung und Geschwindigkeit (von Bewegungen), der damit einhergehenden G-Kräfte, der Rotation, der Stärke des Erdmagnetfelds und der Orientierung,¹⁰² sowie der Höhe (anhand des Luftdrucks). Zusätzlich dazu, besitzen jene nicht nur eine rückseitige- sondern meist auch eine Frontkamera, die neben der Möglichkeit Videosignale abzurufen, auch Daten über die Helligkeit ausliest. All diese Sensoren lassen sich dabei, sowohl unter dem Blickwinkel des Affective Computing, als auch mit dem anthropomorphen Anspruch Mori's als nachgebildete (oder erweiterte) menschliche Sinne begreifen. Einerseits können Sensoren auf funktionaler Ebene nämlich – wie menschliche Sinne – als das ermächtigende Potenzial zur Selbsterforschung der Um- und Außenwelt begriffen werden. und andererseits bilden selbst die modernsten Sensoren, im Grund genommen menschliche Sinne nach oder erweitern sie. – Dabei muss es noch gar nicht darum gehen, Menschen (oder gar Emotionen) zu vermessen. Sondern die Smartphone-Kamera, kann, ganz fundamental als verlängertes (und zusätzliches) Auge verstanden werden.¹⁰³

All diese Sensoren erlauben es jedoch erst im Zusammenspiel mit der Software, aus den gewonnenen Daten eine messbare Logik abzuleiten. Hier gilt es zwei Ebenen zu unterscheiden: Die Datenbank, die auch als Framework bezeichnet wird und die zentrale Schnittstelle des gesamten Systems darstellt. Und die Computermodelle (sowie Cloud-gestützte Berechnung) mit denen smarte Kommunikationsassistenten die im Framework gesammelten Daten auswerten und interpretieren. Im Betriebssystem des Smartphones verankert, münden alle Datensätze zuerst im Framework, das die Daten von etwaigen Messfehlern bereinigt und bereits statistisch aufbereitet. In einem weiteren Schritt können dann diverse Programme – und damit auch smarte Assistenten – auf diese aufbereiteten Daten zugreifen. Allerdings ist bereits diese Auswertung von Datensätzen nur bedingt überprüfbar. Zwar können die im Framework gespeicherten Daten im Regelfall durch ein (vor-)installiertes Dienstprogramm betrachtet werden. Doch ob das Framework die Daten angleicht, rundet oder anderweitig verändert, kann nicht mehr nachvollzogen werden, da

¹⁰² Angabe in Grad, ob waagrecht, senkrecht und Abhängigkeit von Norden

¹⁰³ Dieses Phänomen, lässt schließlich als Imitation der Performance, im Sinne Weizenbaums, fassen. Weizenbaum, Joseph, Computer Power and Human Reason. From Judgment to Calculation; 1976; San Francisco: W.H. Freeman S. 164 f.

die Software proprietär gehalten wird, und der Quelltext¹⁰⁴ nicht einsehbar ist. Dem kommt hinzu, dass besonders externe IoT-Peripheriegeräte ihre Daten vor-aufbereiten. Auch hier ist unklar (nicht nachvollziehbar) welche Datensätze eventuell verworfen werden, und wie die verbleibenden Daten ausgewertet werden.¹⁰⁵ Außerdem stellt sich bei dieser, proprietären Software ebenso die Frage, wie akkurat und, im Falle der Schlafvermessung, medizinisch wertvoll die erhobenen Daten sind. Um mehr als nur Gimmick zu sein, müssten die jeweiligen Sensoren und Software nämlich auch eine medizinische Zertifizierung besitzen. In welchem Ausmaß dabei Schlaftracker überhaupt medizinisch relevante Aussagen treffen können, soll schließlich am Beispiel der Software *Sleepwatch™ by Bodymatter* untersucht werden.

Davor müssen jedoch jene Computermodelle thematisiert werden, die die analysierten Daten interpretieren, und Nutzer*innenprofile angleichen, sowie die notwendige Flexibilität besitzen um affektive Kommunikation zu ermöglichen. Diese Algorithmen sind dabei meist eine Form von Künstlichen Neuronalen Netzwerken.

Nutzer*innenprofile und Künstliche Neuronale Netzwerke

Das Ziel hinter Künstlichen Neuronalen Netzwerken (KNN) ist folgendes: Bereits bei den Erörterungen von Turing zur Universal Machine zeichnet sich ein Konzept ab, wie Lernverhalten bei Maschinen erzeugt werden kann. Und auch Weizenbaum führt aus, dass lernende Maschinen explizit Menschen imitieren, aber das Gelernte nie so detailliert verstehen könnten, wie Menschen dies tun. Doch die Möglichkeit über die Bedeutung von Datensätzen zu lernen, und diese dann ohne äußere Einwirkungen zu verstehen, wie auch dieses Wissen dann anwenden zu können, ist ein wesentlicher Schritt hin zu intelligentem Verhalten. Rosalind W. Picard (2000) beschreibt das Problem, das mit KNNs gelöst werden kann, wie folgt:

¹⁰⁴ Das heißt: Die Programmierung.

¹⁰⁵ Außerdem gibt es keine einheitlichen Standards für die Datenaufbereitung. Siehe dazu: Miraz, Mahdi H. et al.; A review on Internet of Things (IoT), Internet of Everything (IoE) and Internet of Nano Things (IoNT). In: Internet Technologies and Applications (ITA); 2015; Wrexham: IEEE; S. 219

»Systems that proactively adapt to you are easy to think of but hard to produce. Two of the hardest problems are (1) computers lack common sense, and (2) they are not good learners.«¹⁰⁶

Künstliche Neuronale Netzwerke, die in smarten Assistenten eingesetzt werden, imitieren das Denken und Verhalten von spezifischen Nutzer*innen, indem sie diese durch einen Algorithmus nach-modellieren. Somit können durch die Technologie von KNNs Verhaltensweisen besser approximiert und interpretiert werden. Doch wie funktionieren Künstliche Neuronale Netzwerke? Im wesentlichen handelt es sich dabei um eine Abfolge vergleichsweise schlichter, mathematischer Funktionen, die gemeinsam einen Algorithmus ergeben. Wichtig ist jedoch, dass beinahe alles an diesem Algorithmus variabel ist. Das Ziel von KNNs ist es nämlich, Ergebnisse zu approximieren und verschiedenste Abweichungen einzuberechnen, ohne speziell dafür programmiert worden zu sein. Möglich ist dies aufgrund der Fähigkeit sich selbst anzugleichen; dieses Angleichen – als Deep Learning bezeichnet – funktioniert dabei anhand einer bestimmten Logik.

Raúl Rojas beschreibt diese Logik mithilfe folgender Konzeption: »Berechnung = Speicherung + Übertragung + Verarbeitung«¹⁰⁷ Das bedeutet, dass die Neuronen – die einzelnen Bauteile des Netzwerks – spezifische Aufgaben übernehmen. Zum einen werden die Daten, die an die Neuronen weitergeleitet werden (Inputs), verarbeitet und zu einem Ergebnis (Output) zusammengefasst. Dieser Prozess erlaubt es, Informationen zusammenzufassen, und in ein quantitatives Verhältnis zu anderen möglichen Inputs zu stellen.¹⁰⁸ Zum anderen können Neuronen dazu benutzt werden, Schwellenwerte zu berechnen. Dazu werden die Inputs (oder die berechneten Datensätze) auf verschiedene Weise gewichtet. So können Ja/Nein Ergebnisse (also Werte von 0 und 1) berechnet werden. Mithilfe dieser Schwellenwertberechnung können qualitative Verhältnisse und auch boolesche Operatoren hergestellt werden.¹⁰⁹

¹⁰⁶ Picard, Rosalind W.; *Affective Computing*; 1997; Massachusetts: MIT Press; S. 103

¹⁰⁷ Rojas, Raúl; *Künstliche Neuronale Netze als neues Paradigma der Informationsverarbeitung*. In: Pauen, Michael; Roth, Gerhard (Hg.); *Neurowissenschaften und Philosophie: eine Einführung*; 2001; Frankfurt am Main: UTB; S. 2

¹⁰⁸ Diese Art von Neuron wird als generisches Neuron bezeichnet. Ebd. S. 5 f.

¹⁰⁹ Diese Art von Neuron wird Perzeptron bezeichnet. Ebd. S. 6 f.

Ausgehend von dieser Fähigkeit zur quantitativen und qualitativen Einordnung, sowie zur Transformation von Datensätzen, werden schließlich Ergebnisse zu spezifischen Fragestellungen berechnet.

Entscheidend bei der Entwicklung des Netzwerks (mittels Deep Learning) ist, dass diese Fragestellungen bereits im Vorhinein manuell (also im Regelfall durch Menschen) kodiert werden müssen. Dadurch ergibt sich in der Berechnung eine messbare Differenz zwischen dem manuell kodierten, und dem vom KNN berechneten Ergebnis. Diese Differenz wird als Fehler bezeichnet. Nach der Berechnung des Fehlers, wird durch den Prozess der Gegenberechnung (also der umgekehrten Annäherung vom korrekten Ergebnis hin zum originalen Output) der Algorithmus (also die Neuronen in der Kette) transformiert. Diese Gegenberechnung, die als Back-Propagation¹¹⁰ bezeichnet wird, wiederholt sich solange, bis die Fehlerquote, einen, vorher festgelegten, akzeptablen Wert erreicht.

Nach dem positiven Abschluss dieser einen Berechnung (und der Back-Propagation) springt das KNN zur nächsten Berechnung, samt zugehörigem Ergebnis weiter. Dieser Prozess wird entweder für eine bestimmte Zeitdauer, oder Anzahl an Berechnungen wiederholt, im Extremfall jedoch so lange, »bis das System in der Lage ist, quasi autonom für unbekanntem Input (ein bis dato nicht gelerntes Beispiel eines Fuchses oder Hundes) den passenden Output (das richtige Label Fuchs/Hund) mit gewünschter Genauigkeit (das heißt mit einem ausreichenden Annäherungswert) vorherzusagen.«¹¹¹

Aus der Betrachtung dieses Lern- und Berechnungsprozesses ergeben sich schließlich mehrere wesentliche Eigenarten von KNNs: Erstens müssen im Lernprozess KNNs immer mit bereits kodierten Daten gefüttert werden. Zwar müssen jene nicht zwingend von Hand berechnet werden – um maschinelle Fehler auszuschließen, ist das jedoch meist der Fall.

¹¹⁰ Siehe dazu: Le Cun, Yann; A theoretical framework for back-propagation. In: Touretzky, David et. al. (Hg.); Proceedings of the 1988 Connectionist Models Summer School; 1988; San Mateo: Morgan Kaufmann; S. 21-28

¹¹¹ Sudmann, Andreas; Künstliche neuronale Netzwerke als Black Box. Verfahren der Explainable AI. Medienwissenschaftliche Perspektiven. In: Klimczak, Peter e. al. (Hg.); Maschinen der Kommunikation. Interdisziplinäre Perspektiven auf Technik und Gesellschaft im digitalen Zeitalter; 2020; Wiesbaden: Springer Vieweg S. 190 f.

Bei von Hand kodierten Daten existieren zwar genauso Fehler, die Idee dahinter ist es aber, dass die Daten im Idealfall von vielen verschiedenen Personen kodiert werden, und damit etwaige Fehler die Genauigkeit des KNNs nicht gefährden. Das wäre zumindest der Gedanke hinter der Verwendung von per Hand kodierter Daten. Trotzdem können sich durch den Bias¹¹² der kodierenden Menschen tiefgreifende Fehler einschleusen:

»For instance, in 2017, Amazon decided to abandon a recruiting recommendation tool after finding out that it discriminated against women [...] [because] developers had trained the system using data from more than 10 years of incoming past applications and hiring decisions, in which males dominated. This bias went unnoticed [...], because it did not derive and justify its recommendations based on reasoning, but applied a complex statistical model that developers and users could not interpret.«¹¹³

Im Laufe des Lernverhaltens des KNN werden die spezifischen Formeln, und die Funktion der Neuronen so oft transformiert, dass am Ende des Lernprozesses die Berechnungsweise des Netzwerks unklar geworden ist. Da sich das Netzwerk mit jeder Berechnung weiter angleicht, lassen sich die Ergebnisse auch nicht reproduzieren (und damit wissenschaftlich erfassen). Dieses Phänomen lässt sich als Blackbox bezeichnen.¹¹⁴

Eine weitere Eigenart von KNNs ist es, dass, sie weder neue Verknüpfungen bilden, noch neue Neuronen erschaffen können. Grundsätzlich ist in KNNs die Verknüpfung der Neuronen (als Kettenglieder) und die Anzahl der Neuronen (die Länge der Kette) durch die Programmierung determiniert. Das heißt: Ein Künstliches Neuronales Netzwerk, dessen Neuronenzahl sehr klein ist, kann nur ein bestimmtes Level an Differenzierung erreichen,

¹¹² Der Bias der in den manuell kodierten Daten vorkommt, kann dabei ähnlich wie der bereits erwähnte Bias von Programmierer*innen verstanden werden. Vor allem weil die jeweiligen manuell kodierten Daten als Programmierungsleitfaden für KNNs dienen. Siehe dazu: Friedman, Batya; Nissenbaum, Helen; Bias in Computer Systems. In: ACM Transactions on Information Systems, Volume 14, Issue 3; 1996; New York: Association for Computing Machinery; S. 330-347

¹¹³ Maedche, Alexander et. al.; AI-based Digital Assistants: Opportunities, Threats, and Research Perspectives. in: (Journal) Business and Information Systems Engineering, Issue 4; 2019; Wiesbaden: Springer Fachmedien; S. 10

¹¹⁴ Siehe dazu: Sudmann, Andreas; Künstliche neuronale Netzwerke als Black Box. Verfahren der Explainable AI. Medienwissenschaftliche Perspektiven. In: Klimczak, Peter e. al. (Hg.); Maschinen der Kommunikation. Interdisziplinäre Perspektiven auf Technik und Gesellschaft im digitalen Zeitalter; 2020; Wiesbaden: Springer Vieweg S. 192

dafür ist es aber auch schneller trainiert. Die Komplexität des Netzwerks, sowie Dauer und Umfang des Trainings, bestimmen damit die Effizienz der KNN. Zuletzt benötigt eine KNN auch eine Rahmensoftware, die mit dem Netzwerk selbst interagiert, Inputs und Outputs zuweist, und die Ergebnisse abbildet. Eine KNN ist damit keine eigene Entität, sondern eine Berechnungsmethode, die sich die biologische Struktur von Nervenzellen zum Vorbild nimmt, und als Möglichkeit zur Berechnung von Fragestellungen nützt. Wenn allerdings die Ergebnisfindung dieser Berechnungen nicht mehr nachvollzogen werden kann, und zusätzlich tiefgreifende Auswirkungen mit sich bringt, dann ist diese Dynamik problematisch.

Nachdem Künstliche Neuronale Netzwerke schließlich fertig trainiert worden sind, können jene (beziehungsweise Kopien davon) – je nach Trainingszweck – etwa zur Herstellung von Computermodellen über menschliches Verhalten, oder die Spracherkennung und -synthese verwendet werden. So werden im Regelfall nutzer*innenspezifische Profile angelegt, in denen Besonderheiten im Verhalten – wie etwa die Verwendung bestimmter Begriffe, Tagesabläufe oder Präferenzen – gespeichert sind. Für die Beziehung zwischen smarten Kommunikationsassistenten und ihren Nutzer*innen bedeuten die Besonderheiten von Künstlichen Neuronalen Netzwerken dabei folgendes: Erstens basieren die Ergebnisse von KNNs, wesentlich auf den spezifischen Algorithmen – und damit der Art und Weise wie diese Algorithmen trainiert werden. Somit ist die manuelle Kodierung (und der darin inkludierte Bias) entscheidend für die Berechnung. Zweitens ist die Berechnung mittels Künstlichen Neuronalen Netzwerken eine Blackbox, deren genauer Berechnungsweg nicht nachvollzogen werden kann, und die keine reproduzierbaren Ergebnisse liefert. Drittens sind KNNs dann einsatzbereit und verwendbar, wenn sie als genau genug beurteilt werden.¹¹⁵

¹¹⁵ Wenn die Genauigkeit einen gewissen Schwellenwert überschritten hat, und hält. Allerdings kann weder garantiert werden, dass es Ausreißer mehr gibt, noch dass etwaige Ausreißer keine schwerwiegenden Folgen haben könnten. Die einzige Lösung besteht darin die Bedeutung von Berechnungen durch KNNs einzuschränken.

Künstliche Neuronale Netzwerke in der Praxis

Schließlich werden Künstliche Neuronale Netzwerke, als technologische Berechnungsmethode, bereits aktiv für Affective Computing und besonders für die Spracherkennung und -synthese eingesetzt. So etwa bei den smarten Kommunikationsassistenten Apple Siri¹¹⁶ und Amazon Alexa.¹¹⁷ Amazon etwa, setzt seit 2020 Künstliche Neuronale Netzwerke ein, um die Spracherkennung (und -verarbeitung) von Alexa zu verbessern:

»Nadim Fresko, vice president of Alexa Devices and Developers, [...] expects that Alexa skills that use the new DNN-based technology will see an average increase in accuracy of 15%, and the company will continue to expand use of [Deep Neural Networks] to 400 eligible skills [...] by later this year. [...] DNN models are better quipped to recognize natural language. DNNs, combined with unsupervised pretraining, have led to state-of-the-art performance in a variety of natural-language-processing domains. [...] DNNs can generalize from ›buy me an apple‹ to ›order an orange for me‹, as apple/orange and buy/order have similar representations [within the neural networks] [...].«¹¹⁸

Künstliche Neuronale Netzwerke (beziehungsweise Deep Neural Networks) sind somit etwa dafür zuständig, Begriffen Zusatzinformationen zuzuweisen, und die jeweiligen Bedeutungen – dass etwa Äpfel und Orangen beide Teil der Kategorie Obst sind – durch Deep Learning laufend anzupassen. Doch auch hier können sich tiefgreifende Fehler oder Vorurteile einschleichen. Wenn nun beispielsweise Kaffeebohnen unter die Kategorie Obst fallen, dann ist der Fehler vergleichsweise harmlos. Wenn allerdings – wie bereits erörtert – die Hautfarbe die eine Überwachungssoftware¹¹⁹ erkennt, darüber entscheidet ob ein Person gerade zuhause ist, oder an einem fremden Ort einbricht, dann haben entweder Künstliche Neuronale Netzwerke (mit ihrer gegenwärtigen Ungenauigkeit) die Grenzen

¹¹⁶ Siehe dazu: (<https://machinelearning.apple.com/research/hey-siri> Abgerufen am 27. November 2022)

¹¹⁷ Beide Hersteller referenzieren dabei den Begriff der *Deep Neural Networks* (DNN). Dies beschreibt dieselbe Technologie und Ausführung wie KNNs und stellt eine Verschränkung der Begriffe Deep Learning und Künstliches Neuronales Netzwerk dar. Eine wesentliche Besonderheit ist jedoch, dass DNNs im Regelfall eine hohe Komplexität – und damit eine große Neuronenanzahl – besitzen.

¹¹⁸ Ananthakrishnan, Shankar; Amazon Scientists applying deep neural networks to custom skills; 2020; Amazon Science (<https://www.amazon.science/blog/amazon-scientists-applying-deep-neural-networks-to-custom-skills> Abgerufen am 26. November 2022)

¹¹⁹ Dies könnte etwa auch Heimüberwachungs-Software betreffen

ihrer Verwendbarkeit überschritten. Oder die Verwendung von KNNs müsste besser reglementiert, und somit überprüfbarer werden. Daher muss entweder die Handlungsfähigkeit von KNNs begrenzt und die Blackbox von KNNs abgebaut, oder die Systeme vertrauenswürdiger gemacht werden.

Vertrauen und Überprüfbarkeit

»Einerseits geht es darum, sicherzustellen, dass Vorhersagen des KNN-Systems ausreichend präzise sind. Andererseits besteht ein Interesse daran, nachzuvollziehen, warum ein System eine bestimmte Vorhersage trifft. Beide Aspekte sind [...] relevant, um KI-basierten Entscheidungen vertrauen zu können.«¹²⁰

Schließlich geht es nämlich nicht nur bei Künstlichen Neuronalen Netzwerken, oder smarten Kommunikationsassistenten, sondern bei jedweder Art von Technologie, die von den individuellen Nutzer*innen nicht mehr selbst überprüft werden kann, um Vertrauen. Vertrauen ist ebenso notwendig um etwa eine Waschmaschine – ob smart oder nicht – ohne Hintergedanken benutzen zu können:

»Eine sichere Waschmaschine ist [...] eine Waschmaschine, die man bei laufender Schleuder nicht öffnen kann, weil eine gerätetechnische Vorrichtung die Tür verriegelt hat. Ein Schild mit der Bitte, so etwas nicht zu tun, könnte den gleichen Effekt haben, ist aber sozial anspruchsvoller: Man muß erwarten dürfen, daß solche Bitten immer berücksichtigt werden, daß der Benutzer lesen kann, daß der Benutzer die Sprache des Herkunftslandes der Maschine versteht etc. Hier ist ein technischer Garant [...] vertrauensvoller als die Verlässlichkeit eines menschlichen Benutzers.«¹²¹

Aus dieser Erörterung von Gerald Wagner (1994) zu den sozialen Auswirkungen des Vertrauens in diverse Technologie, lassen sich dabei mehrere Aspekte ableiten:

¹²⁰ Sudmann, Andreas; Künstliche neuronale Netzwerke als Black Box. Verfahren der Explainable AI. Medienwissenschaftliche Perspektiven. In: Klimczak, Peter e. al. (Hg.); Maschinen der Kommunikation. Interdisziplinäre Perspektiven auf Technik und Gesellschaft im digitalen Zeitalter; 2020; Wiesbaden: Springer Vieweg S. 193

¹²¹ Wagner, Gerald; Vertrauen in Technik. In: Zeitschrift für Soziologie, Jg. 23, Heft 2, 1994; Stuttgart: Ferdinand Enke Verlag; S. 154

Erstens ist ein technisch garantierter Qualitätsstandard – sei es die Akkuranz von Künstlichen Neuronalen Netzwerken, oder die Verriegelung von Waschmaschinentüren – ein von vermeintlich objektiven Autoritäten gefälltes Urteil, dass der jeweiligen Technologie vertraut werden kann. Zweitens eliminieren technologische Reglementierungen daher die Notwendigkeit zur individuellen Beurteilung. Entweder weil, wie bei KNNs, ein individuelles Urteil gar nicht mehr möglich ist, oder damit weniger Vorwissen und Einarbeitungszeit für die Verwendung von Waschmaschinen benötigt wird. Somit werden zwar durch Nutzer*innen verursachte Fehler minimiert. Im Umkehrschluss werden aber auch Kompetenzen im Umgang mit Technologien durch Vertrauen¹²² ersetzt.

Für die Beziehung zwischen smarten Kommunikationsassistenten und ihren Nutzer*innen bedeutet dies, dass Vertrauen die entscheidende Währung zur Akzeptanz und Verwendung von smarten Assistenten ist. Ziel ist es eine Beziehung zu erzeugen, in der die Nutzer*innen den jeweiligen Technologien gesamtheitlich vertrauen. Denn nur dann können die Nutzer*innen, guten Gewissens, etwa die Kontrolle über den Schlaf an die Maschine abgeben. Konträr zum Uncanny-Valley-Effekt – und der Bedrohlichkeit von HAL 9000, geht es schließlich darum diesen Eindruck, eines vertrauenswürdigen Assistenten, zu erwecken. Dabei lassen sich zwei Zugänge aus den bisherigen Ausführungen zu smarten Kommunikationsassistenten ableiten: Die Assistenten werden entweder vertrauenswürdig, weil sie akkurat arbeiten, und etwa die rationalen und gegebenenfalls auch emotionalen Bedürfnisse ihrer Nutzer*innen besonders genau abbilden, modellieren und zufriedenstellen können; hier stellt sich die Frage wer dies überprüfen soll. Oder aber die Assistenten erzeugen Vertrauen durch die Illusion¹²³ von Kompetenz, ohne jedweden Einblick in die zugrundeliegenden Mechaniken. Hier bleibt letztlich nur die Aussage des Herstellers (und von etwaigen unabhängigen Beobachtern) um zu entscheiden, ob die jeweiligen Nutzer*innen dem smarten Assistenten Vertrauenswürdigkeit (oder Intelligenz) zusprechen. Gegenwärtige smarte Kommunikationsassistenten bewegen sich dabei zwischen diesen Positionen.

¹²² Sowohl Vertrauen in die Technologien, als auch in die Institutionen die die Technologien überprüfen. (In vielen Fällen findet dies mittlerweile innerhalb einer einzelnen Korporation statt).

¹²³ Gegebenenfalls sogar eine bewusste Täuschung.

Zwar können die von Smart Assistant-Software verwendeten Methoden und Berechnungsweisen theoretisch gefasst und ihre Effekte analysiert werden, der genau Einblick in die tatsächliche Programmierung, bleibt aber verborgen.

Server und Clients

Ähnliches gilt für das Zusammenspiel zwischen lokalen Clients¹²⁴ von smarten Kommunikationsassistenten und den Servern,¹²⁵ mit denen sie in permanentem Austausch stehen. So synchronisieren sich die Clients, sofern möglich, mit den zentralen Servern, und übermitteln die erfassten (und verarbeiteten) Daten, oder auch etwaige Anfragen nach Daten oder Netzwerkadressen.¹²⁶ Alok Sinha (1992) beschreibt dabei die Synchronisation zwischen Server und Clients wie folgt:

»A Server provides [...] an accounting data retrieval and processing service to the Client. A service provided by a Server may require minimal Server-based computation (e.g., print servers or file Servers) to intensive computations (e.g., database servers or Image Processing servers). [...] A Server [...] does not initiate a conversation with any Client. It merely acts either as a repository of data (e.g., file Server) or knowledge (e.g., database Server) or as a service provider (e.g., print Server).«¹²⁷

Als Erweiterung der Clients übernehmen Server somit diverse Aufgaben. Um das Zusammenspiel dabei so effizient wie möglich zu gestalten, werden Berechnung und Datenspeicherung aufgeteilt. Die Spracherkennung und -verarbeitung von Amazon's Alexa versucht etwa den lokalen Speicher so wenig wie möglich zu belasten:

¹²⁴ Wird auch als Front End bezeichnet. Siehe dazu: Modi, Kirit J. et al.; Cloud computing - concepts, architecture and challenges. In: International Conference on Computing, Electronics and Electrical Technologies (ICCEET), 2012; Nagercoil: IEEE; S. 878

¹²⁵ Wird auch als Back End bezeichnet. Ebd. S. 878

¹²⁶ die automatisierbaren IoT-Geräte im lokalen Netzwerk werden im Regelfall ebenso über die jeweiligen Netzwerkadressen angesteuert.

¹²⁷ Sinha, Alok; Client-Server Computing. In: Communications of the ACM, Volume 35, Issue 7, 1992; New York: Association for Computing Machinery; S. 79

»Storing word embeddings, however, is extremely memory intensive. To make DNNs practical for more than 100,000 skills, the Amazon scientists combine a large store of shared, generic embeddings with small, local stores of skill-specific embeddings.«¹²⁸

So wird etwa ein reduzierter (Standard-)Wortschatz heruntergeladen, damit die Spracherkennung und -ausgabe weiterhin offline verfügbar bleibt. Damit es jedoch nicht (beziehungsweise kaum) auffällt, dass Amazon's Alexa ohne Internetzugriff nur in abgespeckter Form verfügbar ist, werden etwa spezifische Zusatz-Wortschätze angelegt. Denn das Ziel ist es, dass, zusätzlich zu standardisierten Aufgaben, von den jeweiligen Nutzer*innen verwendete, komplexe Begriffe oder Besonderheiten in der Aussprache, weiterhin erkannt werden können. Und auch Berechnungen werden zwischen Server und Client aufgeteilt. Wenn die Rechenkapazität von Servern maßgeblich für das Funktionieren der lokalen Clients verantwortlich ist – etwa weil Teile der Berechnung in Echtzeit durchgeführt werden müssen – dann kann von Cloud Computing gesprochen werden.¹²⁹

Zusätzlich zum Zusammenspiel zwischen Server und Client, hat bei smarten Kommunikationsassistenten auch die Kommunikation mit IoT-Geräten eine entscheidende Rolle. Denn wie bei smarten Assistenten Teile der Datensätze und/oder Berechnungen ausgelagert werden, kommuniziert auch ein Großteil an IoT-Technologien mit der Cloud.¹³⁰ Folglich gilt es mehrere Modelle zur Kommunikation zwischen smarten Assistenten und IoT-Peripheriegeräten zu unterscheiden. »Device-to-Device Communication«¹³¹ ist dabei der konzeptionelle Ausgangspunkt, bei der die jeweiligen IoT-Geräte ihre Daten direkt an die Zentralstelle – also etwa an das Framework von smarten Kommunikationsassistenten – übermittelt werden.

¹²⁸ Ananthakrishnan, Shankar; Amazon Scientists applying deep neural networks to custom skills; 2020; Amazon Science (<https://www.amazon.science/blog/amazon-scientists-applying-deep-neural-networks-to-custom-skills> Abgerufen am 26. November 2022)

¹²⁹ Siehe dazu etwa: Modi, Kirit J. et al.; Cloud computing - concepts, architecture and challenges. In: International Conference on Computing, Electronics and Electrical Technologies (ICCEET), 2012; Nagercoil: IEEE; S. 877

¹³⁰ Das heißt: Eine Form von zentralen Servern, die (auch) die jeweiligen IoT-Peripheriegeräte unterstützen.

¹³¹ Singhania, Vivek; The Internet of Things: An Overview. Understanding the Issues and Challenges of a More Connected World; 2015; Genf, Reston: The Internet Society (ISOC); S. 13

»The device-to-device communication model represents two or more devices that directly connect and communicate between one another, rather than through an intermediary application server. These devices communicate over many types of networks, including IP networks or the Internet.«¹³²

Darüber hinaus besteht die Möglichkeit für jedwede Kommunikation zwischen IoT-Gerät und Zentralstelle, einen Server als Vermittlungsinstanz zwischenzuschalten. Dies wird als »Device-to-Cloud Communication«¹³³ bezeichnet. Bei dieser Anwendung profitiert die Kommunikation zwar von der (durch die Cloud) verfügbaren Rechenleistung und möglichen Daten-Zwischenspeicherung, allerdings steht und fällt die Kommunikation mit der intakten Internetverbindung. Die am weitesten verbreitete Kommunikationsmethode ist jedoch das »Device-to-Gateway Model«.¹³⁴ Hier kommunizieren Server und IoT-Gerät über eine lokale Software, die etwa in der Zentralstelle – also auf dem Smartphone oder Smart Hub – installiert ist. Die Kommunikation über eine Gateway-Software funktioniert dabei folgendermaßen: Die Daten die vom IoT-Gerät erhoben (und gegebenenfalls vorausgewertet) wurden, werden zuerst an die Gateway-Software übermittelt. Diese kommuniziert, sofern möglich, mit den zugehörigen Servern, kümmert sich um die Datenspeicherung und koordiniert wann, wo und wie Berechnungen durch die Cloud übernommen werden.¹³⁵ Zusätzlich dazu sind Gateways, nach der Abstimmung mit den zentralen Servern, für das Einspeisen der jeweiligen Daten in das Framework von smarten Kommunikationsassistenten zuständig. Dieser Zugang, eine lokale Gateway-Software als Schnittstelle zwischen dem IoT-Gerät, den zentralen Servern und dem Framework von etwa smarten Assistenten zu nutzen, findet dabei besonders in Smartphones und bei Wearables seine Anwendung:

»This is often the model employed with popular consumer items like personal fitness trackers. These devices do not have the native ability to connect directly to a cloud service, so they

¹³² Ebd. S. 13

¹³³ Ebd. S. 14

¹³⁴ Ebd. S. 15

¹³⁵ Falls keine Verbindung zu den Servern aufgebaut werden kann, ist die Gateway-Software schließlich auch dafür zuständig, die Daten lokal zwischenspeichern.

frequently rely on smartphone app software to serve as an intermediary gateway to connect the fitness device to the cloud.«¹³⁶

Hierin liegt auch der wesentliche Nutzen von Begleitsoftware für diverse IoT-Geräte (etwa Wearables). So muss aus technischer Perspektive auch nur die Gateway-Software Datensätze in das jeweilige Framework importieren können. Die Kommunikation zwischen Begleitsoftware und IoT-Gerät muss, abgesehen vom jeweiligen Netzwerkprotokoll, nicht standardisiert werden. Sowohl die Kommunikation zwischen IoT-Geräten und Zentralstelle, als auch die Kommunikation von beiden Technologien mit zentralen Servern, findet dabei im Rahmen von digitalem Schlaf-Tracking (Schlafüberwachung) ihre Anwendung.

Assistierter Schlaf

Um nun – im Rahmen der Frage wie sich die Beziehung zwischen smarten Kommunikationsassistenten und ihren Nutzer*innen gestaltet – assistierten Schlaf als Beispiel heranziehen zu können, müssen zuerst jene Besonderheiten rekapituliert werden, die die Beziehung und damit auch die Schlaf-Assistenz beeinflussen. Zum ersten implementieren smarte Assistenten Methoden des Affective Computing und erfassen damit so viele (und komplexe) Daten wie möglich. Zum zweiten interpretieren sie diese erhobenen Daten (und gemessenen Emotionen), sei es durch lokal gespeicherte Computermodelle, zentrales Cloud Computing oder Künstliche Neuronale Netzwerke. Drittens ist die Art der Datenerhebung (und Vor-Auswertung) entscheidend, selbst wenn ein Einblick aufgrund von proprietärer Software nur schwer möglich ist. Zum vierten ist schließlich das Vertrauen in die Genauigkeit der Assistenten und das Urteil der Intelligenz entscheidend, um der Ergebnisfindung (aufgrund mangelnder Überprüfbarkeit) zu glauben – und es bestimmt, ob die jeweiligen Assistenten überhaupt benutzt werden.¹³⁷ In der Verwendung von Sleeptracking-Software (und zugehörigen IoT-Geräten) geht es aber

¹³⁶ Singhanian, Vivek; *The Internet of Things: An Overview. Understanding the Issues and Challenges of a More Connected World*; 2015; Genf, Reston: The Internet Society (ISOC); S. 15 f.

¹³⁷ Wenn etwa Sleeptracking-Software als nicht vertrauenswürdig gelesen wird, dann wird sie, ähnlich dem Uncanny-Valley-Effekt, auch nicht verwendet werden.

ebenso um die Funktionalität, also etwa gesundheitliche und/oder identitätsstiftende Effekte.

Ein Beispiel ist die Software *Sleepwatch*TM by *Bodymatter*, ein verfügbares Programm¹³⁸ für Apple's (Smartphone-)Betriebssystem IOS. Dabei sammelt IOS zuerst alle verfügbaren Daten und importiert sie in das zugehörige Gesundheitsdaten-Framework Health. Über das Framework können ebenso Daten von diversen IoT-Geräten, wie etwa Smartwatches, verwendet werden. *Sleepwatch* übernimmt schließlich diese Daten ohne zusätzliche Prüfung, und wertet sie in der lokalen App aus.

Dabei erlaubt die Sleeptracking-Software folgende Funktionen: Sie bereitet die durchschnittliche Schlafdauer statistisch auf und weist jedem Schlaf einen Prozentsatz zu, der sich aus diversen Faktoren, wie Schlafdauer, Schlafrythmus, und etwaigen Schlafunterbrechungen zusammensetzt. Die Schlafdauer lässt sich vergleichsweise einfach erheben: *Sleepwatch* erfordert es, das Tracking jeden Abend manuell zu aktivieren, während die Aufzeichnung, durch die Smartphone-Weckfunktion am nächsten Morgen, automatisch gestoppt wird. Schlafunterbrechungen werden durch das Tool des »Snore Tracking«¹³⁹ mitgemessen. So zeichnet jenes Laute (ab einer Schwellenlautstärke) auf und hält damit etwaiges Schnarchen fest. – Im Nachhinein können dann alle angeschlagenen Aufzeichnungen angehört werden.

Die Software ermittelt über diese Aufzeichnungen aber auch Bewegungsgeräusche, die als Schlafunterbrechungen dargestellt werden. Die Ermittlung des Schlafrythmus ist schließlich nur über die Nutzung eines Wearables, während dem Schlaf, möglich. Denn *Sleepwatch* ermittelt den Schlafrythmus durch Veränderungen in der Herzfrequenz. All diese Möglichkeiten erinnern dabei an eine Form der medizinischen Schlafvermessung. Doch die Nutzungsbedingungen zeigen, dass die Software keine medizinisch verwertbaren Informationen produziert.

¹³⁸ Programme für Mobilgeräte werden im Regelfall als Apps (Kurzform für Applications) bezeichnet.

¹³⁹ Introducing Snore Tracking and Sleep Sound Recording for SleepWatch (<https://www.sleepwatchapp.com/blog/introducing-snore-tracking-and-sleep-sound-recording-sleepwatch/> Abgerufen am 24. August 2022)

»The Services and [...] Content are not intended to diagnose, treat, cure or prevent any disease. If you have a medical, sleep or heart condition, consult your doctor before using the services [...]. We are not responsible for any health problems that may result from the lifestyle choices, exercise activity, consultations, products, or events you learn about through the Services.«¹⁴⁰

Diese rechtliche Absicherung impliziert dabei zwei Besonderheiten bei der Nutzung von *Sleepwatch*: Einerseits ist die App in keiner Weise verpflichtet die Informationen zugunsten der Nutzer*innen darzustellen,¹⁴¹ oder die Genauigkeit zu überprüfen. Und andererseits müssen, da sich der Hersteller der medizinischen Überprüfung entzieht, die Visualisierungen, statistischen Auswertungen und Schlafempfehlungen weder gesundheitlich relevant, noch hilfreich sein. Die Funktion dieser Sleeptracking-Software ist es folglich identitätsstiftend zu sein, und Nutzer*innen, im Sinne eines »Quantified Self«,¹⁴² bei der eigenen Maßregelung und vermeintlichen Selbstoptimierung zu unterstützen. Um diese Funktion erfüllen zu können, ist es jedoch notwendig, der Software (und der zugrundeliegenden Datenerhebung und -aufbereitung) zu vertrauen.

Ein zweites Beispiel für assistierten Schlaf findet sich in der Einleitung zu *Unterwachen und Schlafen: Anthropophile Medien nach dem Interface* (2018). Dort beschreiben Michael Andreas, Dawid Kasprowicz und Stefan Rieger das Phänomen, dass Technologien zur Schlafvermessung nicht mehr bewusst oder aktiv genutzt werden, sondern unauffällig und subtil mit ihren Nutzer*innen interagieren und damit die »Praxis des Schlafens auf das Feld medialer Operationen« heben.¹⁴³

¹⁴⁰ SleepWatch Terms of Service (<https://www.sleepwatchapp.com/terms-of-service/in-app/>)
Abgerufen am 20. August 2022)

¹⁴¹ Ein mögliches Beispiel wäre hier ein gezielter Placebo-Effekt, indem die Software bei wiederholter Nutzung stetig die Gesundheitsdaten verbessert, egal ob dies die Realität widerspiegelt oder nicht. – Denn bis zu einem gewissen Punkt muss davon ausgegangen werden, dass der Erhalt der (zahlenden) Nutzer*innen wichtiger ist, als etwa medizinische Genauigkeit.

¹⁴² Quantified Self bezeichnet die Aufzeichnung des eigenen Verhaltens. Dies bedeutet im Regelfall, dass etwa Sport, Schlaf oder Ernährung, detailliert aufgezeichnet und optimiert werden. Siehe dazu: Selke, Stefan (Hg.); *Lifelogging. Digital self-tracking and Lifelogging – between disruptive technology and cultural transformation*; 2016; Wiesbaden: Springer VS; S. 1-3

¹⁴³ Andreas, Michael et. al. (Hg.); *Unterwachen und Schlafen: Anthropophile Medien nach dem Interface*; 2018; Lüneburg: meson press; S. 11

Das bedeutet, dass Andreas et. al. das Zusammenspiel von Sleeptracking-Software und Nutzer*innen nicht nur als schiere Datenerfassung (und nachfolgende Selbstbetrachtung), sondern auch als Wechselbeziehung erkennen. Dafür ist es jedoch entscheidend, dass die Sleeptracking-Software über (etwa auditive Signale) die passive, schlafende Nutzer*in delegieren kann:

»So bietet unter dem Namen ›Sleep Shephard‹ eine [...] Firma ein Stirnband an, das sowohl die Hirnaktivität im Schlaf ermitteln als auch den optimalen Schlaf ermöglichen soll. Gemessen werde per integriertem EEG, dessen Verlaufskurven später auf dem Handy eingesehen werden können, während das biokybernetische Feedback durch ein sogenanntes ›Binaural Beat System‹ gesteuert wird. [...] Das Besondere – und damit von der Hörphysiologie über die Schlafforschung bis in den ›Sleep Shephard‹ reichende Versprechen – ist die direkte Adressierung des Gehirns jenseits der bloß auditiven Bespielung beider Hörkanäle. Takt um Takt fahren die leicht versetzten Frequenzen herunter, bis sich die – durch ein Feedback der EEG-Werte ermittelte – [vermeintlich ideale] Schlaffrequenz der Person ergeben hat.«¹⁴⁴

Hier besitzt die Software die Handlungsfähigkeit, das Schlafverhalten zu regulieren, und damit direkt auf die Nutzer*innen einzuwirken. *Sleep Shephard* unterscheidet sich dabei wesentlich von *Sleepwatch*. Denn wo *Sleepwatch*, eher die Schlafüberwachung – und spätere Auswertung – zum Ziel hat, dort wirkt *Sleep Shephard* regulativ; und besitzt dabei sowohl ein eigenes, proprietäres Framework, als auch medizinische Zertifizierung. Letztlich bedeutet das aber nur, dass mehr Vertrauen (etwa in die Zertifizierung und die stärker verschlüsselte Technologie) notwendig ist, und sich die Effekte, sowie etwaige Fehler, gravierender auf die Nutzer*innen auswirken können. *Sleepwatch* lässt sich damit als die harmlosere, aber auch weniger nützlichere Variante von *Sleep Shephard* begreifen.

Ein drittes Beispiel für assistierten Schlaf ist schließlich die Software *Sleeptracker-AI*[®], gemeinsam mit den zugehörigen proprietären IoT-Geräten. Hier handelt es sich um ein medizinisches Produkt, das mithilfe eines Sleeptracking-Systems¹⁴⁵ Bewegungen (und Geräusche) während dem Schlaf aufzeichnet, und aus diesen Datensätzen die Schlafqualität ermittelt.

¹⁴⁴ Ebd. S. 11

¹⁴⁵ Sleeptracker-AI nutzt dafür ein »Sensor System«, das aus zwei Sensoren besteht, die auf/in der Matratze montiert werden. (<https://www.sleeptracker.com/support/tempur> Abgerufen am 13. Dezember 2022)

Dabei gibt es jedoch mehrere Kontexte wie *Sleeptracker-AI* genutzt werden kann. So besteht die Möglichkeit, die Sensoren (und das zugehörige IoT-Peripheriegerät) vom Hersteller installieren zu lassen, und etwa in kleineren medizinischen Studien zu Schlafverhalten und -qualität zu verwenden. Um dies zu ermöglichen, kann die Software gezielt eingestellt,¹⁴⁶ und etwa nur als Tool zur Aufzeichnung genutzt werden; um beispielsweise den Schlaf der Studienteilnehmer*innen nicht zu beeinflussen.

Außerdem können die erhobenen Daten über das Gateway – eine Begleitsoftware die auf den Smartphones der Nutzer*innen installiert werden muss – in Echtzeit an etwaige Studienleiter*innen übermittelt werden. Die Software (und IoT-Hardware) kann aber auch individuell genutzt werden. Erstens können die Nutzer*innen die jeweiligen Sensoren auch selbst anbringen, jedoch verfliegt dann die medizinische Zertifizierung aufgrund mangelnder Überprüfbarkeit durch den Hersteller. Zweitens kann die Software (unter verlässlichen Bedingungen), neben der passiven Aufzeichnung, auch mittels zwei wesentlichen Steuerungsmechanismen auf den Schlaf der Nutzer*innen einwirken: Entweder durch gezielte Vibrationen, die über das IoT-Gerät abgegeben werden, oder durch die Anpassung der Schlafdauer anhand von dynamischer Weckanpassung.¹⁴⁷ Einerseits sollen die, vom IoT-Peripheriegerät ausgegebenen Vibrationsmuster, von den schlafenden Nutzer*innen durch die Matratze hindurch empfunden werden. Sinn dahinter ist es, so der Hersteller, dass damit Schnarchen und Schlafapnoe minimiert werden können.

Andererseits besitzt die Begleitsoftware von *Sleeptracker-AI* die Möglichkeit, den Aufwachzeitpunkt innerhalb von einem, vorher festgelegten Zeitraum anzupassen,¹⁴⁸ um ein sanfteres Aufwachen zu ermöglichen. Außerdem besteht die Option nicht nur durch einen Klingelton, sondern ebenso durch Vibrationsmuster geweckt zu werden. Die an der Matratze befestigten Sensoren müssen folglich minimalste Bewegungen und Geräusche aufzeichnen, um die Atemfrequenz, Herzfrequenz und schließlich Schlafphasen ermitteln

¹⁴⁶ Eine tiefgreifende Anpassung der lokalen Software-Clients ist dabei nur durch den Hersteller möglich.

¹⁴⁷ Diese Handlungsfähigkeit ist dabei die einzige Möglichkeit, wie die Software direkt auf die Nutzer*innen einwirken können, ohne dass dies von den den Nutzer*innen gestoppt werden könnte. Siehe dazu: (<https://www.sleeptracker.com/faq/tempur> Abgerufen am 13. Dezember 2022)

¹⁴⁸ Dieser Zeitraum wird als »Sleep Cycle Alarm window« bezeichnet. (<https://www.sleeptracker.com/faq/tempur> Abgerufen am 13. Dezember 2022)

zu können. Im Rahmen einer kontrollierten Studie (die auch im eigenen Zuhause möglich ist) und einer Vorinstallation durch Hersteller (oder Fachpersonal) sind medizinisch nutzbare Ergebnisse auch vorstellbar. Doch sobald die Installation und Kalibrierung den Nutzer*innen selbst überlassen wird, ist eine medizinische Verifikation einfach nicht mehr möglich. Ab diesem Punkt wird *Sleeptracker-AI* zum Lifestyle Produkt.

Ähnliche Tendenzen finden sich dabei auch bei *Sleepwatch* und *Sleep Shephard*. So lassen sich gegenwärtige Technologien zur Schlafassistenz anhand von drei Achsen begreifen: Erstens ist es entscheidend wie, und wie detailliert, die jeweiligen technologischen Produkte Daten erheben – gepaart mit der Frage ob (und wie) sich diese Datenerhebung (etwa unter der Zuhilfenahme von IoT-Peripheriegeräten) verifizieren lässt. Zweitens geht es um die Frage ob die technologischen Produkte schlicht passiv, als umfangreiche Aufnahmegeräte mitlaufen, oder ob sie die Möglichkeit besitzen aktiv in den Schlaf (und damit den geistigen und körperlichen Zustand) ihrer Nutzer*innen einzugreifen.¹⁴⁹ Darüber hinaus ist es entscheidend, die Handlungsfähigkeit, also die möglichen Folgen von etwa fehlerhafter Software/Hardware zu thematisieren. Drittens ist nicht nur die Funktion, sondern ebenso die Erwartungshaltung (und der Vertrauensvorschuss, der damit einher geht) entscheidend, da besonders beim Umbruch zum Lifestyle Produkt die Kund*innenzufriedenheit wichtiger wird, als akkurate Ergebnisse.

Im Wesentlichen zeigen sich bei gegenwärtigen Technologien zur Schlafassistenz damit Ähnlichkeiten, aber auch Unterschiede, zu smarten Kommunikationsassistenten. Erwartbarerweise modellieren die jeweiligen technologischen Produkte etwa den Zustand der Nutzer*innen durch die erhobenen Daten nach, setzen allerdings (noch) eine klare Grenze: Denn gegenwärtige Ansätze zu assistiertem Schlaf fokussieren sich weitgehend auf physiologisch messbare Aspekte, wie etwa die Atmung,¹⁵⁰ Herzfrequenz und Bewegung während dem Schlaf. Die Berechnung von Tiefschlafphasen, und möglichen Aufwachzeitpunkten lässt sich dann relativ verlässlich ableiten, sofern die Datensätze akkurat sind.

¹⁴⁹ Und damit während dem Schlaf der Nutzer*innen die Kontrolle übernehmen.

¹⁵⁰ So auch Schnarchen und etwaige Schlafapnoe.

Selbst obskurere Methoden, wie die Messung der Hirnaktivität samt »biokybernetischem Feedback«,¹⁵¹ fokussieren sich auf genaues Messen, und etwaiges Einwirken auf die Körper der Nutzer*innen. Hier geht es, ganz im Gegensatz zu smarten Kommunikationsassistenten, um rein physiologische Symptome.¹⁵² Trotzdem ist dieser Messung ein Verständnis über gesunden und ungesunden Schlaf eingeschrieben. Selbst wenn durch physiologische Symptome die jeweiligen Schlafphasen vermeintlich fehlerfrei identifiziert werden können, liegt der Software ein von Menschen gemachtes Urteil¹⁵³ über idealen Schlaf zugrunde. – Indem zwischen Norm, und Abweichungen unterschieden wird.

Schließlich geht es jedoch ebenso um Vertrauen. Denn sobald die Anwendung nicht mehr durch den Hersteller oder etwaiges Fachpersonal überprüft wird, ist die medizinische Zertifizierung – mangels Überprüfbarkeit durch die individuellen Nutzer*innen – in erster Linie ein Marketinggag. Dieselbe Problematik stellt sich dabei auch bei Lifestyle Produkten, denn auch hier kann nicht überprüft werden, wie akkurat die Datensätze sind, und ob diese gegebenenfalls sogar verfremdet werden. Das hier notwendige Vertrauen findet nur selten bedingungslos statt, sondern wird – bei in der Nutzung von digitalen Technologien – im Regelfall unter der Prämisse von zusätzlicher Funktionalität gewährt, und ist immer an eine bestimmte Form von Verantwortlichkeit geknüpft. Dieses Zusammenspiel von Vertrauen (angesichts mangelnder Überprüfbarkeit) unter der Prämisse von zusätzlicher Funktionalität und gepaart mit der Handlungsfähigkeit sowie der Verantwortlichkeit bei eventuellen Fehlern charakterisiert schließlich die Beziehung zwischen smarten Kommunikationsassistenten und ihren Nutzer*innen.

¹⁵¹ Andreas, Michael et. al. (Hg.); Unterwachen und Schlafen: Anthropophile Medien nach dem Interface; 2018; Lüneburg: meson press; S. 11

¹⁵² Smarte Assistenten messen hingegen nicht nur physiologische Symptome, sondern versuchen schließlich auch mittels Computermodellen die Motivationen und Gedankengänge ihrer Nutzer*innen zu berechnen – um etwa unter dem Aufhänger einer besseren Interaktion angemessener reagieren zu können.

¹⁵³ Schlussendlich lässt sich ebenso festhalten, dass jedwede (auch medizinische) Ansicht über vermeintlich idealen Schlaf, von Menschen über Indizien, etwa die Konzentrationsfähigkeit oder die körperliche Ausdauer nach dem Aufwachen, beurteilt wird.

Dieses Zusammenspiel soll nun an Markus Krajewskis (2010) Konzept der digitalen Dienerschaft und an Isaac Asimov's Konzept des Frankenstein Complex¹⁵⁴ erörtert werden.

Digitale Dienerschaft

Um digitale Dienerschaft im Kontext von smarten Kommunikationsassistenten und ihren Nutzer*innen darstellen zu können, müssen zuerst die Überschneidungen zwischen den Anwendungsbereichen von smarten Assistenten und Markus Krajewskis Ausführungen zu digitalen Dienern dargestellt werden. Krajewski unterscheidet dabei zwischen diversen Ausprägungen von Dienerschaft und zeichnet die historische Entwicklung bis in die digitale Sphäre nach:

»Wenn sich im 19. Jahrhundert mit dem Übergang von [menschlichen Dienern] auf ›stumme Diener‹ und handelnde Dinge eine Transformation der Handlungsmacht von Subjekten auf Objekte vollzieht [...], so ereignet sich an dieser Stelle, im Rahmen der *Client-Server*-Architektur und ihrer Entwicklung [vermeintlich] intelligenter nicht-menschlicher Wesen im Virtuellen nichts weniger als eine weitere Faltung der Subjekt-Objekt-Differenz. Und zwar kommt den einstmals unbeseelten Dingen [...] (erneut) ein spezifischer Subjektcharakter zu.«¹⁵⁵

Das Zusammenspiel zwischen Herrscher und Diener ist somit wesentlich an die Bedingung geknüpft, dass Diener für die Herrscher primär zweckgebunden sind (und tendenziell immer auf die Objektebene reduziert werden) – egal ob menschliche Diener, Telefonapparate,¹⁵⁶ oder smarte Assistenten. Darüber hinaus stellt sich die Frage nach dem Subjektcharakter.

¹⁵⁴ Siehe dazu: Asimov, Isaac; *The Machine and the Robot*. In: Warrick, Patricia et. al.; *Science Fiction: Contemporary Mythology*; 1978; New York: Harper and Row; S. 252

¹⁵⁵ Krajewski, Markus; *Der Diener. Mediengeschichte einer Figur zwischen König und Klient*, 2010, Frankfurt am Main: Fischer: S. 536

¹⁵⁶ Krajewski beschreibt, wie die Verlagerung von menschlicher Dienerschaft auf technologische Objekte einen neuen Parameter zutage treten ließ: Den Aspekt des »Self service«, also der Grad bis zu dem die Herrscher selbst den Dienst verrichten (also die Maschine bedienen) müssen. So dient das Telefon der Kommunikation über weite Distanzen, aber nur wenn beide Nutzer*innen aktiv das Telefon benutzen und instand halten. Siehe dazu: Ebd. S. 496

Eine Möglichkeit wäre es dabei, die vermeintliche (erneute) Subjekthaftigkeit als Illusion von anthropomorphem Auftreten, gepaart mit dem Anschein von intelligentem Verhalten zu erklären. Doch der Begriff beschreibt mehr als nur die Imitation von menschlichen Dienern. Denn für Krajewski ist die Rolle von Subjekt oder Objekt auch immer mit einer Form von Hierarchie (und Gefälle) verbunden. So rekurriert er etwa auf Hegels Argumentation über Herrscher und Diener,¹⁵⁷ nach der die Dienerschaft eine durchaus günstige Position sein kann, die wider Erwarten auch mit Macht verbunden ist. Ausgehend davon beschreibt Krajewski dann etwa Clients als »Quasi-Objekte«¹⁵⁸, da sie zwar als Objekte auftreten, aber trotzdem Einfluss auf ihre Nutzer*innen, und das Netzwerk ausüben können.

»Dieses Quasi-Objekt vertritt einerseits die Interessen seines Herrn, andererseits verfügt es jedoch mit eigener Souveränität über die Kommunikationskanäle, indem es je nach Bedarf die im Netzwerk [...] bereitstehenden *Server* kontaktiert, koordiniert und kontrolliert, [...].«¹⁵⁹

Wenn gegenwärtige smarte Kommunikationsassistenten also Subjektcharakter besitzen, bedeutet das, dass jene Einfluss auf die Nutzer*innen (also auf ihre vermeintlichen Herrscher*innen) ausüben können. Doch dieser Einfluss ist – zumindest gegenwärtig – stark begrenzt. Denn die herausragende Funktion, die smarte Kommunikationsassistenten erfüllen, ist die eines (mehr oder weniger emotional intelligenten) Interfaces. Außerdem ist die Verwendung von smarten Assistenten keineswegs verpflichtend. Jedwedes Scheitern, oder fehlerhafte Funktionen können nämlich immer durch zusätzliche Interfaces, sei es ein Touchscreen, Maus oder Tastatur, behoben werden. Selbiges gilt auch für die Verantwortlichkeit. Umfangreiche und undurchsichtige Geschäftsbedingungen¹⁶⁰ skizzieren eine beschränkte Haftung, bei etwaigen Fehlern der Technologie.

¹⁵⁷ Hier stellt Krajewski auch die Frage, wer nun eigentlich wen beherrscht. Eine Frage, die sich durch jedweden Diskurs über Herrscher und Diener zieht. Ebd. S. 80

¹⁵⁸ Ebd. S. 542

¹⁵⁹ Ebd. S. 542

¹⁶⁰ Wie etwa: SleepWatch Terms of Service (<https://www.sleepwatchapp.com/terms-of-service/in-app/> Abgerufen am 20. August 2022)

Diese Undurchsichtigkeit ist insofern kaum problematisch, als ein Großteil der Fehler manuell behoben werden kann, und gegenwärtige smarte Assistenten keine tiefgreifende Einflussnahme erlauben. Folglich lassen sich smarte Kommunikationsassistenten in ihrer gegenwärtigen Ausprägung als Lifestyle Produkte fassen, die sich angesichts der begrenzten Möglichkeiten zwischen anthropomorpher Liebäugelei und der Illusion von Intelligenz befinden.

Dem gegenüber steht das konträre Bild des unheimlichen HAL 9000, der über das Wohlbefinden einer gesamten Astronautencrew entscheiden kann, und in einer Welt existiert, in der eine Maschine, ohne große Probleme das Leben der schlafenden Astronauten beenden kann. Dieser Weltraumschlaf – gesteuert durch eine allmächtige Maschine, die vor niemandem für ihre Entscheidungen Rechenschaft ablegen muss – ist dabei eine Extremform der Handlungsfähigkeit. So ist zwar hohe Funktionalität gegeben, doch dies inkludiert auch die Möglichkeit zum Machtmissbrauch. Demnach ist HAL 9000 (auch hier) ein Negativbeispiel für zu viel Handlungsfähigkeit bei Maschinen:

»[A] key technology, that never came to pass the way 2001 envisioned was the science of cryogenics, or artificial hibernation. Back in the 1960s it was assumed that the then-burgeoning study of cryogenics would continue to develop and become mainstream. [...] The idea of putting people into a state of suspended animation, where bodily functions were reduced [...], has long been an old chestnut in science fiction. [...] Though there was quite a bit of research done over the years, it was ultimately curtailed for moral as much as for safety reasons;«¹⁶¹

Während sich die Sicherheitsbedenken darauf beziehen, wie erheblich eine fehlerhafte Maschine das Überleben der Nutzer*innen beeinflussen könnte, beziehen sich die moralischen Bedenken vor allem auf die Frage von Verantwortung: Denn wer soll im Schadensfall haften? Der Hersteller oder das Fachpersonal, das für die Wartung zuständig ist? Oder sind die Nutzer*innen gar selbst verantwortlich weil sie sich (mehr oder weniger) wissentlich in Gefahr begeben? – Dies ist ein wesentliches Dilemma, das vor allem bei autonom agierenden Technologien (also Maschinen ohne menschliche Überwachung) auftritt.

¹⁶¹ Frinzi, Joe R.; Kubrick's Monolith. The Art and Mystery of 2001: A Space Odyssey; 2017; Jefferson: McFarland & Company; S. 59

Schließlich kann festgehalten werden, dass Funktionalität und Gefährlichkeit immer in direkter Verbindung zueinander stehen. Je mehr (und stärker) die Maschine Einfluss nehmen kann, um so weniger self service ist notwendig. Im Umkehrschluss können aber auch fatalere Fehler auftreten. Digitale Dienerschaft ist dabei ein Aushandlungsprozess, der sich zwischen belanglosen Tools und totaler Fremdkontrolle bewegt, wobei gegenwärtige smarte Kommunikationsassistenten vor allem als Lifestyle Produkte verstanden werden können. Letzten Endes spielt dabei auch das Vertrauen eine entscheidende Rolle: Denn egal wer haftet, und wie stark die Einflussnahme und fatal etwaige Fehler sein können, ohne Vertrauen in die Maschine – vor allem angesichts mangelnder Überprüfbarkeit – wird sie nicht verwendet werden. Somit soll jedwedes Misstrauen¹⁶² vermieden werden. Nirgendwo wird dies so deutlich wie bei Isaac Asimovs Konzept des Frankenstein Complex.

Science Fiction und Realität

Was also ist der Frankenstein Complex? Im Rahmen seiner Roboterromane¹⁶³ beschäftigte sich der Science-Fiction-Autor Isaac Asimov ausführlich mit der Außenwirkung von fiktiven Robotern. Denn der Autor wollte, dass Roboter als nützliche und den Menschen dienliche Maschinen wahrgenommen werden, und nicht wie Mary Shelley's *Frankenstein*, als Monster und Abschreckung vor technologischem Fortschritt auftreten. Dies bezieht sich zum einen auf die Rezipient*innen, um in ihnen keine (zusätzliche) Furcht vor wildgewordenen Maschinen zu initiieren.

»Isaac Asimov coined the term [...] to describe the fear that the general public has towards human made technologies when they invade the realm commonly considered to be God's domain.«¹⁶⁴

¹⁶² Wie etwa das Phänomen des Uncanny Valley.

¹⁶³ Dabei handelt es sich um 37 Kurzgeschichten und 5 Romane, wobei *I, Robot* eine Kombination aus 9 bereits veröffentlichten Kurzgeschichten darstellt. (https://asimov.fandom.com/wiki/Robot_series Abgerufen am 17. Dezember 2022)

¹⁶⁴ McCauley, Lee; Countering the Frankenstein Complex. In: Multidisciplinary Collaboration for Socially Assistive Robotics, Papers from the 2007 AAAI Spring Symposium, Technical Report SS-07-07, March 26-28; 2007; Stanford: AAAI; S. 42

Zum anderen ist der Frankenstein Complex aber ebenso an die Vorstellung gekoppelt, dass die betroffenen Maschinen vermeintlich belebt wären, also ein eigenes Bewusstsein, Ziele und Bedürfnisse besäßen. Im Rahmen von Science-Fiction-Handlungen ist dies häufig der Fall,¹⁶⁵ in der Realität scheitert dies bereits am Versuch Maschinen mit Emotionen zu versehen. Doch wie die Darstellung von HAL 9000 im Rahmen der Handlung von *2001: A Space Odyssey* zeigt, besitzen die fiktiven (belebten) Maschinen, eine dystopische (oder auch utopische) Vorbildwirkung, und können damit das Vertrauen oder Misstrauen gegenüber realen Technologien stark beeinflussen. Um also das Problem zu vermeiden, dass Nutzer*innen vor allem negative Vorbehalte von Science-Fiction-Handlungen auf reale Maschinen projizieren, gab ihnen Asimov, im Rahmen seiner Roboterromane eine unmissverständliche Eindeutigkeit: Die »three laws of robotics«¹⁶⁶ sollen sowohl Vertrauen auf der Seite der Leser*innen (also den späteren Nutzer*innen) erzeugen, als auch eine Handlungsanweisung sein, wie Roboter konstruiert werden müssten.

»One, a robot may not injure a human being, or, through inaction, allow a human being to come to harm. [...] Two, [...], a robot must obey the orders given it by human beings except where such orders would conflict with the First Law. [...] And three, a robot must protect its own existence as long as such protection does not conflict with the First or Second Laws.«¹⁶⁷

Diese drei Gesetzmäßigkeiten – die die Maschinen zumindest in der Science Fiction von Isaac Asimov erfüllen – sollen die Interaktion nachvollziehbar und vertrauensvoll werden lassen. Natürlich ist diese (Roboter-)Moral in der Realität nicht eins zu eins in tatsächliche Software oder Hardware übersetzbar. Trotzdem, laut den Zukunftsvisionen des Autors sollen Roboter gesellschaftliche Verbreitung erreichen, und zwar ohne Misstrauen – und mit dem Ziel, dass die Maschinen trotz umfassender Funktionalität keine tatsächliche Gefahr für Menschen darstellen.

¹⁶⁵ Neben HAL 9000 in *2001: A Space Odyssey* wären auch *Terminator* (1984, 1991, 2003), *Do Androids Dream of Electric Sheep?* (1968), oder *Westworld* (1974) prominente Beispiele in denen Maschinen belebt sind.

¹⁶⁶ Die drei Robotergesetze erscheinen dabei erstmals 1947 in Asimovs *I, Robot*. – Einer Sammlung an Kurzgeschichten mit und über Roboter; Murphy, Robert R. et. al.; *Beyond Asimov: The Three Laws of Responsible Robotics*. In: *IEEE Intelligent Systems*, Volume 24, Issue 4; 2009; Piscataway: IEEE; S. 14

¹⁶⁷ Asimov, Isaac; *I, Robot*; 2004; New York: Bantam Dell; S. 37

Dabei stellt sich die Frage, ob reale Maschinen die drei Gesetze überhaupt stringent einhalten könnten. Um etwa das erste Gesetz tatsächlich nutzbar machen zu können, müssten Roboter zum einen die notwendige Wahrnehmung und Interpretationsfähigkeit besitzen, um Menschen (vermeintlich fehlerfrei) erkennen zu können. Darüber hinaus müssten die jeweiligen Maschinen einen Wertekodex besitzen, der etwaige Schäden für Menschen eindeutig definiert.¹⁶⁸ Beides ist dabei, zumindest gegenwärtig, nicht real umsetzbar. Das zweite Gesetz scheitert nicht nur an der Problematik, dass die Einigung auf eine lingua franca notwendig wäre (die dann alle Nutzer*innen erlernen müssten).¹⁶⁹ Sondern es ist ebenso unrealistisch, dass ökonomische, soziale und politische Motivationen hintangestellt würden, um jede Maschine durch jede*n Nutzer*in steuerbar zu machen.¹⁷⁰ Das dritte Gesetz scheitert schließlich, sobald es um mehr als etwaigen ökonomischen Verlust geht. Denn damit die Existenz der jeweiligen Roboter mehr als nur finanzielles Investment wäre,¹⁷¹ müssten sich Maschinen als, mehr oder weniger, lebendige Entitäten begreifen, für die ein Schaden mehr als nur kaputte Hardware bedeutet.¹⁷²

Demnach ist es nicht nur auf technologischer Ebene unmöglich, die drei Gesetze in ihrer gegenwärtigen Form in realen Maschinen umzusetzen. – Die Etablierung der three laws würde im Umkehrschluss auch keine adäquate (Roboter-)Moral entstehen lassen, geschweige denn dafür sorgen, dass Maschinen nachvollziehbarer oder vertrauensvoller werden. Ein weitaus geeigneterer Ansatz ist es, die three laws lediglich als Denkanstoß für die Forschung und Entwicklung zu betrachten, und zu fragen, ob und wie die jeweilige Maschine Menschen schaden könnte, wie sie gesteuert werden soll, und wie sie sich (ob mit oder ohne Absicht) selbst vernichten könnte.

¹⁶⁸ Siehe dazu auch: Murphy, Robert R. et. al.; Beyond Asimov: The Three Laws of Responsible Robotics. In: IEEE Intelligent Systems, Volume 24, Issue 4; 2009; Piscataway: IEEE; S. 15

¹⁶⁹ Also eine vergleichsweise eindeutige Verkehrssprache, der sich alle Beteiligten bedienen. Theoretisch könnte dabei auch eine computer language (also eine Form von Computercode) zur lingua franca avancieren.

¹⁷⁰ Abgesehen davon würde damit jedwede Zertifizierung ungültig werden, da jede*r Nutzer*in die Aufgaben und Genauigkeit der Maschine verändern könnte.

¹⁷¹ Selbst Lifestyle Produkte sollen sich durch Langlebigkeit bezahlt machen.

¹⁷² Die Figur des HAL 9000 äußert etwa die Angst vor dem Tod. Siehe dazu: Clarke, Arthur C.; 2001: A Space Odyssey; 2018; London: Orbit S. 162 f.

Zusätzlich zur Funktion als Denkanstoß (und als impraktikables Modell für reale Maschinen) muss schließlich vor allem auf die Effekte der three laws fokussiert werden. Denn obwohl die drei Gesetze nicht wirklich in die Realität übersetzbar sind, haben sie doch einen positiven Nebeneffekt. Wie HAL 9000 in *2001: A Space Odyssey* als Warnung vor einer – wenn auch unrealistischen – dystopischen Zukunft fungiert, so schaffen es die three laws of robotics trotz ihrer Realitätsferne, Kompetenzen im Umgang mit realen Maschinen zu erzeugen. Besonders im Spektrum der Roboterromane wird mit verschiedenen Ausprägungen (und Auslegungen) derselben drei Gesetzmäßigkeiten experimentiert. Dies erzeugt schließlich eine Vielzahl an Blickwinkeln, wie sich die Logik von Maschinen ausprägen kann, und wo diese Ausprägungen problematisch werden. Die Darstellung von verschiedensten Ausprägungen von Maschinen, und der plurale Blick auf technologische Chancen und Risiken (sowie mögliche Effekte) kann auch als genereller Positiveffekt von Science Fiction gelesen werden. Selbst die wildgewordenen Maschinen reihen sich in die Bandbreite an möglichen, technologischen Nova ein. Und je mehr Zukunftsvisionen skizziert werden, desto mehr Referenzen zu sozialen, ökonomischen, politischen und auch kulturellen Implikationen von technologischen Entwicklungen gibt es. Davon profitieren letzten Endes nicht nur Forschung und Entwicklung, sondern auch die Nutzer*innen: Entweder direkt, weil ihr Blick durch Science Fiction auf die Möglichkeiten und Probleme hin geschult ist, oder indirekt, da Dystopien auch *bad for business* sind.

Konklusion

Abschließend gilt es also, ausgehend von der Definition von smarten Kommunikationsassistenten, die Frage zu beantworten wie die Interaktion mit smarten Assistenten funktioniert, und sich die Dynamik zwischen jenen und ihren Nutzer*innen gestaltet. Zusätzlich zum Interfacecharakter besitzen smarte Assistenten diverse Besonderheiten, wie etwa den Zugang zu diversen Datenbanken des Herstellers und ein individuelles (Client-spezifisches) wie auch kollektives (zentralisiertes, Server-gestütztes) Lernverhalten. In der Verwendung von smarten Assistenten entweder als Lifestyle-Produkte, oder auch für medizinische Zwecke, ergeben sich schließlich zwei Eigenschaften, die die Dynamik wesentlich beeinflussen.

Erstens ist das Potenzial zur Automation entscheidend, das sich durch die Integration von zusätzlichen IoT-Peripheriegeräten in das Framework des smarten Assistenten und der daraus möglichen Funktionalität ergibt. Hier gilt es sowohl seitens der Hersteller, als auch seitens der Nutzer*innen nicht nur neue Möglichkeiten der Automation zu beschreiben, sondern auch damit einhergehende Risiken abzuschätzen und zu minimieren. Dies kann teils durch bessere, (also umfassendere) Datenerhebung erreicht werden, teils geht es jedoch darum, die jeweiligen Funktionen zu zertifizieren. Um diese Zertifizierung gewährleisten zu können, darf die Programmierung dabei nicht von außen beeinflusst werden, weswegen Software und Berechnungsmethoden von smarten Assistenten meist verschlüsselt sind. Das bedeutet aber im Umkehrschluss, dass auch etwaige Fehler oder der Bias von den Nutzer*innen selbst nur schwer zu erkennen und kaum zu beheben sind.

Zweitens geht es in der Interaktion mit smarten Assistenten besonders um Vertrauen. Dieser notwendige Vertrauensvorschuss beginnt bei der Definition von intelligenten Maschinen, und erstreckt sich über die Technologien des Affective Computing, wo Nutzer*innenprofile (die den Nutzer*innen selbst verborgen bleiben) maßgeblich für die Genauigkeit der Assistenten sind. Die Akkuranz der Nutzer*innenprofile basiert einerseits auf der Genauigkeit der erhobenen Daten, und andererseits auf der Auswertung durch etwa (schwer einsehbare) Künstliche Neuronale Netzwerke. Das bedeutet: Je umfassender Technologien, wie etwa smarte Assistenten, auf das Wohl ihrer Nutzer*innen Einfluss nehmen können, desto wesentlicher wird die Genauigkeit dieser Technologien und das Vertrauen, das ihnen (beziehungsweise den Zertifizierungen) entgegengebracht werden muss.

Darüber hinaus muss ebenso die Rolle von Science Fiction thematisiert werden, die sowohl in der Forschung und Entwicklung von smarten Technologien als Prototypen betrachtet werden, als auch Gefühle der Bedrohung und Unheimlichkeit von der Fiktion in die Realität tradieren kann. So kann die Narration des wildgewordenen HAL 9000 Angst auf smarte Kommunikationsassistenten projizieren, sei es als (mehr oder weniger) bewusste Illusion, oder als unbewusste Angst vor der Belebtheit von Maschinen. Außerdem gibt es methodische Zugänge, wie solche Unheimlichkeiten umgangen, beziehungsweise reduziert werden können.

Das Uncanny-Valley-Phänomen – der uneindeutigen Maschine, die zwischen (belebten) fiktiven Maschinen und realen Maschinen oszilliert – lässt sich entweder durch die Abstraktion, die aus der Maschine wieder ein Ding macht, oder durch einen konsistenten Auftritt (also eine eindeutige Zwischenrolle) minimieren. Letztlich bezieht ich diese Unheimlichkeit jedoch kaum auf das Sein (also die Technologie), sondern weitgehend auf die Rezeption (also das Auftreten) von Maschinen.

Denn auf technologischer Ebene lässt sich festhalten: Selbst falls ein smarterer Assistent einen vermeintlichen Nervenzusammenbruch erleiden würde; – was unmöglich ist, da Emotionen vielleicht funktional quantifizierbar sind, aber nicht von Maschinen (nach-)empfunden werden können, sondern ausschließlich als Berechnungsdaten zur Wahrscheinlichkeitsrechnung benutzt werden. – Selbst wenn der Nervenzusammenbruch möglich wäre, so könnte die Maschine uns kaum schaden. Jedoch nicht aufgrund von etwaigen moralischer Gesetzen oder irgendwelchen Hardware- oder Software-Limitationen, sondern vor allem weil die Hersteller nicht bereit sind, für solch tiefgreifende Fehler zu haften. Und gerade diese Frage der Verantwortung ist entscheidender, als die tatsächliche Bedrohlichkeit.

Ausblick

Abschließend soll der Aspekt der Verantwortung an zwei Beispielen diskutiert werden, und zwar bei den Softwarelösungen *Chat GPT* und *Dall-E 2*. Bei *Chat GPT* handelt es sich dabei um ein Künstliches Neuronales Netzwerk, das darauf trainiert ist, aus stichwortartigen Konzeptanweisungen eigenständig Texte zu generieren, oder Fragestellungen zufriedenstellend¹⁷³ zu beantworten.¹⁷⁴ Die Software reicht dabei bereits so weit, dass einige Stimmen meinen, *Chat GPT* könne auf Knopfdruck publikationsfertige Inhalte zaubern.¹⁷⁵ Ähnliche Ansichten gibt es bezüglich *Dall-E 2*. Diese Software generiert aus denselben Handlungsanweisungen keinen Text, sondern ein Bild.¹⁷⁶ Erstaunlicherweise hält sich (zum Verfassungs-Zeitpunkt dieser Arbeit) die Ansicht, dass etwa Grafiker*innen aufgrund dieses technologischen Fortschritts ihre Jobs verlieren würden, und nun alles über Künstliche Neuronale Netzwerke generiert werden sollte.¹⁷⁷

Wenn etwa der Auftritt von Firmen durch Maschinen generiert werden soll, beschreibt dies ein ähnliches Zusammenspiel von Funktionalität, Risiko und Verantwortung, sowie Vertrauen. Und auch hier gilt es entweder der jeweiligen Software das Vertrauen entgegenzubringen – also anzunehmen, *Dall-E 2* könne etwa ebenso gut illustrieren wie ein*e Grafiker*in. Oder es tritt der Fall ein, dass entweder der Hersteller vermeintlich publikationsfertige Inhalte garantiert, und anderweitig dafür haftet. – Diese Option ist jedoch relativ unrealistisch.

¹⁷³ Hier geht es also nicht nur darum, korrekte Quellenlagen zu verwenden, sondern Inhalte bereits adäquat aufzubereiten.

¹⁷⁴ Siehe dazu: ChatGPT: Optimizing Language Models for Dialogue; 2022; OpenAI; (<https://openai.com/blog/chatgpt/> Abgerufen am 9. Jänner 2023)

¹⁷⁵ Siehe etwa: Kurz, Yael; How I Used Chat GPT and One AI to Write and Package my SpongeToad Spec Script; 2022; One AI; (<https://www.oneai.com/learn/chat-gpt-3-text-analysis> Abgerufen am 9. Jänner 2023)

¹⁷⁶ Siehe dazu: DALL·E 2. DALL·E 2 is a new AI system that can create realistic images and art from a description in natural language.; 2023; OpenAI; (<https://openai.com/dall-e-2/> Abgerufen am 9. Jänner 2023)

¹⁷⁷ Siehe: Patel, Ankur A.; Sostarec, Dina; The Future Of DALL-E 2 And What It Means For The Graphic Design Industry; 2022; Ankur's Newsletter; (<https://www.ankursnewsletter.com/p/the-future-of-dall-e-2-and-what-it> Abgerufen am 9. Jänner 2023)

Und auch der Anspruch, dass Künstliche Neuronale Netzwerke das Niveau von menschlichem Fachpersonal erreichen können (oder sollen) ist fehlgeleitet. Dies zeigt sich am besten an einem (dystopischen) Beispiel.

Man nähme den Fall, eine Software – etwa *Dall-E 2* – würde damit beauftragt werden, ein Poster zu entwerfen. Und obwohl dieses Poster vor Print und Publikation kontrolliert werden würde, ist dafür kein Fachpersonal mehr zuständig. Es ginge lediglich darum, vermeintliche offensichtliche Fehler (etwa bei der Grammatik) zu erkennen. Bei allem anderen (wie Layout, Illustration, Typographie, etc.) würde der Software vertraut. Obwohl der Hersteller nicht für etwaige Fehler haftet, hätte sich die Software als fehlerfrei bewiesen. Was würde nun passieren, wenn sich ein Fehler einschleust, und dieser Fehler von den Kontrollierenden, die allesamt keine Grafiker*innen, beziehungsweise Texter*innen mehr wären, nicht erkannt wird?

Nun hängt es davon ab, wie gravierend diese Fehler sein können. Unpassende Zeilenumbrüche, oder auch Rechtschreibfehler sind noch vergleichsweise harmlos. Wenn aber nun Begriffe oder Symbole benutzt werden, die etwa politisch oder kulturell belegt sind, dann können Fehler schnell problematisch werden, und weitreichende Folgen haben. Um diesen Extremfall zu vermeiden, muss die Funktionalität eingeschränkt werden – die Software darf also nicht autonom agieren. Oder es kann die Zertifizierung (und damit verbunden Haftung) verbessert (beziehungsweise überhaupt etabliert) werden. Und schließlich gibt es noch eine dritte Möglichkeit: Die Fehler können als notwendiges Übel, um funktionale Technologien zu erzeugen, toleriert werden.¹⁷⁸

In jedem Fall ist jedoch die Funktionalität von Künstlichen Neuronalen Netzwerken ab einem gewissen Niveau immer an Vertrauen, Verantwortung und zwangsläufige Haftung geknüpft.

¹⁷⁸ Fragwürdige Symbole oder Begriffe, sind auf einer Skala des Scheiterns dabei noch vergleichsweise tolerabel, man denke nur an das Gefahrenpotential von selbstfahrenden Automobilen, die spielende Kinder übersehen.

Literaturverzeichnis

Andreas, Michael et. al. (Hg.); Unterwachen und Schlafen: Anthropophile Medien nach dem Interface; 2018; Lüneburg: meson press

Asimov, Isaac; I, Robot; 2004; New York: Bantam Dell

Asimov, Isaac; The Machine and the Robot. In: Warrick, Patricia et. al.; Science Fiction: Contemporary Mythology; 1978; New York: Harper and Row

Caracciolo, Marco; Bones in Outer Space: Narrative and the Cosmos in 2001: A Space Odyssey and Its Remediations. In: Image [&] Narrative; Vol. 16, No. 3 (2015); Löwen: Katholieke Universiteit Leuven;

Chattopadhyay, D.; MacDorman, K. F.; Familiar faces rendered strange: Why inconsistent realism drives characters into the uncanny valley. 2019; Journal of Vision, 16(11):7

Chattopadhyay, D.; MacDorman, K. F.; Reducing consistency in human realism increases the uncanny valley effect; increasing category uncertainty does not. 2016; Cognition, Volume 146

Clarke, Arthur C.; 2001: A Space Odyssey; 2018; London: Orbit

Evans, Dave; The Internet of Things. How the Next Evolution of the Internet Is Changing Everything; 2011; Cisco IBSG

Friedman, Batya; Nissenbaum, Helen; Bias in Computer Systems. In: ACM Transactions on Information Systems, Volume 14, Issue 3; 1996; New York: Association for Computing Machinery

Frinzi, Joe R.; Kubrick's Monolith. The Art and Mystery of 2001: A Space Odyssey; 2017; Jefferson: McFarland & Company

Hauswald, Johann et. al.; Designing Future Warehouse-Scale Computers for Sirius, an End-to-End Voice and Vision Personal Assistant. In: ACM Transactions on Computer Systems, Volume 34; Issue 1; 2016; New York: Association for Computing Machinery

Knote, Robin et. al.; Value Co-Creation in Smart Services: A Functional Affordances Perspective on Smart Personal Assistants. In: Journal of the Association for Information Systems, Volume 22, Issue 2, 2021

Krajewski, Markus; Der Diener. Mediengeschichte einer Figur zwischen König und Klient, 2010, Frankfurt am Main: Fischer

Kubrick, Stanley; 2001: A Space Odyssey; 1970; Los Angeles: Warner Brothers

Le Cun, Yann; A theoretical framework for back-propagation. In: Touretzky, David et. al. (Hg.); Proceedings of the 1988 Connectionist Models Summer School; 1988; San Mateo: Morgan Kaufmann

Maedche, Alexander et. al.; AI-based Digital Assistants: Opportunities, Threats, and Research Perspectives. in: (Journal) Business and Information Systems Engineering, Issue 4; 2019; Wiesbaden: Springer Fachmedien

Manyika, James et.al.; The Internet of Things: Mapping the Value beyond the Hype; 2015; McKinsey Global Institute: McKinsey & Company

Masahiro, Mori; Das Unheimliche Tal. Übersetzung: MacDorman, Karl et. al.; 2019; Hamburg: Textem

McCauley, Lee; Countering the Frankenstein Complex. In: Multidisciplinary Collaboration for Socially Assistive Robotics, Papers from the 2007 AAAI Spring Symposium, Technical Report SS-07-07, March 26-28; 2007; Stanford: AAAI

Mehnert, Wenzel; The Future is Near: Schnittstellen einer negativen Zukunft. 2019; Hamburg: Textem

Miraz, Mahdi H. et al.; A review on Internet of Things (IoT), Internet of Everything (IoE) and Internet of Nano Things (IoNT). In: Internet Technologies and Applications (ITA); 2015; Wrexham: IEEE

Modi, Kirit J. et al.; Cloud computing - concepts, architecture and challenges. In: International Conference on Computing, Electronics and Electrical Technologies (ICCEET), 2012; Nagercoil: IEEE

- Murphy, Robert R. et. al.; Beyond Asimov: The Three Laws of Responsible Robotics. In: IEEE Intelligent Systems, Volume 24, Issue 4; 2009; Piscataway: IEEE
- Perazzo, Pericle et al.; Performance Evaluation of Attribute-Based Encryption on Constrained IoT Devices. In: Computer Communications, Volume 170; 2021; Sciencedirect: Elsevier
- Picard, Rosalind W.; Affective Computing; 1997; Massachusetts: MIT Press
- Rojas, Raúl; Künstliche Neuronale Netze als neues Paradigma der Informationsverarbeitung. In: Pauen, Michael; Roth, Gerhard (Hg.); Neurowissenschaften und Philosophie: eine Einführung; 2001; Frankfurt am Main: UTB
- Salovey, Peter; Mayer, John D.; Emotional Intelligence. In: Imagination, Cognition and Personality, Volume 9, Issue 3; 1990; Sage
- Salovey, Peter; Mayer, John D.; The Intelligence of Emotional Intelligence. In: Intelligence, Volume 17, Issue 4; 1993; Elsevier
- Selke, Stefan (Hg.); Lifelogging. Digital self-tracking and Lifelogging – between disruptive technology and cultural transformation; 2016; Wiesbaden: Springer VS
- Singhania, Vivek; The Internet of Things: An Overview. Understanding the Issues and Challenges of a More Connected World; 2015; Genf, Reston: The Internet Society (ISOC)
- Sinha, Alok; Client-Server Computing. In: Communications of the ACM, Volume 35, Issue 7, 1992; New York: Association for Computing Machinery
- Steinmüller, Karlheinz; Science Fiction: eine Quelle von Leitbildern für Innovationsprozesse und ein Impulsgeber für Foresight. 2010; Bonn: iFQ
- Sudmann, Andreas; Künstliche neuronale Netzwerke als Black Box. Verfahren der Explainable AI. Medienwissenschaftliche Perspektiven. In: Klimczak, Peter e. al. (Hg.); Maschinen der Kommunikation. Interdisziplinäre Perspektiven auf Technik und Gesellschaft im digitalen Zeitalter; 2020; Wiesbaden: Springer Vieweg
- Turing, Alan M.; I. – Computing Machinery and Intelligence. In: Mind, Volume LIX, Issue 236; 1950; Oxford: Oxford University Press

Turing, Alan M.; Intelligent Machinery; 1948; In: Meltzer, Bernard Et. al. (Hsg.); Machine Intelligence 5; 1969; Edinburgh: Edinburgh University Press

Wagner, Gerald; Vertrauen in Technik. In: Zeitschrift für Soziologie, Jg. 23, Heft 2, 1994; Stuttgart: Ferdinand Enke Verlag

Walsh, Toby; The troubling future for facial recognition software. In: Communications of the ACM, Volume 65, Issue 3; 2022; New York: Association for Computing Machinery

Weizenbaum, Joseph, Computer Power and Human Reason. From Judgment to Calculation; 1976; San Francisco: W.H. Freeman

Williamson, James, et. al.; Data sensing and analysis: Challenges for wearables; In: 20th Asia and South Pacific Design Automation Conference, ASP-DAC; 2015

Yorke-Smith, Neil et. al.; The Design of a Proactive Personal Agent for Task Management. In: International Journal on Artificial Intelligence Tools, Volume 21, Issue 1, 1250004; 2012; World Scientific Publishing Company;

Digitalquellen

Ananthakrishnan, Shankar; Amazon Scientists applying deep neural networks to custom skills; 2020; Amazon Science

(<https://www.amazon.science/blog/amazon-scientists-applying-deep-neural-networks-to-custom-skills> Abgerufen am 26. November 2022)

Apple Developer Documentation; Healthkit;

(<https://developer.apple.com/documentation/healthkit> Abgerufen am 1. Dezember 2022)

Apple Machine Learning Research; Hey Siri: An On-device DNN-powered Voice Trigger for Apple's Personal Assistant;

(<https://machinelearning.apple.com/research/hey-siri> Abgerufen am 27. November 2022)

Asimov Wiki; Robot Series;

(https://asimov.fandom.com/wiki/Robot_series Abgerufen am 17. Dezember 2022)

ChatGPT: Optimizing Language Models for Dialogue; 2022; OpenAI;

(<https://openai.com/blog/chatgpt/> Abgerufen am 9. Jänner 2023)

DALL·E 2. DALL·E 2 is a new AI system that can create realistic images and art from a description in natural language.; 2023; OpenAI;

(<https://openai.com/dall-e-2/> Abgerufen am 9. Jänner 2023)

Delaney, John R.; Colon, Alex; Moscaritolo, Angela; What Is a Smart Home Hub (And Do You Need One)?;

(<https://uk.pcmag.com/home-automation-hubs/35777/what-is-a-smart-home-hub-and-do-you-need-one> Abgerufen 23. November 2022)

Kurz, Yael; How I Used Chat GPT and One AI to Write and Package my SpongeToad Spec Script; 2022; One AI;

(<https://www.oneai.com/learn/chat-gpt-3-text-analysis> Abgerufen am 9. Jänner 2023)

Patel, Ankur A.; Sostarec, Dina; The Future Of DALL-E 2 And What It Means For The Graphic Design Industry; 2022; Ankur's Newsletter;

(<https://www.ankursnewsletter.com/p/the-future-of-dall-e-2-and-what-it> Abgerufen am 9. Jänner 2023)

Sharma, Chetan; Correcting the IoT History;

(<http://www.chetansharma.com/correcting-the-iot-history/> Abgerufen am 22. November 2022)

Sleeptracker-AI; Device FAQ;

(<https://www.sleeptracker.com/faq/tempur> Abgerufen am 13. Dezember 2022)

Sleeptracker-AI; Select Your Model;

(<https://www.sleeptracker.com/support/tempur> Abgerufen am 13. Dezember 2022)

SleepWatch by Bodymatter; Introducing Snore Tracking and Sleep Sound Recording for SleepWatch;

(<https://www.sleepwatchapp.com/blog/introducing-snore-tracking-and-sleep-sound-recording-sleepwatch/> Abgerufen am 24. August 2022)

SleepWatch by Bodymatter; SleepWatch Terms of Service;

(<https://www.sleepwatchapp.com/terms-of-service/in-app/> Abgerufen am 20. August 2022)

The Apollo guidance computer: Hardware. In: Computers in Spaceflight: The NASA Experience; Chapter Two. Computers On Board The Apollo Spacecraft;

(<https://history.nasa.gov/computers/Ch2-5.html> Abgerufen am 1. Dezember 2022)

Verification and Validation. In: Computers in Spaceflight: The NASA Experience; Chapter Two. Computers On Board The Apollo Spacecraft;

(<https://history.nasa.gov/computers/Ch2-6.html> Abgerufen am 1. Dezember 2022)

Abkürzungsverzeichnis

AGC	Apollo Guidance Computer
DNN	Deep Neural Network
IOS	Apple's Mobil-Betriebssystem
IoT	Internet of Things
IPA	Intelligent Personal Assistant
KNN	Künstliches Neuronales Netzwerk
L.C.M.	Logical Computing Machine
SPA	Smart Personal Assistant



CC BY-NC-ND 4.0 International
Namensnennung - Nicht-kommerziell - Keine Bearbeitung 4.0 International