

Digitaler Humanismus in Wirtschaft, Wissenschaft
und Gesellschaft

Georg Krause
Martin Giesswein *Hrsg.*

Management Guide Digitaler Humanismus

Unternehmensführung
in der digitalen Welt



Springer Gabler

Digitaler Humanismus in Wirtschaft, Wissenschaft und Gesellschaft

Reihe herausgegeben von

Georg Krause, msg Plaut AG, Wien, Österreich

Martin Giesswein, WU Executive Academy, Wien, Österreich

Die Buchreihe *Digitaler Humanismus in Wirtschaft, Wissenschaft und Gesellschaft* widmet sich den Fragen, wie digitale Technologien im Einklang mit humanistischen Werten – wie sie der Digitale Humanismus und die digitalen Rechte und Grundsätze der Europäischen Union vorgeben – gestaltet und eingesetzt werden können. Die einzelnen Bände setzen sich mit den Grundgedanken des Digitalen Humanismus auseinander, zeigen neue Forschungsergebnisse, Entwicklungen sowie Anwendungsbeispiele auf und vermitteln Handlungsempfehlungen für die wirtschaftliche Praxis.

Die Reihe bietet ein interdisziplinäres Forum zum inhaltlichen Diskurs, wie es gelingen kann, den Menschen und die Gesellschaft in den Mittelpunkt der digitalen Transformation zu stellen und damit einen nachhaltigen, menschenzentrierten Weg der Digitalisierung zu entwickeln. Sie richtet sich an Fach- und Führungskräfte in Unternehmen, an Expertinnen und Experten aus Wissenschaft und Forschung sowie an Studierende und Lehrende, die an einer ethisch fundierten digitalen Transformation interessiert sind.

Georg Krause • Martin Giesswein
(Hrsg.)

Management Guide Digitaler Humanismus

Unternehmensführung
in der digitalen Welt



Springer Gabler

Hrsg.
Georg Krause
Wien, Österreich

Martin Giesswein
Wienersdorf, Österreich

ISSN 3091-3225

ISSN 3091-3233 (electronic)

Digitaler Humanismus in Wirtschaft, Wissenschaft und Gesellschaft

ISBN 978-3-658-49868-9

ISBN 978-3-658-49869-6 (eBook)

<https://doi.org/10.1007/978-3-658-49869-6>

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

© Der/die Herausgeber bzw. der/die Autor(en), exklusiv lizenziert an Springer Fachmedien Wiesbaden GmbH, ein Teil von Springer Nature 2026

Das Werk einschließlich aller seiner Teile ist urheberrechtlich geschützt. Jede Verwertung, die nicht ausdrücklich vom Urheberrechtsgesetz zugelassen ist, bedarf der vorherigen Zustimmung des Verlags. Das gilt insbesondere für Vervielfältigungen, Bearbeitungen, Übersetzungen, Mikroverfilmungen und die Einspeicherung und Verarbeitung in elektronischen Systemen.

Die Wiedergabe von allgemein beschreibenden Bezeichnungen, Marken, Unternehmensnamen etc. in diesem Werk bedeutet nicht, dass diese frei durch jede Person benutzt werden dürfen. Die Berechtigung zur Benutzung unterliegt, auch ohne gesonderten Hinweis hierzu, den Regeln des Markenrechts. Die Rechte des/der jeweiligen Zeicheninhaber*in sind zu beachten.

Der Verlag, die Autor*innen und die Herausgeber*innen gehen davon aus, dass die Angaben und Informationen in diesem Werk zum Zeitpunkt der Veröffentlichung vollständig und korrekt sind. Weder der Verlag noch die Autor*innen oder die Herausgeber*innen übernehmen, ausdrücklich oder implizit, Gewähr für den Inhalt des Werkes, etwaige Fehler oder Äußerungen. Der Verlag bleibt im Hinblick auf geografische Zuordnungen und Gebietsbezeichnungen in veröffentlichten Karten und Institutionsadressen neutral.

Springer Gabler ist ein Imprint der eingetragenen Gesellschaft Springer Fachmedien Wiesbaden GmbH und ist ein Teil von Springer Nature.

Die Anschrift der Gesellschaft ist: Abraham-Lincoln-Str. 46, 65189 Wiesbaden, Germany

Wenn Sie dieses Produkt entsorgen, geben Sie das Papier bitte zum Recycling.

Geleitwort

Der digitale Humanismus stellt die begrifflichen, philosophischen und ethischen, Grundlagen bereit, damit die digitale Transformation in Politik, Wirtschaft und Gesellschaft gelingen kann. Gelingen im Sinne einer Verbesserung der menschlichen Lebensbedingungen und der menschlichen Praxis. Der digitale Humanismus setzt dabei dort an, wo der klassische Humanismus aufgehört hat: Beide Varianten vertrauen auf die menschliche Vernunftfähigkeit und Kooperationsbereitschaft. Beide sind davon überzeugt, dass die volle Entfaltung menschlicher Fähigkeiten zur Humanisierung der Welt beiträgt. Beide sind in diesem Sinne optimistisch.

Pico della Mirandola hatte in *De Dignitate Hominis* argumentiert, dass der Mensch Gott ebenbildlich sei, auch insofern, als er von sich aus kreativ ist, die Künste und die Wissenschaften zur vollen Entfaltung bringen kann und diese wiederum die menschlichen Fähigkeiten vervollkommen. Die Technik ist ein Instrument, um uns auf dem Weg einer humaneren, besseren, Welt zu begleiten, in der wir mit den natürlichen Ressourcen nachhaltiger umgehen und möglichst alle mitnehmen. Die Technik darf den Menschen nicht überwältigen, ihn nicht hilflos oder ängstlich werden lassen, sie ist nicht das *Movens* der Geschichte, sondern ein Instrument, das wir einsetzen, um unsere Lebensbedingungen zu verbessern.

Für mich ist das Zentrum humanistischen Denkens seit der Antike die Idee menschlicher Autorschaft: die Fähigkeit, aus eigenen Gründen zu handeln, mit anderen zu klären, was für uns gemeinsam gut ist, Gründe auszutauschen und sie in eine kohärente Praxis umzusetzen. Eine nach humanistischen Prinzipien gestaltete digitale Transformation weitet die Fähigkeit zu menschlicher Autorschaft aus, schränkt sie nicht ein oder schafft sie gar ab. Der Mensch ist kein schlecht funktionierendes Softwaresystem, das durch besser funktionierende ersetzt werden sollte, vielmehr sollten Softwaresysteme unsere Urteilsfähigkeit, zum Beispiel in der Medizin verbessern, den Lernerfolg in der Interaktion zwischen Schülerinnen und Lehrern erleichtern, ressourcensparender wirtschaften, auch effektiver mit unserer Zeit umgehen.

Nur Menschen tragen Verantwortung für ihr Tun, sie können und müssen sich gegebenenfalls rechtfertigen, sie müssen sich auch selbstkritisch prüfen, ob das, was sie tun, angemessen ist. Nur wir sind auf diesem Planeten Handelnde. Auch wenn wir die Sprache des Handelns, der Interaktion und Kommunikation, der Beratung, ja, des Vertrauens

auf sogenannte digitale Agenten anwenden, sollten wir uns immer bewusst sein, dass dies nur metaphorisch sinnvoll ist. Es gibt dort niemanden, der sich kritisch überprüft, der uns wohlwollend berät, der uns empathisch begleitet. Das ist eine Verirrung des Geistes, die uralt ist, und die darin besteht, das Unbelebte zu beleben, weil dann sein Verhalten erklärbar erscheint. Die moderne Zivilisation ist aus einer animistischen Kultur, die sich über Jahrtausende entwickelt hat, hervorgegangen und hatte auch vor der digitalen Transformation Mühe, das animistische Denken so weit zurückzudrängen, dass Wissenschaft und Technik möglich wurden. Ein Opponent des digitalen Humanismus ist der KI-Animismus, also die Belebung des Unbelebten, die Annahme, dass Künstliche Intelligenzen, sofern sie sich so verhalten wie Menschen, menschliche Eigenschaften haben. Das haben sie nicht. Das ist eine vielleicht für manche schöne, für andere bedrohliche Illusion und ein kindlicher Rückfall in eine magische Welt.

Der zweite große ebenso bedrohliche Opponent des digitalen Humanismus ist der Mechanismus, also die Vorstellung, dass Menschen nichts anderes seien als Maschinen. Auch das eine alte Sichtweise (LLM-Maschinen). Die jeweils neueste technologische Entwicklung hat regelmäßig dazu geführt, dass das Weltbild dieser jeweiligen Technik angepasst wurde. In der Zeit des Europäischen Rationalismus im 17. Jahrhundert wurde die Welt als Uhr begriffen und in ihr der Mensch als ein Uhrwerk. Ein mechanistisches Weltbild. Gegenwärtig haben wir es mit einem digitalen Mechanismus zu tun, wonach menschliches Denken, menschliches Empfinden, menschliches Urteilen in den Begriffen von Software und Hardware, sowie algorithmischer Steuerung zu erfassen versucht wird. Der Mensch ist aber keine schlecht funktionierende Software, wie es in dem Bestseller von Yuval Harari *Homo Deus* heißt, sondern Autor oder Autorin des eigenen Lebens. Softwaresysteme simulieren Teilbereiche menschlichen Äußerungsverhaltens, in der Robotik auch Teilbereiche menschlichen Handelns. Aber man darf Simulation nicht mit Realisierung verwechseln.

Die beeindruckenden Large Language Models, nicht nur Chat GPT, sondern auch Google Bard, Perplexity und zuletzt DeepSeek können Fragen ganz unterschiedlichster Art beantworten und Aufgaben ausführen. Sie können im Stil bestimmter Autorinnen oder Autoren schreiben, Geschichten komponieren, Ratschläge geben, sogar Software entwickeln. Sie sind beeindruckend leistungsfähig und ihre Leistungsfähigkeit wird weiter steigen. Es ist naheliegend, sie als Instanzen eines umfassenden Weltwissens zu interpretieren, ihnen Intelligenz, gar Weisheit zuzuschreiben. Tatsächlich handelt es sich um probabilistische Sprachmodelle, die auf gigantischen Mengen im digital verfügbaren Textkorpora trainiert wurden. Zehntausende kuratieren diese Sprachmodelle, um zu verhindern, dass auch der große Unsinn, der digital verfügbar ist, sich in den Reaktionen niederschlägt, nicht immer mit Erfolg. Man sollte die Language Models als Amalgam von im Internet verfügbaren Texten und Daten, auf denen die Systeme trainiert wurden, interpretieren, nicht als Prozessor, der Beschreibungen, Empfehlungen, Beurteilungen und anderes generiert. In der Hochzeit der Aufklärung hatten d'Alembert und Diderot gehofft, das gesamte Weltwissen der unterschiedlichen Disziplinen und Erfahrungsbereiche enzyklopädisch vollständig zu erfassen. Wikipedia ist die zeitgenössische Variante dieser

aufklärerischen Idee, diesmal als gemeinnütziges Projekt, geschrieben von denjenigen, die sich für bestimmte Themen interessieren, unabhängig von Rang und Namen.

In den Bildungsinstitutionen sollten wir uns darauf rasch einstellen, dass die Nutzung von Chat GPT und Co für Unterrichtende und Lernende, Lehrende, Forschende und Studierende, selbstverständlich wird. Da die Übernahme fremden Gedankengutes ohne Quellenangabe ein Plagiat, oft genug auch eine Urheberrechtsverletzung darstellt, ist die nicht markierte Übernahme von Texten, Ideen, Daten, die Large Language Models generiert haben, keine eigene Leistung. Auch wenn man die entsprechenden Responses für eigene Zwecke modifiziert, wird daraus noch keine umfassend eigene Leistung. Es gehört zu den ethischen Implikationen des digitalen Humanismus, dass wir am Ethos der Autorschaft festhalten und diejenigen, die künstlerische, wissenschaftliche, technische und andere kreative Leistungen vollbringen, nicht enteignen. Es ist eine berechtigte Frage und Gegenstand einer Reihe von Auseinandersetzungen, ob LLMs selbst nicht aufgrund ihrer Genese eine Ansammlung von Urheberrechtsverletzungen bilden, die allerdings deswegen schwer nachzuweisen sind, weil die Spuren der Texte und Daten, auf denen sie trainiert wurden, oft kaum noch zurückverfolgt werden können.

Wie jede technologische Dynamik früherer Zeiten bringt auch die digitale Transformation massive Veränderungen in der Arbeits- und Berufswelt mit sich, sie verändert unser Alltagsverhalten und beeinflusst kulturelle Praktiken. Diese Änderungen in einer Form zu gestalten, die die menschliche Autorschaft individuell und kollektiv stärkt, neue Handlungsoptionen schafft und zur Humanisierung von Wirtschaft und Gesellschaft beiträgt, ist das oberste Ziel des digitalen Humanismus. Die Beiträge in diesem Band geben dazu in vielfältiger Weise Orientierung. Als Philosoph und derjenige, der diesen Begriff des digitalen Humanismus geprägt hat, bin ich froh, dass er die akademischen Gefilde längst verlassen hat, sich gewissermaßen auch verselbständigt und in die wirtschaftliche und gesellschaftliche Praxis Eingang findet. Dieses Buch wird dazu einen weiteren Beitrag leisten. Ich wünsche ihm viel Erfolg.

München
Juni 2025

Julian Nida-Rümelin

Geleitwort

Digitalisierung im Dienst des Menschen – Wiens Weg des Digitalen Humanismus

Wir leben in einer Zeit tiefgreifender Umbrüche. Die digitale Transformation prägt unsere Gesellschaft in einer Geschwindigkeit, wie es noch niemals der Fall war. Sie durchdringt nahezu jeden Lebensbereich – von der Kommunikation über die Verwaltung bis hin zur Wirtschaft, Bildung und Kultur. Die Auswirkungen reichen dabei von praktischen Erleichterungen im Alltag bis hin zu grundsätzlichen Fragen wie soziale Gerechtigkeit, demokratische Mitbestimmung und ethische Verantwortung. Vor diesem Hintergrund ist es unsere Aufgabe, diesen Wandel politisch zu gestalten – nicht bloß zu begleiten. Wien bekennt sich mit voller Überzeugung zum Digitalen Humanismus. Für mich persönlich bedeutet das: Der Mensch muss im Zentrum technologischer Entwicklung stehen – und nicht umgekehrt.

Fortschritt, der den Menschen dient

Technologie ist nicht neutral. Sie ist Ausdruck von Werten, Interessen und auch von Machtverhältnissen. Ob soziale Medien, künstliche Intelligenz oder algorithmische Entscheidungsprozesse – sie alle können Nutzen stiften oder Schaden anrichten. Sie können Freiheit fördern – oder auch Kontrolle verstärken. Deshalb darf Digitalisierung kein Selbstzweck sein. Im Gegenteil: Sie muss dem Gemeinwohl verpflichtet sein. Der Digitale Humanismus ist Wiens Antwort auf diese Herausforderung. Er ist mehr als ein ethisches Ideal – er ist Leitprinzip und Handlungsmaxime unserer Digitalpolitik. Wir haben den Anspruch, dass Wien eine Modellstadt für menschenzentrierte Digitalisierung ist – eine Stadt, in der technologische Innovation mit sozialer Verantwortung einhergeht. Das ist keine abstrakte Zukunftsvision, sondern bereits heute gelebte Realität: in der wertebasierten „Smart City Strategie“ der Stadt Wien genauso wie in ganz konkreten Einzelprojekten.

Struktur schaffen, Werte sichern

Um diesen Anspruch langfristig strukturell zu verankern, haben wir im aktuellen Regierungsprogramm die Etablierung eines „Hubs für Digitalen Humanismus“ vorgesehen. Dieser „Hub“ wird als Plattform fungieren, auf der interdisziplinäre Werteforschung, Digitalisierungsexpertise, zivilgesellschaftliches Engagement und Verwaltungspraxis zusam-

menkommen. Unser Ziel ist es, Wien damit nicht nur als Ort der Reflexion, sondern auch der Umsetzung zu positionieren – in Zusammenarbeit mit anderen Gebietskörperschaften, wissenschaftlichen Institutionen, Unternehmen und NGOs. Darüber hinaus setzen wir gezielt auf einen großangelegten internationalen Diskurs – etwa durch Vernetzungs- und Informationsaustausch-Formate zu den Auswirkungen von künstlicher Intelligenz in Kunst, Kultur und Wissenschaft. Denn gerade im internationalen Austausch wächst das Bewusstsein dafür, dass Europa seine eigenen Antworten auf die Herausforderungen der Digitalisierung finden muss. Wenn ich diesbezüglich einen Wunsch äußern dürfte, dann wäre es dieser: eine gemeinsame, eigenständige Strategie der EU, die den europäischen Werten klar Rechnung trägt. Dazu gehört auch die gezielte Förderung europäischer Technologien, um die Abhängigkeit von globalen Tech-Konzernen zu verringern.

Digitalisierung braucht Rahmen, Regeln – und Haltung

Die Auswirkungen von neuen Technologien sind gestaltbar. Schon in den 1960er Jahren wurde mit der Technikfolgenabschätzung ein Instrument geschaffen, um Veränderungen durch innovative Technologien frühzeitig zu erkennen und gesamtgesellschaftlich zu bewerten. Damals ging es um Autoverkehr oder Haushaltsgeräte. Heute geht es um nicht weniger als das: die politische Gestaltung eines Fortschritts, der Freiheit, Teilhabe und Gerechtigkeit nicht nur stärken, sondern auch garantieren soll. Wien leistet hierzu seinen Beitrag – von der Digitalen Agenda Wien über Beteiligungsplattformen bis hin zu ethischen Leitlinien für KI. Auch in der Verwaltung denken wir Digitalisierung ganzheitlich. Beispielsweise bei der Modernisierung von archivischen oder dokumentarischen Prozessen, einschließlich verbindlicher rechtlicher Rahmenbedingungen.

Eine Politik für die nächste Generation

Digitalisierung wird unsere Städte, unsere Demokratien und unseren Alltag grundlegend verändern – das ist sicher. Offen ist die Richtung. Wien entscheidet sich bewusst für einen Weg, der den Menschen im Zentrum sieht. Dabei geht es nicht nur um Technologiepolitik, sondern um eine Vision für ein gutes Leben im 21. Jahrhundert. Mein persönlicher Beitrag als Bürgermeister ist es, die notwendigen Ressourcen und Mittel bereitzustellen, um diesen Weg konsequent zu verfolgen. Und dafür zu sorgen, dass den Menschen mit ihren Anliegen – auch und gerade im digitalen Raum – jederzeit zugehört wird. Denn digitaler Humanismus ist kein Zustand. Er ist ein Prozess – getragen von Haltung, Werten und der Bereitschaft, Verantwortung zu übernehmen.

Wien zeigt: Digitalisierung und Humanität müssen kein Widerspruch sein. Im Gegenteil – sie können sich gegenseitig stärken. Dieses Buch liefert dazu weitere wichtige Impulse. Ich wünsche Ihnen eine anregende Lektüre und viele neue Perspektiven auf die drängendsten Fragen unserer Zeit.

Wien, Juni 2025

Michael Ludwig

Bürgermeister der Stadt Wien

Executive Manual AKA Vorwort

Der Digitale Humanismus und seine Prinzipien haben in der europäischen Politik, Wirtschaft und Wissenschaft eine hohe Bedeutung, die in den letzten Jahren durch technische, aber vor allem durch geopolitische Entwicklungen noch deutlich gesteigert wurde.

Die Menschen erkennen, dass es bei diesem Thema um nicht mehr oder weniger als um unsere zukünftige Lebensqualität, unsere Sicherheit, unseren Wohlstand und nicht zuletzt um unsere Demokratie geht, wie wir sie heute kennen.

Die Frage ist daher, wie sieht der Weg in eine digitale Welt aus, die unseren Werten entspricht, wie wir sie aus der physischen Welt kennen, wo wir uns sicher und vertrauensvoll bewegen können und wir aber gleichzeitig die großartigen Vorteile der Digitalisierung nutzen können.

Um das zu erreichen, bedarf es des Zusammenspiels mehrerer Akteure. Eine Schlüsselrolle kommt der Politik und dem Gesetzgeber zu, die für die politische Festlegung der Werte, der Ziele und der daraus abgeleiteten Gesetze sind. Dieses europäische, politische Zielbild ist der Kompass und die Gesetze sind die Leitplanken für alle Staaten, Unternehmen und Bürger für deren Aktivitäten im digitalen Raum.¹

Die zweite Schlüsselrolle kommt den Unternehmen zu, die bei ihren Digital-Aktivitäten einen Weg finden müssen, Innovation und digitalhumanistische Werte natürlich im Rahmen der geltenden Gesetze zum Wohle der Kund:innen und Mitarbeitenden in Einklang zu bringen. Auf dieser Ebene geht es also um die Umsetzung, um die Integration mit den eigenen Werten, der eigenen Unternehmenskultur, dem eigenen Geschäftsmodell und den eigenen Unternehmenszielen.

Dabei stehen Unternehmen vor der großen Herausforderung, da die technologische Entwicklung der letzten Jahre so rasant war, dass es heute sehr herausfordernd ist, direkte und indirekte Wirkungen von IT-Lösungen abzuschätzen und nicht selten gut gemeinter Einsatz von Technologien zu unerwünschten Effekten führt oder massiven Reputationsverlust verursacht.

¹ Die EU hat bereits zahlreiche Aktivitäten im Rahmen der sogenannten Digitalen Dekade gesetzt. Als wesentliche Meilensteine sind wohl die von allen Mitgliedsstaaten 2022 verabschiedeten „Digitalen Rechte und Grundsätze“, sowie der Digital Markets Act, der Digital Services Act, die DSGVO und der AI Act zu erwähnen.

Verstärkt wird dieser Effekt dadurch, dass der Einsatz von Technologie in den letzten Jahren immer weitreichendere Auswirkungen auf unser aller Leben hat und in unseren Alltag unauflöslich integriert ist. Damit ist Technologieeinsatz, wie im „Wiener Manifest für Digitalen Humanismus“ (Wien 2019) ausgeführt, nicht mehr wertneutral und jeder, der Lösungen entwickelt, hat eine Verantwortung und leistet einen Beitrag zu unserer zukünftigen Lebenswelt.

Anleitung zur effektiven Nutzung des Buches

Wir wollen mit diesem Buch Menschen, die in ihren Organisationen Verantwortung tragen, Anregungen geben, wie sie in ihrem Wirkungsbereich Innovation und Werte erfolgreich in Einklang bringen können, worauf es zu achten gilt, wo man „hinschauen“ muss und welche Kompetenzen, Methoden und Rahmenbedingungen dafür relevant sind.

Das Buch ist von Praktikern und praxisnahen Wissenschaftlern geschrieben, damit möglichst konkrete Hinweise zur Umsetzung, sowie zahlreiche Beispiele aus der unternehmerischen Praxis dem:der Leser:in konkrete Ideen und Impulse für den eigenen Wirkungsbereich geben.

Der Aufbau des Buches ist so gestaltet, dass der:die Leser:in jene Beiträge gezielt auswählen kann, die für die eigene Arbeit, aktuelle Projekte etc. am besten geeignet sind. Im Folgenden werden stichwortartig die Beiträge beschrieben, um dem:der Leser:in einen raschen Überblick für den persönlichen „Einstieg“ in das Buch zu geben.

Im **ersten Teil des Buches** (Beiträge 1–7) steht die Sichtweise auf die Umsetzung der digitalhumanistischen Werte **unterschiedlicher Rollen** im Unternehmen im Vordergrund:

1. **CEO:** Nutzendarstellung für das Unternehmen; Governance für Digitalen Humanismus; Herangehensweisen; Praxisbeispiele u. a. Whitebook Digitaler Humanismus;
2. **CFO:** Erweiterung der Messkriterien für Projekterfolg; praktische Erfahrungen aus Value Based Engineering Projekt; Kosten- und Nutzen wertorientierter Entwicklung/ Projekte;
3. **CIO:** best practices der Stadt Wien bei Umsetzung des Digitalen Humanismus; „digitalhumanistischer“ KI-Einsatz; menschenzentrierte Systementwicklung am Beispiel von Lösungen der Stadt Wien;
4. **CIO (Daten):** Datenräume, Datensouveränität im Konnex mit KI; Lösungsansätze der EU u. a. Gaia-X; Technologie als Enabler menschenzentrierter Lösungen;
5. **CIO (Projektmanagement):** mit Pragmatismus Digitalen Humanismus im Projektmanagement anwenden; Ansatzpunkte je Projektphase mit konkreten Fragestellungen und Beispielen;

6. **CHRO (KI):** Auswirkungen von KI auf Arbeitsbedingungen, Arbeitsplatz und -inhalte; Frageliste für Kauf und Einsatz von KI-Lösungen; Frageliste zur Vorbereitung der Beschäftigten auf digitalhumanistischen KI-Einsatz;
7. **CHRO (Kompetenzen):** Verbindung zwischen Digitalen Humanismus und Digitalkompetenz (DigComp) -Framework der EU; Maßnahmen zur Verankerung des Digitalen Humanismus in Aus- und Weiterbildung, sowie zur Etablierung in der eigenen Führung und Kultur.

Der **zweite Teil des Buches** (Beiträge 8–13) geht auf **spezifische Fragestellungen ausgewählter Branchen** bei der Umsetzung des Digitalen Humanismus ein und spricht daher Leser:innen an, die entweder in dieser Branche tätig sind oder die von anderen Branchen lernen wollen:

8. **Versicherungen:** konkrete Anregungen und Beispiele zur menschenzentrierten Digitalisierung in einem Versicherungsunternehmen; Zusammenspiel Mensch und Technologie in der Kundeninteraktion; regulatorische Anforderungen;
9. **Fallbeispiel Tatra banka Slovakia:** wie Vertrauen und Sicherheit für Kunden durch außergewöhnliches Projekt verbessert wurde; Beispiel für ganzheitliche, menschenzentrierte Sichtweise zur Problemlösung; messbare Reduktion von Fraud erreicht;
10. **Medien (Rundfunk):** Zusammenhang zwischen Unternehmenswerten, Ethik Kodex und Digitalisierung; besondere Rolle der Medien; Beitrag zur digitalen Kompetenz; niederschwelliger Zugang zu Informationen; ausgezeichnetes Praxisbeispiel: AiDitor; KI-gestütztes Musik-Fingerprinting;
11. **Medien (Presse):** Entwicklung KI-Leitlinien für Presseagentur; das Prinzip Check-Recheck-Doublecheck im Zeitalter der KI; Ethik Board und Governance; User Needs ins Zentrum stellen (mit Projektbeispiel);
12. **Medizin:** Erwartungen der Ärzt:innen und Patient:innen; Rollenverteilung Ärzt:innen und KI; politische und rechtliche Rahmenbedingungen und Vorgaben u. a. eHealth Strategie, Helsinki Deklaration; Forderungen des Bio-Ethikrates; Erfolgsbeispiele und Nutzen der Digitalisierung; ethischer Imperativ für Manager:innen im Gesundheitswesen;
13. **Informations- und Kommunikationsindustrie:** Umgang eines IKT-Unternehmens mit KI im Allgemeinen und im Speziellen im Bereich Mitarbeiter und Kunden um; Erstellung der Die A1 Fair AI-Guidelines; Digitaler Humanismus und Vertrauen als Basis des langfristigen Unternehmenserfolgs.

Der **dritte Teil des Buches** (Beiträge 14–19) beschäftigt sich mit **Methoden, Tools und Standards**, die zur Umsetzung des Digitalen Humanismus geeignet sind bzw. spezifisch dafür entwickelt wurden und zeigt in Fallbeispielen deren Praxiseinsatz auf:

14. **Methodenüberblick:** Überblick zu Methoden und Standards für digitalhumanistische Organisations- und IT-Entwicklung, gesamtheitliche Unternehmensentwicklung, sowie KI spezifische Methoden; Beschreibung von über 10 Methoden als Einstiegspunkt für Auswahl und Vertiefung;
15. **Value-based engineering (VBE) Fallbeispiel ibis acam:** Darstellung und Anwendungstipps der Methode des Value Based Engineerings; Beschreibung eines der ersten Umsetzungsprojekte der Methode mit Erkenntnissen und Beispielen aus dem Projekt;
16. **System-Theoretic Process Analysis (STPA):** Beschreibung der STPA (Methode aus dem Hochsicherheitsbereich); Evaluierung von Folgewirkungen einer Lösung mithilfe von STPA; Einsatz der STPA im Rahmen des VBE; Fallbeispiel;
17. **Corporate Digital Responsibility (CDR) und das Fallbeispiel Merck:** Überblick CDR Managementsystem und Rollen; Beschreibung der systematischen Herangehensweise bei Merck; Digitalethikodex – Ethikboard – Risikobewertung; mehrere Beispiele u. a. Merck Digital Ethics Check;
18. **Fallbeispiel – Prüfkatalog für KI:** Prüfkatalog und sein Einsatz vor- und während der Projekte; Modell zur Operationalisierung der KI Ethik; Praxisbeispiel: Projekt zur Bias-Mitigation; Nutzenbetrachtung;
19. **Kybernetik:** Anwendung des kybernetischen Steuerungsmodells zur Umsetzung des Digitalen Humanismus; systemische Fragetechniken zur Vermeidung kognitiver Verzerrungen; Einsatz von Transformationsmodellen.

Der **vierte Teil des Buches** (Beiträge 20–26) prüft die aktuelle Unterstützung des Digitalen Humanismus in Unternehmen durch Rahmenbedingungen, Förderungen, Recht und Wissenschaft:

20. **Wertebasiertes Management:** wachsende Bedeutung ethischer, sozialer und ökologischer Faktoren über Design-, Effizienz- und Funktionsmaximierung hinaus; Integration von Stakeholderinteressen; Notwendigkeit eines wertebasierten Managements zur Bewältigung komplexer Herausforderungen;
21. **Innovationsstrategie:** Digitaler Humanismus als Instrument der Innovationsstrategie; Balance zwischen unternehmerischer Effizienz und gesellschaftlicher Verantwortung; Regulierung und Ethik als Innovationsförderer; Risikominimierung durch wertebasierte Technologiegestaltung;
22. **EU AI Act Umsetzung:** den AI Act als Chance nutzen; Kombination aus Compliance und ethischer Gestaltung; der AI Leadership Compass als praxisorientierte Orientierung; digitale Souveränität und nachhaltige Wettbewerbsfähigkeit stärken;

23. **Öffentliche Beschaffung:** Ethische und innovative Vergabestrategien nutzen; neue Wege in der Beschaffung zu Verankerung der humanistischen Werte und als Treiber für die Umsetzung in Unternehmen;
24. **Jenseits von Human-in-the-Loop:** Darstellung der Grenzen und Risiken des Human-in-the-Loop Ansatzes; Vorstellung eines alternativen Konzeptes zur Erfüllung der Menschenzentrierung durch partizipative Methoden wie Co-Creation zur Förderung der Akzeptanz neuer Technologien und Ermöglichung einer breiten Einbindung von Perspektiven;
25. **Digitaler Humanismus in der Wirtschaftsförderung:** Menschenzentrierte Digitalisierung als Innovationsfaktor für Unternehmen; Best Practices zeigen vielfältige Ansätze von barrierefreier Kommunikation bis zu ethischer KI; gezielte Strategien und unternehmerische Bewusstseinsbildung
26. **Der Wiener Weg zum Digitalen Humanismus:** Interdisziplinäre Forschung als Schlüssel zur digitalen Gestaltung; Wien als Modellstadt für eine sozial verantwortliche Digitalisierung; langfristige Förderschiene des wwtf – Wiener Wissenschafts-, Forschungs- und Technologiefonds.

Danksagung

Ein Buch ist immer ein Gemeinschaftswerk und es gibt viele Menschen, die durch ihren Einsatz und ihre Motivation zum Gelingen dieses Werkes beigetragen haben:

Danke an alle Autor:innen, die sich für den Digitalen Humanismus, der für uns seit Jahren ein großes Anliegen darstellt, in ihrem Wirkungsbereich einsetzen und bereit waren ihre Erfahrungen in einem Beitrag zu teilen, um andere zu inspirieren, zu eigenen Aktivitäten zu motivieren und sich einfach mit dem Thema zu beschäftigen.

Ein großer Dank geht an das Team des Springer Verlags rund um Fr. Treitschke, die uns nicht nur zur Herausgabe dieses Buches motiviert haben, sondern uns inhaltlich und organisatorisch während des gesamten Herstellungsprozesses hervorragend unterstützt haben.

Nicht zuletzt ein großes Dankeschön an Fr. Monika Zvonic, die uns beide in allen organisatorischen Belangen der Buchherstellung, der Einholung und Durchsicht von Beiträgen und Autor*innenverträgen unterstützt hat.

Wir wünschen viel Freude und Inspiration und vor allem Erfolg bei der Umsetzung im eigenen Wirkungsbereich. Wenn Sie eigene Umsetzungsbeispiele publizieren möchten, freuen wir uns über eine Kontaktaufnahme: herausgeber@digitalerhumanismus.business.

Wien
im Mai 2025

Georg Krause
Martin Giesswein

Inhaltsverzeichnis

Teil I Umsetzung digitalhumanistischer Werte aus Unternehmenssicht

1	CEO-Perspektive: Warum und wie sollte man Digitalen Humanismus im Unternehmen umsetzen	3
	Georg Krause	
2	CFO-Perspektive: Projekterfolge durch Einbeziehung von Werten sicherstellen	19
	Leopold Fischl	
3	CIO-Perspektive 1: Wiener (digitale) Lebensqualität	29
	Klemens Himpele	
4	CIO-Perspektive 2: Nachhaltige Datensouveränität	39
	Helmut Leopold und Lisa Höllbacher	
5	CIO-Perspektive 3: Digitaler Humanismus im Projektmanagement	49
	Mario Tokarz	
6	CHRO-Perspektive: Künstliche Intelligenz im Personalmanagement	63
	Tünde Fülöp	
7	Digitale Kompetenzen zur Förderung des Digitalen Humanismus	75
	Alexander Schmölz	

Teil II Branchenspezifische Herausforderungen und Lösungsansätze

8	Ethische Digitalisierung in der Versicherungsbranche	85
	Gerhard Lahner und Thomas Mara	
9	Vertrauen schaffen und Sicherheit geben im Digitalen Banking	93
	Natalia Major	
10	Digitaler Wandel im ORF: Leichter Leben durch Technologie	101
	Astrid Zöchling	
11	Nutzerzentrierte Innovationsstrategien aus der Perspektive der APA . . .	113
	Verena Krawarik	
12	Digitaler Humanismus im Gesundheitswesen	127
	Siegfried Meryn	
13	Digitaler Humanismus bei A1	137
	Michael Höfler	

Teil III Methoden, Tools und Standards

14	Methoden zur Umsetzung des Digitalen Humanismus	147
	Martin Giesswein	
15	Value Based Engineering und die Umsetzung bei ibis acam	159
	Svenja Schröder und Florian Wurzer	
16	Mit der System-Theoretic Process Analysis (STPA) Zusammenhänge verstehen und Werte schützen	173
	Florian Wagner	
17	Corporate Digital Responsibility als ganzheitlicher Ansatz für Digitalen Humanismus	183
	Karin Dietl, Simon Lucas und Jean Enno Charton	
18	Der BRZ-Prüfkatalog für vertrauenswürdige KI	201
	Wolfgang Kabelka	
19	Kybernetische Steuerungsmodelle zur Umsetzung des Digitalen Humanismus	215
	Armin Preis	

Teil IV Rahmenbedingungen, Förderungen und andere Perspektiven

20 Willkommen im Zeitalter des wertebasierten Managements 239
Sarah Spiekermann, Monika Zvonic und Martin Giesswein

21 Digitaler Humanismus als Innovationsstrategie 247
Erich Prem und Hannes Werthner

22 Der EU AI Act: Aus der Pflicht eine Tugend machen 261
Stefan Tauber und Martin Giesswein

23 Technologie im Dienst der Gesellschaft 275
Martin Schiefer und Maria Troger

24 Digitaler Humanismus jenseits von „Human-in-the-Loop“ 285
Peter Parycek, Diogo Sasdelli und Noella Edelmann

25 Digitaler Humanismus in der Wiener Wirtschaftsförderung 301
Kristina Maurer und Eva Czernohorszky

26 Der Wiener Weg zum Digitalen Humanismus 313
Michael Stampfer

Abbildungsverzeichnis

Abb. 1.1	Wertekompass der msg Plaut	10
Abb. 1.2	Beschreibung der Werte	11
Abb. 1.3	Digitalisierung erzeugt neue Handlungsfelder für Unternehmensverantwortung. (Quelle: Lautermann & Frick, 2023)	12
Abb. 1.4	Ethik Checkliste	15
Abb. 7.1	Digitaler Humanismus in DigComp 2.2 AT und 2.3 AT	76
Abb. 7.2	Digitaler Humanismus in DigComp 2.2 AT und 2.3 AT Teil II	77
Abb. 8.1	Investments Cyber Security. (Quelle: Gartner Inc., 2025. GARTNER is a trademark of Gartner, Inc. and/or its affiliates)	87
Abb. 8.2	Business Continuity (Quelle: Gartner Inc., 2024. GARTNER is a trademark of Gartner, Inc. and/or its affiliates)	89
Abb. 9.1	Gesamtschäden der Kunden ab 2021 vor und nach Einsetzung eines bereichsübergreifenden Teams. (Quelle: Tatra banka Slowakei)	95
Abb. 9.2	Bildungsvideo in der mobilen App	97
Abb. 9.3	Push-Benachrichtigungen	97
Abb. 9.4	Ereignisgesteuerte Kundenbildungskampagnen im Jahr 2024 Teil 1	98
Abb. 9.5	Ereignisgesteuerte Kundenbildungskampagnen im Jahr 2024 Teil 2	98
Abb. 9.6	Externe Kommunikation	99
Abb. 9.7	Entwicklung von Kundenverlusten und verhinderten Beträgen	100
Abb. 10.1	Ethische Herausforderungen durch GenKI	104
Abb. 11.1	Überblick über die Kernelemente der Trusted AI-Strategie APA	117
Abb. 11.2	User Needs mit Leitfragen	122
Abb. 15.1	VBE Prozess	161
Abb. 15.2	Abgrenzung des Gesamtsystems (SOS) vom System of Interest (SOI)	167
Abb. 16.1	„Control Structure“ des Systems und dem Systemkontexts	178
Abb. 16.2	Argumentative Kette für die Begründung und Bewertung von Systemanforderungen und deren Nutzen zum Schutz von Werten	180
Abb. 18.1	Übersicht teilnehmende Organisationen pro Land	203
Abb. 18.2	Sektorale Übersicht der Organisationen	204
Abb. 18.3	Übersicht nach Organisationstyp	205
Abb. 18.4	Bewertungsregeln für Prüfkatalog vertrauenswürdige KI	209

Abb. 18.5	Vorurteile vorprogrammiert: Analyse Herausforderungen	210
Abb. 19.1	Ausprägungen der kybernetischen Steuerungszyklen in Anlehnung an Schwaiger. (Quelle des Cybernetic Control Cubes: Schwaiger, 2012) . .	219
Abb. 19.2	The General Model of Systemic Control. (Quelle: Schwaninger, 2001) .	220
Abb. 19.3	The Recursive Cybernetic Management Model (CRMM). (Quelle: Preis, 2013)	221
Abb. 19.4	Value Ontology	223
Abb. 19.5	Kybernetischer Regelkreis für ethisches Systemdesign	223
Abb. 19.6	Holistisches Transformationsmodell	228
Abb. 19.7	Generisches rekursives Metamodell zur Wissensgenerierung	230
Abb. 19.8	Kombination der Theory U mit dem rekursiven Metamodell für Wissensgenerierung. (Quelle für das Theory U Modell: Scharmer, 2009, Hinz und Garz, o. J.)	231
Abb. 19.9	Fragestellung im Holistischen Transformationsmodell im Sinne der Kybernetik zweiter Ordnung	232
Abb. 21.1	Business Models Canvas	256
Abb. 21.2	Erweiterung Business Models Canvas. (Quelle: nach Prem, 2015)	257

Tabellenverzeichnis

Tab. 11.1	User Needs im Überblick. (Quelle: eigene Darstellung nach Whitepaper Smartocto, 2023, User Needs 2.0)	121
Tab. 15.1	Leitfragen für den Workshop	163
Tab. 15.2	Ergebnisse aus dem Workshop zu Utilitarismus	169
Tab. 16.1	Schlagwörter (Keywords) für die UCA-Identifizierung	175
Tab. 16.2	Gegenüberstellung von Definitionen aus der ISO/IEC/IEEE 24748-7000:2022 und der STPA	176
Tab. 16.3	Auszug identifizierter „Losses“ für das System „Lade-App“	178
Tab. 16.4	Auszug identifizierter „Hazards“	178
Tab. 16.5	Auszug identifizierter „UCAs“	179
Tab. 16.6	Auszug identifizierter „Loss Scenarios“	179
Tab. 16.7	Wertedemonstratoren abgeleitet aus „Loss Scenarios“	180

Herausgeber und Autorinnen/Autoren

Über die Herausgeber



Dr. Georg Krause ist CEO der IT-Beratung msg Plaut AG, die in Österreich, Zentraleuropa und Zentralasien aktiv ist. 2023 wurde er als einer der 1000 einflussreichsten Manager Österreichs ausgezeichnet. Seine Karriere führte ihn über Horváth & Partners, Deloitte und SAP – stets mit dem Fokus auf Innovation und digitale Transformation. Als Lehrender an der FH Technikum Academy und der Donau-Universität Krems bringt er Theorie und Praxis zusammen. Darüber hinaus engagiert er sich als Keynote-Speaker, Buchherausgeber, Podcast-Host und Mitglied im Verband fit4internet für digitale Kompetenzen und eine menschenzentrierte Digitalisierung. Sein Buch und Podcast „Digitaler Humanismus in der Praxis“ sind Ausdruck dieses Engagements.



Martin Giesswein ist Fakultätsmitglied der WU Executive Academy mit den Schwerpunkten KI, Digitalökonomie und Leadership. Er ist Autor und Herausgeber mehrerer Bücher zu den Themen Künstliche Intelligenz, Digital Leadership und Zukunftsgestaltung. 2025 wurde er mit einem ersten Preis des Awards „Digitaler Humanismus in der Praxis“ ausgezeichnet. Er ist als AI Ethics Advisor Mitglied des Advisory Committee im Projekt DRAIGON. Martin Giesswein ist einer der Mitinitiatoren der DigitalCity.Wien, einer seit über 10 Jahren bestehenden Kooperation der Digitalwirtschaft mit der Stadt Wien. Zuvor war er CEO und General Manager in der IKT-Branche.

Die Autorinnen und Autoren

Jean Enno Charton Merck KGaA, Darmstadt, Deutschland

Eva Czernohorszky Wirtschaftsagentur, Wien, Österreich

Karin Dietl VALUE DESIGN, Wien, Österreich

Noella Edelmann Danube University Krems, Krems, Österreich

Leopold Fischl ibis acam, Wien, Österreich

Tünde Fülöp Arbeiterkammer, Wien, Österreich

Martin Giesswein WU Executive Academy, Wien, Österreich

Klemens Himpele Magistratsdirektion Stadt Wien, Wien, Österreich

Michael Höfler A1 Telekom Austria Group, Wien, Österreich

Lisa Höllbacher Unbuzz Consulting, Wien, Österreich

Wolfgang Kabelka BRZ, Wien, Österreich

Georg Krause msg Plaut, Wien, Österreich

Verena Krawarik APA, Wien, Österreich

Gerhard Lahner Vienna Insurance Group, Wien, Österreich

Helmut Leopold AIT Austrian Institute of Technology, Wien, Österreich

Simon Lucas Merck KGaA, Darmstadt, Deutschland

Natalia Major Tatra banka Slowakei, Bratislava, Slowakei

Thomas Mara Vienna Insurance Group, Wien, Österreich

Kristina Maurer Wirtschaftsagentur, Wien, Österreich

Siegfried Meryn Future Health Lab, Wien, Österreich

Peter Parycek Fraunhofer-Institut, Berlin, Deutschland

Armin Preis ÖBB, Wien, Österreich

Erich Prem Verein zur Förderung des Digitalen Humanismus, Wien, Österreich

Diogo Sasdelli Danube University Krems, Krems, Österreich

Martin Schiefer Schiefer Rechtsanwälte GmbH, Wien, Österreich

Alexander Schmölz FH des BFI Wien, Wien, Österreich

Svenja Schröder msg Plaut, Wien, Österreich

Sarah Spiekermann WU Wien, Wien, Österreich

Michael Stampfer WWTF, Wien, Österreich

Stefan Tauber Schiefer Rechtsanwälte GmbH, Wien, Österreich

Mario Tokarz RightMinded AI, München, Deutschland

Maria Troger Schiefer Rechtsanwälte GmbH, Wien, Österreich

Florian Wagner msg Plaut, Wien, Österreich

Hannes Werthner TU Wien, Wien, Österreich

Florian Wurzer msg Plaut, Wien, Österreich

Astrid Zöchling ORF, Wien, Österreich

Monika Zvonic msg Plaut, Wien, Österreich

Teil I

Umsetzung digitalhumanistischer Werte aus Unternehmenssicht

CEO-Perspektive: Warum und wie sollte man Digitalen Humanismus im Unternehmen umsetzen

1

Ein Leitfaden zur Implementierung ethischer Prinzipien in der digitalen Transformation

Georg Krause

Zusammenfassung

In diesem Beitrag wird dargestellt, wie die Umsetzung des Digitalen Humanismus in der eigenen Organisation gelingen kann. Wichtigste Voraussetzung dazu ist Klarheit über den Nutzen und Mehrwert für die eigene Organisation zu erlangen. Der Beitrag führt dazu konkrete Beispiele auf, in welchen Bereichen messbarer Nutzen erreicht werden kann. Weiters geht er darauf ein, wie es gelingt, einen Wertekompass als Messlatte für die eigenen Aktivitäten zu definieren und zeigt auf, wie ein Top-Down-Ansatz und ein Bottom-Up Ansatz für die Umsetzung erfolgen kann und welche Verantwortlichkeiten für den erfolgreichen Einsatz zu definieren sind. In dem Beitrag werden zu allen Themen Praxistipps aus der eigenen Umsetzung dargestellt und abschließend wesentliche Lessons Learned aufgezeigt.

Als Führungskräfte müssen wir jetzt die Weichen stellen, damit sich die Digitalisierung zum Vorteil der Menschen entwickelt und nicht zu einer Bedrohung wird.

1.1 Relevanz und Motivation

„Das Internet ist nicht kaputt – es ist, wie es ist, und als solches haben wir es tief in unsere Lebenswelt eindringen lassen. Dabei ist es allerdings mit dem Welpenschutz, in dessen Genuss alles Digitale lange gekommen war, endgültig vorbei – es ist ernst geworden mit dem

G. Krause (✉)
msg Plaut, Wien, Österreich
E-Mail: georg.krause@msg-plaut.com

digitalen Wandel. Seine Konsequenzen vorwiegend mit ökonomischer Brille zu betrachten wäre fatal – es ist eine zutiefst politische Frage.“ (Reichl, 2020)

Dieses Zitat, das eine Replik auf die Schlagzeile der Titelseite der Times „The Internet is broken“ vom Jänner 2018 darstellt, bringt die aktuellen Herausforderungen im Zusammenhang mit der zunehmenden digitalen Durchdringung unserer Welt sehr gut auf den Punkt.

Es gibt kaum einen Bereich in der Wirtschaft, im Gesundheitswesen, in der öffentlichen Verwaltung, sowie in praktisch allen privaten Belangen – vom Einkauf bis zur Freizeitgestaltung von der Informationsbeschaffung bis zur Mobilität – der nicht von der Digitalisierung erfasst und oft maßgeblich verändert wurde. Diese Veränderung hat auch immer direkte oder zumindest indirekte Auswirkungen auf uns und unser Leben.

Dieser Beitrag richtet sich an Führungskräfte und Manager, die als Leiter von Organisationen Verantwortung für die Ausgestaltung digitaler Lösungen – egal ob intern oder für Kunden – haben und damit auch Verantwortung übernehmen, wie unser aller Leben morgen aussehen wird.

Dabei geht es nicht darum, den Einsatz von Fake News oder süchtig machenden Algorithmen zu vermeiden. Das dürfen wir als selbstverständlich voraussetzen. Es geht vielmehr um die Schaffung von Bewusstsein für die ethische Dimension bei der Umsetzung digitaler Lösungen. Um durch die Beachtung von deren Auswirkungen, einen, wenn auch manchmal kleinen Beitrag, für eine digitale Welt zu leisten, die die positiven Potenziale zur Verbesserung des Lebens der Menschen nutzt und gleichzeitig negative Auswirkungen und Risiken mitigierte.

Wie dies gelingen kann und welchen Mehrwert ein Management, das Digital-Humanistische Prinzipien in seinen Überlegungen berücksichtigt, für die jeweilige Organisation hat, ist der Gegenstand der folgenden Ausführungen.

1.2 Nutzen schaffen – start with why

„Digital Humanism is directed at strengthening human responsibility, at realizing the potentials of digitalization that relief the burden of unnecessary knowledge and calculations in order to give the people the possibility to concentrate on what is essential and contribute to a more humane and just future for humanity.“ (Nida-Rümelin, 2022)

Die erste und wichtigste Frage bei der Auseinandersetzung mit dem Thema des Digitalen Humanismus ist die Frage nach dem **Warum**. Warum soll sich ein Unternehmen, eine Organisation mit diesem Thema beschäftigen? Nur wenn hier ein klares Verständnis geschaffen werden kann, wird die Energie in der Organisation aufgebracht werden können, um die Umsetzung voranzutreiben und allfällige Schwierigkeiten bei der Umsetzung zu lösen.

Dabei ist die ethische Dimension relativ leicht zu beantworten, da es normalerweise außer Frage steht, dass Digitalisierung einen wesentlichen Beitrag zu einer besseren Ge-

sellschaft leisten kann und soll. So wie es im Eingangszitat von Nida-Rümelin, einem Vorreiter des Digitalen Humanismus klar ausgedrückt wird.

Spezifischer muss man hingegen auf die beiden folgenden Fragestellungen eingehen, die oft am Anfang bei der Beschäftigung mit dem Digitalen Humanismus auftauchen:

1. „Wir sind uns der Werte unserer Organisation bewusst (haben diese vielleicht auch definiert bzw. in einem Whitebook festgeschrieben) und handeln ohnehin danach. Wozu brauchen wir darüber hinaus den Digitalen Humanismus?“
2. „Das ist wieder ein Mehraufwand für unsere Organisation, dessen ethisch moralische Zielsetzung wir zwar teilen, der für uns im Wettbewerb aber Nachteile bringt und daher können wir eine Umsetzung nicht rechtfertigen.“

Ad 1 Zum ersten Thema lässt sich feststellen, dass Digitale Lösungen oft weitreichende Auswirkungen auf das Leben von Menschen haben, die nur mit einer systematischen Betrachtungsweise erkannt und beachtet werden können (Stichwort Wirkungsfolgenabschätzung). Dies erfordert eigene Prozessschritte und Herangehensweisen, die üblicherweise in Unternehmen nicht verankert sind, auch bei Unternehmen, die sich mit Werten und ethischen Fragestellungen beschäftigt haben.

Ein einfaches Beispiel ist die Fragestellung bei der Erhebung von Daten und dem nachfolgenden Umgang mit diesen bei einem Portal zur Registrierung von Benutzern. Hier geht es um die gezielte Abwägung zwischen Firmeninteressen, Privatsphäre und (Entscheidungs-)Autonomie des Benutzers und die damit ausgelöste Wirkung auf Kundenbeziehung, Image und Vertrauen. Dieses abzuwägen und die Auswirkung auf die unterschiedlichen Bereiche und Stakeholder und die möglichen sich daraus ergebenden (Reputations-) Risiken abzuschätzen, erfordert Bewusstsein in der Breite der Organisation und eine Ausbildung, damit es nicht nur rechtzeitig erkannt wird, sondern auch systematisch analysiert und entschieden werden kann. Diese Entscheidungen werden bei großen Organisationen oft auf operativer Ebene getroffen, werden vom Top Management nicht wahrgenommen, können aber massive Auswirkungen (insbesondere Vertrauensverlust) für das gesamte Unternehmen haben, wie zahlreiche Fälle in den Medien in den letzten Jahren gezeigt haben. Und dies passiert, selbst wenn alle Datenschutzgesetze eingehalten werden.

Genauso wesentlich, wie die interne Ausbildung und Bewusstmachung zur Mitigation dieser Risiken ist die Einbindung unterschiedlicher Stakeholder (z. B. Vertreter*innen unterschiedlicher Kundengruppen, aber auch indirekt betroffene Personen). Diese sollten gerade in den frühen Phasen bei der Gestaltung neuer digitaler Lösungen und Services eingebunden werden, um die Wirkung und die Folgen besser abschätzen zu können.

Ad 2 Das zweite Thema stellt die zentrale Frage nach dem Mehrwert für das Unternehmen, die Organisation. Dies ist eine der zentralen Fragen, da die Umsetzung ethischer Ziele naturgemäß wesentlich effektiver erfolgen wird, wenn auch ein klarer Nutzen für die Organisation dadurch erreicht werden kann und dieser auch den Entscheidungsträger*innen bekannt ist und von ihnen mitgetragen wird.

Durch die Umsetzung der Prinzipien des Digitalen Humanismus gibt es wesentliche Bereiche, in denen Mehrwert für das Unternehmen oder eine Organisation geschaffen werden können. Einige davon werden im Folgenden beispielhaft und ohne Anspruch auf Vollständigkeit nach Stakeholdern differenziert betrachtet:

1. **Kund:innen:** Gewinnung und Bindung

a. Vertrauenszuwachs

Die Selbstverpflichtung zur Einhaltung der Prinzipien des Digitalen Humanismus führt zu einer Stärkung des Vertrauens in das Unternehmen z. B. aufgrund des Umgangs mit Kundendaten, der Autonomie, die der Nutzer über die Leistungen/Daten erhält, sowie die Sicherheit, dass keine versteckten „Fallstricke“ in die Lösungen eingebaut sind, die der Nutzer leicht übersehen könnte;

b. Imagegewinn

Durch die Kommunikation des Beitrags zu digitalen Lösungen, die sich an den Bedürfnissen und Werten der Menschen orientieren und zur Verbesserung der Lebensqualität beitragen, kann ein positives Image am Markt erzeugt werden, da das Unternehmen seine gesellschaftliche Aufgabe in dieser für viele beunruhigenden Phase der digitalen Revolution aktiv wahrnimmt.

2. **Mitarbeitende:** Gewinnung und Bindung

a. Identifikation durch Sinnstiftung

Durch die bewusste Beachtung der Wirkungen der eigenen digitalen Lösungen und Geschäftsmodelle auf die Gesellschaft, wird für die Mitarbeitenden und allfällige Bewerber:innen der gesellschaftliche Beitrag des Unternehmens und der Sinn vermittelbar. Mitarbeitende und Kandidat:innen, die sich mit diesen Werten identifizieren, werden motiviert, sich selbst an der Umsetzung und Gestaltung digitaler Lösungen zu beteiligen, die dazu beitragen, eine menschenzentrierte digitale Welt zu schaffen.

b. Bindung durch Vertrauen

Organisationen, die sich selbst zur Einhaltung der ethischen Standards in der digitalen Welt verpflichten, erzeugen auch gegenüber neuen oder bestehenden Mitarbeitenden Vertrauen, da sich diese darauf verlassen können, dass digitale Lösungen auch nicht zu Lasten der Mitarbeitenden missbräuchlich für Überwachung oder ähnliches verwendet werden.

3. Andere Stakeholder:

a. **Gesetzgeber:** Grundlage für Compliance

Durch die Beachtung der Prinzipien des Digitalen Humanismus kann davon ausgegangen werden, dass eine gute Grundlage für die Compliance mit wesentlichen Teilen der Regulatorik betreffend Digitale Leistungen und Märkte gegeben ist. Die diesbezüglichen europäischen Regelwerke (wie AI Act, Digital Services Act, Digital Markets Act, etc.) basieren nämlich in ihrer Zielsetzung auf den „Europäischen digitalen Rechten und Grundsätzen“, die sich in den Prinzipien des Digitalen Humanismus wiederfinden (Europäische Kommission, 2022).

b. **Gesellschaft:** Verbesserung der gesellschaftlichen Wahrnehmung

Insbesondere große Unternehmen und öffentliche bzw. öffentlichkeitsnahe Organisationen haben oft eine wichtige Stellung in der Gesellschaft. Es besteht eine erhöhte – nicht selten sehr kritische – Aufmerksamkeit hinsichtlich ihrer Leistungen und ihrer Geschäftsgebarung. Durch eine ethisch orientierte Unternehmensführung nimmt die Organisation bewusst ihre Rolle und ihre Verantwortung in der Gesellschaft in ihren Entscheidungen auf. Dadurch werden mögliche, meist unbeabsichtigte Reputationsschäden – beispielsweise durch Einsatz von KI-Lösungen, die wesentliche Entscheidungen ohne Letztverantwortung durch einen Menschen treffen – vermieden und es kann sogar eine positive Gestaltungs- und Vorreiterrolle zur Stärkung der eigenen Rolle in der Gesellschaft angeführt werden.

c. **Kapitalgeber*innen**

Für Kapitalgeber ist die langfristige Wertentwicklung von Unternehmen bei gleichzeitig möglichst geringen Risiken oft ein relevanter Entscheidungsfaktor bei einer Investitionsentscheidung. Durch eine Unternehmensführung, die die humanistischen Werte mit Innovationen und Weiterentwicklung in Einklang bringt, wird eine nachhaltige Entwicklung begünstigt, da zu erwarten ist, dass diese Lösungen langfristig eine höhere gesellschaftliche Akzeptanz haben. Gleichzeitig werden dadurch Risiken deutlich reduziert, da davon ausgegangen werden darf, dass auch zukünftige Regularien oder gesellschaftliche Entwicklungen nicht konträr zu diesen Prinzipien erfolgen werden.

In der Diskussion wird oft als eines der Gegenargumente gegen den Digitalen Humanismus die **Verhinderung von Innovationen** gebracht. Hier gilt es in der Umsetzung in der Organisation darauf zu achten, dass verstanden wird, dass durch Digitalen Humanismus eine nachhaltige Innovation gefördert wird. Das bedeutet, dass in der Ideenfindungs- und Entwicklungsphase neuer digitaler Lösungen und Modelle, Auswirkungen frühzeitig und von verschiedenen Perspektiven betrachtet werden, damit die Lösung dann entsprechend die positiven Potenziale realisiert und mögliche negative Auswirkungen durch (technische) Lösungen hintanhält.

Das bedeutet, dass es zwar gewisse Mehraufwände im Prozess der Entwicklung und Umsetzung innovativer Ideen entstehen, das Ergebnis dann aber besser abgesichert ist und Risiken durch negative Auswirkungen oder auch spätere Mehrkosten, weil ein Umbau erforderlich wird, reduziert werden.

Fallbeispiel msg Plaut

Die msg Plaut Austria ist Teil der 5. größten IT-Beratungsgruppe (msg) im deutschen Sprachraum. Diese hat über 11.000 Mitarbeitende, den Stammsitz in Ismaning (bei München) und befindet sich im Besitz der Gründerfamilien.

Im Hinblick auf den digitalen Humanismus ergeben sich daraus einige wesentliche positive Voraussetzungen:

- Die Gründerfamilien leben seit der Gründung eine stark werteorientierte Kultur. Vertrauen ist wesentlicher Grundpfeiler für den Erfolg, was sich einer menschenzentrierten Mitarbeitendenführung und in langjährigen Kundenbeziehungen widerspiegelt.
- Durch die Eigentümerstruktur steht eine langfristige Entwicklung des Unternehmens im Vordergrund und nicht eine kurzfristige Maximierung.
- Die Vision der Unternehmensgruppe ist, „we create ecosystems for enabling a better life“. Damit wird eine der wesentlichsten Forderungen des Digitalen Humanismus, nämlich die Nutzung der Digitalisierung zur Verbesserung des Lebens, adressiert.

Auf dieser Basis wurden folgende Ziele – beziehungsweise auf die o. a. Kategorien – für die Umsetzung des Digitalen Humanismus in der Organisation formuliert:

- **Vertrauen und Bindung von Mitarbeitenden und Kund*innen**

Der Digitale Humanismus unterstützt dabei, Werte, von denen viele bereits in der Organisation gelebt werden und in der Vision verankert sind, für Kund*innen und Mitarbeitende kommunizierbar und verständlich zu machen und durch eine systematischere Umsetzung noch besser zu verankern. Dies soll zur Positionierung des Unternehmens beitragen und das Vertrauen der Kund*innen und Mitarbeitenden in die Firma stärken und die Bindung bzw. Attraktivität erhöht werden.

- **Sinnstiftung**

Durch die Beschäftigung mit der Frage nach den Folgen und Auswirkungen der Projekte auf die Gesellschaft wurden nicht nur mögliche Risiken und Gefahren aufgezeigt, sondern auch die positiven Effekte, die realisiert werden konnten (wie z. B. der persönliche Beitrag zu weniger Toten auf Europas Straßen, zur sicheren Medikamentenversorgung, uvm. mehr Infos in „msg Plaut Heldenstory“ (Krause, 2023)). Dadurch konnte das Bewusstsein aller Mitarbeitenden über ihren Beitrag für ein besseres Leben vieler deutlich gemacht werden und der Stolz auf die eigene Arbeit gesteigert werden.

- **Mehrwert für Kund*innen**

IT-Dienstleister haben eine besondere Verantwortung für die menschenorientierte Digitalisierung, da sie im Rahmen ihrer Projekte den Kunden dabei unterstützen können, die Konsequenzen der Projekte zu beachten und Risiken zu mitigieren. Werden nämlich Lösungen eingesetzt, die diesen Werten und Standards nicht entsprechen, ist die Gefahr von Vertrauensverlust, Imageschäden und Abwanderung von Mitarbeitenden oder Kund*innen gegeben. Ziel der msg Plaut ist es daher, all-fällige Problembereiche in Projekten rechtzeitig aufzuzeigen und gemeinsam mit den Kund*innen einer Lösung zuzuführen.

- **Beitrag zur Gesellschaft**

Um das Zielbild der menschenzentrierten Digitalisierung zu erreichen, ist es wichtig, dass die Umsetzung auf breiter Basis in den Organisationen erfolgt. Trotz zahlreicher Aktivitäten und Publikationen zum Thema Digitaler Humanismus gab es zu Beginn der Beschäftigung mit dem Thema wenig konkreten Hinweise und Erfahrungen bezüglich der Umsetzung des Digitalen Humanismus in Organisationen insbesondere in der Privatwirtschaft. Daher hat es sich die msg Plaut auch zum Ziel gesetzt durch ihre diesbezüglichen Aktivitäten im Rahmen von Projekten, Vorträgen, Publikationen, Podcasts und Vernetzungsplattformen einen Beitrag zu leisten, damit das Thema breit in der Wirtschaft und Verwaltung aufgenommen und umgesetzt wird. ◀

1.3 Den eigenen Wertekompass definieren

Digitaler Humanismus ist ebenso wie der Begriff des Humanismus kein klar definierter Begriff. Umso wichtiger ist es, für die Arbeit in der eigenen Organisation Klarheit zu schaffen und die Werte und Prinzipien konkret und passend zu den Werten und Schwerpunkten der eigenen Organisation zu definieren.

Dieser Schritt ist von allergrößter Bedeutung, da die definierten Werte der Bezugsrahmen, die Messlatte für die Beurteilung sind, ob die eigenen Lösungen und Handlungen konform sind oder nicht. Gleichzeitig stellen diese Werte auch die Grundlage für die Kommunikation nach innen und außen dar und werden zum inneren Kompass der Organisation für alle Mitarbeitende für ihre Aktivitäten.

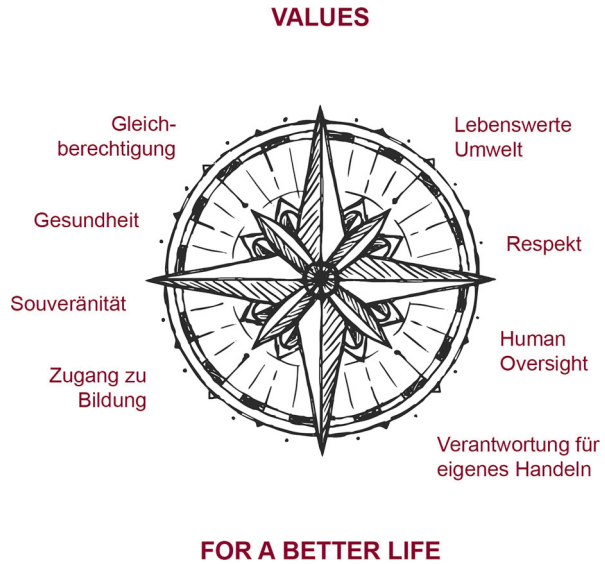
Bei der Erarbeitung der Werte ist ein breit angelegter Diskussionsprozess mit Vertretern unterschiedlicher Bereiche im Unternehmen und unterschiedlicher Managementebenen für die inhaltliche Qualität, die weitere Akzeptanz und das Verständnis hilfreich.

Als Unterstützung bei der Erarbeitung des eigenen Wertekompass im Sinne des Digitalen Humanismus können insbesondere folgende Quellen verwendet werden:

- Das Wiener Manifest für Digitalen Humanismus (Werthner et al., [2020](#))
- Die „Europäischen digitalen Rechte und Grundsätze“ (Europäisches Parlament et al., [2022](#))
- Ethik-Leitlinien für eine vertrauenswürdige KI (Europäische Kommission, [2018](#))
- IEEE: Corporate Digital Responsibility (CDR) Securing Our Digital Futures (IEEE, [2024](#))

Idealerweise sollten diese Werte mit den Werten des eigenen Unternehmens abgestimmt werden, um daraus einen organisationsspezifischen Wertekompass abzuleiten, der zur spezifischen Situation, Positionierung und dem Leistungsspektrum der Organisation passt.

Abb. 1.1 Wertekompass der msg Plaut



Fallbeispiel Wertekompass msg Plaut

Als Beispiel wird hier der Wertekompass (Top Ebene der Werte) der msg Plaut dargestellt, der aus einigen der o. a. Quellen und den Werten der msg Gruppe unter Einbindung von Management, Führungskräften und Mitarbeitenden erarbeitet und im Aufsichtsrat abgestimmt wurde und im Digitaler Humanismus Whitebook der msg Plaut dokumentiert ist (Abb. 1.1).

Jeder dieser Begriffe wird durch weitere Begriffe auf der nächsten Ebene konkretisiert, die dann jeweils kurz beschrieben sind (Abb. 1.2).

1.4 Umsetzung starten

Die Handlungsfelder, wo im Zuge der Digitalisierung ethische Fragestellungen in der Organisation auftauchen können, gibt es sehr viele und man findet sie entlang der gesamten Wertschöpfungskette bzw. gehen sie aufgrund der indirekten Wirkungen digitaler Lösungen sogar weit über diese hinaus.

Eine gute Übersicht über relevante Handlungsfelder gibt Abb. 1.3, die zeigt, dass die Verantwortlichkeit für die Einhaltung der definierten Digital-Humanistischen Werte bei der **Unternehmensleitung** beginnt, die für die Definition derselben, die Verabschiedung von Richtlinien, einer Strategie und einer Governance verantwortlich ist.

Entlang der gesamten **Wertschöpfungskette**, vom Lieferanten bis zum Kunden gilt es auf die Einhaltung zu achten. Das beginnt beispielsweise bei der Beschaffung einer



Abb. 1.2 Beschreibung der Werte

KI-Software mit deren Überprüfung, ob sie Prinzipien wie Transparenz und Erklärbarkeit abbilden kann, über die Anwendung und das Training dieser KI unter Achtung auf nicht-diskriminierende Trainingsdaten und schließlich der Bereitstellung für die Kund:innen unter Wahrung der Autonomie der Kunden bei Entscheidungen im Umgang mit ihren Daten.

Es geht aber nicht nur um die eigene Leistungserstellung, sondern auch um ethische Fragestellungen bezüglich der Kunden. Welche Kunden werden mit den digitalen Lösun-

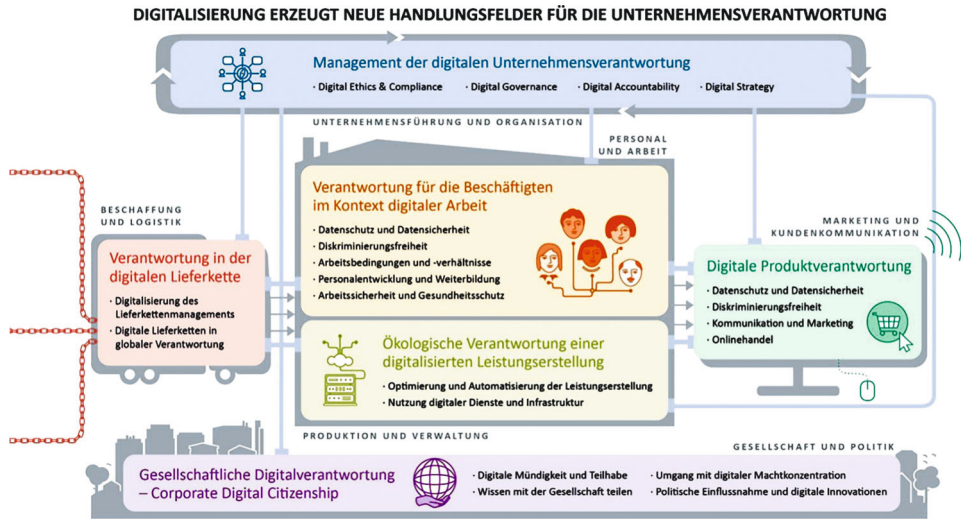


Abb. 1.3 Digitalisierung erzeugt neue Handlungsfelder für Unternehmensverantwortung. (Quelle: Lautermann & Frick, 2023)

gen der Organisation beliefert und ergeben sich möglicherweise aus dem Tätigkeitsfeld des Kunden Widersprüche zu den definierten Werten. Weiters ist auch zu hinterfragen, wie die Lösungen vom Kunden eingesetzt werden (könnten) und welcher Missbrauch bzw. welche Risiken sich daraus ergeben.

Im Rahmen der Leistungserstellung gilt es auch auf die Einhaltung der Werte im Kontext der **digitalen Arbeit** zu achten, also auf Themen, wie Schutz der Privatsphäre der Mitarbeitenden und Nutzung der digitalen Möglichkeiten zur Gestaltung effektiver und gesundheitsunterstützender Arbeitsplätze (z. B. Mensch-Maschine Kooperation bei schweren Arbeiten).

Die **gesellschaftliche Verantwortung** bezieht sich auf die Wirkungen und das Zusammenspiel der digitalen Aktivitäten der Organisation mit der Gesellschaft, also ob durch eigene digitale Aktivitäten, wie neue Geschäftsmodelle, negative Konsequenzen (z. B. prekäre Arbeitsverhältnisse) ausgelöst werden, aber auch, wie durch Digitalisierung positive Beiträge für die Gesellschaft geschaffen werden können. Ein Beispiel sind die Aktivitäten der A1 Telekom in Österreich, die zur Stärkung der digitalen Kompetenzen einen „A1 digital.campus“ ins Leben gerufen haben und in den letzten Jahren jährlich fast 50.000 Österreicher*innen digitale Ausbildung ermöglichen (Arnoldner, 2023).

Daraus ergibt sich, dass es praktisch keinen Bereich im Unternehmen gibt, der nicht von diesen Fragestellungen betroffen ist. Es ist also kein Thema, für das die IT allein die Verantwortung übernimmt, sondern es muss gewährleistet werden, dass alle Beteiligten entlang der Wertschöpfungskette, aber auch alle Servicebereiche (wie HR, Marketing, Finanzen) eingebunden werden.

Aufgrund des Umfangs der möglichen Themenfelder im Unternehmen und der vielen betroffenen Bereiche stellt sich die Frage, wo man beginnen soll, wenn man die Prinzipien des Digitalen Humanismus im Unternehmen umsetzen und verankern will.

Dabei gibt es folgende grundsätzliche Herangehensweisen:

- **Gesamtheitlicher Ansatz (Top-Down):** Man beginnt mit der Umsetzung auf Gesamtunternehmensebene durch Entwicklung der Zielvorstellungen (Warum?), von Guidelines (Wie?), einer Governance Struktur (Wer? z. B. Ethik Board) etc. und tritt frühzeitig in die Kommunikation und den Dialog mit den Mitarbeitenden, Kunden und anderen Stakeholdern ein. Aus der Beschäftigung mit dem Thema leitet man die priorisierten Schwerpunktfelder und -themen, sowie die für das Unternehmen am besten passende Vorgehensweise ab und beginnt diese gezielt zu implementieren.
- **Pilotierung (Bottom-Up):** Man greift ein konkretes Projekt/Entwicklungsvorhaben oder eine Abteilung heraus und beginnt die Umsetzung durch die Auswahl verwendeter Standards, Erweiterung der bestehenden Prozesse um die Anforderungen zur Beurteilung der ethischen Auswirkungen, Training der involvierten Mitarbeitenden etc. Die Erfahrungen dieses Piloten werden dann genutzt, um festzulegen, wie eine Ausbreitung im Unternehmen erfolgen kann.

Fallbeispiel Ansatz msg Plaut

In der msg Plaut wurde der Gesamtheitliche Ansatz verfolgt. In den ersten Monaten erfolgte eine intensive Auseinandersetzung mit dem Thema Digitaler Humanismus im Management und einem Kreis, der sich aus Vertretern der verschiedensten Bereiche und Hierarchiestufen des Unternehmens zusammensetzt (den sogenannten „Ambassadors“, also den Botschaftern für Digitalen Humanismus).

Dabei wurden die Fragen geklärt, warum ist dieses Thema für die Firma wichtig, wie ist die Resonanz bei Mitarbeitenden¹, was sind die Ziele, die damit erreicht werden sollen und letztlich, wie kann die Umsetzung gestaltet werden. Dann erfolgte die Entscheidung für die Anwendung des IEEE7000 Standards für Value Based Engineering, die Ausbildung von Mitarbeitenden in dieser Methode und letztlich die Auswahl eines ersten Umsetzungsprojekts in Abstimmung mit dem Kunden, der dafür gewonnen werden konnte, diese innovative Methode auszuprobieren und umzusetzen².

1.5 Verantwortlichkeiten klären – Governance

Hinsichtlich der Frage nach der Verantwortlichkeit für die Einhaltung der definierten Werte im Sinne des Digitalen Humanismus gibt es folgende wesentliche Themenstellungen, über die man rasch Klarheit schaffen sollte:

¹ Anmerkung: das Management Team war überrascht, wie stark das Interesse an und die Unterstützung für das Thema war.

² Ein Erfahrungsbericht zu dieser Methode und dem Pilotprojekt folgt in einem späteren Kapitel.

1. Verantwortung der obersten Führungsebene

Der obersten Entscheidungsebene kommt eine Schlüsselrolle bei der Umsetzung der Werte und Prinzipien zu. Erstens, da derartig grundlegende Themen, wie Ethik und Werte nur durch das Commitment der Führung unternehmensweite Gültigkeit und Bedeutung erlangen und zweitens, weil es entscheidend für die Glaubwürdigkeit ist, dass wesentliche Entscheidungen in sorgfältiger Abwägung zwischen wirtschaftlichen Interessen und Ethischem Anspruch getroffen werden und entsprechend auch von der Leitung unterstützt werden. Werden derartige Entscheidungen von der Leitung immer durch rein wirtschaftliche Überlegungen übersteuert, verliert das ganze Vorhaben intern und extern jede Glaubwürdigkeit.

2. Verantwortung der Mitarbeitenden und Führungskräfte

Oft sind es gerade die kleinen, alltäglichen Entscheidungen, z. B. bei der Entwicklung einer Oberfläche für eine Kundenapplikation oder das Training eines KI Chatbots, die darüber entscheiden, ob die angestrebten Werte auch eingehalten werden. Es ist daher unabdingbar, dass alle Mitarbeitenden und Führungskräfte über die Werte, aber auch über allfällige bereit gestellte Tools, wie Checklisten, Guidelines etc., informiert sind, damit sie sich dieser in der täglichen Arbeit bewusst sind und bei ihren Entscheidungen berücksichtigen können.

3. Verantwortung des Ethik Boards

Eine zentrale Rolle in der Umsetzung kommt dem Ethik Board zu, da ethische Fragestellungen sehr oft nicht eindeutig zu beantworten sind und von verschiedenen Perspektiven (z. B. Auswirkungen auf Mitarbeitende, auf Kunden, auf Gesellschaft, auf Kunden oder Mitarbeitende mit Beeinträchtigungen, etc.) betrachtet werden sollen. Daher ist eine frühzeitige Einrichtung eines Ethik Boards, das sich aus Vertretern unterschiedlicher Bereiche intern und idealerweise externer Unterstützung/Begleitung zusammensetzt. Das Ethik Board hat eine beratende Aufgabe, da die Verantwortung für die Entscheidung bei den Mitarbeitenden und Führungskräften, verbleiben sollte. Es bietet aber eine wesentliche Unterstützung durch die strukturierte Auseinandersetzung mit dem Thema und durch die Erfahrung aus bisherigen Entscheidungen.

Fallbeispiel Commitment msg Plaut

In der msg Plaut wurde die Umsetzung des Digitalen Humanismus durch den Vorstand und die Geschäftsführung mit Information/Abstimmung des Aufsichtsrates eingeleitet. Daher war von Beginn an, das Commitment der obersten Leitungsorgane gegeben.

Es wurde früh ein Ethik Board eingerichtet, das unter der Leitung der HR Managerin steht und in das Mitarbeitende und Führungskräfte aus den verschiedenen Bereichen des Unternehmens eingebunden sind. Extern wird dieses durch Prof. Christopher Frauenberger unterstützt, der diesbezügliche Erfahrungen aus anderen Organisationen einbringt. Er ist Professor für Human-Computer Interaction an der Universität Salzburg und gestaltet mit seinem Konzept des Entanglement HCI digitale Technologien inklusiv, ethisch und menschenzentriert. Er hat auch klar die Empfehlung dafür ausgesprochen, dass das Ethik Board eine rein beratende Aufgabe hat und die Entscheidung letztlich durch die Linie zu treffen ist. Dadurch wird nicht die Verantwortung vom Mitarbeitenden weggenommen,

wenn es um die Interessensabwägung geht, sondern die Verantwortlichen werden durch das Ethik Board bei ihren Entscheidungen unterstützt.

Damit die Mitarbeitenden und Führungskräfte ihre Verantwortung kennen und auch wahrnehmen können, wurden zahlreiche Informationsveranstaltungen abgehalten, u. a. Workshops im Rahmen von All Hands Meetings, Abteilungstreffen u. v. m. Dort wurden Beispiele bearbeitet und Prozesse diskutiert. Darüber hinaus wurde durch das o. a. Ambassador Konzept und die Ausbildung von einigen Berater*innen auf den IEEE7000 Standard auf Multiplikatoren gesetzt. Weiters wurde allen Mitarbeitenden toolbasierte Checklisten zur Verfügung gestellt, mit Hilfe derer anhand einiger weniger Fragen schnell erkannt werden kann, ob in einem konkreten Projekt die Gefahr der Verletzung der definierten Werte besteht (s. Abb. 1.4).



Human Oversight		Gibt es Klarheit für Nutzer:innen/Stakeholder über die Grenzen und Einschränkungen des Systems (Halluzinieren im Falle von KI-Einsatz, verwendete Datenquellen/-stände etc.)	☐
	Letzt-verantwortung	Ist sichergestellt, dass bei Fragen, die für das Wohlergehen der Menschen bedeutend sind, Menschen die Letztentscheidung treffen?	☐
	Organisatorische Voraussetzungen	Gibt es Klarheit darüber, was zu tun ist und wer entscheidet, wenn es bei der Entwicklung der Lösung zu Abweichungen von den angestrebten Werten kommt?	☐
		Wurden ethische Grundsätze im Entwicklungs- bzw. Auswahlprozess (bei Make-Entscheidung) berücksichtigt?	☐
		Sind die mit der Entwicklung der Lösung (von Idee/Anforderung bis Betrieb) betrauten Personen mit den angestrebten ethischen Werten und mit ihrer diesbezüglichen Verantwortung für deren Umsetzung vertraut und haben sie die erforderlichen Kompetenzen dafür?	☐
	Überwachung	Gibt es Notfallpläne, wenn die Lösung unerwartet agiert oder Probleme auftreten?	☐
		Wird die Lösung im Betrieb laufend überwacht, um rechtzeitig auftretende Risiken/Probleme für die Werte (Bias, Missbrauch, erschwerter Zugang etc.) zu erkennen und eingreifen zu können? Werden dazu sowohl systeminterne Warnmeldungen/ Benachrichtigungen berücksichtigt, als auch Feedback der Anwender (Stakeholder)?	☐
Verantwortung für eigenes Handeln	Folgenabschätzung	Wurden vor der Entwicklung (spätestens vor der Einführung) der Lösung die Auswirkungen auf die Mitarbeitenden, die Kund:innen, die Gesellschaft und die Umwelt evaluiert und sorgfältig abgeschätzt? (z.B. Verlust sinnstiftender Aufgaben, Angst vor Überwachung, Schaffung prekärer Arbeitsverhältnisse, Gefahr von Manipulationen...)	☐
	Accountability	Werden die Entscheidungen im Projekt dokumentiert, um diese im Nachhinein überprüfen zu können (gilt auch für Entscheidungen hinsichtlich ethischer Fragestellungen)?	☐
		Gibt es klar definierte und dokumentierte Ziele und Verantwortlichkeiten für die Lösung, die allen Teammitgliedern klar sind?	☐

Tabelle 1: Ethik-Checkliste

Abb. 1.4 Ethik Checkliste

1.6 Praxistipp: Whitebook „Digitaler Humanismus im Alltag“

In der msg Plaut wurde ein praxisorientiertes Whitebook Digitaler Humanismus unter dem Titel „Digitaler Humanismus im Alltag“ erstellt und veröffentlicht. Dieses definiert die mit dem Digitalen Humanismus verfolgten Ziele, die vereinbarten Werte und Prozesse. Es ist ein Referenzdokument für bestehende und natürlich vor allem auch für neue Mitarbeitende, damit diese einerseits ein gutes Verständnis über die Umsetzung des Themas in der Organisation erhalten, aber auch die entsprechenden Tools (wie Checklisten etc.) einfach und schnell im Zugriff haben. Dafür wurden viele erarbeitete Themen zusammengetragen und die eine oder andere inhaltliche Lücke noch geschlossen. Dabei besteht das Whitebook aus folgenden Bereichen:

1. **Vorwort des Vorstands:** Ausdruck des Commitments der obersten Führungsebene und Darstellung der Bedeutung für das Unternehmen
2. **Grundlagen:** Kurze Erläuterung des Begriffs des Digitalen Humanismus wie er in der msg Plaut angewendet wird, sowie die damit verbundenen Zielsetzungen für die Organisation und die Mitarbeitenden und die Kund*innen
3. **Werte:** Dieser Teil konzentriert sich auf die Werte, die nach den Prinzipien des Digitalen Humanismus gelebt werden. Dazu gehören Gleichberechtigung, Gesundheit, Zugang zu Bildung, Sicherheit, eine lebenswerte Umwelt und Respekt.
4. **Referenzbereich für Mitarbeitende:** Hier sind alle wesentlichen Tools und Prozessschritte, sowie Ansprechpartner und die Governance praxisnah und mit Beispielen dargestellt. Hier finden Mitarbeitende auch die Checkliste, zur Prüfung ethisch relevanter Fragen für Projekte. Mit einer Ampellogik erkennen sie rasch, ob es Risiken gibt, die eine weitere tiefergehende Analyse erfordern. Weiters sind hier die Aufgaben und die Mitglieder des Ethik Boards angeführt.

1.7 Lessons Learned

Aus den Erfahrungen der letzten Jahre lassen sich einige Schlüsselfaktoren ableiten, die wesentlich zum Erfolg der Umsetzung des Digitalen Humanismus beitragen:

- **Ehrliche und offene Kommunikation und Einbindung der Mitarbeitenden vom Beginn an.**

Das Thema Digitalisierung beschäftigt viele Menschen sehr persönlich und emotional. Oft gibt es Ängste und Sorgen, bezüglich negativer Auswirkungen auf unser eigenes Leben. Daher ist das Interesse am Digitalen Humanismus groß, viele Fragen werden in den Diskussionen aufgeworfen und der Arbeitgeber, der dieses Thema ernst nimmt, kann auf die Unterstützung und die Motivation der Mitarbeitenden zählen.

- **Mindset zählt.**

Outsourcing und „Ethik Washing“ sind der falsche Weg. Beim Digitalen Humanismus geht es in erster Linie um die Bewusstmachung möglicher Risiken und einer gezielten Lösungssuche bzw. Entscheidungsfindung. Erfolgreich sind jene Organisationen, die ihre diesbezüglichen Werte und Ziele klar definiert haben und diese glaubhaft vermitteln und (vor)leben. Eine Auslagerung von Entscheidung an ein Ethik Board oder der Fokus auf ein Zertifikat oder ein spezielles Reporting sind der falsche Weg. All das sind gute und wichtige Tools bei der Umsetzung, aber entscheidender Faktor ist und bleibt das Mindset bei der Erstellung digitaler Lösungen, bei der Prüfung der Auswirkungen etc. und die Wahrnehmung der Verantwortung durch alle Mitarbeitenden und alle Führungskräfte bis hinauf zur obersten Leitungsebene.

- **Nicht jedes Problem ist ein Nagel – auf Relevanz achten**

In der praktischen Umsetzung zeigt sich sehr rasch, welche Aktivitäten, Projekte, Bereiche im Unternehmen besondere Relevanz hinsichtlich der Fragestellungen des Digitalen Humanismus haben und welche wenig oder keine. So werden beispielsweise in einem SAP-Transformationsprojekt wesentlich weniger ethisch relevante Fragestellungen auftreten als in der Entwicklung einer Software zur Digitalisierung der Pflege in einem Spital. Daher ist es wichtig im Rahmen der Umsetzung auf die Verhältnismäßigkeit zu achten, um nicht überbordende Aufwände zu generieren und die Mitarbeitenden zu demotivieren.

Literatur

- Arnoldner, T. (2023). Die Praxis des Digitalen Humanismus. In G. Krause (Hrsg.), *Die Praxis des Digitalen Humanismus* (S. 93). Springer.
- Europäische Kommission (2018). *Hochrangige Expertengruppe für künstliche Intelligenz. „Ethik-Leitlinien für eine vertrauenswürdige KI.“*
- Europäische Kommission (2022). „Digitale Prinzipien.“ *Digital Strategy*
- Europäisches Parlament Europäischer Rat, Kommission und Mitgliedsstaaten. „Die Europäischen digitalen Rechte und Grundsätze.“ <https://digital-strategy.ec.europa.eu/de/policies/digital-principles> (Erstellt: 12.2022).
- IEEE (2024). *Corporate Digital Responsibility (CDR) Securing Our Digital Futures*. IEEE Standard
- Krause, G. (2023). Die Praxis des Digitalen Humanismus. In G. Krause (Hrsg.), *Die Praxis des Digitalen Humanismus* (S. 231–240). Springer.
- Lautermann, C., & Frick, V. (2023). *Übersicht Corporate Digital Responsibility Handlungsfelder*. Schriftenreihe des IÖW, Nr. 227/23.
- Nida-Rümelin, J. (2022). Perspectives on Digital Humanism. In H. Werthner, E. Prem, E. A. Lee & C. Ghezzi (Hrsg.), *Perspectives on Digital Humanism* (S. 74). Springer.
- Reichl, P. (2020). Digitaler Humanismus. In F. Tomaschek (Hrsg.), *Digitaler Humanismus* (S. 23–24).
- Werthner, H., et al. (2020). Das Wiener Manifest für Digitalen Humanismus. In M. Hengstschläger (Hrsg.), *Digitaler Wandel und Ethik* (S. 384–412). Ecwin.

Weiterführende Literatur

- A1 Telekom Austria (2023). A1 digital.campus. In *Umsetzung starten*.
- Ethik-Leitlinien für eine vertrauenswürdige KI. (2023). In: Den eigenen Wertekompass definieren.
- Europäische digitale Rechte und Grundsätze (2023). *Den eigenen Wertekompass definieren*
- Frauenberger, C. (2023). Unterstützung des Ethik Boards. In *Verantwortlichkeiten klären – Governance*.
- IEEE: Corporate Digital Responsibility (CDR) Securing Our Digital Futures (2023). *Den eigenen Wertekompass definieren*
- msg Plaut Austria (2023a). Fallbeispiel msg Plaut. In *Relevanz und Motivation*.
- msg Plaut Austria (2023b). Digitaler Humanismus im Alltag. In *Praxistipp Whitebook „Digitaler Humanismus im Alltag“*.
- Nida-Rümelin, J. (2018). Digital Humanism: Strengthening Human Responsibility. In *Relevanz und Motivation*.
- Times (2018). *The Internet is broken*
- Wiener Manifest für Digitalen Humanismus (2023). *Den eigenen Wertekompass definieren*

Dr. Georg Krause ist CEO der IT-Beratung msg Plaut AG, die in Österreich, Zentraleuropa und Zentralasien aktiv ist. 2023 wurde er als einer der 1000 einflussreichsten Manager Österreichs ausgezeichnet. Seine Karriere führte ihn über Horváth & Partners, Deloitte und SAP – stets mit dem Fokus auf Innovation und digitale Transformation. Als Lehrender an der FH Technikum Academy und der Donau-Universität Krems bringt er Theorie und Praxis zusammen. Darüber hinaus engagiert er sich als Keynote-Speaker, Buchherausgeber, Podcast-Host und Mitglied im Verband fit4internet für digitale Kompetenzen und eine menschenzentrierte Digitalisierung. Sein Buch und Podcast „Digitaler Humanismus in der Praxis“ sind Ausdruck dieses Engagements.

CFO-Perspektive: Projekterfolge durch Einbeziehung von Werten sicherstellen

2

Wie durch wertorientiertes Vorgehen die Effizienzziele erreicht und nachträgliche Kosten vermieden werden

Leopold Fischl

Zusammenfassung

Der Beitrag beleuchtet die erfolgreiche Implementierung von KI bei ibis acam, um die Effizienz zu steigern und Mitarbeitende zu entlasten. Durch die Neuentwicklung eines ERP-Systems und die Integration von KI werden wiederkehrende Aufgaben automatisiert. Der Fokus liegt auf der Benutzerakzeptanz und der Berücksichtigung menschlicher Werte. Erfahrungen aus früheren Digitalisierungsprojekten und der Einsatz von Value-Based Engineering (VBE) fließen in die Entwicklung ein, um eine positive Nutzererfahrung zu gewährleisten. Aus der Perspektive des CFO wird besonders hervorgehoben, wie die Automatisierung und Digitalisierung zu erheblichen Produktivitätsgewinnen und einer schnellen Amortisation der Investitionen führen.

Die Implementierung von KI in Systemen und Arbeitsprozesse ist nicht mehr nur eine Möglichkeit, sondern eine Notwendigkeit, um mit der rasanten Entwicklung und Transformation der Arbeitswelt Schritt halten zu können und um mittelfristig wettbewerbsfähig zu bleiben.

L. Fischl (✉)
ibis acam, Wien, Österreich
E-Mail: leopold.fischl@ibisacam.at

2.1 Transformation durch Technologie: Die Neuentwicklung unseres ERP-Systems

In der aktuellen Zeit, in der Technologie unaufhörlich voranschreitet, stehen auch wir bei ibis acam vor einer bedeutenden Transformation. Unser bisheriges ERP-System hat über zwei Jahrzehnte lang unsere Mitarbeiter*Innen dabei unterstützt Menschen durch Bildung und Coaching wieder in den Arbeitsmarkt zu integrieren. Doch mit der Zeit hat unser aktuelles System an technischem Glanz verloren und wurde von den technischen Möglichkeiten überholt. Die Entscheidung auf der grünen Wiese ein neues Produkt zu entwickeln, markiert somit für uns einen Wendepunkt. Unser zukünftiges System wird eine komplette Neuentwicklung, die darauf abzielt, die Schwächen des Vorgängers zu überwinden und eine Plattform zu schaffen, die bereit für die Herausforderungen der Zukunft ist. Neben der Automatisierung wiederkehrender und monotoner Abläufe wird die Integration von künstlicher Intelligenz (KI) eine entscheidende Rolle einnehmen, um die Effizienz zu steigern und Mitarbeiter*Innen zu entlasten damit sie unserer ureigensten Aufgabe, für unsere Teilnehmer*Innen neue Perspektiven und Möglichkeiten zu finden, nachgehen können.

Die Implementierung von KI in Systemen und Arbeitsprozesse ist nicht mehr nur eine Möglichkeit, sondern eine Notwendigkeit, um mit der rasanten Entwicklung und Transformation der Arbeitswelt Schritt halten zu können. Das wird auch durch zahlreiche Studien untermauert. Beispielsweise soll lt. einer Studie von McKinsey bis 2030 jede dritte Arbeitsstunde, unabhängig von der Branche, durch Automatisierung ersetzt werden können. Einen großen Anteil wird bei dieser Entwicklung der generativen KI zugeschrieben. Solche Möglichkeiten müssen von einem Unternehmen genutzt werden, um sich für die Zukunft robust aufzustellen.

In unserem spezifischen Kontext bedeutet es einerseits das Potenzial, durch die Implementierung von KI, Produktivitätsgewinne zu realisieren und andererseits der Verpflichtung gegenüber unseren Unternehmenswerten auch die Verantwortung zu übernehmen, eine Software zu schaffen, die menschliche Werte in der Zusammenarbeit fördert und in der Ausgestaltung eben dieser Software bereits berücksichtigt. Somit ist zur komplexen Aufgabe einer Systemablöse, mit all den bekannten Herausforderungen auch das Spannungsfeld, der Implementierung von KI unter Berücksichtigung menschlicher Werte und Bedürfnisse, hinzugekommen.

2.2 Woran werden erfolgreiche Softwareprojekte gemessen?

Wenn Sie ChatGPT fragen, woran man erfolgreich eingeführte Softwareprojekte oder Systemabläsen erkennt, bekommt man eine Liste von verschiedenen Kriterien und KPIs retour. Bei meiner Abfrage waren es zehn, und an zweiter Stelle, noch vor ROI und tech-

nischer Stabilität wurde die Benutzerakzeptanz angeführt. ChatGPT hat die Kriterien für hohe Benutzerakzeptanz folgendermaßen beschrieben¹:

- Positive Nutzererfahrungen: Mitarbeitende nutzen die Software aktiv und äußern sich positiv zu ihrer Benutzerfreundlichkeit und Leistungsfähigkeit.
- Hohe Nutzungsrate: Die Software wird von den vorgesehenen Nutzern regelmäßig eingesetzt und nicht durch inoffizielle Alternativen ersetzt.
- Geringer Widerstand: Es gibt nur wenige oder keine Widerstände gegen die neue Lösung.

Wir haben für die Ablöse unseres ERP-Systems natürlich auch verschiedenen Kennzahlen und Kriterien definiert. Diese orientierten sich vorwiegend am wirtschaftlichen Erfolg des Projektes (ROI, Effizienzsteigerungen durch eingesparte Arbeitsminuten, etc.) sowie technischen Maßstäben und Kriterien.

Das Thema Benutzerakzeptanz ist für uns ebenfalls von großer Bedeutung, da uns bewusst ist, dass die Systemablöse aufgrund von Widerständen oder negativen Nutzererfahrungen ebenso scheitern kann. Aufgrund der Komplexität der Systemablöse planen wir von Beginn an einen hybriden Betrieb beider Systeme. Das neue System entsteht und löst schrittweise das Alte ab. Da Beide daher eine Zeitlang parallel bestehen und genutzt werden haben wir für eine erfolgreiche Benutzerakzeptanz des neuen Systems das Ziel so formuliert: Das neue System muss so perfekt sein, dass es die Nutzer*Innen nicht erwarten können, neue Funktionen zu bekommen. Oder salopp gesagt: Jede*r sollte „gierig“ auf die neue Software werden.

Und letztendlich ist die Benutzerakzeptanz meiner Meinung nach auch einer der wichtigsten Erfolgsfaktoren: Die Systemablöse scheitert nämlich auch und vor allem sehr schmerzlich, wenn ein technisch einwandfreies Produkt mit allen Vorteilen für Effizienz und Produktivitätssteigerung aufgrund fehlender Akzeptanz nicht (volumfänglich) angewendet wird. Ein derartiges Scheitern ist daher unter allen Umständen zu verhindern.

Daher stellten sich folgende Fragen für uns:

- Welche Anforderungen an ein System bestehen aus Nutzersicht neben einer einfachen Bedienbarkeit noch?
- Wie zahlen diese auf die Benutzerfreundlichkeit und Akzeptanz ein?
- Und vor allem: Wie erhebt man diese bzw. wie fördert man diese in einem strukturierten Prozess zu Tage?

¹ Für die Softwareentwicklungsexperten unter ihnen gibt es auch die Gartner Kriterien.

2.3 In die Vergangenheit horchen – Welche Schlüsse ziehen wir aus den Erfahrungen vergangener Digitalisierungsprojekte für den digitalen Assistenten?

Bevor ich auf die Antwort auf diese Fragen eingehe, möchte ich kurz einen Blick auf vergangene Digitalisierungsprojekte werfen und herausstreichen, welche Learnings für das aktuelle Projekt mitgenommen wurden.

Als Dienstleistungsunternehmen mit einer stundenbasierten Abrechnung unserer Bildungsmaßnahmen haben wir seit jeher sehr hohen Ansprüchen an die Qualität unserer Zeitaufzeichnungen und daher durchlaufen diese auch einen mehrstufigen Prüfprozess. Jede Stufe wird dabei durch persönliche Unterschrift dokumentiert und frei gegeben. Mit rund 1000 Mitarbeiter*Innen an 80 Standorten ist das sowohl ein erheblicher Zeit- als auch logistischer Aufwand, da nicht alle Prüfschritte immer an einem Standort vorgenommen werden können und letztendlich alle Zeitaufzeichnungen in der Zentrale in Wien gesammelt werden. Bis März 2020 war dieser Vorgang vollkommen analog und manuell organisiert. Sie merken schon, auch wir haben die Pandemie für Digitalisierung genutzt.

Um diesen Prozess deutlich effizienter zu organisieren, entschieden wir uns bereits im Herbst 2019 einen digitalen Unterschriftenprozess einzuführen. Ziel war es, den administrativen Aufwand zu reduzieren, die Prozessgeschwindigkeit zu erhöhen und die Datensicherheit zu verbessern. Durch die Integration in unser bestehendes ERP-System wurde das Genehmigungsverfahren deutlich vereinfacht und beschleunigt. Wir konnten damit die durchschnittliche Durchlaufzeit von 15 Tagen auf 5 Tage reduzieren und den Arbeitsaufwand pro Zeitaufzeichnung von durchschnittlich 15 min auf ca. 2 min verringern. Also hochgerechnet pro Monat rund 13.000 min administrative Arbeitszeit durch Digitalisierung für wertschöpfende Tätigkeiten gewonnen. Ein enormer Fortschritt und der ROI für dieses Projekt war damit sehr schnell erreicht. Aus CFO-Perspektive eine win-win-win Situation.

Die Einführung begann im Jänner 2020 mit einem sehr vielversprechenden Testlauf. Die gesamte Umstellung auf den digitalen Prozess wurde durch die Pandemie und den wiederkehrenden Lock downs erheblich beschleunigt. Die Akzeptanz für die Umstellung und die geänderten Abläufe war bei unseren Mitarbeiter*Innen anfangs sehr hoch, zumal es während der Pandemie auch keine Alternativen dazu gab. Aufgrund des Erfolges haben wir uns noch im ersten Halbjahr 2020 sehr schnell dazu entschlossen sämtliche internen Unterschriften (Dienstverträge, etc.) ebenfalls auf diesem Weg zu digitalisieren.

Mit der Rückkehr in den Normalbetrieb änderte sich jedoch sukzessive die Zufriedenheit der Mitarbeiter*Innen mit den veränderten Abläufen. Diese Veränderung war für mich nicht nachvollziehbar, lagen doch die Vorteile der Änderungen auf der Hand. Einfacher, schneller, weniger Aufwand, durchgängig transparent etc. Warum also dieser Umschwung? Was hat sich gegenüber der Pandemiesituation verändert?

Nach einigen Gesprächen mit Anwender*Innen und Auswertung des gesammelten Feedbacks konnten wir die Problematik folgendermaßen eingrenzen. Mit den neuen digitalen Prozessen haben sich ebenfalls einige Kommunikationswege verändert und die Bildschirmarbeitszeit ist quer durch alle Funktionen und Hierarchien angestiegen. Wurden gerade bei den Zeitaufzeichnungen Korrekturen früher handschriftlich direkt auf den Ausdrucken vermerkt und anschließend im System eingetragen, war es nun notwendig Fehler textlich zu beschreiben und Änderungen im System mittels E-Mails zu veranlassen. Gleichzeitig ist der persönliche Kontakt zwischen Mitarbeiter*In und Führungskraft zurück gegangen. Konnte man vor der Umstellung bei der Abgabe und Kontrolle der Zeitaufzeichnungen ein kurzes privates oder berufliches Thema besprechen verlagerte sich diese auf Meetings bzw. mussten diese Gespräche formalisiert oder anderswertig geführt werden. Kurzum: Die Änderungen der Abläufe und Prozesse hat ebenso Änderungen im Miteinander der Menschen nach sich gezogen. Dieser Eingriff in ein bestehendes Gefüge und Wertesystem (Kommunikation, persönlicher Austausch) hat nach der Rückkehr, in das vor den Lock Downs gewohnte Arbeitsumfeld, bei vielen zu einer teilweisen Ablehnung des digitalen Unterschriftensystems geführt.

Diese Änderungen der informellen Abläufe, Strukturen und sozialen Interaktion waren vor und während der Umstellung nicht betrachtet worden. Sie hatten aber letztendlich eine große Auswirkung auf die Zufriedenheit mit der Systemänderung bzw. dann letztendlich mit dem neuen System.

Bei den Planungen und Überlegungen zur Ablöse unseres ERP-Systems war somit von Anfang an klar, dass diese Aspekte berücksichtigt werden müssen. Mit der Einführung eines digitalen Assistenten und den Möglichkeiten, die durch KI-Anwendungen entstehen, wird es erhebliche Auswirkungen und Änderungen auf die bisherigen Abläufe, Kommunikationswege und die gewohnte Art der Zusammenarbeit geben.

2.4 Zusammenspiel von Autonomie, Motivation und Produktivität

Wir kennen bzw. wissen es und handeln auch in unserem Arbeitsalltag danach: Menschen sind produktiver, kreativer und motivierter, wenn Freiräume und Gestaltungsspielräume bei der Bewältigung von Arbeitsaufträgen vorhanden sind². Digitalisierte Prozesse mit den notwendigen strikten und wiederkehrenden Abläufen engen manchen Spielraum jedoch ein. Es stellt sich daher die Frage, wie erzielt man den maximalen Trade off zwischen Autonomieeinschränkung und Produktivitätsgewinn von KI-Tools?

Dieser Zusammenhang ist kognitiv und logisch – denke ich – vollkommen nachvollziehbar. Spür- und erlebbar ist er vermutlich im Arbeitsalltag für viele von uns, mich ein-

² Anführen von Studien und Beispielen.

geschlossen, vielleicht nicht mehr so deutlich. Daher möchte ich ein kurzes und Erlebnis aus meinem privaten Alltag schildern, bei dem ich die Einschränkung meiner Autonomie emotional sehr deutlich gespürt habe.

Stellen sie sich folgende Szene vor: einen freien Arbeitstag will ich nutzen, um ein gutes Buch zu lesen, ein wenig Sport zu treiben und notwendige Hausarbeiten zu erledigen. Konkret warten mehrere Waschmaschinenladungen auf mich. Ein dichtes Programm mit jedoch ausreichend Flexibilität den Tagesablauf produktiv und erholsam zu gestalten.

Mit Ende der ersten Waschladung unserer vernetzten Waschmaschine passiert dann folgendes: Gleichzeitig mit dem Ende des ersten Wachganges der in der App lautstark angezeigt wird, läutet auch mein Telefon und meine Frau meldet sich mit folgender Botschaft: „Du, ich sehe gerade, die Waschmaschine ist fertig, gib mir kurz Bescheid, wenn sie wieder befüllt ist, und ich schalte sie dann von unterwegs wieder ein.“

Diese kurze Episode hat mich im weiteren Tagesverlauf noch sehr beschäftigt. Ersetzen wir nämlich den charmanten Anruf meiner Frau durch einen KI gestützten digitalen Assistenten, der den Tagesrhythmus vorgibt, Arbeitsschritte anleitet und auf Abweichungen von Prozessabläufen aufmerksam macht, sind wir sehr nahe an den Überlegungen unseres digitalen Assistenten bei ibis acam. Zusätzlich hat sich die App mit einem Vorschlag für das Programm der nächsten Waschladung gemeldet. Dieses Angebot des vorausgewählten Programms hat an meinem Stolz ordentlich gekratzt. Habe ich doch den Anspruch an mich selbst, die richtige Programmierung der Waschmaschine selbständig und ohne fremde Hilfe hinzubekommen. Die angebotene Hilfestellung bzw. letztendlich die Abnahme einer Entscheidung hat bei mir negative Emotionen ausgelöst.

Sie schwankten somit von Dankbarkeit für die Erinnerung zur Erledigung der anstehenden Aufgaben bis hin zu unangenehme kontrolliert bzw. fremdgesteuert. Also eine Bandbreite von positiven und negativen Emotionen die sich in einem Arbeitsumfeld auf Produktivität und Arbeitszufriedenheit auswirken werden.

Natürlich ist das eine höchst persönliche Reaktion auf Basis meiner Werte und Einstellungen. Bei anderen Menschen kann und wird es völlig andere und vielleicht vermehrt positive Reaktionen und Emotionen auf ähnlichen Situationen geben.

Jedenfalls habe ich anhand dieses kurzen Erlebnisses sehr deutlich gespürt, wie groß der Einfluss digitaler Tools und KI auf unsere Werte und Haltungen und letztendlich Emotionen sein kann. Ein digitaler Assistent, der bei ibis acam ähnlichen Funktionen – automatische Erinnerungen, Entscheidungen vorbereiten, Abweichungen aufzeigen etc. – übernehmen wird, wie in meinem Beispiel, sollte in keinem Fall bei 1000 Mitarbeiter*Innen negative Emotionen auslösen, die zu Demotivation und Frust führen. Jegliche Produktivitätsgewinne aufgrund von Automatisierungen und vereinfachten Abläufen wären durch demotivierte Mitarbeiter*Innen möglicherweise an anderer Stelle wieder verloren.

Den optimalen Grad zwischen dem Einsatz von KI und der Erzielung von Produktivitätsgewinnen unter Wahrung der persönlichen Werte eines jeden einzelnen und unserer Unternehmenswerte muss daher ebenso in die Systementwicklung miteinbezogen werden.

2.5 In der Gegenwart handeln – ein neuer Ansatz in der Softwareentwicklung

Anhand dieser Erfahrungen war klar, dass weitere Dimensionen wie menschliche Werte, unser soziales Gefüge bei ibis acam, unsere informelle „Schattenorganisation“ sowie der optimale Grad zwischen Automatisierung, dem Einsatz von KI und individueller Gestaltungsspielräume für das Gelingen unserer Neuentwicklung bereits bei der Anforderungsanalyse berücksichtigt werden müssen. Dies vor allem auch aufgrund wirtschaftlicher Überlegungen. Eine nachträgliche Änderung der Software, um die oben beschriebenen Aspekte in der Software später zu integrieren oder bereits bestehende Funktionen entsprechend anzupassen ist mit erheblichen Mehrkosten in der Entwicklung verbunden. Zusätzlich verzögern sich die positiven Effekte aus der Systemverbesserung und damit entsteht ein doppelt negativer Impact.

Eine sehr gute Möglichkeit diese Anforderung ans Licht zu bringen, bietet das sogenannte Value-Based Engineering (VBE) mit der Methode des IEEE7000 Standards der durch unseren Software-Entwicklungspartner, die msg Plaut, vorgeschlagen und in unserem Projekt erstmalig angewendet wurde. Im Wesentlichen ist dieser Standard ein risikobasierter Ansatz und erhebt in einer sehr frühen Phase der Systementwicklung die negativen und positiven Effekte von Funktionen einer Software auf unser humanistisches Wertesystem. Die konkrete Vorgehensweise wird in Kap. 15 beschrieben. Zentraler Ansatz dieser Methode ist die Einbeziehung verschiedener Stakeholder in den Entwicklungsprozess. Durch diese frühzeitige Integration der Endbenutzer – im Wesentlichen waren dies unsere Mitarbeiter*Innen³ – konnten wir sicherstellen, dass das neue System nicht nur technologisch fortgeschritten ist, sondern auch die oben angeführten Kriterien für eine breite und durchgängige Benutzerakzeptanz erfüllt.

In unserem Fall haben wir mit einer Gruppe von 12 Personen aus den verschiedensten Unternehmensbereichen und Hierarchieebenen an 3 Workshoptagen unsere Vorstellungen und Funktionen des digitalen Assistenten auf den Prüfstand gestellt. Dabei wurden von den Workshopteilnehmer*Innen negative und positive Auswirkungen der Funktionen des digitalen Assistenten auf Werte wie Freiheit, Selbständigkeit, Gerechtigkeit, soziale Interaktionen etc. beurteilt.

Für alle Mitwirkenden war das eine neue und sehr spannende Erfahrung. Vor allem auch deswegen, weil es völlig andere Erwartungen an den Workshop und die Vorgehensweise gab, wie KI-Anwendungen und Softwarefunktionen beurteilt werden sollten. Das Feedback einer Teilnehmerin bringt es sehr schön auf den Punkt: „Wir haben uns drei Tage mit KI und Software beschäftigt, ohne ein einziges Mal über KI selbst oder technische Möglichkeiten zu sprechen. Sondern es ging ausschließlich um den Menschen.“

Bei der Auswahl der Personen war uns eine heterogene Gruppenzusammenstellung wichtig. Ältere, jüngere, lange gediente, technikaffin und weniger versierte mit dem Umgang von Softwareanwendungen von Trainer*Innen bis zur Geschäftsführung. Durch die

³ Verschiedenste Funktionen über alle Hierarchieebenen hinweg.

dadurch unterschiedlichen Einstellungen, Werte und Sichtweisen haben wir ein sehr breit gefächertes Ergebnis und sehr viele Rückmeldungen zu den geplanten KI-Anwendungen bekommen. Diese sind ein sehr wertvoller Input für die weitere Systementwicklung und die Gestaltung des digitalen Assistenten.

2.6 Kosten und Nutzen an das Value Based Engineering

Mit VBE haben wir in unserem Projekt unterschiedliche Erwartungen und Zielsetzungen verfolgt.

- Für die vorhin aufgezeigten Fragen war es notwendig Antworten und Handlungsalternativen zu erarbeiten.
- Es sollten zusätzliche Entscheidungskriterien für die Auswahl des dem digitalen Assistenten zugrundeliegenden KI-Modells gewonnen werden.
- Weiterer Input für die Anforderungserhebung und Beschreibung und Gestaltung des digitalen Assistenten
- Ein positiver Einfluss auf die Benutzerakzeptanz einerseits durch die frühe Einbeziehung der Mitarbeiter*Innen in den Entwicklungsprozess und andererseits vor allem durch die Berücksichtigung der in den Workshops formulierten Ängste der Mitarbeiter*Innen.

Neben dem Input für die Softwareentwicklung haben die Workshops noch zu einem weiteren positiven Effekt beigetragen. Die Auseinandersetzung von Softwarefunktionen und deren Einfluss auf menschlichen Werten anhand ethischer Modelle wurde in der internen Kommunikation verwendet und zusätzlich hat sich der Ansatz im Unternehmen positiv herumgesprochen. Insgesamt hat das zu einer positiven Grundhaltung und Vorfreude auf die bevorstehende Systemablöse bei den Mitarbeiter*Innen beigetragen.

Die Workshops wurden im ersten Drittel des Entwicklungsprozesses durchgeführt und die Ergebnisse daraus werden entlang des gesamten Projektes genutzt. So werden Rückmeldungen bereits bei der Anforderungserhebung berücksichtigt und geben gleichzeitig wichtige Impulse für die Implementierung und Einführung der neuen Anwendungen. Auch wenn es schwierig mess- oder vergleichbar sein wird, erwarte ich durch den VBE-Ansatz letztendlich ein besseres Softwareprodukt und somit ist der Ansatz ein wesentlicher Baustein für das Gelingen der Softwareablöse.

Abschließend noch ein kurzer Blick auf die zusätzlichen Kosten, die mit der Implementierung des IEEE7000 Standards in unserem Projekt angefallen sind. Neben den internen Personalkosten für die Workshopgruppe kommen noch die Kosten für die Moderation sowie die Aufbereitung und Präsentation und des Outputs aus den Workshops hinzu. Zusätzliche Kosten in der Softwareentwicklung fallen nicht an, da die Workshopergebnisse im Zuge der ohnehin stattfindenden Anforderungserhebungen in den Entwicklungsprozess miteinfließen. Insgesamt ein sehr überschaubarer Aufwand gemessen am erwarteten Nutzen für die Software und den Gesamtkosten des Projektes.

Mag. Leopold Fischl ist seit 2014 Geschäftsführer und CFO der ibis acam Bildungs GmbH, dem größten Unternehmen innerhalb der Aspire Education Gruppe. Er verantwortet die Bereiche Finanzen, IT & Digitalisierung. Seine berufliche Laufbahn begann er nach dem Studium der Betriebswirtschaft an der Wirtschaftsuniversität Wien bei PwC in der Wirtschaftsprüfung, bevor er in führenden Finanzfunktionen tätig wurde – unter anderem bei Casinos Austria International, Robust Plastics GmbH und zuletzt in der Aspire Education Gruppe. Seit über 20 Jahren in verschiedenen Funktionen ist es ihm stets ein Anliegen, durch Prozessoptimierungen und Digitalisierung zur Wertsteigerung von Unternehmen beizutragen.

CIO-Perspektive 1: Wiener (digitale) Lebensqualität

3

Technologischer Fortschritt und Digitaler Humanismus in Wien

Klemens Himpele

Zusammenfassung

Der Beitrag beleuchtet, wie Wien durch die Integration von Technologie die Lebensqualität seiner Bürger*innen verbessert. Historische und aktuelle Beispiele zeigen, wie technologische Entwicklungen, insbesondere im Bereich der IT, gezielt eingesetzt werden, um den Alltag zu erleichtern. Der Digitale Humanismus steht dabei im Mittelpunkt, indem er sicherstellt, dass technologische Fortschritte stets dem Wohl der Menschen dienen und ethische Werte berücksichtigt.

Wien hat sich immer dadurch ausgezeichnet, neue Technologien zu nutzen, um den Alltag der Menschen zu erleichtern. Technologischer Fortschritt war nie Selbstzweck, sondern immer ein Mittel, um das Leben zu verbessern.

3.1 „Stadt der Ideen. Als Wien die moderne Welt erfand“

Dies ist der (deutsche) Titel eines Buches des englischen Historiker Richard Cockett. Darin beschreibt er, dass die Gestaltbarkeit der Stadt – etwa Bildung und Hygiene – und damit die Gestaltungsmöglichkeiten des eigenen Lebens, ein Herzstück des Roten Wien waren, das später durch die Vertreibung vieler – vor allem jüdischer – Intellektueller aus Wien in die Welt getragen wurde: „This transformative belief in scientifically grounded

K. Himpele (✉)
Magistratsdirektion Stadt Wien, Wien, Österreich
E-Mail: klemens.himpele@wien.gv.at

human progress characterised almost everything that the Viennese attempted in Austria and later in the West“ (Cockett, 2023, S. 40). Vieles aus dieser Zeit ist tief in der DNA der Stadt verankert, die mit einer konsequent menschenzentrierten Stadtpolitik zur Stadt mit einer der höchsten Lebensqualitäten weltweit wurde. Dazu wurden immer sozialwissenschaftliche Erkenntnisse mit Erkenntnissen aus der Technik verbunden, um Lösungen für die Stadt umzusetzen. Lebensqualität heißt in Wien eben auch, sich den Alltag möglichst einfach organisieren zu können – durch gute und zuverlässige „Öffis“, durch eine hohe Zahl und auf Vollzeiterwerbstätigkeit der Eltern ausgelegt Kindergartenplätze, durch verfügbaren und leistbaren Wohnraum oder durch Hochwasserschutzkonzepte, die auch extreme Starkregenereignisse bewältigen können.

Hieran ließen sich die Versprechen der Digitalisierung anschließen, soll doch auch durch diese Technologien das Leben der Menschen erleichtert werden. „Die Herausforderung besteht demnach darin, die in Wien geschätzte hohe analoge Lebensqualität auch im digitalen Raum zu realisieren.“ Wie so oft bei technologischen Entwicklungen kommt es jedoch auf die Rahmenbedingungen an, um tatsächlich den Nutzen für die Menschen zu erreichen. So unstrittig es ist, dass der technologische Fortschritt auch in der IT die Basis unseres Wohlstandes ist (vgl. hierzu Himpele, 2022), so klar liegen die Herausforderungen der Digitalisierung auf dem Tisch: Hass im Netz (vgl. Brodnig, 2016; Pörksen, 2018) und Fake News (vgl. Brodnig, 2018; Volland, 2024), Datenmissbrauch und Überwachungstechniken (vgl. Schrems, 2014; Zuboff, 2018), IT-Sicherheit (vgl. Wolfangel, 2022) und ganz allgemein die Technikfolgenabschätzung (vgl. Grunwald, 2022). Gerade die Entwicklungen der generativen künstlichen Intelligenz verschärfen die Befürchtungen weiter (vgl. Himpele, 2024). Der Philosoph Harari hat in einem Buch ausführlich das Thema Informationsnetzwerke untersucht. Darin schreibt er: „Contrary to what the mission statements of corporations like Google and Facebook imply, simply increasing the speed and efficiency of our information technology doesn’t necessarily make the world a better place. It only makes the need to balance truth and order more urgent“ (Harari, 2024, S. 39). Nicht einmal das Argument der Zugänglichkeit zu Wissen ist also ein eindeutiges Plus der Informationsgesellschaft. Doch all das darf nicht über den enormen Nutzen hinwegtäuschen, den der technologische Fortschritt bringt – insbesondere dann, wenn die Rahmenbedingungen stimmen.

Wien hat sich immer dadurch ausgezeichnet, neue Technologien zu nutzen, um den Alltag der Menschen zu erleichtern. Technologischer Fortschritt war nie Selbstzweck, sondern immer ein Mittel, um das Leben zu verbessern. Das gilt auch für die IKT (Informations- und Kommunikationstechnologien), wie der Wiener Bürgermeister, Michael Ludwig, schreibt: „(...) unser Ziel ist nicht, möglichst viel Technologie, sondern eine möglichst hohe Lebensqualität für alle Wienerinnen und Wiener.“ Und weiter: „Digitalisierungshauptstadt“ (...) ist also jene Stadt, die digitale Tools für die Verbesserung der Lebensbedingungen der Menschen nutzt und moderne Technologien auch tatsächlich einsetzt“ (Ludwig, 2023, S. 38). Es geht also darum, die Technik gezielt einzusetzen. „The idea that technology alone can solve social and political problems is a dangerous delusion. But the idea that they can be solved without technology is also wrongheaded“ (Suleyman,

2023, S. 150). Das Wiener Gegenmodell ist es den Menschen auch bei der Digitalisierung in den Mittelpunkt zu stellen – darum geht es beim Digitalen Humanismus.

3.2 Digitaler Humanismus im digitalen Wien

Der Digitale Humanismus (Nida-Rümelin, 2022; Stadt Wien, 2022; Strassnig et al., 2019; Werthner et al., 2022; Krause, 2024) ist ein Konzept, das den Nutzen für die Menschen in den Mittelpunkt der Überlegungen stellt. Der Mensch muss gleichsam die Autorenschaft über sein Leben behalten, wie es Nida-Rümelin formuliert. Auch die im Oktober 2024 durch den Wiener Gemeinderat beschlossene Digitale Agenda Wien hält fest: „Technologien sind nicht neutral und haben eine ordnende Kraft. Sie sind von den Menschen, die sie gestalten, geprägt und beinhalten deren Werte, Normen, wirtschaftliche Interessen und Annahmen darüber, wie unsere Gesellschaft ist oder sein sollte“ und weiter: „In Wien folgen wir bei der Gestaltung der Digitalisierung dem Weg des digitalen Humanismus [...] Wir wollen das Wohl der Menschen mit ihren sozialen und gesellschaftlichen Bedürfnissen in das Zentrum der Entwicklungen stellen. Digitale Technologien müssen das Leben der Wiener*innen verbessern – durch konkreten Nutzen wie leichtere Antragstellungen, bessere Interaktion mit der Verwaltung und durch die Nutzung vorhandener Informationen zur automatischen Abwicklung ihrer Anliegen“ (Stadt Wien, 2024a, S. 6).

Diese Forderung nach „dem Menschen im Mittelpunkt“ klingt recht einfach – im Konkreten müssen hier jedoch permanent Wertungen vorgenommen werden. Wenn beispielsweise die Frage im Raum steht, ob für Patientengespräche Übersetzungstools verwendet werden sollen, dann kollidieren unter Umständen Datenschutz und Patientenrechte miteinander. Die Orientierung am Digitalen Humanismus betrachtet Werte hier gezielt – und kann daher entsprechende Priorisierungen vornehmen. Die Werte-Polarisierung entwickelt sich zu einer Werte-Priorisierung: also nicht der Datenschutz wird absolut gesetzt und der Patient versteht den Arzt dann nicht, sondern es ist nach den Wertvorstellungen des Menschen im Zentrum der Überlegung zu gewichten. Die Stadt Wien hat daher bereits 2019 eine Studie beauftragt, die die Akteure des Digitalen Humanismus in Wien untersucht hat (Strassnig et al., 2019).

Digitaler Humanismus ist ein Leitfaden, der uns helfen kann, Fehler der Vergangenheit zu vermeiden und den Menschen, seine Würde von Anfang in das Zentrum unserer Entwicklungen und Fortschritte zu stellen. Denn auch die Industrialisierung mit ihrem enormen Pauperismus wurde erst zu einem echten gesellschaftlichen Erfolg, als es der Arbeiterbewegung mit ihren Parteien und Gewerkschaften, aber auch kirchlichen Einrichtungen – man denke an die Sozialenzyklika „Rerum Novarum“ von Papst Leo XIII – gelang, die negativen Auswirkungen einzuhegen – durch Arbeitszeitregelungen, Arbeitnehmerschutz, Urlaubsansprüche und der Sozialversicherung.

3.3 Die IT nutzen

Die öffentliche Verwaltung agiert, an Gesetze gebunden, zum Wohle der Bürger*innen. In diesem Sinne ist schon in der Aufgabenbeschreibung das Prinzip des „Menschen im Mittelpunkt“ verankert. Konkret geht es darum, technische Tools so zu entwickeln und einzusetzen, dass das Leben der Bürger*innen leichter wird. Jeder Online-Antrag erspart einen Weg, jede Information, die gut auffindbar ist, erspart Zeit. Und je effizienter und effektiver Verwaltungsprozesse abgearbeitet werden, umso zügiger kommen Bürgerinnen und Bürger sowie Unternehmen zu den Ergebnissen ihrer Verwaltungsprozesse. Kurzum: IT kann und muss genutzt werden, die Kernaufgabe der Verwaltung zu stärken – nämlich das Leben in einer Stadt organisieren zu helfen.

3.4 Was ist relevant?

Die Tätigkeiten einer Stadtverwaltung sind vielfältig: Parkpickerl, Kindergartenplätze, Unterstützungsleistungen, Statistik, Gesundheitsdienste, Schulen, Sauberkeit, Wasserver- und -entsorgung, Hochwasserschutz, Verkehr, Sicherheit usw. Nicht alles davon ist im Sinne eines Digitalen Humanismus relevant, sieht man davon ab, dass selbstredend alle digitalen Prozesse im Sinne einer einfachen User Journey bereitgestellt werden müssen. Das ist wichtig, denn wir sollten die zentralen Debatten auf die relevanten Themenfelder lenken. Relevant heißt: Ich kann sie gestalten und die ethischen Werte sind explizit zu bewerten und zu gewichten. So wird die Erkennung von Fundgegenständen durch Künstliche Intelligenz (durchaus mit ergänzenden organisatorischen Maßnahmen) vergleichsweise unproblematisch sein, weil keine Persönlichkeitsrechte betroffen sind.

Noch etwas ist wichtig: In den wenigstens Fällen hilft ein radikaler Ansatz, auch wenn große Worte wie Disruption gerade im digitalen Kontext schnell fallen. Der Soziologe Armin Nassehi (2024, S. 21) weist zu Recht darauf hin, dass bei den großen Worten oft das Dringliche mit dem Möglichen verwechselt wird. Er plädiert dafür, die Menschen bei den Entwicklungen mitzunehmen – und lieber kleine mögliche als große unmögliche Schritte zu machen. Auch dieses „Mitnehmen“ kann als Teil einer Strategie des Digitalen Humanismus interpretiert werden.

Die Frage des Digitalen Humanismus wird bereits länger diskutiert als die Folgen der generativen KI, etwa bei der Frage der Programmierung von Algorithmen in einer Art und Weise, die die Abhängigkeit von bestimmten Plattformen fördert. Dennoch ist die rasante Entwicklung der KI der Grund, warum die Diskussion derzeit verschärft geführt wird und zu Recht eine größere Bedeutung erhalten hat. Generell ist hier vieles an Unsicherheiten und Unklarheiten auszuräumen. Es geht daher zunächst darum, Mitarbeiterinnen und Mitarbeiter zu schulen und die Öffentlichkeit zu den Positionen der Stadt zu informieren (etwa in der Digitalen Agenda 2030, vgl. Stadt Wien (2024a), aber auch in einer Broschüre zum Digitalen Humanismus, vgl. Stadt Wien (2022) oder in Veranstaltungen wie den Digital Days oder den Wiener Vorlesungen).

3.5 Klare Regeln und Unterstützung für die Beschäftigten

Die Frage ethischen Handelns ist in der IT auf Grund der massenhaften Verarbeitung von Informationen nicht erst mit dem Aufkommen der Künstlichen Intelligenz ein großes Thema – sie hat sich aber durch die enorme Leistungsfähigkeit der neuen Systeme deutlich verschärft. Die Stadt Wien hat bereits 2019 eine erste KI-Strategie formuliert, die einen zielführenden Umgang mit dieser Technologie unterstützt (Stadt Wien, 2019). 2024 wurde diese Strategie überarbeitet (Stadt Wien, 2024b). Damit sind die strategischen Rahmenbedingungen abgesteckt, mit der aktuellen Version der KI-Strategie will man „vertrauenswürdige KI in Europa und darüber hinaus [...] fördern, indem sichergestellt wird, dass durch KI-Systeme die Grundrechte, die Sicherheit und die ethischen Grundsätze geachtet und die Risiken sehr leistungsfähiger und wirkungsvoller KI-Modelle adressiert werden.“

Der Magistrat der Stadt Wien hat daneben aber die Notwendigkeit erkannt, der Belegschaft klare Rahmenbedingungen für ihre tägliche Arbeit an die Hand zu geben. Damit wurde Sicherheit geschaffen und überhaupt erst die Möglichkeit eröffnet, dass sich die Mitarbeiterinnen und Mitarbeiter der Verwaltung aktiv mit den neuen Technologien und ihren Auswirkungen befassen. Der „Kompass für den dienstlichen Umgang mit generativer Künstlicher Intelligenz“ wurde erstmals im August 2023 veröffentlicht (Stadt Wien, 2023). Wesentlicher Punkt des Kompasses ist die explizite Erlaubnis für die Bediensteten der Stadt Wien, generative künstliche Intelligenz einzusetzen (hierzu und im Folgenden vgl. Himpele, 2024). Hintergrund ist der sinnvolle Einsatz neuer Technologien zur Entlastung der Mitarbeiterinnen und Mitarbeiter bzw. zur Etablierung besserer Services für Bürger*innen und Unternehmen: „Die Nutzung von im Internet verfügbaren generativen KI-Werkzeugen für den dienstlichen Gebrauch ist erlaubt und erscheint im Sinne einer modernen und effizienten Stadtverwaltung unumgänglich.“ Gleichzeitig formuliert der Kompass einige Rahmenbedingungen, wie bspw.:

- Das Grundprinzip Eigenverantwortung: Bei der Verwendung von KI-generierten Inhalten trägt der Mensch die Verantwortung.
- Auf einen rechtskonformen Umgang mit personenbezogenen Daten und dem Amtsgeheimnis ist zu achten.
- Kennzeichnungspflicht mittels KI erstellter Inhalte.

Mit dem Kompass sollte Klarheit und damit die Möglichkeit geschaffen werden, die neuen Technologien auszuprobieren und damit zu experimentieren.

Um experimentieren zu können, müssen die Mitarbeiterinnen und Mitarbeiter dazu in die Lage versetzt werden. Das geschieht einerseits durch klassische Weiterbildung – die Stadt Wien stellt eLearnings etwa zu KI, aber auch zu Fake News bereit – aber auch durch stadtinternes Community Building. Ein Kernstück der Wiener KI-Initiative ist „KITT“

(Künstliche Intelligenz Technologie und Trend), ein umfassendes Austausch- und Fortbildungsprogramm. Die drei Grundbausteine von KITT sind dabei:

- Die KITT-Einführung: Sie bietet Neueinsteiger*innen einmal im Monat die Möglichkeit, neu in das Thema einzusteigen.
- Der KITT-Erfahrungsaustausch dient dazu, bei der rasanten Entwicklung Schritt zu halten und die Neuigkeiten im Blick zu behalten. Mit den KI-News erfolgt auch eine monatliche Zusammenfassung in schriftlicher Form.
- Das KITT PromptTogether findet zwei Mal im Monat statt und dient dem praktischen Üben. Prompts sind die Spracheingaben, mit denen die KI angesteuert wird – je besser diese Prompts formuliert sind, desto brauchbarer sind die Antworten der KI. Hier wird gemeinsam, „geprompted“ und Mitarbeiter*innen der Stadt haben die Möglichkeit, ihre Anwendungsfälle herzuzeigen oder gemeinsam auszuprobieren.

Die Stadt Wien hat mit dieser Initiative die Grundlage für die Bildung einer KI-Community geschaffen – die auch in der Literatur als Best Practice genannt wird (Mader, 2024, S. 154 ff.). Die Früchte von mindestens einem Jahr Vorarbeit werden jetzt langsam sichtbar.

Die Stadt Wien setzt KI bereits seit Jahren ein – der WienBOT mit heute rund 16 Mio. gegebenen Antworten ist seit 2017 in Betrieb. Aber auch bei generativer KI ist Wien unter anderem mit KITT vorne dabei – und kann damit einen Teil des Lobes des Think Tanks der TU München und der Firma accenture auf sich beziehen: „Damit steht die öffentliche Verwaltung hier an vorderster Experimentierfront“ (TUM Think Tank & accenture, 2024, S. 8).

Der Digitale Humanismus lässt sich nicht nur auf Systeme mit künstlicher Intelligenz anwenden, gerade bei diesen mächtigen Instrumenten braucht es aber Leitlinien in der konkreten Anwendung. In welcher digitalen Welt wollen wir leben? Diese Fragestellung muss die Stadt Wien für sich auch im Umgang mit der KI beantworten. Der digitale Humanismus bietet uns hier einen hilfreichen Rahmen. Es zeigt auf, wie wir KI sozial gerecht und sinnvoll einsetzen wollen. Er ist eine Leitlinie, die sicherstellt, dass sich öffentliche Verwaltungen nicht ausschließlich mit Themen wie Optimierung, Effizienz und wirtschaftlichen Interessen auseinandersetzen.

3.6 Value Based Engineering bei der Stadt Wien

Die öffentliche Verwaltung hat – wie oben ausgeführt – per se den Auftrag, im Sinne der Menschen zu agieren. Dies ist jedoch in den relevanten IT-Projekten zu konkretisieren, weshalb beschlossen wurde, exemplarisch anhand von Beispielen zu verstehen, wie unsere Werte in IT-Systeme integriert werden können. Ein zentrales System der Stadtverwaltung ist mein.wien als Plattform für Bürgerinnen und Bürger. Digitale Systeme – und damit potentiell auch „mein.wien“ – haben gewollte und ungewollte Auswirkungen

auf grundlegende Werte und Rechte der Nutzerinnen und Nutzer. Die Stadt Wien hat sich daher mit der Frage auseinandergesetzt, wie man „mein.wien“ noch besser am Digitalen Humanismus ausrichten kann. Denn der Digitale Humanismus umfasst Gebote (etwa: Was sollte man tun, weil es den Bürger*innen einen Nutzen stiftet?) und Verbote (was sollte man lassen, weil der Nutzen zu gering ist/der Schaden überwiegt?). Genau diese ethische Bewertung wird auf Basis der weltweit anerkannte Norm ISO/IEEE 7000 vorgenommen. Dieser Standard ermöglicht eine ethische Analyse von Werten, die für die Wiener*innen rund um mein.wien von Bedeutung sein können – als „Value Based Engineering“. In einem dreitägigen Workshop wurden der Einstieg in mein.wien, die Verwaltung der eigenen Daten und die Nutzung von Amtswegen aus ethischer Sicht analysiert. Im Anschluss wurden identifizierte Werte wie Gleichheit, Würde oder Fairness in System-Anforderungen übersetzt, dies dient als Leitfaden für zukünftige technische Optimierungen (Stadt Wien/Digitales Wien 2024).

Um ein breites Bild auf die ethischen Fragen zu erhalten waren die Teilnehmerinnen und Teilnehmer des Workshops breit gestreut ausgewählt worden. Beteiligt waren unter anderem Vertreter*innen der Helfer Wiens, der Wiener Antidiskriminierungsstelle für LGBTIQ-Angelegenheiten, der MA 13, des Fachbereichs Datenschutz, E-Government und Informationsrecht der MA 63, des Fonds Soziales Wien, der Magistratsdirektion, des Wiener Gesundheitsverbundes und der Wiener Kinder- und Jugendhilfe (MA 11). Identifiziert wurden fünf Kernwerte (Inklusion, Privatsphäre, Selbstbestimmung, Transparenz und Vertrauen), die sich in 33 Wertqualitäten (bspw.: Zugang ist unabhängig von IT-Kompetenz möglich, Datensparsamkeit oder Vermeidung von Filterblasen) weiter unterteilen lassen. Die Ergebnisse des Vorhabens und darüberhinausgehende Ressourcen stehen in einem Abschlussbericht mit Zusammenfassung entlang des Standards ISO/IEC/IEEE 24748-7000 (IEEE 7000) zur Verfügung und können zur Weiterentwicklung von mein.Wien genutzt werden.

Dieses Vorgehen ist ein exemplarischer Ansatz, IT-Systeme systematisch menschenzentriert zu gestalten und so in einer großen Organisation auch entsprechende Erkenntnisse zu gewinnen und diese zu implementieren.

3.7 Internationaler Austausch

Neben den konkreten Anwendungen in der Stadt selbst – wie etwa KITT, KI-Kompass und Value Based Engineering – geht es auch darum, international voneinander zu lernen und gemeinsam Herausforderungen anzugehen. Die Stadt Wien ist daher eine der Leading Cities im Netzwerk CC4DR (Cities Coalition für digital Rights) die sich der Übersetzung des Digitalen Humanismus in die Verwaltung sowie der Förderung der digitalen Grundrechte verschrieben hat. Die Grundsätze der Koalition sind dabei:

1. Universeller und gleichberechtigter Zugang zum Internet und digitale Kompetenz.
2. Privatsphäre, Datenschutz und Sicherheit.

3. Transparenz, Rechenschaftspflicht und Nicht-Diskriminierung von Daten, Inhalten und Algorithmen.
4. Partizipative Demokratie, Vielfalt und Inklusion.
5. Offene und ethische Standards für digitale Dienste.

Wien arbeitet bspw. aktiv an der Frage der Weiterbildung im Rahmen einer Digital Rights Academy mit, denn letztlich stehen alle Städte international vor den gleichen Herausforderungen.

Auch andere internationale Formate werden für den Austausch genutzt. So führt Wien regelmäßig Vienna Science Days in anderen europäischen Städten durch, um die Wissenschaft und die Verwaltung städteübergreifend zu vernetzen. Bei den Vienna Science Days 2024 in Berlin stand auch der Digitale Humanismus auf dem Programm.

3.8 Eine europäische Lösung muss her

All diese Initiativen dürfen über eines nicht hinwegtäuschen: Europa ist derzeit abhängig von großen US-amerikanischen und/oder asiatischen Anbietern im Bereich der IT. Auch durch den Einsatz von Open Source Systemen etwa im Rechenzentrum der Stadt ändert sich an dieser Grundaussage nichts. Wien setzt Open Source breit ein – man kann das unter oss.wien.gv.at nachlesen – aber in vielen Bereichen gibt es keine oder keine gleichwertigen Software- oder Hardwarekomponenten, die Open Source sind und/oder aus Europa kommen. Wenn man diese Abhängigkeit etwa im Sinne einer erhöhten Resilienz in der „Zeitenwende“ (Olaf Scholz) verringern will, dann bedarf es einer europäischen Industriepolitik mit dem Ziel der Schaffung europäischer, skalierbarer Angebote. Die Regulierungen, die die EU in vielen Bereichen geschaffen hat (bspw. durch die Datenschutzgrundverordnung, den Digital Services Act, den Digital Markets Act, die Netzwerk- und Informationssicherheits-Richtlinie oder den AI Act) können hier nur der Anfang sein. Zu glauben, dass das zusammenfassen von (bestehenden) Open Source Lösungen das Problem löst, verkennet die enormen Investitionen und Innovationskraft, die anderswo in diesen Bereichen getätigt werden und auf europäischer Ebene nötig wären.

Die Methoden und Konzepte des Digitalen Humanismus können dabei helfen, eigene Entscheidungen und Umsetzungsvarianten besser beurteilen zu können. Im Rahmen eigener Handlungsmöglichkeiten kann IT daher präziser und besser im Sinne der Nutzerinnen und Nutzer umgesetzt werden. Dabei unterstützen (europäische) Regelungen, da sie – jedenfalls in der Theorie – einen großen Binnenmarkt bespielen, der auf Grund seiner Nachfragemacht für viele IT-Anbieter nicht zu ignorieren ist. All das reicht jedoch nicht aus. Wir müssen Rahmenbedingungen schaffen, damit die europäischen Lösungen auch weltweit konkurrenzfähig werden. Das ist die Aufgabe einer klassischen (europäischen) Industriepolitik.

Literatur

- Brodnig, I. (2016). *Hass im Netz. Was wir gegen Hetze, Mobbing und Lügen tun können*. Wien. Brandstätter
- Brodnig, I. (2018). *Lügen im Netz. Wie Fake News, Populisten und unkontrollierte Technik uns manipulieren* (2. Aufl.). Wien. Brandstätter
- Cockett, R. (2023). *Vienna – how the city of ideas created the modern world*. New Haven. Yale University Press
- Grunwald, A. (2022). *Technikfolgenabschätzung. Einführung* (3. Aufl.). Baden-Baden. Edition Sigma
- Harari, Y. N. (2024). *Nexus. A brief history of information networks from the stone age to AI*. New York. Penguin
- Himpele, K. (2022). Den technologischen Wandel gestalten. Die Verbesserung der Lebensbedingungen als zentrale Prämisse. In E. Kaiser & M. Schober (Hrsg.), *Digitale Wohlstandsgesellschaft. Der Weg in eine digitalisierte Zukunft* (S. 58–71). Wien. Verlag des Österreichischen Gewerkschaftsbundes
- Himpele, K. (2024). Generative KI und die öffentliche Verwaltung. In Wien Holding (Hrsg.), *Digitale Melange. Beiträge zur künstlichen Intelligenz aus Wirtschaft, Forschung, Politik und Verwaltung* (S. 16–22). Wien. Bohmann
- Krause, G. (2024). *Die Praxis des Digitalen Humanismus: Welchen Beitrag Unternehmen dazu leisten und wie sie davon profitieren können*. Wiesbaden: Springer Vieweg.
- Ludwig, M. (2023). Der Wiener Weg des digitalen Humanismus. In G. Krause (Hrsg.), *Die Praxis des Digitalen Humanismus. Welchen Beitrag Unternehmen dazu leisten und wie sie davon profitieren können* (S. 29–43). Wiesbaden. Springer
- Mader, I. (2024). *Next Generation Digital. Digitalisierung & KI strategisch einsetzen, mit 10 Best Practice Cases*. Wien. Excellence Edition
- Nassehi, A. (2024). *Kritik der großen Geste. Anders über gesellschaftliche Transformation nachdenken*. München. CH Beck
- Nida-Rümelin, J. (2022). Digital Humanism and the Limits of Artificial Intelligence. In Werthner, H. et al. (Hrsg.), *Perspectives on Digital Humanism* (S. 71–75). Cham. Springer.
- Pörksen, B. (2018). *Die große Gereiztheit. Wege aus der kollektiven Erregung*. München. Hanser
- Schrems, M. (2014). *Kämpf um Deine Daten*. Wien. edition a
- Stadt Wien / Digitales Wien (2024). *Workshop zum Value-Based Engineering*. Wien: Stadt Wien. <https://digitales.wien.gv.at/workshop-value-based-engineering/> (abgerufen am 5. April 2025)
- Stadt Wien (2019). *Künstliche Intelligenz Strategie*. Wien.
- Stadt Wien (2022). *Digitaler Humanismus in Wien*. Wien.
- Stadt Wien (2023). Kompass für den dienstlichen Umgang mit generativer Künstlicher Intelligenz. <https://digitales.wien.gv.at/ki-kompass-fuer-bedienstete-der-stadt-wien/>
- Stadt Wien (2024a). *Digitale Agenda 2030 der Stadt Wien. Der Mensch im Mittelpunkt der Digitalisierung*. Wien.
- Stadt Wien (2024b). *KI-Strategie der Stadt Wien. Teilstrategie der Digitalen Agenda 2030 der Stadt Wien*. Wien.
- Strassnig, M., Mayer, K., Stampfer, M., & Zingerle, S. (2019). *Akteure, Instrumente und Themen für eine Digital Humanism Initiative in Wien*. Studie im Auftrag der Stadt Wien – Magistratsabteilung 23. Wien. WWTF – Wiener Wissenschafts-, Forschungs- und Technologiefonds
- Suleyman, M. (2023). *The coming wave. Technology, power, and the 21st century's greatest dilemma*. New York. Crown
- TUM Think Tank, & accenture (2024). *Von Pionieren und Potenzialen. Transformation des öffentlichen Sektors mit GenAI*. München. TUM Think Tank, & accenture

- Volland, H. (2024). *Overload. Die KI-Medienflut kommt. Was ist noch echt, was Fake?* Hamburg. brand eins books
- Werthner, H., et al. (2022). *Manifesto on Digital Humanism*. Wien: Springer.
- Wolfangel, E. (2022). *ein falscher Klick. Hackern auf der Spur: Warum der Cyberkrieg uns alle betrifft*. München. Penguin
- Zuboff, S. (2018). *Das Zeitalter des Überwachungskapitalismus*. Frankfurt am Main, New York. Campus

Diplom-Volkswirt Klemens Himpele ist seit 2020 CIO der Stadt Wien. Davor leitete er acht Jahre lange die Abteilung Wirtschaft, Arbeit und Statistik (MA 23). Er wurde in Baden-Württemberg geboren, studierte Volkswirtschaftslehre an der Universität Köln und war danach zunächst wissenschaftlicher Mitarbeiter am Forschungsinstitut für Bildungs- und Sozialökonomie in Berlin. Von 2007 bis 2010 arbeitete Klemens Himpele als Projektleiter in der Bundesanstalt Statistik Austria in Wien. Zuletzt war er als Referent im Bereich Hochschule und Forschung im Hauptvorstand der Gewerkschaft Erziehung und Wissenschaft in Frankfurt am Main tätig.

CIO-Perspektive 2: Nachhaltige Datensouveränität 4

Souveränes Datenmanagement als Grundlage für Innovation und Fortschritt

Helmut Leopold und Lisa Höllbacher

Zusammenfassung

Der Beitrag beleuchtet die Bedeutung einer nachhaltigen Datensouveränität als Fundament für Innovation und wirtschaftliche Prosperität, sowie insbesondere für eine positive gesellschaftliche Weiterentwicklung. Die Autoren diskutieren, wie globale Technologie- und Plattformentwicklungen neue Ansätze der Datenkommunikation und -verarbeitung ermöglichen. Der Fokus liegt auf der Vermeidung von Datenverzerrungen und der Sicherstellung ethischer Grundwerte in datengetriebenen Systemen.

Die Verfügbarkeit großer Datenmengen unter Sicherstellung einer souveränen Nutzung der Daten ist ein grundlegender Innovationstreiber für alle Märkte, Industrien und Lebensbereiche der globalen Digitalwirtschaft.

4.1 Einleitung

Die globalen Technologie-, Infrastruktur und Plattformentwicklungen der letzten 30 Jahre schaffen die Grundlage für neue, innovative Ansätze der Datenkommunikation und -verarbeitung mit gravierenden disruptiven Effekten für unsere globale Wirtschaft und

H. Leopold (✉)
AIT Austrian Institute of Technology, Wien, Österreich
E-Mail: helmut.leopold@ait.ac.at

L. Höllbacher
Unbuzz Consulting, Wien, Österreich
E-Mail: lisa@unbuzz-consulting.com

Gesellschaft. Während Güter global gehandelt und verarbeitet werden, ist der Handel und der Austausch von Daten (Data Sharing) immer noch sehr begrenzt und wird von vielen Unternehmen skeptisch beurteilt. Dabei benötigen Unternehmen immer mehr Daten, die ursprünglich nicht mehr direkt aus ihrem Betrieb stammen, um die Produktqualität zu verbessern, die Produktionseffizienz zu steigern, oder die Kundenbetreuung zu optimieren. Darüber hinaus ermöglicht die Verknüpfung von Daten die Generierung neuer Informationen und Mehrwerte und bildet damit die Grundlage für neue datengetriebene Geschäftsmodelle. Schließlich ist die Verfügbarkeit großer Datenmengen für statistische Analysen und Ableitungen (Datenwissenschaften), aber auch die Basis für maschinelles Lernen für neue Anwendungen von sogenannter Künstlicher Intelligenz (KI) (Funk, 2024) ein grundlegender Innovationstreiber für alle Märkte, Industrien und Lebensbereiche der globalen Digitalwirtschaft.

Darüber hinaus hat die erreichte vollständige Vernetzung der Menschen auf unserem Planeten durch das globale Internet in Verbindung mit leistungsstarken mobilen Endgeräten für alle Teilnehmenden, mit denen auf einfachste Weise multimediale Inhalte generiert werden können, zur Entwicklung von sozialen Medienplattformen geführt, die die Art und Weise der Kommunikation und des Informationsaustausches völlig verändert und die globale Medienlandschaft vom früheren Massenmedium Fernsehen hin zum interaktiven Austausch von Inhalten auf den sozialen Medienplattformen transformiert haben.

4.2 Herausforderungen von digitalen Systemen

Durch diese disruptiven Entwicklungen ergeben sich für uns als Benutzende von digitalen Systemen grundlegende Herausforderungen, welchen wir uns durch einen nachhaltigen Dialog interdisziplinär stellen müssen:

- Verhinderung von Missbrauch und Sicherstellung ethischer Grundwerte in unseren datengetriebenen technischen Systemen
- Sicherstellung einer nachhaltigen Datensouveränität als Grundlage für Innovation und wirtschaftliche Prosperität
- Bildung und digitale Werkzeuge für einen verantwortungsvollen und vertrauensvollen Umgang mit der digitalen Informationsflut

Diese Aspekte werden in den folgenden Abschnitten näher erörtert.

Verhinderung von Missbrauch und Sicherstellung ethischer Grundwerte in unseren daten-getriebenen technischen Systemen

Die schier unbegrenzten Möglichkeiten unserer digitalen Existenz als Folge der rasanten technologischen Entwicklung verstellen oft unseren Blick auf parallel verlaufende Fehlentwicklungen, die unseren gesellschaftlichen Zusammenhalt, unsere soziale Empathie

und letztlich auch unsere Freiheit und Menschenwürde in nie zuvor gekannter Weise bedrohen. Eine kaum noch zu bändigende Markt- und Verhandlungsmacht globaler Social-Media-Plattformen stellt im zeitlichen Zusammenfallen mit großen geopolitischen Umbrüchen eine der großen Herausforderungen für unsere Gesellschaften dar. Damit stellt sich für uns die Frage, wie wir in diesem Umfeld einen digitalen Humanismus nachhaltig sichern können, in dem moralische und ethische Grundwerte nicht durch die Technik untergraben werden (Grabuschnig et al., 2023).

Ein entscheidender Punkt in der digitalen Kommunikationswelt ist, wie Datenverzerrungen – also Data Bias – unsere Informationsverbreitung und die öffentliche Meinungsbildung beeinflussen. Wenn Datensätze nicht repräsentativ sind oder systematische Fehler aufweisen, können bestimmte Gruppen unbewusst benachteiligt oder bevorzugt werden. In Kombination mit einseitigen und für uns oft nicht transparenten Algorithmen, die unser digitales Verhalten laufend analysieren und interpretieren, führt dies zu potentiellen Manipulationen der digitalen Nutzer:innen.

In den sozialen Medien führt dies oft dazu, dass manche Inhalte übermäßig hervorgehoben werden, während andere unter den Tisch fallen. So entstehen Echokammern und Filterblasen, in denen wir vor allem das sehen, was unsere bestehenden Überzeugungen bestätigt. Das Ergebnis ist ein Verlust an Vielfalt, die für Innovation und Weiterentwicklung so grundlegend wichtig ist, eine verstärkte Polarisierung in der Gesellschaft und eine erhöhte Anfälligkeit für Desinformation und Fake News.

Für CIOs und CDOs ist es daher von entscheidender Bedeutung, die Auswirkungen von Datenverzerrungen auf ihre Informationssysteme und Kommunikationskanäle zu erkennen und ihnen aktiv entgegenzuwirken. Dies bedeutet, Mechanismen einzuführen, die Data Bias in allen Phasen der Datenverarbeitung identifizieren, messen und minimieren. Durch den Einsatz von Algorithmen, die auf Fairness, Transparenz und Verantwortlichkeit ausgelegt sind, können Verzerrungen in Datensätzen erkannt und korrigiert werden. Regelmäßige Audits und Überprüfungen der genutzten Datenquellen und Modelle sind unerlässlich, um unbeabsichtigte Diskriminierung zu vermeiden. Darüber hinaus sollten Unternehmen Richtlinien und Schulungen anbieten, die das Bewusstsein für Datenverzerrungen schärfen und einen verantwortungsvollen Umgang mit Daten fördern.

Sicherstellung einer nachhaltigen Datensouveränität als Grundlage für Innovation und wirtschaftliche Prosperität

Angesichts immer kürzerer Innovationszyklen und der zunehmenden Dynamik neuer Datenmärkte und der immer umfassenderen Nutzung von Daten für die in aller Munde befindlichen neuen Anwendungsszenarien Künstlicher Intelligenz brauchen wir neue Ansätze, um unsere ethischen Werte, rechtsstaatlichen Prinzipien und die Mechanismen freier Demokratien als Gestaltungsmaxime mitzubedenken (Leopold, 2018).

Ein verantwortungsvoller Einsatz von digitalen Technologien wird dann sichergestellt, wenn man sich bei aller Euphorie über das Machbare stets am menschlichen Wohl orientiert. Man könnte auch sagen: Naturwissenschaftliche Brillanz braucht das geisteswissenschaftliche Korrektiv.

Um das volle Potenzial von Künstlicher Intelligenz wirklich ausschöpfen zu können, müssen wir über den reinen Besitz von Daten hinausdenken und uns damit befassen, wie wir Daten teilen und nutzen. Hier kommen die sogenannten Datenräume (Data Spaces) ins Spiel. Datenräume sind kollaborative Datenökosysteme bzw. Datenmärkte, in denen verschiedene Akteure – Unternehmen, Forschungseinrichtungen, öffentliche Institutionen – ihre Daten sicher, transparent und souverän miteinander austauschen können. Datenräume sind ein wichtiges Konzept zur Unterstützung der EU-Strategie zur Datensouveränität (EC Data Space, 2022). Sie ermöglichen es uns, Daten aus unterschiedlichen Quellen zu verknüpfen und auf diese Weise völlig neue Erkenntnisse und Innovationen zu generieren. Für KI-Anwendungen sind Datenräume ein echter Game-Changer. Künstliche Intelligenz ist auf große Mengen qualitativ hochwertiger und vielfältiger Daten angewiesen. Je mehr und je unterschiedlichere Daten zur Verfügung stehen, desto präziser und verlässlicher werden die Ergebnisse der KI-Systeme. Durch die Teilnahme an Datenräumen können Unternehmen auf einen reichhaltigen externen Datenpool zugreifen, der ihre eigenen Datensätze ergänzt und erweitert.

Ein Austauschen von Daten in Datenräumen erfolgt allerdings unter Einhaltung von regulatorischen Rahmenbedingen, sowohl in kommerziellen Modellen als auch in solidarischen Ansätzen von Open-Data-Konzepten und „Data Donation“, d. h. der freiwilligen Datenbereitstellung für die Nutzung durch andere („Data Donation“ wird z. B. im Data Governance Act [DGA] für öffentliche Daten gefordert)¹.

Datenräume bieten aber noch mehr als nur den Zugang zu zusätzlichen Daten. Sie ermöglichen es Unternehmen, gemeinsam an Projekten zu arbeiten, Innovationspartnerschaften einzugehen und neue Geschäftsmodelle zu entwickeln. Beispielsweise könnten Unternehmen aus der Gesundheitsbranche anonymisierte Patientendaten sicher austauschen, um gemeinsam neue Behandlungsmethoden zu erforschen. Oder Hersteller in der Industrie könnten Produktionsdaten sicher und kontrolliert austauschen, um Lieferketten effizienter zu gestalten und Ressourcen zu schonen.

Verschiedene EU-Initiativen wie Gaia-X (Gaia-X, 2023), International Data Spaces Association (IDSA, 2024), Big Data Value Association (BDVA, 2024), Fireware (FIWARE Foundation, 2024), als auch direkt die EU-Kommission durch mehrere Förderprogramme und durch die Entwicklung von speziellen Open-Source Lösungen wie SIMPL (SIMPL, 2024; Goyet, 2024), arbeiten daran, solche sicheren und vertrauenswürdigen und vor allem verteilten Dateninfrastrukturen zu etablieren. Sie setzen auf europäische Werte wie Datenschutz, Datensouveränität und Compliance und entwickeln Standards und Technologien, die einen sicheren und vor allem kontrollierten souveränen Datenaustausch ermöglichen und vor allem die prinzipielle Unabhängigkeit und damit Souveränität als Dateneigentümer sicherstellen und Lock-In-Effekte durch monopolartige Datenverwaltungsstrukturen reduzieren.

Dabei werden Architekturen, Spezifikationen, Protokolle, aber auch Softwaremodule nach dem Open-Source Prinzip entwickelt, um eine technische Implementierung von

¹ Seit September 2023 formal in den EU Mitgliedsstaaten in gültiger Regulierung in Kraft.

souveränen Datenräumen zu unterstützen, Unternehmen den Einstieg in diese neue Datenwirtschaft zu erleichtern und die Marktdynamik zu beschleunigen. Gaia-X fokussiert dabei auf die Mechanismen der Authentifizierung und Einhaltung von Sicherheitsvorschriften beim Management der Daten (Trust Framework), und IDSA stellt Datenaustauschmodule (Connectors) zur Verfügung. Das Data Space Support Center (DSSC, 2024) ist eine von der EU geförderte Initiative zur Harmonisierung und Sicherstellung eines kontinuierlichen Informationsaustausches der mittlerweile zahlreichen Datenraumprojekte in den unterschiedlichsten Domänen in den verschiedenen EU-Mitgliedsstaaten.

Insgesamt werden damit in der EU über 3 Mrd. € in den Aufbau von Datenräumen investiert, sowohl durch die Mitgliedstaaten als auch durch die verschiedenen Förderprogramme wie Horizon und das Digital Europe Programm (DEP)².

Schließlich treibt insbesondere die Gaia-X Initiative die Harmonisierung der verschiedenen Aktivitäten zur Implementierung von Datenräumen voran und zielt auf die Entwicklung einer gemeinsamen Architektur und vor allem auf die notwendige Standardisierung ab.

Weiters bieten diese Entwicklungen eine wichtige Grundlage, um zukünftige neue EU Regulierungsansätze für eine souveräne und unseren demokratischen Prinzipien folgende Datenverwendung entsprechend effektiv technisch umzusetzen. Der Data Governance Act (DGA) sowie der kommende Data Act (DA) oder auch der neue EU AI Act definieren neue Pflichten und Rechte für unser Datenmanagement. Datenräume können damit auf effektive technische Plattformen bauen, welche genau diese Regularien entsprechend „by design“ berücksichtigen.

Zur Unterstützung des Wissensaufbaus in Österreich und vor allem der Klein- und Mittelbetriebe im Umgang mit diesen neuen Entwicklungen, hat das Bundesministerium für Innovation, Mobilität und Infrastruktur (BMIMI) und das Staatssekretariat für Digitalisierung im Bundeskanzleramt den Gaia-X Hub Austria eingesetzt. Der Hub bietet multimediale Lerninhalte an, um den Einstieg in die Welt der Datenräume zu erleichtern und eine raschere Entwicklung der Datenraum-Stakeholder-Community in Österreich zu forcieren. Ebenso wurde eine spezielle Data Space Testplattform implementiert, um entsprechende praktische Demonstrationen zu ermöglichen und KMUs eine Testplattform für neue Entwicklungen anzubieten. Für CIOs und CDOs ist es daher wichtig, diese Entwicklungen aktiv zu verfolgen und ihr Unternehmen entsprechend zu positionieren. Das bedeutet, die eigene IT-Infrastruktur so zu gestalten, dass sie die Teilnahme an Datenräumen ermöglicht und klare Richtlinien für den sicheren und verantwortungsvollen Umgang mit Daten zu etablieren.

Gleichzeitig müssen wir uns der Bedeutung von Datenbias bewusst sein, besonders wenn es um KI geht. Wenn die Daten, mit denen wir unsere KI-Systeme trainieren, Verzerrungen enthalten, spiegeln sich diese Vorurteile in den Ergebnissen wider. Das kann zu Diskriminierung führen und das Vertrauen in die Technologie untergraben. Deshalb ist es entscheidend, dass wir in den Datenräumen nicht nur die Verfügbarkeit von Daten in

² Präsentation am Gaia-X Summit 2024, November 14–15, 2024 in Helsinki, Finnland.

verteilten Systemen erreichen, sondern auch auf die Qualität und faire Benutzung dieser Daten achten. Regelmäßige Audits, der Einsatz von Fairness-Algorithmen³ und Schulungen des Teams können helfen, Datenbias zu erkennen und zu minimieren (HLEG, 2019; van den Hoven et al., 2015).

Bei all diesen rasanten technischen Entwicklungen dürfen wir den Menschen nicht aus den Augen verlieren. Das Konzept des „Human Dignity by Design“ erinnert uns daran, dass Technologien stets die Würde und die Rechte des Menschen respektieren sollten. Prof. Sarah Spiekermann von der Wirtschaftsuniversität Wien hat mit ihrem „Values-based Engineering“ (VbE) einen Ansatz entwickelt, der ethische Werte von Anfang an in den Entwicklungsprozess integriert (Spiekermann & Winkler, 2020). Anstatt Ethik erst im Nachhinein zu berücksichtigen, werden hier von Beginn an Werte wie Transparenz, Fairness und Datenschutz in die Technologie eingebettet.

Für CIOs und CDOs bedeutet das, dass sie in ihren Projekten nicht nur technische und wirtschaftliche Aspekte berücksichtigen sollten, sondern auch die ethischen Implikationen. Das kann bedeuten, Transparenz über die Funktionsweise von Algorithmen zu schaffen, den Datenschutz ernst zu nehmen und sicherzustellen, dass Technologien nicht zur Diskriminierung beitragen. Indem wir VbE und „Human Dignity by Design“ in unsere Arbeit integrieren, können wir Technologien entwickeln, die nicht nur innovativ, sondern auch verantwortungsvoll und menschenzentriert sind.

Insgesamt bieten Datenräume enorme Chancen für Innovation und Zusammenarbeit, um neue datengetriebene Geschäftsmodelle zu entwickeln, sowie neue effektive und vor allem verantwortungsvolle KI-Lösungen zu erhalten (Leopold, 2023a). Sie ermöglichen es uns, die Kraft der Daten zu nutzen, um neuartige Projekte zu realisieren und gesellschaftliche Herausforderungen anzugehen. Aber es liegt an uns, diese Möglichkeiten verantwortungsvoll zu nutzen und sicherzustellen, dass der Mensch immer im Mittelpunkt steht. Indem wir uns auf hochwertige und möglichst Bias-freie Daten konzentrieren, eine vertrauensvolle und faire sowie souveräne Datenverwendung sicherstellen, aktiv an Datenräumen teilnehmen und ethische Prinzipien in unsere Arbeit einbeziehen, können wir eine digitale Zukunft gestalten, von der alle profitieren.

Bildung und digitale Werkzeuge für einen verantwortungsvollen und vertrauensvollen Umgang mit der digitalen Informationsflut

Die enorm große Menge an Informationen und die unüberschaubare Vielfalt an Informationsquellen im Internet, mit denen wir laufend konfrontiert werden, machen es für uns immer schwerer, die Vertrauenswürdigkeit von Inhalten im Netz zu beurteilen. Wir alle sind zu Medienproduzentinnen und -produzenten geworden und partizipieren am Segen des freien und offenen Kommunikationsprozesses im Internet. Es wird aber immer schwieriger, in der Fülle der ständig neu produzierten Informationen, echte von gefälsch-

³ Algorithmische Fairness beschreibt einen Prozess, der sicherstellt, dass Algorithmen und ihre Ergebnisse unparteiisch sind. Algorithmen dürfen Einzelpersonen oder Gruppen nicht aufgrund sensibler Merkmale wie z. B. Rasse, Geschlecht oder Alter diskriminieren.

ten Inhalten zu unterscheiden. Dies ist für uns aber unerlässlich, um Manipulationsversuche unseres Verhaltens und unserer Einstellungen klar zu erkennen. Auch wenn es dafür Regulierungen und gesetzliche Vorschriften gibt, müssen wir zusätzlich sicherstellen, dass diese auch eingehalten werden.

Angesichts der großen Informationsmengen und der Geschwindigkeit, mit der neue Informationen auftauchen, ist die Verwendung von digitalen Hilfswerkzeugen unerlässlich. Digitale Werkzeuge in diesem Bereich setzen wiederum auf KI-Lösungen, um Veränderungen an Fotos und Videos zu erkennen, Texte analysieren zu können, um rassistische, oder moralisch verwerfliche Inhalte rasch zu identifizieren, und auch um sehr echt wirkende, von KI generierte Deep-Fake-Inhalte, eindeutig erkennen zu können.

Technologische Kompetenzen und moderne digitale Werkzeuge sind eine wichtige Grundlage für eine digitale Souveränität und Datensouveränität und sind für eine verantwortungsvolle journalistische Arbeit ebenso unerlässlich wie für eine effektive Arbeit der Behörden, um in der neuen Medienwelt nicht in eine verhängnisvolle Abhängigkeit von anderen Staaten oder Monopolbetreibern zu geraten.

Um einen digitalen Humanismus im alltäglichen Umgang mit den Medien nachhaltig sicherzustellen, bedarf es daher einerseits einer grundlegenden Medienkompetenz für Alle mit dem Verständnis für die Bedeutung von Vielfalt und der ständigen Überprüfung von Informationen mit alternativen Informationsquellen. Andererseits braucht es auch Werkzeuge für jede und jeden digitalen Nutzer:in, die bei der Verifikation von Inhalten unterstützen und so einen wichtigen Beitrag zu einem medienkompetenten Verhalten für Alle leisten (Leopold, 2023b).

So gibt es beispielsweise mit dem vielfach ausgezeichneten Fake-Shop-Detector eine in Österreich entwickelte und international führende KI-Lösung, mit der Online-Nutzer:innen vor betrügerischen Online-Shops geschützt werden⁴ (Fake-Shop Detector, 2025).

Es liegt an uns, die enormen Möglichkeiten der digitalen Transformation als Mehrwert für uns alle zu nutzen, dabei aber sicherzustellen, dass der Mensch immer im Mittelpunkt steht. Dies erreichen wir, indem wir die Technik kontinuierlich so gestalten, dass sie verstanden wird und sich unsere Grundwerte, aber auch gesetzliche Vorgaben inhärent und „by-design“ in der Technik widerspiegeln, und indem wir die Technik so souverän beherrschen, dass wir sie kontinuierlich an unsere neuen Anforderungen adaptieren und stets als effektives Werkzeug einsetzen können.

⁴ Der Fake-shop-Detector ist eine KI Entwicklung des österreichischen Konsortiums mit AIT Austrian Institute of Technology, X-Net und dem Österreichisches Institut für angewandte Telekommunikation (ÖIAT).

4.3 Fazit

Die digitale Transformation hat unsere Welt grundlegend verändert und bietet immense Chancen für Innovation und Fortschritt. Doch mit diesen Möglichkeiten kommen auch Herausforderungen, die wir aktiv angehen müssen. Ein verantwortungsvoller Umgang mit Daten, die Sicherstellung ethischer Grundwerte und die Förderung von Medienkompetenz sind entscheidend, um die Vorteile der digitalen Welt voll auszuschöpfen und gleichzeitig ihre Risiken zu minimieren. Indem wir uns auf hochwertige und möglichst bias-freie Daten konzentrieren, eine vertrauensvolle und faire sowie souveräne Datenverwendung sicherstellen und ethische Prinzipien in unsere Arbeit einbeziehen, können wir eine digitale Zukunft gestalten, von der alle profitieren. Es liegt an uns, die Technik so zu gestalten, dass sie unsere Grundwerte und gesetzlichen Vorgaben widerspiegelt und uns als effektives Werkzeug dient, um die enormen Möglichkeiten der digitalen Transformation als Mehrwert für uns alle zu nutzen. Der Mensch muss dabei immer im Mittelpunkt stehen.

Literatur

- BDVA – Big Data Value Association (2024). Offizielle Webseite. Brüssel: BDVA. <https://bdva.eu/>. Zugriffen: 6. Apr. 2025.
- DSSC – Data Spaces Support Centre (2024). Offizielle Webseite. Brüssel: Europäische Kommission / Data Spaces Support Centre. <https://dssc.eu/>. Zugriffen: 6. Mai 2025.
- EC Data Space – European Commission Staff working document on data spaces. <https://digital-strategy.ec.europa.eu/en/library/staff-working-document-data-spaces> Erstellt: 14. Febr. 2022; zugriffen: 23. Nov. 2024.
- Fake-Shop Detector Fake-Shop Detector – Schutz vor betrügerischen Online-Shops. Wien: ÖIAT / Forschungsprojekt SINBAD. <https://www.fakeshop.at/>. Zugriffen: 6. Mai 2025.
- FIWARE Foundation (2024). Über uns – FIWARE (About Us). <https://www.fiware.org/about-us/>. Zugriffen: 6. Mai 2025.
- Funk, M. (2024). *Künstliche Intelligenz, Verkörperung und Autonomie*. Wiesbaden: Springer. ISBN 978-3-658-41105-3.
- Gaia-X – European Association for Data and Cloud AISBL (2023). Offizielle Webseite. Brüssel: Gaia-X. <https://gaia-x.eu/>. Zugriffen: 6. Apr. 2025.
- Goyet, M. (2024). Manuel: SIMPL – Vision and Engagement. Vortrag, European Commission DG Connect E2. https://www.data-spaces-symposium.eu/wp-content/uploads/2024/04/13.55-Manuel-Mateo-Goyet-SIMPL_EC-Slides_DSSCSymposium24_Final.pdf. Zugriffen: 23. Nov. 2024.
- Grabuschnig, W., Leopold, H., & Mürling, M. Safeguarding European Values and Global Competitiveness in the Digital Space. White Paper, A Gaia-X Alpbach Lab elaboration for shaping the EU data strategy. <https://www.gaia-x.at/wp-content/uploads/2024/01/EFA23-AIT-LAB-Digital-Sovereignty-Review.pdf> (Erstellt: 08.2023). Zugriffen: 23. Nov. 2024.
- HLEG – High-Level Expert Group on Artificial Intelligence Ethics guidelines for trustworthy AI. Europäische Kommission. <https://digital-strategy.ec.europa.eu/en/library/ethics-guidelines-trustworthy-ai>. Erstellt: 8. Apr. 2019; zugriffen: 23. Nov. 2024.
- IDSA – International Data Spaces Association (2024). Offizielle Webseite. Dortmund: IDSA. <https://internationaldataspaces.org/>. Zugriffen: 6. Apr. 2025.

- Leopold, H. (2018). Is AI a safe Technology, and Can We Trust it? In H. Androsch, W. Knoll & A. Plimon (Hrsg.), *Discussing Technology – Artificial Intelligence*. Alpbach Technology Symposium Yearbook 2018. Wien: Holzhausen Verlag. ISBN 978-3-903207-24-0.
- Leopold, H. (2023a). Mastering Trustful Artificial Intelligence. In R. Schmidpeter & R. Altenburger (Hrsg.), *Responsible Artificial Intelligence: CSR, Sustainability, Ethics & Governance*. Cham: Springer. https://doi.org/10.1007/978-3-031-09245-9_6. ISBN 978-3-031-09244-2.
- Leopold, H. (2023b). Digitale Souveränität für eine resiliente Gesellschaft. In: OVE Newsletter Social Media. <https://www.ove.at/ove-news/details/digitale-souveraenitaet-fuer-eine-resiliente-gesellschaft/> (Erstellt: 22. Nov. 2023). Zugriffen: 23. Nov. 2024.
- SIMPL (2024). SIMPL: Cloud-to-edge-Föderationen zur Stärkung der EU-Datenräume. Brüssel: Europäische Kommission. <https://digital-strategy.ec.europa.eu/en/policies/simpl>. Zugriffen: 6. Mai 2025.
- Spiekermann, S., & Winkler, T. (2020). Value-based Engineering for Ethics by Design. *SSRN Electronic Journal*. <https://doi.org/10.2139/ssrn.3598911>.
- van den Hoven, J., Vermaas, P. E., & van de Poel, I. (Hrsg.). (2015). *Handbook of ethics, values, and technological design*. Dordrecht: Springer.

Dr. Helmut Leopold hat über 30 Jahre Managementenerfahrung. In seiner Zeit als Technologiechef der Telekom Austria war er als Infrastrukturverantwortlicher der wesentliche Innovator für die digitale Transformation. Der Digitalstandort Österreich, wie wir ihn heute kennen, trägt in hohem Maße seine Handschrift. Derzeit leitet er am AIT Austrian Institute of Technology das Center Digital for Safety & Security. Unter seiner Leitung hat sich das Center im Bereich Cyber Security, Künstliche Intelligenz und Quantentechnologien zu einem anerkannten globalen Technologieführer entwickelt. Leopold absolvierte das Informatikstudium an der TU Wien und promovierte zum Doktor der Philosophie in Computer Science an der Universität Lancaster in England.

Lisa Höllbacher ist Gründerin von Unbuzz Academy & Consulting, spezialisiert auf Technologie-Transformation, Workshops und Keynotes zu Daten und KI. Zuvor war sie CEO & Co-Founder beim SaaS-Startup nexyo für dezentrales Data Sharing. Ihre Beratungs-Karriere startete sie 2018 bei Pioneers als Senior Innovation Consultant. An der WU Wien und der UNU Merit/Maastricht University spezialisierte sie sich auf Public Policy, Social Innovation Design und Innovation Financing. Berufliche Einblicke sammelte sie u. a. bei PwC Advisory Austria, Pioneers und als Research Assistant an der United Nations University MERIT. Ehrenamtlich engagiert sie sich u. a. im Club Alpbach Salzburg, bei Austrian Startups und in der Bundesjugendvertretung.

CIO-Perspektive 3: Digitaler Humanismus im Projektmanagement

5

Ethische Prinzipien von der Planung bis zur Umsetzung im Projekt verankern

Mario Tokarz

Zusammenfassung

Der Beitrag beleuchtet **die Integration wertorientierter Ansätze in das Projektmanagement**. Dabei wird aufgezeigt, wie ethische Prinzipien mithilfe des bewährten Plan-Do-Check-Act-(PDCA)-Ansatzes systematisch in die einzelnen Phasen von IT-Projekten integriert werden können, um Spannungen zwischen Produktanforderungen und ethischen Überlegungen auszugleichen. Im Mittelpunkt stehen praxisorientierte Methoden sowie anschauliche, anwendungsnahe Use Cases.

Technologie darf nicht nur funktional sein – sie muss im Einklang mit menschlichen Werten gestaltet werden.

5.1 Einleitung

Die Veränderungen durch neue Technologien, allen voran die künstliche Intelligenz (KI), prägen unsere Zeit wie kaum ein anderes Thema. Die Diskussion um ihre Folgen wirft zentrale gesellschaftliche Fragen auf: Welche Grenzen hat die Meinungsfreiheit, etwa auf Plattformen wie Facebook oder TikTok? Und wie wird sich die Rolle der menschlichen Arbeitskraft in einer zunehmend automatisierten Welt verändern? Diese Fragen verdeutli-

M. Tokarz (✉)
RightMinded AI, München, Deutschland
E-Mail: mario.tokarz@rightminded.ai

chen, dass technologische Innovationen nicht nur Chancen bieten, sondern auch ethische Herausforderungen mit sich bringen, die es zu adressieren gilt.

Um den digitalen Humanismus in Projekten zu verankern, müssen Projektleiter*innen eine klare Verantwortung übernehmen: Sie sind gefordert, bei der Entwicklung von IT-Systemen ethische Faktoren strukturiert mit einzubeziehen. Spannungen zwischen Produktanforderungen und ethischen Belangen müssen im Sinne der Stakeholder*innen fair und lösungsorientiert ausgeglichen werden.

Doch wie lässt sich dieser Ausgleich in der Praxis realisieren? Klar ist: Es erfordert Fokus. Während auf gesellschaftlicher Ebene über die großen Fragestellungen diskutiert wird, liegt es an den Projektleiter*innen die Balance zu wahren: Einerseits müssen sie exzellente und ethisch fundierte Produkte entwickeln, andererseits sind die zeitlichen und Ressourcen bezogenen Rahmenbedingungen einzuhalten.

Dieser Ausgleich gelingt nur mit einem pragmatischen Vorgehen: Nicht alle großen Fragestellungen müssen vorab gelöst sein – oft wäre dies auch nicht möglich. Vielmehr ist es entscheidend, dass Führungskräfte und Projektleiter*innen die verfügbaren Lösungsansätze verstehen und diese konsequent auf die Herausforderungen im Projekt anwenden. Mit klaren Methoden und einem strukturierten Vorgehen kann der digitale Humanismus im Projektalltag integriert und professionell gemanagt werden.

Zum Aufbau dieses Artikels

Der vorliegende Artikel fasst erste Erkenntnisse und Lessons Learned aus Projekten zusammen, die basierend auf den Prinzipien des digitalen Humanismus umgesetzt werden. Die Inhalte orientieren sich an der bewährten Struktur von Management Systemen und folgen dem PDCA-Zyklus (*Plan, Do, Check, Act*).

Die vier Phasen sind wie folgt ausgerichtet:

- **Projektplanung (Plan):** In der Planungsphase werden die Grundlagen für die Integration des digitalen Humanismus in das Projekt geschaffen. Dazu gehört ein klares Bild des organisatorischen Kontextes, sowie die Definition notwendiger Kompetenzen und Ressourcen.
- **Projektdurchführung (Do):** Hier geht es um die konkrete Umsetzung, etwa durch die Anwendung von Vorgehensmodellen wie Value-based Engineering (VBE) und weiteren geeigneten Methoden.
- **Überprüfung (Check):** Diese Phase umfasst die Überprüfung der Umsetzung der ethischen Anforderungen, sowie das fortlaufende Monitoring. Das Monitoring ist insbesondere bei KI-basierten Produkten essentiell.
- **Verbesserung (Act):** Erkenntnisse aus der Entwicklung und dem Betrieb sollen genutzt werden, um das Produkt und auch die Fähigkeiten der Organisation nachhaltig zu verbessern.

Der Schwerpunkt liegt dabei auf der praktischen Umsetzung im Rahmen einzelner Projekte, wobei relevante Bezüge zur organisatorischen Ebene hervorgehoben werden.

Die genannten Beispiele beziehen sich häufig auf KI-Systeme, da diese nahezu immer ethische Fragestellungen aufwerfen und somit auch ein Treiber für Regulatorik, wie etwa den EU AI Act, sind. KI-Technologien, wie Large Language Models, stehen im Zentrum der medialen Berichterstattung und der öffentlichen Diskussion. Dennoch ist es wichtig zu betonen, dass auch IT-Systeme ohne KI-Komponenten ethische Herausforderungen mit sich bringen können. Ein prägnantes Beispiel hierfür ist der „Like-Button“ in sozialen Medien: Das Ausbleiben von „Likes“ kann bei Nutzer*innen starke emotionale Reaktionen hervorrufen. Dies macht deutlich, wie selbst ein simpler „Zähler“ Einfluss auf das persönliche Wohlergehen und auch auf die soziale Dynamik nehmen kann.

Es sei an dieser Stelle betont, dass dieser Artikel keinen vollständigen oder gar umfassenden Ansatz bieten kann. Vielmehr ergänzt er Ansätze wie Value-based Engineering (VBE), indem er die übergeordnete Ebene des Gesamtprojektes in den Fokus rückt. Ziel ist es, praxisnahe Erfahrungen und Einsichten zusammenzufassen, die Projektleiter*innen dabei unterstützen, die Prinzipien des digitalen Humanismus strategisch und strukturiert in ihren Projekten zu verankern.

5.2 Die Projektplanung

Projektleiter*innen stehen in der Planungsphase vor der Herausforderung, trotz unsicherer Datenbasis einen Projektauftrag abzustimmen, der eine tragfähige Grundlage für den Projekterfolg schafft. Eine sorgfältige Planung und Vorbereitung auf mögliche Risiken und Herausforderungen sind entscheidend, um die spätere Umsetzung zu erleichtern. In diesem Abschnitt werden zentrale Aspekte vorgestellt, die bei der Planung eines Projekts berücksichtigt werden sollten, um die Prinzipien des digitalen Humanismus erfolgreich zu integrieren.

5.2.1 Den Entwicklungs-Kontext verstehen

Ein tiefgehendes Verständnis des Kontexts, in dem die Produktentwicklung stattfindet, ist zentral für die erfolgreiche Umsetzung des digitalen Humanismus. Dabei geht es nicht nur um die technischen Anforderungen. Projektleiter*innen müssen genauso die organisatorischen und lokalen Rahmenbedingungen betrachten, da diese die Herausforderungen und entsprechenden Überlegungen im Projektverlauf maßgeblich beeinflussen.

5.2.2 Die Rahmenbedingungen frühzeitig verstehen

Projekte werden nie isoliert durchgeführt, sondern sind stets in einen größeren Rahmen eingebettet. Dieser wird unter anderem durch die durchführende Organisation, ortsspezifische und rechtliche Rahmenbedingungen, sowie weitere interne und externe Faktoren

bestimmt. Diese Rahmenbedingungen beeinflussen die Projektplanung und -durchführung erheblich. Projektleiter*innen müssen sie daher frühzeitig analysieren und gezielt in die eigene Planung mit einbeziehen.

Für die organisatorische Ebene bietet der Standard ISO/IEC 42001 Anhaltspunkte, wie der organisatorische Kontext beschrieben werden kann. Er unterscheidet zwischen:

- Externe Faktoren: Dazu gehören rechtliche Anforderungen, kulturelle Rahmenbedingungen und das Wettbewerbsumfeld.
- Interne Faktoren: Hierzu zählen organisatorische Strukturen, vertragliche Verpflichtungen, sowie der Zweck der Entwicklung und der Nutzung eines geplanten (KI-)Systems.

Sofern die eigene Organisation bereits ein Management System implementiert hat, kann auf die darin enthaltene Beschreibung des organisatorischen Kontexts zurückgegriffen werden. Andernfalls empfiehlt es sich, eigenständig tätig zu werden und relevante Aspekte dieses Kontextes zu dokumentieren. Solche Annahmen schaffen eine wichtige Grundlage für die Planung und können den Projektauftrag durch eine klare Definition von Rahmenbedingungen ergänzen.

Die Kontextdefinition der ISO/IEC 42001 beinhaltet bereits ortsspezifische Faktoren, wie etwa rechtliche Rahmenbedingungen. Es ist jedoch essenziell, dass neben dieser organisatorischen Perspektive auch eine spezifische projekthafte Sicht eingenommen wird. Hierbei sollten Faktoren wie die Zielmärkte des Produkts und insbesondere ortsspezifische kulturelle, sowie rechtliche Gegebenheiten mit einbezogen werden.

Diese mehrdimensionale Betrachtung hilft, mögliche Herausforderungen frühzeitig zu identifizieren und gezielt anzugehen.

5.2.3 Annahmen über das Produkt treffen

Der digitale Humanismus geht über Mindestanforderungen, wie sie sich etwa aus den Gesetzen ergeben, hinaus. Ziel ist es, mögliche negative Effekte von Technologie so früh wie möglich zu verstehen und diese – soweit möglich – zu vermeiden.

Mit Methoden wie VBE stehen Werkzeuge zur Verfügung, mit denen ethische Risiken detailliert analysiert werden können, insbesondere durch die Befragung von Stakeholder*innen. Solche Analysen setzen jedoch eine gewisse Planungsreife im Bezug auf das zu entwickelnde System (in VBE: System-of-Interest), weitere verbundene Systeme (in VBE: System-of-Systems) und die zu implementierenden Use Cases voraus.

In der Planungsphase vor dem Kick-off eines Projekts liegen diese Informationen oftmals nicht in der notwendigen Tiefe vor. Gleiches gilt häufig für die Details einer zu entwickelnden Funktion. Unternehmen nutzen daher häufig Werkzeuge wie etwa Mock-Ups oder Click-Dummys – vereinfachte, grafische Modelle, die einen ersten Eindruck des zukünftigen Produkts vermitteln. Darauf aufbauend können detailliertere Use Cases und Systementwürfe abgeleitet werden.

Solche Modelle sind oft technologie-agnostisch, da die Auswahl der Zieltechnologie sowie der Umfang von Eigenentwicklung und Zukauf typischerweise Teil der späteren Umsetzung sind. Dies führt zu einem scheinbaren „Henne-Ei-Problem“: Die ethische Analyse benötigt Input aus dem Systementwurf, während der Systementwurf seinerseits Input aus der ethischen Analyse benötigt.

Um diese Herausforderung zu bewältigen, muss das Projekt mit fundierten Annahmen arbeiten. Ein „wahrscheinlicher Systementwurf“, der alle verfügbaren Informationen zusammenfasst, sollte im Projekt explizit verankert und iterativ verfeinert werden. Er soll alle Informationen enthalten, die auch Teil einer Systembeschreibung nach VBE sind. In Verbindung mit Informationen zu möglichen Stakeholder*innen, können geschulte Projektmitarbeiter*innen bereits Überlegungen zu ethischen Fallstricken in der Systemgestaltung anstellen.

Soll in einem Projekt KI eingesetzt werden, ist es sinnvoll, zusätzliche Annahmen zu dokumentieren, wie zum Beispiel:

- Welche Teilfunktionen sollen durch KI realisiert werden?
- Warum ist KI für diese Funktion besonders geeignet?
- Gibt es nicht-KI-basierte Alternativen?
- Welche KI-Technologie kommt zum Einsatz (z. B. neuronale Netze)?
- Wird die KI als Eigenentwicklung umgesetzt oder extern zugekauft?

Diese Annahmen ermöglichen es, Hypothesen über den KI-Einsatz zu formulieren, die im Verlauf der Projektentwicklung weiter verfeinert oder hinterfragt werden können. Der geplante Einsatz von KI, insbesondere in Verbindung mit personenbezogenen Daten, wirft häufig moralische Fragen auf. Daher lohnt es sich, bereits in der frühen Planungsphase präzise Annahmen zu treffen.

In der Praxis besteht die Versuchung, Problemstellungen mit komplexen Technologien wie neuronalen Netzen (Deep Learning) zu lösen, obwohl einfachere Modelle oder klassische Methoden gleichwertige Ergebnisse liefern können. Solche simpleren Ansätze sind in der Regel mit weniger ethischen Herausforderungen verbunden – etwa hinsichtlich ihrer Erklärbarkeit oder des Energieaufwands für Training und Betrieb. Das bewusste Prüfen von KI-Alternativen ist daher nicht nur sinnvoll, sondern oft auch effizienter und nachhaltiger.

5.2.4 Kompetenz aufbauen

Die Leitung von Projekten erfordert neben handwerklichen Fähigkeiten vor allem Erfahrung. Erfahrung hilft, Situationen und Risiken besser einzuschätzen. Neue Technologien werden jedoch mit einer Geschwindigkeit entwickelt, die das Zurückgreifen auf Erfahrung unmöglich macht.

Besonders deutlich wird dies am Beispiel des Datenschutzes: Der Schutz der Privatsphäre ist bereits seit vielen Jahrzehnten ein in der breiten Öffentlichkeit beachteter Wert.

Seitdem sind zahlreiche Standards und Best-Practices entstanden, die Projektleiter*innen vertraut sind. Trotzdem bleibt die Umsetzung anspruchsvoll, gerade angesichts des technologischen Wandels.

Im Gegensatz dazu stehen Werte wie die Erklärbarkeit von Algorithmen oder die Freiheit von Bias, zu denen es bisher kaum Erfahrung gibt. Projektleiter*innen können daher nicht darauf vertrauen, solche Herausforderungen im Projektverlauf intuitiv zu erkennen. Dies macht den gezielten Kompetenzaufbau – sowohl individuell als auch im Team – absolut unverzichtbar.

Drei zentrale Rollenprofile zeichnen sich aus der Perspektive des digitalen Humanismus ab:

- **Methodikkenntnisse:** Ethische Herausforderungen können nicht basierend auf Bauchgefühl identifiziert oder gelöst werden. Die Betrachtung muss entlang klarer Methoden wie VBE erfolgen. Detaillierte Kenntnisse solcher Vorgehensmodellen, sind daher für Projektleiter*innen und Menschen in Schlüsselrollen unerlässlich.
- **Ethikkenntnisse:** Prozesswissen allein reicht nicht aus, um komplexe Zielkonflikte aufzulösen. Ein einfaches Beispiel eines solchen Konfliktes ist eine mögliche Spannung zwischen unterschiedlichen Werten, wie etwa Privatsphäre und Transparenz. Im Standard ISO/IEC/IEEE 24748-7000 ist daher die Rolle eines Value Leads definiert. Diese Person muss sowohl über methodisches als auch über ethisches Fachwissen verfügen. Sie ist für die Lösung von Zielkonflikten nicht etwa allein verantwortlich, führt aber sehr wohl im Projekt die Lösungsfindung.
- **Technologiekenntnisse:** Für Entwickler*innen sind Kenntnisse zu technologischen Aspekten ethischer Werte entscheidend. Für bestimmte Werte, wie die Freiheit von Bias, existieren gesetzliche Vorschriften, aber auch technische Standards. Allein die Analyse eines Algorithmus auf möglichen Bias kann eine anspruchsvolle Aufgabe sein. Kenntnis zu existierenden gesetzlichen Anforderungen und Lösungsansätzen, wie sie etwas in neuen Standards vorgeschlagen werden, können daher sinnvoll sein.

Soweit Organisationen Anbieter oder Betreiber von KI-Systemen im Sinne des EU AI Acts sind, müssen diese zusätzlich für ausreichende KI-Kenntnisse der eigenen Belegschaft Sorge tragen. Diese Anforderung richtet sich zwar primär an die Organisation als Ganzes, sollte jedoch auch Projektleiter*innen bekannt sein. Sie kann als Mindestanforderung verstanden werden, die jeder Mitarbeiter*in im Projekt erfüllen sollte.

5.2.5 Projekt- und Ressourcenplanung durchführen

Die Projekt- und Ressourcenplanung bildet den Kern der Aufgabe von Projektleiter*innen. Aus den im vorherigen Abschnitt beschriebenen Überlegungen lassen sich unmittelbar Konsequenzen für die Planung eines Projekts sowie den Bedarf an notwendigen Ressourcen ableiten.

Auf der inhaltlichen und zeitlichen Ebene muss die Umsetzung des digitalen Humanismus, beispielsweise mit der VBE-Methodik, frühzeitig in den Projektplan integriert werden. Insbesondere die Liste relevanter Stakeholder*innen muss frühzeitig erstellt werden. Auf diese Art können die Workshops rechtzeitig terminiert werden und es gibt genug Vorlauf für das Versenden von Einladungen. Von den Terminen der Stakeholder-Workshops hängt es auch ab, wann notwendige Artefakte (z. B. SOI- und SOS-Beschreibung) vorliegen müssen. Die Erstellung dieser Inhalte, sowie die notwendigen Abstimmungen müssen zeitlich eingeplant werden.

Zusätzlich müssen Aufwände für die Vorbereitung, Durchführung und Nachbereitung von Analyse eingeplant werden. Das betrifft sowohl logistische Aufgaben (wie die Organisation von Terminen, Örtlichkeiten, Verpflegung), als auch die Auswertung der Ergebnisse. Nicht alle genannten Aufwände sind dabei Zusatzaufwände. Die Erstellung einer SOI-Beschreibung ist beispielsweise eine ohnehin zentrale Aufgabe im Projekt. Daher sollte die Anwendung der VBE-Methodik frühzeitig verankert werden, damit die entsprechenden Artefakte bereits mit erstellt werden können.

Bei der Planung von Arbeitspaketen und Ressourcen ist die Unterstützung durch den Value Lead wertvoll. Die Rekrutierung dieser und anderer erforderlicher Kompetenzen kann herausfordernd sein, da die Methoden noch relativ neu sind. In der Praxis wird es oft notwendig sein, auf eine Kombination aus der Weiterbildung eigener Mitarbeiter*innen und der Zusammenarbeit mit erfahrenen Dienstleistern zu setzen. Besonders bei Erstprojekten sollte zudem explizit Zeit und Budget für den Aufbau und die Verankerung der benötigten Kompetenzen in der Organisation eingeplant werden.

5.2.6 Leadership Commitment einholen

Es ist leicht ersichtlich: Die Anwendung strukturierter Methoden zur Umsetzung des digitalen Humanismus erfordert Aufwand und folgt einer Lernkurve. Dies liegt in der Natur der Sache und es zu leugnen wäre wirklichkeitsfremd. Dem gegenüber stehen jedoch greifbare Vorteile, wie die Erleichterung der Compliance mit rechtlichen Rahmenbedingungen und die Verbesserung von Qualität und Akzeptanz des Produkts. Darüber hinaus nimmt ein Unternehmen durch die Berücksichtigung ethischer Prinzipien die eigene gesellschaftliche Verantwortung wahr.

In Pilotprojekten und darüber hinaus ist es entscheidend, dass Projektleiter*innen das geplante Vorgehen, die erwarteten Aufwände und den angestrebten Nutzen transparent darlegen, um ein klares Commitment und die Unterstützung des Senior Managements einzuholen. Dieses Commitment sollte die Bereitstellung der notwendigen Ressourcen sowie die Bestätigung der Planung beinhalten. Darüber hinaus ist es wichtig, ein gemeinsames Verständnis zu schaffen, um partnerschaftlich aufkommende Herausforderungen im Projektverlauf anzugehen und zu bewältigen.

Besonders herausfordernd kann dabei die Umsetzung notwendiger organisatorischer Maßnahmen sein, hierzu mehr in den folgenden Abschnitten.

5.3 Die Projektdurchführung

Die Berücksichtigung der Prinzipien des digitalen Humanismus ändert nichts daran, wie Projekte erfolgreich geleitet werden. Die im Abschnitt über Projektplanung aufgeführten Erwägungen und Empfehlungen sollen dabei helfen, einen, um Aspekte des digitalen Humanismus erweiterten Projektplan zu erstellen. Dieser kann mit den bekannten Werkzeugen der Projektleitung abgearbeitet werden. Im nachfolgenden Abschnitt werden spezifische Punkte adressiert, die den Projektverlauf beeinflussen.

5.3.1 Value-based Engineering anwenden

VBE bietet einen systematischen Ansatz, um ethische Prinzipien in die Entwicklung von IT-Systemen zu integrieren. Die Methodik ermöglicht es, relevante Werte systematisch zu identifizieren und im Projekt gezielt zu adressieren. Damit kann der digitale Humanismus strukturiert und zuverlässig umgesetzt werden. Ein zentraler Bestandteil dieser Methodik ist die Zusammenarbeit mit Stakeholder*innen, um deren Werteempfinden und die damit verbundenen Erwartungen an das zu entwickelnde Produkt frühzeitig zu verstehen.

5.3.2 Wertebasierte Anforderungen entwickeln

Das Verständnis darüber, wie das Ergebnis einer VBE-Analyse den Projektverlauf beeinflusst, ist essenziell. Die Methodik zielt darauf ab, aus der Werte-Analyse zusätzliche Anforderungen abzuleiten, die als *Ethical Value Requirements* (EVR) bezeichnet werden. Diese Anforderungen adressieren ethische Risiken und federn sie angemessen ab. Die praktische Umsetzung dieser Methode wird besonders anschaulich im Use-Case ibis acam (Kap. 15).

Aus den EVRs werden im nächsten Schritt konkrete technische und organisatorische Maßnahmen abgeleitet. Somit folgt bereits aus der Methodik, dass im Projektverlauf mit zusätzlichen Anforderungen auf Basis der VBE-Analyse zu rechnen ist.

Die Ausgestaltung dieser Anforderungen passiert zwar mit Blick auf das konkrete Projekt und die praktische Umsetzbarkeit, letztlich sind und bleiben neue Anforderungen jedoch Änderungen, die das Projekt adressieren muss.

5.3.3 Technische Anforderungen adressieren

Handelt es sich um neue technische Anforderungen, so gibt es in der Regel in jedem Projektvorgehensmodell etablierte Methoden, um sie zu adressieren.

Projekte, die einem Wasserfall oder V-Modell folgen, arbeiten häufig mit sogenannten *Change Requests*. Diese sind Beschreibungen von Änderungen, die unter anderem

bzgl. des zu erwartenden Umsetzungsaufwands und der -dauer bewertet werden. Ein später Change Request stellt natürlich ein Risiko für das Projektbudget und die -zeitleiste dar. Daher ist es offensichtlich, dass das Feedback von Stakeholder*innen so früh wie möglich eingeholt werden sollte. Zusätzlich sollten ethische Fragestellungen, wie dargestellt, auch aus dem Projektteam heraus von Anfang an betrachtet werden, so dass die notwendigen Änderungen so gut wie möglich minimiert werden. Ganz vermeidbar sind Änderungen nicht, da der digitale Humanismus es erfordert, auf ethische Risiken und Erkenntnisse natürlich auch dann zu reagieren, wenn dies mit Aufwand verbunden ist.

Als Lösungsansatz könnte der Einsatz von agilen Methoden ins Spiel gebracht werden. „*Move fast, break things*“ ist ein viel zitierter Satz, der jedoch mittlerweile einen zynischen Unterton hat. Geschwindigkeit kann kein Argument dafür sein, beim Design eines Produktes seine ethischen Risiken zu vernachlässigen. Agile Ansätze finden jedoch ihre Berechtigung darin, dass Flexibilität bei der Entwicklung von neuen Produkten oftmals eine Notwendigkeit ist, da man ohne sie nicht auf neue Erkenntnisse reagieren könnte.

Die Lösung für diesen scheinbaren Konflikt zwischen einer detaillierten Analyse von Risiken und der Notwendigkeit von Flexibilität liegt in einem guten Management von Anforderungen: VBE sieht als eine Anforderung die Nachverfolgbarkeit jeder Anforderung zurück zum Kernwert vor. Diese *Traceability* kann helfen zu verstehen, ob und inwieweit Änderungen auch Anforderungen betreffen, die auf ethischen Erwägungen basieren. So vermeidet man, dass unbedarft die Umsetzung etablierter EVRs aus dem Produkt verschwindet.

Sind die VBE-Workshop bereits abgeschlossen, so besteht ein Risiko, dass spätere Änderungen nicht mehr ausreichend durchleuchtet werden. Eine Lösung hierfür kann es sein, mit den Workshop-Teilnehmer*innen weiterhin in Kontakt zu bleiben, um sie auch mit ergänzenden Fragestellungen zu befassen. Die persönliche Erfahrung des Autors zeigt, dass Teilnehmer*innen häufig sehr interessiert am weiteren Verlauf des Projekts sind, insbesondere wenn es sich um ein im Betrieb eingesetztes Software-Produkt handelt. Man sollte die VBE-Analyse also nicht als Einmalübung ansehen, sondern als fortlaufende Aktivität.

5.3.4 Organisatorische Anforderungen adressieren

Organisatorische Anforderungen sind in der Regel nicht im Umfang des Projektes handhabbar. Was ist unter einer organisatorischen Anforderung zu verstehen? Es handelt sich dabei um Maßnahmen, die nicht technischer Natur sind. Ein einfaches Beispiel ist die Verfügbarkeit des Kundenservice: Bei der VBE-Analyse für die Webseite einer Firma wird vorgeschlagen, dass Kunden immer mit einem menschlichen Mitarbeiter sprechen können müssen. Daraus lässt sich ableiten, dass die Kontaktaufnahme möglich sein muss, was eine technische Anforderung an das System darstellt. Zusätzlich ist es jedoch eine organisatorische Anforderung, dass Personal zur Beantwortung solcher Anfragen zur Verfügung steht. Dieses Personal wird meist nicht aus dem Projekt heraus organisiert, sondern von einer anderen Abteilung – beispielsweise dem Kundenservice.

Um solche Anforderungen aufzulösen, müssen Projektleiter*innen entsprechend Vereinbarung mit anderen Abteilungen außerhalb des eigenen Projektes treffen. Solche Anforderungen an den laufenden Betrieb können nicht geplante Aufwände bei den Schnittstellenpartnern erzeugen. Aus diesem Grund wird die Vereinbarung der Umsetzung organisatorischer Anforderungen häufig die Einbindung des Senior Managements erfordern.

5.3.5 Risiken analysieren und adressieren

Viele Unternehmen sind nach Standards zum Qualitätsmanagement, wie etwa ISO 9001, zertifiziert. Diese Standards beinhalten die Notwendigkeit, Risiken zu identifizieren und zu adressieren. Die Ableitung von Anforderungen im Rahmen von VBE ist eine Risikoanalyse, deren Ergebnisse auch in das betriebliche Risikomanagement mit einfließen können.

Projektleiter*innen sollten jedoch auch berücksichtigen, dass das Ignorieren ethischer Risiken selbst ein erhebliches Risiko darstellt – sowohl monetär (z. B. durch schlechte Akzeptanz des Produkts) als auch für die Reputation des Unternehmens (z. B. durch negative Bewertungen). Es ist daher mindestens empfehlenswert, wenn nicht zwingend erforderlich, diese Risiken ebenfalls zu erfassen, um eine ganzheitliche Betrachtung zu gewährleisten.

5.4 Die Überprüfung

Das Monitoring einer Anwendung über den Produktlebenszyklus hinweg ist in vielen Fällen entweder dringend zu empfehlen oder wird zukünftig sogar gesetzlich verpflichtend. Bei KI-Systemen hat das Monitoring auch den Hintergrund, dass diese Systeme ihr Verhalten über die Zeit ändern können. Dies liegt daran, dass KI-Systeme häufig datenbasiert entwickelt werden. Verändert sich der Einsatzkontext, kann die Qualität der Vorhersagen eines solchen Systems dadurch schlechter werden.

Dieses auch als *Model Drift* bezeichnete Problem ist der KI-Technologie inhärent und wird häufig thematisiert. Der Fokus des vorliegenden Abschnitts liegt deshalb auf der ethischen Dimension: Wie kann man als Organisation die ethische Produktqualität dauerhaft sicherstellen?

5.4.1 Ethische Faktoren beobachten

Auch ethische Faktoren, die ein Produkt betreffen, können sich über die Zeit ändern. Daher ist es, genau wie bei der allgemeinen Produktqualität, wichtig, solche Veränderungen zu erkennen und zu adressieren. Wie genau ein solches Monitoring stattfindet, hängt vom konkreten Produkt und dessen Einsatzkontext ab.

Obwohl es sich um ein neues Feld handelt, existieren erste spannende Ansätze für verschiedene Einsatzszenarien:

- Wertemonitor¹ für Chatbots: Dieser Ansatz wurde auf Basis von VBE entwickelt. Der Wertemonitor verwendet die Wertesprache von VBE. Er ermöglicht es zu bewerten, inwieweit Chatbots bestimmte Werte, wie etwa Empathie und Wahrhaftigkeit, in ihren Dialogen beachten. Die relevanten Werte zu erkennen, sinnvolle Zielwerte zu definieren und diese zu beobachten ist wichtig: Bei ChatBots kommt es entscheidend auf den Kontext des Einsatzes an. Die Verwendung auf der Webseite eines Baumarktes erfordert beispielsweise weniger Empathie im Dialog, als dies auf der Webseite eines Gesundheitsanbieters der Fall ist.
- Monitoring mit IEEE 7010: Der Standard IEEE 7010 beinhaltet Empfehlungen, um die Auswirkungen von autonomen Systemen auf das menschliche Wohlbefinden zu beurteilen. Dabei liegt der Fokus auf der Definition konkreter Metriken, die ein fortlaufendes Monitoring ermöglichen sollen.

Des Weiteren kann das Monitoring des in der Planungsphase analysierten Kontextes hilfreich sein: Ergeben sich hier Änderungen, so könnten auch Überlegungen aus Sicht der Ethik eine erneute Überprüfung der EVRs und weitere Annahmen sinnvoll erscheinen lassen. In jedem Fall muss es Teil des Projektes sein, zu bestimmen, wie ein Monitoring des Werteprofiles eines Produktes erfolgreich umgesetzt werden kann.

5.4.2 Zertifizierungen anstreben

Assessments und die daraus folgenden Zertifizierungen helfen Projekten und Organisationen auf verschiedenen Ebenen: Zum einen können sie helfen, die Qualität der eigenen Prozesse und Produkte zu überprüfen, zum anderen kann das eigene Commitment zu ethischen Praktiken gegenüber den Kund*innen demonstriert werden.

Mit ISO/IEC 42001:2023 existiert ein Standard, der ein organisatorisches AI Management System beschreibt. Für die Umsetzung des EU AI Acts wird dieser gemeinhin als nicht ausreichend bezeichnet, was zu vielen Diskussionen in Fachkreisen geführt hat. Diese Einschätzung mag überraschen, ist jedoch nachvollziehbar: ISO/IEC 42001 ist ein organisatorischer Standard, während der EU AI Act auf die Produktebene abzielt. Diese Trennung ist auch in anderen Management Systemen üblich. So sind Automobilhersteller etwa nach ISO 9001 zertifiziert, obwohl für die Produkte zusätzliche Standards zu beachten sind.

¹ Ein Wertemonitor ist ein Konzept, das im Rahmen des Value-based Engineering (VBE) entwickelt wurde. Es handelt sich dabei um ein Instrument, das verwendet wird, um zu bewerten, inwieweit bestimmte Systeme oder Technologien, wie zum Beispiel Chatbots, bestimmte Werte wie Empathie und Wahrhaftigkeit in ihren Dialogen beachten. Der Wertemonitor hilft dabei, relevante Werte zu erkennen, sinnvolle Zielwerte zu definieren und diese kontinuierlich zu beobachten.

Daher wird es kein „entweder ... oder“ für die Zertifizierung von Organisation und Produkt geben. Es wird ein „und“ sein. Unternehmen werden organisationsweite Prozesse für das Management von KI benötigen, die insbesondere auch helfen, die Chancen des KI-Einsatzes realistisch zu bewerten und über die Strategie zu deren Realisierung zu entscheiden. Gleichzeitig müssen gesetzliche und ethische Anforderungen berücksichtigt und mit dem globalen Management System verknüpft werden. ISO/IEC/IEEE 24748-7000 definiert VBE und ist eine solche produktorientierte Methode, die in ein organisatorisches AI Management System integrierbar ist.

Darüber hinaus existieren Produktzertifizierungen, wie IEEE CertifAIED™. Dieser Ansatz beurteilt ein Produkt hinsichtlich der Verwirklichung bestimmter Aspekte des werte-basierten Systemdesigns. So existieren aktuell vier Kriterienkataloge zu Transparenz, Accountability, Datenschutz und algorithmischen Bias. Ein Assessment mit der CertifAIED™ kann wertvolles Feedback für die Entwicklung und Verbesserung von Produkten generieren. Das für konforme Produkte vergebene Prüfzeichen ermöglicht es, ein Produkt entsprechend als ethisch durchdacht im Sinne des betrachteten Wertes zu positionieren.

5.4.3 Feedback einholen und berücksichtigen

Ein weiterer wichtiger Faktor ist das Einholen und Berücksichtigen von Feedback, das Verwender des Produkts geben. Organisationen sollten regelmäßig den Dialog suchen, beispielsweise durch Befragungen von Kund*innen und weiteren Stakeholder*innen. Darüber hinaus sollten Feedbacks aus dem Feld, wie Kundenbeschwerden oder Supportanfragen, analysiert und bei neuen Produktversionen berücksichtigt werden.

Um dies zu ermöglichen, muss das eigene Personal für ethische Aspekte sensibilisiert werden. So können Schwierigkeiten bei der Nutzung des Produkts durch bestimmte Personengruppen etwa auf mangelhafte Barrierefreiheit und somit eine reduzierte Teilhabe dieser Personen hindeuten. Der beschriebene Wertemonitor kann auch hier ein hilfreiches Werkzeug sein.

5.5 Verbesserungen umsetzen

Methoden und Standards wie VBE existieren erst seit vergleichsweise kurzer Zeit. Insofern dürfen sich alle, die den digitalen Humanismus proaktiv in ihre Arbeit einbringen, immer noch als Pioniere betrachten. Die gewonnenen Erfahrungen aus Planung, Umsetzung und schließlich „aus dem Feld“ sind wertvoll und sollen sowohl in die Verbesserungen bestehender Produkte als auch in die Entwicklung neuer Produkte einfließen.

Für die Berücksichtigung von Erkenntnissen im eigenen Produkt, also Updates und Wartungen, existieren in der Regel unternehmensinterne Prozesse. Dabei sind zwei Aspekte besonders wichtig:

- Neue Erkenntnisse und Anforderungen: Wie beschrieben, können sich neue Erkenntnisse aus dem Einsatz des Produkts ergeben, die wiederum neue Features und Anforderungen nach sich ziehen. Soweit sich das Produkt weiterentwickelt, ist es auch notwendig, neu entstehende Risiken zu analysieren und mit den beschriebenen Ansätzen zu behandeln. Dies kann etwa auch die Notwendigkeit weiterer VBE-Analysen und Stakeholder-Workshops bedeuten.
- Kontinuität ethischer Anforderungen: Es sollte darauf geachtet werden, dass Änderungen nicht negativ auf bereits bekannte und adressierte ethische Anforderungen wirken. Die durch VBE entwickelten Ethical Value Requirements müssen auch nach jeder Produktaktualisierung weiterhin erfüllt sein. Dies kann als Teil des Requirements-Engineerings adressiert werden.

Darüber hinaus lohnt es sich das im Projekt gewonnene Wissen auch der Gesamtorganisation zur Verfügung zu stellen. Die Rückmeldungen der Stakeholder*innen sind oft stark von deren Lebensrealität geprägt. So werden Menschen in Europa ein IT-System oftmals anders beurteilen als Menschen, die außerhalb Europas leben. Eine Organisation operiert jedoch häufig im gleichen Kontext. Insofern können Erkenntnisse aus einem Projekt für andere Projekte derselben Organisation hilfreich sein. Die Erkenntnisse sollten deshalb in Form von Best Practices dokumentiert werden.

5.6 Ausblick

Standards wie ISO/IEC/IEEE 24748-7000 existieren erst seit wenigen Jahren – wir alle sind noch Pioniere auf dem Weg, ethisches Handeln systematisch in die Praxis zu integrieren. Der vorliegende Beitrag zeigt erste Ansätze und Erkenntnisse: Es gibt Möglichkeiten, ethische Lösungen frühzeitig und strukturiert in Projekte einzubringen. Doch dies ist erst der Anfang.

Die bisherigen Praxiserfahrungen zeigen, dass es sich lohnt, ethische Überlegungen nicht nur bei der Produktentwicklung, sondern insbesondere dort voranzutreiben, wo IT die Arbeits- und Lebensrealität von Menschen beeinflusst. Die Akzeptanz der Menschen ist entscheidend für den Erfolg solcher (Digitalisierungs-)Projekte. Betroffene frühzeitig einzubinden und ethische Spannungsfelder klar zu adressieren, ist deshalb zwingend geboten.

Der Weg in die Zukunft erfordert nicht nur Offenheit und Lernbereitschaft, sondern auch eine klare Haltung: Technologie sollte nicht nur funktional sein, sondern stets im Einklang mit menschlichen Werten gestaltet werden. Methoden wie VBE bieten dafür die notwendige Struktur. Doch ihre erfolgreiche Umsetzung hängt von uns allen ab – insbesondere von der Bereitschaft, Verantwortung für die Auswirkungen von Technologie zu übernehmen. Indem wir aus Erfahrungen lernen und neue Erkenntnisse einfließen lassen, können wir eine Technologie gestalten, die das Leben von Menschen tatsächlich verbessert und damit die Bezeichnung „Fortschritt“ wirklich verdient.

Mario Tokarz ist Gründer und Geschäftsführer von RightMinded AI, einem Unternehmen, das Organisationen bei der verantwortungsvollen Nutzung von Künstlicher Intelligenz und der Entwicklung nachhaltiger Softwarelösungen unterstützt. Als Mitbegründer und Geschäftsführer der VBE Academy engagiert er sich für die Vermittlung von Kompetenzen in der ethischen Systemgestaltung auf Basis des internationalen Standards ISO/IEC/IEEE 24748-7000. Nach seinem Informatikstudium war er langjährig für die BMW AG tätig, zuletzt als Abteilungsleiter für KI-Strategie. Anschließend übernahm er als Engineering- und Program Manager in einem führenden Startup für autonomes Fahren maßgebliche Verantwortung für die Entwicklung und Skalierung KI-gestützter Fahrfunktionen.

CHRO-Perspektive: Künstliche Intelligenz im Personalmanagement

6

Menschenzentrierter Ansatz und Compliance in der Arbeitswelt

Tünde Fülöp

Zusammenfassung

Der Beitrag beleuchtet die Herausforderungen und Chancen des Einsatzes von Künstlicher Intelligenz (KI) im Personalmanagement. Es wird diskutiert, wie KI-Systeme zu mehr Produktivität und Innovation beitragen können, während gleichzeitig ethische und rechtliche Aspekte sowie die Einbindung der Beschäftigten berücksichtigt werden. Der menschenzentrierte Ansatz der EU und praktische Empfehlungen zur sicheren und rechtskonformen Nutzung von KI im Personalmanagement stehen im Fokus.

KI-Einsatz ohne einen menschenzentrierten Ansatz droht Arbeitslose und Ungleichheit zu produzieren. Verzichteten Sie nicht auf echte Innovation, die auf der Integration von Technologie und menschlichen Fähigkeiten beruht.

6.1 Künstliche Intelligenz und Arbeitswelt

6.1.1 Herausforderungen beim Einsatz künstlicher Intelligenz im Personalmanagement

Künstliche Intelligenz (KI) ist bereits in zwei Drittel der österreichischen Unternehmen im Bereich „Human Resources“ im Einsatz, überwiegend als integraler Bestandteil von

T. Fülöp (✉)
Arbeiterkammer, Wien, Österreich
E-Mail: tuende.fueloep@akwien.at

Software: Zum Beispiel bei der Personalbeschaffung, zum Schreiben von Stellenanzeigen mit generativer KI, zum Filtern von Bewerbungen, bei der Leistungsbewertung von Mitarbeitenden, bei der Zuweisung von Aufträgen an Mitarbeitende und der Personaleinsatzplanung, bei der Berechnung von Gehältern und Boni, bei der Auswahl und Zuteilung von Schulungen, oder bei der Identifikation von Betrug und Regelverstößen (RTR, 2025). Gleichzeitig ist KI in zwei Drittel der österreichischen Unternehmen bislang kaum ein Thema (Der Standard, 2023, 2024, Statistik Austria, 2023). Gemäß ersten Ergebnissen aus einer laufenden Monitoring-Studie der Arbeiterkammer Wien könnte das unter anderem am mangelnden Wissen über die Technologie und ihre potenziellen Vor- und Nachteile liegen (AK Wien, 2025).

Damit sind verschiedene Herausforderungen verbunden: Für viele Einsatzmöglichkeiten gilt die europäische KI-Verordnung, die Beschäftigung und Personalmanagement als Hochrisiko-Bereich qualifiziert, siehe Anhang III der Verordnung (EU) 2024/1689 des Europäischen Parlaments und des Rates zur Festlegung harmonisierter Vorschriften für künstliche Intelligenz (KI-Verordnung, 2024). Einige KI-Systeme am Markt sind nicht rechtskonform, oder können nicht leisten, was sie versprechen. Technische Kompetenzen zur Bewertung von KI-Tools, aber auch für den korrekten Einsatz, fehlen oft sowohl auf Arbeitnehmer- als auch auf Arbeitgeberseite. Besonders, wenn KI ungefragt und ungebeten in Software integriert wird, stellen sich Fragen der Compliance mit Datenschutz und arbeitsrechtlichen Bestimmungen. Insbesondere im algorithmischen Management (das ist der Einsatz von Algorithmen und künstlicher Intelligenz zur Überwachung, Steuerung und Optimierung von Arbeitsprozessen und Mitarbeiterleistungen) finden unethische und zum Teil rechtswidrige Praktiken zunehmend Verbreitung, z. B. Diskriminierung und Bias, verbotene Überwachung, illegale Datenverarbeitung (Christl, 2024).

Problematisch ist dies besonders dann, wenn sich Unternehmen beim KI-Einsatz ausschließlich am Effizienzgewinn orientieren, wie das häufig der Fall ist. Nicht nur, weil so oft auf die Interessen von Beschäftigten „vergessen“ wird, denen Risiken wie Eingriffe in die Menschenwürde durch ständige Beobachtung und Kontrolle, die Verdichtung der Arbeit und damit verbundener Stress und Burnout, und der Verlust von Skills/Kompetenzen drohen. Auch aus unternehmerischer Sicht ist das kurzsichtig, da Automatisierung durch Effizienzgewinne nicht gleich Produktivität bedeutet, und weil die Chance auf echte Innovation vertan wird, die nur aus dem erfolgreichen Zusammenspiel von Menschen, Technologie und Organisation entstehen kann.

Unternehmen und Personalverantwortliche müssen sich deshalb mit verschiedenen Fragen im Zusammenhang mit dem Einsatz von KI-Systemen auseinandersetzen: Einerseits, ob sie bestimmte KI-Systeme und Anwendungen ankaufen und im Personalmanagement überhaupt einsetzen dürfen, und worauf sie dabei achten sollten. Andererseits fragen sich viele: Wie wird die Zukunft mit KI das Arbeitsumfeld verändern, und wie bereiten wir uns, unsere Beschäftigten und unsere Organisation darauf am besten vor? Wird algorithmisches Management bald Berufe verdrängen (vielleicht sogar den eigenen)?

Diesen Fragen widmen wir uns, indem wir zunächst einen Blick auf Trends und Einsichten aus aktuellen Studien zu KI in der Arbeitswelt, und zu Wirkungen der Einführung

von KI-Systemen auf Beschäftigte werfen. Aus verschiedenen Untersuchungen wird deutlich: Der Einsatz von KI-Systemen funktioniert besser und führt zu mehr Produktivität, wenn – im Sinne des digitalen Humanismus – Menschen in den Mittelpunkt gestellt und Beschäftigte eingebunden werden. Am Beispiel des menschenzentrierten Ansatzes der EU möchte ich dann Vorgaben skizzieren und praktische Empfehlungen geben, die den Einsatz von KI-Systemen im Personalmanagement im Sinne des digitalen Humanismus möglich machen.

6.1.2 Trends zu KI und Beschäftigung

Aus verschiedenen Studien, die sich mit den Auswirkungen von KI und algorithmischem Management in verschiedenen Branchen und Ländern beschäftigen, geht hervor: KI-Technologien wirken sich auf eine große Bandbreite von Beschäftigten und ihren Aufgaben aus. Im Unterschied zu früheren technologischen Revolutionen sind diesmal auch „kreative“ Aufgabenbereiche und „White Collar Work“, d. h. praktisch alle Arbeitsbereiche, einschließlich Wissenschaft, Medizin und Recht, betroffen. Nicht klar ist bislang, welche Berufsbilder betroffen sind oder ob tatsächlich Berufe verschwinden werden (OECD, 2023, ILO/JRC 2024, FEPS 2024).

Die Beschäftigungsniveaus sind trotz der Einführung von KI in den untersuchten Unternehmen stabil geblieben, mit Einführung von KI steigt das Beschäftigungsniveau tendenziell sogar zunächst an. Auf 15–30 Jahre gesehen, mit zunehmender Etablierung der neuen Technologie, sind größere Verwerfungen zu erwarten (ähnlich wie durch Dampfmaschine, Elektrizität und Computer), Genauer ist allerdings nicht vorhersagbar.

Die KI-Nutzung führt eher zur Umorganisation von Tätigkeiten als zu tatsächlichen Arbeitsplatzverlusten. Komplementäre KI-Technologie steigert die Leistungsfähigkeit, die Automatisierung bewirkt eine Verlagerung auf jene Aufgaben, bei denen Menschen komparative Vorteile aufweisen. Beschäftigte übernehmen besonders solche Tätigkeiten, die Empathie, soziale Interaktion, oder bestimmte Arten der Entscheidungsfindung erfordern.

Erforderliche Kompetenzen ändern sich, neue Berufsbilder entstehen. Oft erfordert die Einführung von KI-Technologien sowohl ein höheres Kompetenzniveau als auch ein breiteres Kompetenzspektrum. Analytische Fähigkeiten und zwischenmenschliche Kompetenzen sind gefragt. Es gibt eine starke Nachfrage nach spezialisierten KI-Kompetenzen, und hier besonders nach gut ausgebildeten Mitarbeitenden. Aber auch bei anderen, nicht spezifisch KI-bezogenen Tätigkeiten bringt KI- bzw. digitale Kompetenz große Vorteile, da Arbeit leichter wird und Ressourcen freigesetzt werden.

Kurz- und mittelfristig wird der Einsatz von KI also einen massiven Schulungs- bzw. Umschulungsbedarf mit sich bringen. Betroffen sein werden unterschiedliche Jobs auf unterschiedliche Art: Zunächst steht alles, was automatisierbar ist, an. Langfristig werden voraussichtlich viele Berufsbilder wegfallen, aber auch neue entstehen – genauso, wie zum Beispiel der PC eine ganze Bandbreite an neuen Berufen gebracht hat.

Was spannend ist: Die Wirkung von KI auf Beschäftigte kann sehr unterschiedlich ausfallen. In manchen Fällen wird z. B. der Wegfall monotoner Tätigkeiten sehr positiv, als Verbesserung der Beschäftigungsqualität beurteilt. In anderen scheint er zu Arbeitsverdichtung, einer Steigerung der Arbeitsintensität und dadurch zu psychosozialem Druck und Stress zu führen. Kompetenzanforderungen können durch KI-Einführung steigen, oder, wie in einigen Fällen im verarbeitenden Sektor, fallen. Ähnliche Technologien wirken sich unterschiedlich aus, wie eine Studie in Unternehmen in zwei europäischen und zwei außereuropäischen Ländern gezeigt hat (ILO/JRC, 2024).

6.1.3 Produktivität und gute Arbeitsbedingungen durch KI-Einsatz

Die Studien von OECD (2023) und ILO/JRC (2024) zeichnen ein differenziertes Bild der Vorteile und Herausforderungen von KI auf Produktivität und Wohlergehen der Beschäftigten. Was macht den Unterschied aus? Aus den Ergebnissen lässt sich ableiten: Die Auswirkungen des KI-Einsatzes sowohl auf Arbeitnehmerzufriedenheit als auch auf Produktivität sind abhängig von organisatorischen und rechtlichen Rahmenbedingungen, dem Kompetenzniveau und der Beteiligung der Beschäftigten, sowie der Adaptierung der Organisation und der Beschäftigten an die neue Technologie durch begleitende Maßnahmen in der Arbeitsorganisation und durch Schulungen. Das heißt, im Detail:

1. Die Arbeitsorganisation (Aufgaben, Rollenbilder, Prozesse) bestimmt, ob KI zu höherer Produktivität und besseren Arbeitsbedingungen führt. Es macht einen Unterschied, ob und wie sich das KI-System in bestehende Arbeitsprozesse fügt oder integriert wird, bzw. wie die Arbeitsorganisation an die neuen technologischen Möglichkeiten angepasst wird.
2. Die Beteiligung der Beschäftigten an der Umsetzung von KI-Lösungen verringert Ängste und stärkt die Bereitschaft, sich mit KI und neuen Aufgaben auseinanderzusetzen. Mitarbeitende kennen den tatsächlichen Bedarf, die Optimierungsmöglichkeiten und die Arbeitsabläufe (auch jene, die nicht den formalen Vorgaben entsprechen). Sie können dazu beitragen, dass die Technologie im gegebenen Arbeitskontext bestmöglich eingesetzt werden kann.
3. Berufsbild und Gehalt müssen den Aufgaben entsprechend angepasst werden. Eine Aufwertung der Arbeitsplätze, Wertschätzung und Bildungsmaßnahmen, die Kompetenzen erhöhen (Upskilling, Reskilling) verhindern Frustration und helfen, wertvolle Arbeitskräfte an die Organisation zu binden.
4. Klare rechtliche Rahmenbedingungen (u. a. Arbeitsrecht, Arbeitnehmerschutz, Datenschutz) haben positive Auswirkungen auf Arbeitsbedingungen und Arbeitszufriedenheit, die sich wiederum positiv auf die Produktivität auswirken. Das zeigt der Vergleich zwischen Unternehmen in der EU bzw. Südafrika und Indien (ILO/JRC, 2024).
5. Ganz essenziell sind Kompetenzen. Ob sich Unternehmen und ihre Beschäftigten an die Einführung von KI gut und schnell anpassen können, hängt davon ab, wie sie die

Tätigkeiten und Prozesse umorganisieren, welches Kompetenzniveau die Beschäftigten besitzen und wie sehr sich die Unternehmen bemühen, die Qualifikation ihrer Beschäftigten bei Bedarf durch Schulungsmaßnahmen zu erhöhen (OECD, 2023).

Deshalb setzt die KI-Verordnung auf KI-Kompetenzen (Art 4) als Voraussetzung für effektive menschliche Aufsicht und als wichtiges Element des menschenzentrierten Ansatzes. In diese Richtung sollen auch nationale Programme zur Förderung digitaler Kompetenzen wirken (DKO Strategie, 2023).

6.2 Der menschenzentrierte Ansatz im EU-Recht: Digitaler Humanismus durch Beachtung rechtlicher Vorgaben

6.2.1 Was ist der „menschenzentrierte Ansatz“?

Im jüngeren EU-Digitalrecht ist der menschenzentrierte Ansatz nicht nur eine Methode für Technologieentwicklung und -Einsatz, sondern auch die Grundlage für ein alternatives, wertebasiertes Wirtschaftsmodell, das sich vom US- bzw. chinesischen Modell (der „Surveillance Economy“, Zuboff, 2019) absetzen will und die eine ethische und wertebasierte Haltung voraussetzt. Daten, KI und digitale Technologien sollen weder ausschließlich digitalen Großkonzernen mit Monopolstellung in der Datenökonomie (wie Google oder Meta) zugutekommen, noch für gesellschaftliche Kontrolle eingesetzt werden (Nemitz & Pfeffer, 2020). Sie sollen dem Gemeinwohl, europäischen Unternehmen (insbesondere KMU) und Menschen insgesamt Vorteile bringen und das Fundament für einen breiteren Wohlstand und europäische Wettbewerbsfähigkeit werden.

Der menschenzentrierte Ansatz dient in diesem Sinn dem Erhalt der Menschenwürde, der menschlichen Autonomie und der Selbstbestimmung auf Grundlage der europäischen Werte und Grundrechte, sowie der Förderung von speziell menschlichen Bedürfnissen: Technologie sollte „den Menschen als Instrument dienen und letztendlich das menschliche Wohl verbessern“ (Erwägungsgrund 6 der KI-Verordnung). Er spiegelt sich u. a. in den „Ethics Guidelines for Trustworthy AI 2019“ der von der Europäischen Kommission eingesetzten „High Level Expert Group“ (Ethics Guidelines, 2019) oder in der „Europäischen Erklärung zu den digitalen Rechten und Grundsätzen für die digitale Dekade“ (Europäische Erklärung, 2023), und baut auf die Grundsätze „Vorrang menschlichen Handelns und menschlicher Kontrolle“, „technische Robustheit und Sicherheit“, „Privatsphäre und Datenqualitätsmanagement“, „Transparenz und Erklärbarkeit“, „Vielfalt, Nichtdiskriminierung und Fairness“, „Gesellschaftliches und ökologisches Wohlergehen“ sowie „Rechenschaftspflicht“. Menschenzentrierter KI-Einsatz sollte transparent sein, partizipativ, die informationelle Selbstbestimmung achtend, für und mit Beschäftigten, Kund*innen und Bürger*innen, vertrauenswürdig, sicher und ethisch.

6.2.2 Gute KI: Sicher, vertrauenswürdig und menschenzentriert

Dementsprechend schreibt die KI-Verordnung vor, dass hochriskante KI-Systeme, zum Beispiel im Bereich Beschäftigung und Personalmanagement, nur eingesetzt werden dürfen, wenn bestimmte Vorgaben befolgt werden (Fülöp, 2025). Diese dienen der Minimierung von Risiken für Menschen – Risiken für Gesundheit, Sicherheit und Grundrechte. Die Vorgaben betreffen u. a. Risikomanagement, Dokumentation und Transparenz, Daten-Governance, menschliche Aufsicht und Cybersicherheit. Personalverantwortliche sollten sich daher bewusst sein, dass sie sich bei der Beschaffung, aber vor allem bei der Eigenentwicklung in einem strengen rechtlichen Korsett bewegen. Beim derzeitigen rasanten technologischen Fortschritt macht das, angesichts der Unsicherheit und der potenziellen Gefahren, die mit KI verbunden sind, Sinn: Das Recht soll Risiken minimieren und Menschen schützen.

Wie können Sie sich im rechtlichen Dickicht zurechtfinden, und bestmöglich, im Sinne des digitalen Humanismus, Entscheidungen zum KI-Einsatz treffen? Die folgenden Tipps können helfen.

6.3 Praxistipps: Wie kann menschenzentrierte KI im Unternehmen verankert werden?

Achten Sie darauf, dass Sie „gute“ KI-Systeme ankaufen und einsetzen, die den Anforderungen des komplexen Regelwerks der KI-Verordnung entsprechen. Binden Sie dazu idealerweise Spezialisten ein, und bringen Sie Ihr eigenes Fachwissen in die Diskussion ein. Gute KI gibt Ihnen eine ausreichende Entscheidungsgrundlage: Idealerweise erfüllt ein der KI-Verordnung entsprechendes KI-System bereits alle Anforderungen für den menschenzentrierten KI-Einsatz. Die Betriebsanleitung (die gemäß Art 13 Abs. 2 KI-Verordnung vom Anbieter bereitzustellen ist) muss transparent sein und ausreichend Informationen enthalten, damit Sie beurteilen können, ob die KI für den geplanten Einsatzbereich geeignet ist. Diese vorgeschriebene, Betreibern von Hochrisiko-KI-Systemen zur Verfügung zu stellende Dokumentation bzw. vorvertragliche Information muss verschiedene Elemente enthalten, u. a.:

- eine ausführliche Erklärung bezüglich „Merkmale, Fähigkeiten und Leistungsgrenzen des Hochrisiko-KI-Systems“, einschließlich der erwartbaren Genauigkeit, Funktionsweise, Möglichkeiten des Einsatzes, Cyber- und sonstige Risiken und potenzielle Fehlanwendungen, sowie Informationen, um die Ausgaben der KI zu interpretieren,
- Erläuterungen zu den Trainingsdaten, und den während des Einsatzes verarbeiteten Daten („lernt“ die KI mit? Wer bekommt allenfalls geschützte Daten – Geschäftsgeheimnisse und personenbezogene Daten?) sowie
- Hinweise zum Lebenszyklus und zur Kooperation mit dem Anbieter, unter anderem bei der verpflichtenden Meldung von riskanten Vorfällen.

6.3.1 Wie können Sie den KI-Einsatz ethisch, rechtskonform und sicher mitgestalten?

Um den KI-Einsatz menschenzentriert und grundrechtskonform zu gestalten, sollten Sie auf Zweckmäßigkeit, Verhältnismäßigkeit, Transparenz und menschliche Aufsicht achten. Die Einhaltung dieser Grundsätze in der Praxis stellt Menschen in den Mittelpunkt. Die folgenden Fragen können Ihnen bei der Compliance und bei der Umsetzung der relevanten rechtlichen Vorgaben aus der KI-Verordnung und dem Datenschutz helfen.

Ziele und Zweckmäßigkeit:

- Überlegen Sie sich noch vor der Beschaffung, was genau Ihre Ziele sind. Was ist der Zweck des KI-Einsatzes, was genau brauchen Sie?
- Was brauchen Ihre Kund*innen, was Ihre Beschäftigten?
- Welche mittleren und langfristigen, positiven oder negativen Folgen könnten mit dem KI-Einsatz für die Organisation, aber auch für die Menschen und die Umwelt entstehen?
- Gibt es ausreichend Garantien hinsichtlich Lebensdauer/Leistung/Rechtskonformität in der vertraglichen Information? Ist die Betriebsanleitung (gemäß Art 13 KI-Verordnung) informativ genug?
- Verschaffen Sie sich einen Überblick über die Produktlandschaft, suchen Sie nach „guter KI“, die Ihren Anforderungen entspricht, und mit der die Einhaltung von Datenschutz, Arbeitnehmerrechten und der Schutz der Grundrechte einfacher wird.

Erforderlichkeit und Verhältnismäßigkeit:

- Kann das, was Sie bezwecken, tatsächlich mit KI-Systemen so erreicht werden, dass Sie (kurz-/mittel-/langfristig) einen Vorteil haben? Oder müssen Sie mehr Ressourcen einsetzen, als das Ergebnis bringt?
- Kann die KI was sie verspricht? Können die intern gesetzten Ziele damit erreicht werden? Können sie besser (kostengünstiger, mit weniger Eingriff in die Rechte von Beschäftigten oder Kund*innen) oder gleich gut mit anderen Mitteln erreicht werden?
- Wird ein zusätzlicher (Verwaltungs-)Aufwand mit dem KI-Einsatz geschaffen, der den erwarteten Vorteil aufwiegt (erforderliche Schulungen, Kontrollen, Ergebnisprüfung)?
- Ist es besser abzuwarten, bis Produkte heranreifen?

Ethische und rechtliche Überlegungen:

- Wem kommt das erwartete Ergebnis bzw. die Veränderung zugute, wer hat was davon: Mitarbeitende, Bürger*innen, Kund*innen?
- Gibt es einen Diversitätsaspekt oder mögliche Diskriminierungen?
- Wie kann dem sog. „Automation Bias“ (übermäßiges Vertrauen in automatisierte Hilfsmittel und Systeme zur Entscheidungsfindung) entgegengewirkt werden?
- Wie kann effektive menschliche Aufsicht sichergestellt werden?

- Gibt es Trade-Offs zwischen widersprüchlichen Anforderungen im Risikomanagement, bzw. zwischen unternehmerischen und ethischen Anforderungen?
- Sind rechtliche Vorgaben wie die Erstellung einer Grundrechte- oder einer Datenschutz-Folgenabschätzung zu beachten?
- Sind Beschaffungs- oder Vergabekriterien zu beachten oder zu definieren?

Organisatorische Begleitmaßnahmen:

- Gibt es ausreichend interne Vorgaben und Prozesse für den sicheren und ethischen KI-Einsatz, oder müssen neue geschaffen werden, z. B. zur Überprüfung von Ergebnissen/Ausgaben der KI, oder zur Gewährleistung der menschlichen Aufsicht?
- Ist eine Einbeziehung der Rechtsabteilung, der Datenschutzbeauftragten, des IT-Teams oder anderer Expert*innen im Vorfeld der Beschaffung bzw. Vergabe erforderlich?
- Können oder müssen Stakeholder befragt oder eingebunden werden (einschließlich der allenfalls rechtlich verpflichtenden Einbindung des Betriebsrats)? Die Einbindung hilft, um spätere Ablehnung, Ängste oder negative Haltungen zu vermeiden, und um zu sehen, was wirklich gebraucht wird.
- KI-Kompetenz der Beschäftigten: Sind begleitende Richtlinien und Weisungen, Anwendungshinweise, eine technische Hilfestellung, oder allfällige Schulungen (intern oder durch den KI-Anbieter) erforderlich?
- Sind rechtliche bzw. datenschutzrechtliche Begleitmaßnahmen erforderlich, z. B. Erstellung einer Datenschutz-Information, oder sonstige Transparenz- und Informationspflichten aus Datenschutz-Grundverordnung und KI-Verordnung? Ist die Arbeitnehmer*innen-Vertretung einzubinden, sind Betriebsvereinbarungen abzuschließen?
- Sind, mit Blick auf den gesamten Lebenszyklus des Systems, Richtlinien und Vorgaben für den Einsatz, die laufende Beobachtung und die Kontrolle des KI-Systems erforderlich (die KI-Verordnung enthält u. a. gesetzliche Meldepflichten bei Krisen/Vorfällen, sowie eine diesbezügliche Kooperationspflicht zwischen Anbieter und Betreiber, einschließlich offene Kommunikationskanäle)? Sind Mitarbeitende entsprechend zu schulen?

6.3.2 Was können Sie sonst tun, um Ihre Beschäftigten im Sinne des digitalen Humanismus auf die Zukunft mit KI vorzubereiten? Empfehlungen für Personalverantwortliche.

Der menschenzentrierte Ansatz verlangt einen breiteren Blickwinkel als das Unternehmensinteresse oder die streng notwendige Compliance mit rechtlichen Vorgaben. Er bedeutet auch und insbesondere die Berücksichtigung der Interessen von Menschen und der Umwelt. In Bezug auf Beschäftigte können folgende Empfehlungen unterstützen:

- Sehen Sie zunächst das Potenzial Ihrer Mitarbeitenden (und nicht der KI). Echte Innovation und Produktivität entstehen nicht aus reiner Automatisierung, sondern vor allem aus der Motivation und der Begeisterung kompetenter Mitarbeitender.
- Setzen Sie auf Upskilling, Reskilling, Human Skills, digitale und KI-Kompetenz! Es besteht jetzt schon ein Mangel an ausgebildeten Fachkräften – entwickeln Sie interne Ressourcen und geben Sie Ihren Beschäftigten die Möglichkeit zu wachsen. Sie brauchen menschliche Expert*innen, Mitarbeitende die auf die Technologie vorbereitet sind, und die mit Skills und Ressourcen ausgestattet sind. Der Einsatz wird sich mittel- und langfristig lohnen.
- Sehen Sie menschliche Kompetenz und Skills als Wettbewerbsvorteil. KI wird bald integraler Bestandteil von fast jedem Arbeitsplatz sein. Abheben können Sie sich mit dem besonderen menschlichen „Plus“ – Kreativität, Kommunikationsfähigkeit und Teamwork.
- Bauen Sie interdisziplinäre und diverse Teams. Flexibilität sowie verschiedene Blickwinkel und Kompetenzen werden für die Entwicklung und den Einsatz von KI-Systemen, für die interne Organisation und in der Compliance essenziell sein.
- Investieren Sie in Expert*innen für KI, Daten- und Wissensmanagement und Compliance. Grundlage für erfolgreichen KI-Einsatz sind vor allem gute Daten und das Wissen, wo sie vorhanden sind.
- Binden Sie die Mitarbeitenden ein. Nicht nur wenn es eine arbeitsrechtliche Pflicht ist, sondern weil sie meist selbst besser wissen, was ihnen bei ihrer Arbeit tatsächlich hilft. Und weil hoffentlich Menschen auch in Zukunft im Mittelpunkt stehen, Vorteile aus der Technologie ziehen, und über sich selbst bestimmten, und nicht von KI bestimmt werden sollten.
- Seien Sie skeptisch bei großen Versprechungen von KI-Anbietern, und achten Sie auf das (menschliche) größere Ganze.

6.4 Fazit

KI-Einsatz kann im Personalmanagement zu mehr Produktivität und Innovation, zur Vereinfachung und Streamlining von Prozessen, zur Abnahme von repetitiven, ermüdenden Tätigkeiten und zu besseren Arbeitsbedingungen beitragen. Sie kann aber auch die menschliche Autonomie und die Menschenwürde sowie die Arbeitsqualität (sowohl die Qualität der Ergebnisse als auch der empfundenen Befriedigung) beeinträchtigen. KI-Einsatz ohne eine ethische Haltung und den menschenzentrierten Ansatz droht, mit zunehmender Verbreitung, Arbeitslose und Ungleichheit zu produzieren. Damit gehen auch Kund*innen für die effizient gewonnenen Produkte und Dienstleistungen verloren. Verzichteten Sie schon deshalb nicht auf echte Innovation und Wettbewerbsfähigkeit, die auf der Integration von Technologie und menschlichen Fähigkeiten beruht. Mit dem menschenzentrierten Ansatz fördern Sie Ihre Beschäftigten und Ihr Unternehmen.

Literatur

- AK Wien (2025). Kimperler S., Kirstgen M., Zukunftsthemen: KI und digitale Transformation am Arbeitsplatz. Wien: Verlag Arbeiterkammer Wien. <https://emedien.arbeiterkammer.at/viewer/fullscreen/AC17548061/6/>. Zugegriffen: 23. Dezember 2025.
- Christl, W. (2024). *Surveillance and digital control at work*. Wien: Cracked Labs. <https://wolfie.crackedlabs.org/pub>. Zugegriffen: 15. Dezember 2024.
- Der Standard (2023). „Österreichische Unternehmen setzen künstliche Intelligenz noch kaum ein“. <https://www.derstandard.at/story/3000000195512/oesterreichische-unternehmen-setzen-kuenstliche-intelligenz-noch-kaum-ein>. Erstellt: 16. November 2023, zugegriffen: 15. Dezember 2024.
- Der Standard (2024). „Für fast zwei Drittel der Unternehmen in Österreich spielt KI – noch – keine Rolle“. <https://www.derstandard.at/story/3000000234253/fuer-fast-zwei-drittel-der-unternehmen-in-oesterreich-spielt-ki-noch-keine-rolle>. Erstellt: 24. August 2024, zugegriffen: 15. Dezember 2024.
- DKO Strategie (2023). Strategie Digitale Kompetenzen. Wien: Bundesministerium für Finanzen. <https://www.digitalaustria.gv.at/kompetenzen/strategien.html>. Zugegriffen: 15. Dezember 2024.
- Ethics Guidelines (2019). High Level Expert Group on AI, Ethics guidelines for trustworthy AI. Brüssel: Europäische Kommission. <https://digital-strategy.ec.europa.eu/en/library/ethics-guidelines-trustworthy-ai>. Zugegriffen: 15. Dezember 2024.
- Europäische Erklärung (2023). *Europäische Erklärung zu den digitalen Rechten und Grundsätzen für die digitale Dekade*. Amtsblatt 2023/C 23/01. Brüssel: Europäische Union. <https://digital-strategy.ec.europa.eu/de/library/european-declaration-digital-rights-and-principles>. Zugegriffen: am 15. Dezember 2024.
- FEPS (2024). Cox, T., & Oosterwijk, G.R., Algorithmic Management in the Workplace – Case Studies on the Impact of Algorithmic Technologies in Seven Sectors in the Nordics. FEPS Policy Study. Brüssel: Foundation for European Progressive Studies. <https://feps-europe.eu/publication/algorithmic-management-in-traditional-workplaces/>. Zugegriffen: 15. Dezember 2024.
- Fülöp, T. (2025). Digitale Rechte – mehr als Datenschutz. In A. Siegl & S. Klocker (Hrsg.), *KI verstehen und gestalten. Künstliche Intelligenz und die Arbeitswelt*. Wien: ÖGB-Verlag.

- ILO/JRC (2024). Rani, U., Pesole, A., & González Vázquez, I. *Algorithmic management practices in regular workplaces: Case studies in logistics and healthcare*. Luxemburg: Publications Office of the European Union. <https://data.europa.eu/doi/10.2760/712475>. Zugegriffen: am 15. Dezember 2024.
- KI-Servicestelle (2025). *Künstliche Intelligenz im Human Resource Management: Ergebnisse einer repräsentativen Unternehmensbefragung vor dem Hintergrund des europäischen AI Acts*. Wien: Rundfunk und Telekom Regulierungs-GmbH (RTR). <https://www.rtr.at/rtr/startseite.de.html>. Zugegriffen: am 15. Dezember 2024.
- KI-Verordnung (2024). Verordnung (EU) 2024/1689 des Europäischen Parlaments und des Rates vom 13.6.2024 zur Festlegung harmonisierter Vorschriften für künstliche Intelligenz (KI-Verordnung), ABI L 2024/1689 vom 12.7.2024, ELI <https://eur-lex.europa.eu/eli/reg/2024/1689/oj>. Zugegriffen: 15. Dez. 2024.
- Nemitz, P., & Pfeffer, M. (2020). *Prinzip Mensch: Macht, Freiheit und Demokratie im Zeitalter der Künstlichen Intelligenz*. Bonn: Dietz.
- OECD (2023). The Impact of AI on the Workplace: Evidence from OECD Case Studies of AI Implementation. Social, Employment and Migration Working Papers, Nr. 288. Paris: Organisation for Economic Co-operation and Development (OECD). <https://www.oecd.org/els/workingpapers>. Zugegriffen: 15. Dez. 2024.
- Statistik Austria (2023). *11 % der österreichischen Unternehmen nutzen künstliche Intelligenz. Pressemitteilung vom 17. Oktober 2023*. Wien: Statistik Austria. <https://www.statistik.at/fileadmin/announcement/2023/10/20231017IKTU2023.pdf>. Zugegriffen: 15. Dezember 2024.
- Zuboff, S. (2019). *The age of surveillance capitalism: the fight for a human future at the new frontier of power*. London: Profile Books.

Tünde Fülöp ist Expertin für Digitalisierung und künstliche Intelligenz im Büro für digitale Agenden der Arbeiterkammer Wien. Ihre Schwerpunkte sind das EU-Digitalrecht, KI, Daten-Governance, digitale Märkte, Datenschutz und Datensicherheit, die Digitalisierung am Arbeitsplatz sowie digitale Rechte. Sie ist Juristin mit Background im Europarecht und erfahrene Trainerin und Vortragende. Ihre Begeisterung für den digitalen Humanismus entdeckte sie als COO am Research Institute – Digital Human Rights Center. Davor übte sie verschiedene Positionen im öffentlichen Dienst aus, u.a. als Datenschutzbeauftragte bzw. Referatsleiterin und stellvertretende Abteilungsleiterin im BMEIA-Völkerrechtsbüro, sowie als Rechtsberaterin im BKA-Verfassungsdienst und in der Parlamentsdirektion.

Digitale Kompetenzen zur Förderung des Digitalen Humanismus

7

Eine praktische Anleitung für betriebliche Weiterbildung

Alexander Schmölz

Zusammenfassung

Der Beitrag beleuchtet die Notwendigkeit digitaler Kompetenzen in einer zunehmend digitalisierten Welt. Er bietet eine praktische Anleitung für Unternehmen, um sicherzustellen, dass Mitarbeitende nicht nur technische Fertigkeiten erwerben, sondern auch humanistische Werte im Umgang mit digitalen Technologien verinnerlichen. Der Fokus liegt auf der Vermeidung von Datenverzerrungen und der Sicherstellung ethischer Grundwerte in datengetriebenen Systemen.

Die Verfügbarkeit großer Datenmengen ist grundlegender Innovationstreiber und zentrale Herausforderung für Menschen, Märkte, Industrien sowie Lebens- und Arbeitsbereiche der globalen Digitalwirtschaft.

7.1 Einleitung

In einer Zeit, in der digitale Technologien zunehmend alle Lebens- und Arbeitsbereiche durchdringen, wird die Notwendigkeit digitaler Kompetenzen immer zentraler. Doch wie können Unternehmen sicherstellen, dass ihre Mitarbeitenden nicht nur die technischen Fertigkeiten erwerben, sondern auch humanistische Aspekte im Umgang mit digitalen Technologien verinnerlichen? Dieser Beitrag liefert eine praktische Anleitung für die betriebliche Weiterbildung, um die Prinzipien des Digitalen Humanismus nachhaltig in Organisationen zu verankern.

A. Schmölz (✉)

FH des BFI Wien, Wien, Österreich

E-Mail: alexander.schmoelz@fh-vie.ac.at

7.2 Bedeutung des Digitalen Humanismus und digitaler Kompetenzen in Österreich

Digitaler Humanismus (Nida-Rümelin & Weidenfeld, 2018) verbindet technologische Innovation mit den Grundwerten des Menschseins. Im Fokus steht die Frage, wie Technologie genutzt werden kann, um menschliche Potenziale zu entfalten, soziale Gerechtigkeit zu fördern und ethische Verantwortung zu übernehmen. Der Digitale Humanismus ist mittlerweile weit über die Grenzen von Wien (Werthner et al., 2019) bekannt und zielt darauf ab, technologische Entwicklungen an menschlichen Werten auszurichten, anstatt den Menschen der Technologie unterzuordnen.

Die Prinzipien des digitalen Humanismus (Schmoelz, 2020) umfassen:
Setzen von Grenzen und Werten für digitale Technologien. Diese Dimension zielt darauf ab, Technologien so zu gestalten, dass sie im Einklang mit humanistischen Werten und Grenzen stehen. Ein Beispiel hierfür ist die EU-Datenschutz-Grundverordnung (DSGVO), die die Rechte von Individuen in der digitalen Welt stärkt. Unternehmen wie

Digitaler Humanismus	DigComp 2.2 AT / 2.3 AT
Grenzen und Werte für die digitale Technologie setzen	0. Grundlagen, Zugang und digitales Verständnis 0.4 Suchen Sie die Auseinandersetzung mit der Digitalität und entwickeln Sie ein Urteilsvermögen.
	1. Umgang mit Informationen und Daten 1.2 Daten, Informationen und digitale Inhalte kritisch bewerten und interpretieren.
	4. Sicherheit und nachhaltige Nutzung von Ressourcen 4.2 Schützen Sie persönliche oder vertrauliche Daten und die Privatsphäre. 4.3 Gesundheit und Wohlbefinden schützen 4.4 Schutz vor Betrug und Missbrauch von Verbraucherrechten 4.5 Umwelt schonen und IT nachhaltig betreiben
Ko-Kreative Gestaltung (neuer) Modelle digitaler Technologien	3. Erstellung, Produktion und Veröffentlichung 3.1 Digitale Entwicklung von Inhalten und Objekten 3.2 Digitale Integration und Neuerstellung von Inhalten und Objekten 3.4 Programmierung und Automatisierung von Prozessen 3.5 Inhalte und Gegenstände in verschiedenen Öffentlichkeiten auf rechtskonforme Weise digital produzieren und veröffentlichen
	5. Problemlösung, Innovation und Weiterbildung

Abb. 7.1 Digitaler Humanismus in DigComp 2.2 AT und 2.3 AT

Digitaler Humanismus	DigComp 2.2 AT / 2.3 AT
	5.3. Kreativer und innovativer Umgang mit digitalen Technologien
Inklusives Design, das multimodale Kompetenzen und integrierte Sinnstiftung im digitalen Raum für alle ermöglicht	0. Grundlagen, Zugang und digitales Verständnis 0.3 Kenntnis, Nutzung und Bereitstellung inklusiver Formen des Zugangs zu digitalen Inhalten
	2. Kommunikation, Interaktion und Zusammenarbeit 2.3 Nutzung digitaler Technologien für die gesellschaftliche Teilhabe
	5. Problemlösung, Innovation und Weiterbildung 5.2 Ermittlung des Bedarfs und der technologischen Antworten darauf 5.4 Lücken in der digitalen Kompetenz erkennen und schließen

Abb. 7.2 Digitaler Humanismus in DigComp 2.2 AT und 2.3 AT Teil II

IBM haben eigene ethische Leitlinien zur KI-Entwicklung eingeführt, um sicherzustellen, dass Algorithmen nicht diskriminieren und transparente Entscheidungen treffen. Der europäische Accessibility Act verpflichtet, dass auch alle digitalen Produkte und Service barrierefrei sein müssen.

Ko-kreative Transformation von Modellen der Digitalisierung. Ko-Kreativität bezeichnet die kollaborative und kreative Gestaltung und Verbesserung digitaler Systeme. Ein Beispiel für ko-kreative Transformation ist das Open-Source-Entwicklungsmodell von Linux, bei dem Entwickler weltweit gemeinsam Betriebssysteme erstellen. Dieses Vorgehensweise fördert Innovation und demokratisiert die Technologieentwicklung.

Inklusive Gestaltung digitaler Räume, die multimodale Literacies fördern. Digitale Räume sollten barrierefrei gestaltet sein und alle Nutzergruppen ansprechen. Ein erfolgreiches Beispiel ist die Plattform „Be My Eyes“ (Be My Eyes, 2025), die sehbehinderten Menschen mithilfe von Freiwilligen visuelle Unterstützung bietet. Dieses Modell zeigt, wie inklusive Technologien Barrieren abbauen können.

Zusätzlich gilt der Nationale Referenzrahmen für digitalen Kompetenzen (Bundesministerium für Finanzen & Plattform Digitale Kompetenzen, 2022) in ganz Österreich. Dieser basiert auf dem Kompetenzmodell DigComp 2.3 AT (Nárosy et al., 2022). Eine Einordnung der zentralen Kompetenzen des Digitalen Humanismus in das Kompetenzmodell ist daher zielführend. Wie in Abb. 7.1 dargestellt, werden die Prinzipien des Digitalen Humanismus in den Kompetenzbereichen des DigComp-Modells größtenteils abgebildet. Die Fortschritte in der Version 2.3 AT zeigen eine stärkere Berücksichtigung humanistischer Prinzipien:

Abb. 7.1 und 7.2 zeigen den Vergleich zwischen DigComp 2.2 AT und 2.3 AT und, dass die Iteration 2.3 AT des DigComp-Modells deutliche Fortschritte bei der Einbindung

der Prinzipien des Digitalen Humanismus gemacht hat. Sie bildet eine Grundlage für Weiterbildungskonzepte, die ethische, soziale und technologische Aspekte beim Erwerb von Kompetenzen in den Mittelpunkt rücken.

7.3 Schritte zur Weiterbildung für Digitalen Humanismus

Um die Prinzipien und den Kompetenzen des Digitalen Humanismus in der Weiterbildung zu verankern, sollten folgende Schritte berücksichtigt werden:

- **Menschlichkeit in unserem Unternehmen:** Eingangs ist es wichtig herauszufinden, was Menschlichkeit im Unternehmen bedeutet. Welche zentralen menschlichen Eigenschaften sind für unsere Unternehmenskultur wichtig? Eine gründliche Analyse des unternehmensspezifischen Humanismus bildet die Basis jeder Weiterbildung. Vielleicht ist ihrem Unternehmen die Ko-Kreativität eine wichtige menschliche Eigenschaft. Das gemeinsame Schaffen von Neuem steht oft im Mittelpunkt ihrer Arbeit? Ist die innerbetriebliche Technologie so ausgerichtet, dass sie Ko-Kreativität behindert, zulässt oder vielleicht sogar fördert?
- **Kompetenzen und Lernergebnisse zur zentralen menschlichen Eigenschaft:** Effektive Weiterbildungen brauchen Kompetenzorientierung, Dazu gehört die Formulierung von konkreten Lernergebnissen zur wünschenswerten menschlichen Eigenschaft und den in Abb. 7.1 und Abb. 7.2 systematisierten Kompetenzen.
- **Interdisziplinärer Ansatz:** Die Integration von Expert:innen aus verschiedenen Disziplinen, darunter Technologie, Ethik, Soziologie und Philosophie, sorgt für eine ganzheitliche Betrachtung der Herausforderungen. Dieser Ansatz fördert ein breiteres Verständnis der komplexen Wechselwirkungen zwischen Mensch und Technologie.
- **Praxisbezug:** Weiterbildung sollte stets praxisnah gestaltet sein. Fallstudien und reale Szenarien (Wirtschaftsagentur Wien, 2023) helfen den Teilnehmenden, das Gelernte direkt in ihren Arbeitsalltag zu übertragen. Unternehmen können zudem Projektarbeiten einbinden, die konkrete betriebliche Herausforderungen adressieren.
- **Einschläge Unterlagen:** Erfolg einer Weiterbildung kann nur durch einschlägige Unterlagen und qualitativ hochwertige Lektüre sichergestellt werden. Neben diesem Buch sind u. a. die folgenden Unterlagen empfehlenswert:
 - Eichinger, A., Knees, P., & Werthner, H. (2024). DIGITALISIERUNG UND WIR. LEHRBUCH ZUM DIGITALEN HUMANISMUS MIT PRAKTISCHEN ÜBUNGEN: Digitaler Humanismus als Modell für die Zukunft. Residenz.
 - Krause, G. (Hrsg.). (2023). Die Praxis des Digitalen Humanismus: Welchen Beitrag Unternehmen dazu leisten und wie sie davon profitieren können. Springer Vieweg.
 - Nida-Rümelin, J., & Weidenfeld, N. (2018). Digitaler Humanismus: Eine Ethik für das Zeitalter der künstlichen Intelligenz. Piper.
 - Schmoelz, A. (2020). Die Conditio Humana im digitalen Zeitalter. Zur Grundlegung des Digitalen Humanismus und des Wiener Manifests. MedienPädagogik.

Zeitschrift für Theorie und Praxis der Medienbildung, 20, 208–234. <https://doi.org/10.21240/mpaed/00/2020.11.13.X>

- Schmoelz, A., & Bauer, V. (2021). Digitalisierung, Humanismus und die Zukunft von Arbeit und Berufsbildung: Zentrale Prognosen aus der Berufsbildungsforschung. In R. Löffler, P. Schlögl, & A. Schmoelz (Hrsg.), 50 Jahre Berufsbildungsforschung in Österreich im Spannungsfeld von Wissenschaft, Politik und Praxis (S. 221–236). wbv. <https://doi.org/10.3278/6004915w221>
- Werthner, H., Lee, E., Akkermans, H., Vardi, M., & et al. (2019). Wiener Manifest für Digitalen Humanismus. https://www.informatik.tuwien.ac.at/dighum/wp-content/uploads/2019/07/Vienna_Manifesto_on_Digital_Humanism_DE.pdf
- **Auswahl von Trainer:innen und Expert:innen:** Die genaue Auswahl einschlägiger Personen für die innerbetriebliche Weiterbildung ist von großer Relevanz. Eine Liste von potenziellen Expert:innen und Trainer:innen kann aus den Vortragenden vergangener Events (<https://digitalhumanism.at/en/past-events/>) oder Autor:innen bestehender Publikationen (<https://digitalhumanism.at/en/publications/>) entnommen werden.
- **Reflexion und Feedback:** Reflexionsphasen sind entscheidend, um den Lernfortschritt zu evaluieren und Anpassungen vorzunehmen. Feedback sollte nicht nur von den Teilnehmenden, sondern auch von Trainern und Führungskräften eingeholt werden, um die Qualität der Weiterbildung kontinuierlich zu verbessern. Hat die Weiterbildung zur Förderung von Ko-Kreativität im Betrieb geholfen? Gibt es nun einen besseren Blick auf die innerbetriebliche Technologie und ist klar geworden, ob diese Ko-Kreativität behindert, zulässt oder vielleicht sogar fördert?

7.4 Beyond Weiterbildung: Wege zur Kultur des digitalen Humanismus im Betrieb

Weiterbildung darf nicht isoliert betrachtet werden, sondern muss Teil einer langfristigen Strategie sein. Hierzu gehören regelmäßige Follow-up-Maßnahmen, die den Wissenstransfer sichern, sowie die Integration der erlernten Kompetenzen in bestehende Unternehmensprozesse, wie zum Beispiel:

1. **Strategische Verankerung:** Die Integration des Digitalen Humanismus sollte auf der höchsten Unternehmensebene verankert werden. Dazu gehört die Festlegung klarer Werte und Prinzipien in Leitlinien und Strategiedokumenten. Diese Werte müssen in der gesamten Organisation kommuniziert werden, um eine einheitliche Ausrichtung sicherzustellen.
2. **Schulung der Führungskräfte:** Führungskräfte spielen eine Schlüsselrolle bei der Umsetzung des Digitalen Humanismus. Sie sollten in ethischen Fragestellungen, sozialer Verantwortung und technologischen Grundlagen geschult werden. Programme wie Workshops, Mentoring und Peer-Learning-Sitzungen können diesen Wissenstransfer unterstützen.

3. **Analyse der eigene Technologieverwendung:** Eine weitreichende Analyse der innerbetrieblichen Beschaffung, Anwendung bzw. Entwicklung wird zeigen, ob diese die zentralen humanistischen Werte fördern oder eher behindern. Behindernde Elemente der Verwendung von Technologie werden identifiziert und abgeschafft.
4. **Kulturelle Verankerung:** Um den Digitalen Humanismus in die Unternehmenskultur einzubetten, sollten die Prinzipien in alltägliche Abläufe integriert werden. Dazu gehört die Anerkennung und Belohnung von Verhalten, das den Werten des Digitalen Humanismus entspricht, sowie regelmäßige Schulungen und Diskussionen zu relevanten Themen.
5. **Verankerung im unternehmerischen Umfeld:** Zentral ist die Zusammenarbeit mit anderen Unternehmen, Bildungsinstitutionen bis hin zu NGOs und der Verwaltung. In einem kooperativen System kann gute Praxis ausgetauscht werden und die gegenseitige Orientierung an humanistischen Werten verstärkt sich.

7.5 Fazit

Die gezielte Förderung digitaler Kompetenzen im Sinne des Digitalen Humanismus bietet Organisationen die Möglichkeit, eine Kultur zu schaffen, in der Technologie im Dienst der Menschen steht. Durch die Integration humanistischer Werte wie Werte und Grenzen setzen, Ko-Kreativität und Inklusion kann die Digitalisierung ethisch gerecht gestaltet werden.

Der zentrale Erfolgsfaktor ist die langfristige Verknüpfung von ethischen, sozialen und technischen Kompetenzen in der Weiterbildung und darüber hinaus. Dies erfordert eine strategische Verankerung im Unternehmensleitbild, kontinuierliche Anpassungen an technologische Entwicklungen sowie die Einbeziehung interdisziplinärer Perspektiven. Langfristig stärkt dies nicht nur die Mitarbeiter:innenbindung, sondern auch die Resilienz von Unternehmen in einer sich schnell verändernden digitalen Welt.

Die dargestellten Wege verdeutlichen, dass es möglich ist, die Prinzipien des Digitalen Humanismus in die Praxis umzusetzen. Erfolgreiche Ansätze zeigen, dass eine enge Zusammenarbeit zwischen verschiedenen Akteuren – von Unternehmen über Bildungsinstitutionen bis hin zu NGOs – entscheidend ist, um die Digitalisierung inklusiv und zukunftsorientiert zu gestalten. Der Digitale Humanismus bietet somit eine ethisch fundierte Grundlage, um die digitale Transformation aktiv und verantwortungsvoll zu gestalten.

Literatur

- Be My Eyes (2025). Be My Eyes – accessibility technology for blind & low vision people. <https://www.bemyeyes.com/>. Zugegriffen: 1. Oktober 2025.
- Bundesministerium für Finanzen / Plattform Digitale Kompetenzen (2022). *Nationaler Referenzrahmen für Digitale Kompetenzen (DigComp 2.3 AT)*. Wien: Republik Österreich. <https://www.digitalekompetenzen.gv.at/kompetenzen.html> Zugegriffen: 1. Oktober 2025.
- Eichinger, A., Knees, P., & Werthner, H. (2024). *Digitalisierung und wir. Lehrbuch zum Digitalen Humanismus mit praktischen Übungen: Digitaler Humanismus als Modell für die Zukunft*. Residenz.
- Krause, G. (Hrsg.). (2023). *Die Praxis des Digitalen Humanismus: Welchen Beitrag Unternehmen dazu leisten und wie sie davon profitieren können*. Springer Vieweg.
- Nida-Rümelin, J., & Weidenfeld, N. (2018). *Digitaler Humanismus: Eine Ethik für das Zeitalter der künstlichen Intelligenz*. Piper.
- Nárosy, T., Schmölz, A., Proinger, J., & Domany-Funtan, U. (2022). Digitales Kompetenzmodell für Österreich: DigComp 2.3 AT. *Medienimpulse*, 60(4), 103. <https://doi.org/10.21243/mi-04-22-23>.
- Schmoelz, A. (2020). Die *Conditio Humana* im digitalen Zeitalter. Zur Grundlegung des Digitalen Humanismus und des Wiener Manifests. *MedienPädagogik. Zeitschrift für Theorie und Praxis der Medienbildung*, 20, 208–234. <https://doi.org/10.21240/mpaed/00/2020.11.13.X>.
- Schmoelz, A., & Bauer, V. (2021). Digitalisierung, Humanismus und die Zukunft von Arbeit und Berufsbildung: Zentrale Prognosen aus der Berufsbildungsforschung. In R. Löffler, P. Schlögl & A. Schmoelz (Hrsg.), *50 Jahre Berufsbildungsforschung in Österreich im Spannungsfeld von Wissenschaft*.
- Werthner, H., Lee, E., Akkermans, H., Vardi, M., et al. (2019). Wiener Manifest für Digitalen Humanismus. https://www.informatik.tuwien.ac.at/dighum/wp-content/uploads/2019/07/Vienna_Manifesto_on_Digital_Humanism_DE.pdf
- Wirtschaftsagentur Wien (2023). *Whitepaper Digitaler Humanismus – Best Practices aus der Unternehmenspraxis*. Wien: Wirtschaftsagentur. https://wirtschaftsagentur.at/crmfiles/event_a22ac97d-c432-ef11-9288-0050560000f0/WA_DigiHum_BestPractices.pdf (zuletzt abgerufen am 2. Oktober 2025)

Dr. Alexander Schmölz ist Professor für digitalen Humanismus an der FH des BFI Wien, geschäftsführender Leiter des ÖIBF, Chefredakteur der Zeitschrift Digital Culture & Education und Researcher-in-Residence bei fit4internet. Er forscht und lehrt zu Humanismus und Digitalisierung mit Fokus auf Inklusion, Ko-Kreativität und Kompetenzen in Bildung sowie förderliche didaktische und politische Rahmenbedingungen. Seit 20 Jahren ist er in nationale und internationale Entwicklungsprojekte zur Digitalisierung der Bildung und Förderung digitaler sowie didaktischer Kompetenzen involviert.

Teil II

Branchenspezifische Herausforderungen und Lösungsansätze

Ethische Digitalisierung in der Versicherungsbranche

8

Chancen und Risiken in der digitalen Transformation

Gerhard Lahner und Thomas Mara

Zusammenfassung

Der Beitrag beleuchtet die Bedeutung des Digitalen Humanismus in der Versicherungsbranche, indem er praxisnahe Ansätze zur Verbindung von Technologie und menschlichen Werten aufzeigt. Bei dem Ziel, den Menschen in den Mittelpunkt technologischer Entwicklungen zu stellen besteht ein wesentliches Spannungsfeld in der Handhabung von umfassenden Kundendaten und -informationen, insbesondere in Bezug auf Datenschutz und die Wahrung der Privatsphäre. Zudem wird die Rolle von regulatorischen Anforderungen und internationalen Wettbewerbsbedingungen hervorgehoben. Der Digitale Humanismus wird als zentraler Erfolgsfaktor für eine ethisch fundierte, digitale Zukunft gesehen.

Der Digitale Humanismus stellt den Menschen in den Mittelpunkt der digitalen Transformation.

G. Lahner (✉) · T. Mara
Vienna Insurance Group, Wien, Österreich
E-Mail: gerhard.lahner@vig.com

T. Mara
E-Mail: Thomas.Mara@vig.com

8.1 Kann digitaler Humanismus das Vertrauen der Kunden stärken?

Versicherungsgeschäft beruht auf gegenseitigem Vertrauen aller Marktteilnehmer*innen. Indem Technologie so eingesetzt wird, dass der Mensch im Mittelpunkt steht und ethische Prinzipien bei der Entwicklung und Nutzung digitaler Lösungen priorisiert werden, stärkt angewandter digitaler Humanismus dieses gegenseitige Vertrauen. Der Versicherungsbereich gilt als ein traditionell personenbezogenes Geschäft, das stark auf Vertrauen und menschliche Interaktion angewiesen ist, und befindet sich in einem Spannungsfeld zwischen technologischer Innovation und der Bewahrung dieser Werte. Das Bewahren dieser Werte wird oft als konservativ wahrgenommen, entspricht aber der besonderen Verantwortung für dieses Vertrauensverhältnis.

In diesem Artikel betrachten wir die Chancen und Risiken des digitalen Humanismus im Versicherungsbereich anhand von Beispielen. Wir analysieren, wie Technologie sinnvoll eingesetzt werden kann, um das Leben von Kunden und Mitarbeitenden zu verbessern, und welche Herausforderungen mit dieser Entwicklung einhergehen. Wir werden auch strategische Ansätze vorstellen, wie Versicherungsunternehmen den digitalen Humanismus erfolgreich umsetzen können.

8.2 Beispiele von digitaler Humanismus im Versicherungsbereich

8.2.1 Transparenz und Verständlichkeit

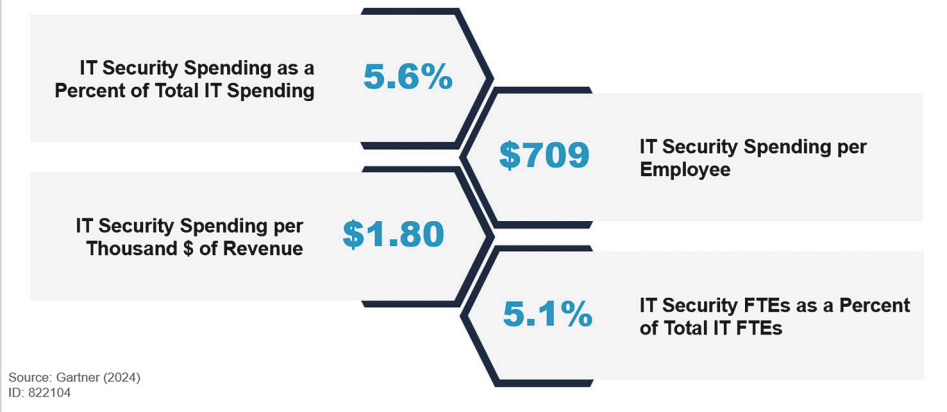
Transparente Algorithmen: Beim Einsatz von Algorithmen oder KI-gestützter Systeme zur Risikobewertung oder Schadensregulierung, erklären Versicherer proaktiv und transparent deren Funktionsweise allen Beteiligten, um die Sorge potenzieller Vorurteile oder Diskriminierung zu nehmen.

Einfache und verständliche Kommunikation: Sowohl auf den klassischen Kommunikationskanälen aber auch in digitalen Plattformen und durch ChatBots werden Informationen zu Versicherungsprodukten, Vertragsbedingungen und Datenschutz klar und verständlich dargestellt, sodass Kunden informierte Entscheidungen treffen können. Einfache und für alle am Prozess Beteiligten nachvollziehbare Beschreibung von Leistungen und Produkten stärkt das Vertrauen bei Abschluss und im Leistungsfall.

8.2.2 Datenschutz und Sicherheit

Ethischer Umgang mit Daten: Marktteilnehmer*innen wird klar kommuniziert und dargestellt, dass ihre sensiblen Daten sicher gehalten und nur für legitime Zwecke genutzt werden. In der Versicherungsbranche kommt oft das „Need-to-know-Prinzip“ zum Einsatz. Dieses hat das Ziel, Datenzugriffe nur dann für einen Benutzer freizugeben, wenn die Information unmittelbar für die Erfüllung einer Aufgabe benötigt wird.

Executive Summary IT Security



Gartner.

Abb. 8.1 Investments Cyber Security. (Quelle: Gartner Inc., 2025. GARTNER is a trademark of Gartner, Inc. and/or its affiliates)

Sicherheit und im Besonderen Cyber Security wird in der Versicherungsbranche große Aufmerksamkeit gewidmet. Der Gesetzgeber schreibt durch Regelungen, wie zum Beispiel GDPR – General Data Protection Regulation¹, IDD – Insurance Distribution Directive² und DORA³ – Digital Operational Resilience Act, aktuell Resilienz und Schutzmechanismen zur Datensicherheit vor. Versicherer investieren darüber hinaus signifikante Aufwände, um krimineller Energie entsprechende Gegenmaßnahmen entgegenstellen zu können (s. Abb. 8.1).

Zielgerichtete Kommunikation (also nicht Panik verbreitende Schlagzeilen) an die Marktteilnehmer über die tiefgreifenden und umfassenden Maßnahmen steigert das Vertrauen und erhöht die Bereitschaft zur Weitergabe von persönlichen Daten.

Datensoeveränität: In Portalen oder Mobile Apps geben Versicherer ihren Kunden die Kontrolle über ihre Daten, in denen diese eingesehen und aktiv verwaltet werden können. Durch diese Souveränität bleiben Kunden „Teil des Systems“, können darin aktiv teilnehmen und sind nicht potentiell intransparenter Manipulation der eigenen Daten ausgesetzt.

¹ GDPR – General Data Protection Regulation REGULATION (EU) 2016/679 of the European parliament and of the Council of 27 April 2016.

² IDD – Insurance Distribution Directive DIRECTIVE (EU) 2016/97 of the European Parliament and of the Council of 20 January 2016.

³ DORA – Digital Operational Resilience Act REGULATION (EU) 2022/2554 of the European Parliament and of the Council of 14 December 2022 on digital operational resilience for the financial sector and amending Regulations (EC) No 1060/2009, (EU) No 648/2012, (EU) No 600/2014, (EU) No 909/2014 and (EU) 2016/1011.

8.2.3 Empathische Kundeninteraktion

Menschzentrierte Technologien: Der Einsatz von Mobile Apps, Chatbots und KI wird dann nicht unpersönlich wahrgenommen, wenn eine unterstützende Rolle eingenommen und der Zugang zu menschlichen Ansprechpartnern ermöglicht wird, wenn Kunden dies wünschen. Abläufe in Kundeninteraktionen (persönliche Datenänderungen, Schadensmeldungen, allgemeine Anfragen, Vertragsänderungen oder -ergänzungen, etc.) werden „customer-centric“ gestaltet, also aus Kundenperspektive und nicht aus verwaltungstechnischer Sicht.

Emotionale Intelligenz: KI-Systeme werden so gestaltet, dass sie empathisch reagieren und die Bedürfnisse der Kunden erkennen. Neue KI-Systeme erlauben zunehmend auf Kunden-Reaktionen einzugehen und so die Barriere zwischen Mensch und Maschine zu verkleinern. Die so akzeptierte und nahezu ununterbrochene (Schlagwort: 24/7) Verfügbarkeit von Auskunft- und Unterstützungssystemen kommt Kunden und Versicherern gleichermaßen zugute.

8.2.4 Individuelle Lösungen

Personalisierte Angebote: Versicherer bieten mit Hilfe digitaler Technologie maßgeschneiderte Produkte, die die spezifischen Bedürfnisse und Lebenssituationen der Kunden berücksichtigen. Durch die gesamtheitliche Betrachtung der Kunden können dynamisch Produkte angepasst oder auch von Kunden aktiv, flexibel Verträge gestaltet werden. Diese Flexibilität von Produkten und Verträgen, unter Wahrung der Wirtschaftlichkeit, stellt die Bedürfnisse der Kunden in den Mittelpunkt – Kunden partizipieren, sind gestaltender Teil und identifizieren sich somit weitaus mehr mit dem finalen Vertrag bzw. der gewählten Dienstleistung. Eine besondere Herausforderung stellt in diesem Zusammenhang das allgemeine Versicherungsprinzip dar, welches darauf aufbaut, dass durch die Bildung einer Gefahrengemeinschaft ein Risikoausgleich von gleichartigen Gefahren (z. B. Naturkatastrophen) zwischen den einzelnen Versicherten sicherstellt. Dieses Prinzip wird durch überbordende Individualisierung ad absurdum geführt. Hier ist die Funktion des Versicherers – unabhängig ob digital – eine Balance zwischen den Bedürfnissen des Individuums und Gefahrengemeinschaft herzustellen, um den Grundsatz des Versicherungsprinzips aufrecht erhalten zu können.

8.2.5 Ethische Grundsätze und Fairness

Vermeidung von Diskriminierung: Verhaltensweisen und Algorithmen, die faire Entscheidungen treffen, stärken das Vertrauen. Zum Beispiel basieren Risikobewertungen auf nachvollziehbaren und ethischen Kriterien. Digitale Prozesse stellen sicher, dass alle Kunden, unabhängig von Geschlecht, Alter, Herkunft oder technologischem Know-how, gleichbehandelt werden.

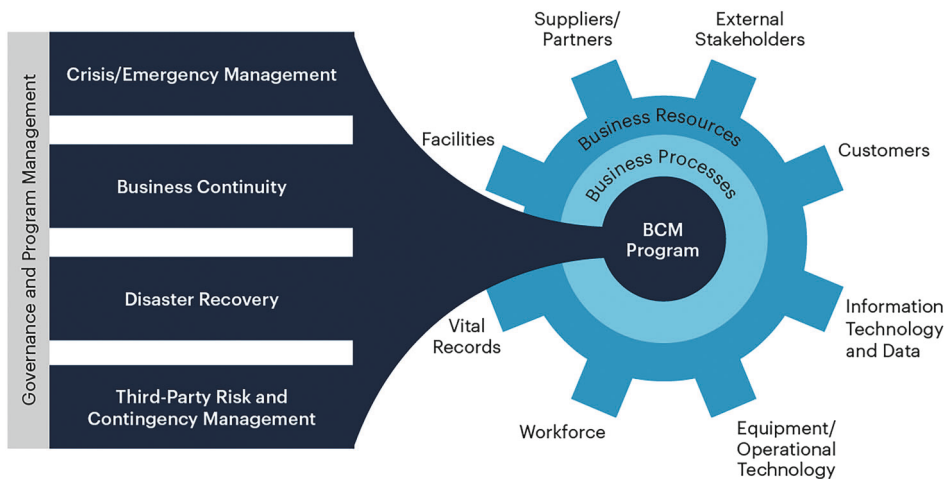
8.2.6 Resilienz und Verlässlichkeit

Zuverlässige Technologie: Systeme, die auch in Krisenzeiten oder bei erhöhtem Bedarf stabil funktionieren, zeigen Kunden, dass sie sich auf ihren Versicherer verlassen können. Signifikante Investments in „Business Continuity Management“, wie auch vom Gesetzgeber gefordert, zeugen von den laufenden Bemühungen der Versicherer dieser sich immer ändernden Herausforderung zu stellen (Abb. 8.2).

Durch regelmäßige Ernstfallübungen (Übungsannahmen wie: Pandemie, Atomarer Zwischenfall, Erdbeben, Flut, Blackout, Cyber Attacken, Major IT-Incidents ...) werden die Unternehmen „fit“ gehalten, um im Ernstfall die notwendigen Geschäftsbereiche aufrecht zu erhalten. Test- und Simulationssysteme unterstützen beim Erstellen von Krisenszenarien.

Auch hier kann durch gezielte Information und Kommunikation das Kundenvertrauen und das Vertrauen aller Marktteilnehmer gestärkt werden – wir Versicherer sind auf Krisen gut vorbereitet, halten uns „fit“ und passen die Krisenszenarien den globalen Herausforderungen an – und nehmen so unsere Verantwortung als systemrelevanter Player wahr. Die Weiterentwicklung von digitalen Lösungen führt dazu, dass Services zur Abdeckung finanzieller Risiken durch den z. B. Klimawandel erst versicherbar bleiben. So werden laufend – unter Zuhilfenahme von immer mehr Daten – die Modelle für Flut-, Sturm- Erd-

Business Continuity Management Disciplines



Source: Gartner
721184_C

Gartner

Abb. 8.2 Business Continuity (Quelle: Gartner Inc., 2024. GARTNER is a trademark of Gartner, Inc. and/or its affiliates)

bebenbedrohungen laufend weiterentwickelt. Hier finanziert die Gefahrengemeinschaft gemeinsam die Erhöhung der Vorhersagbarkeit von Katastrophenereignissen und auch die Berechenbarkeit. Beides sind unbedingte Eingangsparameter für die Möglichkeit diese Risiken überhaupt zu versichern. Auch die lenkende Funktion von Investitionen aus dem öffentlichen Bereich (Kommunen, Länder) ist hier zu erwähnen – welche letzten Endes im Sinne der Wohnbarkeit von Gebieten Ausschlag geben kann.

8.2.7 Partizipation und Feedback

Kundeneinbindung: Kunden werden aktiv in die Weiterentwicklung digitaler Dienste einbezogen, z. B. durch Umfragen oder Beta-Tests, um so eine Plattform schaffen, auf der Erfahrungen geteilt und Verbesserungsvorschläge gemacht werden.

8.3 Fazit

Digitaler Humanismus fördert eine vertrauensvolle Beziehung zwischen Versicherern und Kunden, indem er digitale Lösungen menschenzentriert, transparent und fair gestaltet. Die besondere Verantwortung der Versicherungswirtschaft im Rahmen der gesellschaftlichen Entwicklung drückt sich schon allein dadurch aus, dass es hier besondere Anforderungen von Seiten der Gesetzgeber und Aufsichtsinstanzen gibt. Insofern ist im Sinne des digitalen Humanismus auch darauf hinzuweisen, dass die Versicherungswirtschaft bei technologischer Entwicklung sicher kein First-Mover sein kann und soll, sondern die Bedürfnisse der Kunden und Mitarbeiter im Auge behalten muss bei gleichzeitiger Abwägung der involvierten Chancen und Risiken, sei es technologischer oder gesellschaftlicher Natur. Wenn Kunden das Gefühl haben, dass Technologie nicht als Selbstzweck dient, sondern ihren Bedürfnissen gerecht wird, stärkt dies das Vertrauen in den Versicherer und dessen digitaler Transformation.

Literatur

- Gartner Inc (2024). Mitigate Operational Dependencies to Strengthen Business Resilience (Report-ID 5544495). <https://www.gartner.com/en/documents/5544495>. Erschienen: 27.06.2024; zugegriffen: 05.01.2026.
- Gartner Inc (2025). IT Key Metrics Data 2025: IT Security Measures – Analysis (Report-ID 5938807). <https://www.gartner.com/en/documents/5938807>. Erschienen: 5.12.2024; zugegriffen: 05.01.2026.

Gerhard Lahner, Chief Operating Officer der Vienna Insurance Group (VIG), studierte Betriebswirtschaft an der Wirtschaftsuniversität Wien und ist seit 2002 in unterschiedlichen Funktionen für die VIG tätig. Er war Vorstandsmitglied der österreichischen Versicherungsgesellschaften Donau und Wiener Städtische sowie Vorstandsmitglied der tschechischen Kooperativa und ČPP. Seit 2020 ist Gerhard Lahner Mitglied des Vorstandes der VIG Holding und unter anderem als COO für technologische Themen zuständig.

Thomas Mara ist seit 2019 für die Vienna Insurance Group (VIG) tätig und Geschäftsführer der VIG IT Digital Services in Wien und General Manager der VIG Management Services, in Bukarest. Davor nahm er Managementaufgaben in der Generali Gruppe wahr, unter anderem als Chief Information and Data Officer der Generali CEE in Prag (2018), als Chief IT Officer Corporate der Generali Assicurazioni in Trieste/Milano (bis 2016) und als Geschäftsführer der Generali VIS Informatik in Wien (2008–2013). Sein technologischer Background und frühere berufliche Engagements unterstützen sein internationales Wirken in den Gesellschaften der VIG in Zentral- und Osteuropa.



Vertrauen schaffen und Sicherheit geben im Digitalen Banking

9

Innovative, menschenzentrierte Ansätze zur Betrugsprävention und zum Schutz der Kund:innen

Natalia Major

Zusammenfassung

Der Beitrag beleuchtet die Herausforderungen und Lösungen im Umgang mit Phishing und Online-Betrug bei Tatra banka Slowakei. Die Bank hat ein agiles, funktionsübergreifendes Team eingeführt, um Betrugsfälle zu minimieren und die Kundenzufriedenheit zu erhöhen. Durch den Einsatz von IT-Systemen, Datenmodellen und gezielter Kundenaufklärung konnte die Bank die Anzahl der Betrugsfälle signifikant reduzieren.

Die Verfügbarkeit großer Datenmengen ist ein grundlegender Innovationstreiber für alle Märkte, Industrien und Lebensbereiche der globalen Digitalwirtschaft.

9.1 Einleitung

Transaktionsbetrug im Bankwesen stellt ein operatives Risiko dar, das in den letzten Jahren im gesamten Finanzsektor erheblich zugenommen hat. Dabei spiegeln sich weniger die Verluste der Bank wider als vielmehr die Anzahl der geschädigten Kunden und der Anstieg der ihnen verursachten Schäden. Bankkunden sind ein unverhältnismäßig leichtes Ziel für Hacker als die Bank selbst mit ihren Strukturen, Prozessen und Werkzeugen. In diesem Zusammenhang betont der Digitale Humanismus die Bedeutung der Gestaltung

N. Major (✉)
Tatra banka Slowakei, Bratislava, Slowakei
E-Mail: natalia_major@tatrabanka.sk

von Technologie, die menschliche Werte und Sicherheit priorisiert und sicherstellt, dass Kunden vor solchen betrügerischen Aktivitäten geschützt sind.

Die Tatra banka Slowakei ist auf dem slowakischen Markt als Innovationsführer etabliert, mit Hunderttausenden von Kunden, die digitale Kanäle nutzen. Kunden erwarten, dass ihre Bank sie in der digitalen Welt schützt. Die Bank entschied sich, ihr Markenversprechen der Innovationsführerschaft mit dem Attribut der Cybersecurity zu bereichern. Das Ziel war nicht nur, ein positives Kundenerlebnis zu schaffen, sondern im Kern die Perspektive des Kunden einzunehmen, während er digitale Kanäle nutzt.

Die Tatra banka Slowakei, ein Mitglied der RBI-Gruppe, wurde 1990 gegründet. Von Anfang an hatte sie ein klar definiertes Markenversprechen, das der Innovationsführerschaft. Die Mission der Tatra banka ist es, ihren Kunden durch Innovation ein außergewöhnliches Erlebnis zu bieten. Cybersecurity nimmt beim Online-Banking eine immer wichtigere Rolle ein. Diese Fallstudie bietet einen detaillierten und umfassenden Überblick über die Art und Weise, wie die Tatra banka Slowakei auf den wachsenden Trend des Online-Betrugs im digitalen Raum reagiert hat.

9.2 Fallbewertung

Der Kunde nimmt die Tatsache, dass er im Online-Bereich Opfer eines Betrugs geworden ist, insbesondere bei einer Bank, die als technologischer Innovationsführer auf dem Markt wahrgenommen wird, sehr emotional und sensibel wahr. Bildung, Alter, Geschlecht oder Segment des Kunden spielen keine Rolle. Unter den Opfern gibt es in den meisten Fällen keine auffälligen gemeinsamen Merkmale, die es ermöglichen würden, einen Archetyp des Kunden zu bestimmen, der ein potenzielles Betrugsopfer ist.

Kunden verlieren oft ihre Ersparnisse, in der Regel sind die Schäden speziell im Premiumsegment um ein Vielfaches höher. Es handelt sich oft um lebenslange Ersparnisse oder Eigentum im Wert von Zehntausenden bis Hunderttausenden von Euro. Immer häufiger gibt es Fälle, in denen Kunden nicht nur ihrer Gelder beraubt werden, sondern auch genehmigte Kredite, die ihnen von Hacker gestohlen werde.

Die Raffinesse der Hacker nimmt zu, und Kunden, die keine Gelder auf ihren Konten haben, können als Verteilungskanäle genutzt werden, um einen anderen Kunden zu berauben.

Die Bank versuchte, das Problem des Online-Transaktionsbetrugs durch die Initiierung prozessgesteuerter Projekte anzugehen. Sie erkannte jedoch bald, dass ein projektbasierter Ansatz für ein so komplexes Problem unzureichend war. Online-Transaktionsbetrug betrifft mehrere Abteilungen innerhalb der Bank, und eine effektive Koordination zwischen diesen Abteilungen fehlte. Betrug ist ein integraler Bestandteil der Kundenerfahrung mit der Finanzinstitution und erfordert eine permanente, zentrale Verwaltung und Koordination, im Gegensatz zur vorübergehenden Natur eines Projekts.

Das Problem des Online-Betrugs wird nicht mehr verschwinden; die Methoden des Betrugs aber werden sich kontinuierlich weiterentwickeln und verbessern. Ohne syste-

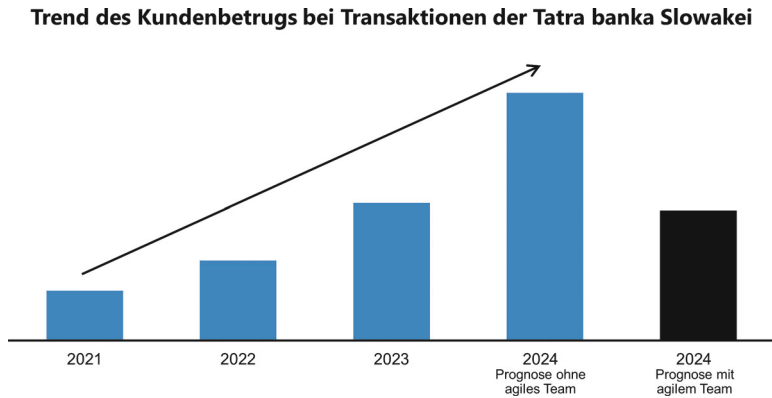


Abb. 9.1 Gesamtschäden der Kunden ab 2021 vor und nach Einsetzung eines bereichsübergreifenden Teams. (Quelle: Tatra banka Slowakei)

matische Maßnahmen zur Betrugsprävention wird die Anzahl der betroffenen Kunden wachsen, was die Kundenzufriedenheit beeinträchtigt und die Arbeitsbelastung der Mitarbeiter im Callcenter und im Filialnetz erhöht. Geschädigte Kunden nehmen die Bank als Teil des Problems wahr, weil sie die betrügerische Transaktion nicht verhindert hat, und fordern finanzielle Entschädigung. Aufgrund der großen Anzahl von Fällen war es notwendig, sie standardisiert, einheitlich und effizient zu behandeln, wobei stets das Interesse des geschädigten Kunden berücksichtigt wurde.

Aufgrund des enormen Anstiegs der Anzahl der Fälle sowie der Schäden an Kunden, die die Bank seit dem Sommer 2021 beobachtet, widmet sich seit 2023 ein spezifisches bereichsübergreifendes agiles Team – Transaktionsbetrug – dieser Angelegenheit (siehe Abb. 9.1). Der Product Owner ist ein Kundenerfahrungsspezialist, und die Team-Komposition beinhaltet Vertreter von Compliance-Untersuchungen, Online-Kanälen, Cybersecurity, Marketing und Datenwissenschaftlern.

Die Aufgabe des agilen Teams besteht darin, die Kundenverluste zu minimieren, die durch Transaktionsbetrug entstehen, sei es unbefugt oder vom Kunden autorisiert.

Unbefugter Betrug umfasst Fälle, in denen der Zahlungsauftrag nicht vom Kunden eingegeben oder bestätigt wurde, obwohl sein digitaler Zugang genutzt wurde. Hacker verschaffen sich in der Regel Zugang zu den digitalen Kanälen der Kunden durch Phishing; Pharming; Spoofing; Smishing; Vishing; Remote Access Tools; Malware; usw. Die Intensität und Gefahr eines Angriffs, bei dem digitale Zugangsdaten gelockt werden, variiert je nachdem, ob es sich um einen organisierten Angriff handelt und welche Technologien der Hacker verwendet.

Ein Hacker nutzt oft eine Kombination von Zahlungskkanälen, um Gelder von einem Kundenkonto zu stehlen und maximale Effizienz zu erzielen. Beispielsweise erfolgt häufig die Übertragung von Geldern von einem Sparkonto, die Nutzung einer Kreditkarte, die Aufnahme eines vorab genehmigten Kredits, wenn eine mobile Banking-Anwendung von

einem Hacker missbraucht wird. Überweisungen, Kartenzahlungen, Geldautomatenabhebungen mit einem Mobiltelefon, Generierung eines E-Tokens sind weitere Beispiele für einen Hacker-Missbrauch.

Die zweite Art des Transaktionsbetrugs ist autorisierter Betrug, der hauptsächlich durch die Manipulation des Kunden verursacht wird. Am häufigsten wird die Bank mit folgenden Fällen konfrontiert: CEO-Betrug; Umleitungsbetrug; Rechnungsbetrug; Online-Shopping, Investitionen, Romantik, Erbschaften, Lotteriegewinne; Ransomware; Erpressung; gefälschte Wohltätigkeitssammlungen usw.

9.3 Vorgeschlagene Lösungen

Jeder Kunde ist einzigartig und erfordert einen individuellen Ansatz. Umgekehrt benötigt die Bank eine standardisierte und effiziente Methode. Bei der Tatra banka Slowakei werden Kunden-Personas verwendet, um ein besseres Verständnis des Kunden zu gewinnen. Im Einklang mit den Prinzipien des digitalen Humanismus helfen diese Personas der Bank, menschliche Werte und Sicherheit zu priorisieren und sicherzustellen, dass die Technologie den besten Interessen der Kunden dient. Weitere Personas helfen der Bank, die Kunden besser zu verstehen, für die sie Produkte und Dienstleistungen erstellt. Sie bieten Einblicke nicht nur in Daten, sondern auch in das Verhalten, die Bedürfnisse, Meinungen, Wahrnehmungen und Vorlieben der Kunden. Personas unterstützen die Bank dabei, sich individueller an die Kunden anzupassen und alles, was die Bank anbietet, genau auf ihre Bedürfnisse abzustimmen.

Zusammen mit der Einführung eines bereichsübergreifenden agilen Teams hat die Bank Maßnahmen in mehreren Bereichen eingeführt, wie Prozesse, IT-Systeme, Datenmodelle und Sensibilisierung durch systematische Weiterbildung der Mitarbeiter und Kunden.

Die Einführung von End-to-End-Prozessen in der gesamten Bank hat die Erkennung von Betrugsszenarien und die Umsetzung von Echtzeitmaßnahmen erheblich beschleunigt und so Schäden an Kunden verhindert.

Die Implementierung einer robusten IT-Lösung hilft, die Genauigkeit bei der Identifizierung betrügerischer Aktivitäten zu verbessern, zusammen mit den Datenmodellen, die verwendet werden, um Szenarien zu erstellen und das Verhalten von Opfern zu erkennen.

Und schließlich hat die Bank die erwähnte systematische und gezielte Weiterbildung für Kunden eingeführt. Die Tatra banka Slowakei investiert jedes Jahr erhebliche Mittel in die Cyber-Literacy der Kunden.

Die menschliche Natur im Cyberspace neigt dazu, von Gier, Naivität und mangelndem Bewusstsein getrieben zu werden. Bei der Vorbereitung von Bildungskampagnen für Kunden geht die Bank auf die häufigsten Hackerangriffe ein. Beispielsweise waren es im Jahr 2024 gefälschte Anzeigen in sozialen Medien, die bekannte Unternehmen oder Personen imitieren. Sie nutzen oft die Gier, Naivität und den guten Glauben der Menschen aus.

Abb. 9.2 Bildungsvideo in der mobilen App

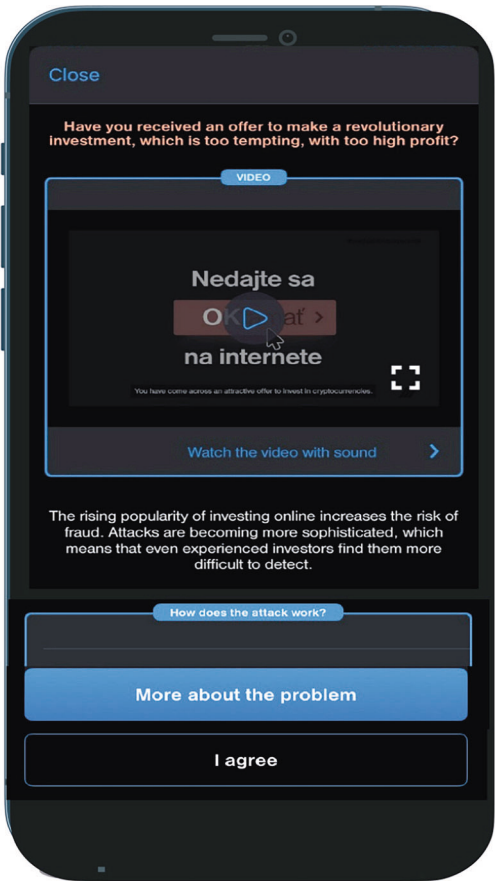
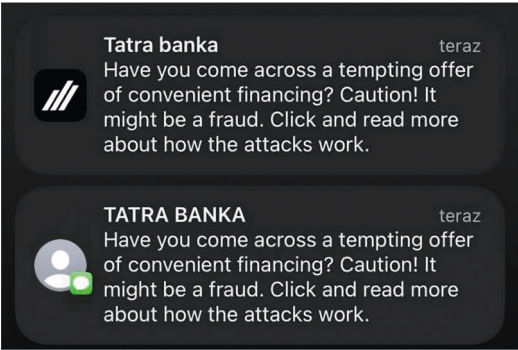


Abb. 9.3 Push-Benachrichtigungen



Im Jahr 2024 brachte die Bank jedes Quartal ein neues Bildungsvideo in die mobile App (siehe Abb. 9.2). Es wurde allen Kunden über Online-Kanäle und E-Mail gestellt. Zur Erhöhung der Sichtbarkeit verwendet die Bank auch Push-Benachrichtigungen (siehe Abb. 9.3).

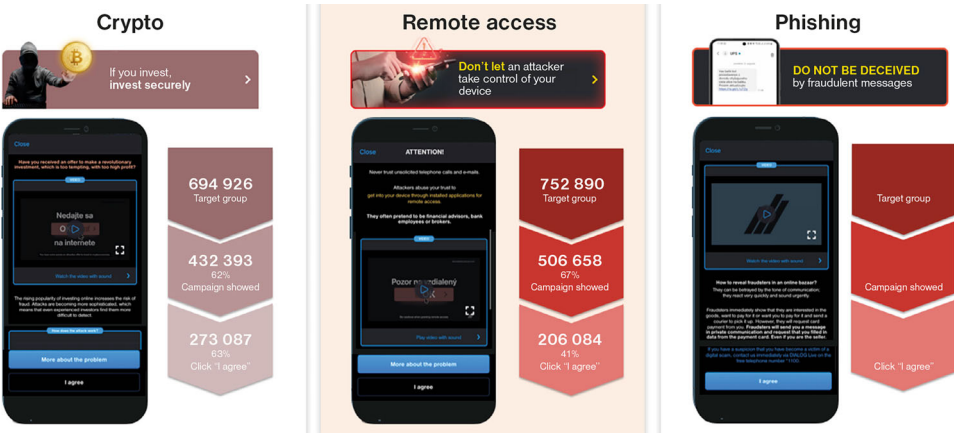


Abb. 9.4 Ereignisgesteuerte Kundenbildungskampagnen im Jahr 2024 Teil 1

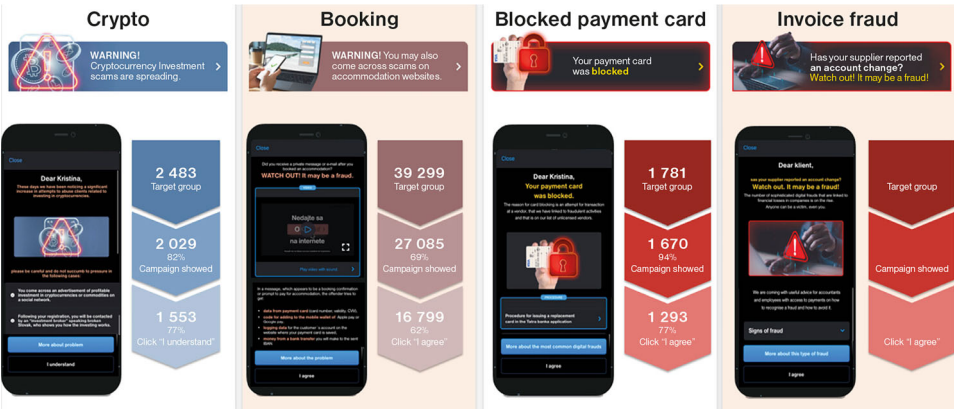


Abb. 9.5 Ereignisgesteuerte Kundenbildungskampagnen im Jahr 2024 Teil 2

Die Wahrnehmung einer solchen Kampagne liegt normalerweise über 64 %. Die Bank misst auch die Zeit, die der Kunde der Bildungskampagne widmet. Eine solche allgemeine Bildung zielte darauf ab, das Bewusstsein der Kunden für Investitionsbetrug, Online-Shopping-Betrug, Missbrauch von Fernzugriffstools und vor allem für Phishing zu schärfen (siehe Abb. 9.4).

Die Nutzung von Data-Science-Modellen ist nicht nur leistungsstark bei der Verbesserung der Präzision von Betrugserkennungsszenarien, sondern auch bei der frühzeitigen Erkennung potenziellen Opferverhaltens. Die Ergebnisse der Data-Science-Modelle werden auch für ereignisgesteuerte Kundenbildung verwendet. Die Bank erstellt personalisierte und hochgradig zielgerichtete Bildungskampagnen, die auf einen bestimmten Kundentyp abzielen, mit einer Wahrnehmungsquote von mehr als 76 % (siehe Abb. 9.5).



Abb. 9.6 Externe Kommunikation

Die Bank engagiert sich konsequent für Cybersecurity, ein zentrales Merkmal ihrer Marke. Dieses Engagement zeigt sich auch in der externen Kommunikation, die sowohl Kunden als auch die breite Öffentlichkeit anspricht. Im Jahr 2024 passte Tatra banka Slowakei bewusst ihre externe Kommunikation in diesem Bereich an und setzte verstärkt auf eine aufmerksamkeitsstarke Ansprache. Dabei kamen direkte Formulierungen zum Einsatz, die klar vor unerwünschtem Verhalten warnten, sowie ein boulevardesker Präsentationsstil mit intensiven, kontrastreichen Farben (siehe Abb. 9.6).

9.4 Fazit

Seit der Einführung der Maßnahmen kommen fast keine Phishing-Vorfälle mehr vor. Der Trend im Volumen der Kundenschäden durch Manipulation konnte umgedreht werden und nimmt weiterhin ab. Der Betrag der verhinderten Schäden ist höher als der Betrag der eingetretenen Kundenverluste (siehe Abb. 9.7).

Schäden durch unbefugte Transaktionen nehmen ständig ab und zeigen einen von der Bank gewünschten Trend in dieser Kategorie.

Das gleiche Muster des abnehmenden Schadens zeigt sich auch bei betrügerischen autorisierten Transaktionen, bei denen Kunden manipuliert wurden und die Zahlungen selbst veranlasst haben.

Die Entscheidung, ein bereichsübergreifendes agiles Team einzuführen, das sich auf die Verhinderung von Online-Betrug und gleichzeitig auf die Steigerung der Kundenzufriedenheit konzentriert, trägt bereits nach sehr kurzer Zeit Früchte und ist unerlässlich. Dieser Ansatz steht im Einklang mit den Prinzipien des digitalen Humanismus, die sich für Technologie einsetzen und menschliche Werte sowie Sicherheit priorisieren.

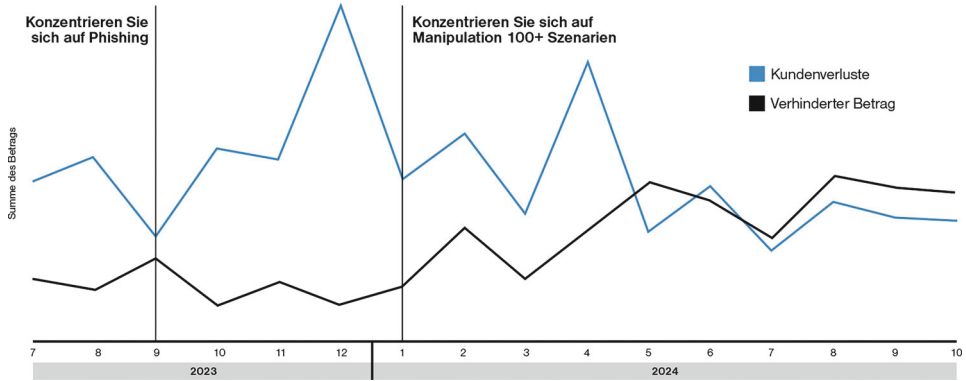


Abb. 9.7 Entwicklung von Kundenverlusten und verhinderten Beträgen

Um die Häufigkeit von Online-Betrug zu reduzieren, ist es ebenso wichtig, die Kundenbasis proaktiv zu schulen, damit die Kunden potenziellen Betrug selbst erkennen können. Mit dem Ziel, die Cyber-Literacy ihrer Kunden zu fördern, bereitet die Bank regelmäßig Bildungskampagnen vor. Durch diese Kampagnen kommuniziert die Bank spezifische Bedrohungen im Online-Umfeld. Die Investition in Bildungskampagnen ist erheblich und ihre Auswirkungen werden gründlich gemessen.

Natalia Major ist Chief Operating Officer (COO) der Tatra banka Slowakei, mit Schwerpunkt auf operative Exzellenz, digitale Transformation und Innovation im Bankensektor. Zuvor leitete sie die Abteilungen Transaktionsabwicklung und elektronische Vertriebskanäle und war maßgeblich an der Einführung des ersten slowakischen Internet-Banking-Dienstes sowie des ersten 24-Stunden-Kontaktzentrums „Dialog“ beteiligt. Ihre Karriere bei der Tatra banka begann 1994. Akademisch absolvierte sie ein Masterstudium an der Comenius-Universität Bratislava und vertiefte ihre Expertise durch ein Postgraduiertenprogramm an der University of Bristol sowie ein Executive Management Programm an der Harvard Business School.

Digitaler Wandel im ORF: Leichter Leben durch Technologie

10

Wie der ORF in seinem Angebot und seinen internen Prozessen Werte des digitalen Humanismus unterstützt

Astrid Zöchling

Zusammenfassung

Der ORF verfolgt in seinem Wirkungskreis Ziele und Prinzipien, die auch dem Konzept des digitalen Humanismus innewohnen. Technologie unterstützt die Arbeit der Mitarbeiter*innen, ohne den Menschen aus zentralen Aufgaben – insbesondere dem Journalismus – zu verdrängen oder dessen Handlungsfähigkeit einzuschränken. Der vorliegende Artikel beleuchtet, wie dieser Ansatz in der Praxis gelebt wird, welche Grenzen dabei bestehen und wie der Bildungsauftrag des ORF durch die Förderung der AI-Literacy, sowie der Medien- und Digitalkompetenz der Bevölkerung erfüllt wird. Gleichzeitig wird das Leben der Mitarbeiter*innen durch neue Technologien an vielen Stellen einfacher und effizienter gestaltet. Ein Fokus des Beitrags liegt auf den Vorteilen, die die Digitalisierung im Zusammenhang mit dem Ausbau der Barrierefreiheit bietet, etwa durch die automatische Erstellung von Untertiteln oder die Bereitstellung von Beiträgen in Einfacher Sprache. Durch konkrete Anwendungsbeispiele wird der praktische Nutzen von KI im Unternehmen veranschaulicht. Abschließend werden aktuelle Initiativen und Projekte vorgestellt, mit denen der ORF seine Mitarbeiter*innen in den digitalen Transformationsprozess einbindet. Der ORF ist davon überzeugt, dass Digitalisierung als Mammutaufgabe unserer Gesellschaft und Zeit nur gemeinsam bewältigt werden kann, und dass nur jene Technik, die ein echtes Problem für die Menschen löst, Akzeptanz und Nutzung finden wird.

Hinweis – KI-Verwendung: Teile dieses Beitrages wurden durch KI (Chat GPT 4.0) überarbeitet und gekürzt.

A. Zöchling (✉)

ORF, Wien, Österreich

E-Mail: astrid.zoechling@orf.at

Digitalisierung, als Mammutprojekt unserer Zeit, kann ausschließlich gemeinsam bewältigt werden. Nur Software oder KI, die ein echtes Problem löst und Menschen dabei unterstützt, Aufgaben besser zu bewältigen, wird Akzeptanz finden und Nutzen stiften können.

10.1 Wie betrifft digitaler Humanismus den ORF?

Im Verständnis des Österreichischen Rundfunks (ORF) bedeutet digitaler Humanismus, den technologischen Fortschritt gezielt in den Dienst von Publikum und Belegschaft zu stellen. Im Fokus steht dabei die Vereinfachung der zusehends komplexer und schneller werdenden Lebenswelt und des Arbeitsalltags der Menschen: Technologie soll unterstützen und entlasten, nicht ersetzen oder kontrollieren. Digitalisierung und Künstliche Intelligenz (KI) werden folglich nur dort eingesetzt, wo sie tatsächlich einen Mehrwert schaffen, sei es durch die Verbesserung der Zugänglichkeit des Programms für das Publikum oder durch die Bereitstellung praktischer Tools, die Routinearbeiten beschleunigen oder administrative Aufgaben übernehmen. Dadurch wird mehr Raum für kreative und wertschöpfende Tätigkeiten geschaffen. Im Sinne des digitalen Humanismus begreift der ORF Technologie folglich nicht als Selbstzweck, sondern als Instrument, das den Menschen und seine Bedürfnisse in den Mittelpunkt stellt. Im Rahmen seines öffentlich-rechtlichen Bildungsauftrags fördert der ORF außerdem ein kritisches Bewusstsein für neue technologische Möglichkeiten in seinen Programmen. Weiters geht es darum, Berührungsängste mit der Technologie, die oft aus fehlender Information resultieren, abzubauen.

10.1.1 Eine Frage der Werte

Der ORF hat als öffentlich-rechtliches Medienunternehmen einen gesetzlich genau definierten Auftrag im Dienste der Allgemeinheit. Vor diesem Hintergrund verweisen auch die zentralen unternehmerischen Ziele des ORF, wie er sie etwa in seiner aktuellen Unternehmensstrategie „ORF 2030“ (2024) beschreibt, auf ein von humanistischen Werten geprägtes Selbstbild und einen entsprechenden Gestaltungsauftrag. Auch digitaler Humanismus kann als Wertegerüst verstanden werden, aus dem sich normative Handlungsanweisungen ableiten lassen. Dabei überschneiden sich die journalistischen und unternehmensinternen Prinzipien des ORF mit den Werten des digitalen Humanismus.

Exemplarisch soll hier auf einige der wichtigsten Unternehmensziele eingegangen werden:

- **Relevanz der Themenauswahl und Meinungspluralität in der Berichterstattung:** Der ORF versucht, die vielfältigen Interessen und Meinungen der Bevölkerung umfassend zu berücksichtigen. Auch Digitalisierungsthemen werden daher aus diversen Blickwinkeln beleuchtet, um die Inklusion bildungsferner Milieus oder älterer Mitmenschen voranzutreiben, womit ein Beitrag zu einem breiten Diskurs über technologische Veränderungen geleistet wird. Komplexe Inhalte (etwa über technische Anwendungsfälle neuer Technologien wie Robotik oder KI) werden dabei allgemein verständlich aufbereitet.
- **Akzeptanzsteigerung:** Ziel des ORF ist die nachhaltige Stärkung seiner Publikumsbeziehungen durch Dialog und neue Partizipationsmöglichkeiten, um möglichst viele Menschen mit öffentlich-rechtlichen Qualitätsinhalten zu erreichen. Je größer die Akzeptanz eines Mediums in der Bevölkerung ist, desto besser können Mythen aufgeklärt und Ängste genommen werden.
- **Innovationsführerschaft:** Aus Sicht des ORF braucht es mutige Offenheit gegenüber technologischen Innovationen, um den Publikumserwartungen in einer zunehmend digitalisierten und individualisierten Medienwelt gerecht zu werden. Wenn Innovation darauf abzielt, Technologien zur Erleichterung menschlicher Arbeit und zur Förderung qualitativ hochwertiger Inhalte zu nutzen, stärkt dies auch die Ziele des digitalen Humanismus. Gerade beim Einsatz von Automatisierung und Künstlicher Intelligenz liegt die ethische Messlatte des ORF besonders hoch, wie in den folgenden Kapiteln ausgeführt wird.

Die Technische Direktion und im speziellen die IT des ORF sieht Digitalisierung als Gemeinschaftsprojekt, das nur erfolgreich ist, wenn es echte Probleme löst und dem Menschen nützt. Dieser Betrachtung liegt eine grundsätzlich positive Auffassung des technologischen Fortschrittes und der Digitalisierung im Allgemeinen zugrunde. Dessen ungeachtet muss aber immer zusätzlich hinterfragt werden, wo potenzielle Risiken und Gefahren einer bestimmten Technologie liegen. In einer aktuellen Public Value Studie „Öffentlich-rechtliche Qualität im Diskurs. Digitale Innovation öffentlich-rechtlicher Medien“ beschäftigt sich der ORF daher intensiv mit der Frage, welche negativen Auswirkungen KI auf ethische Fragestellungen und unsere Gesellschaft hat.

Darin vertreten die beiden Wissenschaftler Bernd Carsten Stahl und Damian Eke die These, dass viele Herausforderungen im Zusammenhang mit neuen Angeboten aus dem Bereich der generativen KI direkt mit dem öffentlichen Auftrag des ORF verbunden sind und daher im Bewusstsein der Nutzer*innen und des Managements verankert sein sollten (vgl. Public Value Studie, 2024, S. 69 f.). Gerade hinsichtlich eines möglichen Bias, der mit dem heutigen Stand des Wissens über die technologischen Möglichkeiten kaum detektiert und auch nicht wirksam bekämpft werden kann, werden nicht nur der ORF, sondern viele Unternehmen, die sich mit Werten des digitalen Humanismus identifizieren, ihre Anwendungsgebiete künftig besonders kritisch prüfen.



Abb. 10.1 Ethische Herausforderungen durch GenKI

Für den ORF als öffentlich-rechtliches Medienunternehmen ist die Frage der ethischen Herangehensweise an journalistische Arbeit zentral – vergleiche dazu Abb. 10.1. Im Ethikkodex des ORF wird daher unter anderem folgendes festgelegt:

1. **Genauigkeit und Wahrhaftigkeit:** Journalist*innen tragen die Verantwortung, der Öffentlichkeit genaue und zuverlässige Informationen zur Verfügung zu stellen. Die Überprüfung von Fakten vor der Berichterstattung ist daher essenziell.
2. **Fairness und Objektivität:** Journalist*innen stellen Informationen ausgewogen und fair für die Öffentlichkeit bereit und ermöglichen so dem Publikum, sich eine eigene Meinung zu bilden.
3. **Unabhängigkeit und Freiheit:** Die Pressefreiheit ist ein essenzieller Grundpfeiler der Demokratie.
4. **Transparenz und Rechenschaftspflicht** sind entscheidend, um das Vertrauen in die Medienberichterstattung sicherzustellen (vgl. ORF Ethikkodex, 2024).

Insgesamt tragen diese umfassenden ethischen Leitlinien dazu bei, das hohe Ansehen und die Glaubwürdigkeit des ORF zu erhalten und zu stärken.

10.1.2 Mensch und Maschine: Technologie, die uns unterstützt, nicht ersetzt

Der ORF hat sich intensiv mit aktuellen Digitalisierungsthemen beschäftigt, insbesondere mit der Einführung von KI in den journalistischen Arbeitsprozess. Hierzu wurden redaktionelle Richtlinien zur Verwendung von KI in den Medien des ORF in einem interdisziplinären Team erarbeitet. Darin wird betont, dass der Einsatz von KI stets den Werten eines öffentlich-rechtlichen Rundfunks entsprechen soll. Der Fokus liegt auf den Kompetenzen und Talenten, sowie auf der Kreativität und der Vielfalt der Journalist*innen. Technische Möglichkeiten, rechtliche Rahmenbedingungen und ethische Aspekte werden abgewogen, wobei nicht alles, was in diesen Bereichen realisierbar ist, auch journalistisch sinnvoll erscheinen muss. Entscheidend ist, welchen Mehrwert die Technologie für die Mitarbeiterinnen und Mitarbeiter, das Publikum und die Erfüllung des öffentlich-rechtlichen Auftrags bringt (vgl. ORF KI-Guidelines, Februar 2025).

Besonderes Augenmerk wird auf die Wahrung der Authentizität der Berichterstattung gelegt: Transparenz in Bezug auf die Verwendung von KI wird großgeschrieben, wobei mit Sorgfalt und Augenmaß vorgegangen wird. Der ORF nutzt KI beispielsweise zur Unterstützung gängiger Arbeitsschritte bei der Bearbeitung von Texten, Bildern, Audios und Videos. KI-Tools, die redaktionsintern zur Transkription und Übersetzung verwendet werden, beeinflussen die Authentizität der Inhalte jedoch nicht und können daher ohne spezielle Hinweise eingesetzt werden. Rechtliche Bestimmungen rundum Datenschutz, Informationssicherheit und Urheberrecht bleiben ungeachtet der neuen technologischen Möglichkeiten weiter bestehen (vgl. ORF KI-Guidelines, Februar 2025).

Dabei treten auch klare Limits für den Einsatz von Digitalisierung und KI zutage: Wo die Grenze der Unterstützungsleistung für den Menschen überschritten wird und die inhaltliche bzw. gestalterische Hoheit der Technologie beginnen würde, ist die Beschränkung der technologischen Möglichkeiten aus ORF-Sicht geboten.

10.1.3 Einblick in die betriebliche Praxis: Journalismus und Bildungsauftrag des ORF als Beitrag zu einer funktionierenden Demokratie

Der ORF übernimmt innerhalb seines breit gefächerten Programms Verantwortung für die digitale Bildung der Bevölkerung und trägt so zur Einbindung unterschiedlichster Bevölkerungsschichten in die digitale Transformation bei. Mit hunderten Beiträgen in Radio, Fernsehen und Online wird über sämtliche Kanäle des ORF ein breites Publikum erreicht, das über neue technologische Entwicklungen und deren gesellschaftliche Auswirkungen informiert wird. Populäre TV-Formate wie „Bundesland heute“ mit Berichten über regionale Digitalisierungsinitiativen erzielen regelmäßig hohe Reichweiten von weit über eine Million Zuschauer*innen.

Besonders bemerkenswert ist die Akzeptanz dieser Inhalte: 40 % der befragten Österreicher*innen wünschen sich laut einer Publikumsstudie des ORF aus dem Jahr 2017 eine verstärkte Berichterstattung über technologische Themen, während 47 % den Umfang als angemessen beurteilen. Dieses Engagement ermöglicht es nicht nur technologisch affinen Menschen, sondern auch Bevölkerungsgruppen mit geringer digitaler Kompetenz, sich aktiv mit dem digitalen Wandel auseinanderzusetzen und dessen Chancen zu verstehen. Damit leistet der ORF einen wesentlichen Beitrag zur Förderung der digitalen Inklusion und stärkt die demokratische Teilhabe, indem er sicherstellt, dass niemand von dieser Transformation ausgeschlossen bleibt.

10.1.4 Inklusion und Barrierefreiheit in Zeiten der Digitalisierung

Digitale Technologien können bei richtiger Anwendung Hürden und Barrieren abbauen. Im Programmangebot des ORF sorgen Untertitel (neben weiteren Maßnahmen) dafür, dass möglichst viele Menschen gleichberechtigt am Gemeinschaftserlebnis Fernsehen bzw. Streaming teilhaben können.

Untertitel sind eine wesentliche Unterstützung für Menschen mit Hörbeeinträchtigungen, indem sie gesprochene Inhalte visuell erlebbar machen. In Österreich sind laut Gehörlosenverbänden etwa 450.000 Menschen von Kommunikationsbeeinträchtigungen aufgrund einer Hörbehinderung betroffen, darunter 8000 bis 10.000 Gehörlose. Untertitel ermöglichen es diesen Personen, Inhalte zu verstehen, die sie sonst nicht wahrnehmen könnten. Doch der Nutzen von Untertiteln reicht weit über diese Zielgruppe hinaus: Sie fördern den Spracherwerb, indem sie das Sprachverständnis unterstützen, und erleichtern den Zugang zu Inhalten in öffentlichen Räumen, beispielsweise in Situationen, in denen die vorhandene Tonspur nicht genutzt werden kann. Während in Ländern wie Großbritannien und den USA Untertitel auf Bildschirmen in öffentlichen Bereichen – etwa in Pubs oder auf Flughäfen – längst Standard sind, hat sich diese Praxis in Österreich bisher kaum etabliert. Dadurch bleibt ein bedeutendes Potenzial zur Förderung von Barrierefreiheit und Inklusion noch ungenutzt.

Der ORF geht hier mit gutem Beispiel voran und bietet weitreichende Untertitelungen für Fernsehsendungen sowie Programme und Online-Angebote in einfacher Sprache, die insbesondere Menschen mit Lernschwierigkeiten oder geringer Bildung adressieren. Seit 1980 wurden Untertitelungsleistungen im ORF von Menschen erbracht, was stets eine natürliche Kapazitätsgrenze bedeutet hat. Ziel des ORF ist es mittlerweile, 100 % seiner audiovisuellen Inhalte bis 2030 zu untertiteln. Hier leistet Künstliche Intelligenz, die verantwortungsvoll von der Online Technik des ORF entwickelt wurde, einen Beitrag, der

im Sinne des digitalen Humanismus als Bereicherung gelten kann. Der ORF zielt darauf ab, wesentlich mehr Inhalte als bisher für das Publikum zu untertiteln, wobei die KI gute Ergebnisse bei einfacher Skalierbarkeit bietet.

Seit vergangenem Jahr testet der ORF die automatische Transkription von „Bundesland heute“-Sendungen mit KI-Unterstützung. Diese Funktion ist derzeit auf ORF ON verfügbar und arbeitet nach dem Human-in-the-Loop-Prinzip, bei dem die Letztverantwortung immer bei einem Redaktionsmitglied liegt. Ein Vorteil dieser Lösung ist die Möglichkeit, die KI mit Redaktionstexten zu trainieren, um typische Fehler – wie etwa bei der Aussprache von österreichischen Ortsnamen – zu reduzieren. Dies senkt die Fehleranfälligkeit und zeigt die Umsetzung von digitalem Humanismus in der Praxis.

10.1.5 Sprachbarrieren beseitigen

Seit Ende Mai 2020 gibt es auf ORF.at die Möglichkeit, Informationen und Beiträge in Einfacher Sprache zu konsumieren. Diese Angebote richten sich an Personen mit geringer Lesekompetenz, kognitiven Beeinträchtigungen, Lernschwierigkeiten oder Menschen, die die Landessprache erst erlernen. Sie reduzieren sprachliche Barrieren, indem komplexe Inhalte vereinfacht und leicht verständlich aufbereitet werden. Dadurch wird sichergestellt, dass alle Gesellschaftsschichten – unabhängig von Bildungsgrad oder Sprachkenntnissen – Zugang zu relevanten Nachrichten und gesellschaftlichen Diskussionen erhalten. Gerade in demokratischen Gesellschaften ist es essenziell, dass alle Bürger*innen gut informiert sind und am öffentlichen Diskurs teilnehmen. Nachrichten in einfacher Sprache verbessern daher die Chancengleichheit und gesellschaftliche Integration, indem sie sprachliche Barrieren abbauen und alle Menschen einbeziehen. Relativierend zu den oben genannten Aspekten sei jedoch erwähnt, dass Digitalisierung auch das Risiko birgt, bestehende soziale und ökonomische Ungleichheiten zu verstärken. Bevölkerungsgruppen, die nicht über die nötigen Kompetenzen verfügen oder keinen Zugang zu digitaler Infrastruktur haben, könnten von wesentlichen Informationen und damit von der gesellschaftlichen Teilhabe ausgeschlossen bleiben. Dies verdeutlicht die Notwendigkeit, digitale Angebote nicht nur technisch verfügbar, sondern stets auch inklusiv und niedrigschwellig zu gestalten, um sicherzustellen, dass alle Menschen – unabhängig von Alter, Einkommen oder Bildungsgrad – an der digitalen Transformation teilhaben können. Nur so kann die Digitalisierung ihr Potenzial als Werkzeug der Inklusion vollständig entfalten.

10.2 Praktische Anwendungsbeispiele von digitalem Humanismus und KI im ORF für seine Mitarbeiter*innen

10.2.1 AiDitor: Journalistische Arbeit im digitalen Wandel

Der AiDitor, ein vom Innovationsteam des ORF eigens entwickelter KI-Assistent des ORF, eröffnet Journalist*innen mittels Künstlicher Intelligenz zahlreiche Möglichkeiten, ihre Arbeit effizienter zu gestalten und repetitive Tätigkeiten zu automatisieren. Eine seiner Hauptfunktionen ist die Verbesserung und Anpassung von Audio- und Videoinhalten, bei der etwa Hintergrundgeräusche in Aufnahmen mit einem einfachen Klick entfernt werden können. Besonders häufig wird diese Funktion genutzt, um Störgeräusche zu eliminieren und so die Qualität der Ausstrahlungen zu optimieren. Seit der Einführung des Features im Jahr 2024 hat der AiDitor bereits über 10.000 solcher Audio-Verbesserungen vorgenommen, was die Bedeutung dieser Funktion im journalistischen Alltag unterstreicht.

Darüber hinaus bietet der AiDitor eine Kürzungs- und -bearbeitungsfunktion von Texten. Die intuitive „One-Click-Operation“ erlaubt es geschulten Redakteurinnen und Redakteuren, diese Anpassungen mit minimalem Aufwand vorzunehmen und ihre Aufmerksamkeit stattdessen verstärkt auf die inhaltliche Tiefe und Kreativität ihrer Arbeit richten zu können.

Ein weiterer zentraler Anwendungsbereich ist die Transkription von Audio- und Videoinhalten sowie deren Übersetzung und Untertitelung. Diese Funktionen sind vor allem im redaktionellen Alltag wertvoll, da sie komplexe und zeitintensive Arbeiten übernehmen und es den programmgestaltenden Abteilungen so ermöglichen, schneller auf Interviewmaterial zuzugreifen und es zielgerichtet einzusetzen. Die Kontrolle über den Inhalt und seine Qualität bleibt dabei stets in den Händen der Anwender*innen: Dem „Human in the Loop“-Konzept entsprechend, steht der Mensch stets am Anfang und Ende eines jeden Gestaltungsprozesses, für den er oder sie anschließend auch die volle Verantwortung trägt. Dieter Bornemann, Vorsitzender des ORF Redaktionsrates, fasst im Dialog Forum zusammen: „ChatGPT hat im Oktober 2022 eingeschlagen – schnell war klar, das ist etwas ganz Großes. Die KI als Wundertüte für alle Kreativen. Und für die Faulen. Und für alle dazwischen. Er betont jedoch, dass die KI mit Wahrscheinlichkeiten arbeitet und man immer im Bewusstsein haben muss, das Trainingsdaten unzureichend oder falsch sein können und deswegen Verzerrungen oder Halluzinationen entstehen können. Wird dieses zu 80 % richtige Ergebnis mit einem erfundenen Rest vermischt, entsteht Desinformation“, (vgl. ORF Public Value, [2023/24](#)).

Mit dem AiDitor demonstriert der ORF die verantwortungsvolle Nutzung von KI im Journalismus, indem er menschliche Arbeitskräfte unterstützt, nicht ersetzt. Das fördert professionellen Journalismus und bietet Redakteur*innen mehr Freiräume für qualitativ hochwertige Arbeit, ohne journalistische Standards zu gefährden.

10.2.2 AiKM: KI-gestütztes Musik-Fingerprinting

Der KI-gestützte Musik-Fingerprint-Prozess unter dem Titel „AiKM“ im ORF zeigt, wie digitale Technologien genutzt werden können, um die Arbeitsbelastung zu reduzieren und die Effizienz zu steigern. Ziel der Initiative ist es, die manuelle Erstellung der AKM-Meldungen für im medialen Gestaltungsprozess verwendete Musik zu automatisieren und gleichzeitig die Qualität der Musikstückerkennung zu verbessern. Die rechtlichen Rahmenbedingungen hierfür waren günstig, da die verwendeten Musikdaten nicht personenbezogen und ohnehin öffentlich verfügbar sind, und zudem keine generative KI zum Einsatz kommt, sondern klassische Machine-Learning-Algorithmen. Das System basiert auf erprobter open-source KI-Technologie, userzentrierte Entwicklung und ein technisches fundiertes Testing des Modells stellen die Ergebnisqualität sicher.

Die Einbindung der Belegschaft erfolgte schrittweise: Erst gab es gemeinsame Customer-Journey-Mapping-Workshops, wo eine gemeinsame Vision eines vereinfachten Prozesses erarbeitet wurde. Anschließend wurde das Modell im Hintergrund technisch getestet und auf Treffgenauigkeit analysiert, bevor es den Nutzerinnen und Nutzern zur Verfügung gestellt wurde. So konnten grundlegende Fehler vermieden und zugleich Vertrauen in die neue Technik aufgebaut werden. Weitere Automatisierungs- und Optimierungsschritte erfolgen iterativ, was dem Projektteam die Möglichkeit gibt, die Entwicklung aktiv und laufend mitzugestalten. Bislang bewährt sich das Tool durch die zuverlässige Erkennung von Musikstücken, unabhängig von Gesang oder Störgeräuschen. Der manuelle Aufwand in den Redaktionen konnte damit um rund 90 % reduziert werden, was das Projekt zu einem Paradebeispiel für die Werte des digitalen Humanismus macht: Technologie unterstützt die Menschen und nicht umgekehrt.

Das Projekt traf von Anfang an auf Unterstützung in allen involvierten Einheiten des Hauses. Auch hier scheint sich die These zu bestätigen, dass eine Software, die in diesem Fall eine ungeliebte Arbeit übernimmt, die nicht zu den Kernkompetenzen einer Redakteurin oder eines Redakteurs zählt, auf Akzeptanz stößt. Durch den Einsatz der Technologie konnte nicht nur die Qualität der AKM-Meldungen verbessert, sondern auch die Arbeitsaufwände für nicht-redaktionelle, bürokratische Tätigkeiten deutlich gesenkt werden. Bei Bedarf stehen den Redakteur*innen ergänzende Materialien wie Checklisten oder Guidelines zur Verfügung. Es ist ein gelungenes Beispiel dafür, wie der ORF digitale Innovationen gezielt einsetzt, um die Zusammenarbeit zwischen Menschen und Technik zu optimieren und einen kleinen Teil der Mammutaufgabe Digitalisierung gemeinsam zu bewältigen.

10.2.3 Von Widerstand zu Akzeptanz: Digitalisierung durch gemeinschaftliches Lernen gestalten

Datenzentrierte Entscheidungsprozesse setzen Menschen voraus, die ein gutes Verständnis für Zahlen haben und mit den notwendigen Tools und Techniken vertraut sind, um relevante Daten im Unternehmen zu generieren und aufzubereiten. Da das Wissen über modernes Datenmanagement und Business-Intelligence-Anwendungen im ORF noch nicht lückenlos vorhanden war, wurde vor einigen Jahren eine Daten- und BI-Community ins Leben gerufen. Mitarbeiterinnen und Mitarbeiter wurden dabei aktiv eingebunden: Die Vereinbarung etwa, dass jede Sitzung mit der Präsentation eines Dashboards aus unterschiedlichen Einheiten des Hauses zu verbinden ist, schafft praxisnahe Anreize und fördert den Lernprozess durch das gezielte Involvement der Teilnehmenden.

Die Community ist bewusst offen und inklusiv gestaltet, um allen interessierten Personen im ORF Zugang zu bieten. Monatliche Treffen dienen als Plattform für den Wissensaustausch und die Vorstellung neuer Features, zudem wird ein Teams-Channel als begleitende Austauschmöglichkeit angeboten. Der kooperative Austausch wird seither als wertvolle Unterstützung im Arbeitsalltag wahrgenommen, wodurch konkrete Probleme schneller gelöst und neue Funktionen effizienter genutzt werden können. Es entstehen dadurch neue Arbeitsbeziehungen zwischen Expert*innen.

Das Beispiel zeigt, wie Digitalisierung menschenfreundlich vermittelt werden kann. Durch die Schaffung eines unterstützenden Raums für Wissensaustausch und Zusammenarbeit wird der Lernprozess erleichtert. Neue Kolleginnen und Kollegen müssen komplexe Programme nicht mehr mühsam auf eigene Faust erlernen, sondern erhalten vom ersten Tag an die notwendige Unterstützung der Community. Diese trägt auch dazu bei, Widerstände gegen neue Technologien zu minimieren: Der direkte Nutzen, der aus der Teilnahme resultiert, stärkt die Bereitschaft zur weiteren Beteiligung und vermindert Unsicherheiten. Dies entspricht den Prinzipien des digitalen Humanismus, bei dem Technologie an den menschlichen Bedürfnissen ausgerichtet wird.

10.3 Fazit

Der ORF verfolgt einen ganzheitlichen Ansatz, um digitale Technologien gezielt und verantwortungsvoll im Unternehmen einzusetzen. Im Mittelpunkt steht dabei einerseits die verbesserte Zugänglichkeit des öffentlich-rechtlichen Programms für das Publikum, sowie andererseits die Unterstützung der Mitarbeiterinnen und Mitarbeiter in ihrem Arbeitsall-

tag. Durch Maßnahmen wie die automatisierte Untertitelung, Nachrichten in Einfacher Sprache und KI-gestützte Tools gelingt es dem ORF, technische Innovationen praxisnah und nutzerorientiert einzubinden.

Besonderen Wert legt der ORF dabei auf die Einhaltung hoher ethischer Standards und auf einen vertrauensvollen Umgang mit neuen Technologien. Zahlreiche Initiativen zur Förderung digitaler Kompetenzen innerhalb der Belegschaft sowie eine umfassende Kommunikation zu aktuellen Entwicklungen im Bereich der Künstlichen Intelligenz und den Nutzungsmöglichkeiten neuer KI-Tools verdeutlichen das Bestreben, Digitalisierung als gemeinschaftliches Projekt voranzutreiben.

Die praktischen Beispiele aus dem Arbeitsalltag zeigen, dass technologischer Fortschritt nicht nur Effizienzgewinne im Unternehmen ermöglicht, sondern auch den Handlungsspielraum jedes Einzelnen erweitern kann. Der ORF liefert damit einen Nachweis, wie Digitalisierung mit einem klaren Fokus auf den Menschen gestaltet werden kann, ohne die Chancen neuer Technologien aus den Augen zu verlieren.

Literatur

- ORF Ethikkodex (2024). <https://der.orf.at/unternehmen/leitbild-werte/ethikkodex/ethikkodex102.pdf>. Zugegriffen: 10. Jan. 2025.
- ORF KI-Guidelines, Stand Februar 2025
- ORF Public Value (2024). *Public Value Bericht 2023/24 – 46 – Wie viel KI ist drin im ORF? – Stefan Kollinger; ORF Technik | Dieter Bornemann, ORF-Redaktionsrat*. <https://zukunft.orf.at/modules/orfpv2023/build/documents/WEB-ORF-PV-Bericht-23-24.pdf>. Zugegriffen: 05. Jan. 2026.
- ORF Publikumsratsstudie (2017). *INTEGRAL, Digitaler Wandel*
- Public Value Studie (2024). *Öffentlich-rechtliche Qualität im Diskurs. Fast Forward. Digitale Innovation öffentlich-rechtlicher Medien in Europa*
- Strategie ORF 2030, Stand Dezember 2024

Astrid Zöchling ist CIO des ORF und leitet dort seit 2023 den Bereich IT und Netzwerk. Davor baute Zöchling über zehn Jahre in unterschiedlichen beruflichen Stationen als Head of Analytics und Head of Customer Experience Plattform IT- und Datenteams der BAWAG Group und ÖVAG auf. Die IT-Managerin ist gebürtige Oberösterreicherin, hat in Wien Politikwissenschaften, Wirtschaft und Recht studiert und bietet Trainings im Bereich digitale Transformation und Leadership an. Sie brennt für Digitalisierung und Möglichkeiten, durch Technologie das Leben der Menschen einfacher zu machen. 2025 wurde sie von Confare mit dem „CIO of the year“ Award ausgezeichnet.

Nutzerzentrierte Innovationsstrategien aus der Perspektive der APA

11

Von User Needs bis Trusted AI

Verena Krawarik

Zusammenfassung

Nachrichtenagenturen wie die APA – Austria Presse Agentur stehen selten im Mittelpunkt der Öffentlichkeit. Und doch stellen sie die journalistische Grundversorgung für die heimischen Medienschaffenden sicher und sorgen dafür, dass Entscheidungsträger*innen jederzeit am Puls der Zeit sind. Was die APA als relevant erachtet, findet sich alsbald auf vielen Ausspielkanälen wieder oder wird im Rahmen von Softwareprojekten in die eigene Organisation integriert. So ist es nicht verwunderlich, dass das Interesse an der Art und Weise gestiegen ist, wie die APA Nachrichten und daraus folgende Produkte produziert, welchen Leitlinien die Agentur in Zeiten von Künstlicher Intelligenz (KI) folgt, und wo die Konzepte für einen nutzenzentrierten Journalismus herkommen. Der vorliegende Beitrag zeigt anhand von zwei Projekten wie in der österreichischen Nachrichtenagentur die Bedürfnisse von Kund*innen, Leser*innen und Mitarbeiter*innen gezielt ins Zentrum der Entwicklungen geholt werden.

Wenn sich Medienunternehmen Daten orientiert mit ihrem Angebot beschäftigen, müssen sie nach Strategien suchen, die das Empowerment von Nutzer*innen in das Zentrum ihres Handels stellen und diese auch messbar machen.

V. Krawarik (✉)
APA, Wien, Österreich
E-Mail: verena.krawarik@apa.at

11.1 Einleitung – Medienwandel und Digitaler Humanismus

„Geglückte Zukunft verlangt halt eben auch: ‚digitalen Humanismus‘“. Mit diesen Worten beendete der Journalist Jakob Steuerer zu Beginn des Jahres 1998 seinen Artikel zum Thema „Digital Citizens“ in der Tageszeitung „Die Presse“ (Steuerer, 1998). Es ist der zeitlich erste Eintrag im großen Medienarchiv der APA – Austria Presse Agentur, der diesen Begriff explizit benutzt. Seither hat sich das dahinter liegende Konzept substanziell weiterentwickelt und auch im Medienbereich zahlreiche Fürsprecher gefunden. Die Denkschule harmoniert gut mit den Werten von Qualitätsmedien, und mit der dem Konzept inhärenten Forderung nach der Vermittlung von Medienkompetenz als einem wesentlichen Schlüssel zur verantwortungsvollen Nutzung digitaler Technologien, ist auch ein strategischer Match gelungen (vgl. Nida-Rümelin & Weidenfeld, 2018).

Wie aber sieht nun die Übersetzung des Konzeptes bei der Nachrichtenagentur APA in der Praxis aus und welche Rolle spielt das im APA-medialab angesiedelte Innovationsmanagement dabei? Am besten lässt sich dies anhand von zwei größeren Projekten der letzten drei Jahre erklären – allen voran die Entstehung der Trusted AI-Strategie und ihrer organisatorischen Begleitmaßnahmen sowie der Umsetzung des Frameworks der „User Needs“, die bereits den Ausblick auf eine maßgeschneiderte Informationsbeschaffung liefert.

In beiden Fällen spielte eine Forschungsgeleitete Herangehensweise im Zuge des Innovationsprozesses eine entscheidende Rolle. Mit Blick auf das Thema Künstliche Intelligenz wurde der Aufgabenbereich des Innovationsmanagement nach und nach auch um Aspekte einer Corporate Digital Responsibility (CDR) ergänzt, die sich im Spannungsfeld zwischen notwendiger Agilität, profunder Kenntnis kommender Normen und Regularien und ethischen Überlegungen behaupten muss.

11.2 Leitplanken für das Zeitalter der Künstlichen Intelligenz

Künstliche Intelligenz begleitet die APA seit vielen Jahren. In der F&E-Abteilung der APA-IT, einer Tochtergesellschaft der Nachrichtenagentur, sind Entwicklungen rund um die Themen Intelligente Suche, Medienmonitoring und automatische Codierung ein fester Bestandteil des Tagesgeschäftes. In der Geschäftseinheit Digitalisierung, den IT-Units weiterer APA-Töchter, dem Faktencheck-Team sowie im Innovationshub des Hauses, dem APA-medialab, stehen vertiefende Experimente, Studien oder Implementierungen an.

Im Medialab etwa wurde ab 2018 der Grundstein für eine automatisierte Berichterstattung gelegt. Das Ziel der ersten Prototypen, die noch lange vor dem Release von ChatGPT im Rahmen von Design Sprints entstanden, war Einsatzzwecke für die Nutzung von Natural Language Generation-Technologie rund um Daten intensive Themen zu erproben oder neuartige Algorithmen zur Erstellung von Texten zu entwickeln und bei Erfolg in die Produktlandschaft zu integrieren. Positioniert am Front End of Innovation durfte hier auch nicht die Begleitforschung in Form von Literaturstudien zu Red Lines und Best Practices

anderer Marktteilnehmer fehlen. Damit konnten ethische Fragestellungen zum Einsatz der digitalen Technologien schon in den Entwicklungsprozess hineingeholt werden.

Im Jahr 2019 launchte die APA-Redaktion im Zuge der Wahlberichterstattung ihre ersten maschinengenerierten Texte für über 2000 Gemeinden in Österreich – eine Innovation, die bis dahin aufgrund des Ressourceneinsatzes nicht möglich gewesen war (Schell, 2021a). Die Texte waren als „automatisiert erstellt“ gekennzeichnet, da es vor der Publikation nicht mehr möglich war, noch ein prüfendes Auge auf die Kurzartikel zu werfen.

Die menschliche Aufsicht lag in der Aufbereitung des Datensatzes, der perfekten Produktion von variantenreichen Templates und in der Kontrolle nach dem Release der Texte. Man hatte Vertrauen in die Fähigkeiten des Systems und die eigene Sorgfalt. Die notwendige Transparenz schuf einen vertrauensvollen Rahmen für die Kund*innen. Dass Maschinen auch „halluzinieren“ und das Ergebnis des KI-Einsatzes einmal nicht „beherrschbar“ sein könnte, war damals noch nicht bekannt. Wer weiß, ob die APA schon eine hyperlokale Wahl-Berichterstattung hätte, wäre damals als erstes die LLM-Technologie genauer unter die Lupe genommen worden.

Im Zuge der Recherchen für eine groß angelegte Studie (vgl. Bailer et al., 2021) zum Einsatz von Künstlicher Intelligenz und künftigen Herausforderung im Medienbereich für das Technologieministerium wurde allerdings schnell klar, dass hier große Umwälzungen in den nächsten Jahren zu erwarten waren. Zum einen schienen Deep Fakes, deren Produktion immer einfacher und kostengünstiger wurde, die Demokratie zu bedrohen, weshalb bereits intensiv an Methoden geforscht wurde diese zu entlarven. Zum anderen ließen die Empfehlung der High Level Expert Group für AI der europäischen Kommission für eine Vertrauenswürdige Künstliche Intelligenz erahnen, dass es künftig einen Governance-Rahmen für den Einsatz von KI brauchen würde (High-Level Expert Group on Artificial Intelligence, 2019).

Die ersten Ansätze für eine Regulatorik offenbarten zudem, dass die weitreichenden Ausnahmebestimmungen, die Medien in der DSGVO auf Basis des sogenannten Medienprivilegs genießen, wohl nicht auf das Thema KI umlegbar wären. Da schon die Umsetzung der DSGVO deutliche Auswirkungen auf den Ideen- und Innovationsprozess gehabt hatte, wollte sich das Innovationsmanagement rechtzeitig mit den kommenden Neuerungen beschäftigen, um einen nachvollziehbaren Rahmen für die Produktentwicklung zu haben.

11.2.1 Die Entstehung der ersten KI-Leitlinie

Aus diesem Grund nutzte die APA ein im Jahr 2021 gestartetes Forschungsprojekt zur Entwicklung eines Bilderkennungsalgorithmus für Personen öffentlichen Interesses dazu nicht nur die entsprechende Technologie hervorzubringen, sondern auch ihre erste Leitlinie zum Umgang mit Künstlicher Intelligenz zu erstellen. Da zu diesem Zeitpunkt nur sehr wenige Guidelines aus dem Medienbereich vorhanden waren, musste sich das Projektteam viele Grundlagen mit Unterstützung der Beratungsfirma KPMG selbst erarbeiten. Von

Vorteil war, dass die Idee eines vertrauenswürdigen Einsatzes von Künstlicher Intelligenz perfekt zum Wertversprechen der APA passte. Aus „Trusted Content“ sollte „Trusted AI“ werden (vgl. Pig, 2023, S. 137 ff.).

Zur Operationalisierung des Vorhabens erarbeitete das Projektteam ein Vorgehensmodell für die von der High Level Expert Group adressierten Grundsätze:

1. Achtung der menschlichen Autonomie
2. Schadensverhütung
3. Fairness
4. Erklärbarkeit

Es wurde unter anderem festgelegt, dass im Rahmen jeder KI-Entwicklung zu Beginn eine Risikoeinschätzung vorgenommen werden muss. Weiters sollte für das KI-Training bevorzugt mit eigenen Daten in einem klar definierten Testsetting gearbeitet werden. Für das Training in diesem risikobehafteten Bereich musste das Vier-Augen-Prinzip angewendet werden. Einem etwaigen Bias in den Daten wollte man durch die Schaffung eines möglichst umfassenden, inklusiven Datensatzes begegnen. Die Privatsphäre von zufällig auf den Bildern auftauchenden Unbeteiligten sollte gewahrt bleiben, indem auf den Bildern ausschließlich Personen des öffentlichen Interesses indexiert und ein Opt-Out ermöglicht werden sollte.

Wesentlich war auch die Bestellung eines Visual Trust Officer und damit einhergehend eine klare Rollenverteilung sowie die Erstellung einer Checkliste zur Inbetriebnahme. Last but not least verpflichtete sich die APA zu einer umfassenden Transparenz durch Kennzeichnung des KI-Filters und die Beschreibung der Wirkmechanismen der zugrunde liegenden Technologie auf der Website. Mit der Verabschiedung der Leitlinie durch das Executive Board der APA im Frühjahr 2022 wurde das Regelwerk in Kraft gesetzt. Ein erster großer Wurf war gelungen.

11.2.2 Eine Taskforce als Wegbereiter

Und dann kam im November 2022 der Release von ChatGPT. Gemeinsam mit den kurz davor gelaunchten Bildgeneratoren à la Stable Diffusion und Midjourney war hier eine neue Dimension der KI auf den Plan getreten, die in den kommenden Monaten wesentlich den Diskurs bestimmen, den EU AI Act noch einmal auf den Kopf stellen und unmittelbar in das Herz von Medienunternehmen treffen sollte. Für die Geschäftsleitung war klar, dass die Beschäftigung mit dem Thema KI nun gehörig an Fahrt aufnehmen müsse.

Zu Jahresbeginn 2023 schlug daher die Geburtsstunde der Taskforce AI, deren Mitglieder sich zu einem großen Teil aus Teilnehmer*innen der früheren Governance-Workshops rekrutierten sowie weiterer Kolleg*innen aus dem Haus. Insgesamt zehn Mitglieder bekamen den Auftrag eine KI-Strategie für die APA auszuarbeiten, Schulungsmaßnahmen für die Mitarbeiter*innen aufzusetzen und die Leitlinie um das Kapitel „Generative AI“



Abb. 11.1 Überblick über die Kernelemente der Trusted AI-Strategie APA

zu ergänzen. Zudem sollten Stakeholder*innen aus dem Eigentümerumfeld zu ihren Bedürfnissen befragt werden. Damit wollte APA-CEO Clemens Pig den Grundstein für die Entwicklung eines „neuartigen österreichischen AI-Medienhubs operated by APA als gemeinsamer Wissensraum der heimischen Medienunternehmen“ legen. Hier sollte künstliche Intelligenz auf „Basis faktenbasierter Informationen trainiert und genutzt werden“ können (vgl. Pig, 2023, S. 152). Die organisatorische Klammer für die schnelle Eingreiftruppe bildete das APA-medialab.

Das interdisziplinäre Team startete damit sich die Grundsätze der Trusted AI-Strategie zu erarbeiten, die sowohl für Produkte aber auch für Prozesse zum Einsatz kommen sollte. Die nachfolgend abgebildeten Elemente sind auch heute noch gültig (Abb. 11.1).

Aufbauend auf den Kernelementen wurden Ableitungen für die Produktwelt getroffen, für die Produktion wurden rasch sichere Experimentierräume geschaffen. Die APA wollte ihren Mitarbeiter*innen möglichst früh Zugriff auf getestete Tools in einer sicheren Umgebung geben, anstatt den Einsatz zu verbieten. Verpflichtende Basisschulungen zu den Möglichkeiten und Grenzen der generativen Technologie aber auch zum Erkennen von manipulierten oder erstellten Bildern schufen die notwendigen Voraussetzungen für den Praxiseinsatz. Hackshops, kreative Workshop-Formate, sowie Unit-Trainings, die am konkreten, für die jeweiligen Teams relevanten Use Case die Technologie erprobten, halfen Schwellenängste abzubauen. Die überarbeiteten Leitlinien zeigten klar, welcher Einsatz befürwortet und welcher nicht gewünscht wurde.

Nun liegt es in der Natur der Sache, dass nicht in allen Bereichen die gleiche Intensität an Regeln gelten kann, trotzdem war das Bestreben da, ein Rahmenwerk für alle Units des APA-Konzern zu schaffen. Arbeiteten andere Medienhäuser nur Verhaltensregeln für

die Redaktion aus, so wurde in der Nachrichtenagentur das Marketing ebenso erfasst wie die IT. Dabei wurde mit definierten Freiheitsgraden für bestimmte Anwendungszwecke gearbeitet, um eine innovative Nutzung nicht zu behindern.

Wichtig war ein grundsätzlich offenes Mindset: Die Mitarbeiter*innen wurden zum Ausprobieren verschiedener Technologien aufgefordert. Für die Taskforce selbst sowie einzelne IT-Abteilungen, die neue Tools oder Technologien bewerteten und testeten, gab es größtmögliche Freizügigkeit bei Beachtung wesentliche Faktoren wie der Schutz des urheberrechtlich geschützten Contents und der Wahrung von Persönlichkeitsrechten und Geschäftsgeheimnissen. Die bestehende Regelung, die eine Weitergabe einer APA-Meldung an Dritte untersagt, wurde auf die Geschäftspraxis großer KI-Hersteller in der freien Nutzung der Tools angewandt.

11.2.3 KI und Mensch in Wechselwirkung

Die strengste Auslegung herrscht naturgemäß in der APA-Redaktion vor, wo generierte Bilder oder auch genuine KI-Artikel nicht veröffentlicht werden dürfen. In modifizierender und assistierender Form kommt KI aber auch dort etwa beim Lektorat oder in der Kürzung von Texten zum Einsatz. Sogar für die Adaption von Artikeln für das Medium Radio wird KI genutzt und verhalf einem vor vielen Jahren aus Ressourcengründen eingestellten Produkt zu neuem Dasein. Die Letztkontrolle durch die APA-Expert*innen ist hier anders als bei der alten, mittels NLG generierten Meldung obligatorisch.

Ebenfalls in Verwendung ist Speech-to-Text-Technologie, die von der APA-Tochter APA-Comm auch bei der Beobachtung von TV- und Radio-Beiträgen genutzt wird. Lange Zeit war hier die Qualität der Transkription zwar für das Medienmonitoring ausreichend, aber nicht für die Redaktion. Mit besseren Modellen, die auch österreichisches Deutsch gut interpretieren können, ist dieses Hilfsmittel jetzt auch im journalistischen Alltag einsetzbar.

Als Bereicherung haben sich Methoden erwiesen, die LLM-Systeme so gut wie möglich anleiten und Faktentreue erzwingen sowie Referenzen zur Überprüfung bereitstellen. Dies kommt dem journalistischen Arbeitsprozess des Checks-Rechecks-Doublechecks sehr entgegen. Retrieval Augmented Generation (RAG) ist eine dieser Techniken, die wie die semantische Suche auch für den Praxiseinsatz in den Distributionsplattformen der APA geeignet scheint. Hier soll sie mit entsprechenden „Guardrails“ versehen die Kund*innen künftig bei der Recherche unterstützen.

Um sich im Dschungel der Vorschriften für die Kennzeichnung von KI zurechtzufinden hat die APA gemeinsam mit dem KI Medien-Roundtable bei der österreichischen Rundfunk und Telekom Regulierungs-GmbH (RTR) Handlungsanleitungen für die Implementierung entwickelt und wird sich auch in der Erstellung eines nationalen AI Media Code of Conduct einbringen. Auch wenn einzelne Häuser schon früh mit dem Thema gestartet sind, so ist jetzt doch die Zeit gekommen, um zu harmonisieren, damit Nutzer*in-

nen nicht in einem Wald von undurchsichtigen KI-Lables ratlos zurückgelassen werden, sondern die KI-Kompetenz insgesamt gehoben werden kann. Medien können dabei einen wesentlichen Beitrag leisten, als verantwortliche Umsetzer aber auch als Erklärer der neuen Werkzeuge.

Digitale Technologien nicht um jeden Preis einzuführen, sondern nur dann, wenn sie wirklich einen echten Nutzen stiften und die Arbeit erleichtern bzw. einen Mehrwert für Kund*innen schaffen, ist eine wichtige Maxime. Der Mensch darf nicht zum Diener der Maschine werden, hochqualifizierte Redakteur*innen und Kund*innen nicht zu Faktencheckern unzuverlässiger Sprachmodelle. Und mit Blick auf die kommenden KI-Agenten-Systeme ist auch klar, dass Human-in-the-Loop nicht heißen kann, dass das letzte Glied der Kette die ganze Last möglicher fehlerhafter Zwischenschritte tragen muss. Das Design des Prozesses, der zum gewünschten Ergebnis führt, ist von entscheidender Bedeutung.

11.2.4 Organisatorische Weichenstellungen für die Transformation

Wie dieser Prozess gestaltet wird, darüber zu entscheiden obliegt seit Winter 2024 dem neu geschaffenen AI Board der Nachrichtenagentur, dem drei interne Teams zugeordnet sind. Diese bereiten die Umsetzung des EU AI Act vor, setzen Trainingsprogramme auf, entwickeln die KI-Produkte der neuen Generation und machen die Organisation fit für die Transformation. Die Ablöse der kleinen und wendigen Taskforce AI, die während ihres Bestehens ein Brennglas auf jeden dysfunktionalen Prozess gelegt hat und die Etablierung einer neuen Governance-Struktur machen deutlich, dass ein Konzern wie die APA gesamthaft Verantwortung für das Gelingen der Trusted-AI-Strategie übernehmen muss.

Nicht jede Frage lässt sich in kleinen Teams beantworten, bei wichtigen Themen ist sogar eine möglichst hohe Pluralität in der Entscheidungsfindung notwendig. Digitale Ethik ist ein Aushandlungsprozess zwischen den Stakeholdern und immer ein Abwägen von vielen Parametern. Das gilt insbesondere dann, wenn die Reife von Systemen vor dem Einsatz in der Praxis beurteilt werden soll. Gerade im Innovationsbereich muss sorgfältig abgewogen werden, wo eine KI-Anwendung in Sandbox-Umgebungen getestet werden muss und wo sie breit ausgerollt werden kann.

Das Prinzip der Technologie-Giganten Massentest mit völlig unfertigen Produkten zu fahren, muss stark hinterfragt werden. Ein Fauxpas wie etwa automatisch generierte Abstracts mit irreführenden Titeln, die noch dazu unter dem Markennamen von Medien publiziert werden, wie es der renommierten BBC in der Einführungsphase von Apple Intelligence passierte, untergräbt gleichermaßen die Glaubwürdigkeit des medialen Angebots UND das Vertrauen in eine leistungsfähige Technologie (BBC News, 2024). Der Mensch steht bei diesem Vorgehen gewiss nicht im Mittelpunkt. Hier müssen Unternehmen wie die APA sehr viel behutsamer vorgehen.

Ganz anders ist dies beim Konzept der „User Needs“ gelagert, jener Daten orientierten Vermessung von Nutzerbedürfnissen, die sich als gestaltendes Prinzip innovativer Medienhäuser etabliert hat. Im nachfolgenden Abschnitt soll darauf näher eingegangen werden.

11.3 User Needs im Zentrum

Nachrichtenagenturen als Akteure am Anfang der Lieferkette der medialen Produktion haben in der Regel keinen Zugang zu Lesern und Leserinnen. Ob die Nachrichtenartikel, Erklärstücke oder Interviews, die jeden Tag im APA-Newsroom sorgfältig erstellt werden, auch interessierte Abnehmer*innen bei den Kunden der Agentur, den TV-Anstalten und Medien verlegerischer Herkunft finden, blieb lange Zeit opak. Trotz der Messbarkeit der Performance von Artikeln im digitalen Raum und der hohen Übernahmequote von Ready-Made-News wurden Nutzungsdaten der Endkonsument*innen nur in seltenen Fällen geteilt. So blieb für Leitende Redakteur*innen der Agentur einzig der Daten orientierte Blick auf die von den Medienkunden verwerteten Artikel als Messgröße für die Steuerung – ein zwar aussagekräftiger, aber nicht vollständiger Key Performance Indikator. Als sich im Zuge von Paid Content-Strategien der Medienhäuser der Blick auf User*innen substantiell veränderte und deren Interessen immer mehr in den Vordergrund rückten, war die Zeit für gemeinsame Projekte gekommen.

11.3.1 Gemeinsames Lernen bei DRIVE

Das APA-medialab schloss sich im Jahr 2021 im Rahmen einer zweijährigen Projektpartnerschaft der von der Deutschen Presse-Agentur dpa und dem Beratungsunternehmen Schickler initiierten Digitale Revenue Initiative (DRIVE) an, bei der Medienhäuser aus Deutschland und Österreich mit den beiden teilnehmenden Nachrichtenagenturen die Performancedaten ihrer Artikel austauschten, um gemeinsam Erkenntnisse für die Entwicklung von Leserstrategien zu gewinnen. Aus dem Datenpool konnten nicht nur unter Wahrung der DSGVO-Vorschriften umfangreiche, statistische Datenanalysen über Lese- und Nutzungsgewohnheiten gewonnen, sondern auch Algorithmen für die Personalisierung trainiert werden. Letztere wurden dann auf den jeweiligen Portalen getestet und in Fokusgruppen weiter optimiert.

Begleitend zu den Analysen wurden laufend, erfolgreiche Konzepte für die Leserbindung evaluiert, die ihrerseits Daten orientiert arbeiteten. Einer der vielversprechendsten Ansätze war das vom Digital-Strategen Dmitry Shishkin für die britischen BBC World entwickelte User-Needs-Modell, das unterschiedliche Nutzerbedürfnisse identifizierte und kategorisierte. Es diente in der BBC-Redaktion der Planung von Themen und wurde im Jahr 1997 erstmals publiziert. Ziel war es neben dem knappen Nachrichtenformat noch weitere Möglichkeiten der Aufbereitung von Geschichten zum Einsatz zu bringen, das Engagement der Nutzer*innen zu steigern und sogenannte „Geisterartikel“ zu vermeiden.

Das Modell wurde seit Bestehen durch zahlreiche Redaktionen auf die eigenen Bedürfnisse angepasst. Im März 2023 launchte Shishkin eine überarbeitete 2.0-Version gemeinsam mit dem auf redaktionelle Datenanalysen spezialisierten Unternehmen Smartocto. Es beschreibt, wie in der Tab. 11.1 dargestellt, in vier Kategorien die acht Grundbedürfnisse bei der News-Nutzung (siehe Smartocto, 2023).

Tab. 11.1 User Needs im Überblick. (Quelle: eigene Darstellung nach Whitepaper Smartocto, 2023, User Needs 2.0)

Kategorie	User Need
Know	„Update me“ (Klassische Nachrichten)
Fakten-orientierten Wissens	„Keep me engaged“ (Nachricht mit speziellem Fokus auf interessante oder kuriose Aspekte)
Understand	„Educate Me“ (Factsheets und Hintergründe)
Kontext-orientiertes Verstehen	„Give me perspective“ (Analysen und Expertengespräche)
Feel	„Divert Me“ (Soft News)
Emotions-orientiert	„Inspire Me“ (Features, Porträts)
Do	„Connect Me“ (Community)
Aktionsorientiert	„Help me“ (Ratgeber und Service)

Studien bei der BBC und nachfolgend bei den DRIVE-Verlagen, die eine Variante des Modells mit sechs Dimensionen einsetzen, ergaben, dass die neuen am „Solution and Constructive Journalism“ ausgerichteten Formate die Nutzungsdauer und die Bindung an das journalistische Produkt erhöhten. So zeigte sich etwa, dass Artikel aus den Segmenten „Inspire me“- und „Help me“ das höchste Engagement erzielten, wobei letzteres vor allem durch eine hohe Media Time punktete. Die Conversion ins digitale Abonnement aber brachte zusätzlich zu „Inspire Me“ vor allem das einordnende Format „Give me Perspective“ (alles Fleischmann, 2023).

11.3.2 Einführung des User Needs Modells bei der APA

Der Erfolg des Modells war hinreichend belegt und schien geeignet auch in der täglichen Nachrichtenproduktion, vor allem aber bei der Themen-Planung der Nachrichtenagentur zum Einsatz zu kommen. Aus diesem Grund startete die APA-Chefredaktion im Herbst/ Winter 2023 eine groß angelegte Schulungsmaßnahme im Newsroom, um den eigenen Redakteuren und Redakteurinnen das Konzept näherzubringen und es mittels für Kunden auslesbare Metadaten auch für die Distribution im APA-NewsDesk fit zu machen.

Klare Ziele waren die Qualitätsverbesserung der eigenen Inhalte durch mehr Diversität, sowie bei entsprechender Datenlage die Erstellung eines Algorithmus für die Beschlagwortung von User Needs. Zur besseren Orientierung produzierte die Redaktion unter anderem das in Abb. 11.2 wiedergegebene Poster, das die Kategorien mit den wichtigsten journalistisch motivierten begleitenden Leitfragen zeigt.

Wenige Monate nach Einführung des Modells zeigte sich, dass vor allem in Krisenzeiten die einordnenden Artikel der Kategorien „Educate me“ und „Give me perspective“ deutlich an Fahrt aufnehmen. Hier ist die Agentur stark gefragt schnell entsprechendes Material bereitzustellen. Aber auch in der Berichterstattung zum Thema „Künstliche Intelligenz“ sind diese erklärenden Kategorien gut vertreten. Die Komplexität der Entwicklungen in diesem Bereich journalistisch gut zu begleiten, ist ein Gebot der Stunde,



Abb. 11.2 User Needs mit Leitfragen

kann hier doch gleich die KI-Kompetenz der Leser*innen angehoben werden. Freilich, für die tägliche Nachricht bleibt aufgrund des nicht unbeträchtlichen Ressourcenaufwands vorerst noch „Update Me“ das Mittel der Wahl.

11.3.3 Training der Algorithmen

Auch im Bereich der Daten orientierten Arbeit mit dem Modell starteten recht früh die Bestrebungen zur Entwicklung eines Algorithmus zur automatischen Verschlagwortung, um User Needs auch über den gesamten Bestand des APA-Mediendatenbank analysieren zu können.

Zugute kam der APA dabei das Forschungsprojekt „FAIRMedia“, das unter anderem die Kreation von qualitativ hochwertigen Mediendatensets auf Basis von User Needs vorsah (Joanneum, 2023). Der Fokus des FFG-Projektes „FAIRMedia“ lag auf der Entwicklung eines Werkzeugkastens für die effiziente Erstellung von vertrauenswürdigen Datensets aus aktuellen und archivierten Inhalten österreichischer Medienorganisationen sowie deren Nutzung für das Training von KI-Modellen. Es wurde im Winter 2025 abgeschlossen. Gemeinsam mit den Forscher*innen vom Institut für Innovation und Digitalisierung im Recht an der Universität Wien und anderen namhaften Innovator*innen arbeitete das Projektteam zuerst Regeln für eine faire Erstellung – wie etwa die Anonymisierung – aus und annotierte danach über 5000 Beiträge österreichischer Qualitätsmedien. Dabei zeigte sich, dass einige Kategorien unterrepräsentiert waren, weshalb erhebliche Anstrengungen unternommen wurden, um eine ausgewogenes Datenset sicherzustellen.

Der erstellte Datensatz wurde in weiterer Folge nicht nur für das Training eines Classifiers herangezogen, sondern auch im Rahmen der Entwicklung eines Dialog orientierten Tools zur Beantwortung von Fragen von Nutzer*innen auf Basis von Medieninhalten. Das Projektteam überprüfte dabei die These, ob in Geschichten mit besonders hoher Informationsdichte wie etwa in „Educate Me“ auch ein Nutzen für neue Q&A-Tools im Journalismus zu finden ist. Diese Angebote werden in naher Zukunft eine neue Produkt- oder Servicekategorie für Medienunternehmen darstellen.

Sie benötigen neben der semantischen, oft multimodalen Erschließung der Inhalte durch Embedding-Modelle und der Einbettung in Medien zentrierte Wissensarchitekturen (Knowledge Graphen) eben auch die richtige Auswahl an Inhalten für die Beantwortung der Nutzer*innen-Fragen. Neben der Aktualität könnte auch die Informationsdichte bestimmter User Needs-Kategorien hier eine entscheidende Rolle spielen.

11.3.4 Reflexionen und Ausblick

Ein zentrales Anliegen des Digitalen Humanismus ist es, menschenzentrierte Innovationen zu fördern. Er postuliert, dass digitale Technologien so gestaltet sein müssen, dass sie Demokratie und Inklusion fördern. Wenn sich Medienunternehmen vor diesem Hintergrund also Daten orientiert mit ihrem Angebot beschäftigen, müssen sie nach Strategien suchen, die das Empowerment von Nutzer*innen in das Zentrum ihres Handels stellen und diese auch messbar machen. Das User Needs Model ist dazu in besonderer Weise geeignet, führt es doch zu einer Win-Win-Situation: Medienhäuser können ihr Angebot besser planen und monetarisieren, die User*innen bekommen ein besseres Nutzungser-

lebnis. In Kombination mit Personalisierungsstrategien rückt die Idee von „Tailor Made News“, ein maßgeschneidertes Informationsangebot, das Jeden und Jede dort abholt, wo er oder sie gerade steht, in greifbare Nähe. Die Idee von „Tailor Made News“, heute oft unter dem Schlagwort „Liquid Media“ diskutiert, wurde in der Studie AI.AT.Media beschrieben (Bailer et al., 2021).

Da bei der Erstellung der Angebote immer häufiger Künstliche Intelligenz im Spiel ist und in Zukunft auch noch mehr in allen Phasen der Wertschöpfungskette genutzt werden wird, sind Medienunternehmen in besonderem Maße gefordert, diese transparent, nachvollziehbar und zum Wohle der Rezipient*innen einzusetzen. Dafür müssen laufend neue technische und organisatorische Leitplanken entwickelt werden, die die Mitarbeiter*innen (Journalist*innen, Developer*innen u. a.) in die Lage versetzen diesem Anspruch gerecht zu werden.

In der APA gelingt dies durch eine sorgfältige Auswahl der Tools und Technologien, einem Entwicklungsfokus, der Produktionsprozesse als auch Produkte gleichermaßen berücksichtigt und passend zum Automatisierungsgrad die jeweiligen Maßnahmen ableitet sowie eine Reihe von Schulungen und Weiterbildungsangebote wie etwa die Zertifizierung zur KI-Managerin bzw. zum KI-Manager. Denn die durch KI beschleunigte Transformation gelingt nur durch eine lernende Organisation.

Literatur

- Bailer, W., et al. (2021). *AI.AT.Media. AI and the Austrian Media Sector: Mapping the Landscape, Setting a Course*. Wien: Bundesministerium für Klimaschutz, Umwelt, Energie, Mobilität, Innovation und Technologie. <https://apa.at/wp-content/uploads/2022/04/AI-and-the-Austrian-Media-Sector.pdf>. Zugegriffen: 14. Dez. 2024.
- BBC News (2024). BBC complains to Apple over misleading shooting headline. <https://www.bbc.com/news/articles/cd0elzk24dno>. Zugegriffen: 14. Dez. 2024.
- Joanneum Research. FAIRMedia – Faire und vertrauensvolle Anwendung von künstlicher Intelligenz in den Medien. <https://www.joanneum.at/digital/projekte/fairmedia/> (Erstellt: 12. Nov. 2025). Zugegriffen: 30. Nov. 2025
- Fleischmann, K. DRIVE: Warum „Inspire me“-Artikel die wertvollsten Digital-Inhalte sind. Deutsche Presse-Agentur (dpa). <https://innovation.dpa.com/2023/03/02/drive-inspire-me-artikel-die-wertvollsten-digital-inhalte/> (Erstellt: 2. März 2023). Zugegriffen: 14. Dez. 2024.
- High-Level Expert Group on Artificial Intelligence (EU): Ethics Guidelines for Trustworthy AI. Europäische Kommission. <https://digital-strategy.ec.europa.eu/en/library/ethics-guidelines-trustworthy-ai> (Erstellt: 04.2019). Zugegriffen: 14. Dez. 2024.
- Nida-Rümelin, J., & Weidenfeld, N. (2018). *Digitaler Humanismus. Eine Ethik für das Zeitalter der Künstlichen Intelligenz*. München: Piper.
- Pig, C. (2023). *Democracy dies in Darkness, Fake News, Big Tech, AI: Hat die Wa(h)re Nachricht eine Zukunft?* Wien: Brandstätter.
- Schell, K. (2021a). Alltag von Datenjournalisten. Zum Potenzial der automatisierten Textberichterstattung. Whitepaper, APA – Austria Presse Agentur. <https://apa.at/blog/alltag-von-datenjournalisten/>. Zugegriffen: 14. Dez. 2024.

- Smartocto (2023). Whitepaper User Needs 2.0. <https://smartocto.com/research/userneeds/>. Zugegriffen: 14. Dez. 2024.
- Steurer, J. (1998). Die Spezies des digital-vernetzten Bürgers. In: *Die Presse*, 10. Jänner (S. 8). Abgerufen am 14. Dez. 2024 in der Mediendatenbank des APA-NewsDesk.

Weiterführende Literatur

- Schell, K. (2021b). Whitepaper Journalistische Textautomatisierung. Wien. <https://apa.at/blog/alltag-von-datenjournalisten/>. Zugegriffen: 14. Dez. 2024.

Mag. Verena Krawarik ist Leiterin der Stabsstelle Innovationsmanagement der APA – Austria Presse Agentur und führt das APA-medialab. Sie hat ihre beruflichen Wurzeln im Journalismus und war lange Jahre in der APA-Redaktion tätig. Aktuell verantwortet sie im Rahmen der Trusted AI-Strategie des Unternehmens die Themen Governance und AI Literacy. Seit Februar 2024 ist sie zudem Mitglied des KI-Beirates der österreichischen Bundesregierung, einem Beratungsgremium zur Umsetzung der KI-Strategie in Österreich. Die Innovationsexpertin wirkt derzeit an EU-Projekten wie der Etablierung eines „Trusted European Media Data Space“ mit sowie an der Umsetzung des Austrian Lab for AI Trust (ALAIT).

Ethik und Technologie im Einklang für eine bessere Patientenversorgung

Siegfried Meryn

Zusammenfassung

Die Digitalisierung des Gesundheitswesens bietet großes Potenzial für präzisere Diagnosen, bessere Behandlungen und leichten Zugang zu Gesundheitsdiensten. Zugleich entstehen ethische Herausforderungen, etwa bei der Wahrung von Autonomie, Datenschutz und sozialer Gerechtigkeit. Die revidierte Helsinki-Deklaration von 2024 und die österreichische eHealth-Strategie 2024 fordern einen menschenzentrierten Ansatz, der technologische Innovationen mit ethischen Prinzipien verbindet. Ärzt*innen, Patient*innen, Pflegekräfte, Führungskräfte und die öffentliche Hand müssen gemeinsam an Lösungen arbeiten, um digitale Technologien verantwortungsvoll einzusetzen und den menschlichen Aspekt der Medizin zu bewahren. Transparenz, Fairness und digitale Kompetenz aller Beteiligten sind dabei zentral.

Die Digitalisierung des Gesundheitswesens bietet enormes Potenzial, die Qualität der medizinischen Versorgung zu verbessern. Gleichzeitig stellt sie aber auch neue ethische Herausforderungen.

S. Meryn (✉)
Future Health Lab, Wien, Österreich
E-Mail: meryn@meryn.at

12.1 Rahmenbedingungen

Die zunehmende Digitalisierung des Gesundheitswesens bietet enormes Potenzial, medizinische Prozesse zu optimieren und die Gesundheitsversorgung zu verbessern. Fortschritte in der Künstlichen Intelligenz (KI), Robotik, Big Data und der Telemedizin bieten völlig neue Möglichkeiten, endlich Prävention in den Mittelpunkt zu stellen, Krankheiten früher und präziser zu diagnostizieren, Behandlungen zu individualisieren und damit die Erfolgchancen zu verbessern und den Zugang zu Gesundheitsdiensten zu erleichtern. Gleichzeitig wirft die Digitalisierung jedoch auch erhebliche ethische Fragen auf. Wie lassen sich technologische Fortschritte mit der Wahrung der Menschenwürde und der Autonomie von Patient*innen vereinbaren? Diese und andere Fragen stehen im Zentrum des Digitalen Humanismus, der darauf abzielt, den technologischen Fortschritt im Gesundheitswesen menschenzentriert und ethisch verantwortlich zu gestalten.

Ein zentraler ethischer Leitfaden, der in diesem Zusammenhang oft herangezogen wird, ist die revidierte Helsinki-Deklaration von 2024. Diese internationale Vereinbarung, die ursprünglich 1964 verabschiedet und mehrfach überarbeitet wurde, legt ethische Prinzipien für die medizinische Forschung und Praxis fest. In ihrer neuesten Fassung von 2024 berücksichtigt die Deklaration explizit die Herausforderungen der Digitalisierung und betont, dass technologische Innovationen stets im Einklang mit ethischen Grundsätzen wie Autonomie, Gerechtigkeit und dem Schutz der Privatsphäre stehen müssen (World Medical Association, 2024). Zusätzlich formuliert die eHealth-Strategie Österreich (BMSGPK, 2024) konkrete Ziele für die digitale Transformation im Gesundheitswesen, darunter der Sicherstellung eines digitalen Zugangs zum Gesundheitssystem (Ziel S1), der Zugänglichkeit von Innovationen (Ziel S7) und der Stärkung digitaler Kompetenzen (Ziel S8). Diese strategischen Ziele spielen eine zentrale Rolle bei der Umsetzung eines humanen, gerechten und zugänglichen digitalen Gesundheitssystems in Österreich. Weiters gibt es die KI – Strategie der Stadt Wien, die eine Teilstrategie der Digitalen Agenda 2030 der Stadt Wien ist. Dabei wird auch festgehalten, dass der digitale Humanismus jenen Zugang zur Digitalisierung beschreibt, der nicht die Technologie als Selbstzweck, sondern die Bedürfnisse der Menschen in den Fokus der Digitalisierung rückt. Außerdem gibt es die Strategie der Bundesregierung für Künstliche Intelligenz (o. J.), Annex, AIM AT 2030 (Artificial Intelligence Mission Austria) bei der auf Seite 17 und 18 explizit auf die KI im Gesundheitssystem eingegangen wird. Die potentiellen Anwendungsfelder wird festgehalten, erstrecken sich über das gesamte Spektrum von Gesundheitsförderung über Prävention, Diagnose, Therapie, Nachsorge bis hin zu systemischen Aufgaben wie Public Health und Pandemiemanagement.

12.2 Erwartungen von Ärzt*innen und Patient*innen zum Thema Digitalen Humanismus in der Medizin

Die Digitalisierung des Gesundheitswesens hat das Potenzial, Ärzt*innen dabei zu unterstützen, ihre Arbeit effizienter und präziser zu gestalten. Gleichzeitig bestehen jedoch klare Erwartungen, dass diese technologischen Fortschritte nicht auf Kosten der menschlichen Interaktion und der ethischen Verantwortung gehen dürfen. Ärzt*innen sehen sich selbst nicht nur als Expertinnen für medizinische Diagnosen und Behandlungen, sondern auch als Ansprechpartnerinnen für die persönlichen und emotionalen Bedürfnisse ihrer Patient*innen.

Ein zentrales Thema ist also die Wahrung der Patientenautonomie. Wie in der eHealth-Strategie Österreich 2024 zurecht hingewiesen, wird in naher Zukunft nicht die Patient*in zur Medizin, sondern die Medizin zur Patient*in kommen. Dies bedeutet zum Beispiel „HomeSpital“, dass Basisdiagnose- und Therapiegeräte zu Hause verfügbar sind und dass die entsprechenden Daten elektronisch übermittelt und über Video-Konsultation mit ärztlichem oder Gesundheitspersonal besprochen werden. Ein weiterer technologischer Entwicklungsbereich sind Plattformen für virtuelle Welten mit Gesundheitsbezug, in denen virtuelle Gesundheits- und Arztbesuche, durch Virtual Reality unterstützte Physiotherapie oder andere Formen gesundheitsbezogener Interaktionen möglich werden. Die Medizin wird einen kompletten Paradigmenwechsel von „Sick Care, episodisch, physisch, one size fits all, Bezahlung-pro-Service“ zu „Health Care, kontinuierlich, virtuell, personalisiert, Smart Hospitals, Robotik, automatisierte Diagnostik, Crowd Sourcing, AI-enhanced und Bezahlung-nach-Ergebnis“ erleben.

Die österreichische Ärztekammer fordert eine nationale Roadmap für Telemedizin sowie eine Breitbandmilliarde, um den Zugang zur Telemedizin flächendeckend zu ermöglichen. Weiters den begleitenden Einsatz von Gesundheits-Apps für Patienten, etwa im psychischen Bereich wie Stressbewältigung oder Burn-out, aber auch bei Diabetes, Bluthochdruck oder Schlafproblemen. Gleichzeitig werden große Hoffnungen in den neuen Europäischen Gesundheitsdatenraum (EHDS) gesetzt, in dem Gesundheitsdaten für alle Patienten europaweit abrufbar sein werden. Durch die Verknüpfung der Informationen der nationalen Gesundheitssysteme soll es in Zukunft auch möglich sein, dass etwa der behandelnde Arzt in Italien die Krankengeschichte eines Urlaubers aus Österreich am Computer abrufen kann (Österreichische Ärztezeitung, 2024).

Technologien wie KI, telemedizinische Plattformen und Gesundheits-Apps müssen daher so gestaltet sein, dass sie die Patient*innen in die Lage versetzen, fundierte Entscheidungen über ihre Gesundheit und Daten zu treffen. Dies bedeutet, dass die Informationen klar, transparent und leicht zugänglich sein müssen. Gleichzeitig müssen die Anwendungen so gestaltet sein, dass sie die Privatsphäre und die Daten der Patient*innen schützen. Besonders der Datenschutz ist in einer zunehmend digitalisierten Welt ein zentrales Anliegen des Gesundheitswesens.

Hier kommt auch die eHealth-Strategie Österreich (Ziel 8) ins Spiel, die darauf abzielt, digitale Kompetenzen bei Patient*innen und Fachkräften zu stärken. Durch die Stärkung der digitalen Fähigkeiten wird sichergestellt, dass sowohl Ärztinnen als auch Patient*innen in der Lage sind, die technologischen Werkzeuge effektiv und verantwortungsvoll zu nutzen (BMSGPK, 2024). In Kombination mit den ethischen Leitlinien der Helsinki-Deklaration von 2024 schafft dies die Grundlage für eine verantwortungsbewusste Nutzung digitaler Innovationen im Gesundheitswesen.

12.3 Rollenverteilung zwischen Ärzt*in und KI

Die Debatte über den Einsatz künstlicher Intelligenz (KI) in der Medizin hat eine neue Wendung genommen. KI-gestützte Systeme können laut neuester M.I.T.-Harvard und Harvard-Stanford Studien in bestimmten medizinischen Anwendungen unabhängig oft präzisere Diagnosen stellen als in Kombination mit menschlichen Ärzten. Diese Erkenntnis stellt die klassische Rollenverteilung zwischen Ärzt*innen und KI infrage und erfordert eine Neubewertung der optimalen Zusammenarbeit. Während lange davon ausgegangen wurde, dass eine Kombination aus menschlicher Expertise und maschineller Intelligenz stets zu besseren Ergebnissen führt, zeigen aktuelle Entwicklungen, dass dies nicht immer der Fall ist. Eine der Herausforderungen besteht darin, dass medizinisches Personal noch nicht ausreichend mit den Möglichkeiten und Grenzen von KI vertraut ist. Zudem kann eine gewisse Skepsis gegenüber automatisierten Entscheidungshilfen bestehen, die als „Automation Neglect“ bezeichnet wird. Darüber hinaus ist es wichtig, die methodischen Einschränkungen bisheriger Untersuchungen zu berücksichtigen. Die M.I.T.-Harvard-Studie hat gezeigt, dass Radiologen dazu neigen, die von KI bereitgestellten Informationen zu unterschätzen und stattdessen ihrer eigenen, oft weniger präzisen Einschätzung zu vertrauen. Dies verdeutlicht, dass eine effektive Integration von KI in den medizinischen Alltag nicht nur technologische, sondern auch kulturelle und bildungspolitische Maßnahmen erfordert.

Vielversprechende Ansätze zur Arbeitsteilung sind:

1. Ärzt*innen führen die Patientenanamnese durch und sammeln klinische Daten, während KI diese Informationen analysiert und Diagnosevorschläge macht.
2. KI wird zur Erstbewertung und zum Screening genutzt, während komplexe Fälle weiterhin von Ärzt*innen behandelt werden.
3. KI-Systeme unterstützen medizinisches Personal durch Entscheidungsfindungshilfen, ohne dass die finale Diagnose automatisch von der KI übernommen wird.

Die Zukunft der Medizin wird nicht von einer simplen Entscheidung zwischen menschlicher und künstlicher Intelligenz geprägt sein. Vielmehr muss ein optimales Zusammenspiel entwickelt werden, das die Stärken beider Seiten nutzt und die Patientenversorgung nachhaltig verbessert.

Dazu wird es aber auch notwendig sein, dass die Ärzteschaft sich mit diesen neuesten Entwicklungen intensiv beschäftigt, ob dies bereits im Medizinstudium der Fall ist oder aber auch zwingend in der nachfolgenden Aus-, Fort- und Weiterbildung. Es wird auch notwendig sein, dass man hier an der Entwicklung aktiv mitarbeitet und diese ständig unter Berücksichtigung ethischer Aspekte mit dem Menschen und Patient*innen im Zentrum des Handelns und Wirkens mitentwickelt. Die Etablierung von eigenen Fachärzten wie zum Beispiel „Digital Doctors“ und Data-Scientists sowie Spezialisten für ethisch-digitale Themen für die Kooperation mit KI-Systemen ist dringend angezeigt.

12.4 Zugang, Qualität und Leistbarkeit: Konkrete Beispiele

Beim Zugang zu medizinischen Leistungen spielt die eHealth-Strategie Österreich wieder eine zentrale Rolle. Durch das strategische Ziel 7, das den Zugang zu Innovationen fördert, wird sichergestellt, dass auch benachteiligte Bevölkerungsgruppen von digitalen Fortschritten profitieren. Diese Zielsetzung steht im Einklang mit der revidierten Helsinki-Deklaration von 2024, die fordert, dass alle Bevölkerungsgruppen Zugang zu medizinischen Innovationen erhalten und dass technologische Fortschritte auf eine gerechte und zugängliche Weise genutzt werden (World Medical Association, 2024).

Umsetzungen und konkrete Beispiele dieser Rahmenvorgaben, von der alle Bevölkerungsgruppen profitieren, sind in folgenden Punkten schon sichtbar und sollten umgesetzt werden:

- Ausbau der elektronischen Gesundheitsakte ELGA in Österreich: Patient*innen und Gesundheitsdienstleister haben Zugriff auf wichtige Gesundheitsdaten, wobei die Sicherheit und der Schutz dieser Informationen höchste Priorität haben.
- Telemedizinische Angebote: Ab 2026 soll die österreichische Gesundheitshotline 1450 Terminbuchungen und Videokonsultationen ermöglichen. Dies verbessert nicht nur den Zugang zur Gesundheitsversorgung, sondern erhöht auch die Qualität der Beratung, da die Mitarbeiter auf die elektronische Gesundheitsakte zugreifen können.
- Digitale Gesundheitsanwendungen: Für chronisch Kranke, wie beispielsweise Diabetiker, werden Apps entwickelt, die bei der Überwachung von Ernährung und Zuckerniveaus helfen. Diese Anwendungen ermöglichen eine kontinuierliche Betreuung und können die Behandlungsqualität verbessern.
- E-Rezept: Die Einführung des E-Rezepts erleichtert den Zugang zu Medikamenten, reduziert den administrativen Aufwand und kann Kosten senken und damit langfristig den Zugang für einkommensschwache Bevölkerungsschichten unterstützen.
- Telemedizin für den ländlichen Raum: Durch den Ausbau telemedizinischer Angebote können auch Menschen in abgelegenen Regionen Zugang zu spezialisierter medizinischer Versorgung erhalten, insbesondere dort, wo eine Ausdünnung der niedergelassenen Allgemeinmedizin zu beobachten ist.

Auch die Strategie der österreichischen Bundesregierung für künstliche Intelligenz aus dem Jahr 2021, meiner Meinung nach aufgrund der rasanten Entwicklungen im Bereich der KI-Techniken dringend revisionsbedürftig, legt einen klaren digitalhumanistischen Fokus auf die vielfältigen Anwendungsfelder sowie den ethischen Kontext der Technologie. Es zeige sich das Potenzial von KI im Sinne der Qualität und Leistbarkeit von Medizin in folgenden Einsatzgebieten:

- die Erstellung von individuellen oder bevölkerungsweiten Risikoprofilen für Diagnose und Prognose (siehe z. B. die Vielzahl an Modellen zu Covid-19 Diagnose und Krankheitsverlauf)
- KI-unterstützte Datenaggregation (etwa für epidemiologisches Monitoring)
- die Verwendung von künstlichen neuronalen Netzen in der diagnostischen Auswertung von medizinischen Bilddaten, und Audiodaten
- KI-unterstützte Medikamentenentwicklung und personalisierte Medizin
- therapeutische Intervention unter Rückgriff auf Datenquellen aus neuartiger Sensorik (z. B. Wearables)
- der Einsatz von KI-Algorithmen für Patient*innen-Interaktion (z. B. Chatbots, Entscheidungsunterstützungssysteme, Studienrekrutierung etc.)

Insbesondere der Bereich der Patient*innen-Interaktion verdeutliche, dass maschinelles Lernen nicht nur bestehende Prozesse optimiert, sondern auch tiefgreifende Transformationen in traditionellen Berufsbildern des Gesundheitswesens vorantreiben könne.

Zu den großen Herausforderungen der Anwendung von KI im Gesundheitsbereich zähle – über die technischen und medizinischen hinausgehend – die Klärung der erforderlichen Rahmenbedingungen, wie z. B.: Datenschutz und Datenverfügbarkeit, sozioökonomische Fragen, wie die Auswirkungen von KI auf die Gesundheitsberufe und Interaktionsformen im Gesundheitswesen, ethische Aspekte (Algorithmen- und Datenethik, Transparenz und explainable AI) und rechtliche Anforderungen (Marktzulassung/ Medizinproduktrecht, Haftungsfragen).

12.5 Die Kluft zwischen digitaler Verantwortung und kommerziellen Interessen

Ein zentrales Problem im Bereich der digitalen Medizin ist die Diskrepanz zwischen Gesundheitswesen und Gesundheitswirtschaft, beziehungsweise der ethischen Verantwortung gegenüber den Patient*innen und den kommerziellen Interessen von Unternehmen und Ärzt*innen. Während Ärzt*innen und medizinische Einrichtungen meistens das Wohl der Patient*innen in den Mittelpunkt stellen, stehen für Unternehmen häufig wirtschaftliche Ziele im Vordergrund. Diese Kluft wird besonders im Bereich der KI deutlich, wo die Entwicklung neuer Technologien oft von kommerziellen Interessen geleitet wird.

Ein Beispiel dafür ist der Einsatz von KI in der Arzneimittelentwicklung. KI-basierte Systeme können den Prozess der Medikamentenentwicklung erheblich beschleunigen und die Kosten senken, was dazu führt, dass neue Medikamente schneller auf den Markt kommen und potenziell mehr Menschen zugutekommen. Allerdings besteht mehr denn je die Gefahr, dass der stetig steigende kommerzielle Druck dazu führt, dass ethische Überlegungen, wie der gerechte Zugang zu diesen Medikamenten, vernachlässigt werden.

Um diese Kluft zu überbrücken, sind klare ethische Leitlinien erforderlich, die sicherstellen, dass technologische Innovationen sowohl den wirtschaftlichen Anforderungen als auch den ethischen Prinzipien gerecht werden. Die revidierte Helsinki-Deklaration von 2024 fordert, dass der Einsatz von KI und anderen digitalen Technologien im Gesundheitswesen stets mit einer klaren ethischen Verantwortung einhergehen muss. Insbesondere der Schutz der Patientenautonomie und die Gewährleistung eines gerechten Zugangs zu neuen Technologien sind zentrale Forderungen der Deklaration (World Medical Association, 2024).

Ein weiteres wichtiges Thema in diesem Zusammenhang ist die Verantwortung, die mit dem Einsatz von KI und anderen Technologien einhergeht. Wer trägt die Verantwortung, wenn eine KI eine falsche Entscheidung trifft? Diese Frage wird zunehmend in der medizinischen Ethik diskutiert. Die Helsinki-Deklaration von 2024 fordert klare Richtlinien für den Einsatz von KI, die sicherstellen, dass die Verantwortlichkeiten eindeutig festgelegt sind. Dies gilt sowohl für die Entwickler der Technologien als auch für die Ärzt*innen, die diese Technologien in ihrer Praxis einsetzen.

12.6 Der österreichische Bioethikrat und der Digitale Humanismus

Der österreichische Bio-Ethikrat hat in den letzten Jahren mehrfach Stellung zu den ethischen Herausforderungen der Digitalisierung im Gesundheitswesen bezogen. Besonders im Bereich der neuen digitalisierten Arbeitswelt, der Künstlichen Intelligenz und der Robotik sieht der Bioethikrat große ethische Herausforderungen, die es zu bewältigen gilt. In einer seiner jüngsten Stellungnahmen zur Rolle von KI in der Medizin betont der Rat, dass der Einsatz dieser Technologien nur unter strenger Beachtung ethischer Prinzipien erfolgen darf. Ein zentrales Thema ist der bereits angesprochene Schutz der Patientenautonomie. KI-basierte Systeme und andere digitale Technologien dürfen nicht dazu führen, dass Patient*innen die Kontrolle über ihre Gesundheitsdaten verlieren. Dies ist besonders wichtig, da viele KI-Systeme auf einer „Black-Box“-Struktur basieren, bei der die Entscheidungsfindung der KI für die Ärztinnen und Patient*innen nicht transparent ist. Der Bioethikrat fordert daher, dass KI-Systeme so gestaltet werden müssen, dass sie für alle Beteiligten nachvollziehbar sind und dass die endgültige Entscheidung über eine Behandlung immer bei den Ärztinnen und Patient*innen liegt.

Auch die eHealth-Strategie Österreich unterstützt dieses Ziel, indem sie sicherstellen will, dass alle Patient*innen einen fairen und transparenten Zugang zu digitalen Gesundheitsdiensten im Bereich Gesundheit und Pflege erhalten. Es wird dabei auch festgehalten,

dass im Hinblick auf die Nutzung digitaler Gesundheitsangebote die sozioökonomischen sowie geschlechts- und altersbezogenen Ungleichheiten, die auf Unterschiede im Zugang zu Internet beziehungsweise Technologien, aber auch auf die Unterschiede in der digitalen Gesundheitskompetenz unterschiedlicher Bevölkerungsgruppen zurückzuführen sind, noch stärker zu berücksichtigen sind (World Health Organization. Regional Office for Europe, 2023). Dies deckt sich mit den ethischen Grundsätzen des Bioethikrats, der fordert, dass technologische Innovationen im Gesundheitswesen im Einklang mit Gerechtigkeit und Transparenz stehen müssen (BMSGPK, 2024).

Eine weitere Herausforderung in Bezug auf den Digitalen Humanismus im Gesundheitswesen ist die Schaffung eines regulatorischen und organisatorischen Rahmens, der Sicherheit für alle Akteur*innen bietet und dadurch die Umsetzung und Nutzung innovative digitaler Lösungen auf allen Versorgungsebenen fördert (Wong et al., 2022).

12.7 Ethische Imperative für Manager*innen in einer digitalisierten Gesundheitsbranche

Auch Führungskräfte im Gesundheitswesen stehen vor der Herausforderung, die Digitalisierung in ihre Einrichtungen zu integrieren, ohne dabei die ethischen Prinzipien des Digitalen Humanismus zu vernachlässigen. Managerinnen müssen sicherstellen, dass technologische Innovationen nicht nur die Effizienz steigern, sondern auch die Qualität der Patientenversorgung verbessern. Dies erfordert die Entwicklung von ethischen Leitlinien und die enge Zusammenarbeit mit Ärztinnen und Ethikkommissionen.

Ein weiterer zentraler Imperativ für Führungskräfte ist es, sicherzustellen, dass der Einsatz von KI und anderen Technologien transparent und nachvollziehbar ist. Dies umfasst den Datenschutz ebenso wie die Sicherstellung, dass Patient*innen umfassend über die Funktionsweise und die möglichen Risiken neuer Technologien informiert werden. Darüber hinaus müssen Manager*innen darauf achten, dass digitale Innovationen gerecht verteilt werden und allen Patient*innen gleichermaßen zugutekommen.

Die revidierte Helsinki-Deklaration von 2024 hebt die besondere Verantwortung von Führungskräften im Gesundheitswesen hervor. Sie fordert, dass technologische Innovationen immer unter Berücksichtigung ethischer Standards entwickelt und eingesetzt werden müssen. Manager*innen müssen dafür Sorge tragen, dass der Einsatz von KI und anderen Technologien stets transparent und nachvollziehbar ist und dass die ethischen Grundsätze des digitalen Humanismus in der Praxis umgesetzt werden (World Medical Association, 2024).

Zusätzlich zur ethischen Verantwortung fordert die eHealth-Strategie Österreich mit Ziel 8 die Stärkung der digitalen Kompetenzen bei Führungskräften und medizinischem Personal, um sicherzustellen, dass sie die technologischen Innovationen nicht nur verstehen, sondern auch effizient und ethisch verantwortungsvoll einsetzen können (BMSGPK, 2024).

12.8 Fazit

Im Kontext der Digitalisierung des Gesundheitswesens ist es also unerlässlich, dass alle Beteiligten – Ärzt*innen, Patient*innen, Pflegebereich, Führungskräfte und die öffentliche Hand – gemeinsam daran arbeiten, technologische Fortschritte verantwortungsvoll und ethisch fundiert zu integrieren. Die analysierten Strategien bieten hierbei wertvolle Leitlinien, die sicherstellen, dass der Mensch im Mittelpunkt bleibt. Transparenz, Fairness und digitale Kompetenz sind zentrale Prinzipien, die es zu fördern gilt, um die Potenziale der Digitalisierung in der Medizin voll auszuschöpfen und gleichzeitig die ethischen Herausforderungen zu meistern. Nur so kann eine zukunftsfähige und gerechte Gesundheitsversorgung gewährleistet werden, die den menschlichen Aspekt der Medizin bewahrt und die Autonomie und Würde der Patient*innen respektiert.

Literatur

- BMSGPK (2024). *eHealth-Strategie Österreich. v1.0 im Juni 2024*. Wien: Bundesministerium für Soziales, Gesundheit, Pflege und Konsumentenschutz.
- Österreichische Ärztezeitung Ausgabe 23/24; erschienen 16. Dezember 2024, Wien
- Strategie der Bundesregierung für Künstliche Intelligenz, Annex, AIM AT 2030
- Wong, J., Meskó, B., Topol, E. J., & Matheny, M. E. (2022). Digital Health in the Era of Artificial Intelligence: Opportunities and Challenges. *The Lancet Digital Health*, 4(7), e526–e537.
- World Health Organization. Regional Office for Europe (2023). *Digital Health in the WHO European Region: the ongoing journey to commitment and transformation*. WHO Regional Office for Europe. https://cdn.who.int/media/docs/librariesprovider2/data-and-evidence/english-ddh-260823_7amcet.pdf. Zugegriffen: 23.12.2025
- World Medical Association (2024). *Deklaration von Helsinki – Ethische Prinzipien für medizinische Forschung am Menschen*. Verabschiedet von der 75. Generalversammlung der WMA, Helsinki.

Weiterführende Literatur

- Agarwal, N., Moehring, A., Rajpurkar, P., & Salz, T. (2023). *Diagnostic AI: Experimental Evidence from Radiology*. MIT Department of Economics Working Paper No. 23–19.

Em. Univ. Prof. Dr. Siegfried Meryn, geboren am 6. März 1954 in Wien, ist einer der führenden Internisten in Österreich und seit langen Jahren Facharzt und Professor für Innere Medizin und Additivfacharzt für Gastroenterologie, Hepatologie und Endoskopie an der Medizinischen Universität. Nach seiner Matura und dem Medizinstudium an der Universität Wien promovierte er 1977. Er absolvierte Facharztausbildungen in Wien, Birmingham und London und forschte bei Nobelpreisträgerin Rosalyn Sussman Yalow in New York. Meryn ist spezialisiert auf Innere Medizin, Gastroenterologie, Hepatologie und Endoskopie sowie Long COVID. Er war Gastprofessor in Kanada und ist Autor zahlreicher Fachpublikationen und Bestseller. Seit 2025 ist er Chefarzt des Wiener Roten Kreuzes. Meryn beschäftigt sich intensiv mit der Verbindung von Technologie und menschlichen Werten. Seine Arbeit im Bereich Digitaler Humanismus in der Medizinbranche, insbesondere die ethische Verantwortung und die Qualität der medizinischen Versorgung, sind zentrale Themen.

Vier Thesen als Fundament und Überblick über die A1 AI-Guidelines

Michael Höfler

Zusammenfassung

Der Artikel beleuchtet, wie A1 den digitalen Humanismus in der Praxis umsetzt. Es wird diskutiert, wie KI und Digitalisierung genutzt werden, um die Effizienz zu steigern, ohne den menschlichen Beitrag zu vernachlässigen. Der Fokus liegt auf der Balance zwischen technologischen Fortschritten und ethischen Überlegungen, um sicherzustellen, dass die Technologie den Menschen zugutekommt. Beispiele wie die A1 Fair AI-Guidelines zeigen, wie A1 mögliche Bias minimiert und humanistische KI-Modelle entwickelt.

Technologie sollte nicht nur funktional sein, sondern stets im Einklang mit menschlichen Werten gestaltet werden.

13.1 Einleitung

Die rasante Entwicklung Künstlicher Intelligenz (KI) stellt Unternehmen aller Branchen vor fundamentale Fragen – besonders jene, wie die A1, die selbst im Bereich Informations- und Kommunikationstechnologie (IKT) tätig sind. Wie geht man als Technologieunternehmen verantwortungsvoll mit KI um – nicht nur im Hinblick auf unsere Produkte und Dienstleistungen, sondern auch im Umgang mit unseren Mitarbeitenden und Kund*innen?

M. Höfler (✉)

A1 Telekom Austria Group, Wien, Österreich

E-Mail: michael.hoefler@a1.group

Während KI-Systeme zunehmend Aufgaben übernehmen, die einst menschlicher Intelligenz vorbehalten waren, zeigt sich zugleich die Notwendigkeit zur Unterscheidung: zwischen einer „tumb“ agierenden KI, die lediglich auf Datenbasis Muster erkennt und Entscheidungen trifft, und einer KI-Anwendung, die in einem größeren ethischen, sozialen und kulturellen Rahmen reflektiert wird. Genau hier setzt der Begriff des *Digitalen Humanismus* an – als Leitidee für ein technologiegestütztes, aber zugleich menschenzentriertes Handeln.

In der aktuellen KI-Debatte ist der Digitale Humanismus kein akademischer Luxus, sondern ein praktischer Kompass. Gerade IKT-Unternehmen stehen in der Verantwortung, ihre Innovationskraft mit gesellschaftlicher Gestaltung zu verbinden. Denn die entscheidende Frage ist nicht, ob wir KI nutzen – sondern wie wir sie gestalten: mit Blick auf Werte, Verantwortung und den unverzichtbaren menschlichen Beitrag. Hier vier Schritte einer Herleitung dieser Notwendigkeit.

13.2 Das Fundament: Warum Digitaler Humanismus für IKT-Unternehmen eine Notwendigkeit ist

Technologie-Skepsis ist unangebracht: IKT ist der Motor des Humanismus

Was bewegt die Welt und treibt sie voran.	Neue (Kommunikations-)Technologie waren immer schon Katalysator, also Beschleuniger für die Entwicklung der Menschheit und des Humanismus.
Neue Technologien sind der Basis und Beschleuniger menschlicher Entwicklung.	<i>Beispiel Buchdruck:</i> Die Reformation des 16. Jahrhunderts wäre nicht möglich gewesen ohne Buchdruck. Hätte Martin Luther seine Thesen nur in Worms von der Kanzel gepredigt, seine Wirkung wäre wesentlich geringer gewesen. Es war die gedruckte Übersetzung der Bibel, die die Reformation in Gang gesetzt hat.
Das Glasfasernetz ist der Buchdruck des 21. Jahrhunderts.	So wie vor 500 Jahren der Buchdruck die menschliche Entwicklung vorangetrieben (und überhaupt erst ermöglicht) hat ist es heute die Digitalisierung. Natürlich sind damit auch „Gefahren“ verbunden. Aber die Chancen überwiegen. Auch der Buchdruck war eine Gefahr für die Amtskirche – trotzdem will heute niemand mehr im Mittelalter leben. Im 21. Jahrhundert ist die Glasfaser ist die Lebensader der Gesellschaft und Basis für alles: Für Bildung, für Forschung, für wirtschaftliche Entwicklung, für die Energiewende (Smart Grid), für Fernsehen und Musik-Streaming . . .

KI/Digitalisierung wird Menschen ersetzen, aber nicht überflüssig machen

KI ist der „nur“ der nächste Schritt der Automatisierung.	<i>Wird KI, Menschen ersetzen? Ja! Wird KI, Menschen überflüssig machen? Nein!</i> Aber das ist keine neue Entwicklung: Schon die Dampfmaschine hat die menschliche Arbeitskraft ersetzt. 1977 sang Wolfgang Ambros durchaus wehmütig den Fahrscheinautomaten: „Schaffner sein, das war einmal was, die Zeit ist vorbei, heute fährt man schaffnerlos.“
Beispiel Fließband	Ob wir als Menschen, der Technik dienen (statt umgekehrt). Oder welche Auswirkungen neue Technologien auf unser (Arbeits)leben habe. Diese Frage hat Charlie Chaplin schon 1936 in <i>Modern Times</i> thematisiert. Die Ängste bei Einführung des Fließbandes sind nicht weit weg von der Diskussion über KI. Bei Charly Chaplin werden die Menschen zu Zahnrädern in der Produktion, die jegliche Individualität verlieren und am Fließband die immer gleichen Arbeitsabläufe reproduzieren (ohne selbe zu denken).
Die Entwicklung	Die Geschichte hat gezeigt, dass die Dystopie in <i>Modern Times</i> nicht eingetreten ist. Die Automatisierung der Industrie hat unser (Arbeits)leben verbessert und vereinfacht. Arbeit ist heute effizienter, gesünder und im Allgemeinen auch sinnstiftender. In manchen Fällen haben neue Technologien tatsächlich menschliche Arbeitskraft ersetzt, dennoch sind wir als Menschen nicht überflüssig geworden. Wir haben nur andere Aufgaben übernommen.
Die Nebenwirkungen	So wie beim Fließband müssen wir aber auch bei der Digitalisierung mögliche Nebenwirkungen mitdenken: z. B. wie stellen wir sicher, dass Mitarbeiterinnen nicht permanent „online“ sind und sich nicht überwacht fühlen. Die permanente Überwachung des Arbeiters durch den <i>Fabrikbesitzer</i> ist ebenso schon ein Motiv in <i>Modern Times</i> wie die <i>Work-Life-Balance</i>. Auch das ist nicht neu. Die Digitalisierung und <i>Remote-Working</i> verwischen die Grenze zwischen Arbeit und Freizeit nur noch ein weiteres Stück. <i>MS Teams</i> zeigt permanent meinen Status als verfügbar/beschäftigt/abwesend. Hier müssen wir aufpassen, dass wir Menschen nicht überwachen oder sich diese nicht überwacht fühlen. Sich das bewusst zu machen, ist schon der erste Schritt.

Was ist Leistung im Zeitalter der KI:

KI kann Aufgaben lösen, aber keine Lösungen finden

KI ist (noch) nicht kreativ	KI ist nicht kreativ. KI experimentiert nicht. KI optimiert vorhanden Lösung, sie findet keine neue Lösung.
Beispiel Hochsprung	Eine kleine Anekdote. Hochsprung, Olympische Spiele Mexiko 1968. Olympiasieger Dick Fosbury war der erste Mensch, der rückwärts über die Latte sprang. Obwohl zuerst belächelt, war der „Fosbury Flop“ schon in paar Jahre später die Standard-Technik im Hochsprung – und ist es bis heute. Fosburys Idee war eine neue kreative Antwort auf den technischen Fortschritt. Kurz zuvor waren dicke weiche Matten eingeführt worden, sodass die Sportler nicht mehr auf den Füßen landen mussten. KI – auch generative KI – greift auf Datenbanken zurück, erkennt Muster und reproduziert diese. Sie generiert „neues“ aus schon vorhandenem. Dick Fosbury – und das ist Kreativität – hat aber nicht die vorhandenen Muster reproduziert und optimiert, sondern durchbrochen. Das ist in der DNA von KI nicht vorgesehen.
Was ist in Zukunft Leistung?	Arbeit und Leistung wird in Zukunft anders bewertet werden: – Leistung heißt: Finden eines neuen Lösungswegs. – Kreativität heißt, vorhandene Muster zu durchbrechen. Leistung ist in Zukunft also genau das, was KI nicht kann.

KI übernimmt keine Verantwortung und denkt nicht mit

KI übernimmt keine Verantwortung	Digitalisierung und KI übernehmen immer mehr Aufgaben, aber sie macht das nur auf Befehl, KI setzt keine Eigeninitiative und übernimmt keine Verantwortung. Auch das wird Leistung in Zukunft sein: Die Fähigkeit und Bereitschaft Verantwortung zu übernehmen.
Attitude wird wichtiger, Fachwissen verliert an Bedeutung	Wir haben gerade ein A1 Internes Learning zum Thema „Prompt Writing“ gestartet. Das wird eine Schlüsselqualifikation der Zukunft. Wie „kommandiere“ ich KI richtig. Wir folgen im Recruiting immer öfter dem Prinzip „Hire for Attitude“. Auch das wird in den nächsten Jahren immer mehr Teil unseres Verständnisses von Leistung. Wer denkt mit, wer zieht gemeinsam am gleichen Strang und übernimmt Verantwortung. Da sich fachliche Anforderungen – gerade in unserer Branche – immer schneller ändern werden diese unwichtiger. Beim Programmieren und Erstellen von Code wird KI ein Game-Changer werden. Dafür wird die richtige Einstellung noch wichtiger.

Diese Herleitung ist ein Fundament und eine Einordnung für humanistischen KI-Einsatz in einem Unternehmen wie A1. Um aber wirksam zu werden, braucht es eine konkrete Umsetzung in Projekten, Leitlinien und Applikationen. Diese sind die Nagelprobe für gelebten Digitalen Humanismus.

13.3 Konkret: Digitaler Humanismus im Live Einsatz bei A1: Die A1 Fair AI-Guidelines

Künstliche Intelligenz wird oftmals als Synonym für maschinelles Lernen verstanden. Dabei lernt die KI anhand von Beispielen aus großen Datenmengen, um daraus selbstständig Muster und Regeln zur Lösung einer bestimmten Aufgabe zu finden. KI lernt und verknüpft also vorhandene Daten. Das ist ein Vorteil bei wiederkehrenden standardisierten Aufgaben wie beispielsweise der Analyse eines Blutbildes. Der Nachteil beim maschinellen Lernen ist, dass dadurch ein Hang zur statistischen Mitte entsteht, welcher Ränder in zu geringem Ausmaß oder gar nicht berücksichtigt. Man kann also davon ausgehen, dass KI voreingenommen oder gebiased ist. In diesem Kontext bedeutet Digitaler Humanismus auch das zu inkludieren, was nicht der Norm entspricht.

Um Fairness beim Einsatz von KI-Modellen sicherzustellen, wurden die „A1 Fair AI-Guidelines“ ins Leben gerufen. Ein Team von A1 hat sich bereits im Jahr 2022 mit dem Thema „Fairness“ bei dem Einsatz von KI-Lösungen beschäftigt. Also: Bereits vor dem Einsatz der Modelle beschäftigen wir uns also mit möglichen Auswirkungen des Modells auf unterschiedliche Kundengruppen. Im Sinne der Transparenz müssen dabei im Vorhinein folgende Fragen beantwortet werden:

- Was ist die Anforderung, welche das Modell lösen soll
- Welche Daten und Zielgruppendefinition werden für das Training des Modells verwendet (mögliche Bias in Trainingsdaten)
- Welche Attribute werden verwendet, wie sensitiv sind diese und wie stark beeinflussen sie die Entscheidung
- Welcher KI-Algorithmus wird verwendet
- Was ist der mögliche Schaden des Modells auf unterschiedliche Zielgruppen unter besonderer Berücksichtigung von Randgruppen
- Was wurde unternommen, um den Bias zu minimieren

In Bezug auf sensitive Attribute ist besondere Vorsicht in Bezug auf mögliche diskriminierende Ergebnisse geboten. Dazu zählen unter anderem Geschlecht, Alter, sozial benachteiligte Gruppen oder eine bestimmte Herkunft. So kann die Information der Herkunft zu unfairer Behandlung oder gar einem Ausschluss einer bestimmten Gruppe an Kund:innen führen. Damit das nicht passiert, muss das beim Erstellen des Modells entsprechend berücksichtigt werden. Dafür wurde ein Prozess definiert, welche klare Schritte bei möglichen kritischen Auswirkungen vorsieht.

Der zunehmende Einsatz von KI führt dazu, dass mehr und mehr Entscheidungen von Menschen an datengetriebene Algorithmen abgegeben werden. Als Leitbetrieb in Zentral- und Osteuropa ist es die Verantwortung von A1 sich diesem Thema anzunehmen. Das ist absolut im Sinne des digitalen Humanismus und sichert, dass Technologie den Menschen zugutekommt. Und damit meinen wir wirklich alle Menschen. Für A1 bedeutet Digitaler Humanismus das Mitdenken des Anderen.

Kritische Entscheidungen mit potenziellem Bias bei A1:

- Migration: Wer muss eine Anzahlung in welcher Höhe für das neue Handy zahlen?
- Klassische Kundenstruktur: Familienvater bestellt Festnetzleitung für den Haushalt bei A1. Angebote müssen aber alle Familienmitglieder umfassen.
- Geschlechterrollen: Wer verfügt über ein höheres Einkommen für höhere Tarife und höherwertige Hardware?

Die Frage, die wir uns als A1 zu Beginn stellen: Woher kommt der Bias? Nur, wenn man versteht, warum es einen gibt und warum KI Systems auf diesen zugreifen, kann man ihn auch korrigieren. Das nicht berücksichtigen birgt die Gefahr, dass die Ränder der Gauß'schen Normalverteilung runterfallen. Es ist unsere Verantwortung als Unternehmen gegenzusteuern, damit die KI nicht biased wird und sich eine Self-fulfilling Prophecy baut.

Auch A1 Mitarbeiter*innen werden mit neuartigen KI-Tools geschult. Bereits 2023 haben wir eine interne „A1 Chat GPT“ Anwendung gestartet, die unseren Kolleg*innen die Möglichkeit geben soll, in einem geschützten Rahmen auch mit sensiblen Daten die Vorzüge KI-generierter Texte zu nutzen.

Nicht zuletzt bedarf es einer offenen, praxisorientierten und konkreten Kommunikation im Unternehmen, am besten mit einem klaren Commitment des Top-Managements.

13.4 Fazit: Warum beschäftigt sich A1 mit Digitalem Humanismus?

Erstens, weil A1 als Leitbetrieb eine besondere Verantwortung gegenüber der Gesellschaft tragen. Die A1 Produkte prägen den Alltag von Millionen Menschen – deshalb ist es unsere Pflicht, ihre Entwicklung und Anwendung im Sinne gemeinsamer Werte mitzugestalten.

Zweitens, weil Vertrauen ein zentraler Markenwert für in diesem Unternehmen ist. In einer zunehmend digitalisierten Welt, in der Algorithmen Entscheidungen beeinflussen, möchte A1, dass die Kund*innen nachvollziehen können, wie und warum wir KI einsetzen. Digitaler Humanismus schafft dafür den ethischen und kommunikativen Rahmen.

Letztlich dient diese Haltung nicht nur gesellschaftlichen, sondern auch betriebswirtschaftlichen Zielen: Denn Vertrauen ist kein weicher Faktor – es ist die Grundlage für nachhaltige Kundenbeziehungen, langfristige Innovationsfähigkeit und wirtschaftlichen Erfolg. Wer den Menschen in den Mittelpunkt stellt, schafft nicht nur faire digitale Lösungen, sondern auch ein stabiles Fundament für unternehmerisches Wachstum in einer komplexen Zukunft.

Danksagung Danke an die A1 Kolleg:innen Michaela Barta-Müller, Simone Scholz, Christoph Ulbrich und das A1 ADA (Analytics, Data & AI)-Team für die inhaltlichen Inputs.

Dr. Michael Höfler studierte berufsbegleitend an der FH Marketing & Sales in Wien sowie an der Fachhochschule der Wirtschaft in Berlin und promovierte 2010 am Institut für Volkswirtschaftslehre an der Alpen-Adria Universität Klagenfurt. Er startete seine Berufslaufbahn 2004 beim damaligen PR-Marktführer PLEON Publico und war von 2008 bis 2017 zusätzlich Universitätslektor an der WU-Wien sowie an der FH-Wien. 2011 gründete er das Beratungsunternehmen pantarhei corporate advisors. 2018 verkaufte Höfler das Unternehmen an seine Mitarbeiter und Partner und ist seit 2019 Group Director Corporate Communications & Sustainability (ESG Österreich) sowie Konzernsprecher der A1 Telekom Austria Group. Er ist Gerichtssachverständiger für Medienwesen, Marketing; Öffentlichkeitsarbeit und Lobbying.

Teil III

Methoden, Tools und Standards

Methoden zur Umsetzung des Digitalen Humanismus

14

Praktische Ansätze für digitale Verantwortung und Ethik

Martin Giesswein

Zusammenfassung

Der Beitrag bietet eine umfassende Orientierungshilfe zur Umsetzung des Digitalen Humanismus in Organisationen. Es werden methodische Ansätze vorgestellt, die digitale Verantwortung und Ethik in Unternehmensstrukturen und IT-Systemen verankern. Der Fokus liegt auf allgemeinen Methoden zur Organisationsentwicklung, spezifischen Methoden für KI-Systeme und ganzheitlichen Ansätzen zur Gestaltung von IT-Systemen. Ziel ist es, dem Management praktische Werkzeuge zur Umsetzung der Werte des Digitalen Humanismus zu bieten.

Die Kunst der Ingenieursarbeit liegt darin, humane Werte im System digital umzusetzen.

14.1 Einleitung

Bei der Umsetzung des digitalen Humanismus steht das Management vor der zentralen Frage: Welche Methode, welcher Ansatz oder welche Checkliste soll angewendet werden? Dieses Kapitel bietet Orientierung, indem es eine Auswahl der praktikabelsten und bewährtesten Strategien vorstellt. Auch wenn keiner dieser Ansätze die vollständige Erreichung eines Idealzustandes garantieren kann, so liefern sie doch wertvolle Ansatzpunkte,

M. Giesswein (✉)
WU Executive Academy, Wien, Österreich
E-Mail: martin@giesswein.org

damit das Management den für das jeweilige Unternehmen optimalen Weg einschlagen kann.

Die Auswahl basiert auf dem „Atlas der Methoden“ der DigitalCity.Wien – einem Verein, der die Zusammenarbeit zwischen Digitalwirtschaft, Verwaltung und Wissenschaft in Wien fördert. Der Atlas umfasst mehr als 40 Methoden, von denen hier die am häufigsten verwendeten und relevantesten vorgestellt werden.

Die beschriebenen Ansätze sind in drei Kategorien unterteilt:

1. Allgemeine Vorgehensweisen für Organisations- und IT-Entwicklungen
2. Gesamtheitliche, KI-spezifische Vorgehensweisen
3. Vorgehensweisen für spezifische KI-Aspekte

Jede Beschreibung enthält auch einen Link zu weiterführenden Informationen.

14.2 Allgemeine Vorgehensweisen für Organisations- und IT-Entwicklungen

CDR Building Bloxx

Die CDR Building Bloxx sind ein praxisorientierter Rahmen, der Unternehmen, politische Institutionen, wissenschaftliche Organisationen und die Zivilgesellschaft dabei unterstützt, ein einheitliches Verständnis von Corporate Digital Responsibility (CDR) zu entwickeln und konkrete Maßnahmen umzusetzen. Das Framework bietet einen strukturierten Ansatz, um digitale Verantwortung nachhaltig in Unternehmenskultur, -struktur und -strategie zu verankern.

Es besteht aus allgemeinen und spezifischen Bausteinen: Die allgemeinen Building Bloxx definieren grundlegende Prinzipien und organisatorische Aspekte, während die spezifischen Bausteine zentrale Themenfelder der digitalen Verantwortung adressieren, wie den verantwortungsvollen Umgang mit Daten, den Beitrag digitaler Technologien zum Klimaschutz oder die Förderung sozialer Gerechtigkeit.

Die Anwendung des CDR Building Bloxx ermöglicht es Unternehmen, ihre Aktivitäten im Bereich der digitalen Verantwortung systematisch zu analysieren, zu priorisieren und weiterzuentwickeln. Gleichzeitig erleichtert das Framework Organisationen, die erstmals CDR-Strategien umsetzen wollen, den Einstieg und hilft Fortgeschrittenen, bestehende Maßnahmen zu optimieren. Durch die Standardisierung von Vorgehensweisen und die transparente Dokumentation von Fortschritten unterstützt es die Vergleichbarkeit und Weiterentwicklung von CDR-Praktiken in unterschiedlichen Branchen.

Treibende Kraft hinter der Methode ist der Bundesverband Digitale Wirtschaft (BVDW). In Deutschland unterstützt das Bundesministerium für Umwelt, Naturschutz, Reaktorsicherheit und Verbraucherschutz die CDR-Initiative. Zum Zeitpunkt der Erstellung dieses Artikels haben folgende Unternehmen CDR-Berichte veröffentlicht: BARMER, Deut-

sche Telekom AG, Deutsche Kreditbank AG, ING Deutschland, Otto Group, Telefónica Deutschland, Weleda AG, Zalando SE.

- Weitere Informationen: <https://www.cdr-building-bloxx.com/>.

Gesamtunternehmens-Vorgehen gemäß IÖW

Eine Studie des Instituts für Ökologische Wirtschaftsforschung hat 62 Unternehmen analysiert und Vorschläge für ein methodisch-empirisches Vorgehen entwickelt. Untersucht wird, wie das Management die digitale Verantwortung von Unternehmen grundlegend gestalten kann.

Im Fokus stehen Unternehmen als Nachfrager und Nutzer digitaler Dienstleistungen, nicht als deren Produzenten. Denn die Digitalisierung wird häufig weniger durch eigene Innovationen als durch Konsumententscheidungen für bestimmte digitale Angebote geprägt. Unternehmen stehen vor der Herausforderung, bewusst abzuwägen, auf welchen sozialen Voraussetzungen digitale Konsumoptionen basieren und welche ökologischen Konsequenzen sie nach sich ziehen.

Die Betrachtung umfasst sechs zentrale Handlungsfelder, die sowohl Managementpraktiken als auch Stakeholderbeziehungen von Unternehmen adressieren. Diese sind Unternehmensführung, Verantwortung gegenüber Mitarbeiter*innen, die ökologische Verantwortung eines Unternehmens, Produktverantwortung, Verantwortung in der Lieferkette und Corporate Digital Citizenship.

- Weitere Informationen: https://www.ioew.de/publikation/corporate_digital_responsibility.

Digital Impact Methode der WU Executive Academy

Zielgruppe dieser Methode der WU Executive Academy sind Unternehmen aller Branchen und Größen, die öffentliche Verwaltung sowie wissenschaftliche und zivilgesellschaftliche Organisationen. Die Digital Impact Methode dient der Planung und Umsetzung von Maßnahmen des Digitalen Humanismus im Unternehmen. Sie ermöglicht eine erste Annäherung von Organisationen an den Digitalen Humanismus.

Die Anwendung der Digital Impact Methode erfordert keine Änderung der Gesamtstrategie, sondern fokussiert auf die kurzfristige Erstellung einer organisationsindividuellen Leitlinie zum Digitalen Humanismus und die Umsetzung erster Maßnahmen zur Optimierung des Digitalen Impacts für alle Stakeholder*innen. Wenn bereits Initiativen (z. B. Nachhaltigkeitsberichterstattung) im Unternehmen existieren oder Managementsysteme eingeführt sind, gibt die Methode Anregungen, wie diese Initiativen und Systeme um Ziele und Maßnahmen des Digitalen Humanismus ergänzt werden können.

Die Schritte der Methode sind: Gemeinsames Verständnis, Analyse und Stakeholder-Map, Maßnahmendefinition, Priorisierung und Vermeidung von „Humanismwashing“, Leitbildformulierung, Review und laufende Vertiefung.

- Weitere Informationen: <https://humanism.digital/>.

Digitaler Ethik-Standard: Value-based Engineering: ISO/IEC/IEEE 24748-7000:2022

Value-based Engineering (VBE), wie im Standard ISO/IEC/IEEE 24748-7000 beschrieben, ist ein praktischer, bewährter und zertifizierbarer Ansatz für die Erstellung und den Betrieb ethischer IT-Systeme. Er bietet Organisationen einen klar strukturierten und transparenten Prozess, um Wertprinzipien in Anforderungen an IT-Systeme und organisatorische Rahmenprozesse zu übersetzen. Dies inkludiert auch KI. VBE stellt sicher, dass Produktmanagement und technische Einheiten gemeinsam an den Werten der Stakeholder arbeiten. Der Standard definiert Prozesse, die es Organisationen ermöglichen, ethische Überlegungen systematisch in allen Phasen der IT-Konzeptentwicklung und des Systemdesigns zu berücksichtigen.

Ein Schwerpunkt liegt dabei auf der transparenten Kommunikation zwischen Management, Ingenieuren und ausgewählten Stakeholdern, um Werte zu identifizieren, zu priorisieren und ihre Umsetzung im Systemdesign nachvollziehbar zu machen. Dies beinhaltet unter anderem die Rückverfolgbarkeit von Werten durch operative Konzepte, Wertvorschläge und deren Umsetzung in Designentscheidungen; z. B. auch zur Vorlage bei Rechtskonformitätsprüfungen. Das Value-based Engineering (VBE) mit ISO/IEC/IEEE 24748-7000 ist für Organisationen aller Größen und Branchen anwendbar und kann flexibel in bestehende Lebenszyklusmodelle integriert werden.

Value-based Engineering wurde in verschiedenen Projekten ganzheitlich oder in Teilaspekten eingesetzt: KI-Einsatz bei ibis acam, Yoma-Projekt der UNICEF, Dialoge des AMS-Berufsinformat, Mein Stadt Wien Konto, datahub.tirol und in der Automobilindustrie.

Austrian Standards und die VBE Academy (in Kooperation mit IEEE) bieten eine nationale und internationale Personenzertifizierung zum VBE Ambassador an (https://www.austrian-standards.at/de/produkte-loesungen/standards-anwenden/zertifizierung/personenzertifizierung/value_based_eng_ambassador). Der „VBE Ambassador“ und im nächsten Schritt „Value Lead“ ist in der Lage digitale Ethik ebenso wie KI-Ethik holistisch in der Organisation zu adressieren und eine ethische, Werte-basierte Governance der Systeme zu gewährleisten.

► Weiterführende Informationen: <https://www.vbe.academy/>.

14.3 Gesamtheitliche, KI-spezifische Vorgehensweisen

KI-Standard ISO/IEC 42001:2023

Während der oben genannte Value-based Engineering-Ansatz eine starke inhaltliche Wertebasis für jedes IT-System einschließlich KI-Systemen bietet, ist die ISO/IEC 42001:2023 stark auf Überwachung und Bewertung (KI Governance) ausgerichtet. Beide Normen ergänzen sich ideal und können gemeinsam angewendet werden. Die Govern-

ance der zu adressierenden KI-, „Issues“ in ISO/IEC 42001:2023 entsprechen den Werten in ISO/IEC/IEEE 24748-7000.

ISO/IEC 42001:2023 ist ein internationaler Standard für KI-Managementsysteme, der Organisationen einen strukturierten Ansatz für die verantwortungsvolle Nutzung und Verwaltung von KI-Systemen bietet. Der im Dezember 2023 veröffentlichte Standard zielt darauf ab, vertrauenswürdige und effektive Managementpraktiken für KI zu etablieren, die ethische, rechtliche und gesellschaftliche Anforderungen berücksichtigen.

Ein zentrales Element des 42001-Standards ist der risikobasierte Ansatz, der Organisationen dazu verpflichtet, Risiken während des gesamten Lebenszyklus von KI-Systemen zu identifizieren, zu bewerten und zu managen. Um eine kontinuierliche Verbesserung zu gewährleisten, fordert der Standard regelmäßige Überprüfungen und Optimierungen von KI-Systemen und -Prozessen. Darüber hinaus berücksichtigt ISO/IEC 42001:2023 die langfristigen Auswirkungen von KI-Anwendungen auf Umwelt, Gesellschaft und Wirtschaft und integriert Nachhaltigkeitsaspekte in die Systemgestaltung. Der Standard ist kompatibel mit bestehenden Managementsystemen wie ISO 9001 (Qualitätsmanagement) und ISO/IEC 27001 (Informationssicherheit).

Die Zertifizierung nach ISO/IEC 42001:2023 bietet zahlreiche Vorteile. Sie stärkt das Vertrauen von Kunden, Partnern und der Öffentlichkeit, indem sie signalisiert, dass die Organisation KI-Systeme verantwortungsvoll und sicher einsetzt. Gleichzeitig ermöglicht sie die frühzeitige Erkennung und Minimierung potenzieller Risiken und hilft bei der Einhaltung rechtlicher und regulatorischer Anforderungen, wie z. B. dem zu erwartenden EU AI Act. Der strukturierte Rahmen ermöglicht es Unternehmen, Innovationen voranzutreiben, ohne dabei Sicherheits- und Ethikaspekte zu vernachlässigen.

► Weitere Informationen: <https://www.iso.org/standard/81230.html>.

KI-Zertifizierungsregime IEEE CertifAIEd™

Während die oben genannte ISO/IEC 42001:2023 einen umfassenden Managementrahmen für KI-Managementsysteme bietet, zielt IEEE CertifAIEd™ auf die ethische Zertifizierung mehrerer Schlüsselbereiche der KI-Ethik ab, darunter KI-Transparenz, -Privatsphäre, -Verantwortlichkeit und Minimierung algorithmischer Verzerrungen. Es handelt sich somit um ein Zertifizierungsprogramm, das sich auf die Förderung und Bewertung der zentralen technischen Wertprinzipien fokussiert, wie sie auch in der KI Verordnung angesprochen werden.

In der Anwendung ist IEEE CertifAIEd™ einem Value-based Engineering (VBE) Projekt idealerweise nachgelagert. Wenn etwa ein VBE-Projekt herausfindet, dass algorithmische Verzerrungen für das betrachtete KI-System von Bedeutung sind, stellt IEEE CertifAIEd™ ein Zertifizierungsprogramm zur Verfügung, das diese Eigenschaft nachweisbar adressiert. Mit aktuellen Ansätzen analysiert das Programm Treiber und Hindernisse, die die Entwicklung der entsprechenden ethischen Qualität(en) im jeweiligen KI-System beeinflussen können.

IEEE bietet die Ausbildung von CertifAIED™ Assessoren an. Dieses Training qualifiziert Fachleute, als autorisierte Gutachter zu agieren. Die Teilnehmenden erwerben ein fundiertes Verständnis der benannten technischen Kernprinzipien für die KI-Ethik sowie des Zertifizierungsprozesses und können nach Abschluss der Ausbildung KI-Produkte, Dienstleistungen und Systeme auf diese oft auch vom Gesetzgeber geforderten Werte durchführen.

- Weitere Informationen: <https://standards.ieee.org/products-programs/icap/ieee-certifaied/>.

Assessment List for Trustworthy Artificial Intelligence (ALTAI)

Die Assessment List for Trustworthy Artificial Intelligence (ALTAI) ist ein praktisches Selbstbewertungsinstrument, das von der High-Level Expert Group on AI (AI HLEG) der Europäischen Kommission entwickelt wurde. Ziel von ALTAI ist es, Unternehmen und Organisationen bei der Entwicklung und Implementierung vertrauenswürdiger KI-Systeme durch die praktische Anwendung ethischer Prinzipien zu unterstützen.

Das Tool basiert auf den sieben Schlüsselanforderungen für vertrauenswürdige KI, die in den Ethik-Leitlinien der AI HLEG definiert sind: menschliche Kontrolle und Überwachung, technische Robustheit und Sicherheit, Datenschutz und Datenverwaltung, Transparenz, Diversität, Nichtdiskriminierung und Fairness, soziales und ökologisches Wohlergehen sowie Rechenschaftspflicht. Diese Anforderungen bilden die Grundlage einer dynamischen Checkliste, die sowohl als PDF-Dokument als auch als webbasiertes Tool zur Verfügung steht.

Die Anwendung von ALTAI umfasst mehrere Schritte, darunter die Erstellung individueller Bewertungslisten, die Durchführung der Bewertung – idealerweise durch ein multidisziplinäres Team – sowie die Identifizierung von Risiken und die Umsetzung geeigneter Maßnahmen.

ALTAI bietet Organisationen einen praktischen Leitfaden, um die Vertrauenswürdigkeit ihrer KI-Systeme systematisch zu bewerten und deren Übereinstimmung mit ethischen Prinzipien sicherzustellen. Durch seine flexible und anpassbare Methodik ermöglicht ALTAI eine umfassende Bewertung ethischer Aspekte und unterstützt die Umsetzung konkreter Maßnahmen zur Risikominimierung und Vertrauensbildung.

- Weitere Informationen: <https://altai.insight-centre.org/>.

KI-Standard VDE SPEC 90012

Die in Deutschland vom Verband der Elektrotechnik, Elektronik und Informationstechnik entwickelte Spezifikation VDE SPEC 90012 dient der Beschreibung der Vertrauenswürdigkeit von KI-Systemen. Sie wurde entwickelt, um eine Methode zur Charakterisierung sozio-technischer Merkmale von KI-Systemen und -Anwendungen bereitzustellen, insbesondere für Produkte, die ein besonders hohes Maß an Vertrauen erfordern.

Kern der Spezifikation ist das VCIO-Modell (Values, Criteria, Indicators, Observables), das eine systematische Bewertung der Vertrauenswürdigkeit von KI-Systemen ermöglicht.

Das Modell zielt darauf ab, die Einhaltung spezifischer Werte durch KI-Systeme messbar zu machen und bietet damit eine Grundlage für die Entwicklung eines möglichen Vertrauenslabels für KI-Produkte. Die Spezifikation adressiert fünf zentrale Werte: Transparenz, Verantwortlichkeit, Datenschutz, Fairness und Zuverlässigkeit.

Die VDE SPEC 90012 wurde in enger Zusammenarbeit mit Bertelsmann und Industriepartnern wie Bosch, Siemens und SAP sowie wissenschaftlichen Einrichtungen entwickelt. Dabei wurden bestehende Konzepte und EU-Vorgaben zur KI-Ethik berücksichtigt, um sicherzustellen, dass die Spezifikation mit aktuellen Standards und rechtlichen Anforderungen harmonisiert.

- Weitere Informationen: <https://www.vde.com/de/arbeitsfelder/standards/spec/vde-spec-veroeffentlichungen>.

Ethical Impact Assessment (EIA) der UNESCO

Das UNESCO Ethical Impact Assessment (EIA) ist ein Instrument zur Bewertung der ethischen Aspekte der Künstlichen Intelligenz. Es wurde entwickelt, um sicherzustellen, dass spezifische Algorithmen mit den Werten, Prinzipien und Richtlinien der UNESCO-Empfehlung zur Ethik der KI übereinstimmen. Gleichzeitig soll es die Transparenz fördern, indem Informationen über KI-Systeme und deren Entwicklung öffentlich zugänglich gemacht werden.

Das EIA deckt den gesamten Lebenszyklus von KI-Systemen ab und berücksichtigt sowohl Ex-ante- als auch Ex-post-Anforderungen. Das Instrument legt besonderen Wert auf die Qualität und Repräsentativität der Daten, die Diversität der Entwicklungsteams, die Robustheit und Transparenz der Algorithmen, die Verifizierbarkeit sowie die Möglichkeit, Kontrollpunkte während des Entwicklungsprozesses zu implementieren.

Primär richtet sich das EIA an Beschaffer von KI-Systemen, kann jedoch ebenso von Entwicklern im öffentlichen und privaten Sektor genutzt werden. Es besteht aus zwei Hauptteilen: dem Scoping, das ein grundlegendes Verständnis des Systems und erste Fragen umfasst, und der Umsetzung der Prinzipien der UNESCO-Empfehlung. Ziel ist es, ethische Richtlinien in die Praxis zu integrieren und die potenziellen Auswirkungen von KI-Systemen umfassend zu bewerten.

Die EIA ist Teil eines umfassenderen Implementierungsplans für die UNESCO-Empfehlung zur Ethik der KI. Es ergänzt die Readiness Assessment Methodology (RAM), die Regierungen dabei unterstützt, ihre rechtlichen Rahmenbedingungen, Richtlinien und Institutionen im Umgang mit KI-Risiken zu bewerten. Durch diesen integrativen Ansatz bietet das EIA Organisationen und Behörden eine wertvolle Orientierungshilfe, um KI-Systeme verantwortungsvoll und transparent zu gestalten.

- Weitere Informationen: <https://www.unesco.org/ethics-ai/en/eia>.

NIST AI Risk Management

Das NIST AI Risk Management Framework (AI RMF) wurde vom National Institute of Standards and Technology (NIST) entwickelt, einer US-amerikanischen Behörde, die für

die Entwicklung von Standards und Technologien verantwortlich ist. Dieses Framework, zusammen mit dem ergänzenden AI RMF Playbook, dient als freiwillig zu verwendende Ressource, um das Risikomanagement bei der Entwicklung und Nutzung von KI-Systemen zu verbessern und Organisationen bei der verantwortungsvollen und sicheren Implementierung von KI zu unterstützen.

Das Framework basiert auf vier Kernfunktionen, die den gesamten Lebenszyklus von KI-Systemen abdecken. Die erste Funktion, „Govern“, zielt auf den Aufbau von Governance-Strukturen ab und fördert eine Kultur des KI-Risikomanagements. „Map“ ermöglicht die Identifizierung und Bewertung von Risiken im Zusammenhang mit KI-Systemen. Mit „Measure“ wird die Analyse und Nachverfolgung der Risikoexposition ermöglicht, während „Manage“ die Implementierung und Pflege von Kontrollen zur Risikominderung sicherstellt.

Das AI RMF verfolgt das Ziel, eine vertrauenswürdige und verantwortungsvolle Entwicklung von KI-Systemen zu fördern. Es berücksichtigt Aspekte wie Validität, Zuverlässigkeit, Verantwortlichkeit, Transparenz, Sicherheit und Fairness. Das Framework ist anpassungsfähig gestaltet, um den unterschiedlichen Anforderungen von Organisationen und Anwendungsfällen gerecht zu werden.

Das ergänzende AI RMF Playbook bietet konkrete Handlungsempfehlungen für die Umsetzung der Kernfunktionen des Frameworks. Es stellt keine starren Checklisten dar, sondern flexible und anpassbare Anwendungsmöglichkeiten. Regelmäßige Aktualisierungen des Playbooks basieren auf technologischen Fortschritten und Nutzerfeedback, um sicherzustellen, dass es stets aktuell und relevant bleibt.

► Weitere Informationen: <https://www.nist.gov/itl/ai-risk-management-framework>.

14.4 Vorgehensweisen für spezifische KI-Aspekte

Algorithmic Bias Risk Radar (ABRRA)

Der Algorithmic Bias Risk Radar (ABRRA) ist ein Projekt des Startups leiwand.ai in Zusammenarbeit mit der TU Wien, das sich der Vorabeeschätzung von Bias-Risiken in künstlichen Intelligenz-Systemen widmet. Die Technologie basiert auf der Kombination von Knowledge Graphs, statistischen Verfahren und Machine Learning-Methoden. Mit Hilfe dieser Ansätze ermöglicht der ABRRA umfassende Ähnlichkeits- und Musteranalysen, die dazu beitragen, potenzielle Diskriminierungs- und Bias-Risiken in KI-Systemen frühzeitig zu identifizieren. Darüber hinaus unterstützt das Tool die Durchführung von Grundrechtsverträglichkeitsprüfungen, die besonders für Hochrisiko-Anwendungen relevant sind.

Das Projekt zielt darauf ab, negative Auswirkungen von KI-Systemen in sensiblen Anwendungsbereichen wie Personalwesen, Gesundheitswesen, Finanzsektor, öffentlicher Verwaltung und Bildung zu erkennen, zu analysieren und zu mindern. Eine zentrale Rolle

spielt dabei die Entwicklung einer Experten-Datenbank, die umfassende Informationen zu Bias-Risiken bereitstellt und so eine fundierte Bewertung ermöglicht.

- Weitere Informationen: <https://www.leiwand.ai/projects/the-algorithmic-risk-radar-abrra>.

Fundamental Rights Impact Assessment (FRIA)

Der AI Act sieht für Hochrisiko-KI-Systeme die Durchführung einer Fundamental Rights Impact Assessment (FRIA) vor. Diese Folgenabschätzung für Grundrechte ist verpflichtend und muss vor dem ersten Einsatz eines Hochrisiko-KI-Systems erfolgen. Ziel der FRIA ist es, potenzielle Schäden in Bezug auf die Grundrechte von Einzelpersonen zu identifizieren und geeignete Maßnahmen zu deren Minderung zu treffen.

Hochrisiko-KI-Systeme umfassen Anwendungen in besonders sensiblen Bereichen, die erhebliche Auswirkungen auf Einzelpersonen und die Gesellschaft haben können. Dazu zählen Systeme zur biometrischen Identifizierung und Kategorisierung von Personen sowie solche, die in kritischen Infrastrukturen wie dem Verkehr eingesetzt werden. Weitere Beispiele sind KI-Systeme, die in der Bildung und Berufsbildung, im Beschäftigungs- und Personalmanagement oder bei der Berufserlaubnis eingesetzt werden. Auch Anwendungen, die den Zugang zu wesentlichen privaten und öffentlichen Dienstleistungen wie der Kreditbewertung ermöglichen, sowie Systeme in den Bereichen Strafverfolgung, Migrations-, Asyl- und Grenzkontrollmanagement, Justizverwaltung und demokratische Prozesse, gelten als hochriskant.

Die FRIA umfasst mehrere zentrale Inhalte. Dazu gehören die Beschreibung der Prozesse, in denen das System eingesetzt wird, sowie der Zeitraum und die Häufigkeit der geplanten Nutzung. Zudem werden die Kategorien der betroffenen Personen und Gruppen sowie spezifische Schadensrisiken für diese identifiziert. Ergänzt wird dies durch eine Darstellung der Maßnahmen zur menschlichen Aufsicht und der vorgesehenen Reaktionen im Falle der Verwirklichung von Risiken.

Die Verpflichtung zur Durchführung der FRIA betrifft öffentliche Einrichtungen, private Unternehmen, die öffentliche Dienstleistungen erbringen, sowie Betreiber von Hochrisiko-KI-Systemen in den genannten Bereichen. Die FRIA muss nicht nur vor dem erstmaligen Einsatz eines Hochrisiko-KI-Systems erfolgen, sondern bei Änderungen relevanter Faktoren auch aktualisiert werden.

Die Ergebnisse der FRIA müssen der zuständigen Marktüberwachungsbehörde gemeldet werden, wobei hierfür ein standardisiertes Formular verwendet werden muss, das vom AI Office bereitgestellt wird. Darüber hinaus ergänzt die FRIA bei Bedarf die Datenschutz-Folgenabschätzung (DPIA), die gemäß DSGVO erforderlich sein kann.

Lösungen als Compliance- und Ethik-Unterstützung

Neben Standards und Vorgehensmodellen gibt es auch immer mehr Lösungen, die es Unternehmen erleichtern, die Compliance mit dem AI-Act inklusive der ethischen Fragen,

Risiken und Mitigierungen durchzuführen und laufend zu monitoren. Dabei handelt es sich oft um eine Kombination von Beratung, formalisierten Vorgehen, einer technischen Unterstützung, die Speicherung, Berichte und Revisionen ermöglicht.

Beispielsweise bietet pAIper.one eine solche Lösungen in der folgenden Struktur an: Die pAIper.one AI-ACT Compliance Plattform hilft Unternehmen, KI-Anwendungen rechtssicher zu entwickeln und zu nutzen. Sie prüft die Anwendungen auf rechtliche und ethische Risiken, zeigt notwendige Dokumentationspflichten auf und liefert eine Risikobewertung. Die Plattform unterstützt bei der Risikominimierung, der Umsetzung von Maßnahmen zur Einhaltung des EU AI Acts und der DSGVO sowie bei der Dokumentation von Prüfprozessen. Zudem sorgt der Compliance Tracker für Transparenz und Nachvollziehbarkeit. Unternehmen bleiben durch kontinuierliche Überwachung rechtskonform und profitieren von Schulungen und einem Wissensaustausch in der AI Education Community. Die Governance Plattform bietet zudem eine Übersicht über alle KI-Anwendungen im Unternehmen und fördert Innovationen durch einen strukturierten Demand-Prozess zur Bewertung neuer Anwendungsfälle.

14.5 Weitere Methoden und Vertiefung

Nach diesem Überblick über Vorgehensweisen zur Umsetzung des Digitalen Humanismus bieten die folgenden Kapitel dieses Buchabschnitts eine Vertiefung und weiterführende methodische Ansätze. Svenja Schröder und Florian Wurzer zeigen den Einsatz von Value-based Engineering bei der Firma ibis acam. Florian Wagner diskutiert am Beispiel einer Elektromobilitäts-App das Zusammenspiel von VBE und System-Theoretic Process Analysis (STPA). Karin Dietl und Simon Lucas beschreiben den konkreten aktuellen Stand bei der Firma Merck zu den Themen Digitale Ethik und CDR. Wolfgang Kabelka informiert über die Checkliste für vertrauenswürdige KI, die vom Bundesrechenzentrum der Republik Österreich entwickelt und eingesetzt wird. Armin Preis erläutert die Kybernetik als Methode zur Umsetzung des Digitalen Humanismus.

Weiterführende Literatur

- Austrian Standards und VBE Academy (n.d.). Value-based Engineering: ISO/IEC/IEEE 24748-7000:2022. https://www.austrian-standards.at/en/products-solutions/apply-standards/certification/personal-certification/value_based_eng_ambassador. Zugegriffen: 9. Juni 2025.
- Bundesverband Digitale Wirtschaft (BVDW) (n.d.). CDR Building Bloxx. <https://www.cdr-building-bloxx.com/cdr-building-bloxx/?lang=en>. Zugegriffen: 3. Juni 2025.
- High-Level Expert Group on AI (AI HLEG) (n.d.). Assessment List for Trustworthy Artificial Intelligence (ALTAI). <https://op.europa.eu/en/publication-detail/-/publication/73552fcd-f7c2-11ea-991b-01aa75ed71a1>. Zugegriffen: 15. Juni 2025.
- IEEE Standards Association (n.d.). IEEE CertifAIEd™. <https://standards.ieee.org/products-programs/icap/ieee-certifaied/professional-certification/>. Zugegriffen: 13. Juni 2025.

- Institut für Ökologische Wirtschaftsforschung (IÖW) (n.d.). Corporate digital responsibility. https://www.ioew.de/en/publication/corporate_digital_responsibility. Zugegriffen: 5. Juni 2025.
- International Organization for Standardization (ISO) (2023). ISO/IEC 42001:2023. <https://www.iso.org/standard/42001>. Zugegriffen: 11. Juni 2025.
- leiwand.ai, & TU Wien (n.d.). Algorithmic Bias Risk Radar (ABRRA). <https://www.leiwand.ai/projects/the-algorithmic-risk-radar-abrra>. Zugegriffen: 23. Dezember 2025.
- National Institute of Standards and Technology (NIST) (n.d.). AI Risk Management Framework. <https://www.nist.gov/itl/ai-risk-management-framework>. Zugegriffen: 23. Dezember 2025.
- UNESCO (n.d.). Ethical Impact Assessment (EIA). <https://www.unesco.org/ethics-ai/en/eia>. Zugegriffen: 23. Dezember 2025.
- Verband der Elektrotechnik, Elektronik und Informationstechnik (VDE) VDE SPEC 90012 (n.d.). https://www.researchgate.net/publication/360167343_VDE_SPEC_90012_V10_-_VCIO_based_description_of_systems_for_AI_trustworthiness_characterisation. Zugegriffen: 17. Juni 2025.
- WU Executive Academy Digital Impact Methode (n.d.). <https://www.executiveacademy.at/en/knowledge/digitization/are-you-ready-for-digital-transformation>. Zugegriffen: 7. Juni 2025.

Martin Giesswein ist Fakultätsmitglied der WU Executive Academy mit den Schwerpunkten KI, Digitalökonomie und Leadership. Er ist Autor und Herausgeber mehrerer Bücher zu den Themen Künstliche Intelligenz, Digital Leadership und Zukunftsgestaltung. 2025 wurde er mit einem ersten Preis des Awards „Digitaler Humanismus in der Praxis“ ausgezeichnet. Er ist als AI Ethics Advisor Mitglied des Advisory Committee im Projekt DRAIGON. Martin Giesswein ist einer der Mitinitiatoren der DigitalCity.Wien, einer seit über 10 Jahren bestehenden Kooperation der Digitalwirtschaft mit der Stadt Wien. Zuvor war er CEO und General Manager in der IKT-Branche.

Value Based Engineering und die Umsetzung bei ibis acam

15

Erfahrungen aus einem der ersten Projekte mit dem Value Based Engineering Ansatz

Svenja Schröder und Florian Wurzer

Zusammenfassung

Der Digitale Humanismus ist in aller Munde, aber wie kann eine konkrete Umsetzung in Digitalisierungsprojekten aussehen? Der Standard ISO/IEC/IEEE 24748-7000:2022 bzw. Value Based Engineering liefern als Methode eine Antwort: Durch eine systematische Einbeziehung aller Stakeholder und derer ethischen Werte können diese Werte als Anforderungen in die Softwareentwicklung einfließen. Die Vorteile sind vielfältig: nachhaltige Geschäftsmodelle, Entdeckung ungenutzter Innovationen und risikobasiertes Systemdesign. In diesem Beitrag soll die zugrundeliegende Methode erklärt und anhand eines Softwareentwicklungsprojektes bei der ibis acam veranschaulicht werden.

„Handle so, dass du die Menschheit, sowohl in deiner Person, als in der Person eines jeden andern, jederzeit zugleich als Zweck, niemals bloß als Mittel brauchst.“
(Immanuel Kant)

S. Schröder (✉) · F. Wurzer
msg Plaut, Wien, Österreich
E-Mail: svenja.schroeder@msg-plaut.com

F. Wurzer
E-Mail: florian.wurzer@msg-plaut.com

© Der/die Autor(en), exklusiv lizenziert an Springer Fachmedien Wiesbaden GmbH, ein Teil von Springer Nature 2026, G. Krause, M. Giesswein (Hrsg.), *Management Guide Digitaler Humanismus*, Digitaler Humanismus in Wirtschaft, Wissenschaft und Gesellschaft, https://doi.org/10.1007/978-3-658-49869-6_15

159

15.1 Einleitung: Warum brauchen wir ethische Softwareentwicklung?

Hören wir den Begriff „Value“ im Zusammenhang mit Value Based Engineering (VBE), assoziieren viele zunächst „wirtschaftlichen Wert“ oder „Geldwert“. Doch VBE geht über finanzielle Aspekte hinaus: Mit „Values“ sind hier ethisch-moralische Werte gemeint, wie Freiheit, Selbstbestimmung und Gesundheit. Im Zentrum von VBE steht die Einbindung solcher ethischen Werte in die Softwareentwicklung. Dies ist entscheidend, denn ein nachhaltiger Erfolg in der digitalen Transformation verlangt heute nicht nur ökonomische, sondern auch ethische Perspektiven.

Entwickelt von Prof. Dr. Sarah Spiekermann-Hoff, basiert VBE auf dem international anerkannten Standard ISO/IEC/IEEE 24748-7000:2022: „Standard Model Process For Addressing Ethical Concerns During System Design“ (ISO/IEC/IEEE, 2022). Dieser Standard und die VBE-Methode sind in weiten Teilen deckungsgleich; jedoch legt VBE einen besonderen Schwerpunkt auf ein tiefes philosophisches Verständnis, während der Standard die Integration ethischer Prinzipien strukturiert in den Softwareentwicklungsprozess verankert.

Nach einem fünfjährigen Entwicklungsprozess und der Expertise von über 140 Fachleuten wurde der Standard im Jahr 2021 veröffentlicht (IEEE, 2021). Die Arbeitsgruppe stellte sich dabei zentrale Fragen: Wie gestalten wir Software ethisch, um sie zukunftssicher und zugleich innovativ zu halten? Welche ethischen Prinzipien sollten wir einbeziehen, um bestmögliche Lösungen für Endanwender*innen zu entwickeln?

Dieser Artikel liefert Ihnen einen klaren Überblick über die Methoden des ISO/IEC/IEEE 24748-7000 und des Value Based Engineering. Anhand eines Praxisbeispiels verdeutlichen wir, wie Manager*innen den praktischen Nutzen dieser Ansätze für eine wertorientierte und nachhaltige Geschäftsentwicklung ausschöpfen können.

Takeaways

- Value Based Engineering ermöglicht die Entwicklung von Softwareprodukten, die ethische Werte aktiv berücksichtigen.
- Die Methode setzt den Digitalen Humanismus in Digitalisierungsprojekten um.
- Digitale Geschäftsmodelle, die mit VBE durchleuchtet werden, sind langfristig zukunftssicher, innovativ und mit einem geringeren Risiko behaftet.

15.2 Methode: Wie wende ich VBE in meinen Projekten an?

Value Based Engineering (VBE) beginnt bereits mit der ersten Systembeschreibung – noch bevor die erste Codezeile geschrieben wird. Als Teil des Requirements Engineering werden in einem frühen Stadium Workshops mit allen relevanten Stakeholdern durchgeführt. In diesen Workshops werden die ethischen Werte identifiziert, die für das Projekt von zentraler Bedeutung sind.

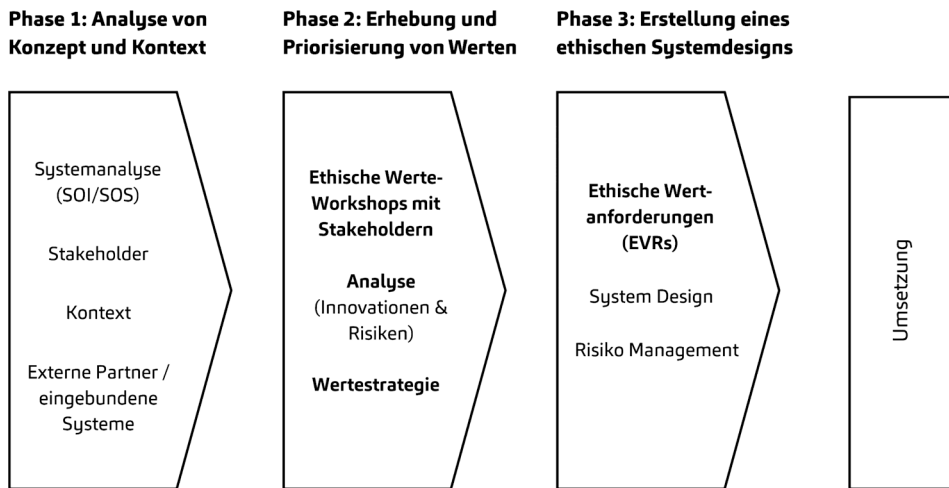


Abb. 15.1 VBE Prozess

Die eigens geschaffene Rolle eines*r „Value Lead“ leitet diesen Prozess an und stellt sicher, dass die gesammelten Werte in eine klare Wertestrategie einfließen. Die Rolle des Value Lead begleitet im Nachgang ebenfalls den Entwicklungsprozess, um sicherzustellen, dass relevante ethische Werte durchgehend berücksichtigt und integriert werden.

Im Folgenden finden Sie eine Übersicht der einzelnen Schritte im VBE-Prozess (s. Abb. 15.1).

15.2.1 Phase 1: Analyse von Konzept und Kontext

In der ersten Phase wird ein umfassendes Verständnis des geplanten Systems entwickelt. Dabei werden folgende zentrale Komponenten identifiziert und analysiert:

- **System of Interest (SOI):** Das Hauptsystem, das im Fokus des Projekts steht.
- **System of Systems (SOS):** Die technische Systemumgebung, in die das SOI eingebettet ist.
- **Systemkomponenten:** Alle Elemente, die innerhalb des Systems oder mit dem System direkt interagieren.
- **Partnerunternehmen:** Externe Firmen, die an der Entwicklung oder an der Bereitstellung der Lösung beteiligt sind.
- **Systemkontext:** Die Umgebung, in der das System später genutzt wird.
- **Stakeholder, insbesondere Endnutzer*innen:** Alle relevanten Akteure, die vom System direkt oder indirekt betroffen sind.
- **Geschäftsmodell:** Das zugrunde liegende Geschäftsmodell, das das System stützt.

Die Dokumentation dieser Elemente erfolgt maßgeschneidert für das Projekt. Hierbei können verschiedene Methoden aus dem Requirements Engineering und der Business Analyse eingesetzt werden:

- **Anforderungsdokumente oder Backlog:** Übersicht der geplanten Funktionen und Features.
- **Use-Case-Szenarien:** Beschreibungen typischer Anwendungsfälle in schriftlicher Form.
- **Visuelle Modellierung:** Einsatz von UML, BPMN oder anderen Modellierungssprachen.
- **Business Model Canvas:** Eine Visualisierung des Geschäftsmodells.
- **Stakeholder-Liste:** Eine detaillierte Liste aller relevanten Stakeholder.

Takeaways

In Phase 1 erfolgt eine detaillierte Beschreibung des geplanten Systems.

Eine umfassende Stakeholder-Liste wird erstellt.

15.2.2 Phase 2: Erhebung und Priorisierung von Werten

In der zweiten Phase von VBE steht die Ermittlung ethischer Werte im Mittelpunkt, gefolgt von der Entwicklung einer klaren Wertestrategie.

Die Rolle des Value Lead

Um den Prozess zu leiten und das Team sowie die Stakeholder sicher durch die ethische Analyse zu begleiten, gibt es die Rolle des „Value Lead“. Diese Person hat eine spezielle Ausbildung in digitaler Ethik und ist in der Lage, souverän mit philosophischen Fragestellungen und ethischen Werten umzugehen.

Werteworkshops mit Stakeholdern

Die ethischen Werte werden gemeinsam mit den Stakeholdern erhoben (Tab. 15.1) und erarbeitet. Dazu finden drei aufeinander abgestimmte Workshops statt, in denen Stakeholder-Vertreter*innen anhand philosophischer Ansätze relevante ethische Werte identifizieren. Diese Workshops basieren auf drei ethischen Disziplinen, die jeweils unterschiedliche Perspektiven auf ethische Werte eröffnen und somit eine möglichst vollständige Werteerhebung gewährleisten. Dabei dient eine Workshop-Checkliste als Leitfaden für die strukturierte Durchführung.

Die Teilnehmer*innen erhalten eine Systembeschreibung und analysieren daraufhin die wesentlichen Merkmale des Systems. Anschließend sammeln sie die ethischen Werte, die sie für besonders relevant für die Stakeholder-Gruppen halten. Diese strukturierte Herangehensweise unterstützt die Teilnehmerinnen dabei, fundierte Werte zu identifizieren, die

Tab. 15.1 Leitfragen für den Workshop

Philosophische Perspektive	Fragen
Utilitarismus	Welche guten und schlechten Konsequenzen würde das System für die Stakeholder haben?
Tugendethik	Was sind die positiven und negativen Implikationen für den Charakter und/oder die Persönlichkeit der Stakeholder?
Pflichtethik	Laut Projektteam, welche persönlichen Maximen oder Wertprioritäten werden durch das System beeinträchtigt, die das Projektteam für wichtig hält?

das System aus ihrer Perspektive prägen. Es wird auf die Beschreibung des Effekts mit Bezug auf hinderlich oder förderlich für den Wert hingewiesen.

Wertestrategie

Im Anschluss an die Workshops werden die gesammelten Werte strukturiert und zu einer übergreifenden Wertestrategie zusammengeführt. Dazu werden Kernwerte aus den Workshopergebnissen extrahiert, sortiert und priorisiert. Ergänzend in die Analyse fließen zusätzliche Wertequellen ein:

- **Unternehmenswerte:** Orientierung an den bereits definierten Werten des Unternehmens.
- **Allgemeingültige Werte:** Berücksichtigung universeller Werte, wie etwa der Europäischen Werte (Europäische Union, 2024).
- **Branchenspezifische Werte:** Einbeziehung von relevanten Werten, etwa im Gesundheitsbereich.

Auf Basis aller gesammelten Werte entsteht eine individuelle Wertestrategie für das zukünftige System. Dabei werden, ähnlich wie im Anforderungsmanagement, manche Werte höher priorisiert, um gezielt in die Software einfließen zu können. Abschließend erfolgt eine Abstimmung der Wertestrategie mit dem Management, um eine durchgängige und strategiekonforme Umsetzung sicherzustellen.

15.2.3 Phase 3: Erstellung eines ethischen Systemdesigns

In dieser Phase werden die erhobenen Werte und ihrer Effekte auf das System in konkrete ethische Wertanforderungen (EVR) an das System und an die Organisation überführt. Warum brauchen wir zusätzlich zu den technischen auch organisatorische Anforderungen? Da technische Systeme stets in einen realen Kontext eingebettet sind, muss das ethische Systemdesign nicht nur technische, sondern auch organisatorische Anforderungen (an die Organisation, an Verantwortliche, an den Kontext, usw.) enthalten, falls dies zur Umsetzung der Werte notwendig ist.

Ethische Wertanforderungen

- sind nach den Prinzipien des Requirements Engineering nachvollziehbar und lassen sich eindeutig auf den zugrunde liegenden Wert zurückführen.
- setzen die erarbeiteten Werte praxisnah und direkt im System um.
- fungieren als Verbindung zwischen philosophischen Werten und ihrer technischen Realisierung.
- werden gemäß der Wertestrategie priorisiert, um eine fokussierte Umsetzung sicherzustellen.

Zusätzlich werden die Anforderungen auf potenzielle Risiken überprüft. Wo erforderlich, werden Gegenmaßnahmen als zusätzliche Anforderungen festgelegt, wodurch ein risikobasierter Ansatz in das Systemdesign integriert wird. Dieser Ansatz ergänzt das traditionelle Risikomanagement im Projektmanagement und adressiert Risiken explizit auch auf Systemebene. Hier gilt das Prinzip: Doppelt hält besser.

Takeaways

- Ein gründliches Requirements Engineering bildet das Fundament für den VBE-Prozess.
- Die Rolle des „Value Lead“ begleitet und führt Team und Stakeholder sicher durch den gesamten VBE-Prozess.
- In Stakeholder-Workshops werden ethische Werte erhoben, und die Stakeholder frühzeitig eingebunden.
- Aus diesen ethischen Werten entsteht eine Wertestrategie, die als Grundlage für die Softwareentwicklung dient.
- Ethische Wertanforderungen sind klar auf die zugrunde liegenden Werte rückverfolgbar und fließen in die Entwicklung ein.

15.3 Vorteile: Wieso sollte ich VBE in meinem Projekt nutzen?

Insgesamt bietet Value Based Engineering (VBE) einen systematischen und klar strukturierten Ansatz im Requirements Engineering, um die Prinzipien des Digitalen Humanismus konkret in Digitalisierungsprojekten umzusetzen. Durch die umfassende und nachvollziehbare Dokumentation sind sowohl Management als auch Projektteam jederzeit optimal auf ein Audit vorbereitet. Die Einhaltung eines international anerkannten Standards macht VBE-Projekte für Unternehmen attraktiv.

Trainings und Zertifizierungen

Für Fachkräfte besteht die Möglichkeit, ihr VBE-Know-how offiziell durch eine Zertifizierung als „Certified Value-Based Engineering Ambassador nach ISO 24748“ bei Austrian Standards anerkennen zu lassen. Schulungen werden von folgenden Einrichtungen angeboten:

- Value-Based Engineering (VBE) Academy (Value-Based Engineering, [2025](#))
- Austrian Standards International – Standardisierung und Innovation (Austrian Standards, [2025](#))

Da VBE auf dem Standard ISO/IEC/IEEE 24748-7000 basiert, können Projekte nach dieser Methode theoretisch zertifiziert werden. Aktuell gibt es jedoch keine offizielle Zertifizierung für Projekte nach VBE.

Relevanz für KI-Projekte

Im Hinblick auf den bevorstehenden EU AI Act und die potenzielle Einstufung von KI-Systemen als Hochrisiko-Anwendungen gewinnt VBE zunehmend an Bedeutung. Durch die Berücksichtigung ethischer Werte gemäß dem Standard können Unternehmen bei der Planung solcher Systeme sicherstellen, dass ethische Kriterien wie Transparenz und Sicherheit integriert werden, was auf die Compliance einzahlt.

Das IEEE CertifAIEd-Programm (IEEE, [2025](#)) ergänzt diesen Ansatz, indem es gezielt Zertifizierungen für KI-Produkte auf Basis der IEEE 7000-Standardreihe anbietet. Diese Zertifizierung bewertet fertige KI-Systeme in den Kategorien Zuverlässigkeit, Fairness, Transparenz und Sicherheit. Somit unterstützt es Unternehmen dabei, ihre ethischen Verpflichtungen nachweislich zu erfüllen.

Takeaways

- VBE ermöglicht mit seinem klaren Ansatz im Requirements Engineering die praktische Umsetzung des Digitalen Humanismus.
- Für Fachkräfte gibt es die Zertifizierung als „Certified Value-Based Engineering Ambassador“ bei Austrian Standards.
- VBE unterstützt die Einhaltung ethischer Anforderungen bei KI-Anwendungen und stärkt deren Compliance, beispielsweise bei Hochrisiko-KI-Anwendungen nach dem EU AI Act.

15.4 Wie sieht VBE in der Praxis aus? Die Implementierung bei ibis acam

15.4.1 Einleitung: Warum Praxis eine Schlüsselrolle spielt

Die zunehmende Digitalisierung und Automatisierung stellt Unternehmen vor die Herausforderung, ethische Werte nicht nur zu definieren, sondern aktiv in den Entwicklungsprozess ihrer Systeme einzubinden. Value Based Engineering (VBE) bietet einen strukturierten Ansatz, um diese Herausforderung zu meistern.

Der Fokus liegt darauf, Werte wie Selbstbestimmung, Fairness und Vertrauen methodisch zu analysieren und in technische Lösungen zu überführen.

Wie dieser Prozess in der Praxis umgesetzt wird, zeigt sich besonders deutlich anhand der Implementierung eines KI Assistenten, dessen Entwicklung durch die Prinzipien des VBE geleitet wurde.

15.4.2 Phase 1: Konzept und Kontext – Den Grundstein legen

Die erste Phase des VBE-Prozesses dient dazu, das Fundament für eine ethisch orientierte Entwicklung zu schaffen. Dabei ist es entscheidend, den Rahmen klar abzustecken und die Komplexität des Gesamtsystems auf das Wesentliche zu reduzieren.

Praktischer Fokus

Zur Abgrenzung des Gesamtsystems – dem System of Systems (SOS) – vom System of Interest (SOI) wurde die bestehende Application Architecture analysiert. Das SOS beschreibt das Gesamtsystem mit seinen Kommunikationsstrukturen, während das SOI, in diesem Fall der digitale Assistent, den spezifischen Fokusbereich für die VBE-Analyse definiert. Diese klare Trennung ermöglicht es, ethisch-moralische Fragestellungen gezielt zu adressieren, ohne die Analyse unnötig zu verkomplizieren.

Abb. 15.2 stellt diesen Prozess anschaulich dar. Sie erläutert die Abgrenzung des Gesamtsystems (SOS) vom System of Interest (SOI) und deren Kommunikationsstrukturen.

Ziel war es, Funktionen zu identifizieren, die den Wertvorstellungen der Stakeholder entsprechen, und diese als Basis für das Systemdesign und die KI-Implementierung zu nutzen.

Die zentralen Funktionen des digitalen Assistenten wurden auf Basis der Anforderungen definiert:

1. Kommunikation in natürlicher Sprache

Benutzer können intuitiv interagieren, ohne aufwändige Schulungen

2. Geführte Prozessanleitung

Unterstützung bei der korrekten und effizienten Durchführung komplexer Aufgaben

3. Proaktive Warnungen und Hinweise

Frühzeitige Benachrichtigungen über potenzielle Probleme, bevor sie eskalieren

4. Autonome Prozessausführung

Automatisierung wiederkehrender Tätigkeiten wie Datenerfassung und E-Mail-Ver-sand

Ein zentraler Bestandteil dieser Phase war die Business Value Analyse. Diese sollte zeigen, dass der KI Assistent nicht nur ethisch, sondern auch kommerziell sinnvoll ist.

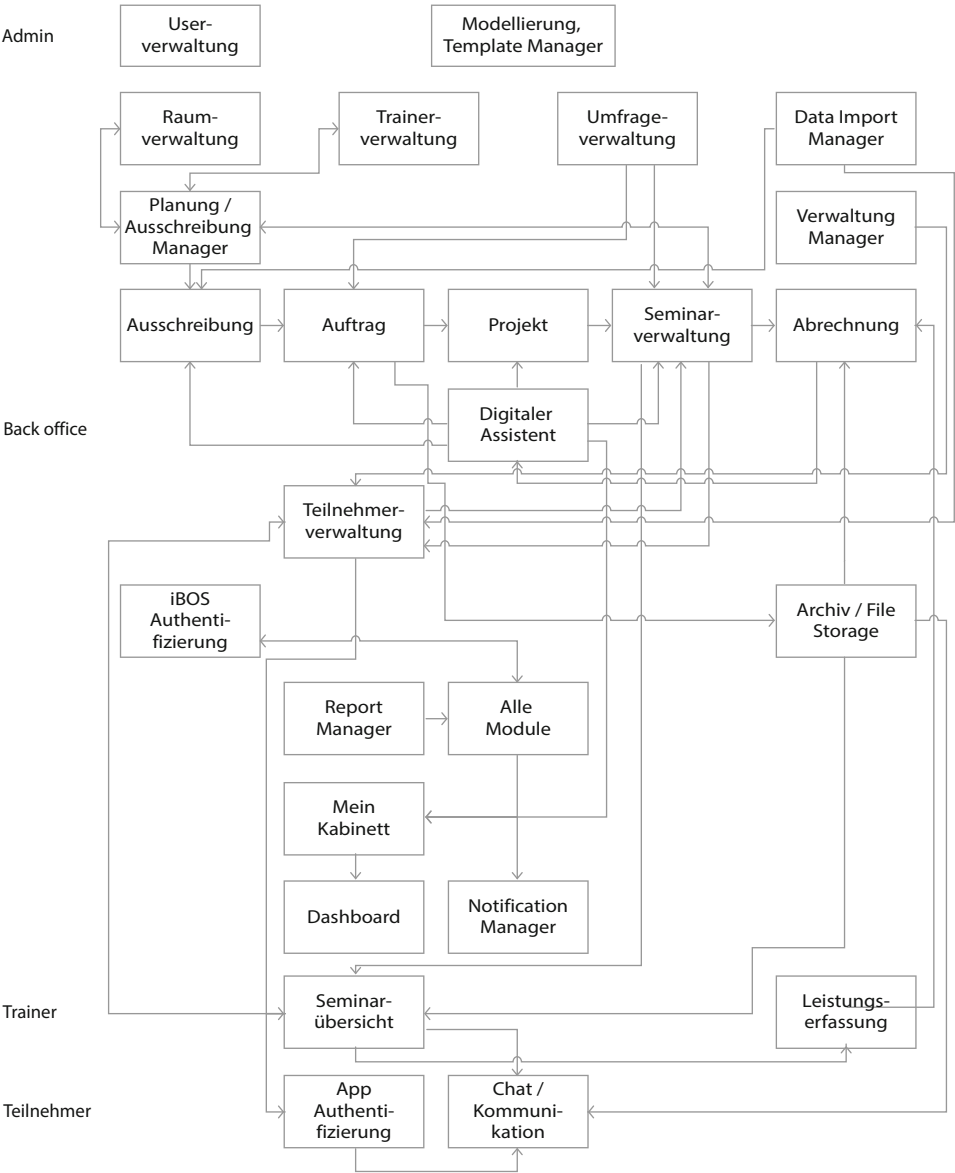


Abb. 15.2 Abgrenzung des Gesamtsystems (SOS) vom System of Interest (SOI)

Der Assistent unterstützt den täglichen Arbeitsablauf und liefert dabei erheblichen Mehrwert für AnwenderInnen und Unternehmen:

- **Produktivitätssteigerung**
Automatisierung entlastet Mitarbeiter und erhöht die Effizienz.
- **Fehler- und Ausfallzeitreduktion**
Präzise Prozesse minimieren Risiken und optimieren Abläufe
- **Verbesserte Benutzerfreundlichkeit**
Intuitive Interaktionen erleichtern den Zugang zu komplexen Funktionen
- **Kosteneinsparungen**
Effizientere Prozesse führen zu geringeren operativen Aufwänden

Mit der Abgrenzung des SOI, der Definition der Kernfunktionen und der Business Value Analyse wurde eine solide Grundlage geschaffen, um in den nächsten Phasen Werte der Stakeholder zu analysieren und ethische Anforderungen systematisch umzusetzen.

15.4.3 Phase 2: Werteanalyse – Von Anforderungen zu Werten

Die zweite Phase des VBE-Prozesses widmet sich der systematischen Erhebung und Analyse der Werte der Stakeholder (Tab. 15.2). Ziel ist es, die ethisch-moralischen Erwartungen aller Beteiligten zu erfassen, diese zu konkretisieren und in den Entwicklungsprozess zu integrieren. Dies erfordert gezielte methodische Ansätze, die über herkömmliche Systemdesign-Methoden hinausgehen.

Praktischer Fokus

Die Werteanalyse begann mit drei ganztägigen Workshops, die einen repräsentativen Querschnitt der Stakeholder einbezogen – darunter Vertreter der Geschäftsleitung, Regionalleitung, des Betriebsrats, HR sowie Trainer und Projektkoordinatoren.

Der erste Workshop vermittelte die grundlegende VBE-Methodik und schuf mit einer Einführung in den digitalen Assistenten und Business Cases eine gemeinsame Ausgangsbasis.

Im zweiten Teil wurden gezielt Werte wie Autonomie und Vertrauen in den Fokus gerückt.

Die Teilnehmer arbeiteten individuell und in Gruppen, um Rückmeldungen zu entwickeln, die in einem standardisierten VBE-Formular festgehalten wurden.

Die methodische Herangehensweise – von offenen Fragen bis zu kollaborativen Diskussionen – führte zu einem dynamischen und produktiven Austausch, dessen Ergebnisse die ethisch-moralischen Erwartungen und potenziellen Konfliktpunkte der involvierten Stakeholder klar sichtbar machten.

Tab. 15.2 Ergebnisse aus dem Workshop zu Utilitarismus

I. Potentialen und Risiken				II. Vorschläge zur Verbesserung
Betroffener Stakeholder	Beschreibung des Effekts	Wert	+ / –	
Trainer*in	Automatische Zeiterfassung ohne Mehraufwand	Zeitersparnis	+	
Trainer*in	Durch Automatisierung werden eigene Denkprozesse vernachlässigt	Reflexionsfähigkeit	–	
Trainer*in	TR fühlen sich durch KPI Monitoring unter Druck gesetzt	Vertrauen	–	
Alle	Mitarbeiter fühlen sich der KI ausgeliefert	Vertrauen, Sicherheit	–	
TR & FK	Wenn KI Aufgaben von FK übernimmt, geht auf der menschlichen Ebene zw. TR & FK etwas verloren	Gemeinschaft, Vertrauen	–	
TR & Teilnehmer*innen	Wenn durchgängig automatisiert wird, hat der TR mehr Zeit & Energie für die TN	Empathie, Menschlichkeit	+	

Die grundlegenden Erkenntnisse

Die Analyse zeigte deutlich, dass die Wahrnehmungen der Stakeholder grundsätzlich in zwei Kategorien unterteilt werden können: klassische *funktionale Erwartungen* und *ethisch-moralische Anforderungen*.

Zu den positiven *funktionalen* Aspekten zählten:

- **Steigerung von Effektivität und Effizienz**
Der KI Assistent wird als wesentliche Unterstützung zur Optimierung von Prozessen wahrgenommen
- **Zeitersparnis**
Die Automatisierung von Aufgaben reduziert den manuellen Aufwand erheblich
- **Verfügbarkeit von Wissen**
Der KI Assistent bietet klare Anleitungen und erleichtert die Navigation durch Geschäftsprozesse

Demgegenüber standen ethisch-moralische Bedenken, die primär *nicht-funktionale* Themen betrafen

- **Autonomie und Selbstbestimmung**
Stakeholder befürchten, dass ihre Entscheidungsfreiheit durch Automatisierung eingeschränkt wird

- **Vertrauen versus Überwachung**

Die automatisierte Erfassung und Verarbeitung von Daten wurde teilweise als Kontrollmechanismus wahrgenommen

- **Bedenken zur Privatsphäre**

Sorgen hinsichtlich des Schutzes persönlicher Informationen waren besonders ausgeprägt

Fazit der Phase 2

Die Werteanalyse offenbarte, dass klassische funktionale Anforderungen wie Effizienz und Produktivitätssteigerung oft mit ethisch-moralischen Erwartungen wie Selbstbestimmung und Privatsphäre in Konflikt geraten können.

Dieser Erkenntnisprozess ist entscheidend, um sicherzustellen, dass das Systemdesign sowohl funktionale als auch nicht-funktionale Aspekte berücksichtigt.

Die Ergebnisse dieser Phase bilden somit die Grundlage, um in der nächsten Phase konkrete ethische und technische Anforderungen für die Umsetzung des digitalen Assistenten abzuleiten.

15.4.4 Phase 3: Systemdesign – Vom Wert zur Funktion

Die dritte Phase des VBE-Prozesses bildet die Schnittstelle zwischen den analysierten Werten und den daraus resultierenden technischen Anforderungen.

Ziel ist es, die ethisch-moralischen Erwartungen der Stakeholder in konkrete Systemfunktionen zu übersetzen, die sowohl den funktionalen Anforderungen als auch den ethischen Prinzipien gerecht werden.

Praktischer Fokus

Im Mittelpunkt dieser Phase stand die Umsetzung der in Phase 2 ermittelten Werte in ein Systemdesign, das die Balance zwischen funktionaler Effizienz und ethischen Anforderungen wahrt.

Dies erforderte die enge Zusammenarbeit zwischen Stakeholdern und Vertretern des Entwicklungsteams, um sicherzustellen, dass die technischen Lösungen den Erwartungen entsprechen.

Ein zentrales Beispiel war das Thema **Autonomie und Selbstbestimmung**. Die Sorge der Stakeholder, durch den Assistenten Entscheidungsfreiheiten zu verlieren, wurde durch eine Funktion adressiert, die den Benutzern stets die letzte Entscheidungsgewalt überlässt. Konkret wurde dies im Bereich des Urlaubsmanagements umgesetzt: Der Assistent schlägt automatisch geeignete Vertretungen vor, überlässt jedoch die finale Genehmigung stets dem Benutzer. Diese Lösung berücksichtigt nicht nur die ethischen Erwartungen, sondern schafft auch Vertrauen in die Automatisierung.

Ein weiteres Beispiel betraf das Thema **Vertrauen und Überwachung**. Um die Transparenz der Datenverarbeitung zu erhöhen, wurden für den Assistenten klare Visualisierungen der Datenflüsse und Entscheidungskriterien entwickelt. So wurden etwa Warnhinwei-

se, die auf potenzielle Konflikte im Arbeitszeitmanagement hinweisen, nicht nur frühzeitig bereitgestellt, sondern auch mit einer verständlichen Erläuterung der zugrunde liegenden Logik versehen. Dies stellte sicher, dass die Nutzer jederzeit nachvollziehen konnten, wie und warum Entscheidungen des Systems getroffen wurden.

Fazit der Phase 3

Phase 3 zeigte eindrücklich, wie technische Lösungen gestaltet werden können, die sowohl funktionale Anforderungen erfüllen als auch ethische Werte berücksichtigen.

Die Beispiele aus dem Projekt verdeutlichen, dass der Erfolg moderner Systeme zunehmend von der Fähigkeit abhängt, die Akzeptanz der Stakeholder durch transparente, nachvollziehbare und wertorientierte Designentscheidungen zu gewinnen.

Diese Phase schloss den Kreis zwischen den Erwartungen der Nutzer und der technischen Umsetzung, wodurch ein nachhaltiges und zukunftssicheres System geschaffen wurde.

15.5 Conclusio

Die Fallstudie bei ibis acam und die Darlegung der Theorie zeigen, wie Value Based Engineering (VBE) als strukturierte Methode in Digitalisierungsprojekten angewendet werden kann, um ethische Aspekte frühzeitig und systematisch zu berücksichtigen. Durch die klare Analyse des Systemkontexts, die Einbindung relevanter Stakeholder und die Ableitung konkreter technischer Anforderungen aus identifizierten Werten wurde ein nachvollziehbarer und sinnvoller Entwicklungsprozess geschaffen.

Dabei wird deutlich, dass funktionale Anforderungen und ethische Erwartungen bewusst aufeinander abgestimmt werden können. Die Umsetzung bei ibis acam verdeutlicht, dass VBE eine praktikable Methode ist, um Softwareprojekte sowohl technisch als auch organisatorisch auf eine nachhaltige und risikobewusste Basis zu stellen.

Literatur

- Austrian Standards International (2025). *Certified Value-Based Engineering Ambassador nach ISO 24748*
- Europäische Union (2024). *Ziele und Werte der Europäischen Union*. Brüssel: Europäische Union. https://european-union.europa.eu/principles-countries-history/principles-and-values/aims-and-values_en (abgerufen am 6. April 2025)
- IEEE (2021). IEEE 7000-2021: IEEE Standard Model Process for Addressing Ethical Concerns during System Design. <https://technologyandsociety.org/what-to-expect-from-ieee-7000-the-first-standard-for-building-ethical-systems/>. Zugegriffen: 15. März 2025.
- IEEE CertifAIEd-Programm (2025). *Zertifizierungen für KI-Produkte auf Basis der IEEE 7000-Standardreihe*
- Internationale Norm (2022). ISO/IEC/IEEE 24748-7000: Systems and Software Engineering – Life Cycle Management – Standard Model Process for Addressing Ethical Concerns during System Design. <https://www.iso.org/standard/84893.html>. Zugegriffen: 15. März 2025.

Value-Based Engineering: Offizielle Webseite. <https://value-based-engineering.com>. Zugriffen: 6. Apr. 2025.

Weiterführende Literatur

Austrian Standards (2024). Offizielle Webseite. Wien: Austrian Standards. <https://www.austrian-standards.at>. Zugriffen: 6. Apr. 2025.

EU AI Act (2025). *Hochrisiko-KI-Anwendungen und ethische Anforderungen*

IEEE Standards Association (2024). IEEE CertifAIEd™ – The Mark of AI Ethics. Piscataway, NJ: IEEE Standards Association. <https://standards.ieee.org/products-programs/icap/ieee-certifaiied/#process-offerings>. Zugriffen: 6. Apr. 2025.

Spiekermann-Hoff, S. (2022). *Value Based Engineering (VBE) und der Standard ISO/IEC/IEEE 24748-7000:2022: Standard Model Process For Addressing Ethical Concerns During System Design*

Dr. Svenja Schröder leitet die Abteilung für Requirements Engineering und User Experience bei msg Plaut. Seit über 15 Jahren ist sie in Wirtschaft und Wissenschaft in IT-Projekten tätig und beschäftigt sich mit der Schnittstelle zwischen Technologie, Mensch und Gesellschaft. Sie studierte Angewandte Kommunikations- und Medienwissenschaften – einen interdisziplinären Studiengang mit Schwerpunkten in Informatik, Psychologie und Medienwissenschaft – an der Universität Duisburg-Essen. Über verschiedene Forschungsaufenthalte führte ihr Weg an die Universität Wien, wo sie im Fach Technische Informatik mit Auszeichnung promovierte. Mit ihrem Team unterstützt sie Unternehmen in den Bereichen Requirements Engineering, Business Analyse, User Experience und Prozessmanagement – stets mit einem Fokus auf menschenzentrierte und werteorientierte Digitalisierung.

Florian Wurzer, MBA, ist Leiter der Softwareentwicklung bei msg Plaut und verantwortlich für die Software- und Produktentwicklung. Mit seiner langjährigen Erfahrung in der Entwicklung komplexer IT-Systeme treibt er die Digitalisierung in verschiedenen Branchen voran. Sein Fokus liegt auf modernen Architekturkonzepten, agilen Methoden und der Integration ethischer Werte in technologische Entwicklungen.

Im Rahmen seiner Tätigkeit beschäftigt er sich intensiv mit Value Based Engineering (VBE), insbesondere bei der Entwicklung KI-gestützter Systeme. Durch praxisnahe Implementierungsmethoden setzt er Impulse für eine nachhaltige, wertorientierte Softwareentwicklung.

Mit der System-Theoretic Process Analysis (STPA) 16 Zusammenhänge verstehen und Werte schützen

Das Zusammenspiel von STPA und VBE am Beispiel einer Lade-App für Elektrofahrzeuge

Florian Wagner

Zusammenfassung

Um Anforderungen des Value-Based Engineerings (VBE) wirtschaftlich und nachvollziehbar in Unternehmen und Produkten zu erfüllen, hat sich die System-Theoretic Process Analysis (STPA) als hilfreich erwiesen. Sie wurde am MIT entwickelt und hat sich vor allem in Hochsicherheitsbereichen wie der Automobilbranche, Luft- und Raumfahrt durchgesetzt. Zu den prominentesten Anwendern zählen NASA, Embraer, American Airlines und Google. Am Beispiel einer Lade-App wird die praktische Anwendung von STPA im Kontext von VBE veranschaulicht. Die Ergebnisse zeigen, dass eine klare Rückverfolgbarkeit zwischen Werten und technischen Lösungen entscheidend ist, um positive Werte in einem System zu stärken.

Die Kunst der Ingenieursarbeit liegt darin, Werte sichtbar zu machen.

16.1 Einleitung

Unternehmen und ihre Produkte müssen zunehmend strengere Anforderungen an Aspekte wie Sicherheit erfüllen. Gleichzeitig steigt auch die Erwartung seitens der Kundschaft und der Gesellschaft, dass auch ethische Gesichtspunkte berücksichtigt werden. Damit diese

AI statement. Ziel der KI-Nutzung: – Textbausteine oder Inspiration, – Bearbeitung oder Überarbeitung; Ergebnis: Der Einsatz von KI betrifft: – geringe Teil des Dokuments.

F. Wagner (✉)

msg Plaut, Wien, Österreich

E-Mail: florian.wagner@msg-plaut.com

Anforderungen und Erwartungen nachvollziehbar und wirtschaftlich erfüllt werden können, benötigt man kluge Methoden, um diese Aspekte „by Design“ in der Entwicklung zu berücksichtigen. Die System-Theoretic Process Analysis (STPA) hat sich in diesem Zusammenhang als hilfreich erwiesen. Die Methode erlaubt die verschiedenen Ziele und Werte, die eingehalten werden sollen, in einer Analyse zu vereinen. Die Ergebnisse zeigen, dass eine klare Rückverfolgbarkeit zwischen Werten und technischen Lösungen entscheidend ist, um positive Werte in einem System zu stärken.

16.2 System-Theoretic Process Analysis

Die System-Theoretic Process Analysis (STPA) ist eine Methode zur Gefährdungs- und Risikoanalyse, die am MIT entwickelt wurde. Sie eignet sich besonders zur Analyse von komplexen, softwareintensiven und dynamischen Systemen. Damit hat sie sich vor allem in Hochsicherheitsbereichen wie der Automobilbranche, Luft- und Raumfahrt durchgesetzt. Zu den prominentesten Anwendern zählen NASA, Embraer, American Airlines und Google. Im Gegensatz zu vielen traditionellen Sicherheitsanalysemethoden werden nicht nur technische, sondern auch organisatorische und menschliche Faktoren berücksichtigt. Deshalb – und wegen der universellen Zielsetzung – eignet sie sich auch besonders gut für Analysen im Kontext des digitalen Humanismus.

Das Ziel der STPA ist es Werte wie Sicherheit, Verlässlichkeit, Transparenz bzw. allgemein Etwas, das einen Wert für einen Stakeholder hat, zu schützen. Das Ziel der Analyse kann somit individuell gesetzt werden. Hier können die Werte, die im Zuge des Value-Based Engineerings (VBEs) Prozesses wie in Kap. 15 mit dem Case Ibis Acam ermittelt und priorisiert wurden, berücksichtigt werden.

16.2.1 Wie die STPA funktioniert

Eine detaillierte Beschreibung der STPA findet sich in Leveson und Thomas (2018) und SAE (2024). Die STPA basiert auf einem systemtheoretischen Ursachenmodell. Dieses Modell berücksichtigt Entscheidungsfindungsprozesse, Interaktionen, Veränderung über die Zeit und weitere Aspekte. Der Schutz der oben erwähnten Werte hängt nicht von einzelnen Komponenten, Menschen oder anderen Elementen in einem System ab, sondern hängt vom Zusammenspiel aller der beteiligten Elemente und Menschen ab.

Im ersten Schritt werden Schadensereignisse (sogenannte „Losses“) formuliert, die abgewendet werden sollen. Sie werden als der Verlust von etwas beschrieben, das für Stakeholder von Wert ist. Im Kontext des VBEs sind Schadensereignisse als Verlust bzw. Schädigung positiver Werte bzw. das Auftreten negativer Werte definiert. Beispiele sind der Verlust von Sicherheit, Reputationsverlust oder Auftreten von Intransparenz.

Ob diese Schadensereignisse eintreten oder nicht kann manchmal außerhalb der eigenen Kontrolle als Firma oder Organisation sein. Deshalb werden sogenannte „Hazards“

Tab. 16.1 Schlagwörter (Keywords) für die UCA-Identifizierung

Originale Definition (Leveson & Thomas, 2018)	Übersetzung bzw. Beschreibung
Not providing causes hazard	Das nicht Bereitstellen, nicht Durchführen einer Control Action bzw. Nicht-Handeln führt zu einer Gefährdung
Providing causes hazard	Das Bereitstellen, Durchführen einer Control Action bzw. eine Handlung führt zu einer Gefährdung
Too early, too late, out of order	Das Bereitstellen, Durchführen einer grundsätzlich korrekten Control Action bzw. Handlung zur Unzeit bzw. verfrüht oder zu spät führt zu einer Gefährdung
Stopped too soon, applied too long	Das Bereitstellen, Durchführen einer grundsätzlich korrekten Control Action bzw. Handlung wird zu früh gestoppt bzw. hält zu lange an und führt zu einer Gefährdung

definiert. Das sind Systemzustände die unter der eigenen Kontrolle (im Entwurfsstadium und/oder in der Betriebsphase) stehen. Unter worst-case Bedingungen – die außerhalb der Kontrolle des zu entwickelnden Systems liegen – führen sie zu einem „Loss“. Daher zielt der abschließende System-Design-Prozess darauf ab, das Auftreten von „Hazards“ zu vermeiden. Einige Beispiele für Hazards:

- Bei einem Auto kann das Unterschreiten eines Mindestabstands zu anderen Verkehrsteilnehmern einen Hazard darstellen, der zum Loss „Tod oder Verletzung von Menschen“ führen kann.
- Bei einem Online-Shop können falsch angezeigte Preise einen Hazard darstellen, die zum Loss „Intransparenz“ können.

Für die weitere Analyse wird ein logisches Modell des Systems benötigt, eine sogenannte „Control Structure“ definiert. Es erfasst funktionale Beziehungen und Interaktionen in einem System. Hierbei sind sowohl menschliche Akteure als auch ihre Interaktionen mit technischen Systemen explizit zu berücksichtigen. Diese Beziehungen und Interaktionen werden durch sogenannte „Control Actions“ und „Feedbacks“ modelliert. Ersteres sind Steuerbefehle, mit der eine Systemeinheit eine andere beeinflusst bzw. Vorgaben setzt. Zweiteres sind Informationen, die eine Systemeinheit bekommt und zur Entscheidungsfindung nützt. Das können Sensordaten, visuelle Signale oder menschliche Kommunikation sein. Die exakte technische Umsetzung von „Control actions“ und „Feedbacks“ ist nicht zwingend relevant. Abb. 16.1 zeigt eine „Control Structure“.

Danach erfolgt eine systematische Analyse, welche „Control Actions“ in bestimmten Situationen zu Gefährdungen führen können. Diese Befehle werden als „Unsafe Control Actions“ (UCAs) bezeichnet, wobei das Wort „unsafe“ (unsicher) historisch bedingt ist. Unsicher in diesem Kontext sind Steuerbefehle, die zu Gefährdungen (Hazards) führen. Die Analyse bedient sich dabei vier Schlagwörtern, um unsichere Control Actions zu identifizieren (siehe Tab. 16.1).

Abschließend werden Verlustszenarien (sogenannte Loss Scenarios) und damit die Ursachen für UCAs identifiziert. Die Ursachen können in folgende Grundklassen eingeteilt werden (die Entscheidungsfindung kann durch einen Menschen oder durch ein Steuergerät stattfinden):

- Klasse 1: Eine UCA wird angewiesen, obwohl Inputs zur Entscheidungsfindung korrekt waren
- Klasse 2: Eine UCA wird angewiesen, da die Inputs zur Entscheidungsfindung inkorrekt waren
- Klasse 3: Ein sicherer Steuerbefehl („Control Action“) wird angewiesen, aber inkorrekt/unsicher durchgeführt (z. B. Problem mit dem Aktuator)
- Klasse 4: Ein Steuerbefehl auf Basis der Inputs wie spezifiziert korrekt angewiesen und ausgeführt, aber der Steuerbefehl führt zu einer Gefährdungssituation

Diese Grundklassen können kombiniert werden. So kann beispielsweise ein inkorrekt durchgeführter Steuerbefehl (Klasse 3) in weiterer Folge zu inkorrekten Inputs führen (Klasse 2). Zudem ist es im Sinne einer umfassenden Suche nach Loss Scenarios ratsam auf einem allgemeineren Level zu beginnen, und die Ursachen immer spezifischer einzugrenzen. Diese hierarchische Definition von allgemein zu spezifisch ist auf mehrfache Weise hilfreich.

16.2.2 Wie die STPA Value-Based Engineering unterstützen kann

Zum Darlegen der Eignung der STPA im Kontext des VBEs, wird in diesem Abschnitt ein Abgleich der Definitionen der STPA mit jenen der ISO/IEC/IEEE 24748-7000:2022-Norm gezeigt (siehe auch IEEE, 2021). In Tab. 16.2 ist eine Gegenüberstellung der Definitionen zu finden.

Wie zuvor erwähnt, definiert die STPA „Losses“ als den Verlust eines Wertes bzw. dem Verlust von etwas, das einen Wert hat/darstellt. Im Kontext des VBEs sind „Losses“ als Verlust bzw. Schädigung positiver Werte bzw. das Auftreten negativer Werte definiert. Beispiele sind der Verlust von Sicherheit, Reputationsverlust oder Auftreten von Intransparenz. Eine Verletzung dieser Werte wird in der IEEE-Norm als „Harm“ bezeichnet.

Tab. 16.2 Gegenüberstellung von Definitionen aus der ISO/IEC/IEEE 24748-7000:2022 und der STPA

ISO/IEC/IEEE 24748-7000:2022 (IEEE, 2021)	STPA (Leveson & Thomas, 2018)
Harm	Loss
	Harm
Value harm	Hazard
Hazard	
Ethical value requirement (EVR)	Constraint

Vergleicht man die Definitionen, kann man „Harm“ mit der STPA-Definition von „Losses“ gleichsetzen.

Der Begriff „Hazard“ wird in beiden Normen erwähnt, jedoch mit unterschiedlicher Bedeutung verwendet. Der Begriff wird typischerweise mit Sicherheit und Risikomanagement verbunden. In der Norm ist er als eine potenzielle Ursache für Personen-, Sach- und Umweltschäden recht klassisch definiert, die Definition ist aus ISO (2018) übernommen. Daher definiert die Norm den Begriff des „Value Harms“ um die oft immateriellen Effekte auf Werte und Ethik der Stakeholder zu berücksichtigen.

Die STPA geht einen umgekehrten Weg und führt keinen neuen Begriff ein, sondern definiert „Hazards“ breiter. Zur Erinnerung: Ein STPA „Hazard“ ist ein Systemzustand, der unter worst-case Bedingungen zu einem „Loss“ führt. Da der „Loss“ von Werten abgeleitet wird, umfasst ein „Hazard“ alle Zustände, die Werte beschädigen. Es stellt einen potenziellen Schaden dar, der jedoch nicht unvermeidlich ist.

Somit stellen die Begriffe „Value Harm“ (im Kontext der Norm) und „Hazard“ (im Kontext der STPA) jeweils ein Potenzial für einen Verlust dar. Beim „Value Harm“ ist von einem Untergraben des Wertes die Rede. Das Untergraben eines Wertes bedeutet noch nicht, dass ein kompletter Werte-Verlust („Harm“) eingetreten ist. Ein STPA „Hazard“ führt nur im Zusammenspiel mit worst-case Bedingungen zu einem „Loss“. Er stellt einen potenziellen Schaden dar, der jedoch nicht unvermeidlich ist. Im Folgenden sind „Hazards“ als STPA „Hazards“ zu verstehen.

Die identifizierten „Loss Scenarios“ bilden darüber hinaus die Grundlage für die Definitionen von Systemanforderungen zur Vermeidung von „Losses“, somit bilden sie auch die Grundlage für die Definition von konkreten ethischen Wertanforderungen („Ethical Value Requirements“, EVR).

16.3 Beispiel: Analyse einer Lade-App für Elektrofahrzeuge

Im Folgenden wird ein Auszug des Ablaufs einer STPA im Kontext des VBEs anhand eines Beispiels gezeigt. Die zugrundeliegende Arbeit auf Englisch durchgeführt. Das betrachtete System ist eine Lade-App für Elektrofahrzeuge, die folgende Funktionen und Services bieten soll:

- Bezahlung der geladenen Energie bei verschiedenen Ladestation-Anbietern
- Bezahlung über die App
- Routenplanung mit KI-basierter Optimierung, z. B. Optimierung auf
 - Gesamtreisezeit: schnelle, aber eventuell teurere Ladestationen werden bevorzugt
 - Ladepreis: kleine Umwege werden akzeptiert, um beim Ladepreis zu sparen
 - Infrastruktur in der Nähe einer Ladestation: Ladestationen mit WCs, Supermärkten oder Restaurants in der Nähe werden bevorzugt, um Ladezeit besser zu nutzen
- Informationen über die Infrastruktur in der Nähe einer Ladestationen

Tab. 16.3 Auszug identifizierter „Losses“ für das System „Lade-App“

ID	Loss	Beschreibung des Wertes dahinter (IEEE, 2021)
L-1	Verlust der Autonomie	Selbstbestimmung von Zielen und Handlungen
L-2	Kontrollverlust	Entscheidungsgewalt, Wahlmöglichkeiten
L-5	Verlust der Nachhaltigkeit	Umweltschutz, Wartbarkeit, Erneuerbarkeit
L-6	Verlust der Transparenz	Nachvollziehbarkeit, Zugänglichkeit
L-7	Verlust der Sicherheit	Keine Gefahr für Leben, Eigentum oder Umwelt
L-8	Vertrauensverlust	Kompetenz, Aufrichtigkeit, Vorhersehbarkeit

Tab. 16.4 Auszug identifizierter „Hazards“

ID	Hazards	Trace zu Losses
H-5	Die App gibt unzuverlässige Informationen und Anweisungen.	L-1, L-6, L-8
H-6	Die Bezahlung kann nicht korrekt abgewickelt.	L-2, L-8
H-7	Die gefahrene Route erfüllt die Optimierungsziele nicht.	L-2, L-5, L-8
H-8	Die Infrastruktur in der Nähe von angefahrenen Ladestationen entspricht nicht den Anforderungen.	L-1, L-2, L-7, L-8

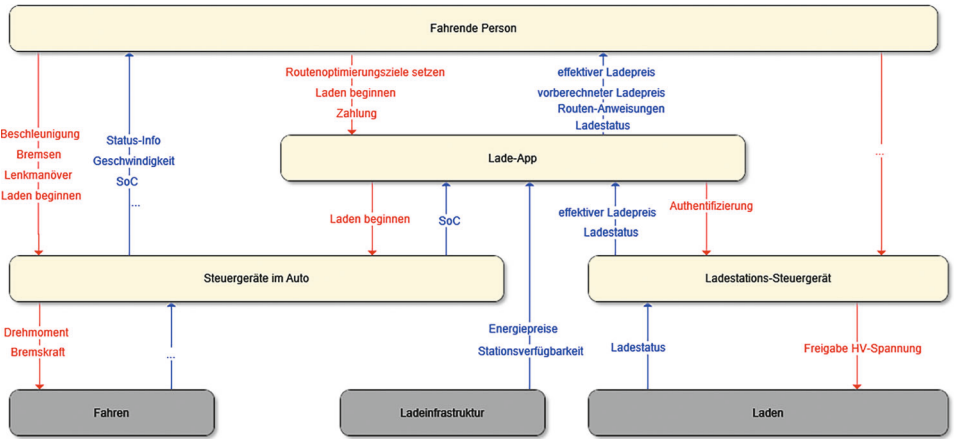


Abb. 16.1 „Control Structure“ des Systems und dem Systemkontexts

Im ersten Schritt müssen die „Losses“ definiert werden. Die für das Beispiel relevanten und auf Basis von „Core Values“ identifizierten „Losses“ sind in Tab. 16.3 aufgelistet.

Im Folgenden wird in diesem Beispiel die Analyse von „Loss“ L-8 „Vertrauensverlust“ im Detail behandelt. „Vertrauen“ ist der Überbegriff für den „Value Cluster Vertrauen“, welcher die Werte Vorhersehbarkeit und Verlässlichkeit in die App beinhaltet. Werden die versprochenen Services nur unzuverlässig angeboten, mindert es das Vertrauen in die App.

Als nächstes werden „Hazards“ identifiziert. Dabei sollen Systemzustände beschrieben werden, die unmittelbar durch das Design der App beeinflusst werden können. In der Regel werden während der Analyse neue „Hazards“ entdeckt. Zum Start ist es auch ausreichend eine initiale Liste an „Hazards“ zu haben, abgeleitet durch die versproche-

Tab. 16.5 Auszug identifizierter „UCAs“

ID	Unsafe Control Actions	Trace zu Hazards
UCA-1	Die fahrende Person führt eine höhere Zahlung durch als den von der App kommunizierten effektiven Ladepreis.	H-5
UCA-26	Die fahrende Person führt Lenkmanöver aus, die das Fahrzeug von der energiepreisoptimierten Route abbringt.	H-5, H-7
UCA-27	Die fahrende Person führt ein Lenkmanöver aus, die das Fahrzeug zu Ladestationen ohne sanitäre Infrastruktur führt, obwohl benötigt.	H-7, H-8

Tab. 16.6 Auszug identifizierter „Loss Scenarios“

ID	Loss Scenarios	Trace zu UCAs
LS-1	Aufgrund eines Fehlers im Preiscalgorithmus der Lade-App wird der fahrenden Person ein falscher effektiver Ladepreis mitgeteilt. Folglich zahlt die fahrende Person mehr als den von der App angegebenen Betrag.	UCA-1
LS-2	Der effektive Ladepreis, den die fahrende Person zahlen muss, ist höher als im Voraus kommuniziert, da die Ladeinfrastruktur der App einen falschen Energiepreis übermittelt hat.	UCA-1
LS-3	Die Routen-Anweisungen der Lade-App leiten die fahrende Person zu Ladestationen, die keine sanitäre Infrastruktur haben.	UCA-27
LS-3-1	Die Routen-Anweisungen der Lade-App leiten die fahrende Person aufgrund falscher Daten zur Routenplanung zu Ladestationen, die keine sanitäre Infrastruktur haben.	UCA-27

nen Services der App. In Tab. 16.4 ist ein Auszug der identifizierten „Hazards“ und die Verlinkung zu den „Losses“ zu finden.

Die dazugehörige Kontrollstruktur ist in Abb. 16.1 dargestellt. Neben der App selbst ist auch das Umfeld zu sehen, mit dem sie interagiert. Diese Struktur bildet die Basis für die Identifikation der „Unsafe Control Actions“ (UCA).

Ein Auszug an identifizierten UCAs ist in Tab. 16.5 und eine Auswahl der daraus abgeleiteten „Loss Scenarios“ ist in Tab. 16.6 zu finden. Die Auswahl fokussiert sich auf App-relevante Informationen. Grundsätzlich ist eine STPA immer auf den gesamten Systemkontext durchzuführen.

Aus den „Loss Scenarios“ (zu deutsch: Verlustszenarien) lassen sich Wertedemonstratoren ableiten. Beispielsweise zeigt das LS-1, dass der Algorithmus zur Preisberechnung ein Wertedemonstrator ist. Der fehlerhafte Algorithmus im negativen, ein korrekter im positiven Sinne. In Tab. 16.7 sind in **fett** weitere Wertedemonstratoren gelistet, die aus den „Loss Scenarios“ abgeleitet wurden.

Die Analyse schafft eine Rückverfolgbarkeit zwischen den definierten Werten und Wertedemonstratoren. Dies ermöglicht argumentative Ketten wie in Abb. 16.2 veranschaulicht. Potenzielle Verlustszenarien können so frühzeitig identifiziert und Maßnahmen zum Schutz der Werte gezielt entwickelt werden. In weiter Folge können EVRs formuliert werden.

Tab. 16.7 Wertedemonstratoren abgeleitet aus „Loss Scenarios“

ID	Wertedemonstratoren Definition	Trace zu Loss Scenarios
VD-1	Autonomie, Transparenz und Vertrauen geschädigt durch einen fehlerhaften Steuerungsalgorithmus	LS-1
VD-3	Autonomie, Transparenz und Vertrauen geschädigt durch falsche Energiedatenpreise	LS-2
VD-9	Autonomie, Kontrolle, Nachhaltigkeit, Sicherheit und Vertrauen geschädigt durch unzweckmäßige Wegbeschreibung durch die Lade-App	LS-3
VD-9-1	Autonomie, Kontrolle, Nachhaltigkeit, Sicherheit und Vertrauen geschädigt durch falsche Daten zur sanitären Infrastruktur	LS-3-1



Abb. 16.2 Argumentative Kette für die Begründung und Bewertung von Systemanforderungen und deren Nutzen zum Schutz von Werten

16.4 Conclusio und Ausblick

Die Einhaltung und Sicherstellung von Werten kann von entscheidender oder gar kritischer Bedeutung für ein System sein (Spiekermann, 2023). Daher sollen Systeme mit hohen Wert-Erwartungen mit umfassenderen Methoden des risikobasierten Systemdesigns analysiert werden. Mit der STPA wurde hier eine Methode für diesen Zweck vorgestellt.

Werte werden manchmal von vordergründigen technischen Lösungen überschattet. So bezeichnet beispielsweise Tesla deren leistungsstarke Batterie als einen ihrer „Core Values“ (Spiekermann, 2023). Tatsächlich unterstützt die Batterie Wertedemonstratoren wie Flexibilität, Zeitersparnis und Reichweite, die für den Wert „Mobilität“ wichtig sind. Es ist eine Verwechslung technischer Voraussetzungen mit den Werten selbst. Eine isolierte Betrachtung von Systemkomponenten ist in komplexen sozio-technischen Systemen nicht wirkungsvoll, um Werte zu schaffen. Aber eine klare Rückverfolgbarkeit zwischen Werten, Wertedemonstratoren und technischen Lösungen ist wichtig. Sie hilft, diese technischen Lösungen zu verbessern und schafft eine Rückverfolgbarkeit. STPA zeigt die Stellschrauben auf an denen gedreht werden muss, um die Vorteile bestimmter Werte zu erhöhen und „Value Harms“ zu reduzieren.

Literatur

- IEEE (2021). *7000-2021 – IEEE Standard Model Process for Addressing Ethical Concerns during System Design*. IEEE – Institute of Electrical and Electronics Engineers.
- ISO (2018). *DIN ISO 31000 – Risikomanagement – Leitlinien, Standard*.
- Leveson, N. G., & Thomas, J. P. (2018). *STPA Handbook*. MIT Partnership for Systems Approaches to Safety and Security (PSASS).
- SAE (2024). *J3307 – System Theoretic Process Analysis (STPA) Standard for All Industries, Standard*.
- Spiekermann, S. (2023). *Value-based engineering: a guide to building ethical technology for humanity*. in *De Gruyter Graduate*. Berlin. Boston: De Gruyter.

Florian Wagner ist seit 2021 bei msg Plaut Austria und dort als Senior Business Consultant tätig und hat sich auf sicherheitskritische Themen sowie Model-Based Systems Engineering spezialisiert. Mit Erfahrung in den Bereichen Automotive, Railway, Aerospace und Maschinenbau unterstützt er Unternehmen dabei, komplexe Systeme sicher und resilient zu gestalten. Nach Abschluss der HTL für Elektronik absolvierte das Bachelorstudium der Elektrotechnik an der TU Wien und das Masterstudium „Safety and Systems Engineering“ der FH Campus Wien. Bereits während seiner Studienzeit sammelte er als Ingenieur wertvolle Erfahrung. Vor seiner Anstellung bei der msg Plaut war er unter anderem für in-tech engineering und Austrian Airlines tätig. Seit 2017 ist Florian Wagner ein aktiver Anwender und Verfechter der System Theoretic Prozess Analysis (STPA). Er hatte die Gelegenheit unter anderem am renommierten STAMP Workshop des MIT zu präsentieren und ist auch in Normierungsgremien zur STPA tätig.

Corporate Digital Responsibility als ganzheitlicher Ansatz für Digitalen Humanismus

17

Aufbau eines CDR-Managementsystems und die Umsetzung bei Merck

Karin Dietl, Simon Lucas und Jean Enno Charton

Zusammenfassung

Die „Triple Transition“ – aus der digitalen, ökologischen und sozialen Transformation – stellt Unternehmen vor neue Herausforderungen. Corporate Digital Responsibility (CDR) dient als verbindendes Element, um ethische, nachhaltige und technologische Ziele miteinander zu verknüpfen. Digitale Ethik etabliert sich nicht nur als neue akademische Disziplin, sondern auch als neue Ausprägung der praxisorientierten Technik- und Unternehmensethik und fordert eine verantwortungsvolle Gestaltung digitaler Innovationen. Diese Entwicklung fördert nicht nur den Erhalt zentraler gesellschaftlicher Werte, sondern stärkt Vertrauen, Compliance und nachhaltige Geschäftsmodelle, die den Wandel aktiv vorantreiben. Angesichts der bevorstehenden globalen Umbrüche und Unsicherheiten in den nächsten Jahrzehnten bietet CDR die Chance, digitale Innovation verantwortungsvoll zu gestalten und Vertrauen in eine sich wandelnde Gesellschaft zu stärken.

Nachhaltigkeit neu gedacht – Die Triple Transition und die Rolle der digitalen Ethik

K. Dietl (✉)

VALUE DESIGN, Wien, Österreich

E-Mail: k.dietl@cdr-austria.com

S. Lucas · J. E. Charton

Merck KGaA, Darmstadt, Deutschland

E-Mail: simon.lucas@merckgroup.com

J. E. Charton

E-Mail: ethics@merckgroup.com

17.1 Corporate Digital Responsibility und Digitale Ethik

17.1.1 Einleitung und Hintergründe

Seit 2019 erhält der Begriff der **Corporate Digital Responsibility (CDR)** zunehmend Beachtung im wissenschaftlichen und unternehmerischen Kontext als ähnliches Konzept anlehnend zur Corporate Social Responsibility (CSR). Obwohl allgemeingültige Definitionen noch ausstehen, herrscht weitgehend Einigkeit über die wesentlichen Inhalte (Nemat et al., 2022). CDR zielt darauf ab, gemeinsame Werte und Normen zu fördern, die Organisationen bei der Nutzung und Entwicklung digitaler Technologien und Daten leiten sollen. Das Konzept der CDR hebt die Bedeutung ethischer Überlegungen und moralischer Normen für den digitalen Bereich hervor, insbesondere in der Erstellung, im Betrieb und in der Weiterentwicklung digitaler Technologien.

Digitale Ethik bildet eine der normativen Grundlagen von CDR. Sie setzt sich mit ethischen Fragestellungen in der Datenerhebung, -nutzung, algorithmischen Entscheidungsfindung (hier eingeschlossen Künstliche Intelligenz, KI) und digitaler Infrastruktur auseinander. Diese Werte und Prinzipien fließen in die CDR-Strategien von Unternehmen ein, um sicherzustellen, dass digitale Technologien verantwortungsvoll und werteorientiert eingesetzt werden (Lautermann, 2024).

Ziel der CDR ist es, dargestellt beispielsweise im „CDR Building Bloxx Framework“ des deutschen Bundesverband Digitale Wirtschaft (BVDW) (2024) berücksichtigen, nationalen CDR Manifest, Stakeholder*innen zu identifizieren, Guidelines, wie Normen der IEEE 7000™-2021 „Addressing Ethical Concerns During Systems Design“ (Spiekermann et al., 2022) zu berücksichtigen um Prozesse zu etablieren, die digital-ethisch kritische Situationen aufzeigen, verhindern oder abmildern können. Aus der Praxis (Carl et al., 2023) haben sich zudem weitere Vorteile und Aspekte des positiven Einflusses der CDR gezeigt, wobei die Motivation hinter CDR-Maßnahmen je nach Unternehmensgröße und -typ variieren. Größere Unternehmen verbindet oft ähnliche externe Treiber wogegen KMUs eher individuell motiviert handeln. Hierbei spielen die Motivatoren der Markenstärkung, Erhöhung des Kundenvertrauens und die Absicherung von geplanten Innovationen, die intern als auch extern zum Einsatz finden sollen, durch eine vorgereifte Technikakzeptanz der Zielgruppe eine bedeutende Rolle.

Im Einklang mit dem sich entwickelnden EU-Datenrecht begegnet die CDR dem kontinuierlichen und schnellen Wandel technologischer Entwicklungen. Dabei entfaltet die CDR ihr volles Potenzial, insbesondere bei der Identifikation und Bearbeitung von „blinden Flecken“ in der Interpretation und Einhaltung regulatorischer Vorgaben wie der DSGVO, dem DMA, dem DSA oder dem AI Act. In solchen Kontexten dient die CDR als Wertekompass und als Grundlage für die Formulierung einer „digitalen DNA“. Dies ermöglicht es, praktische Herausforderungen mit rechtlichen Unschärfen in Einklang zu bringen und bietet präzise Orientierung bei der Interpretation komplexer Vorschriften.

Mit der zugespitzten Formel: *„die USA innoviert, China kopiert und die EU reguliert“*, wird, wenn vielleicht auch scherzhaft dargestellt, ein wichtiger Faktor genannt: wir wer-

den zukünftig nicht alle Fragen zum Umgang mit digitalen Technologien durch Gesetze, Verordnungen und Richtlinien beantworten können und deshalb auf die CDR als Lösung setzen müssen, um ethische Prinzipien mit technologischem Fortschritt zu vereinen und damit verantwortungsvolle Innovationen und nachhaltige digitale Transformation zu ermöglichen.

17.1.2 Ethische Verantwortung – CDR als Wegweiser

CDR wird nicht nur von den Verantwortlichen für Corporate Responsibility oder Nachhaltigkeit getragen, sondern ist ein gemeinschaftlicher Prozess, der in unterschiedlichen Bereichen des Unternehmens entsteht – darunter Produktmanagement, Innovation, Compliance, Einkauf und Marketing. Wesentlich für den Erfolg sind entsprechende CDR-Governance-Strukturen, eine kontinuierliche Messung der Wirkungen und Schulungen zur Sensibilisierung aller Mitarbeitenden. Diese Elemente bilden das Fundament, das durch die Geschäftsleitung getragen und verantwortet wird. Der Einstieg in CDR erfordert eine klare Geschäftsstrategie und das Engagement der Führungsspitze sowie Führungskräfte.

Die CDR ist ein umfassender Ansatz, der weit über die technische Dimension von Digital-, Daten- oder KI-Ethik hinausgeht und Mechanismen der Verantwortung, ethisches Handeln und Transparenz in die Organisation integriert. Gerade in einer Zeit, in der neue Technologien wie Künstliche Intelligenz und datenbasierte Anwendungen unvorhersehbare Herausforderungen mit sich bringen, liegt es in der Verantwortung der Führungskräfte, den Rahmen für verantwortungsvolles und ethisches Handeln zu schaffen. Um CDR effektiv in einem Unternehmen zu verankern, sollte das Thema nicht isoliert betrachtet, sondern in die bestehende Unternehmensstrategie integriert werden. Eine Anknüpfung an vorhandene Corporate Responsibility (CR)-Initiativen ist dabei essenziell, um digitale Verantwortung in alle Geschäftsbereiche zu tragen und bestehende Ansätze sinnvoll zu ergänzen. Sie ermöglicht es Führungskräften und Mitarbeitenden, ethisch zu handeln, eine raschere und treffsichere Entscheidungsfindung zu ermöglichen, auch in Situationen, die aktuell nicht absehbar sind und bildet die Grundlage für digitales Vertrauen in die Organisation.

Es stehen eine Reihe von Leitlinien und Rahmenwerke für das CDR-Management zur Orientierung zur Verfügung, beispielhaft ist hier das CDR Manifesto ([o. J.](#)), das in 20 Sprachen zur Verfügung steht, zu nennen. In einer internationalen Zusammenarbeit wurden sieben Prinzipien aus den bestehenden Definitionen von Corporate Digital Responsibility entwickelt, die Organisationen auf ihrem Weg zur digitalen Verantwortung begleiten sollen. Des Weiteren gibt es bereits Unternehmen, die regelmäßig CDR Berichte über die Deutsche CDR Initiative (die Initiative ist im Mai 2018 durch das seinerzeitige Bundesministerium der Justiz und für Verbraucherschutz [BMJV] gemeinsam mit Unternehmen initiiert worden) veröffentlichen (Deutsche CDR Initiative, [2024](#)), oder Benchmarking Auswertungen (Ethos Stiftung, [2023](#)), wie dies in der Schweiz durch die ETHOS Stiftung alljährlich stattfindet.

17.1.3 Aufbau eines CDR Management Systems

Die Aufgabe von CDR – und folglich der zugrunde liegenden Disziplin der Digitaethik – besteht in einer systematischen Abwägung sozialer, kultureller, ökologischer sowie wirtschaftlicher Interessen als **Frühindikatoren** und **Treiber** von Geschäftschancen und -risiken entlang der digitalen Wertschöpfungskette (Schmidpeter, 2017; Thorun, 2018). Um den **Reifegrad** der CDR im Unternehmen festzustellen, werden Verantwortungs-Cluster oder Themenblöcke gebildet, der den Stand der Aktivitäten und Maßnahmen gesammelt und aggregiert darstellen soll. Dies erfolgt in der Regel durch die Analyse von CDR-Expertinnen und wird in Anschluss mittels systematischer Befragungsmethoden interner und externer Stakeholder ergänzt. Durch Miteinbeziehen von Kommunikationsmaterialien, Nachhaltigkeitsreports oder Strategiepapiere sowie Befragung von verantwortlichen Personen in der Organisation wird ein Zwischenergebnis erarbeitet, das eine Ableitung einer Liste von bereits „aktivierten“ **CDR-Cluster** ermöglicht. Diese werden anhand des Ausprägungsgrades zur Feststellung der „**CDR-Performance**“ tiefergehender bewertet. Die bestimmte Position wird im Folgenden genutzt, um die strategischen Potenziale der CDR für das Unternehmen zu identifizieren. Die CDR-Cluster werden in Bezug auf das Kerngeschäft bewertet (sog. „**Business-Fit**“) und ist Teil der strategischen Meinungsbildung. Für jedes Handlungsfeld liegen somit Einschätzungen zum Reifegrad sowie zur chancenorientierten Passung mit dem Kerngeschäft vor. Diese Bewertungen dienen als Grundlage zur Identifikation von CDR-Potenzialen, mit denen sich strategische Wettbewerbsvorteile entwickeln lassen. Diese Vorgehensweise kann als inhaltliche Vorbereitung einer (aufwändigeren) Wesentlichkeitsanalyse für die Erweiterung von CR dienen. Die Verantwortungsbereiche der Corporate Responsibility (CR) – wie Menschenrechte, Stakeholder-Engagement, Arbeitsrechte, Umwelt, wirtschaftliche Aspekte, Transparenz, lokale Entwicklung sowie Wissenschaft und Technologie – sind auch für CDR relevant. Bestehende CR-Instrumente können als Vorlage für die Umsetzung von CDR dienen und dabei neue Ziele und Indikatoren integrieren, wenn sie mit dem nötigen Fachwissen und der Bereitschaft an die Herausforderungen des digitalen Zeitalters angepasst werden. Auf ein wesentliches Instrument im Rahmen der CR sei hier verwiesen: **digitale Selbstverpflichtungen**, die für eine wirksame Veränderung von Unternehmenshandeln bedeutsam sind (Thorun, 2018).

Unternehmen jeglicher Größe und Branchenausrichtung bietet der Modulare Aufbau des CDR-Konzepts eine schrittweise Umsetzung anhand von Themenfelder, wie Ethics-by-Design, New Work, Digital Wellbeing, Verantwortungsvolle Innovationen, u. v. m. an. Praxisnahe Trust-building-Maßnahmen (wie Code of Ethics, Einsatz eines Ethik-Rates, Digital Wellbeing-Maßnahmen für Mitarbeitende, Humane Tech, Building Ground Truth, u. v. m.) können individuell zu Umsetzungspakete zusammengestellt.

„The Law is the bare Minimum.“

Es muss jedoch klargestellt werden, dass CDR **über das gesetzlich geforderte Maß hinaus** gehen oder zur Unterstützung bei unbestimmten Rechtsbegriffen dient und somit

Unternehmen nicht nur rechtlich korrekt handeln, sondern auch digital-ethisch einwandfreies Wirtschaften entlang der „European Declaration on Digital Rights and Principles“ (European Commission, 2022) umsetzen. Hier steht die zentrale Frage „Rechtlich dürfen wir das, aber wollen wir es auch?“ im Mittelpunkt. Die CDR legt damit den notwendigen Grundbaustein eines Ethischen Kompasses für Unternehmen in der Digitalen (Wirtschafts-)Welt.

17.1.4 Digital Ethics Officer, AI Ethics Officer und CDR Manager

Die Verflechtung und Komplexität der CDR lässt auch ein neues Rollenverständnis entstehen: Wer ist intern für eine Umsetzung zuständig? Wer ist verantwortlich bei „Digitalen Ethikverstößen“? Bei Ablehnung eines digitalen Produktes, Services aufgrund von fehlenden „Digital Trust“, weil im Vorfeld die wesentlichen Fragen zur Vertrauensbildung nicht gestellt wurden? Welche (arbeitsrechtlichen) Konsequenzen kann es für Führungskräfte und Mitarbeiter bei Missachtung interner Vorschriften geben? Wer wird miteinbezogen, wenn es zur Formulierung eines Digital Ethics Kodex kommt?

„DEOs are professionals who navigate the complexities of digitalisation, ensuring ethical practices and compliance with legal frameworks. (EIPA European Institute of Public Administration)“

Auf all diese Fragen geben Best Practice Umsetzungen Antworten, es werden beispielsweise hochqualifizierte Aufsichtsgremien geschaffen, die als One-Stop-Shop Anlaufstelle bei neuen Projekten eine kontrollierende und beratende Funktion einnehmen oder die Ernennung von „Digital Ethics Officers“ (DEO) vorgesehen sowie die Ausbildung von CDR Manager vorangetrieben. Inwieweit die DEO Funktion mit jenen eines Data Protection Officer (DPO), AI Ethics Officer (AEO), Chief Sustainability Officer (CSO), Chief Compliance Officer (CCO) und Chief Information Security Officer (CISO) zusammenspielen wird, ist eine firmenindividuelle Frage der Umsetzung, was jedoch sich bereits abzeichnet ist, dass es zukünftig Fachexperten geben muss, die sich auf ethische Fragen konzentrieren, die mit der gesamten digitalen Transformation einer Organisation einhergehen.

17.2 Ausblick und Weiterentwicklung der CDR

Im Oktober 2024 befasste sich das Deutsche Institut für Normung e. V. (DIN) mittels Sondierungsworkshop mit der Frage der Notwendigkeit einer Normung der CDR und kam zum vorläufigen Ergebnis, dass neben der Weiterverfolgung eines festgestellten Normungsinteresses bestehende Standards angepasst werden könnten. Ein weiterer Schritt zur vertiefenden Etablierung der CDR in der Praxis ist das ebenfalls das im Oktober 2024 gestartete Ausbildungsprogramm der EU-Kommission erkennbar. Das im Jahr 2023 ins Leben gerufene Projekt „Digital Ethics Officer for Public Administration“ (DEO4PA) greift

ein wichtiges Bedürfnis in der heutigen, sich schnell wandelnden digitalen Landschaft auf. Während öffentliche Verwaltungen in der Europäischen Union zunehmend digitalisiert werden, besteht ein Bedarf an ethisch orientierten Fachkräften, die diesen Wandel effektiv begleiten können. Mit solchen Ausbildungsprogrammen und der Auseinandersetzung zur Digitalverantwortung im Öffentlichen Sektor wird dies zu einer zunehmenden Beachtung der CDR in der Unternehmenspraxis führen, insbesondere im Bereich der Vergabe von öffentlichen Aufträgen und durch die Vorbildwirkung der Öffentlichen Stellen, die schlussendlich zu Branchenstandards und zur Konkretisierung des „Stand der Technik“ führen kann.

Auch aus rechtlicher Sicht bleibt eine Auseinandersetzung mit der CDR ebenso spannend. Was bleibt vom CDR-Ansatz, wenn bereits hochregulierte Bereiche mit hohem Schutzniveau wenig Handlungsspielraum für „über das gesetzlich hinausgehende“ Maßnahmen lassen?

Absehbar ist, dass mit zunehmender Regulierung die Spielräume für freiwillige Regelungen kleiner werden, während gleichzeitig eine Verrechtlichung der digitalen Unternehmensverantwortung einsetzt. Die gesetzliche Normbildung geschieht zudem nicht isoliert, sondern in einem dynamischen Wechselspiel mit Stakeholdern, wodurch CDR-Erwägungen verstärkt in die legislative Willensbildung und Interessenvertretung einfließen können. Insbesondere im unionsrechtlichen Rahmen, der derzeit keine spezifische Berichterstattungspflicht zu digitalen oder CDR-bezogenen Themen vorsieht, könnte sich der Fokus der CDR künftig auf rechtliche Compliance verlagern.

Die Aufnahme wichtiger gesetzlicher Anforderungen in freiwillige CDR-Strategien unterstreicht deren besondere Bedeutung und kann gezielt auf eine verbraucherfreundliche Umsetzung abzielen. Damit erfüllt CDR zwei bedeutende rechtliche Funktionen: Einerseits ermöglicht sie **proaktive Selbstregulierung** und **Orientierung** in einem hochdynamischen Bereich, in dem noch nicht alle Fragen gesetzlich geregelt sind oder unternehmerischer Eigenverantwortung unterliegen. Andererseits können Unternehmen durch CDR-Strategien bestehende oder entstehende Lücken und **Unsicherheiten in der rechtlichen Regulierung aktiv schließen**.

Die Vernetzung zu Austausch über CDR-Maßnahmen und deren Weiterentwicklung findet innerhalb der Community regelmäßig statt. Seit der Entstehung der CDR-Community ab 2015 hat sich ein Netzwerk entwickelt, das Unternehmen in den Fokus rückt, deren **herausragende Leistungen** und zukunftsweisende Strategien in der Fachwelt öffentlich anerkannt sind. Diese Unternehmen setzen Maßstäbe für eine verantwortungsvolle und nachhaltige digitale Transformation. Ein herausragendes Beispiel dafür ist **Merck**, das als führendes Best Practice Beispiel im Zuge dieses Beitrages vorgestellt wird. Merck hat mit seiner Digitaletik-Strategie nicht nur Vorbildcharakter innerhalb der Branche bewiesen, sondern auch Impulse für die gesamte Community gesetzt. Dies zeigt sich besonders in der fortgeschrittenen Entwicklung einer Digitaletik-Strategie, die neben der Entwicklung eines **Managementsystems zur Operationalisierung** von Digitaletik – z. B. in Form von konkreten CDR Maßnahmen – auch die Erarbeitung von **KPIs** zur Erfassung von Digitaletik und CDR umfasst.

17.3 Vorstellung Praxisbeispiel Merck

17.3.1 Technologie ist nicht wertneutral und Technologiefirmen tragen moralische Verantwortungen

Eine populäre Auffassung unter Wissenschaftlern besagt, dass Technologie an sich wertneutral sei, das heißt, sie sei an sich weder gut oder schlecht. Stattdessen sei das spezifische Einsatzgebiet oder der Kontext, in dem die Technologie verwendet wird von normativer Relevanz. Slogans von Waffenbefürwortern wie „Waffen töten keine Menschen, Menschen töten Menschen“ lassen sich in dieser Denkweise problemlos auch auf komplexe digitale Technologien übertragen: Man könnte argumentieren, dass KI-Systeme aufgrund ihres Designs nicht inhärent gut oder schlecht seien; vielmehr könnten vermeintlich wertneutrale technologische Artefakte wie KI-Systeme neben ihren beabsichtigten positiven Zwecken auch für böswillige Zwecke missbraucht werden, etwa zur Verbreitung von Fehlinformationen, zur biometrischen Überwachung oder zur subliminalen Manipulation. Infolgedessen betrachten viele Technikentwickler Ethik als ein Instrument, um den angemessenen und unangemessenen *Einsatz* von Technologie zu bestimmen, während sie gleichzeitig häufig als unbegründetes Hindernis für Innovation sowie für die *Herstellung* vermeintlich wertneutraler technologischer Artefakte angesehen wird. Diese Sichtweise ist insbesondere in Unternehmenskreisen stark ausgeprägt, wo Ethik nicht nur als Hemmschuh für Innovationen, sondern oft auch als Hürde für kommerziellen Erfolg wahrgenommen wird.

Die These der Wertneutralität von Technologie und Wissenschaft erweist sich allerdings als philosophisch unzulänglich und wird in der akademischen Diskussion sowie in der Literatur überwiegend als widerlegt angesehen (vgl. Boaz, 2021; Heyndels, 2023), wobei stattdessen häufig die Auffassung vertreten wird, dass Technologie wertbeladen ist. Floridi (2023) bietet ein prägnantes Konzept zur „Wertbeladenheit“ von Technologie an. Er schlägt vor, Technologie nicht als neutral, sondern als „doppelt wertbeladen“ zu betrachten. Statt als wertfrei sollte Technologie als von positiven und negativen Werten durchzogen angesehen werden. Ein Messer mag neutral erscheinen, da sich seine positiven und negativen „Wertvektoren“ ausgleichen. Floridi verdeutlicht jedoch, dass dies vom Design abhängt: Ein stumpfes Buttermesser ist wenig gefährlich, während ein Bajonett, mit dem man zwar Nutella schwer verstreichen kann, genau dafür konzipiert ist, Menschen zu verletzen oder zu töten (ebd.). Das Bajonett ist also mit einem Wertevektor gestaltet, der den Benutzer gezielt dazu einlädt, die Klinge als Waffe einzusetzen – ein Vektor, der bei der Klinge eines Buttermessers weitgehend fehlt.

Floridis Konzept lässt sich überzeugend auf digitale Lösungen wie beispielsweise soziale Medienplattformen anwenden. Einerseits können diese Plattformen durch geeignetes Design mit positiven „Wertvektoren“ versehen werden, die effiziente Kommunikation fördern, die Vernetzung von Gemeinschaften erleichtern, die Verbreitung vielfältiger Ideen unterstützen und globale soziale Bewegungen stärken. Andererseits können dieselben Plattformen entlang negativer „Wertvektoren“ gestaltet werden, wodurch sie als Kanäle

zur Verbreitung von Fehlinformationen dienen, Polarisierung verstärken, suchtartiges Verhalten fördern und potenziell die Privatsphäre der Nutzer gefährden (Nemat & Charton, 2024).

Die Auffassung von Technologie als wertbeladene Artefakte macht deutlich, dass ethische Überlegungen in jeder Phase des technologischen Lebenszyklus und des Designprozesses von zentraler Bedeutung sind – von der Konzeption bis zur Umsetzung und eben nicht nur in der späteren Anwendung. Besonders entscheidend ist dies für Entwickler und Wissenschaftler, die sich in dynamischen Technologiefeldern bewegen, die oft durch regulatorische Unklarheiten und moralische Komplexitäten gekennzeichnet sind. Dies begründet weitreichende moralische Verantwortungen für Technologieentwickler, insbesondere im Unternehmenskontext, da gerade Technologien, die über den Laborstatus hinausgehen, eine erhebliche Wirkung entfalten können, sobald sie erfolgreich vermarktet und von einer großen Nutzerbasis verwendet werden. Die Einbindung digitaler Ethik in die Technologieentwicklung und in unternehmerische Entscheidungsprozesse fördert nicht nur verantwortungsvolle Innovation, sondern ist auch unerlässlich, um der moralischen Verantwortung von Wissenschaftlern und Führungskräften in Technologiekonzernen gerecht zu werden.

17.3.2 Ausgangslage bei Merck

Das Management der moralischen Herausforderungen von Technologie ist besonders in einem Unternehmenskontext komplex. Große Firmen nutzen eine Vielzahl an Technologien und Anwendungsfeldern, was die Komplexität weiter erhöht, insbesondere durch die vielschichtigen Projektinfrastrukturen. Ein zusätzliches Problem ist die Wissenslücke in Bezug auf ethische Expertise in technikfokussierten Teams, die oft durch potenzielle Interessenkonflikte verstärkt wird, etwa zwischen kommerziellen Interessen der Manager und wissenschaftlicher Neugierde oder Ehrgeiz. Eine weitere Herausforderung besteht darin, zu verstehen, dass Ethik nicht nur das Verbot bestimmter Aktivitäten beinhaltet, sondern auch die Notwendigkeit, übermäßige Vorsicht zu vermeiden, wenn sie nicht gerechtfertigt ist. Im Unternehmenskontext dient Ethik also auch dazu, die Vorteile und Beiträge von Technologien zur menschlichen Entfaltung zu fördern. Diese komplexe Aufgabe erfordert eine systematische Herangehensweise.

Seit der Einrichtung eines externen Ethik-Beratergremiums im Jahr 2011 und der Einführung interner Strukturen zur Umsetzung der Empfehlungen und zur Operationalisierung von Ethik-Leitlinien übernimmt Merck KGaA, Darmstadt, Deutschland, als führendes Unternehmen in den Bereichen Gesundheitswesen, Life Sciences und Elektronik eine verantwortungsvolle Rolle. Das Unternehmen kann auf eine langjährige Tradition der systematischen Integration ethischer Überlegungen in seinen Geschäftspraktiken zurückblicken. Dieses Engagement manifestiert sich insbesondere in der intensiven Zusammenarbeit eines internen Ethik-Teams mit externen Beratungsgremien, die unabhängige Expertise zu ethisch anspruchsvollen Fragestellungen wie der Stammzellforschung und

der Genom-Editierung bereitstellen. Die gewonnenen Erkenntnisse flossen bereits in die Ausarbeitung und Weiterentwicklung verschiedener Unternehmensgrundsätze und -richtlinien ein (vgl. Sugarman et al., 2018).

Als ein multinationales Wissenschafts- und Technologieunternehmen hat Merck, wie auch andere Unternehmen, erhebliche digitale Veränderungen und die damit verbundenen Herausforderungen in Bezug auf seine internen Abläufe sowie die Art und Weise, wie es sein Geschäft entwickelt und betreibt, erfahren. Diese Entwicklungen haben zu ethischen Fragestellungen geführt, die durch den etablierten bioethischen Ansatz bei Merck nicht immer zufriedenstellend adressiert werden können (Becker et al., 2023). Prinzipien, die in der praktischen Ethik eine pragmatische Rolle bei der Lösung moralischer Probleme spielen, erhalten im Kontext von Daten und Algorithmen oft eine neue Bedeutung. Ein Beispiel dafür ist der Datenschutz: Im Zusammenhang mit KI-Anwendungen und Big Data wird der Schutz der Privatsphäre zunehmend zu einem kollektiven Handlungsproblem. Dabei kann die Nutzung persönlicher Daten nicht nur individuelle Chancen eröffnen, sondern auch Risiken für andere, potenziell große Personengruppen, mit sich bringen (vgl. Véliz, 2020). Ähnlich verhält es sich mit dem Prinzip der Transparenz, das in der Bioethik traditionell als wesentlicher Faktor zur Sicherung des Autonomieprinzips betrachtet wird. Im Kontext der Datenethik und der Ethik von KI bedarf es jedoch einer eigenständigen Reflexion. Während die Informationsasymmetrie im Arzt-Patient-Verhältnis ein zentraler Aspekt für das Autonomieprinzip in der Medizinethik ist, reicht dieser Ansatz nicht aus, um die Bedeutung von Transparenz in komplexen Datenökosystemen und algorithmischen Systemen vollständig zu erfassen. Diese Systeme bleiben größtenteils unsichtbar und unverständlich für die meisten Menschen, haben jedoch weitreichende Auswirkungen auf zahlreiche Individuen (Becker et al., 2023).

Diese Asymmetrien und Besonderheiten der Ethik im digitalen Kontext haben zunehmend akademische Beachtung gefunden. Aufbauend auf den Grundlagen der Computer- und Informationsethik, die seit der zweiten Hälfte des 20. Jahrhunderts entwickelt wurden (Moor, 1985; Bynum, 2019; Floridi, 2013), hat sich zuletzt das aufstrebende Feld der Datenethik (Floridi & Taddeo, 2016) systematisch mit den ethischen Herausforderungen auseinandergesetzt, die durch die Datenexpansion und die damit verbundenen algorithmischen Fortschritte entstehen (Becker et al., 2023). Mittlerweile hat sich die Digitalethik als eigenständige Disziplin etabliert und verzeichnet eine wachsende Zahl speziell dafür eingerichteter Lehrstühle und akademischer Fachjournale.

Der entscheidende Impuls für Merck, sich systematisch mit digitaler Ethik auseinanderzusetzen, kam 2019 durch die Gründung von Syntropy, einer Partnerschaft zwischen Merck und Palantir Technologies Inc. Dieses Projekt, das sich auf die Integration von Gesundheitsdaten aus verschiedenen Quellen spezialisiert und Wissenschaftlern sowie Forschungszentren Zugang zu einer kollaborativen Technologieplattform für die Krebsforschung bietet, stellte das Unternehmen vor bislang beispiellose ethische Fragen, die einen systematischen Ansatz erforderten. Als Reaktion auf die ethischen Herausforderungen, die sich speziell aus diesem Projekt ergaben, aber auch im Hinblick auf allgemeinere und weitreichendere Herausforderungen der Digitalisierung sowie der Verwendung von

Daten und KI, hat Merck proaktiv ein digitales Ethikmandat eingeführt. Dieses Mandat begründet Digitalethik als beratende Instanz, vergleichbar mit den etablierten bioethischen Strukturen im Unternehmen, und spielt eine zentrale Rolle bei der Entwicklung und Implementierung ethischer Richtlinien im digitalen Bereich. Es bietet einen strukturierten Rahmen für die Bewertung und Lösung ethischer Fragen, die sich aus der zunehmenden Digitalisierung und dem Einsatz von Daten und KI ergeben. Durch die Zusammenarbeit mit verschiedenen Abteilungen soll so gewährleistet werden, dass ethische Überlegungen von Anfang an in alle relevanten Prozesse und Entscheidungen integriert werden. Dieses Mandat unterstreicht so das Engagement des Unternehmens für einen verantwortungsvollen Umgang mit Daten, Algorithmen und Künstlicher Intelligenz und adressiert spezifische ethische Risiken im Zusammenhang mit diesen Technologien. Ziel ist es, die kommerziellen und wissenschaftlichen Chancen, die der technologische Fortschritt bietet, verantwortungsvoll zu nutzen. Im Rahmen dieser Initiative hat sich Merck verpflichtet, sich in der sich schnell entwickelnden Landschaft von Daten und KI zu orientieren und diese Entwicklungen ethisch fundiert zu gestalten.

Das vorliegende Mandat für Digitale Ethik wurde unabhängig von bestehenden Corporate-Responsibility-Initiativen entwickelt und wird folglich bei Merck nicht als CR-Bestandteil erachtet. Vielmehr versteht das Ethikteam – das weitgehend autonom neben den CR-Aktivitäten agiert – diese Aufgabe als eine wissenschaftliche Untersuchung der normativen Grundlagen für den Umgang mit Daten und Algorithmen. Diese Tätigkeit baut auf den bestehenden Aufgaben in der Bio- und Technikethik auf, bei denen das Team ethisch anspruchsvolle Fragestellungen in rechtlich unklaren und moralisch komplexen Szenarien untersucht – etwa bei der Bewertung neuartiger embryoähnlicher Strukturen oder innovativer Technologien in der Reproduktionsmedizin. In dieser Rolle fungiert das Ethikteam als unabhängige beratende Instanz für die Geschäftseinheiten, unterstützt sie bei der Umsetzung ethischer Leitlinien und berät bei der Implementierung von Richtlinien. Die Verbindlichkeit der empfohlenen Maßnahmen variiert je nach technologischer Komplexität, moralischer Tragweite und Relevanz für das Unternehmen. Diese Maßnahmen reichen von verbindlichen Handlungsdirektiven für den Umgang mit Stammzellen und Genom-Editierungstools für die gesamte Firma bis hin zu subtileren Etikette-Regeln für kleinere Teams oder zur ethischen Absicherung bei der Einführung spezifischer Produkte. Bei all diesen Aktivitäten ergeben sich sowohl Überschneidungen mit bestehenden Corporate-Responsibility-Strukturen als auch Aufgaben, die üblicherweise unter der Bezeichnung CDR subsumiert werden, wie etwa die Konzeption und Durchführung von Schulungen zur Sensibilisierung der Belegschaft.

Ethik wird hier als etwas verstanden, das über das hinausgeht, was durch rechtliche Rahmenwerke erfasst werden kann. Das erklärt, warum die digitale Ethik – also die fortlaufende normative Bewertung bestehender und neu entstehender digitaler Lösungen, auch im Hinblick auf entstehende regulatorische Rahmenwerke – einen eigenständigen und unersetzlichen Wert hat, den Gesetze allein nicht bieten können. Ethik liefert Einsichten, die über regulatorische Vorgaben hinausgehen, und adressiert moralische Nuancen, die nicht – bzw. noch nicht – durch Gesetze erfasst werden. Sie fördert ver-

antwortungsbewusstes Handeln und schafft eine Kultur der Verantwortlichkeit, die sich an gesellschaftlichen Werten und technologischen Entwicklungen orientiert, um ihre Relevanz in einem dynamischen Umfeld zu bewahren. Zudem kann Ethik Möglichkeiten aufzeigen, wie Technologien zum sozialen Wohl genutzt werden können, und Innovationen in positive Richtungen lenken, die Gesetze allein nicht unbedingt anregen würden. Ethische Rahmenwerke tragen auch dazu bei, Ängste zu mindern und fundierte Entscheidungen zu unterstützen. Schließlich stärken transparente ethische Praktiken, wie bereits erwähnt, das öffentliche Vertrauen in Technologien wie KI und fördern so eine vertrauenswürdige Entwicklung und Anwendung. Die Arbeit des Ethikteams bildet einen wesentlichen Grundpfeiler der CDR-Initiativen und liefert die fundamentalen normativen Leitlinien, die auf verschiedenen Ebenen des Unternehmens implementiert werden müssen. Diese Leitlinien erfordern jedoch einen anderen Fokus und eine abweichende Ebene der Verbindlichkeit, da es hier darum geht, Governance-Strukturen zu etablieren, die die ethischen Grundlagen in die konkreten Unternehmensstrukturen und -prozesse integrieren.

17.3.3 Die ethischen Grundlagen: DEAP – CoDE – PaRA

Um den ethischen Herausforderungen systematisch begegnen zu können, wurde ein zweigleisiger Ansatz verfolgt. Aufbauend auf der langjährigen Zusammenarbeit mit externen, unabhängigen Ethikberatern erschien es sinnvoll, externes Expertenwissen in die ethischen Überlegungen innerhalb von Merck zu integrieren. Im Rahmen der Entwicklung einer umfassenden Strategie für digitale Ethik etablierte Merck das Digital Ethics Advisory Panel (DEAP), das aus externen Experten in den Bereichen Digialethik, Recht, Big Data, digitale Gesundheit, Datenverwaltung und Patientenvertretung besteht. Zu den wesentlichen Aufgaben des DEAP, wie in dessen Satzung festgelegt, gehören: (1) die Beratung zu spezifischen ethischen Fragen im Zusammenhang mit Daten und algorithmischen Systemen innerhalb von Merck; (2) die Bewertung von Vorschlägen zur Minderung ethischer Risiken, die von Geschäftseinheiten unterbreitet werden, sowie die Integration von Empfehlungen des Panels und die Rückmeldung zur Implementierung; (3) die Bereitstellung einer Plattform zur Diskussion und Bewertung neuer Unternehmensrichtlinien mit ethischen Implikationen sowie die Überprüfung bestehender ethischer Prinzipien und Richtliniendokumente von Merck; und (4) die proaktive Beratung von Merck zu relevanten Entwicklungen und aufkommenden Diskursen im Bereich der digitalen Ethik (Nemat & Charton, 2024).

In seinen ersten Tätigkeiten konzentrierte sich das DEAP auf die Ausarbeitung eines umfassenden Digialethikkodex („Code of Digital Ethics“ bzw. CoDE), der auf einem Rahmenwerk von 20 handlungsleitenden Prinzipien basiert (Becker et al., 2023). Dieser Prozess umfasste eine rigorose Analyse wissenschaftlicher Beiträge zur digitalen Ethik sowie eine umfassende Überprüfung bestehender Ethikrichtlinien. Im ersten Schritt erfolgte die Auswahl relevanter Literatur anhand vordefinierter Kriterien sowie die Identifizie-

rung und Konsolidierung der in diesen Dokumenten enthaltenen Themen und ethischen Prinzipien mittels eines rekonstruktiven sozialwissenschaftlichen Forschungsansatzes. Im zweiten Schritt des Projekts übergeordnete Prinzipien definiert und die verbleibenden Prinzipien diesen in untergeordnete Gruppen zugeordnet. Im dritten und letzten Schritt wurden alle Prinzipien in die Richtlinien übersetzt, die den operationalisierbaren Kern des CoDE bilden (ebd.). Ziel war es, ein anpassungsfähiges, musterbasiertes Instrument zu entwickeln, das im gesamten digitalen Tätigkeitsfeld von Merck effektiv operationalisiert werden kann.

Der CoDE stützt sich auf fünf grundlegende Prinzipien: Gerechtigkeit, Autonomie, Fürsorge, Schadensvermeidung und Transparenz. Diese Prinzipien bieten eine strukturierte Grundlage zur Analyse und Bewertung ethischer Fragestellungen und dienen sowohl den verschiedenen Unternehmensbereichen als auch den einzelnen Mitarbeitenden als Orientierungshilfe in komplexen Situationen, in denen noch keine gesetzlichen oder anderweitigen Regulierungen existieren. Der Kodex fungiert dabei nicht nur als interne Leitlinie, sondern auch als externes Bekenntnis gegenüber Kunden und Geschäftspartnern, indem er die ethischen Handlungsprinzipien festlegt, auf die man sich bei Merck verlassen kann, und die Erwartungen formuliert, die Merck an seine Partner richtet.

Nach der Erarbeitung des CoDE übernahm das DEAP eine doppelte Aufgabe: Zum einen sollte es die Einhaltung der Prinzipien in den eigenen Tätigkeiten sicherstellen, zum anderen ihre umfassende Implementierung im gesamten Unternehmen gewährleisten. Ersteres ist entscheidend, um dem Eindruck von Beliebigkeit bei der Auswahl und Anwendung ethischer Prinzipien entgegenzuwirken und sogenanntes „Ethics Shopping“ zu vermeiden – also die selektive Anpassung oder Übernahme ethischer Prinzipien, Richtlinien oder Rahmenwerke, um bestehende Verhaltensweisen nachträglich zu rechtfertigen, anstatt diese als Orientierung zur Verbesserung neuer Praktiken zu nutzen (Florida, 2019). Um die Sitzungen des DEAP effizient zu gestalten und die Prinzipien des CoDE zu priorisieren, hat Merck daher Strukturen entwickelt, die ethische Prinzipien systematisch mit der Arbeit des Ethik-Panels verknüpfen. Hierzu wurde ein standardisierte Risikobewertungsinstrument, die „Principle-at-Risk-Analyse“ (PaRA) entwickelt, um einen entscheidenden Teil der sogenannten „Operationalisierungs-Kluft“ (Schiff et al., 2021) zu schließen, nämlich die Rolle des Ethik-Panels sowohl als Empfänger als auch als Treiber der Operationalisierung digitaler Ethik (Nemat et al., 2023).

Das PaRA-Tool erfüllt dabei zwei Hauptaufgaben: Erstens dient es der standardisierten Identifikation von Prinzipienkonflikten und der Vorbereitung von Themen für das DEAP, indem es strukturierte Fragebögen bereitstellt, die Hintergrundinformationen zu den einzelnen Anwendungsfällen sammeln. Es überprüft systematisch, ob und wie die Prinzipien des CoDE und die entsprechenden Richtlinien durch das jeweilige Problem betroffen sind. Durch diesen standardisierten Prozess können ethische Fragen, die für den jeweiligen Fall relevant sind, konsistent und zuverlässig identifiziert werden. Zweitens liefert das PaRA klare Leitlinien und Strukturen für die Diskussionen im DEAP in Übereinstimmung mit dem CoDE. Durch die explizite Verknüpfung der anstehenden Themen mit den Prinzipien des CoDE ermöglicht der PaRA-Bericht konstruktive Diskussionen, die

schnell die für spezifische Probleme verantwortlichen Parteien und mögliche Maßnahmen zur Risikominderung identifizieren. Dies ist besonders wichtig für eine erfolgreiche Operationalisierung und die Vermeidung von Beliebigkeit oder „Ethics Shopping“ sowie der Diffusion von Verantwortlichkeiten (Floridi, 2019; Morley et al., 2021; Mökander et al., 2022).

17.3.4 Operationalisierung von Digialethik bei Merck

Die systematische Operationalisierung von Digialethik stellt Unternehmen vor besondere, zusätzliche Herausforderungen. Viele Fragestellungen erfordern weiterhin eine tiefgehende Analyse, möglicherweise unter Einbeziehung von externen Ethik-Gremien. Merck stößt jedoch immer wieder auf Situationen, in denen der traditionelle Ansatz ethischer Überlegungen nicht umsetzbar ist, beispielsweise im Bereich der Datenanalyse oder generell in Bereichen mit einer hohen Anzahl von Daten- oder KI-Anwendungen (Charton et al., 2025). Diese Bereiche erstrecken sich über alle Unternehmensabteilungen, von der Forschung und Entwicklung bis hin zu den Personalabteilungen.

Eine zentrale Herausforderung ergibt sich aus der überwältigenden Anzahl kontinuierlich entwickelter Anwendungsfälle, insbesondere in der Datenanalyse, wo manchmal mehrere tausend Fälle oder Projekte parallel bearbeitet werden. Dies macht eine individuelle ethische Bewertung durch Ethikabteilungen oder externe Beratergremien wie das DEAP äußerst komplex oder sogar unmöglich. Es verdeutlicht die dringende Notwendigkeit, Digialethik in einem solchen Umfeld so zu operationalisieren, dass Anwendungsfälle ressourceneffizient und zuverlässig klassifiziert werden können, etwa durch risikobasierte Kategorisierungsmethoden (Mökander et al., 2023).

Praktische Herausforderungen bei der ethischen Bewertung ergeben sich auch aus der dynamischen Natur des Projektmanagements, die durch inkrementelle und iterative Prozesse in sämtlichen Phasen gekennzeichnet ist. Häufige Änderungen in der Projektstrategie, den Zielen oder den Projektparametern, wie etwa in der Methodik oder den verwendeten Daten, können die ethische Bewertung komplizieren und zu problematischen Fällen von „Scope Creep“ führen, bei denen KI-Systeme in Kontexten eingesetzt werden, die von dem ursprünglich vorgesehenen Anwendungsfall abweichen (Danks & London, 2017; Aizaz et al., 2021). Dies erfordert umfassende und kontinuierliche ethische Bewertungen, die flexibel an sich verändernde Bedingungen angepasst werden müssen, da einmalige Untersuchungen in solchen Fällen oft unzureichend sind.

Diese Herausforderungen verdeutlichen, dass Methoden wie eine risikobasierte Kategorisierung von Anwendungsfällen im Sinne der Ressourceneffizienz von den verantwortlichen Projektteams durchgeführt werden müssen. Dies verschärft jedoch die genannten Probleme durch den offensichtlichen Mangel an Fachwissen im Bereich der Daten- und KI-Ethik. Hinzu kommen Interessenkonflikte und abweichende moralische Einschätzungen, die von lokalen Normen und Sensibilitäten beeinflusst werden, insbesondere bei Teams, die über geografisch verteilte Standorte arbeiten (Charton et al., 2025). Die de-

zentrale Natur vieler Anwendungsfälle kann es zudem erschweren, ethische Probleme zu erkennen und Verantwortlichkeiten klar zu definieren, insbesondere wenn einzelne Entscheidungen oder Verhaltensweisen ethisch unbedeutend erscheinen. Wenn solche Zuweisungen unklar sind, besteht das Risiko, dass die Verantwortung verwischt und letztlich niemand die Verantwortung übernimmt – das Risiko, dass „jedermanns Problem zu niemandes Verantwortung wird“ (Floridi & Sanders, 2004). Daher muss eine effiziente Operationalisierung der Digialethik auch Fragen der Kompetenzentwicklung und Verantwortlichkeit adressieren, die über einmalige Schulungen hinausgehen.

Entwicklung und Einführung des Merck Digital Ethics Check

Ein Beispiel dafür, wie Merck die Herausforderungen der Operationalisierung digitaler Ethik angeht, zeigt die Entwicklung und Implementierung des „Merck Digital Ethics Check“ (MDEC). Dieses neuartige Instrument wurde speziell entwickelt, um die ethischen Herausforderungen im Bereich der Datenanalyse zu adressieren und stellt eine halbautomatisierte Integration digitaler Ethik in den Bereich Datenanalyse dar (Charton et al., 2025). Das Tool wurde ursprünglich für das „Life Sciences Analytics Center of Excellence“ (ACE) von Merck maßgeschneidert, dient jedoch darüber hinaus als praktischer Leitfaden für Organisationen, die ethische Überlegungen in ihre digitalen Prozesse einbetten möchten. Das ACE treibt im Wesentlichen die globale Daten- und Analysestrategie der Life-Science-Sparte voran und fungiert als Serviceeinheit. Der Fokus liegt auf der unternehmensweiten Überwachung und Bereitstellung sicherer, qualitativ hochwertiger und zeitnaher Daten und Anwendungen, die datengetriebenes Entscheiden und Denken unterstützen. Das Portfolio von ACE umfasst derzeit über 1000 Anwendungsfälle, die sich in ständiger Entwicklung befinden und zentral über eine Datenintegrationsplattform verwaltet und bearbeitet werden. Die vom ACE verwalteten Anwendungsfälle sind vielfältig und können Themen wie Prozessoptimierung, Qualitätsverbesserung oder Kosteneinsparungen ansprechen. Ein Beispiel ist die Entwicklung des COMPASS-Tools, einer Datenplattform, die Vertriebsteams dabei unterstützt, den richtigen Kunden zur richtigen Zeit zu finden. Basierend auf der Analyse großer Mengen komplexer Daten aus früheren Kundeninteraktionen nutzt das Tool maschinelles Lernen, um vorherzusagen, welche potenziellen Kunden nach Geschäftspartnern suchen, um Teile ihrer Entwicklungs- und Herstellungsprozesse auszulagern.

Im Kontext von ACE können sich moralische Spannungsfelder ergeben, die einer sorgfältigen Auseinandersetzung bedürfen, wie etwa die Sensibilität der erfassten Kundeninformationen, die Transparenz gegenüber den Kunden hinsichtlich ihres Umgangs mit datengetriebenen Tools oder potenzielle Verzerrungen, die zu einer unfairen Behandlung bestimmter Kundensegmente führen könnten. Diese ethischen Herausforderungen, in Kombination mit der Vielzahl an Projekten, der komplexen Struktur des Projektmanagements und der fehlenden ethischen Kompetenz innerhalb der ACE-Projektteams, machten die Entwicklung und Implementierung eines strukturierten digitalen Ethikrahmens erforderlich. Dieser Rahmen soll eine reibungslose Integration in bestehende Arbeitsabläufe ermöglichen und sicherstellen, dass (1) Risiken ressourceneffizient erkannt werden, um die

operative Belastung und den bürokratischen Aufwand für die Mitarbeiter zu minimieren, (2) Risiken über den gesamten Projektlebenszyklus hinweg kontinuierlich bewertet werden und (3) auch von ethisch ungeschultem Personal eingesetzt werden können, wobei er gleichzeitig zur Weiterentwicklung der ethischen Kompetenz dieses Personals beiträgt.

Der MDEC ist nahtlos in die Projektplattform integriert und erfolgt in zwei Phasen: Zunächst wird der Anwendungsfall halbautomatisiert kategorisiert, was in der Berechnung eines Gesamtrisiko-Scores resultiert. Auf Basis dieses Scores wird das Projektteam in der zweiten Phase dazu angehalten, einen durch das Tool gesteuerten Reflexionsprozess durchzuführen, um potenzielle ethische Risiken zu identifizieren und gezielt anzugehen. Jeder Anwendungsfall wird in zwei verschiedenen Projektstadien mit dem MDEC bewertet, um den spezifischen Dynamiken des Projekts Rechnung zu tragen. Zentral für die halbautomatisierte Klassifizierung ist ein Multiple-Choice-Fragebogen, der vier kritische Kategorien abdeckt: Geschäftseigenschaften, Stakeholder, Daten und Analyse. Die Antworten auf diese Fragen liefern die erforderlichen Daten zur Berechnung des Gesamtrisiko-Scores des jeweiligen Anwendungsfalls.

Die zugrunde liegende Logik und die entsprechenden Prozesse wurden ausführlich an anderer Stelle beschrieben (Charton et al., 2025). Kurz zusammengefasst: Die Prinzipien des CoDE wurden im MDEC in sogenannte „Ethikmarker“ überführt, die gezielt auf die ethischen Herausforderungen von ACE im Kontext der Datenanalyse abgestimmt sind. Diese Ethikmarker werden in messbare „Observables“ übersetzt, die von den Projektverantwortlichen leicht quantifiziert und in den MDEC-Fragebogen integriert werden können.

Ein anschauliches Beispiel ist das Prinzip der „Nachverfolgbarkeit“ (Traceability). In den Datenanalyseprojekten von ACE spielt die Nachverfolgbarkeit eine wichtige Rolle, da sie eng mit der Integrität der Daten verbunden ist. Um dieses Prinzip in die Praxis umzusetzen, werden die Herkunft der Daten und der eingesetzte Machine-Learning-Ansatz untersucht und in den Ethikmarkern „Machine-Learning-Ansatz“ und „Datenquelle“ quantifiziert. Die Wahl des Machine-Learning-Verfahrens ist von Bedeutung, da die Komplexität der Nachverfolgbarkeit je nach Methode variiert. Der gewählte Machine-Learning-Ansatz dient somit als Indikator für die Relevanz des Nachverfolgbarkeitsprinzips. Diese Modelle fungieren als messbare „Observables“ im MDEC-Fragebogen. Die Nachverfolgbarkeit wird auch durch die Datenherkunft beeinflusst. Extern bezogene Daten werden in der Regel als kritischer angesehen, was die ethischen Überlegungen bei deren Nutzung intensiviert. Der MDEC differenziert daher zwischen „extern produzierten Daten“ und „intern produzierten Daten“, um den Grad der Nachverfolgbarkeit adäquat zu bewerten. Basierend auf diesen „Observables“ berechnet der MDEC einen spezifischen Risikoscore für jedes ethische Prinzip des CoDE. Dieser Score quantifiziert das ethische Risiko und klassifiziert den Use Case als niedrig, mittel oder hoch riskant, wobei die Klassifizierung anhand empirisch ermittelter Schwellenwerte erfolgt.

In der zweiten Phase des MDEC liegt der Schwerpunkt auf der Implementierung wirksamer Maßnahmen, um Risiken abzuwenden. MDEC erstellt auf Grundlage der Ergebnisse der ersten Phase einen umfassenden Bericht, der die angewandte Methodik, die erzielten Resultate und die entsprechenden Empfehlungen detailliert zusammenfasst. Bei

mittlerem oder hohem Risiko erfordert diese Phase eine proaktive Risikoanalyse durch einen moderierten Selbstreflexions-Workshop, der anhand vorgegebener Fragen durchgeführt wird. Bei mittlerem Risiko wird das interne ACE-Steering Committee eingebunden, während bei hohem Risiko Vertreter der Ethikabteilung hinzugezogen werden, um Lösungen zu erarbeiten und deren Implementierung zu begleiten. Die CoDE-Prinzipien bilden die Grundlage der Selbstbewertung und zielen darauf ab, das Projektteam zu einer tiefgehenden ethischen Reflexion ihrer Projekte anzuregen. Ergänzende Leitlinien unterstützen das Team dabei, ethische Überlegungen praktisch umzusetzen.

17.3.5 Ausblick

Abschließend lässt sich festhalten, dass die Umsetzung von Digialethik im Unternehmenskontext nicht nur ein Lippenbekenntnis sein darf, sondern tief in die Strukturen und Prozesse eines Unternehmens integriert werden muss. Es reicht nicht aus, allgemeine Prinzipien zu formulieren – diese müssen in konkrete, verbindliche Handlungsanweisungen für die Praxis überführt werden. Nur so können Unternehmen sicherstellen, dass ethische Standards tatsächlich eingehalten und in der täglichen Arbeit gelebt werden. Dies erfordert nicht nur ein starkes Engagement des Managements, sondern auch die Schaffung von Rahmenbedingungen, die die gesamte Belegschaft einbindet und sensibilisiert. Die Entwicklung und kontinuierliche Anpassung von Instrumenten wie PaRA und MDEC zeigt, dass es Wege gibt, ethische Fragestellungen greifbar zu machen. Doch der Schlüssel zum Erfolg liegt in der Bereitschaft, diese Ansätze weiterzuentwickeln und auf die spezifischen Anforderungen verschiedener Abteilungen und Anwendungsfelder zuzuschneiden. Nur ein solcher systematischer Ansatz kann langfristig sicherstellen, dass digitale Ethik nicht nur theoretisch verankert ist, sondern auch praktisch umgesetzt wird – zum Vorteil des Unternehmens, seiner Stakeholder und der Gesellschaft.

Literatur

- Aizaz, F., Khan, S., Khan, J. A., Rehman, I., & Akhunzada, A. (2021). *An empirical investigation of factors causing scope creep in agile global software development context: a conceptual model for project managers*. IEEE.
- Becker, S. J., Nemat, A. T., Lucas, S., et al. (2023). A code of digital ethics: laying the foundation for digital ethics in a science and technology company. *AI & Society*, 38, 2629–2639.
- Boaz, M. (2021). Is technology value-neutral? *Science, Technology, and Human Values*, 46(1), 53–80.
- Bundesverband Digitale Wirtschaft (BVDW) e.V. CDR Building Bloxx. <https://www.cdr-building-bloxx.com/>. Zugriffen: 6. Nov. 2024.
- Bynum, T. (2019). Computer and Information Ethics. In E. N. Zalta (Hrsg.), *The Stanford Encyclopedia of Philosophy (Summer 2019 Edition)*.

- Carl, K. V., Kubach, M., & Mihale-Wilson, C. (2023). The motivation of companies to implement corporate digital responsibility activities voluntarily: An empirical assessment. In *Wirtschaftsinformatik 2023*. Paper No. 50. <https://aisel.aisnet.org/wi2023/50>.
- Charton, J. E., Lucas, S., Vogd, W., Halter, W., & Becker, S. J. (2025). Operationalising digital ethics: establishment of an ethics evaluation tool for data analytics. *AI & Society*, <https://doi.org/10.1007/s00146-025-02621-2>.
- Danks, D., & London, A. J. (2017). *Algorithmic bias in autonomous systems*. Twenty-Sixth International Joint Conference on Artificial Intelligence.
- Deutsche CDR Initiative <https://cdr-initiative.de/>. Zugegriffen: 8. Nov. 2024.
- Ethos Stiftung (2023). Ethos Studie 2023, Digitale Verantwortung der Unternehmen im SMI Expanded. <https://www.ethosfund.ch/fr/cdr>. Zugegriffen: 6. Nov. 2024.
- European Commission (2022). European Declaration on Digital Rights and Principles. <https://digital-strategy.ec.europa.eu/en/library/european-declaration-digital-rights-and-principles>. Zugegriffen: 6. Nov. 2024.
- Floridi, L. (2013). *The ethics of information*. Oxford University Press.
- Floridi, L. (2019). Translating principles into practices of digital ethics: five risks of being unethical. *Philosophy & Technology*, 32(2), 185–193.
- Floridi, L. (2023). On good and evil, the mistaken idea that technology is ever neutral, and the importance of the double-charge thesis. *Philosophy & Technology*, 36, 60.
- Floridi, L., & Sanders, J. (2004). On the Morality of Artificial Agents. *Minds and Machines*, 14, 349–379.
- Floridi, L., & Taddeo, M. (2016). What is data ethics? *Philosophical Transactions of the Royal Society A: Mathematical, Physical and Engineering Sciences*, 374(2083), 1–5.
- Heyndels, S. (2023). Technology and neutrality. *Philosophy & Technology*, 36(4), 75.
- Lautermann, C. Corporate Digital Responsibility. IÖW Impulse; No. 6. 2024. https://www.ioew.de/fileadmin/user_upload/DOKUMENTE/Publikationen/2024/IOEW-Impulse_6_Allianz_fuer_CDR.pdf. Zugegriffen: 6. Nov. 2024.
- Mökander, J., Axente, M., Casolari, F., & Floridi, L. (2022). Conformity assessments and post-market monitoring: a guide to the role of auditing in the proposed European AI regulation. *Minds & Machines*, 32, 241–268.
- Mökander, J., Sheth, M., Watson, D. S., & Floridi, L. (2023). The switch, the ladder, and the matrix: models for classifying AI systems. *Minds & Machines*, 33, 221–248.
- Moor, J. H. (1985). What is computer ethics? *Metaphilosophy*, 16(4), 266–275.
- Morley, J., Elhalal, A., Garcia, F., Kinsey, L., Mökander, J., & Floridi, L. (2021). Ethics as a service: a pragmatic operationalisation of AI ethics. *Minds & Machines*, 31(2), 239–256.
- Nemat, A. T., & Charton, J. E. (2024). Ethics versus technological advancements: partners in crime. In N. Siragusa & A. T. Nemat (Hrsg.), *Bioconvergence revolution*. Springer.
- Nemat, A. T., Panzer-Heemeier, A., & Meckenstock, C. (2022). ESG 2.0 – Digitale Ethik als neue Dimension der Nachhaltigkeit. *ESG*, 4, 104–109.
- Nemat, A. T., Becker, S. J., Lucas, S., Thomas, S., Gadea, I., & Charton, J. E. (2023). The principle-at-risk analysis (paRA): operationalising digital ethics by bridging principles and operations of a digital ethics advisory panel. *Minds & Machines*, 33, 737–760.
- Schiff, D., Rakova, B., Ayesh, A., Fanti, A., & Lennon, M. (2021). Explaining the principles to practice gap in AI. *IEEE Technology and Society Magazine*, 40(2), 81–94.
- Schmidpeter, R. (2017). Digitalisierung – die schöpferische Kraft der Zerstörung mit Verantwortung managen. In A. Hildebrandt & W. Landhäußer (Hrsg.), *CSR und Digitalisierung* (S. 595–602). Berlin: Springer.
- Spiekermann, S., Krasnova, H., Hinz, O., Baumann, A., Benlian, A., Gimpel, H., et al. (2022). Values and Ethics in Information Systems. *Bus Inf Syst Eng*, 64, 247–250.

- Sugarman, J., Shivakumar, S., Rook, M., Loring, J. F., Rehmann-Sutter, C., Taupitz, J., Reinhardt-Rupp, J., & Hildemann, S. (2018). Ethical considerations in the manufacture, sale, and distribution of genome editing technologies. *The American Journal of Bioethics*, 18(8), 3–6.
- The International CDR Manifesto. <https://corporatedigitalresponsibility.net/cdr-manifesto>. Zugegriffen: 6. Nov. 2024.
- Thorun, C. (2018). Corporate Digital Responsibility: Unternehmerische Verantwortung in der digitalen Welt. In C. Gärtner & C. Heinrich (Hrsg.), *Fallstudien zur Digitalen Transformation*. Wiesbaden: Springer Gabler. <https://www.springerprofessional.de/corporate-digital-responsibility-unternehmerische-verantwortung-/15214480>.
- Véliz, C. (2020). *Privacy is power*. London: Penguin (Bantam Press).

Weiterführende Literatur

- Dörr, S. (2020). *Praxisleitfaden Corporate Digital Responsibility, Unternehmerische Verantwortung und Nachhaltigkeitsmanagement im Digitalzeitalter* (S. 123–130). Berlin: Springer.
- Morley, J., Kinsey, L., Elhalal, A., Garcia, F., Ziosi, M., & Floridi, L. (2023). Operationalising AI ethics: barriers, enablers and next steps. *AI & Society*, 38, 411–423.
- Ruff, D. (2024). Daten, Plattformen, Smart Contracts, Aktuelle Rechtsfragen. In F. Bösch, S. Lai-mer, A. C. Mittwoch & T. Müller (Hrsg.), *Daten, Plattformen, Smart Contracts: Aktuelle Rechtsfragen*. Schriften zum Digitalwirtschaftsrecht I / Studies on Digital Business Law. (S. 208–210). Zürich St. Gallen: Dike.

Mag. Karin Dietl ist Expertin für Data Compliance, Digital Governance und Corporate Digital Responsibility (CDR). Sie gründete 2019 die Beratungsagentur VALUE DESIGN und startete 2021 die Businessplattform CDR Austria, Österreichs größtes Netzwerk für digitale Ethik. Zuvor war sie als Datenschutzjuristin in Wiener Wirtschaftskanzleien tätig, legte die Rechtsanwaltsprüfung ab und begann als Lektorin an der FH Wien der WKW sowie als Co-Leiterin des Think-Tanks Datenschutz am IIA Austria tätig zu werden.

Dr. Simon Lucas ist Lead Expert Digital Ethics & Bioethics bei der Merck KGaA. In dieser Funktion arbeitet er daran, ethische Überlegungen in konkrete Maßnahmen innerhalb des Unternehmens zu überführen, mit einem besonderen Fokus auf neuartigen Biotechnologien und digitalen Lösungen. Vor seinem Wechsel zu Merck im Jahr 2018 war er rund ein Jahrzehnt in verschiedenen Rollen in Forschung & Entwicklung sowie strategischer Innovation bei Grünenthal, Boehringer Ingelheim und Evonik tätig. Simon hat Chemie und Bioethik in München, Saarbrücken, Kopenhagen und Mainz studiert. Er hält einen Dokortitel in Naturwissenschaften und erwarb einen Magister mit Arbeiten zur Neuroethik und experimentellen Philosophie.

Dr. Jean Enno Charton ist Director Digital Ethics & Bioethics bei der Merck KGaA. In dieser Funktion ist er für das Thema digitale Ethik und Bioethik im Unternehmen weltweit verantwortlich und berichtet direkt an den Vorstand. Vor seiner derzeitigen Position hatte Jean Enno verschiedene Rollen in der Forschungs- und Entwicklungsabteilung von Merck Healthcare inne, zuletzt als Chief of Staff des Chief Medical Officers. Seit 2014 bei Merck, hat Jean Enno einen Dokortitel in Immuno-Onkologie von der Universität Lausanne, Schweiz, sowie ein Diplom in Biochemie von der Universität Tübingen.

Der BRZ-Prüfkatalog für vertrauenswürdige KI

18

Digitaler Humanismus im öffentlichen Sektor

Wolfgang Kabelka

Zusammenfassung

Der Beitrag beleuchtet die konkrete Umsetzung des digitalen Humanismus im öffentlichen Sektor durch den BRZ-Prüfkatalog für vertrauenswürdige Künstliche Intelligenz (KI). Der Katalog zielt darauf ab, ethische Prinzipien wie Transparenz, Fairness und Verantwortlichkeit in KI-Systemen zu verankern. Durch die Berücksichtigung sozialer Nachhaltigkeit und die Förderung von Diversität und Inklusion wird ein menschenzentrierter Ansatz verfolgt, der die Akzeptanz und Nutzung von KI in der Gesellschaft stärkt.

Vertrauenswürdige KI zeichnet sich durch Transparenz, Fairness und Verantwortlichkeit aus.

18.1 Einleitung

Die digitale Transformation bringt erhebliche Veränderungen in verschiedenen Sektoren mit sich. Dabei gewinnen menschenzentrierte digitale Arbeitsprozesse und Geschäftsmodelle zunehmend an Bedeutung. Diese Entwicklungen müssen im Kontext von vertrauenswürdiger Künstlicher Intelligenz (KI), sozialer Nachhaltigkeit und digitalem Humanismus betrachtet werden, um ihr volles Potenzial zu entfalten.

W. Kabelka (✉)
BRZ, Wien, Österreich
E-Mail: wk@kabelka.com

Vertrauenswürdige KI stellt einen zentralen Aspekt dar. Vertrauen in KI-Systeme ist unerlässlich, um ihre Akzeptanz und Nutzung in der Gesellschaft zu fördern. Vertrauenswürdige KI zeichnet sich durch Transparenz, Fairness und Verantwortlichkeit aus.

Transparenz bedeutet, dass die Entscheidungsprozesse der KI nachvollziehbar und verständlich sind. Fairness stellt sicher, dass die KI keine diskriminierenden oder voreingenommenen Entscheidungen trifft. Verantwortlichkeit bedeutet, dass klare Verantwortlichkeiten definiert sind, wer für die Handlungen und Entscheidungen der KI verantwortlich ist. Diese Prinzipien sind entscheidend, um Vertrauen in digitale Arbeitsprozesse zu schaffen und gleichzeitig ethische Standards zu gewährleisten.

Soziale Nachhaltigkeit ist ein weiterer wichtiger Aspekt, der bei der Implementierung digitaler Geschäftsmodelle berücksichtigt werden muss. Soziale Nachhaltigkeit bezieht sich auf die Fähigkeit eines Systems, zur sozialen Gerechtigkeit und zum Wohlstand beizutragen, ohne die Bedürfnisse zukünftiger Generationen zu gefährden. In diesem Zusammenhang sollten digitale Arbeitsprozesse darauf abzielen, Chancengleichheit zu fördern, die Arbeitsbedingungen zu verbessern und einen positiven gesellschaftlichen Einfluss auszuüben. Dies kann durch Maßnahmen wie die Förderung von Weiterbildungsprogrammen, die Unterstützung von Diversität und Inklusion sowie die Schaffung von flexiblen und gesunden Arbeitsumgebungen erreicht werden.

Digitaler Humanismus spielt ebenfalls eine entscheidende Rolle in der menschenzentrierten Digitalisierung. Der digitale Humanismus stellt den Menschen und seine Werte in den Mittelpunkt der technologischen Entwicklungen. Dies bedeutet, dass technologische Innovationen stets darauf abzielen sollten, das menschliche Wohlbefinden zu fördern und die Grundrechte zu schützen. Digitaler Humanismus fordert eine kritische Auseinandersetzung mit den Auswirkungen der Technologie auf die Gesellschaft und betont die Bedeutung der Menschlichkeit in einer zunehmend digitalisierten Welt. Hierbei ist es essenziell, die Balance zwischen technologischen Möglichkeiten und humanistischen Werten zu finden, um die Technologie als Mittel zum Wohle der Gesellschaft zu nutzen.

Zusammengefasst bieten menschenzentrierte digitale Arbeitsprozesse und Geschäftsmodelle ein enormes Potenzial, um die Effizienz und Zufriedenheit in Unternehmen zu steigern und gleichzeitig soziale und ethische Werte zu wahren. Durch die Berücksichtigung von vertrauenswürdiger KI, sozialer Nachhaltigkeit und digitalem Humanismus können Unternehmen nicht nur ihre Wettbewerbsfähigkeit verbessern, sondern auch einen positiven Beitrag zur Gesellschaft leisten. Die Herausforderung besteht darin, diese Prinzipien in die Praxis umzusetzen und dabei stets den Menschen in den Mittelpunkt der digitalen Transformation zu stellen.

18.2 Auf dem Weg zur sozialen Nachhaltigkeit

Das BRZ hat eine lange Tradition zum Thema Nachhaltigkeit. Zu Beginn des Jahrtausends hat das BRZ erfolgreich in einem Fünfjahresprogramm das Ziel des „grünsten Rechenzentrums“ verfolgt. Dabei war Free-Cooling der zentrale Hebel zu einer deutlich höheren

Energieeffizienz. Aber neben den technologischen und umweltbezogenen Aspekten hat das BRZ immer auch die soziale Nachhaltigkeit verfolgt.

18.2.1 Vorstudie für die BRZ-Methode – ALTAI

2019 nahm das BRZ als einziges österreichisches Unternehmen an der Evaluierung der Bewertungsliste für vertrauenswürdige künstliche Intelligenz (ALTAI), der High-Level Expert Group on Artificial Intelligence (AI HLEG) der EU, teil. Die zentralen Ergebnisse sind in Abb. 18.1 dargestellt.

ALTAI stellte eine Checkliste zur Selbstbewertung dar, die Entwickler und Nutzer von KI bei der praktischen Umsetzung der sieben zentralen EU-Anforderungen für vertrauenswürdige KI unterstützen sollte. Diese Anforderungen waren:

- Menschliche Entscheidungsfreiheit und Aufsicht
- Technische Robustheit und Sicherheit
- Datenschutz und Datenverwaltung
- Transparenz
- Vielfalt, Nichtdiskriminierung und Fairness
- Umwelt- und gesellschaftliches Wohlergehen
- Rechenschaftspflicht

Die Ergebnisse der Evaluierung wurden im Dokument „Deep Dive Interviews Report“, das im Rahmen des AI4EU-Projekts erstellt wurde, zusammengefasst (AI4EU-Projekt, o.J.). Ziel des Berichts war es, die HLEG-Richtlinien für vertrauenswürdige Künstliche Intelligenz (KI) durch tiefgehende Interviews mit 50 europäischen Organisationen zu be-

Übersicht teilnehmende Organisationen pro Land

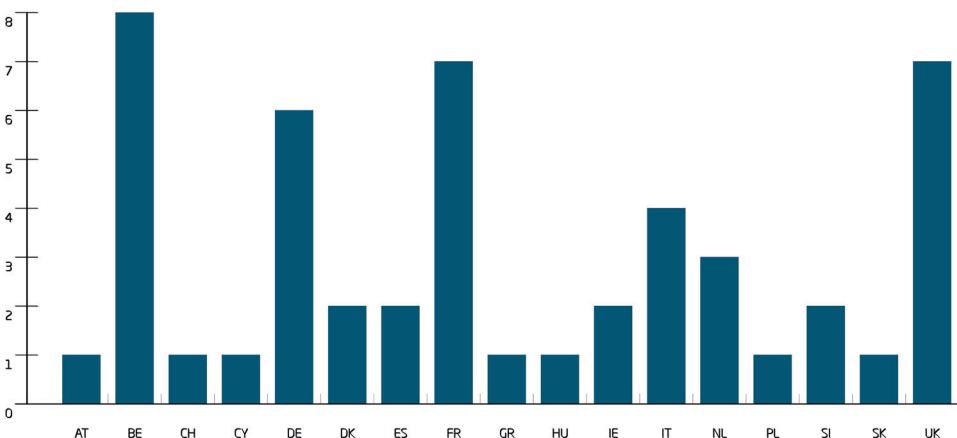


Abb. 18.1 Übersicht teilnehmende Organisationen pro Land

werten. Diese Untersuchung beleuchtete die Verbesserungspotentiale der Richtlinien und bot wertvolle Einblicke in deren praktische Anwendbarkeit und Akzeptanz.

18.2.2 Methodik

Die Methodik umfasste die Auswahl relevanter Organisationen (Abb. 18.2), die Durchführung von Interviews und die Analyse der gesammelten Daten. Die teilnehmenden Organisationen wurden aufgrund ihrer Bereitschaft zur Teilnahme und ihrer Relevanz für KI-Entwicklung und -Einsatz ausgewählt (Abb. 18.3).

Die Interviews konzentrierten sich auf konkrete Anwendungsfälle, um die HLEG-Richtlinien praxisnah zu bewerten. Folgende Verbesserungspotentiale wurden identifiziert:

1. Praktische Relevanz

Eine der zentralen Schwachstellen der HLEG-Richtlinien war der erhebliche Aufwand, der zur Beantwortung der Fragenliste erforderlich ist. Viele Organisationen berichteten, dass die Fragen zu komplex und schwer verständlich seien. Begriffe wie „Nutzer:innen“, „Endnutzer:innen“ und „Stakeholder“ waren nicht klar definiert, was zu Verwirrung führt und den Bewertungsprozess erschwerte.

2. Inhaltliche Redundanz und Branchenrelevanz

Die Analyse ergab, dass viele Fragen redundant oder zu allgemein gehalten waren. Insbesondere für B2B-Szenarien waren einige Fragen nicht anwendbar. Es fehlte an konkreten Beispielen und klaren Definitionen, die den Organisationen helfen könnten, die Anforderungen besser zu verstehen und umzusetzen.

Sektorale Übersicht der Organisationen

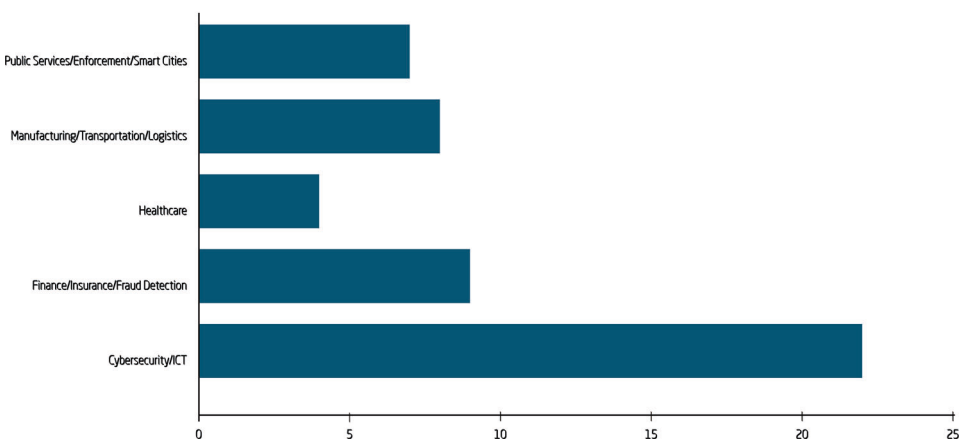


Abb. 18.2 Sektorale Übersicht der Organisationen

Übersicht nach Organisationstyp

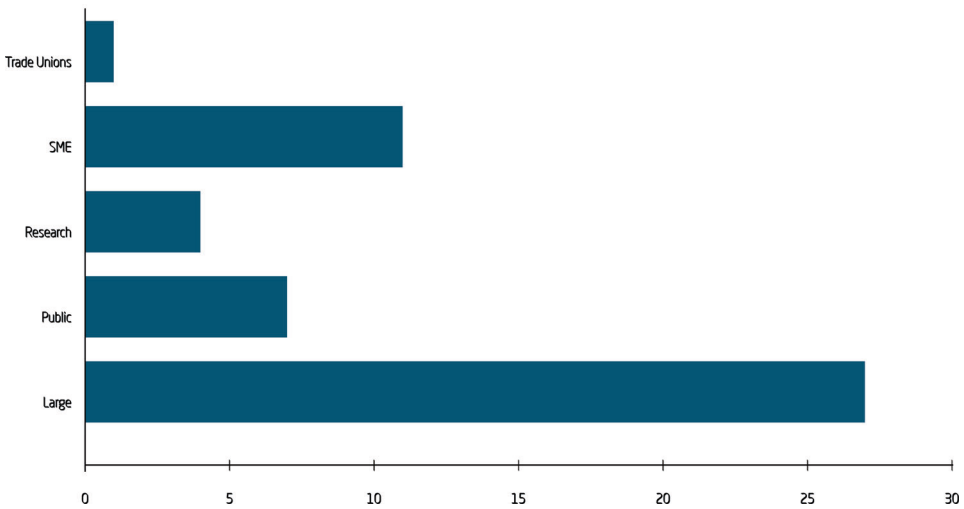


Abb. 18.3 Übersicht nach Organisationstyp

3. Ausrichtung auf bestehende Praktiken

Ein weiteres Problem war die Überlappung der HLEG-Richtlinien mit bestehenden Vorschriften wie der Datenschutz-Grundverordnung (DSGVO, [o.J.](#)). Viele Anforderungen waren bereits durch bestehende Regulierungen abgedeckt, was zu Verwirrung und doppeltem Aufwand führte. Einige Organisationen hatten bereits interne Richtlinien, die gut mit den HLEG-Richtlinien übereinstimmen, was die Notwendigkeit zusätzlicher Bewertungen in Frage stellt.

4. Risiken und Chancen

Die Durchführung der Bewertungen kann die Geschäftsentwicklung verlangsamen und zu einer Überlastung durch zusätzliche Dokumentationsanforderungen führen. Transparenzanforderungen könnten zudem dazu führen, dass innovative Aspekte offengelegt werden müssen, was als Hemmnis für die Innovation gesehen wird. Gleichzeitig bietet die Bewertung jedoch die Chance, das Bewusstsein für verantwortungsvolle KI zu schärfen und als Lerninstrument zu dienen.

5. Ressourcenbedarf

Kleine und mittlere Unternehmen (KMU) benötigen zusätzliche Werkzeuge und Ressourcen, um die Anforderungen zu erfüllen. Große Unternehmen hingegen benötigen geschultes Personal für die Durchführung der Bewertungen. Es besteht ein allgemeiner Bedarf an klaren Leitlinien und unterstützenden Ressourcen, um die Umsetzung der Richtlinien zu erleichtern.

6. Spezifische Anforderungen

Bei Anforderungen an technische Robustheit und Sicherheit kam es oft zu Überschneidungen mit allgemeinen IT-Sicherheitsstandards, welche auch nicht spezifisch für KI waren. Fragen zur Privatsphäre und Datenschutz waren oft redundant zur DSGVO. Die Transparenzanforderungen waren unklar in Bezug auf das erforderliche Detailniveau bei der Erklärung von Entscheidungen. Zudem gab es unterschiedliche Interpretationen von „Fairness“ und fehlende Metriken zur Messung von Fairness.

18.2.3 Fazit

Die HLEG-Richtlinien für vertrauenswürdige KI wurden im Allgemeinen positiv aufgenommen, jedoch waren Anpassungen und Klarstellungen erforderlich, um die Durchführbarkeit und Relevanz zu verbessern. Es wurde empfohlen, die Fragenliste zu vereinfachen, spezifische Beispiele und Definitionen hinzuzufügen und die Anforderungen besser auf bestehende Praktiken und regulatorische Rahmenbedingungen abzustimmen. Diese Maßnahmen könnten dazu beitragen, die Akzeptanz und praktische Anwendbarkeit der Richtlinien zu erhöhen und gleichzeitig die Entwicklung vertrauenswürdiger KI-Systeme zu fördern.

Im weiteren Verlauf zeigte sich, dass viele der Vorschläge im finalen Dokument nicht berücksichtigt wurden. Aus BRZ-Sicht war insbesondere relevant, dass eine stärkere sektorale Abgrenzung erforderlich war und Aspekte wie „Technische Robustheit und Sicherheit“, „Datenschutz und Datenverwaltung“ sowie „Umwelt- und gesellschaftliches Wohlergehen“ einerseits durch Zertifizierungen und Normen wie ISO 27001, ITIL (Infrastructure Library), CoBIT und andererseits durch Maßnahmen wie DSGVO und Green IT Initiative abgedeckt waren.

18.3 Der BRZ-Prüfkatalog

Der Digitale Humanismus widmet sich der Übertragung ethischer Prinzipien und Normen auf konkrete technologische Herausforderungen. In diesem Sinne verfolgt auch der BRZ-Prüfkatalog „Vertrauenswürdige KI“ das Ziel, ethische Maßstäbe in der praktischen Anwendung von Technologie umzusetzen.

Ziel ist dabei die Schaffung der relevanten Voraussetzungen, die auf die Einhaltung bestehender (grund-)rechtlicher Rahmenbedingungen und ethischer Prinzipien abzielen. Dies dient einerseits der Vermeidung und Erkennung von Fehlern, die unbeabsichtigte Auswirkungen sowohl für die Gesellschaft und Betroffene als auch für die einsetzenden Institutionen und ihre Mitarbeiter und Mitarbeiterinnen zur Folge haben. Andererseits zielt es auf eine Softwareentwicklung ab, die by design digitaletische Aspekte einbezieht.

Daher ist die Betrachtung des Digitalen Humanismus kein Add-On, das nach Projektabschluss durchzuführen ist, sondern etwas, das den gesamten Projektprozess begleitet. Es

ist kein Thema, das in kürzester Zeit abgehakt werden kann, vielmehr ist eine kontinuierliche Beschäftigung mit den damit verbundenen Fragen und Herausforderungen gefordert.

Der BRZ-Prüfkatalog basiert auf internationalen Forschungsergebnissen (EU High-Level Expert Group on AI, 2019) und wurde vom BRZ für die spezifischen Anforderungen der öffentlichen Verwaltung adaptiert. Das Vorgehensmodell orientiert sich am Konzept der sozialen Nachhaltigkeit (Algorithmus Accountability Lab et al., 2024). Ziel ist es dabei, die Potenziale von künstlicher Intelligenz für die österreichische Verwaltung mittels eines umfassenden Ansatzes für effektives Risikomanagement und ethische Compliance bestmöglich nutzbar zu machen.

Der BRZ-Prüfkatalog ist primär für den Einsatz im Vorfeld eines KI-Projekts (Product Backlog) vorgesehen und identifiziert Risiken. Jeder Prüfpunkt hat dabei mehrere Merkmale, die in einer Abstufung den Grad der Risikominimierung messen. Der aktuell erstellte Fragenkatalog ist für den Einsatz bei lernenden Systemen optimiert.

Die praktischen Herausforderungen bei der Implementierung des digitalen Humanismus bei der Entwicklung des BRZ-Prüfkatalogs umfassen mehrere Aspekte:

18.3.1 Kontextabhängigkeit der Werteverwirklichung

Die Umsetzung ethischer Werte hängt stark vom Anwendungsfeld und kulturellem Kontext ab. Werte wie Gerechtigkeit und Transparenz müssen je nach Einsatzgebiet und kulturellem Umfeld unterschiedlich interpretiert und priorisiert werden. In Gerichts- oder Verwaltungsverfahren entscheidet ein System mitunter über Strafen, Haft oder existenzielle finanzielle Belastungen, weshalb schon kleine Fehler oder Verzerrungen gravierende, kaum reversible Folgen haben können und deshalb besonders hohe Maßstäbe an Fairness, Transparenz und Datenschutz angelegt werden. In vielen industriellen Anwendungen stehen dagegen vor allem Effizienz, Qualitätssicherung und Arbeitssicherheit im Vordergrund, und Entscheidungen betreffen häufig Prozesse oder Maschinen, nicht unmittelbar grundlegende Individualrechte. Zudem gibt es oft Konflikte zwischen Werten, die als Trade-offs betrachtet werden, wie etwa zwischen Transparenz und Datenschutz.

18.3.2 Sozio-technische Natur der KI-Systeme

KI-Systeme sind komplexe sozio-technische Systeme, deren gesellschaftliche Auswirkungen nicht nur von der Technologie selbst, sondern auch von den zugrunde liegenden Zielen und der Einbettung in organisatorische Strukturen abhängen. Die Umsetzung von Werten wie Gerechtigkeit und Transparenz erfordert daher Maßnahmen während des gesamten Entwicklungs- und Implementierungsprozesses. Dies umfasst sowohl technische Aspekte wie die Erklärbarkeit von Algorithmen als auch organisatorische Maßnahmen wie die Kommunikation und Erklärung von Entscheidungsprozessen.

18.3.3 Unterschiedliche Anforderungen der Stakeholder an die Benutzerfreundlichkeit

Verschiedene Interessengruppen haben unterschiedliche Anforderungen an die Benutzerfreundlichkeit von Ethikrahmenwerken. Entwickler benötigen klare technische Anleitungen, während Organisationen einfache Vergleichsmöglichkeiten und Orientierungshilfen für die Einbettung von KI-Systemen in ihre Strukturen benötigen. Bürger und Verbraucher erwarten verständliche und transparente Informationen über die ethischen Eigenschaften der Systeme, die ihr Leben beeinflussen.

Diese Herausforderungen verdeutlichen, dass die Implementierung des digitalen Humanismus ein vielschichtiger Prozess ist, der sowohl technische als auch organisatorische Maßnahmen erfordert.

18.3.4 Die Operationalisierung der KI-Ethik

Das VCIO-Modell (Values, Criteria, Indicators, Observables) beschreibt einen Ansatz zur Operationalisierung und Messung abstrakter ethischer Prinzipien. Es besteht aus vier Ebenen:

Werte: Diese formulieren allgemeine ethische Anliegen, die unser Handeln leiten sollen. Beispiele sind Gerechtigkeit, Transparenz und Verantwortlichkeit.

Kriterien: Diese definieren, wann die Werte erfüllt oder verletzt werden. Sie spezifizieren, was unter einem bestimmten Wert verstanden wird.

Indikatoren: Diese überwachen, ob die Kriterien erfüllt sind. Sie stellen eine Verbindung zwischen den Kriterien und den beobachtbaren Merkmalen her.

Beobachtbare Merkmale: Diese quantifizieren oder qualifizieren, inwieweit die Indikatoren erfüllt sind. Sie liefern konkrete Daten, die zur Bewertung der Indikatoren verwendet werden.

Das VCIO-Modell hilft dabei, ethische Prinzipien in der Praxis anwendbar, vergleichbar und messbar zu machen. Es ermöglicht die Bewertung der Erfüllung von Werten durch die Definition von Kriterien, Indikatoren und beobachtbaren Merkmalen (Abb. 18.4).

Die Visualisierung der Ergebnisse des Prüfkatalogs beinhaltet quantitative und qualitative Aspekte. Der äußere Kreis liefert Information zum latenten Risiko und Handlungsbedarf. Der innere Kreis stellt qualitative Informationen zur Relevanz und Schärfe des Analyseraster in Bezug auf selbstlernende Algorithmen dar.

Bewertungsregeln für Prüfkatalog vertrauenswürdige KI (pro Prüfbereich)
Gesamtdarstellung nur Risikobewertung

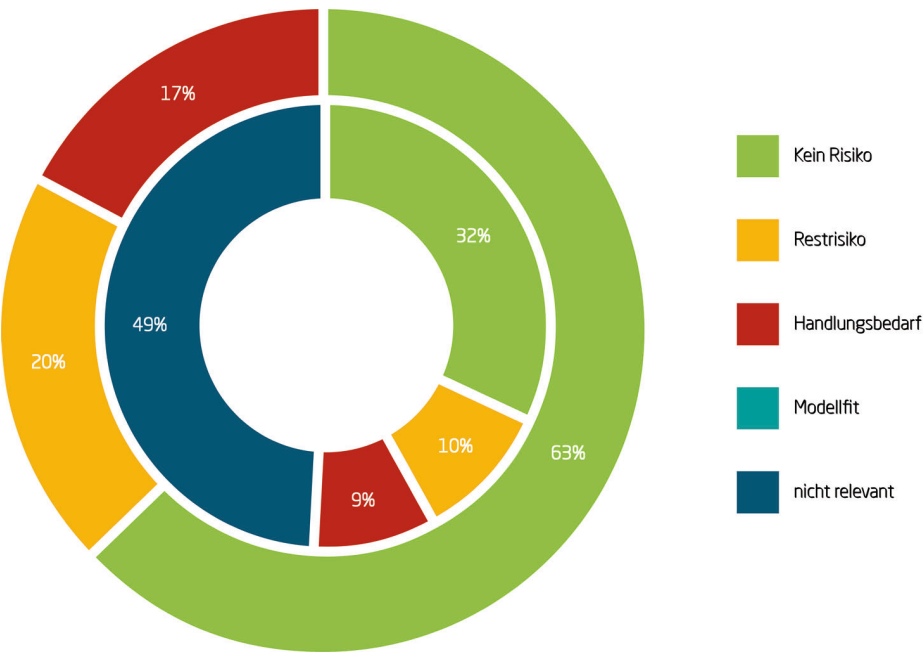


Abb. 18.4 Bewertungsregeln für Prüfkatalog vertrauenswürdige KI

18.4 Vorurteile vorprogrammiert: Eine Analyse der ethischen Herausforderungen und Lösungen in der Künstlichen Intelligenz

18.4.1 Einleitung

In der modernen Gesellschaft werden Entscheidungsprozesse zunehmend durch Algorithmen der Künstlichen Intelligenz (KI) unterstützt. Diese Entwicklung wirft wichtige ethische Fragen auf, insbesondere im Hinblick auf die Fairness und die Vermeidung von Diskriminierung. Dabei werden die Herausforderungen und Lösungsansätze im Umgang mit Vorurteilen (Bias) in KI-Systemen, basierend auf einem Innovationsprojekt des BRZ zur Bias-Mitigation untersucht.

Vorurteile vorprogrammiert:

Eine Analyse der ethischen Herausforderungen und Lösungen in der Künstlichen Intelligenz

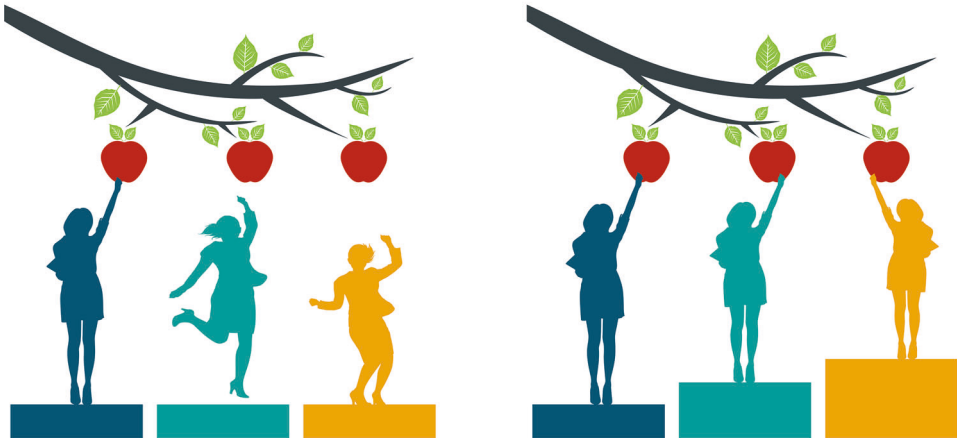


Abb. 18.5 Vorurteile vorprogrammiert: Analyse Herausforderungen

18.4.2 Fairness und Ethik in der KI

Fairness ist ein zentraler Begriff in der Diskussion um ethische KI. Die Definition von Fairness kann jedoch variieren. Ein Beispiel verdeutlicht dies: Drei unterschiedlich große Personen erhalten gleich hohe Podeste, um Äpfel zu pflücken. Obwohl jede Person die gleiche Unterstützung erhält, kann die größte Person mehr Äpfel erreichen (Abb. 18.5). Eine alternative Definition von Fairness würde bedeuten, dass jede Person ein Podest erhält, das ihrer Größe entspricht, sodass alle die gleiche Chance haben, die gleiche Menge an Äpfeln zu pflücken. Diese Analogie zeigt, dass Fairness oft kontextabhängig ist und ethische Entscheidungen erfordert.

18.4.3 Bias in der Künstlichen Intelligenz

Bias entsteht häufig durch die individuellen Wertvorstellungen und kognitiven Verzerrungen der Menschen, die die Daten für KI-Systeme bereitstellen. Diese Verzerrungen können unbewusst in die Algorithmen einfließen und zu diskriminierenden Ergebnissen führen. Ein Beispiel hierfür sind Einstellungsprozesse, bei denen KI-Systeme nach Herkunft oder Geschlecht diskriminieren, weil die Trainingsdaten solche Verzerrungen enthalten.

18.5 Das BRZ-Innovationsprojekt zur Bias-Mitigation

Im Rahmen des BRZ-Innovationsprojekts wurden verschiedene Open-Source-Toolsets getestet, um Bias in KI-Systemen zu erkennen und zu mindern. Ein fiktiver Anwendungsfall war die Abschätzung der Impfbereitschaft anhand von Daten des Austrian Corona Panel Project (ACPP, 2020–2024). Erste Analysen zeigten, dass Personen mit höherem Einkommen eher impfbereit sind. Diese Verzerrung führte dazu, dass Personen mit niedrigerem Einkommen trotz vorhandener Impfbereitschaft benachteiligt wurden. Durch den Einsatz von Bias-Mitigation-Algorithmen konnte die Fairness erhöht und die Zahl der korrekt klassifizierten Personen gesteigert werden.

18.5.1 Ansätze zur Bias-Mitigation

Es gibt verschiedene Ansätze¹ zur Bias-Mitigation. Eine Möglichkeit besteht darin, die Trainingsdaten neu zu gewichten, um unerwünschte Verzerrungen zu reduzieren. Eine andere Methode ist die Anpassung bestehender Algorithmen, indem die Klassifikationsgrenzen für verschiedene Gruppen unterschiedlich gesetzt werden. Diese Ansätze sind vergleichbar mit der Anpassung der Podesthöhen im Fairness-Beispiel.

18.5.2 Ergebnisse und Ausblick

Nach Abschluss des Projekts verfügt das BRZ über ein Analyseraster zur Bewertung der eingesetzten Tools und wertvolles Methodenwissen zur Bias-Mitigation. Künftig kann das BRZ-Trainingsdaten auf kognitive Verzerrungen prüfen, bewerten und abschwächen. Dies ermöglicht Entscheidungsträgern in der öffentlichen Verwaltung, Fairnessabwägungen sichtbar zu machen und Entscheidungen nach ethischen Standards zu treffen.

18.5.3 Fazit zum Thema Bias-Mitigation

Die zunehmende Integration von KI in Entscheidungsprozesse erfordert eine sorgfältige Betrachtung ethischer Aspekte, insbesondere der Fairness und der Vermeidung von Bias. Das BRZ-Innovationsprojekt zeigt, dass durch den Einsatz von Bias-Mitigation-Algorithmen die Fairness in KI-Systemen erhöht werden kann. Zukünftige Entwicklungen sollten darauf abzielen, diese Ansätze weiter zu verfeinern und in der Praxis zu implementieren, um diskriminierungsfreie und faire KI-Systeme zu gewährleisten.

¹ AI Fairness 360 <https://aif360.res.ibm.com/>, SHAP (SHapley Additive exPlanations) <https://shap.readthedocs.io/en/latest/>, sweetviz 2.3.1 <https://pypi.org/project/sweetviz/>.

18.6 Wie ist KI-Ethik im BRZ verankert?

Als Kompetenzzentrum für Digitalisierung in der Bundesverwaltung macht sich das BRZ als Vordenker für Künstliche Intelligenz im öffentlichen Bereich in Österreich stark. Als IT-Partner der Verwaltung stellt sich das BRZ daher künftig als erster Ansprechpartner bei der Prüfung von KI-Applikationen zur Verfügung. Deshalb soll das Thema digitaler Humanismus Teil der BRZ IT-Strategie und der BRZ KI-Strategie werden.

Das BRZ bekennt sich zu folgenden Grundsätzen: Algorithmische Entscheidungen müssen als solche klar erkennbar und für User überprüfbar, nachvollziehbar und im Zweifel anfechtbar sein. Nur eine Vielfalt algorithmischer Systeme kann gesellschaftliche Pluralität angemessen abbilden, Diskriminierung vermeiden und Innovation fördern.

18.7 Nutzen für die Organisation und die öffentliche Verwaltung

18.7.1 Welche Effekte wurden erreicht?

- Sensibilisierung für Governance-Kriterien zum Thema digitaler Humanismus in der öffentlichen Verwaltung
- Definition von Schwerpunktsetzungen
- Verortung der Einsatzmöglichkeiten algorithmischer Assistenzsysteme
- Operationalisierung der Nutzenaspekte
- Maßnahmen zur Stärkung des Vertrauens von Bürger:innen in die Technologie
- Voraussetzungen zur Schaffung von Vorbildern für gemeinwohlorientierte Assistenzsysteme

18.7.2 Welcher Nutzen entsteht für Anwender:innen oder Beteiligte daraus?

- Erkennen kritischer Aspekte in Bereichen von Kundenvorhaben
- Verdeutlichen, was unter einem bestimmten Risiko zu verstehen ist
- Nachvollziehbar erklären, wie zu evaluieren und zu verbessern ist
- Erkennen, erklären und behandeln von Wertekonflikten
- Sensibilisierung zum Thema digitaler Humanismus
- Compliance Vorbereitung für die künftige KI-Strategie des Bundes

Literatur

- AI4EU-Projekt (n.d.). Deep Dive Interviews Report. <https://www.ai4eu.eu/>
- Algorithmus Accountability Lab, TU Kaiserslautern, & Bertelsmann Stiftung (2024). *Forschungsprojekt Algorithmenethik*. Kaiserslautern Gütersloh: TU Kaiserslautern Bertelsmann Stiftung.
- Austrian Corona Panel Project (ACPP) (n.d.). Panelumfrage zur Corona-Krise. Universität Wien, 2020–2024. <https://viecer.univie.ac.at/coronapanel/>. Zugegriffen: 6. Apr. 2025.
- Datenschutz-Grundverordnung (DSGVO) (n.d.). <https://gdpr.eu/>
- EU High-Level Expert Group on AI (2019). *Ethik-Leitlinien für eine vertrauenswürdige Künstliche Intelligenz*. Brüssel: Europäische Kommission.

Weiterführende Literatur

- Austrian Corona Panel Project (ACPP). <https://www.coronapanel.at/>
- CoBIT (n.d.). <https://www.isaca.org/resources/cobit>
- Deutsches Institut für Technologiebewertung, Internationales Zentrum für Ethik in den Wissenschaften, iRights-Lab, VDE, COBIT, & ISACA (2024). *Indikatorenpolitik – Über konsistentes und kohärentes kommunikatives Handeln von Organisationen und in Unternehmen*. TU Darmstadt. Professor Hubig, im Auftrag der Deutschen Chemie-Stiftung Sozialpartner-Akademie
- Green IT Initiative (n.d.). <https://www.greenit.net/>
- High-Level Expert Group on Artificial Intelligence (AI HLEG) (n.d.). Assessment List for Trustworthy Artificial Intelligence (ALTAI). <https://altai.insight-centre.org/>
- International Organization for Standardization (ISO) (n.d.) ISO 27001. <https://www.iso.org/iso-27001-information-security.html>
- ITIL (Infrastructure Library) (n.d.). <https://www.axelos.com/best-practice-solutions/itil>

Wolfgang Kabelka war fast 20 Jahre im Bundesrechenzentrum (BRZ) tätig, zuletzt im Bereich Digital Advisory, wo er an der Weiterentwicklung von eGovernment-Lösungen mitwirkte. Zuvor verantwortete er das Bundes Content Management System (BCMS) und leitete Projekte zu IT- und Informationssicherheit. Weitere Stationen seiner Karriere waren die Braunsberger AG als CIO, die FH Technikum Wien als Studiengangsleiter für Wirtschaftsinformatik sowie die Nomina GmbH als Geschäftsführer. Er studierte Betriebs- und Wirtschaftsinformatik an der Universität Wien.

Kybernetische Steuerungsmodelle zur Umsetzung des Digitalen Humanismus

19

Ein integrativer Ansatz für ethische und technologische Innovation

Armin Preis

Zusammenfassung

Der Beitrag beleuchtet die Anwendung kybernetischer Steuerungsmodelle zur Umsetzung des digitalen Humanismus. Es wird ein Managementmodell vorgestellt, das normative Aspekte wie Werte und Regularien sowie Steuerungsmodelle und Feedbackschleifen integriert. Der Fokus liegt auf der Sicherstellung ethischer Werte in der digitalen Transformation durch Value-Based-Engineering und holistische Veränderungsansätze. Ziel ist es, die menschlichen Werte in der technologischen Entwicklung zu bewahren und weiterzuentwickeln.

Humanismus ist geprägt von der stetigen Weiterentwicklung der Gesellschaft und der Entfaltung des menschlichen Potentials sowie der Nutzung von Innovation und Technologie.

19.1 Einleitung

Humanismus ist geprägt von der stetigen Weiterentwicklung der Gesellschaft und der Entfaltung des menschlichen Potentials sowie der Nutzung von Innovation und Technologie. Dabei muss die Gratwanderung zwischen dem Bewahren der zentralen humanistischen Werte und der Weiterentwicklung derselben gelingen. Hierfür bedarf es eines Managementmodells, das sowohl normative Aspekte wie Werte und Regularien als auch Steue-

A. Preis (✉)
ÖBB, Wien, Österreich
E-Mail: office@arminpreis.at

rungsmodelle und Feedbackschleifen für einen proaktiven Diskurs und vorausschauendes Erleben von Technologien berücksichtigt, um den Humanismus der Zukunft durch bewusste Entscheidungen zu gestalten.

19.2 Das Wesen des digitalen Humanismus

Im Humanismus stehen der Mensch und dessen Potential sowie der Glaube an die Verwirklichung dieses Potentials im Mittelpunkt. Es wird schnell klar, dass Humanismus niemals die Beschreibung eines Zustands sein kann, sondern immer den Fokus auf zielgerichtete Weiterentwicklung legt. Diese Weiterentwicklung ist jedoch kein „wishful thinking“, sondern nach Ernst Bloch „die Dämmerung nach vorne“ (Bloch, 2002), in der die Hoffnung gepaart ist mit Ratio und Technologie als Werkzeug für die Verwirklichung des humanistischen Potentials.

Hierfür braucht es Zuversicht, Vernunft und Gestaltungswillen. Jean-Paul Sartre betont die Eigenverantwortlichkeit des Menschen: „Der Mensch ist dazu verurteilt, frei zu sein. Verurteilt, weil er sich nicht selbst erschaffen hat, und dennoch frei, weil er, einmal in die Welt geworfen, für alles das verantwortlich ist, was er tut.“ Er stellt klar, dass die Existenz der Essenz vorausgehe, weil der Mensch sich selbst erst durch die Gestaltung seines Lebens definiere (Sartre, 2000). Wir müssen also immer handeln und erleben, damit das Erlebte als Feedbackschleife zur Anpassung unserer Werte und deren Erfüllungskriterien nutzbar wird.

Digitalisierung spielt in der Transformation der Menschheit eine zentrale Rolle. Es geht darum, für die Menschheit die Wahlmöglichkeiten zur Lebensentfaltung zu erweitern und gleichzeitig die essenziellen humanistischen Werte zu bewahren und bei Bedarf auch weiterzuentwickeln.

Im Sinne der Aussage „the medium is the message“ von Marshal McLuhan (1964) prägt jede Technologie, jedes Medium auch die Gesellschaft, Technologien sind in Anlehnung an Schellers materielle Werteethik Träger von Werten (Joas, 2017). Durch die Nutzung der Digitalisierung und neuer Technologien verändern sich jedoch auch die Werte der Gesellschaft selbst sowie die jedes Einzelnen. Die vom Menschen geschaffene Technologie wirkt auf den menschlichen Wesenskern und seine Identität zurück. Dies ist ein unbewusster Prozess, in dem Weltbilder, Werte und Grundannahmen unbewusst durch Nutzung von Technologien und Medien verändert werden und im Sinne eines „new normal“ Teil eines neuen Weltbilds werden.

Nachdem jede digitale Transformation rückwirkt auf die Werte der Menschen, muss das Ziel des digitalen Humanismus sein, Achtsamkeit und Bewusstsein für diese Veränderungen zu schaffen, damit es vom unbewussten Verändern zur bewussten Entscheidung und Aufnahme in das neue Weltbild kommen kann. Diese Transformation und Steuerung der Digitalisierung und deren Rückwirkung auf die Gesellschaft bedarf unweigerlich eines umfassenden Steuerungs- oder Managementmodells. Dies muss generisch genug sein,

um auf alle Organisationen, Technologien und Kontexte anwendbar zu sein und doch adaptierbar für den konkreten Zweck.

Mit Blick auf Technologien wie künstlicher Intelligenz, die in unglaublicher Geschwindigkeit und Skalierung Entscheidungen in unser aller Leben treffen kann und wird, ist es wichtig, ein solches Steuerungsmodell präventiv und proaktiv anzuwenden. Ein Blick in die Vergangenheit und auf die Etablierung sozialer Medien zeigt, wie sehr diese Technologie unser Leben verändert hat, diese Veränderung hat sich über Jahre hinweg etabliert, radikale technologische Schritte wie Methoden der künstlichen Intelligenz können dies binnen kürzester Zeit bewirken. Das Konzept der *ambidextren Organisation* von Tushman (2009) erlebt durch künstliche Intelligenz eine Renaissance: klassische Optimierungsmethoden und Informationssysteme entsprachen einem „Exploitation“ Ansatz, unter dem oftmals eine reine Optimierung des Bestehenden durchgeführt wird, mit künstlicher Intelligenz kann binnen kürzester Zeit eine „Exploration“, also radikale Veränderungen in Organisationen, Geschäftsmodellen, Abläufen und Entscheidungsfindungen entstehen im Sinne einer disruptiven Innovation nach Schumpeter (Uctu et al., 2024). Umso mehr ist es erforderlich, dass diese digitalen Methoden proaktiv und präventiv gemanagt werden. Um dies zu ermöglichen, wird im nächsten Abschnitt ein Managementmodell samt ethischer Aspekte eingeführt.

19.3 Kybernetisches Managementmodell

Nachfolgend wird ein generisches kybernetisches Steuerungsmodell für digitalen Humanismus entwickelt und anschließend konkret angewandt in Synthese mit 1.) der Methode systemischer Fragetechniken im Value-Based-Engineering (siehe Abschn. 19.4) zur Sicherstellung der wertorientierten Entwicklung von IT-Services sowie mit 2.) holistischen Veränderungsansätzen in der digitalen Transformation (siehe Abschn. 19.5) zur Sicherstellung der bewussten, zukunftsgerichteten Transformation bestehender Unternehmenskultur und Identität.

19.3.1 Kybernetik und Radikaler Konstruktivismus

Ein Modell zur Steuerung muss stets der Komplexität der Realität gerecht werden, daher wird die Kybernetik als Steuerungslehre verbunden mit dem Modell des radikalen Konstruktivismus, um der Tatsache zu begegnen, dass Interaktionen in Organisationen zutiefst von subjektiven Erfahrungen, Denkweisen und dadurch geprägten Taten von Menschen beeinflusst sind, beide Aspekte werden zusammen adressiert.

Die *Kybernetik* ist nach Scott (2004) ein interdisziplinäres, aber generisches Governance- bzw. Steuerungsmodell, welches von Ingenieuren, Anthropologen, Neurologen, Psychologen und Wirtschaftswissenschaftlern entwickelt wurde mit der Zielsetzung, ein allgemeingültiges Modell zu entwickeln, das auf alle Arten von Systemen anwendbar ist.

Dieser Ansatz ist die Basis für die Steuerung des komplexen Systems von großen Unternehmen in denen sowohl Menschen, Daten, Technologien, Prozesse zusammen und aufeinander wirken. Der *radikale Konstruktivismus* geht laut Scott (2004), nach dem Ansatz von von Foerster und von Glasersfeld (1999) davon aus, dass jeder Mensch, Teil eines Systems und Beobachter desselben ist, kognitiv die Welt um sich herum interpretiert und ein inneres mentales Modell der Realität entwickelt. Wir müssen also bei jeglicher Interaktion zwischen Menschen in Organisationen davon ausgehen, dass diese immer auf Basis mehrerer Abbildungen der Realität erfolgt. Eine wesentliche Aufgabe ist dabei im ersten Schritt das Eintauchen in die Welt des jeweils anderen, um abweichende Interpretationen der vermeintlichen Realität zu verstehen. Im Zuge des gemeinsamen *Sensemaking Prozess* in Anlehnung an Weick (2001) wirkt auch stets das Prinzip der *Zirkularität* nach Glanville (2004). Dieses besteht darin, dass ein Beobachter beobachtet, was in einem System stattfindet und selbst Aktionen am System setzt, wobei jede Handlung an jeglichem Systemteils auf die anderen Teile in nicht linearer Art und Weise wirkt.

Steuernde einer digitalen Transformation sind sowohl Teil eines Systems als auch Beobachter und keine Beobachtung bleibt ohne Effekt im System. Wir müssen uns bewusst sein, dass im digitalen Humanismus und in der digitalen Transformation somit die Menschen und deren Abbild der Realität inklusive Wertvorstellungen und Identitäten im Mittelpunkt unserer Beobachtungen und Aktionen stehen. Wir dürfen bei der Erhebung oder Transformation von Wertvorstellungen niemals von einer objektiven Realität ausgehen, sondern müssen anhand der Transformationsgrammatik von Chomsky (1957) und auch nach Glanville (2004) in die subjektive tatsächliche Repräsentation der einzelnen Menschen eintauchen und dabei eine wertfreie, neutrale Coaching Haltung einnehmen.

„Jeder gute Regulator eines Systems muss ein Modell des Systems sein. [...] In anderen Worten, die Ergebnisse eines Management Prozesses können nicht besser sein als das Modell, auf dem es basiert, außer durch Zufall“ (Schwaninger, 2001). Um ein Management-Modell einzuführen, welches der Komplexität eines Unternehmens gerecht wird, werden nachfolgend die verschiedenen Ausprägungen von kybernetischen Steuerungszyklen erläutert und im Anschluss mit dem Modell der normativen, strategischen und operativen Managementebenen zusammengeführt.

19.3.2 Kybernetische Steuerungszyklen

Kybernetische Steuerungsmodelle ermöglichen unterschiedliche Kombinationen der angeführten Ausprägungen der Steuerungszyklen. Für zukunftsfähige Unternehmen sei darauf zu achten, dass im Managementmodell sämtliche Ausprägungen vorhanden sind:

1. Kybernetik erster Ordnung

Der PDCA-Kreislauf wird hier klassisch durchlaufen. Es werden Vorgaben wie Zielwerte an das System gemacht, Geschäftsprozesse werden abgewickelt, es werden Abweichungen vom Soll gemessen und mittels *korrekativer Adaptionen* das Erreichen der Zielwerte sichergestellt.

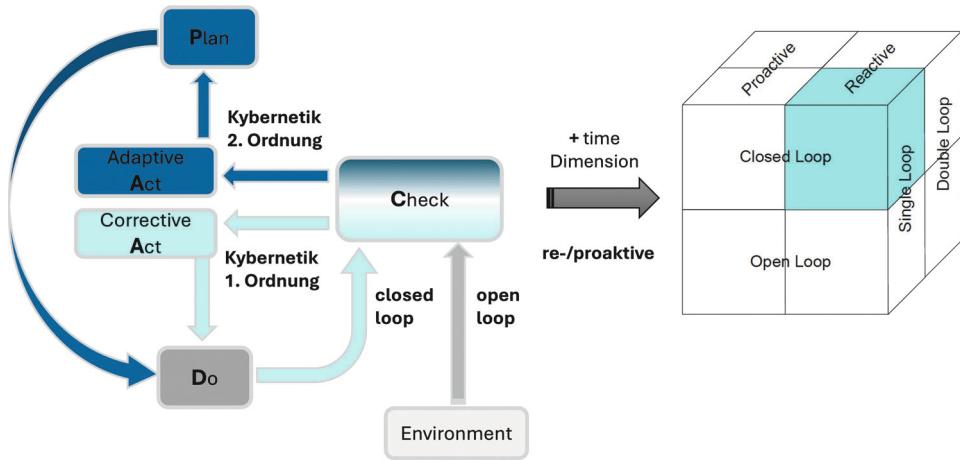


Abb. 19.1 Ausprägungen der kybernetischen Steuerungszyklen in Anlehnung an Schwaiger. (Quelle des Cybernetic Control Cubes: Schwaiger, 2012)

2. Kybernetik zweiter Ordnung

Nach Durchlaufen des PDCA-Kreislaufs erster Ordnung werden auf Grund des Feedbacks bzw. der Abweichung vom Zielwert auch *adaptive Anpassungen* am Steuerungssystem selbst vorgenommen. Im Sinne einer Emergenz wirkt das System auf sich selbst zurück und schafft sich beziehungsweise Teile daraus neu.

3. Closed- und Open-Loop Steuerung

Beim geschlossenen Regelkreislauf wird kein Feedback von außerhalb des Systems eingeholt. Beim der open-loop Regelkreislauf wird sowohl von intern als auch systemexternes Feedback eingeholt.

4. Re- und proaktive Steuerung

Eine reaktive Steuerung regelt lediglich auf Basis von Feedback in Form von tatsächlich gemessenen Werten beziehungsweise Abweichungen von Zielwerten. Bei proaktiver Steuerung wird die Zeittangente im Modell eingeführt, auf Basis von Messwerten und Feedback wird eine Prognose über zukünftige Systemzustände abgegeben und auf Basis derer korrektive oder adaptive Anpassungen im System vorgenommen.

Aus in Abb. 19.1 angeführtem Modell wird ersichtlich, dass es zwei zentrale kybernetische Ansätze gibt, den der ersten und den der zweiten Ordnung. Im Abschnitt III. Kybernetik 1. Ordnung: Value-Based-Engineering wird die Kybernetik erster Ordnung im Hinblick auf werte-getriebenes Systemdesign angewandt und Abschnitt IV. Kybernetik 2. Ordnung: Kultur- und Wertebasierte digitale Transformation wird die Kybernetik zweiter Ordnung angewandt auf Modelle der kulturellen und digitalen Unternehmens-transformation angewandt in denen sich Unternehmen holistisch neu definieren.

19.3.3 Normatives, strategisches und operatives Management

Humanismus impliziert wie in Abschn. 19.2 angeführt Innovation und geht Hand in Hand mit Digitalisierung. Würde das Bewahren im Humanismus in den Vordergrund gestellt, würde er seiner eigenen Existenzgrundlage nämlich der Weiterentwicklung und Entfaltung des menschlichen Potentials beraubt. Dennoch muss die Entwicklung digitaler Systeme und Services vorausschauend erfolgen. Die kybernetischen Steuerungsansätze – insbesondere die proaktiven (Abb. 19.1) und jene der zweiten Ordnung – sollten dabei auch auf der normativen Managementebene von Unternehmen angewandt werden. Systeme haben keinen Selbstzweck, sondern dienen dem Ethos, der Vision und Mission des Unternehmens und materialisieren dessen Wertvorstellungen, die dem Nutzen der betroffenen Menschen wie Kunden und Stakeholdern dient.

Das General Model of Systemic Control (Abb. 19.2) ist ein kybernetisches Management-Modell das die Ebenen des normativen, strategischen Operativen Managements integriert (Schwaninger, 2001).

Preis (2013) führt basierend auf dem MSC ein Modell ein, um Ebenen-übergreifend Unternehmen ganzheitlich zu steuern. Ein Element daraus ist das Management der normativen Unternehmensebene nach dem PDCA Cycle in Kombination mit kybernetischen Steuerungsansätzen.

Die operativen und strategischen Ebenen der Unternehmensführung sind eingebettet in dem normativen Managementansatz. Die Werte und das Ethos leiten sämtliche strategische und operative Entwicklungen (Abb. 19.3). Dieser Nutzen wird auch im digitalen

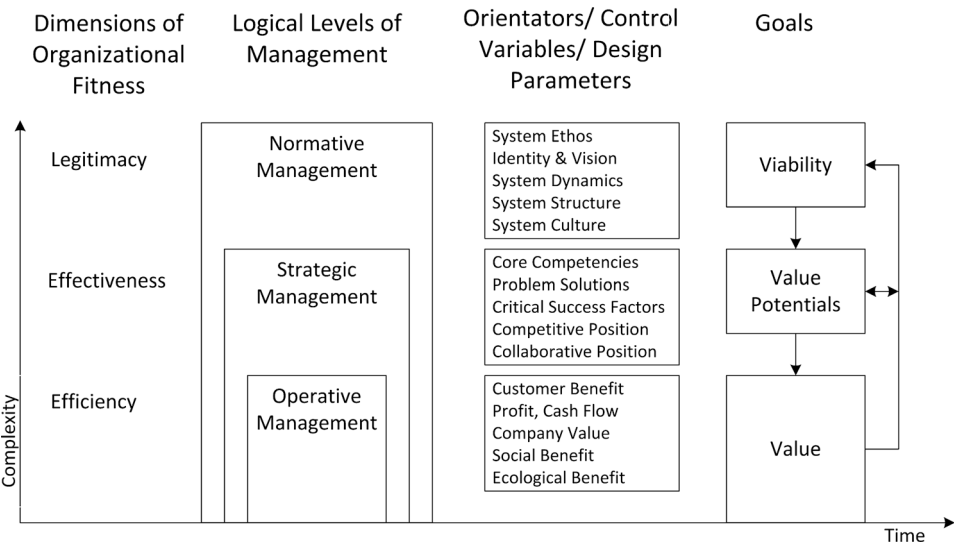
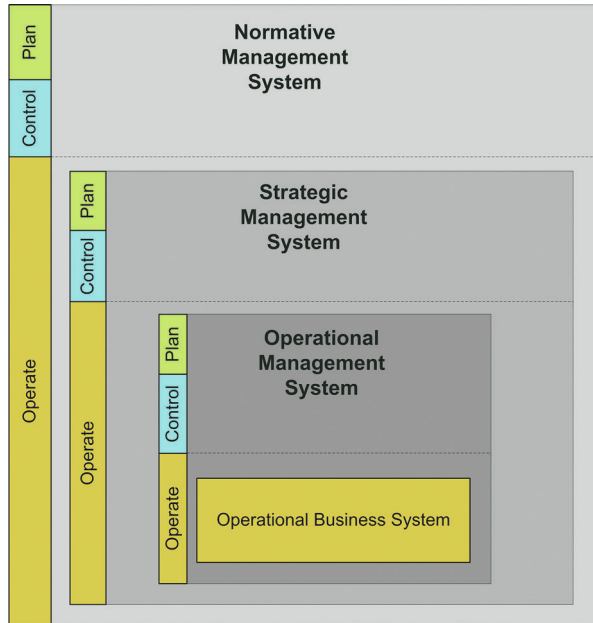


Abb. 19.2 The General Model of Systemic Control. (Quelle: Schwaninger, 2001)

Abb. 19.3 The Recursive Cybernetic Management Model (CRMM). (Quelle: Preis, 2013)



Humanismus schlagend. Es werden digitale Services entwickelt und dabei wirken die normativen Ebenen und prägen die technischen Entwicklungen.

In den zwei folgenden Abschnitten werden zwei Anwendungsfälle für die Kybernetik zur ethischen Steuerung der Digitalisierung skizziert und es finden die kybernetischen Steuerungszyklen sowie die Ebenen aus dem rekursiven kybernetischen Management Modell Anwendung:

1. Abschnitt III. Kybernetik 1. Ordnung: Value-Based-Engineering

Der erste Ansatz des Value-Based-Engineering in Verbindung mit systemischen Frage-techniken, welche die Steuerung von dezidierten Wertvorstellungen entlang der oben angeführten kybernetischen Regelkreise erlauben, stellt eine Kybernetik erster Ordnung im Systemdesign dar.

2. Abschnitt IV. Kybernetik 2. Ordnung: Kultur- und Wertebasierte digitale Transformation

Der zweite Ansatz ist der der weiter gefassten Transformation der strategischen und normativen Ebene in Unternehmen durch die Einführung von Informationssystemen in Verbindung mit Transformationsmodellen wie der Theory U, dem Design Thinking, dem Sensemaking und dem Action Learning. Dieser Ansatz ist eine Anwendung der Kybernetik zweiter Ordnung mit starkem Fokus auf die Transformation und Neuerfindung des gesamten Unternehmenssystems.

19.4 Kybernetik 1. Ordnung: Value-Based-Engineering

In diesem Abschnitt wird die Anwendung der Kybernetik erster Ordnung im Zuge Wertgetriebener Systementwicklung durch Value-Based-Engineering und systemische Frage-techniken zur Erhebung von ethischen Systemanforderungen beschrieben.

19.4.1 Value Based Engineering

Spiekermann und Winkler (2022) beschreiben Value-Based-Engineering (VBE) als „transparente, klar strukturierte, schrittweise Methode, welche Innovations-Management, Risikomanagement und System- und Software-Engineering in ein Prozess-Framework kombiniert“. Sie definieren das Erheben von Werten als Prozess-Modell das vom sogenannten „Value Lead“ geführt wird in Anlehnung an den IEEE 7000 Standard für ethisches Systemdesign (IEEE, 2021). Der Value Lead führt die Organisation durch das Prozessmodell gemeinsam mit dem Management, Stakeholdern, Systemgestaltenden, dem Innovationsmanagement und dem Produktmanagement. Eine der Hauptaufgaben dabei ist, sicher zu stellen, dass Organisationen nicht durch bestehende Werte-Listen geprägt bleiben, laut Spiekermann und Winkler, würde dazu führen, dass unreflektiert bestehende Wertevorstellungen fortgeschrieben und verfestigt werden durch die Implementierung in neuen digitalen Services.

Die konkrete Erhebung der Werte für System Design beinhaltet nachfolgende Elemente (Spiekermann & Winkler, 2022): Core-Values diese werden immer wieder beschrieben von Stakeholdern und sind intrinsisch, also ihrem Selbstzweck dienend; Value-Qualities – diese konkreten Eigenschaften dienen einem Core-Values, sind also instrumentelle Charakteristika beziehungsweise Qualitäten der Core-Values; Durch die Relationen zwischen den unterschiedlichen Core-Values sowie deren Priorisierung ergibt sich eine Ziel-Werte-Ontologie für das zu entwickelnde digitale Service. In Abb. 19.4 sind die relevanten Werte dargestellt, die Priorität der Werte wird anhand der Größe visualisiert, die zugehörigen Werte-Qualitäten sind angeführt, sowie die Relationen der Werte in Sinne einer Ontologie.

Die Core-Values werden durch die Stakeholder priorisiert, diese Priorisierungen bestimmen später die Priorität in der Systementwicklung, dabei werden sowohl Kants Tugend-Ethik als auch die unternehmensinternen und -externen Erwartungshaltungen an die zu entwickelnden digitalen Services im Sinne eines Utilitarismus berücksichtigt (z. B. Gesetze, Normen, Verträge, Unternehmenskultur, Auftrag an das Unternehmen, usw.) wie Spiekermann und Winkler (2022) beschreiben. Joas erläutert in Anlehnung an Scheler, dass sämtliche Artefakte, die uns umgeben, wie Technologien Träger von Wertvorstellungen sind. Um die Wertvorstellungen tatsächlich auf das zu entwickelnde IT-System oder digitale Service zu übertragen wurde ein Prozessmodell entwickelt (IEEE, 2021). Spiekermann und Winkler (2022) beschreiben, dass jedes Tupel aus Core-Value und Value-Quality in ein Ethical-Value-Requirement (EVR) übersetzt wird, welches wiederum in eine Systemanforderung übersetzt wird. Durch die vollständige Überleitung aller Werte

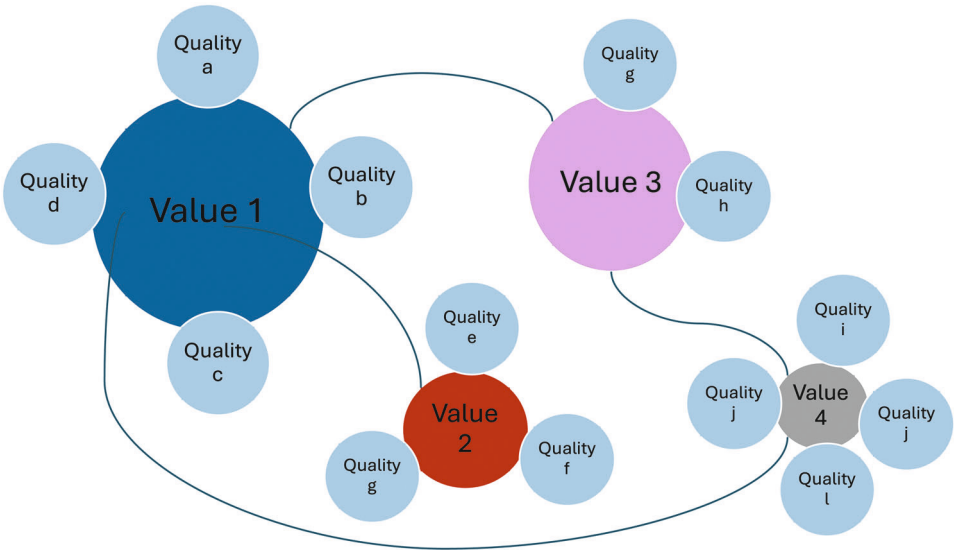


Abb. 19.4 Value Ontology

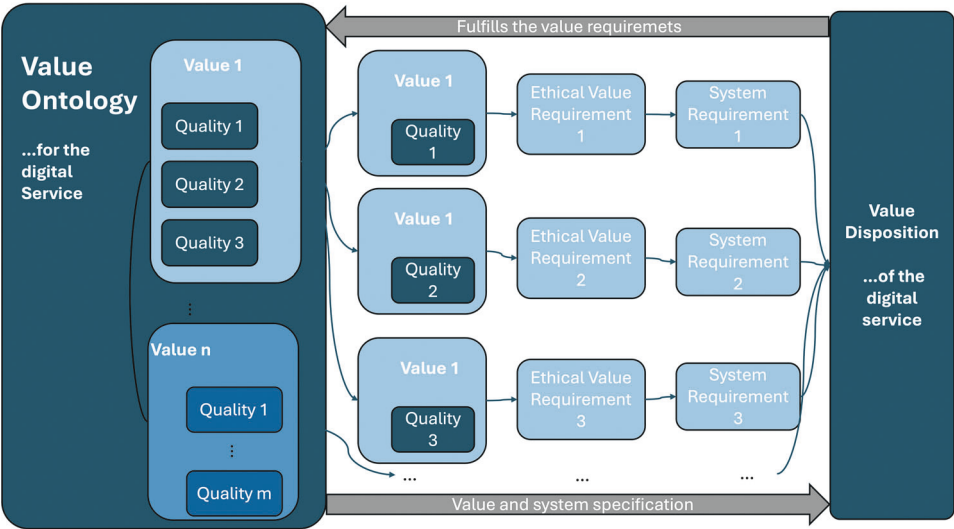


Abb. 19.5 Kybernetischer Regelkreis für ethisches Systemdesign

und Qualitäten aus der Werte-Ontologie wird sichergestellt, dass das entwickelte digitale Service zukünftig genau diejenigen Werte verkörpert, die angefordert wurden.

Es wird ersichtlich, dass der Prozess für ethische Systementwicklung einen kybernetischen Regelkreis (Abb. 19.5) erster Ordnung darstellt. Es werden Ziel-Werte erhoben und in Systemanforderungen übersetzt, das System wird implementiert und es wird gemessen, ob das System die geplanten Werte verkörpert. Ist dem nicht der Fall, wird der Regelkreis

erster Ordnung nochmals durchlaufen und Änderungen an den Vorgaben und Ziel-Werten werden durchgeführt, solange bis die Verkörperung der gewünschten Werte und Qualitäten im jeweiligen IT-System erfolgreich ist. Der Regelkreislauf wird je nach Situation geschlossen, also nur mit Menschen aus dem Unternehmen durchgeführt, es kann aber auch zielführend sein, externe Betroffene wie Kunden und Stakeholder bei der Festlegung der Werte mit einzubinden, je nach Bedarf wird der Regelkreislauf zum closed- oder open-loop Steuerungsmodell. Das Vorgehen stellt eine Mischung aus pro- und reaktivem Vorgehen dar. Es werden im ersten Schritt proaktiv die zukünftig gewünschten Werte definiert und anschließend reaktiv die erfolgreiche Implementierung derselben gemessen.

19.4.2 Coaching Haltung und systemische Fragetechniken

Bei der Erhebung von Werten und zugehörigen Wertequalitäten für digitale Services ist der zentrale Erfolgsfaktor, dass sichergestellt wird, dass tatsächlich die zukünftig benötigten Werte und deren Ausprägungen erhoben werden und nicht bestehende Werte-Kataloge übernommen werden.

Um dies zu unterstützen, wird mittels systemischer Fragetechniken ein möglichst großer Teil der mentalen bzw. kognitiven Verzerrungen, Generalisierungen und Tilgungen der Gesprächspartner vermieden und tatsächlich benötigte Werte eruiert. Der Value Lead versucht möglichst tiefgehende Erkenntnisse zu den mentalen Modellen der Teilnehmenden zu erringen, um von oberflächlichen Aussagen, der Oberflächenstruktur der Sprache (Chomsky, 1957), zu den dahinterliegenden Wertvorstellungen und Überzeugungen vorzudringen. Um diesen Prozess professionell zu gestalten und die Interviewpartner möglichst wenig zu beeinflussen, werden nachfolgend beschriebene Coaching-Interventionen und systemische Fragetechniken angewandt.

Value Leads nehmen eine spezifische Coaching-Haltung ein, welche Radatz (2002) im Kontrast zu alltäglichen Situationen als besonders wertschätzend beschreibt. Der Coach (bzw. Value Lead) hält dabei die eigene Meinung zurück, erteilt keine Ratschläge, sondern stellt offene Fragen, um in das subjektive Weltbild und zugehörige Wertvorstellung der Stakeholder einzudringen und Werte, deren Prioritäten und Zusammenhänge zu verstehen. Für Ungeübte sei gesagt, dass je länger ein Befragter für eine Antwort braucht, desto entscheidender ist dies nach Radatz, da das Vordringen in eigene teils unbewusste Zustände und Wertungen wesentlich länger braucht als das Wiederholen von angelernten Slogans von Unternehmenswerten. Bei der Erhebung der gewünschten Werte, die durch das neue digitale Service verkörpert werden sollen, treffen im Sinne des radikalen Konstruktivismus (von Foerster & von Glasersfeld, 1999) mehrere mentale Modelle aufeinander, das des fragenden Value Leads und das der Antwortenden. Bei der Befragung müssen neben der oben angeführten Coaching-Haltung auch Methoden gewählt werden, um mit unbewusster Voreingenommenheit der befragten Stakeholder umzugehen.

Radatz (2002) macht deutlich, dass der Interviewführende niemals Antworten suggeriert, sondern auf das innere Abbild der Realität des Gesprächspartners fokussiert. Es geht

als nicht um Bestätigung von Hypothesen, sondern mittels Gegenfragen wird der Reflektionsprozess im Gesprächspartner angestoßen. Sie streicht die nachfolgenden Fragetypen als besonders geeignet für systemische Interventionen hervor: Skalenfragen, Differenzfragen, Kontextfragen (auch anwendbar auf jeweilige Technologien), zeitliche Fragen, assoziierende und dissoziierende Fragen (wie das Einnehmen der Rolle einer anderen Person im Unternehmen, eines Kunden, eines Stakeholders, eines Unbeteiligten, eines zukünftig Nutzenden des zu entwickelnden IT-Service) sowie zirkuläre Fragen mit Fokus auf Interaktion der unterschiedlichen Teilnehmenden im System. All diese Ausprägungen von Fragen regen eine Beobachtung, Reflexion und Kommunikation der subjektiven Wahrnehmung an. Im Nachfolgenden werden verschiedene kognitive Verzerrungen (Bias) beschrieben, auf die bei der Erhebung von Wertvorstellungen zu achten ist und es werden die jeweiligen Fragetechniken zur bestmöglichen Vermeidung derselben angeführt.

19.4.3 Kognitive Verzerrungen

Die Erhebung geplanter Wertvorstellungen für zu entwickelnde digitale Services muss weit unter die Oberfläche von Unternehmensslogans vordringen und die tatsächlichen wertebasierten Anforderungen erheben. Damit die befragten Stakeholder auf die benötigten Wertequalitäten für zu entwickelnde Systeme zugreifen können, muss intensiv auf die Vermeidung kognitiver Verzerrungen geachtet werden, da Menschen dazu neigen, ihr Denken und Erinnern als weitgehend objektiver einzuschätzen als dies tatsächlich der Fall ist (Scopelliti et al., 2015). Beim Erheben der Werte im Zuge des Value-Based-Engineering müssen die Befragten dabei unterstützt werden, weil durch den *authority bias*, sowie durch den *availability bias*, den *confirmation bias* (Dobelli, 2011) sowie den *illusory truth effect* (Hassan & Barber, 2021; Udry & Barber, 2024) in Form der Präsenz von Autorität wie dem Top-Management oder durch die Vorgabe von Unternehmenswerten, die laufend kommuniziert werden, die aktuellen Unternehmenswerte häufiger genannt werden als Werte, die für die Entwicklung neuer Services tatsächlich notwendig wären. Werden die Werte und deren Qualitäten in Gruppen erhoben, muss der Value-Lead zusätzlich den *groupthink effect*, der zusätzlich zu obigen Effekten die subjektive Einschätzung noch weiter nivelliert, mitigieren.

19.4.4 Systemische Fragen zur Vermeidung kognitiver Verzerrung

Assoziationsfragen: Der Gesprächspartner wird in eine oder mehrere Situationen versetzt und soll mit der Situation und dem Kontext assoziierte Wahrnehmungen und Wertvorstellungen direkt erleben und unbeeinflusst kundtun. Hierfür kann in *imago* in eine bereits erlebte Situation geführt werden, um erlebte Erwartungshaltungen zum Beispiel von Kunden in realen Situationen reaktiv zu rekapitulieren. Der Gesprächspartner kann aber ge-

nauso mit einer konkreten Situation assoziiert werden, in der das zu entwickelnde digitale Service in Zukunft angewandt wird, um die Erfahrung zu antizipieren.

Differenzfragen: Der Gesprächspartner wird gebeten zu beschreiben, worin der Unterschied zukünftig notwendiger Werte im Gegensatz zu den materialisierten Werten in bestehenden Systemlandschaften besteht. Die Differenzierung von gelebten Werten erfolgt in diesem Fall oft auf der Ebene der Wertequalitäten, sprich es wird vom Fragenden erhoben, welche sind die Erfüllungskriterien (Edmonds, 2014; Dilts, 1995) von Werten auf der Verhaltensebene sind.

Kontextfragen: Wertvorstellungen sind kontextsensitiv. Der Befragte wird ersucht in unterschiedlichen Situationen und auch beim Nutzen unterschiedlicher Informationssysteme die erlebten Wertvorstellungen zu beschreiben, durch den mehrmaligen Kontext-Switch zum Beispiel zwischen bestehenden und zukünftigen digitalen Services wird die differenzierte Wahrnehmung beziehungsweise Antizipation der ethischen Anforderungen erleichtert.

Dissoziierende Fragen: Der Gesprächspartner wird ersucht, sich in eine dissoziiert Beobachterposition zu begeben und das Verhalten und assoziierte Werte durch unterschiedliche Rollen, die bei der Entwicklung und Nutzung des digitalen Services involviert sind zu interpretieren. So löst sich der Gesprächspartner aus gewohnten Konditionierungen und kann die Perspektiven unterschiedlicher Beteiligter am digitalen Service einnehmen.

Skalenfragen und Prioritätsfragen: Diese dienen insbesondere zur Priorisierung der Wertvorstellungen und unterstützen die oben angeführte Differenzierung durch quantitative Einschätzung von Prioritäten auf dezidierten Skalen.

Zusammenfassend wird deutlich, dass, dass im Value-Based-Engineering die Kybernetik erster Ordnung angewandt wird, im closed- oder open-loop Modus und in sowohl in re- als in proaktiver zukunftsgerichteter Weise.

19.5 Kybernetik 2. Ordnung: Kultur- und wertebasierte digitale Transformation

Im Zuge der **digitalen Transformation im Sinne der Kybernetik zweiter Ordnung** werden digitale Services gänzlich neu gedacht und wirken auf die Identität des Unternehmens zurück, wie in Abschn. 19.3 beschrieben wird. Die digitale Transformation ist daher immer ein Veränderungsmodell mit **offenen und proaktiven Regelkreisen**, da das Unternehmensumfeld und die antizipierten zukünftigen Bedingungen die Neugestaltung ermöglichen.

Die digitale Transformation von Unternehmen wirkt nicht nur auf Wertvorstellungen, sondern kann auch Identität, Ethos, Unternehmenswerte, Mission und Vision von Unternehmen radikal verändern. Daher sind neben dem Value Based Engineering auch weiter gefasste Steuerungsmodelle für die Transformation der Unternehmenskultur notwendig, um die digitale Transformation im humanistischen Sinne zu gestalten.

19.5.1 Digitale Transformation

AlNuaimi et al. (2022) beschreiben, dass digitale Transformation alle Ebenen von Unternehmen beeinflusst von der Kultur über Geschäftsmodelle, Organisation und Prozesse, also alles, was notwendig sei, um sich ändernden Marktanforderungen gerecht zu werden. Sie sehen die digitale Transformation als „radikalen institutionelle Change“, welcher „durch kombinierte Effekte von mehreren digitalen Innovationen und Technologien neue Akteure, Strukturen, Praktiken, Werte, Arrangements, und Änderung, Zerstörung und Ersatz von Glaubenssystemen, oder Ergänzung von bestehenden ‚rules of the game‘ in Organisationen, Wirtschaftssystemen, Industrien oder Branchen“ führt. Digitale Transformation formt somit Geschäftsmodelle neu, beeinflusst Kulturen und Unternehmen grundlegend. Auch Ellström et al. (2022) definieren nach Verhoef et al. digitale Transformation als einen „Wandel darin, wie Unternehmen digitale Technologien anwenden, um neue digitale Geschäftsmodelle zu entwickeln, welche [...] mehr Wert für das Unternehmen generieren.“ Sie unterstreichen dabei die Fähigkeiten, die ein Unternehmen hierfür benötigt und betonen die Kompetenz des „Wahrnehmens, Übernehmens und der Rekonfiguration von notwendigen Fähigkeiten“, sie prägen das Konzept der *dynamischen Fähigkeiten* für Unternehmen in Transformation.

Omol (2024) unterstreicht die Interaktion zwischen Technologie und menschlicher Kreativität als evolutionären Prozess, bei dem Technologien das menschliche Potential durch Adaption und Agilität erweitern. Er empfiehlt, ein tieferes Verständnis zu den Dynamiken zwischen technologischer Innovation, Organisation und sozialen Aspekten zu erforschen, um anwendbare Praktiken der Transformation zu entwickeln.

Auf Basis des kybernetischen Managementmodells, des umfassenden Wirkungsbereichs von digitalen Transformation nach AlNuaimi et al. sowie des Fähigkeiten-Modells von Ellström et al. und dem Modell der dynamischen Fähigkeiten nach Omol wird im nachfolgenden ein holistisches Managementmodell vorgestellt, dass alle relevanten Ebenen der Transformation insbesondere aber auch die normative Ebene betrachtet. Auf dieses Modell wird anschließend das Transformationskonzept der Theory U von Scharmer (2009) angewandt, um eine digitale Transformation in der nötigen holistischen Breite zu steuern und Unternehmen neu zu definieren.

19.5.2 Holistisches Transformationsmodell

Managementebenen

Holistische Veränderung und auch digitale Transformation beginnen immer bei der Unternehmensidentität: wofür gibt es das Unternehmen in Zukunft, was ist der wesentliche gesellschaftliche Beitrag? Aus den Antworten auf diese Fragen ergibt sich die Identität und hierdurch werden die Werte, in Form eines klassischen top-down Ansatzes geprägt und verändert. Veränderung erfolgt jedoch auch bottom-up, durch die Veränderung der Umwelt und des täglichen Lebens wie durch die Einführung neuer Informationstechno-

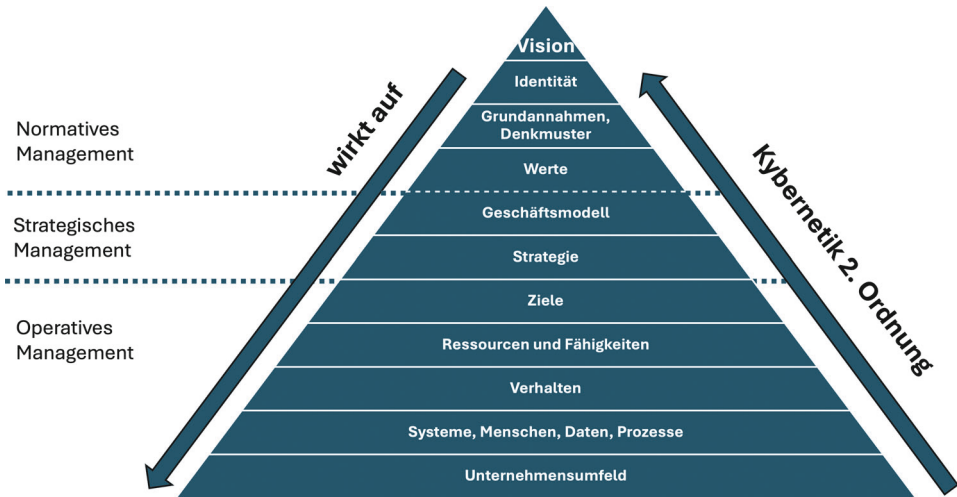


Abb. 19.6 Holistisches Transformationsmodell

logien. Genauso wie die Vision und Identität nach unten wirken, kann insbesondere eine disruptive Technologie nach oben wirken und die Werte verändern bis hin zum Hinterfragen der Unternehmensvision, -mission und -identität.

Daher wird eine digitale Transformation immer auch auf strategischer und normativer Ebene erfolgen. In Anlehnung an die Modelle der Organisationskultur von Schein (2009), den logischen Ebenen von Dilts (1995), sowie dem Model of Systemic Control (Schwanning, 2001) lässt sich folgendes holistisches Transformationsmodell skizzieren.

Entsprechend der Kybernetik zweiter Ordnung wirken sämtliche Veränderungen auf untern Ebenen auf die höheren Ebenen zurück. Das Neu-Denken von Prozessen, IT-Systemen und der Interaktion mit Menschen im Unternehmen sowie mit Kunden und Stakeholdern kann auch eine radikale Veränderung der strategischen und normativen Unternehmensebenen bewirken. Eine digitale Transformation führt zu neuen Geschäftsmodellen, Strategien und ändert auch in einem gewissen Ausmaß das Unternehmen, selbst im Sinne der unbewussten Werte, Grundannahmen, Denkmuster und das Ethos.

Der Begriff der digitalen Transformation wird oft leichtfertig verwendet. Es sollte jedoch kritisch reflektiert werden, ob er auch tatsächlich richtig verstanden wird – eine Transformation die nicht auf die Strategie, Denkmuster, Werte von Unternehmen zurückwirkt, ist des Begriffes der Transformation nicht würdig, dies wäre lediglich ein digitaler MVP.

Das Management dieser ambidextren Organisation, in der das Bewahren und Optimieren des Funktionierenden in Konkurrenz steht mit dem Sich-Neu-Erfinden ist eine der größten Herausforderungen, insbesondere in der digitalen Transformation. Dafür sollten auch für die gänzliche Neugestaltung der digitalen Zukunft von Unternehmen erprobte Prozesse angewandt werden in denen das zukünftige Unternehmen, samt aller oben angeführten Ebenen visualisiert werden kann, um rechtzeitig Leitplanken insbesondere auf normativer Ebene in der Digitalisierung einzuführen.

Metamodell zur Wissensgenerierung

Digitale Transformation beinhaltet einen neuen Blick auf die Welt und die Zukunft des jeweiligen Unternehmens. Dies impliziert sowohl einen Ansatz des Innovationsmanagements als auch einen Kreativitätsprozess. Beide Ansätze werden gemeinsam geleitet im Zuge der Transformation. Hierfür gibt es unterschiedlichste Ansätze wie *Design Thinking* (Mickahail, 2015), *Action Learning* (Pedler, 2024) oder *Theory U* (Scharmer, 2009). Eine Prozessanalyse aller drei Modelle ergibt in Synthese mit dem *SECI Spiralmodell* für Wissensmanagement nach Nonaka (1994) und Takeuchi (2006) ein generisches rekursives Metamodell zur Wissensgenerierung.

Zur Gestaltung der Zukunft sind folgende Schritte notwendig: 1.) bewusstes Wahrnehmen der aktuellen Situation, 2.) Wissen explizit machen, reflektieren, teilen und analysieren, 3.) Synthese für die Zukunft, Durchführung von Planungen, 4.) Umsetzung und wieder von vorne beginnend die Wahrnehmung des „new normal“. Weik beschreibt dabei das Phänomen des *Sensemakings*, welches er als Organisieren des Erlebnisstrom bezeichnet, der entsteht, wenn Menschen im Dialog mit der Umwelt ihre gemeinsame Realität konstruieren (Weick, 2001). Es ermöglicht so die Gestaltung von Koordination, Kohärenz und Change in Organisationen. Er geht somit noch weiter als der *radikale Konstruktivismus* und führt die Idee ein, dass Menschen eine gemeinsame neue intersubjektive Realität und Sinnstiftung gestalten können.

Synthese des Holistischen Managementmodells mit der Theory U

Beispielhaft sei die *Theory U* von Scharmer (2009) der MIT Sloan School of Management angeführt. Sie beschreibt einen klaren Prozess, um organisationale Veränderungsprozesse anzustoßen, die eine Detaillierung des oben angeführten Metamodells darstellen. Im Zuge der nachfolgenden drei virtuellen Räume und sieben Schritte wird der oben skizzierte Wissensmanagement-Prozess (Abb. 19.7) mehrmals spiralförmig durchlaufen. In Anlehnung an Scharmer (2009) sowie Hinz und Garz (o.J.) nachfolgend die Beschreibung der Schritte im *Theory U* Ablauf:

Erster Raum – Raum der Bewusstwerdung

I. Stoppen des Downloading

Im Prozess werden die Teilnehmenden dahin geführt, sich von vergangenheits- und gewohnheitsorientierten Denkmustern zu lösen.

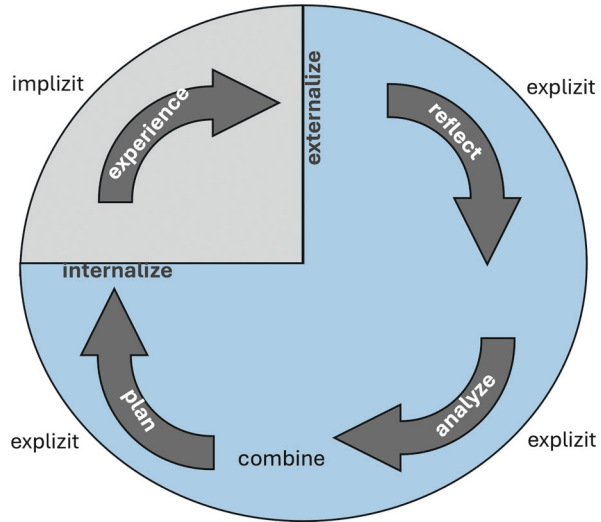
II. Seeing

Im zweiten Schritt werden sie dazu übergeführt, differenziert und unvoreingenommen wahrzunehmen und den Kern der Herausforderungen zu erkennen und detailliert zu beschreiben.

III. Sensing

Durch die Auseinandersetzung mit allen Elementen des holistischen Transformationsmodells (Abb. 19.6) im Zuge der Transformation wird der Kern der Herausforderung und die zu lösenden Fragen klarer. Es werden unterschiedliche Perspektiven eingenommen, es wird gemeinsam eine Zukunftsvision gestaltet.

Abb. 19.7 Generisches rekursives Metamodell zur Wissensgenerierung



Zweiter Raum – Die Zukunft und die inneren Quellen der Kreativität

I. Presencing

Dies ist der entscheidende Raum, der den Beginn einer Transformation einläutet. Es erfolgt ein presencing mit tiefgreifendem Umsetzungsprozess aus kybernetischer Sicht mit Interaktionen im System. Das Wesentliche für die Transformation wird durch Prototyping eine konkrete Idee entwickelt im Sinne der agilen Arbeitsweise und Feedback von allen Stakeholdern eingeholt. Die Erzählungen erfolgen in „first person stories“.

Dritter Raum – Raum der Verwirklichung

I. Crystalizing

Hier werden die notwendigen Schritte für den Transformationsprozess kristallisiert in einem Zusammenspiel zwischen Inspiration und Experiment und im Sinne des generischen rekursiven Modells zur Wissensgenerierung (Abb. 19.7) als sich wiederholender Vorgang durchlaufen.

II. Prototyping

Es wird ein minimum-viable-product inklusive aller notwendigen Teile aus den Managementlevels initiiert, hier wird ein wirkliches Erleben und Erfahren der neuen Situation in interdisziplinären Teams ermöglicht.

III. Realizing

Nach der Weiterentwicklung des Prototyps startet die holistische Umsetzung und die tatsächliche Transformation.

Im Zuge des geleiteten Transformationsprozesses vom Loslassen der Vergangenheit bis hin zur Gestaltung der Zukunft ist insbesondere bei digitalen Veränderungsprozessen ein holistischer Ansatz zu wählen, da eine reine technologische Transformation weder zielführend noch aus kybernetischer Sicht möglich ist (Abb. 19.8).

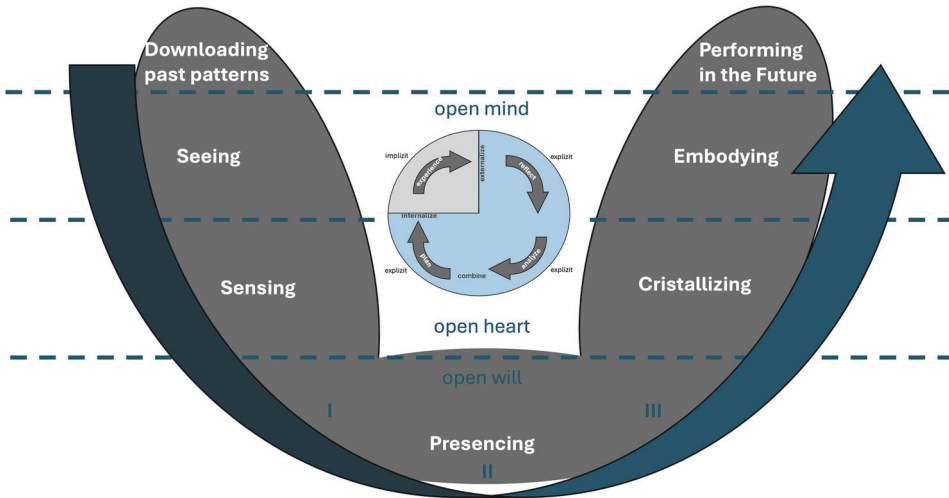


Abb. 19.8 Kombination der Theory U mit dem rekursiven Metamodell für Wissensgenerierung. (Quelle für das Theory U Modell: Scharmer, 2009, Hinz und Garz, o.J.)

Im Zuge des Entwicklungsprozesses können verschiedene Techniken angewandt werden. Diese Techniken führen zur intensiven Beschäftigung und Gestaltung des Unternehmens auf alle Ebenen (Abb. 19.6), insbesondere auf der strategischen und normativen Ebene und nicht lediglich auf der operativen Ebene der Technologien, Prozesse usw. Es werden in der digitalen Transformation allen nachfolgend angeführten Fragen diskutiert und deren Antworten in die Zukunft projiziert werden. Es mag sein, dass manche Ebenen in der digitalen Transformation keine Veränderung erleben wie zum Beispiel die Unternehmensvision und Identität, dennoch müssen die Fragen bewusst beantwortet werden, um den Gestaltungsprozess umfassend sicherzustellen. Der oben angeführte Transformationsansatz am holistischen Managementmodell ergänzt den Ansatz des AI readiness frameworks von Holmström (2022) um weitere notwendige Dimensionen der Unternehmenstransformation.

19.5.3 Mitigation kognitiver Verzerrungen

In der disruptiven Gestaltung der Zukunft durch digitale Transformation sind ähnlich wie im klassischen Value-Based-Engineering kognitive Verzerrungen zu vermeiden. Der *anchoring effect* und der *endowment effect* (Dobelli, 2011) bringen eine gewisse „Verlustaversion“ in der zu transformierenden Organisation mit sich. Die Menschen neigen im Sinne der *Status-Quo Verzerrung* (Kahnemann & Knetsch, 1991) dazu, den Status-Quo aufrechtzuerhalten und zu versuchen, den Verlust desselben zu vermeiden auf Grund der Unsicherheit einer noch nicht bekannten Zukunft, dies wird auch deutlich beim Wunsch zur Aufrechterhaltung selbst entwickelter und mit spezifischen Vorstellungen veranker-



Abb. 19.9 Fragestellung im Holistischen Transformationsmodell im Sinne der Kybernetik zweiter Ordnung

ten IT-Systemen. Ergänzend bringt der *self-serving bias* (Dobelli, 2011) mit sich, dass die Teilnehmenden in der Transformation dazu tendieren, zu beweisen, sie hätten den notwendigen Change längst vollzogen. Um diesen Effekten entlang des Transformationsprozesses der Theory U entgegenzuwirken, können wiederum systemische Fragetechniken im Coaching- bzw. Transformations-Kontext hilfreich sein.

Dissoziations- und Assoziationsfragen: Auch in der digitalen Transformation finden diese beiden Fragetechniken Anwendung. Im ersten Schritt wird die befragte Person von ihrer aktuellen Situation, Verantwortungen, Systemen dissoziiert, um sie dann schrittweise in die Assoziation einer neu zu gestaltenden Zukunft zu führen. Die Dissoziation muss dabei vollständig erfolgt sein, damit bei der Gestaltung der Zukunft nicht die bedingten Konditionierungen schlagend werden.

Wunderfrage: In Verbindung mit der Dissoziationsfrage kann auch die Wunderfrage angewandt werden. Der Coachee wird hier schrittweise in eine Vorstellung geführt, in der „alles möglich sei“. Der Coachee bekommt die Aufgabe, sich vorzustellen, wie die Zukunft unter optimalen Rahmenbindungen aussehen kann ohne jegliche Restriktionen. Die Wunderfrage kann kombiniert werden mit weiteren Methoden aus dem Coaching wie Kreativitätsstrategien und -methoden.

Zirkuläre Fragen: Hier wird geprüft, wie sich die zu gestaltende Zukunft des Unternehmenssystems auf allen Ebenen nach Abb. 19.6 auf die einzelnen Personen im System (Mitarbeitende, Stakeholder, Kunden) auswirkt und wie sich deren Einstellungen und Verhalten ändert und wiederum auf das System samt der in der Skizze angeführten operativen, strategischen und normativen Ebenen des Unternehmens samt Menschen, Daten, Systemen und Prozessen zurückwirkt. Der Coachee wird ersucht, sich in die Schuhe der verschiedensten Personen zu versetzen und mittels *Differenz- und Kontextfragen* wird eruiert, wie sich die neuen Wertvorstellungen und die neue Kultur sowie die materialisierten

Werte der digitalen Services im komplexen System auswirken. Es werden Außen- und Gegenperspektiven eingenommen und die Wirkung von Einstellungen, Handlungen und der Nutzung digitaler Services antizipiert, mit Kontrastfragen wird der Unterschied zur Gegenwart herausgearbeitet.

Zusammenfassend wird ersichtlich, dass eine tatsächliche digitale Transformation ein disruptiver Veränderungsprozess ist, der sowohl auf operativer, strategischer als auch normativer Ebene des Unternehmens gemanagt werden muss. Anhand der eingeführten Modelle der holistischen Transformation in Kombination mit dem generischen Metamodell der Wissensgenerierung und der Theory U wird das Wesen der Kybernetik zweiter Ordnung mit proaktiver Fokussierung auf Zukunftsszenarien und open-loop Feedback aus dem Unternehmensumfeld deutlich. Jegliche vermeintliche digitale Transformation, die nicht auf alle drei Unternehmensebenen wirkt oder keine offene Transformation mit Projektion und holistischer Transformation des Systems Richtung Zukunft beinhaltet, ist keine digitale Transformation und nicht Teil eines wahren digitalen Humanismus.

Literatur

- AlNuaimi, B., Singh, S. K., Ren, S., Budhwar, P., & Vorobyev, D. (2022). Mastering digital transformation: the nexus between leadership, agility, and digital strategy. *Journal of Business Research*, 145, 636–648.
- Bloch, E. (2002). *Das Prinzip Hoffnung*. Suhrkamp.
- Chomsky, N. (1957). *Syntactic Structures*. The Hague: Mouton.
- Dilts, R. (1995). *Identität, Glaubenssysteme und Gesundheit*. Junfermann.
- Dobelli, R. (2011). *Die Kunst des klaren Denkens*. München: Hanser.
- Edmonds, S. C. (2014). *The Culture Engine*. New Jersey: Wiley & Sons.
- Ellström, D., Hölström, J., Berg, E., & Josefsson, C. (2022). Dynamic capabilities for digital transformation. *Journal of Strategy and Management*, 15(2), 272–286.
- von Foerster, E., & von Glaserfeld, H. (1999). *Wie wir uns erfinden. Eine Autobiographie des radikalen Konstruktivismus*. Heidelberg: Car-Auer-Systeme Verlag.
- Glanville, R. (2004). the purpose of second-order cybernetis. *Kybernetes*, 33(9/10), 1379–1386.
- Hassan, A., & Barber, S. J. (2021). *The effects of repetition frequency on the illusory truth effect*. Hassan and Barber Cogn. Research.
- Hinz, H., & Garz, H.-G. (o.J.). Theorie U Methode. https://www.netzwerk-buergerbeteiligung.de/fileadmin/Inhalte/PDF-Dokumente/newsletter_beitraege/nwbb_beitrag_hinz_garz_scholz_120319.pdf
- Holmström, J. (2022). From AI to digital transformation: The AI readiness framework. *Business Horizons*, 65(3), 320–339.
- IEEE (2021). *IEE Standard Model Process for Addressing Ethical Concerns during System Design*
- Joas, H. (2017). *Die Entstehung der Werte*. Frankfurt am Main: Suhrkamp.
- Kahnemann, D., & Knetsch, J. L. R. H. (1991). Anomalies: the Endowment Effects, Loss Aversion and Status Quo Bias. *Journal of Economic Perspectives*, 5(1), 193.
- McLuhan, M. (1964). *Understanding Media: Extensions of Man*
- Mickahail, B. (2015). Corporate Implementation of Design Thinking for Innovation and Economic Growth. *Journal of Strategic Innovation and Sustainability*, 10(2), 67–79.

- Nonaka, I. (1994). A Dynamic Theory of Organizational Knowledge Creation. *Organization Science*, 5(1), 14–37.
- Omol, E. J. (2024). Organizational digital transformation: from evolution to future trends. In *Digital Transformation and Society* (S. 240–256).
- Pedler, M. (2024). *Action Learning in Practice*. Gower.
- Preis, A. (2013). *Integrated enterprise risk and resilience management framework. A cybernetic planning and control framework*. Wien: Technische Universität Wien.
- Radatz, S. (2002). *Beratung ohne Ratschlag. Systemisches Coaching für Führungskräfte und BeraterInnen*. Verlag Systemisches Management.
- Sartre, J.-P. (2000). *Der Existenzialismus ist ein Humanismus*. rororo.
- Scharmer, O. (2009). *Theory U: leading from the future as it emerges*.
- Schein, E. H. (2009). *Corporate Culture Survival Guide*. San Francisco: Wiley.
- Schwaiger, W. S. (2012). Risk management: comprehensive integration into the enterprise Management. In *Asset Management* (S. 420–459).
- Schwaninger, M. (2001). System theory and cybernetics. A solid basis for transdisciplinarity in management education and research. *Kybernetes*, 30(9), 1209–1222.
- Scopelliti, I., et al. (2015). Bias Blind Spot: Structure, Measurement, and Consequences. *Management Science*, 61(10), 2468–2486.
- Scott, B. (2004). Second-order cybernetics: an historical introduction. *Kybernetes*, 33(9/10), 1365–1379.
- Spiekermann, S., & Winkler, T. (2022). *Value-based Engineering with IEEE 7000*. IEEE Technology and Society. pre-print version
- Takeuchi, H. (2006). The News Dynamism of the Knowledge-Creating Company. In H. Takeuchi & T. Shibata (Hrsg.), *Japan Moving Toward a More Advanced Knowledge Economy: Advanced Knowledge-Creating Companies*. Washington D.C.: World Bank Institute.
- Tushman, M. L. (2009). Organizational Ambidexterity: Balancing Exploitation and Exploration for Sustained Performance. *Organization Science*, 20(4), 685–695.
- Uctu, R., Tuluca, N. S., & Aykac, M. (2024). Creative destruction and artificial intelligence: The transformation of industries during the sixth wave. *Journal of Economy and Technology*, 2, 296–309.
- Udry, J., & Barber, S. J. (2024). The illusory truth effect: A review of how repetition increases belief in misinformation. *Current Opinion in Psychology*. <https://doi.org/10.1016/j.copsyc.2023.101736>.
- Weick, K. E. (2001). *Making Sense of the Organization*. Blackwell.

Weiterführende Literatur

- Schein, E. (1992). *Organizational culture and leadership*. Jossey-Bass Publishers.
- Schumpeter, J. A. (1975). *Capitalism, Socialism and Democracy*
- Senge, P. M. (1990). *The Fifth Discipline – The Art & Practice of the Learning Organization*
- Spiekermann, S., & Winkler, T. (2020). *Value-based Engineering for Ethics by Design*. Vienna. arXiv preprint arXiv:2004.13676.

Diplom-Ingenieur Armin Preis, ÖBB-Infrastruktur AG, leitet das Digitalisierungsportfolio des Konzerns und gestaltet seit über 10 Jahren in unterschiedlichsten Bereichen die Transformation des Unternehmens zur kundenorientierten Serviceorganisation. Er ist neben seinen Studien an den

technischen Universitäten Graz und Wien in Softwareentwicklung, Wirtschaft und Wirtschaftsingenieurwesen Informatik ausgebildeter Coach, IT- und Unternehmensberater. Er beschäftigt sich intensiv mit der wertorientierten Gestaltung von Gesellschaft und Digitalisierung, hat 2004 ein Kinderschutz-Zentrum mit aufgebaut. Er hat viele Jahre Unternehmen in der Organisationsentwicklung und im Resilienz-Management beraten.

Teil IV

Rahmenbedingungen, Förderungen und andere Perspektiven

Willkommen im Zeitalter des wertebasierten Managements

20

Von funktionalen Ansätzen zu ethischen und nachhaltigen Strategien

Sarah Spiekermann, Monika Zvonic und Martin Giesswein

Zusammenfassung

Der Beitrag beleuchtet die Entwicklung des Managements von funktionalen Ansätzen hin zu einem wertebasierten Management. Es wird diskutiert, wie Unternehmen durch die Integration von ethischen, ökologischen und sozialen Werten in ihre Strategien und Prozesse langfristigen Erfolg sichern können. Der Fokus liegt auf der Bedeutung von Stakeholder-Dialogen und der Einbindung vielfältiger Interessen, um Produkte und Dienstleistungen zu schaffen, die positive Werte für alle Beteiligten generieren.

Management gestaltet und steuert ein System, das positive Werte für Kunden schaffen soll.

Dieser Artikel entstand aus einem Gespräch mit Sarah Spiekermann.

Monika Zvonic und Martin Giesswein verwendeten KI-Werkzeuge für: Bearbeitung oder Überarbeitung. Der Einsatz von KI betrifft Teile des Dokuments.

S. Spiekermann (✉)

WU Wien, Wien, Österreich

E-Mail: sarah.spiekermann-hoff@wu.ac.at

M. Zvonic

msg Plaut, Wien, Österreich

E-Mail: monika.zvonic@msg-plaut.com

M. Giesswein

WU Executive Academy, Wien, Österreich

E-Mail: martin@giesswein.org

20.1 Einleitung

Das Management einer Organisation ist die treibende Kraft hinter der Gestaltung und Steuerung von Produkten, Dienstleistungen, technischen Systemen oder Organisationsformen. Es definiert Strategien, startet Initiativen, beauftragt Projekte und überwacht deren Umsetzung. Einfacher ausgedrückt: Management gestaltet und steuert ein System, das positive Werte für Kunden schaffen soll. Die Art und Weise, wie das Management diese Systemaufgabe wahrnimmt, hat sich im Laufe der Jahrzehnte grundlegend verändert.

Im letzten Jahrhundert war die Welt des Managements stark von funktionalen Ansätzen geprägt. Unternehmen konzentrierten sich darauf, Produkte zu entwickeln, die vor allem durch ihre Funktionalität überzeugten. Die Idee war klar: Wer die nützlichsten Werkzeuge, Maschinen oder Systeme anbietet, beherrscht den Markt. Es ging um die Gestaltung und Beherrschung der Welt nach dem cartesianischen Prinzip: „Ich denke, also bin ich. Ich denke, also gestalte ich.“ Dies führte zu einem einseitigen Weltbild, in dem die Entwicklung neuer Technologien und Systeme ausschließlich auf die Maximierung von Funktionalität und Effizienz abzielte. Die Bedürfnisse der Menschen wurden kaum berücksichtigt, negative Auswirkungen externalisiert. Das System musste nur funktionieren.

Bereits vor der Jahrtausendwende setzte ein Umdenken ein. Unternehmen wie Apple unter der Führung von Steve Jobs revolutionierten die Welt mit der Erkenntnis, dass nicht nur Funktionalität, sondern auch Ästhetik und Benutzerfreundlichkeit für den Erfolg eines Produktes entscheidend sind. Die Formel „Form follows function“ wurde weiterentwickelt: Form und Funktion sollten von nun an Hand in Hand gehen. Damit begann eine neue Ära, in der Design und User Experience (UX) in den Vordergrund rückten. Produkte sollten nicht nur nützlich, sondern auch schön und intuitiv bedienbar sein. Design Thinking war die Methode der Wahl: Ausgewählte Nutzer beobachten oder befragen – so wurde das Produkt besser. Doch der Fokus auf die direkten Nutzer für ein gutes Produkt reichte nicht immer aus, um den Herausforderungen einer immer komplexer werdenden Welt gerecht zu werden.

Seit einigen Jahren befinden wir uns in einem neuen Zeitalter des Managements: der Werteorientierung in der Wirtschaft und damit auch des wertebasierten Managements. Die Welt ist nicht mehr nur ein Raum, den wir ökonomisch nutzen können, sondern sie konfrontiert uns mit vielschichtige, oft nicht-wirtschaftliche Realitäten, die wir akzeptieren und bewältigen müssen. Komplexe gesellschaftliche Veränderungen, scheinbar unkontrollierbare neue Technologien, ökologische Herausforderungen und die daraus resultierende Vielfalt von Stakeholderinteressen erfordern ein Umdenken. Es geht nicht mehr nur darum, was technologisch möglich ist, sondern was ethisch, ökologisch und sozial verantwortbar ist. Modernes Management erfordert eine kontinuierliche und intensive Auseinandersetzung mit einer Vielzahl von Stakeholdern. Nicht nur die direkten Kunden des Systems stehen im Fokus, sondern auch alle, die in ihren Werten durch das System beeinflusst werden können.

20.2 Die Rolle von Kultur in Unternehmen

Gleichzeitig mit der oben beschriebenen Phase der User Experience und dem Design Thinking auf der Produktentwicklungsseite, entstanden organisationsphilosophische Ansätze wie Corporate Social Responsibility (CSR), Purpose- und Mission-Statements sowie abstrakte Kulturarbeit. Die CSR- und Kulturstrategien der letzten Jahrzehnte waren oft von der Idee geprägt, dass Management-Teams Werte für das Unternehmen definieren und diese symbolisch in Kantinen oder Bürofloren über Plakate kommunizieren. In der heutigen komplexen und dynamischen Welt reicht dieser Ansatz jedoch nicht mehr aus und erzeugt oft das Gegenteil des hehren Ziels: Ablehnung und Entfremdung bei den Mitarbeitenden.

Statt isolierter Top-down-Ansätze ist ein Dialog mit den Stakeholdern – sowohl in direkten als auch in indirekten Gruppen (Interessensvertretungen, Interessensgruppen) – erforderlich. Zu diesen Stakeholdern zählen Kunden, Nichtkunden, Mitarbeitende, Betriebsräte, Lieferanten, Vertriebspartner, Anrainer sowie alle weiteren, die potenziell von den Werten eines Systems beeinflusst werden. Der Fokus liegt auf der Erfassung ihrer Bedürfnisse, Erwartungen und Perspektiven, um Produkte und Dienstleistungen so zu gestalten, dass sie die Werte der Organisation authentisch widerspiegeln. Dieser Ansatz fördert eine hohe Identifikation der Stakeholder mit den angebotenen Lösungen. Dies schafft Vertrauen, das sich auch auf die Kundenbeziehungen auswirkt.

Der so genannte Neue Realismus als philosophische Strömung fordert im wirtschaftlichen Kontext eine Abkehr von klassischen, rein deklarativen CSR-Modellen, in denen Werte von oben nach unten konstruiert werden. Stattdessen wird ein ganzheitlicher Ansatz verfolgt, der auf der aktiven Einbindung aller relevanten Stakeholder basiert. Der Erfolg dieses Ansatzes hängt von der ehrlichen Integration menschlicher Situationen, Perspektiven und Werte in die wirtschaftliche Realität ab.

Humanismus ist das Fundament dieser Entwicklung. Er fordert ein tiefes Vertrauen in die Weisheit und Ehrlichkeit aller Beteiligten. Dies ist weder ein geübter noch ein einfacher Prozess. Die Integration vielfältiger Interessen und Realitäten in Systeme und Prozesse ist eine anspruchsvolle Managementaufgabe, die aber notwendig ist, um in einer zunehmend vernetzten und komplexen Welt langfristig erfolgreich zu sein. Sie fordert Unternehmen heraus, die Lebensrealitäten ihrer Stakeholder zu verstehen und ihre Systeme so zu gestalten, dass sie diese Realitäten ehrlich und nachhaltig berücksichtigen. Wertorientiertes Management ist keine Option mehr, sondern eine Notwendigkeit, um den Herausforderungen des 21. Jahrhunderts gewachsen zu sein.

20.3 Ethik im wirtschaftlichen Kontext

Ein System, das eine Organisation initiiert, ist ethisch, wenn am Beginn des Innovationsprozesses ethische Ansätze und Denkmodelle detailliert, strukturiert und gesamtheitlich durchdacht werden.

Dies bedeutet, dass man sich zunächst mit der Frage auseinandersetzen muss, welche positiven oder negativen Folgen das System für die direkten und indirekten Stakeholder im Sinne des Utilitarismus hat. Der Utilitarismus ist eine ethische Theorie, die besagt, dass eine Handlung dann moralisch richtig ist, wenn sie den größtmöglichen Nutzen oder das größtmögliche Glück für die größtmögliche Anzahl von Betroffenen erzeugt, wobei das Wohlergehen aller Beteiligten gleichermaßen berücksichtigt wird.

Zweitens muss man über den Utilitarismus hinausgehen und die Auswirkungen des Systems auf die Menschen, ihren Charakter, die sozialen Systeme und das Zusammenleben untersuchen. Bei der Definition eines Systems bietet die Tugendethik eine wertvolle Perspektive, indem sie sich auf die Entwicklung und Förderung moralischer Qualitäten innerhalb des Systems konzentriert. Anstatt sich ausschließlich auf festgelegte Regeln oder die Optimierung von Ergebnissen zu konzentrieren, wird das System auf die Werte und Charaktereigenschaften ausgerichtet, die es verkörpern soll. Im Mittelpunkt stehen die inneren Einstellungen und Haltungen der Akteure im System, die ihr Handeln an einem idealen, tugendhaften Zustand ausrichten.

Drittens empfiehlt es sich, aus einer pflichtethischen Perspektive zu reflektieren: Welche Grundwerte wie Freiheit oder Würde sind für unsere Gesellschaft wesentlich? Welche Werte wollen wir in Europa oder in Österreich bewahren und schützen, weil sie für uns von herausragender Bedeutung sind? Im Zentrum steht die Verpflichtung, das Richtige zu tun, unabhängig von den Konsequenzen.

Diese drei ethischen Rahmenmodelle – utilitaristische Folgenethik, Tugendethik sowie pflichtenethische Betrachtung – bilden nicht nur eine solide Grundlage für das Value-based Engineering (im Sinne der Gestaltung eines digitalen Systems), sondern zeigen auch, wie Ethik systematisch in die Prozesse einer Organisation integriert werden kann.

20.4 Konkrete Umsetzung

Ein typisches Beispiel aus der Wirtschaft kann dies verdeutlichen: Ausgangspunkt ist das Kernprodukt oder die Kernleistung eines Unternehmens. In einem Unternehmen, das in erster Linie eine Dienstleistung oder ein Produkt verkauft, geht es darum, dieses Produkt genau zu analysieren. Ziel ist es, es nicht nur funktional zu gestalten, sondern zu einem echten Wertträger zu machen.

Im Mittelpunkt steht die Frage, wie sich das Leben der Kundinnen und Kunden verändert, wenn sie dieses Produkt nutzen. Welche Vorteile bringt es? Welche möglichen Nachteile entstehen – nicht nur für die direkte Zielgruppe, sondern auch für indirekte Stakeholder? Diese Analyse erfordert eine Rückbesinnung auf ethische Dimensionen. Es geht darum, kritisch zu prüfen, ob das Produkt hohe Werte gefährdet oder verletzt. Beispiele sind die Kontrolle über persönliche Daten, die Selbstbestimmung über deren

Verwendung oder die Wahrung der Würde in der digitalen Kommunikation, etwa in sozialen Netzwerken.

Die Unternehmen sind also gefordert, ihre produktbezogenen Hausaufgaben zu machen und ihren Einflussbereich kritisch zu hinterfragen. Es reicht jedoch nicht aus, sich nur auf den eigenen Bereich zu konzentrieren. Vielmehr bedarf es der Zusammenarbeit mit anderen Dienstleistern, die Teil der Wertschöpfungskette sind, um gemeinsam für mehr Qualität zu sorgen. Ethisches und digitalhumanistisches Handeln darf nicht als Kostenfaktor, sondern muss als Qualitätsstrategie verstanden werden. Dieses Verständnis prägt auch das Image ganzer Unternehmen, Städte oder Organisationen, die durch Qualität und nicht durch Billigkeit oder wertigen Schein überzeugen. Ein solcher Ansatz verlangt von Unternehmen, sich aktiv mit den Auswirkungen ihrer Produkte auseinanderzusetzen und dort nachzubessern, wo bisher nur Kostenoptimierung und kurzfristiger Gewinn im Vordergrund standen.

20.5 Europäische Richtlinien und Value-based Engineering

Europa steht vor einer Vielzahl neuer Gesetze, die ein ethisches Systemdesign einfordern. Dazu gehören nicht nur die DSGVO, sondern auch der Digital Markets Act, der Digital Services Act, die KI-Verordnung, der Cybersecurity Act, der Data Act und andere Regelwerke. Unternehmen stehen daher oft vor der Frage, wo sie anfangen sollen. Alle diese Gesetze teilen jedoch ein gemeinsames Ziel: die ethischen Prinzipien unserer Gesellschaft zu schützen.

Hier gibt es zwei grundlegende Ansätze für Firmen, die diesen Regeln unterliegen: Entweder erfüllt ein Unternehmen eine Compliance-Anforderung nach der anderen mechanisch und mit Mindestmaß unter Abwägung der potentiellen Strafen. Man kann in teure rechtliche Absicherungen investieren, um die gesetzlichen Anforderungen zu erfüllen – sinnbildlich als „Feigenblätter“. Oder man entwickelt von Anfang an Systeme, die ethisch korrekt sind. Solche Organisationen handeln tatsächlich ethisch und wertebasiert und machen aus der Not (rechtliche Verpflichtung) eine Tugend (wirtschaftlichen Vorteil): Wer ethisch handelt, erfüllt ganz automatisch die Compliance-Anforderungen.

Doch wie wird dieser Anspruch konkret umgesetzt? Was bedeutet es, wenn der Gesetzgeber beispielsweise fordert, dass KI- oder allgemeine IT-Systeme sicher, datenschutzfreundlich, transparent, robust und diskriminierungsfrei sein sollen?

Hier kommt wieder das ursprüngliche Kerngebiet des Value-Based Engineering ins Spiel. Es basiert auf dem Standard ISO/IEC/IEEE 24748-7000:2022 und bietet Unternehmen eine strukturierte Vorgehensweise. Schritt für Schritt werden die notwendigen Prozesse und Aktivitäten aufgezeigt, um Maßnahmen zur Sicherstellung ethischer KI- und IT-Systeme zu definieren und umzusetzen.

Value-based Engineering ist der weltweit einzige Standard, der Unternehmen methodisch von der Problemdefinition bis zur Systemimplementierung begleitet. Zwischen 2016 und 2021 wurde dieser Standard von einer Arbeitsgruppe mit über 150 Experten entwickelt und kontinuierlich verbessert. Über 1000 Rückmeldungen aus der Praxis flossen in die Optimierung ein. Ergänzend wurde ein Handbuch erstellt, das den Prozess weiter vereinfacht. Austrian Standards bietet eine entsprechende Personenzertifizierung an.

20.6 Digitaler Humanismus braucht exzellentes Management

In der Anwendung des Value-based Engineerings als ein systemischer Ermöglicher des Digitalen Humanismus braucht es Management-Exzellenz. Sei es im Top-Management oder im konkreten IT-System und dessen Projektmanagement. Man muss in der Lage sein, Systeme so zu beschreiben und darzustellen, dass sie einerseits einfach und verständlich sind, andererseits fachlich detailliert und präzise genug, um fundierte Stakeholder-Dialoge führen zu können. Die Kunst besteht darin, das Wesentliche vom Unwesentlichen zu unterscheiden, ohne unnötige Komplexität hinzuzufügen. Die Balance zwischen Pragmatismus und Tiefe ist entscheidend. (Projekt-)Manager müssen mit einem pragmatischen Ansatz arbeiten, der Stakeholdergruppen überschaubar hält, Systeme klar darstellt und gleichzeitig ethische Diskurse ermöglicht.

Wenn an solchen Projekten Personen beteiligt sind, die nicht in der Lage sind, die Spreu vom Weizen zu trennen, relevante Details zu berücksichtigen und relevante Stakeholder einzubinden, entstehen oft faule Kompromisse und „wertlose“ Systeme. Es braucht kluge und erfahrene Menschen, die das Relevante erkennen und priorisieren können. Hierzu passt das lateinische Sprichwort „Sapientis est ordinare“ – „Weisheit besteht darin, zu ordnen“. Diese Fähigkeit des Ordners und Priorisierens ist keine Selbstverständlichkeit und erfordert Erfahrung und Weitsicht. Sie ist der Kern eines wertebasierten Managements, das (digitalen) Humanismus versteht, lebt und fördert.

Dieses Managementverständnis steht im Gegensatz zur reinen Ingenieursperspektive, die sich auf Details fixiert, oder zur Design-Thinking-Mentalität, die zwar kreative Ansätze liefert, oft jedoch zu oberflächlich bleibt und das notwendige Verständnis für Werte und die besonderen Anforderungen von IT Systemen vermissen lässt. Weder die Fixierung auf Details noch ein rein designorientierter Ansatz reichen aus. Vielmehr sind Pragmatismus und situatives Handeln gefragt.

Im Silicon Valley wurde in der Vergangenheit oft propagiert, dass rein datenbasierte Entscheidungen die menschliche Intuition ersetzen könnten. Der Begriff HIPPO (Highest Paid Person's Opinion) wurde dort abwertend verwendet, um zu sagen, dass die Meinung von oft hochbezahlten Managern schädlich sei, da A/B-Tests und statistische Verfahren die bessere Entscheidungsgrundlage bieten würden. Doch gerade diese Sichtweise greift zu kurz. Gerade in komplexen Systemen, in denen ständig neue Entscheidungen auf unbekanntem Terrain getroffen werden müssen, ist es unmöglich, nur datenbasiert vorzugehen.

Es bedarf einer Kombination aus Intuition, Erfahrung und pragmatischer Systembewertung.

Effektives Leadership in diesem Kontext bedeutet, die Fähigkeit zu entwickeln, zuzuhören, entscheidende Faktoren zu identifizieren und auf dieser Basis fundierte Entscheidungen zu treffen. Dieses Leadership der Manager ermöglicht es, Verantwortung zu übernehmen und Entscheidungen zu treffen, die auf einem Gleichgewicht von analytischem Denken und intuitivem Handeln basieren. Diese Kombination aus Zuhören und entschlossenem Handeln ist der Schlüssel zum Erfolg in einer zunehmend komplexen und datengetriebenen Welt. Es ist die neue und notwendige Kunst des wertebasierten Managements.

Sarah Spiekermann leitet seit 2009 das Institut für Wirtschaftsinformatik & Gesellschaft an der Wirtschaftsuniversität Wien (WU). Sie forscht, lehrt und schreibt zur digitalen Ethik und Wertebasierten IT-Innovation. Nach beratenden Funktionen für die EU-Kommission leitete sie die Entwicklung des ersten globalen Standards für ethisches IT-Design (ISO/IEEE 24748-7000) (2016–2021) und prägte das sog. ‘Value-based Engineering’, zu dem 2023 ihr gleichnamiges Lehrbuch erschien. Aufgrund von über 100 Fachartikeln wurde sie 2024 von der Stanford University zu den 2% weltweit am häufigsten zitierten Wissenschaftlern ihres Fachs gezählt. 2024 gründete sie in Wien die VBE-Akademie, die international Fachleute im Bereich ethischer KI – Innovation ausbildet.

Monika Zvonic ist Marketing-Expertin mit Schwerpunkt PR und Kommunikation. Bei GlobeAir leitete sie das Corporate Magazin The Smart Traveller (Auflage 20.000) und entwickelte kreative Marken- und Kommunikationsstrategien. Heute setzt sie ihre Erfahrung als PR & BU Marketing Managerin bei msg Plaut ein und verbindet Storytelling mit strategischer Markenführung.

Martin Giesswein ist Fakultätsmitglied der WU Executive Academy mit den Schwerpunkten KI, Digitalökonomie und Leadership. Er ist Autor und Herausgeber mehrerer Bücher zu den Themen Künstliche Intelligenz, Digital Leadership und Zukunftsgestaltung. 2025 wurde er mit einem ersten Preis des Awards „Digitaler Humanismus in der Praxis“ ausgezeichnet. Er ist als AI Ethics Advisor Mitglied des Advisory Committee im Projekt DRAIGON. Martin Giesswein ist einer der Mitinitiatoren der DigitalCity.Wien, einer seit über 10 Jahren bestehenden Kooperation der Digitalwirtschaft mit der Stadt Wien. Zuvor war er CEO und General Manager in der IKT-Branche.

Erich Prem und Hannes Werthner

Zusammenfassung

Als Initiative für eine menschenorientierte Digitalisierung setzt sich der digitale Humanismus für eine verantwortungsvolle Entwicklung der Informationstechnologie ein, die Demokratie und Inklusion fördert. Er steht damit nicht im Gegensatz zu Innovation, sondern kann sogar als Innovationstreiber wirken. Durch die Berücksichtigung gesellschaftlicher Aspekte hilft der Ansatz des digitalen Humanismus, individuelle und gesellschaftliche Risiken abzufedern. Damit ist er auch ein Instrument der Innovationsstrategie. Als gesamtgesellschaftlicher und gesamtwirtschaftlicher Ansatz zeigt er aber auch die Wechselwirkung bzw. auch Schwierigkeit zwischen einzelunternehmerischer Aktivität und Optimierung auf der einen Seite und gesamtgesellschaftlichem Fortschritt auf der anderen Seite. Letztendlich ist es eine Gesamtsystemfrage, die entsprechende Anreize und Markteingriffe erfordert.

Der digitale Humanismus zielt darauf ab, Mensch und Natur in den Mittelpunkt der digitalen Transformation zu stellen und sicherzustellen, dass digitale Technologie gesellschaftliche Werte wie Freiheit, Gerechtigkeit und Solidarität fördert.

E. Prem (✉)

Verein zur Förderung des Digitalen Humanismus, Wien, Österreich

E-Mail: prem@eutema.com

H. Werthner

TU Wien, Wien, Österreich

E-Mail: hannes.werthner@tuwien.ac.at

21.1 Der digitale Humanismus

21.1.1 Motivation und Ursprung

Der digitale Humanismus ist eine Initiative, die sich mit der Problematik digitaler Technologie und grundlegenden gesellschaftlichen und humanzentrierten Werten beschäftigt. Diese Bewegung nahm ihren Ursprung in Wien. Sie gewinnt insbesondere seit der Veröffentlichung des **Wiener Manifests für Digitalen Humanismus** im Jahr 2019 zunehmend an Bedeutung. Dieses Manifest wurde von einer Gruppe von Wissenschaftler*innen unterschiedlicher Disziplinen initiiert, die sich mit der Frage auseinandersetzen, wie die rasante technologische Entwicklung der Digitalisierung mit gesellschaftlichem Fortschritt in Einklang gebracht werden kann. Der digitale Humanismus zielt darauf ab, Mensch und Natur in den Mittelpunkt der digitalen Transformation zu stellen und sicherzustellen, dass digitale Technologie gesellschaftlichen Werte wie Freiheit, Gerechtigkeit und Solidarität fördert, anstatt diese zu gefährden.

Ein zentrales Anliegen der Wiener Denker*innen ist es, der oft unkontrollierten und undurchsichtigen Entwicklung von Technologien wie Künstlicher Intelligenz, automatisierten Systemen und von Internetplattformen entgegenzuwirken. Sie kritisierten die zunehmende Kommerzialisierung des Internets, die zu Überwachung, einem Verlust an Privatsphäre und einer Entmenslichung und Verschiebung der Machtverhältnisse in Richtung weniger technokratischer Akteure führten. Das Manifest fordert eine transparente, inklusive und auf Mensch und Natur ausgerichtete digitale Zukunft. Es betont, dass Technologie nicht nur als Werkzeug zur Effizienzsteigerung betrachtet werden sollte, sondern als ein Mittel zur Förderung des Gemeinwohls und der sozialen Gerechtigkeit. Das Manifest begreift die digitale Transformation als komplexen sozioökonomischen und technischen Prozess, der aber nicht deterministisch abläuft. In seiner engeren Definition ist der digitale Humanismus ein Ansatz, der *das komplexe Zusammenspiel von Technologie und Mensch beschreibt, analysiert und vor allem beeinflusst, für ein besseres Leben und eine bessere, nachhaltige Gesellschaft im Einklang mit der Natur, unter voller Achtung der Menschenrechte* ist. Er ist also gleichzeitig analytisch und konstruktiv, und daher auch – per Definition – interdisziplinär, von Geistes- über die Sozialwissenschaften bis zu den technischen Disziplinen; oder anders formuliert (als Informatiker): die Informatik ist wichtig, aber ohne die anderen ist sie „nichts“. Die Motivation des digitalen Humanismus liegt daher in der Überzeugung, dass der technologische Wandel nicht auf Kosten der menschlichen Werte und der gesellschaftlichen Integration gehen darf. Die Wiener Bewegung ist zu einer internationalen Initiative geworden, die zahlreiche Wissenschaftler*innen und Intellektuelle erfasst hat. Durch die Betonung von Werten wie Verantwortung, Menschenwürde und gesellschaftliche Solidarität strebt der digitale Humanismus an, eine Gesellschaft zu schaffen, in der technologische Innovationen dem Wohlergehen aller Menschen dienen und nicht nur den Interessen einer einflussreichen Elite. Dazu gehört der Schutz der Umwelt ebenso wie die Anerkennung und Förderung von Diversität und kultureller Identität. Der digitale Humanismus ist zwar bereit, digita-

ler Technologie Grenzen zu setzen, er ruft jedoch vor allem zur aktiven Gestaltung der digitalen Zukunft auf. In diesem Kontext versteht sich der digitale Humanismus als eine kritische, aber zugleich konstruktive Perspektive auf die digitale Zukunft, was ihn von vielen Strömungen unterscheidet, die primär in der Kritik verbleiben.

Der digitale Humanismus als Konzept und auch als Initiative hat schon viel erreicht, insbesondere unter den gegebenen beschränkten Ressourcen (Digital Humanism, 2025; Digital Humanism Initiative, 2025). Das wichtigste Ergebnis ist jedoch, dass es gelungen ist, einen nationalen und internationalen intellektuellen Kern des Digitalen Humanismus zu schaffen, und zwar aus der Wissenschaft, der Wirtschaft, Institutionen und der Zivilgesellschaft, die im gegenseitigen Respekt und aus tiefster Überzeugung das Thema weitertreiben. Ohne sie wäre diese Idee nichts.

21.1.2 Grundsätze

Die Prinzipien des digitalen Humanismus basieren auf den Werten der Aufklärung. Technologie ist im Einklang mit grundlegenden Werten wie Menschlichkeit, Freiheit, Gerechtigkeit zu entwickeln und zu nutzen. Digitale Technologien sollen nicht nur Einzelnen helfen, sondern stets auch den sozialen und ethischen Bedürfnissen der Gesellschaft gerecht werden. Der digitale Humanismus fordert eine faire Gestaltung digitaler Systeme, die allen Menschen zugutekommt und Diskriminierung oder Ausgrenzung verhindert.

Ein weiteres wichtiges Prinzip ist die Förderung von Transparenz und Verantwortung in der digitalen Welt. Entscheidungen, die durch Technologien wie Künstliche Intelligenz beeinflusst werden, sollen nachvollziehbar und überprüfbar sein. Digitale Technologien sollen die Autonomie von Personen unterstützen und in Kooperation mit ihnen funktionieren.

Zusätzlich wird im digitalen Humanismus der Wert der Freiheit und der Selbstbestimmung betont. Menschen sollen die Kontrolle über ihre Daten und ihre digitale Identität behalten, ohne Angst vor ungewollter Überwachung oder Missbrauch. Der Schutz von Autonomie und Privatsphäre und der individuellen Rechte werden als fundamentale Grundlage der Menschenwürde angesehen. Schließlich fordert der digitale Humanismus eine inklusive digitale Zukunft, die den Zugang zu Technologien für alle Menschen ermöglicht, unabhängig von Herkunft, sozialer Stellung oder Bildungsniveau. Dies schließt auch die Förderung von Bildung und kritischem Denken in Bezug auf digitale Technologien ein, damit die Gesellschaft aktiv und verantwortungsbewusst am digitalen Wandel teilnehmen kann. Die Zukunft der digitalen Technologien ist daher auch in demokratischer Weise voranzutreiben. Viele digitale Innovationen verlangen nach einem Diskurs über die einzuschlagende Richtung ihrer Entwicklung und nach möglichen Grenzen, Rahmenbedingungen oder Abfederung der Folgen digitaler Technologie.

Der digitale Humanismus lehnt das mechanistische und das animistische Paradigma ab – Menschen sind keine Maschinen, und Maschinen sind keine Menschen (Nida-Rümelin und Staudacher, 2024; Prem, 2024; Werthner, 2024). Das mechanistische Bild ist

eine Metapher aus der frühen Neuzeit. Der Mechanismus lässt die Welt als ein großes deterministisches Uhrwerk erscheinen und die Menschen als Rädchen in diesem Mechanismus. Nichts wird dem Zufall überlassen. Es gibt keinen Raum für menschliche Freiheit, Verantwortung und Vernunft. Die andere Sichtweise – die Maschine ist wie ein Mensch – spiegelt sich in der anthropomorphen Sichtweise der KI wider, die Maschine „lernt“ oder „weiß“ – beides können „nur“ Menschen. Im animistischen Paradigma wird das KI-System mit geistigen Eigenschaften ausgestattet, wenn sein Output Verhalten dem des Menschen hinreichend ähnlich ist.

21.1.3 Herausforderungen

Das umfassende Programm des digitalen Humanismus unterscheidet sich deutlich von anderen Initiativen, die einzelne Aspekte der Digitalisierung ansprechen. So hat in den letzten Jahren zum Beispiel die Debatte um erklärbare KI (*explainable AI* oder *XAI*) stark zugenommen und wurde sowohl von der Industrie als auch von der Wissenschaft als ein wichtiges Thema aufgenommen. Ebenso gibt es zahlreiche Initiativen, die sich mit dem Schutz der Privatsphäre beschäftigen. Diese haben zu intensiven Forschungen und zur Entwicklung neuer Technologien im Bereich des maschinellen Lernens geführt (*privacy-preserving machine learning*). Ein anderes wichtiges Thema ist die Frage der Regulierung großer Internetplattformen und sozialer Netzwerke, die einen wesentlichen Einfluss auf den Nachrichtenfluss und damit auf die Meinungsbildung in der Gesellschaft haben können. Sie werfen wichtige Fragen nach geopolitischen Machtverhältnissen und wirtschaftlicher Monopolbildung auf.

Eine Besonderheit des digitalen Humanismus ist es, diese Fragen und Problemkreise in ihrer Gesamtheit zu diskutieren und die Zusammenhänge zwischen ihnen zu beleuchten. Es geht also zum Beispiel genau darum aufzuzeigen, wie der Schutz der Privatsphäre mit Fragen der Kontrolle von online Diskursen und der Verschiebung von Machtverhältnissen in einer Gesellschaft zusammenhängt. Und natürlich versucht der Digitale Humanismus Forschung und Innovation voranzutreiben, um auf derartige Fragen konstruktive Antworten zu geben. Dies führt auf der einen Seite zu enormen Herausforderungen für Unternehmen, aber auch zu neuen Fragestellungen für Entwickler*innen von Innovationen. Während das alte Modell der digitalen Innovation primär der Frage der Wirtschaftlichkeit und der Einhaltung gesetzlicher Bestimmungen verpflichtet ist, ergeben sich nun weit über diese traditionellen Rahmenbedingungen hinausgehende Fragen, die häufig überaus komplex sind.

Die Ziele des Digitalen Humanismus sind einfach und schnell erklärt, in der Umsetzung wird es aber kompliziert da die Herausforderungen groß sind (Neidhardt et al., 2022). Es geht nicht nur um Wissenschaft und Forschung. Da sich Technologie nicht von selbst entwickelt, sie wird von ungleichen gesellschaftlichen und wirtschaftlichen Verhältnissen

bestimmt, reicht das Spektrum der Aufgaben bis zur Politik. Der Digitale Humanismus bewegt sich daher in einem mehrdimensionalen Rahmen (Werthner, 2024):

- Verschiedene Problembereiche der Digitalisierung, von der Konzentration im Web bis hin zur staatlichen Souveränität oder Arbeit
- Unterschiedliche Disziplinen: Informatik und technische/ingenieurwissenschaftliche Disziplinen, Sozial- und Geisteswissenschaften; von der Analyse bis zur Systementwicklung
- Unterschiedliche Aktivitäten: angewandte und Grundlagenforschung, Entwicklung und Experimente, Innovation, Bildung, Kommunikation und schließlich politische Intervention
- Verschiedene gesellschaftliche AkteurInnen: von Universitäten und Schulen, Firmen und Interessensvertretungen und öffentlichen Institutionen bis hin zur Zivilgesellschaft.

Dies zeigt sich sehr gut in der aktuellen Entwicklung im Bereich der KI, als die jetzt aktuelle (und sicherlich nicht letzte) Welle der Digitalisierung, mit all ihrer Widersprüchlichkeit von hohem Potenzial versus problematische Entwicklungen. Das Versprechen der KI ist offensichtlich: ein mächtiges Werkzeug, intelligent, flexibel und leicht zu bedienen, dass uns Menschen bei der Entscheidungsfindung, bei technischer Arbeit und kreativen Aufgaben unterstützt bzw. ergänzt, etwa beim Erzeugen und Zusammenfassen von Texten, dem Beantworten von Fragen, beim Programmieren, und vieles mehr. Forschungsergebnisse zeigen, dass KI menschliche Fähigkeiten verbessern kann:

- Peng et al. (2023) zeigen, dass Software-IngenieurInnen mit GitHub Copilot doppelt so schnell sein können.
- Laut Noy und Zhang (2023) können Mitarbeiter mit geringerer Produktivität ihre Leistung bei Schreibaufgaben durch den Zugang zu ChatGPT verbessern.
- Brynjolfsson et al. (2023) diskutieren Verbesserungen der Arbeitsproduktivität, wenn ChatGPT ergänzend eingesetzt wird und zusätzliche Informationen für die MitarbeiterInnen bereitstellt.
- In einer Studie (Humlum & Vestergaard, 2024) mit Beschäftigten aus verschiedenen Berufen wird geschätzt, dass ChatGPT den Zeitaufwand für etwa ein Drittel der Arbeitsaufgaben im Durchschnitt halbieren kann.

Diese konkreten Anwendungsbeispiele zeigen den Charakter der KI als GPT (General Purpose Technology). Die Herausforderung – insbesondere auch für Unternehmen – ist nun, diese Technologie dazu einzusetzen, um menschliche Arbeit zu bereichern, anstatt bloß zu automatisieren. Dies hängt auch (oder vor allem) von den gesellschaftlichen, politischen und rechtlichen Rahmenbedingen ab (Acemoglu & Johnson, 2023).

21.2 Digitaler Humanismus und Innovation

21.2.1 Digitale Innovation und digitaler Fortschritt

Regulierung und ethische Überlegungen stehen nicht im Widerspruch zu Innovation, sondern können diese tatsächlich fördern, indem sie ein stabiles und vertrauenswürdiges Umfeld schaffen, in dem neue Technologien gedeihen können. Klare Vorschriften und ethische Leitlinien setzen Grenzen, die sicherstellen, dass Technologien auf eine Weise entwickelt und genutzt werden, die der Gesellschaft insgesamt zugutekommt, wodurch das Risiko von Schäden, Ausbeutung oder unethischen Praktiken verringert wird. Durch die Festlegung von Standards für Transparenz, Verantwortung und Sicherheit bieten Regulierungen den Innovatoren einen Rahmen, innerhalb dessen sie Produkte entwickeln können, die mit gesellschaftlichen Werten in Einklang stehen. Dies kann auch das Vertrauen der Öffentlichkeit stärken, was entscheidend für die Akzeptanz und den Erfolg neuer Technologien ist. Wenn die Menschen das Gefühl haben, dass eine Technologie sicher, fair und ethisch vertretbar ist, sind sie eher bereit, sie zu unterstützen, was wiederum weitere Innovationen vorantreibt.

Darüber hinaus fördert die Berücksichtigung von Ethik im Innovationsprozess eine langfristige und nachhaltige Entwicklung. Ohne die Berücksichtigung ethischer Aspekte könnten schnelle technologische Fortschritte unbeabsichtigte Folgen wie Datenschutzverletzungen, die Förderung von Ungleichheit oder Umweltschäden durch überhöhten Ressourcenverbrauch nach sich ziehen. Indem entsprechende Überlegungen frühzeitig in die Design- und Entwicklungsphasen einfließen, können Innovatoren diese Herausforderungen proaktiv angehen und Lösungen entwickeln, die positive gesellschaftliche Auswirkungen haben. Statt Kreativität zu ersticken, bieten Regulierungen und eine am Digitalen Humanismus orientierte Entwicklung einen Leitfaden für die Entwicklung von Technologien, die nicht nur die Grenzen des aktuell Möglichen erweitern, sondern auch sicherstellen, dass Innovationen dem Gemeinwohl dienen und der Gesellschaft in bedeutender Weise zugutekommen.

Die grundsätzliche und schwierige Frage ist dabei der Gegensatz zwischen einzelunternehmerischer versus gesamtgesellschaftlicher Innovation bzw. genügt für das allgemeine Wohlbefinden die Menge von Einzeloptimierungsschritten (nach dem bekannten Wohlfahrtsaxiom), oder benötigt es gesamtgesellschaftliche Rahmenbedingungen und Markteingriffe. Damit verbunden ist die Frage nach der Resilienz unseres Wirtschaftens, und zwar über einen längerfristigen Zeitraum (Werthner, 2024). Diese Frage stellte sich auch im Zusammenhang mit der COVID-19-Pandemie, wo viele Stimmen für eine grundlegende Veränderung unseres Wirtschaftssystems plädierten – es wirtschaftlich, sozial und ökologisch nachhaltiger zu gestalten. Organisatorisch, technisch und gesellschaftlich belastbare „Architekturen“ mit Puffern und Fehlertoleranz, die nicht auf Optimalität ausgerichtet sind, waren wahrscheinlich die wichtigsten Lehren. Dies ergibt sich nicht nur aus der Pandemie, sondern gilt insbesondere auch für die weitaus größere Energie- und Umweltkrise, eine der großen Herausforderungen unserer Zeit. Resilienz bedeutet, dass

die Funktionsfähigkeit eines Systems unter störenden internen oder externen Einflüssen aufrechterhalten werden kann und dass die Funktionsfähigkeit eines Systems nach einer Störung so schnell wie möglich wiederhergestellt werden kann. Resilienz ist also ein Gesamtsystembegriff.

Andererseits bedeutet Effizienz, optimale Ergebnisse zu erreichen, wobei die zur Verfügung gestellten Ressourcen bestmöglich verwendet werden; oder einfacher: Etwas ist effizient, wenn nichts verschwendet wird. Dies gilt und galt auch für digitale Wirtschaft, insbesondere dem E-Commerce. Interessanterweise zeigen Studien aber auch, dass bei Empfehlungssystemen das Ziel der kurzfristigen Optimalität, gemessen als hohe Präzision der Empfehlung oder hohe CTR (Click-Through-Rate) – also, wenn man dem Kunden das „richtige“ empfohlen hat und er schnell zugreift, längerfristig die Kunden-Retention senken kann. Oder nehmen wir das Internet als das Beispiel großer technischer und unternehmerischer Innovation, es wurde nach Resilienzkriterien und Prinzipien der Ausfallsicherheit entwickelt. Und es funktioniert. Dazu bedarf es aber quasi „Markt“ externer Anreize und Regulierungen, und als Basis dazu einen gesellschaftlichen Konsens der Nachhaltigkeit.

21.2.2 Innovationstreiber Digitaler Humanismus

Ein großer Teil der vom Digitalen Humanismus aufgeworfenen Fragen ist bis heute unbeantwortet geblieben, d. h. für viele kritische Themen liegen keine fertigen Rezepte vor, mit denen ein digitaler Humanismus einfach für digitale Innovationen garantiert werden kann. Wie oben bereits erwähnt sind viele Fragen so komplex, dass sie eigene Forschungsbereiche geworden sind, z. B. erklärbare KI oder privacy-preserving maschinelles Lernen. An dieser Entwicklung sind zwei Aspekte beachtenswert: Erstens gibt es tatsächlich für viele Fragen heute neue Technologien, die entweder schon Industriereife erlangt haben oder kurz davorstehen. Zweitens haben sich einige der Problemfelder als hartnäckig, aber gleichzeitig als besonders produktiv im technisch-wissenschaftlich Sinn erwiesen.

Für den ersten Bereich – also neue Technologien für Lösungen, die mit dem digitalen Humanismus kompatibel sind – ist es wichtig, neue Ansätze und Strategien auszuprobieren. Hier bedarf es leider immer noch einer gewissen Vorreiterrolle, selbst wenn Technologien de facto seit Jahren verfügbar sind. Ein einfaches Beispiel stellen etwa kryptografische Signaturen, Wasserzeichen und ähnliche Verfahren dar, um die Echtheit und Nachvollziehbarkeit von Nachrichten, Tonaufnahmen, Videos u. v. m. zu untermauern. Derartige Technologien stehen zur Verfügung, werden aber unzureichend verwendet. Auf diese Art kann die Glaubwürdigkeit und Vertrauenswürdigkeit in digitale Innovationen verbessert werden und es können wichtige Impulse gegen ein Ausufern von Falschnachrichten und verzerrende Informationen gesetzt werden. Hier gibt es aber auch noch viel Raum für Innovationen, z. B. um neue Werkzeuge, Methoden und Plattformen zu entwickeln, die diese Technologien nutzer*innenfreundlich verfügbar machen.

Für den zweiten Aspekt ist es unerlässlich, dass sich auch die industrielle Forschung der Fragen des digitalen Humanismus annimmt, um praxisnahe Antworten, Standards und Rahmen zu entwickeln, in der Praxis zu testen und breit auszurollen. Methoden und Ansätze, die helfen sollen, einen digitalen Humanismus zu realisieren, werden derzeit weltweit erforscht. Dazu gehören Fragestellungen, wie dramatische Machtverschiebungen weg von Konsumenten und zu einigen wenigen Plattformen vermieden werden können. Dies betrifft z. B. Empfehlungssysteme und die Fragen, wie die Modellbildung über User durch die UserInnen selbst kontrolliert werden kann. Allerdings ist ein großer Teil dieser Forschung primär wissenschaftlicher Natur und orientiert sich nicht immer an der Praxis. Hier bedarf es eines noch stärkeren Zugehens von Wissenschaft und Industrie aufeinander, z. B. bei Fragen, die gesellschaftliche Aspekte berühren und das Gemeinwohl betreffen. Dazu gehören z. B. KI-Systeme für seltene Sprachen oder KI, die unbedenklich von Kindern verwendet werden kann.

21.3 Digitaler Humanismus als Innovationsstrategie für Unternehmen

21.3.1 Risikomindernde Wirkung des digitalen Humanismus

Unternehmer*innen, die alle Risiken vermeiden, gelten prinzipiell nicht als besonders erfolgreich. Umgekehrt gibt es zahlreiche Beispiele von Unternehmen, die sich durch unbedachte oder unethische Produkte und Dienstleistungen einen schlechten Ruf erworben haben und auch wirtschaftliche Rückschläge hinnehmen mussten. Die Orientierung am digitalen Humanismus kann helfen, Unternehmensrisiken frühzeitig zu erkennen und entsprechend gegenzusteuern. Beispiele sind:

- PR-Risiken durch Verletzung von Rechten, Privatsphäre oder unfairer Praxis
- Gesetzliche Risiken durch Missachtung oder Verletzung von Grundrechten
- Regulatorische Risiken durch neue Regularien infolge öffentlicher Debatten
- Sicherheitsrisiken durch unzureichende Berücksichtigung schutzwürdiger Interessen Dritter
- etc.

Ähnliche Risiken treten etwa auch im Zusammenhang mit dem Einsatz von KI auf. Einen Überblick über die sich daraus ergebenden unternehmerische Risiken bietet (Prem, 2025). Die KI-Risiken sind eng verbunden mit der Verletzung von ethischen Prinzipien, wie z. B. Transparenz, Fairness, Privatsphäre, nicht-Malefizienz, Benefizienz u. v. m. Diese Prinzipien sind auch für den Digitalen Humanismus relevant, müssen aber um weitere relevante Grundsätze auf der gesellschaftlichen Ebene ergänzt werden. Dazu gehören z. B. die Berücksichtigung von demokratischen Prinzipien, der Schutz und die Weiterentwicklung sozialer Standards, staatlicher Souveränität oder die Vermeidung unerwünschter

Machtverschiebungen. Die Forderung nach einem digitalen Humanismus in digital orientierten Unternehmen bedeutet in der Praxis, die Fähigkeit zur Innovation in einen neuen, stärker als bisher ethisch orientierten Kontext unternehmerischen Handelns zu stellen.

21.3.2 Digitaler Humanismus als Kompetenzfaktor im Unternehmen

Die Komplexität der verschiedenen Themen des digitalen Humanismus, die Abschätzung der konkreten Folgen digitaler Innovationen und das Ineinandergreifen verschiedener Themenfelder gehören zu den größten Herausforderungen für Unternehmen, die dem digitalen Humanismus mehr Gewicht verleihen möchten. Die wirksame, d. h. über bloße Lippenbekenntnisse hinausgehende Berücksichtigung des digitalen Humanismus bedarf Kompetenzen, die selbst in großen Unternehmen nur wenig Beachtung fanden. Für viele kleinere Unternehmen ist die Situation aber noch schwieriger, da Personen mit technischer Expertise schon jetzt stark unter Druck sind und sich kaum auch noch zusätzlich mit der Problematik der Folgewirkungen von Innovationen bzw. mit speziellen Fachgebieten wie etwa dem Schutz der Privatsphäre, gerechten Empfehlungssystemen oder Fragen der Diversität in KI-Modellen auseinandersetzen können. Es stellt sich damit die Frage, wie Unternehmen entsprechende Kompetenzen erwerben und erhalten können.

Mit der Verbreitung des digitalen Humanismus beginnt eine neue Generation von Ingenieur*innen heranzuwachsen, die sich zunehmend auch bereits in der Ausbildung mit Fragen der Ethik, von Solidarität und Diversität bzw. des digitalen Humanismus beschäftigt. Eine steigende Anzahl von Ausbildungsstätten bietet Kurse und Kursprogramme an, die neben rein informatischen Kompetenzen auch Kompetenz in der Beurteilung von Technikfolgen, eine kritische Selbstreflexion technologischer Entwicklung und Methoden und Ansätze für Lösungen vermitteln, die mit dem digitalen Humanismus kompatibel sind.

Eine wichtige Voraussetzung für einen erfolgreichen Erwerb dieser Kompetenzen ist, dass Digitaler Humanismus als Wert vom Management vorgelebt wird. Dies kann durch schriftliche Unternehmensleitbilder geschehen. Wichtig ist aber auch die Definition von Strategien, die klare Ziele vorgeben und Methoden zur Fortschrittsmessung im Hinblick auf diese Ziele festlegen. Digitaler Humanismus sollte im Unternehmen keine vage Philosophie bleiben, sondern wird umso wirksamer gelebt werden können, je konkreter Fortschritte in diesem Bereich messbar gemacht werden. Hierzu wurden jüngst auch Indikatorsysteme vorgeschlagen, wie z. B. die *Digital Responsibility Goals* in Anlehnung an die Sustainability Development Goals der Vereinten Nationen (Identity Valley, 2025). Die DRGs wurden entworfen, um neben digitaler Kompetenz auch Faktoren wie die Cybersicherheit, Privatsphäre, Fairness im Bereich Daten, die Vertrauenswürdigkeit von Algorithmen, Transparenz und menschliche Verantwortung und Identität zu messen. Diese Indikatoren bedürfen der Anpassung und auch der Erweiterung, um den digitalen Humanismus in seiner ganzen Breite im Unternehmen zu erfassen. Sie zeigen aber, dass digitaler Humanismus messbar gemacht werden kann und damit auch operationalisierbar wird, um

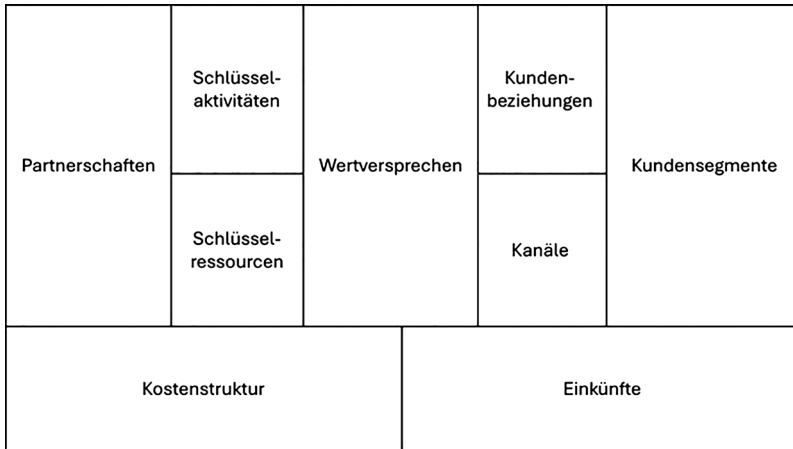


Abb. 21.1 Business Models Canvas

so den Mitarbeiter*innen im Unternehmen konkrete Anhaltspunkte für die eigene Arbeit, für Priorisierungen und den Kompetenzerwerb zu liefern.

Der digitale Humanismus kann auch als Leitfaden für die Entwicklung neuer Produkte/Services und damit verbesserte Wettbewerbsfähigkeit dienen. Als Beispiel diskutieren wir kurz eine mögliche Erweiterung des Business Model Canvas (Abb. 21.1), die zurzeit wahrscheinlich meistbenutzte Methode zur Ideengenerierung und kompakten Darstellung eines neuen Geschäftsmodells (Osterwalder et al., 2010).

Der Canvas veranschaulicht die Komplexität und die verschiedenen Dimensionen einer neuen Idee, und gleichzeitig auch die Möglichkeit einer einfachen Darstellung. Er stellt quasi die Innensicht dar und vernachlässigt daher die gesamtgesellschaftliche Dimension, die im Digitalen Humanismus aber auch wichtig ist. Eine Alternative wäre das i2 value Modell, das die Vernetzungssicht eines Unternehmens widerspiegelt (Gordijn & Akkermans, 2001). Die einzelnen Kategorien des Canvas wie Partner, Kundensegmente oder Wertversprechen machen klar, dass die Prinzipien des Digitalen Humanismus orthogonal zu ihnen stehen, diese also als eigene Kategorie integriert werden können, oder in den einzelnen Kategorien als zusätzliche Kriterien dargestellt werden müssten (mit entsprechenden Auswirkungen auf die Kostenstruktur und Einnahmemöglichkeiten).

Eine Erweiterung des Canvas basierend auf dem adaptierten Modell (Abb. 21.2) von Schallmo und Brecht (2010) wurde in Prem (2015) mit dem Ziel präsentiert, typische Effekte der Digitalisierung zu verdeutlichen, z. B. Effekte wie Co-Kreation von Inhalten oder die Einführung bidirektionaler Kommunikationskanäle. Der erweiterte Canvas liefert auch eine Präzisierung digitalisierter Komponenten des Modells, z. B. digitale Kommunikationskanäle, auf Vorhersage oder Datenanalytik beruhende Wertversprechen oder den Übergang von Produkten zu Dienstleistungen.

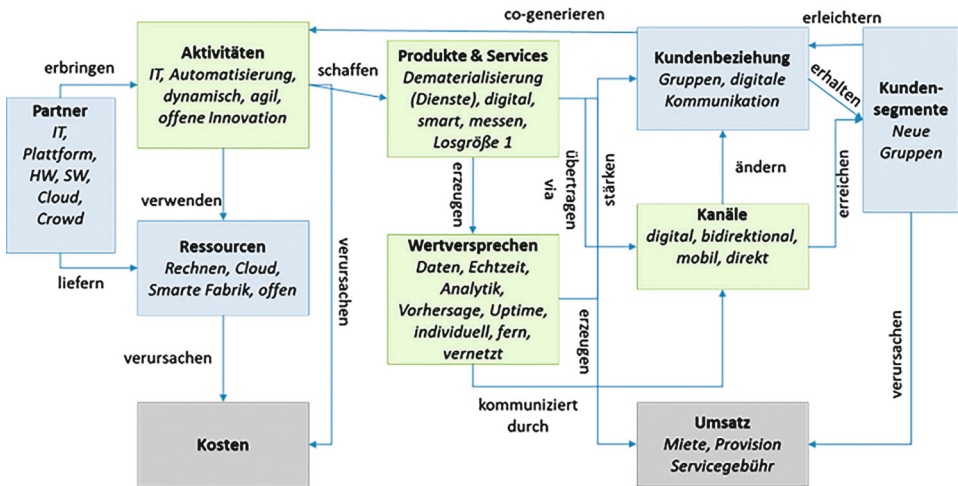


Abb. 21.2 Erweiterung Business Models Canvas. (Quelle: nach Prem, 2015)

Abb. 21.2 verdeutlicht typische Effekte der Digitalisierung, d. h. typische Formen einzelner Komponenten (z. B. Vorhersage als Teil des Wertversprechens – etwa KI oder Big Data) und neue Zusammenhänge wie z. B. die Co-Kreation von Inhalten nach Einführung digitaler Kommunikationskanäle.

Diese Aspekte digitalisierter Modellkomponenten haben wichtige ethische und gesellschaftliche Implikationen, z. B. die Vorhersage menschlichen Verhaltens oder der Verlust an Eigentumsrechten durch den Übergang zu Dienstleistungen. Sie sind im Modell zwar explizit bezeichnet, aber nur schlecht als problematisch zu erkennen, da hier keine gesellschaftlichen Elemente repräsentiert sind und ethisch problematische Beziehungen ebenfalls nicht auf den ersten Blick deutlich werden. Diese Übung zeigt, dass die Berücksichtigung gesellschaftlicher Kriterien und Zielsetzungen eines neuen und erweiterten Ansatzes bedarf, der die auf die traditionellen Elemente des Business Model Canvas beschränkte Sicht einzelner Unternehmen sprengt.

21.3.3 Digitale Technologie- und Gesellschaftspolitik als Verantwortung der Industrie

Abschn. 21.3.2 zeigt, dass es die Prinzipien des digitalen Humanismus in vielen Fällen notwendig machen, dass sich Unternehmen vermehrt mit der gesellschaftlichen Wirkung ihrer Aktivitäten und geplanten Innovationen auseinandersetzen und eine weitere Sicht benötigen. Dies betrifft einerseits die Abschätzung der Wirkungen digitaler Innovation auf Mensch, Umwelt und Gesellschaft. Andererseits gehört dazu auch, Technologieentwicklung als politische Aktivität zu erkennen. Unternehmen im Bereich der Digitalisierung

müssen stärker als bisher Verantwortung für ihr Handeln als zumindest implizite Technologie- und Gesellschaftspolitik übernehmen.

Dies ist eine Folge der immensen Wirkung digitaler Systeme und Netzwerke auf Individuen und gesellschaftliche Strukturen. Da jedoch große Macht nicht nur große moralische Verantwortung bedeutet, sondern auch mit den erwähnten unternehmerischen Risiken einhergeht ist die Diskussion von Fragen des Digitalen Humanismus letztendlich auch auf Unternehmensebene unausweichlich.

21.4 Schlussfolgerung und Empfehlung

Mit der zunehmenden Berücksichtigung des digitalen Humanismus verändern sich die Aufgaben für das Management im Unternehmen. Der digitale Humanismus bietet Unternehmen wichtige Grundwerte, an denen Strategien neu ausgerichtet werden können. Schon bisher ging in vielen Industrien die Formulierung von Unternehmensstrategien, Werten und Missionen über bloße Wirtschaftlichkeitsüberlegungen und Fragen der Compliance hinaus. Unternehmensstrategien haben oft versucht durch Firmenaktivitäten einen Beitrag zu gesellschaftlichem Fortschritt zu leisten. Mit dem digitalen Humanismus erlangen zusätzliche Aspekte Bedeutung. Er liefert wichtige Rahmenbedingungen für Innovationen, weil Digitaler Humanismus gesamtgesellschaftliche Fragen aufwirft, nicht nur einzelunternehmerische.

Dadurch entstehen zusätzlich auch politische Fragen, die in einen demokratischen Diskurs eingebettet werden müssen. Das Thema Innovation zeigt, dass Digitaler Humanismus eine systemische Frage ist, in der einzelne Akteure, die mit moralischen Zielvorstellungen agieren, wichtig sind. Gleichzeitig bleiben zahlreiche Fragestellungen aber auf gesellschaftlicher Ebene als allgemeine Rahmenbedingungen zur regeln.

Literatur

- Acemoglu, D., & Johnson, S. (2023). *Macht und Fortschritt: Unser 1000-jähriges Ringen um Technologie und Wohlstand*. Campus.
- Brynjolfsson, E., Li, D., & Raymond, L. (2023). *Generative Ai at Work*. NBER Working Paper No. w31161.
- Digital Humanism (Verein) (2025). Offizielle Webseite. Wien: Digital Humanism. <https://digitalhumanism.at/>. Zugegriffen: 6. Apr. 2025.
- Digital Humanism Initiative (2025). *TU Wien: The Digital Humanism Initiative*. Wien: TU Wien. <https://dighum.ec.tuwien.ac.at/> (abgerufen am 6. April 2025)
- Gordijn, J., & Akkermans, H. (2001). Designing and Evaluating E-Business Models. *IEEE intelligent Systems*, 16(4), 11–17.
- Humlum, A., & Vestergaard, E. (2024). *The Adoption of ChatGPT*. IZA Discussion Paper, 16992.
- Identity Valley (2025). *Digital Responsibility Goals (DRGs): Die sieben Ziele für eine digitale Verantwortung*. München: Identity Valley. <https://identityvalley.eu/de/drg> (zuletzt abgerufen am 5. April 2025)

- Neidhardt, J., Werthner, H., & Woltran, S. (2022). It Is Simple, It Is Complicated. In H. Werthner, E. Prem, A. Lee & C. Ghezzi (Hrsg.), *Perspectives on Digital Humanism*. Springer.
- Nida-Rümelin, J., & Staudacher, K. (2024). Philosophical Foundations of Digital Humanism. In H. Werthner, C. Ghezzi, J. Kramer, J. Nida-Rümelin, B. Nuseibeh, E. Prem & A. Stanger (Hrsg.), *Introduction to Digital Humanism. A Textbook*. Springer.
- Noy, S., & Zhang, W. (2023). Experimental evidence on the productivity effects of generative artificial intelligence. *Science*, 381(66), 187–192.
- Osterwalder, A., Pigneur, Y., & Clark, T. (2010). *Business Model Generation: A Handbook for Visionaries, Game Changers, and Challengers*. Hoboken: John Wiley & Sons.
- Peng, S., Kalliamvakou, E., Cihon, P., & Demirel, M. (2023). *The impact of ai on developer productivity: Evidence from github copilot*. arXiv preprint arXiv:2302.06590.
- Prem, E. (2015). A digital transformation business model for innovation. In *Proceedings ISPIM Innovation Summit*. Brisbane, 6–9 December 2015. <https://www.ispim.org>.
- Prem, E. (2024). Principles of digital humanism: a critical post-humanist view. *Journal of Responsible Technologies*, 17, 100075.
- Prem, E. (2025). Managing entrepreneurial AI risks. In C. A. Hoffman (Hrsg.), *Artificial Intelligence, Entrepreneurship, and Risk Management: Reflections and Positions at the Crossroads between Philosophy and Management*. Springer.
- Schallmo, D., & Brecht, L. (2010). Business model innovation in business-to-business markets—procedure and examples. In *Proceedings of the 3rd ISPIM Innovation Symposium: Managing the Art of Innovation: Turning Concepts into Reality*. Quebec City. (S. 12–15).
- Werthner, H. (2024). *Digitaler Humanismus – Über Digitalisierung und Künstliche Intelligenz*. Wien: Picus.

DDr. Erich Prem, MBA, ist Chef FTI-Strategie und Geschäftsführer von eutema (www.eutema.com). Er arbeitet wissenschaftlich in den Gebieten AI, Technikethik und Innovationsforschung. Er war Gastforscher am Massachusetts Institute of Technology, Wissenschaftler am österreichischen Forschungsinstitut für AI und an der Universität Wien (Technikphilosophie). Er ist Obmann des Vereins für Digitalen Humanismus und lehrt Digitalen Humanismus an der TU Wien.

Hannes Werthner ist pensionierter Informatikprofessor der TU Wien, wo er auch Dekan der Fakultät für Informatik war. Vor der TU Wien hatte er Professuren an österreichischen und internationalen Universitäten inne. Seine Forschung umfasst verschiedene Bereiche wie Decision Support Systems, E-Commerce, E-Tourismus und Empfehlungssysteme. Neben Forschung und Lehre engagiert er sich aktiv für die Gründung neuer Initiativen wie der Vienna PhD School of Informatics und des i2c (Informatics Innovation Center). Er beobachtet aufmerksam, wie die IT die Welt und uns verändern. Dies ist seine Motivation für den Digitalen Humanismus, wo er 2019 den Wiener Workshop des Digitalen Humanismus organisierte und das Wiener Manifest initiierte. Er ist Autor des Buches „Digitaler Humanismus – über Künstliche Intelligenz und Digitalisierung“, das 2025 im Picus Verlag erschienen ist.

Der EU AI Act: Aus der Pflicht eine Tugend machen

22

AI Leadership Compass zur Kombination von Ethik, Compliance und Betriebswirtschaft

Stefan Tauber und Martin Giesswein

Zusammenfassung

Ausgehend von der humanistischen Digitalpolitik der EU legt dieses Kapitel den Schwerpunkt auf den AI Act und seine Implikationen für Unternehmen. Es beleuchtet die ideale Balance zwischen regulatorischer Verpflichtung und ethischem Gestaltungswillen – eine Verbindung von rechtlichem Sollen und moralischem Wollen. Für das Management ergibt sich daraus ein klarer Imperativ: die bestmögliche Strategie zu gestalten und umzusetzen, die gesellschaftlichen und unternehmerischen Interessen gerecht wird. Als konkrete Hilfestellung wird der AI Leadership Compass strukturiert erklärt und auf vier Ebenen erläutert: strategische Führung, operative Führung, HR-Management und Selbstführung.

Aus den regulatorischen Vorgaben ein ethisches und betriebswirtschaftliches Optimum zu machen ist die hohe Kunst des digitalhumanistischen Managements.

S. Tauber (✉)

Schiefer Rechtsanwälte GmbH, Wien, Österreich

E-Mail: office@schiefer.at

M. Giesswein

WU Executive Academy, Wien, Österreich

E-Mail: martin@giesswein.org

22.1 Die Digitalpolitik der EU als Vorgabe

Die digitalhumanistische Realität unserer Unternehmen wird maßgeblich durch gesetzliche Rahmenbedingungen geprägt. Besonders die Regularien der Europäischen Kommission haben derzeit einen starken Einfluss auf die betriebswirtschaftliche Praxis. Um sich als Manager in dieser komplexen regulatorischen und ethischen Landschaft sicher zu bewegen, ist es essentiell, die Grundlagen der digitalhumanistischen Ausrichtung der europäischen Gesetzgebung und die Zielsetzungen der von der Europäischen Kommission ausgerufenen digitalen Dekade (Ziele für 2030) zu verstehen.

22.1.1 Europäische Leitplanken für eine menschenzentrierte Digitalisierung

Einer der tragenden Referenzpunkte der EU-Legistik ist die Europäische Erklärung zu den digitalen Rechten und Grundsätzen (Europäische Kommission, 2022). Mit diesem Werk aus dem Jahr 2022 hat die Europäische Union eine Vision für die digitale Transformation formuliert, die den Menschen in den Mittelpunkt stellt und auf europäischen Werten beruht. Unterzeichnet von den Präsidenten der Europäischen Kommission, des Europäischen Parlaments und des Rates, unterstreicht sie das hohe politische Engagement der EU und ihrer Mitgliedstaaten für eine verantwortungsbewusste digitale Zukunft.

Die Erklärung basiert auf der EU-Grundrechtecharta (Europäische Union, 2012) und hebt relevante Rechte in der digitalen Transformation hervor, wie die Meinungs- und Informationsfreiheit, den Datenschutz und die Privatsphäre. Sie umfasst sechs zentrale Prinzipien:

1. Den Menschen und seine Rechte in den Mittelpunkt der digitalen Transformation stellen: Technologie soll allen Menschen in der EU dienen und ihre Rechte respektieren.
2. Solidarität und Inklusion unterstützen: Jeder soll Zugang zu erschwinglicher und schneller digitaler Konnektivität haben und die notwendigen digitalen Fähigkeiten erwerben können.
3. Freiheit der Wahl im Internet gewährleisten: Nutzer sollen befähigt werden, informierte Entscheidungen online zu treffen, insbesondere im Umgang mit künstlicher Intelligenz.
4. Teilnahme am digitalen öffentlichen Raum fördern: Aktive Beteiligung am digitalen Leben und an demokratischen Prozessen soll unterstützt werden.
5. Sicherheit, Schutz und Befähigung von Individuen erhöhen: Besonderes Augenmerk liegt auf dem Schutz von Kindern und Jugendlichen im digitalen Umfeld.
6. Nachhaltigkeit der digitalen Zukunft fördern: Digitale Produkte und Dienstleistungen sollen nachhaltig gestaltet, produziert und entsorgt werden, um die Umwelt zu schonen.

Mit dieser Erklärung verpflichtet sich die EU, eine menschenzentrierte, sichere und nachhaltige digitale Transformation sowohl innerhalb der Union als auch international zu för-

dern. Die Kommission überwacht die Anwendung der digitalen Rechte und Grundsätze in der gesamten EU und veröffentlicht jährliche Berichte über den Fortschritt.

22.1.2 Globale Strategien im Vergleich: Europa, USA und China

Die Europäische Union hat in den letzten Jahren mehrere wichtige Gesetze und Verordnungen im Lichte der Erklärung im Bereich der digitalen Regulierung verabschiedet, die in engem Zusammenhang stehen und sich gegenseitig ergänzen.

Die wichtigsten davon sind der Digital Services Act (DSA; Europäische Union, 2022b), der Digital Markets Act (DMA; Europäische Union, 2022a) und der Data Act (Europäische Union, 2023). Diese Gesetze bilden zusammen einen umfassenden Rahmen für die Regulierung digitaler Dienste und Märkte in der EU.

Der Digital Services Act ist am 1. November 2022 in Kraft getreten und gilt seit dem 2. Mai 2023. Der DSA zielt darauf ab, mehr Wettbewerb und Fairness in digitalen Märkten zu schaffen. Insbesondere soll der DSA die bewusste Irreführung der Nutzer verhindern und die freie Meinungsäußerung schützen. Zusätzlich regelt der DSA den Zugang zu bestimmten Daten für Forschende und Behörden, um Transparenz und Rechenschaftspflicht der großen Plattformen sicherzustellen.

Der Digital Markets Act hingegen zielt darauf ab, die Macht von großen Digitalkonzernen zu beschränken, die als sogenannte Gatekeeper (Torwächter) den digitalen Markt beherrschen. Die Europäische Kommission hat für solche Unternehmen einen strengen Verhaltenskodex aufgesetzt, um eine zu dominante Marktposition zu verhindern. Kriterien wie Umsatz, Unternehmenswert und die Zahl der monatlichen Endnutzer sind dabei relevant für die Einstufung als Gatekeeper. Der DMA verpflichtet diese Gatekeeper als sehr große Plattformen und Suchmaschinen, fairen Zugang zu den auf ihren Plattformen erhobenen oder generierten Daten zu gewähren. Dies dient dazu, den Wettbewerb zu fördern und kleinen sowie mittleren Unternehmen die Möglichkeit zu geben, auch innovative Dienstleistungen anzubieten.

Zum Zeitpunkt des Erstellens dieses Kapitels in 2025 gibt es folgende Gatekeeper: Alphabet (Google), Amazon, Apple, ByteDance (TikTok), Meta (Facebook, Instagram, WhatsApp), Microsoft, Booking.

Als eine weitere zentrale legislative Maßnahme der EU-Digitalpolitik wurde der Data Act (DA) im Jahre 2023 vom Rat der Europäischen Union verabschiedet. Als unmittelbar geltendes EU-Recht schafft er die Grundlage dafür, Daten in verschiedenen Lebensbereichen effektiver zu nutzen. Ziel ist es, die Wertschöpfung zu steigern – insbesondere durch neue Geschäftsmodelle, die Förderung von Start-ups sowie die Stärkung kleiner und mittlerer Unternehmen (KMU).

All diese beschriebenen Gesetze stehen in engem Zusammenhang mit der bereits bestehenden Datenschutz-Grundverordnung (DSGVO; Europäische Union, 2016).

Die europäische Digitalpolitik verfolgt also das Ziel, eine digitalhumanistische Entwicklung in den Mitgliedstaaten zu fördern – mit einem klaren Bekenntnis zur Digital-

souveränität und zur Stärkung der europäischen Wirtschaft. Diese Strategie steht in einem grundlegenden Spannungsverhältnis zu den Ansätzen der USA und Chinas:

Die USA setzt auf Deregulierung und nimmt ein Oligopol digitaler US-Großkonzerne in Kauf, um durch die enge Verzahnung von Staat und Digitalgiganten die globale digitale Vorherrschaft zu sichern sowie die Vorreiterrolle weiter auszubauen. Die Amerikanische Digitalpolitik folgt der Idee, dass durch einen möglichst liberalen Markt und dem freien Spiel der Kräfte viel Neues und Innovation entsteht. Nicht immer unbedingt folgenlos: So betrachtet die World Health Organization die Auswirkungen von Social Media auf Jugendliche in Studien durchaus kritisch (WHO Regional Office for Europe, 2024). Anerkannte Wissenschaftler setzen sich immer intensiver mit dieser Thematik auseinander, warnen vor Langzeitfolgen und fordern, dass das digitale Umfeld kindlichen und jugendlichen Bedürfnissen gerecht wird (Aiken, 2018). Diese Überlegungen greift Australien auf. Seit dem 10. Dezember 2025 ist in Australien für Kinder und Jugendliche Social-Media verboten.

China hat ein stark kontrollierendes System etabliert. Dort greift der Staat gezielt in digitale Unternehmen ein, um Transparenz und Steuerungsmöglichkeiten zu gewährleisten – nicht zuletzt im Hinblick auf die Kontrolle der Bevölkerung. Als Negativbeispiel kann hier das chinesische Sozialkredit-System (Social Scoring) angeführt werden. Im Wesentlichen geht es China dabei darum, dass in diesem System eine große Anzahl an Daten zu jeder Person gesammelt und ausgewertet werden. Je nachdem, welches gewünschte bzw. unerwünschte Verhalten gezeigt wird, erhält bzw. verliert eine Person Punkte. Der Punktestand hat Auswirkungen auf praktisch alle Lebensbereiche. Aufgrund des AI Acts in Europa ist ein solches staatliches Vorgehen nicht mehr denkbar, da Social Scoring als verbotene Praktik gemäß Art. 5 Abs. 1 lit. c) AI Act gilt (siehe dazu weiter unten).

Will die interessierte Öffentlichkeit wirklich den europäischen Weg in der Digitalpolitik? Ein Indiz für die Bejahung dieser Frage ist die im Vorfeld des AI Action Summit 2025 durchgeführte globale Konsultation (The Future Society, 2024). Sie lieferte wichtige Erkenntnisse zur KI-Governance. Ziel war es, die Perspektiven von Bürgern, Experten und zivilgesellschaftlichen Organisationen zu integrieren, um eine gerechte und inklusive Entwicklung von Künstlicher Intelligenz (KI) zu gewährleisten. Mit einer breiten Beteiligung von über 11.000 Menschen und 200 Experten wird ein wachsendes Bewusstsein für die Chancen und Risiken von KI sichtbar.

Schlüsselthemen und Empfehlungen:

1. Bürgerengagement und Bildung: Die Forderung nach transparenter Bildung und ethischem Umgang mit KI zieht sich durch alle Konsultationen. Es wird vorgeschlagen, KI-Trainingsprogramme in Schulen, Universitäten und Berufsausbildungen zu integrieren.
2. Globale KI-Governance: Experten fordern ein multilaterales Gremium ähnlich dem Weltklimarat (Intergovernmental Panel on Climate Change, 2025) für den Klimaschutz, um internationale Standards zu setzen, Transparenz zu fördern und die Beteiligung des globalen Südens sicherzustellen. Mechanismen zur Kontrolle großer KI-Modelle, wie Schwellenwerte und Sicherheitsforschung, sollen etabliert werden.

3. Nachhaltige Entwicklung: KI sollte gezielt genutzt werden, um globale Herausforderungen wie den Klimawandel oder soziale Ungleichheiten anzugehen. Gleichzeitig müssen die ökologischen Auswirkungen der Technologie durch effizientere Modelle und nachhaltige Ressourcenstrategien minimiert werden.
4. Vertrauen in KI: Transparenz, Sicherheit und Nachvollziehbarkeit werden als essenziell erachtet. Vorschläge umfassen die Einführung eines Bewertungssystems für KI-Unternehmen und standardisierte Prüfverfahren für KI-Modelle.
5. Zukunft der Arbeit: Angesichts potenzieller Arbeitsplatzverlagerungen durch KI müssen Umschulungen und Übergangsprogramme gefördert werden, um den Menschen neue Chancen in der digitalen Wirtschaft zu bieten.

Der Bericht betont die Notwendigkeit eines langfristigen, inklusiven Ansatzes, um KI als Werkzeug für das Gemeinwohl zu nutzen. Dabei ist es entscheidend, dass ethische Prinzipien, technologische Innovation und soziale Gerechtigkeit im Einklang stehen. Gerade in der korrekten Nutzbarmachung dieser Technologie liegt für Europa eine große Chance.

22.2 Der AI Act im Fokus des Managements

22.2.1 Der AI Act als digitalhumanistischer Rechtsrahmen

Ausgehend von der oben beschriebenen Nachfrage nach KI-Governance und der gegebenen Ausrichtung der EU-Digitalpolitik ist der AI Act (Europäische Kommission, 2025) eine natürliche Umsetzung dieser Prinzipien. Der AI Act zeigt eine klare digitalhumanistische Ausrichtung, indem er den Menschen und seine Rechte in den Mittelpunkt stellt und gleichzeitig Innovation und die verantwortungsvolle Entwicklung von KI-Technologien fördert. Er bietet die Chance, eine ethische und menschenzentrierte KI-Landschaft in Europa zu schaffen und könnte als Vorbild für ähnliche Regelungen weltweit dienen. Folgende Bestimmungen des Gesetzes sind für die Unternehmensführung besonders relevant.

Breite Definition von KI: Der AI Act legt eine breite Definition von KI zugrunde, wonach ein KI-System ein maschinengestütztes System ist, das für einen in wechselndem Maße autonomen Betrieb ausgelegt ist, nach seiner Einführung anpassungsfähig sein kann und aus den erhaltenen Eingaben Ergebnisse wie etwa Vorhersagen, Inhalte, Empfehlungen oder Entscheidungen hervorbringt, die physische oder virtuelle Umgebungen beeinflussen können. Diese umfassende Definition wird dazu führen, dass viele IT-Systeme, insbesondere im Zweifel, vom Management als KI bezeichnet werden müssen.

Anfang 2025 wurde die lang erwartete Leitlinie (Europäische Union, 2024) zur Definition von KI-Systemen gemäß Art. 3 Z 1 iVm Art. 96 Abs. 1 lit. f) AI Act veröffentlicht.

Nach dieser Leitlinie liegen KI-Systeme insbesondere dann vor, wenn maschinelles Lernen, (un-)überwachtes Lernen, selbst überwachtes Lernen, verstärkendes Lernen oder tiefgehendes Lernen als Techniken für Schlussfolgerungen verwendet werden. Logik und wissensbasierte Ansätze, bei denen das KI-System nicht aus Daten, sondern aus Wissen, einschließlich Regeln, Fakten und Beziehungen lernt, sollen auch umfasst sein. Keine KI-Systeme im Sinne des AI Act sind z. B. Systeme zur mathematischen Optimierung, Systeme zur einfachen Datenverarbeitung ohne Lernen, Modellieren oder Schlussfolgern, Systeme zur Anwendung heuristischer Methoden oder Systeme zur einfachen Vorhersage.

Weiters gibt es eine Leitlinie gemäß Art. 5 iVm Art 96, Abs. 1 lit. b) AI Act, welche als Auslegungshilfe zur Einstufung von verbotenen Praktiken herangezogen werden kann. So soll z. B. die Bewertung des Risikos oder die Vorhersage der Wahrscheinlichkeit, dass eine Person ein Verbrechen begeht, als verbotene Praktik i. S. d. Art. 5 Abs. 1 lit. d) AI Act angesehen werden. Wer diese Leitlinie allerdings zur Gänze lesen möchte, braucht etwas Zeit. Sie umfasst 135 Seiten und wird dementsprechend in der einen oder anderen Governance- bzw. Compliance-Abteilung für viel Lesestoff sorgen.

Bei beiden genannten Schriftstücken handelt es sich jeweils um eine unverbindliche Leitlinie, die Klarheit bei der praktischen Umsetzung schaffen soll. Die Praxis wird allerdings zeigen, wie „unverbindlich“ diese Leitlinien tatsächlich sind. Geben sie doch Auskunft darüber, wie der Gesetzgeber den AI Act interpretiert wissen will. Es wird daher davon auszugehen sein, dass sich Behörden und Gerichte an diesen Leitlinien orientieren und ihre Entscheidungen darauf aufbauen werden. Für den Normunterworfenen eine nicht ganz einfache Situation: Gegen diese „unverbindlichen“ Leitlinien bestehen nur in Einzelfällen Rechtsschutzmöglichkeiten. So hat der EuGH in der Vergangenheit formal unverbindliche Rechtsakte wie Empfehlungen und Leitlinien der Europäischen Kommission als Klagegegenstand akzeptiert, wenn diese formal unverbindlichen und damit grundsätzlich nicht klagbaren Handlungen den Anschein der Wiedergabe von unionsrechtlichen Verpflichtungen erweckten, die tatsächlich nicht bestehen; der so entstandene Rechtsschein kann mit Nichtigkeitsklage beseitigt werden (vgl. EuGH 20.02.2018, Rs C-16/16 P, Belgien/Kommission; Europäischer Gerichtshof, [2018](#)). Andernfalls kann nur auf eine Änderungen der Leitlinien hingewirkt werden: Auf Ersuchen der Mitgliedstaaten oder des Büros für Künstliche Intelligenz oder von sich aus kann die Europäische Kommission gemäß Art. 96 Abs. 2 AI Act die verabschiedeten Leitlinien, wenn es von der Europäischen Kommission als notwendig erachtet wird, überarbeiten.

Weitere Leitlinien zu anderen Teilen des AI Acts sind zu erwarten. Der Faktor Rechtssicherheit für Managemententscheidungen ist also nicht im gewünschten Ausmaß gegeben.

Einbeziehung der Betreiber: Der Anwendungsbereich des AI-Acts ist derart weit formuliert, dass praktisch alle Unternehmen in der Wertschöpfungskette betroffen sind, nicht nur die Anbieter von KI-Systemen. Jede nicht-private Nutzung eines KI-Systems, zum Beispiel der Einsatz von Microsoft Copilot im Büroalltag, führt zum Status als „Betreiber“ für das jeweilige Unternehmen, Behörde oder sonstige Organisation.

22.2.2 Risikokategorien, Verbote und Pflichten für Unternehmen

Risikosystem und Verbot bestimmter KI-Anwendungen: Grundsätzlich sieht die Gesetzgebung vor, künstliche Intelligenz in vier Fällen zu verbieten. Zur Auslegungshilfe der Europäischen Kommission siehe oben. Hierzu zählen:

1. KI-Systeme, die das Verhalten von Menschen so beeinflussen, dass ihnen oder anderen dadurch körperlicher oder seelischer Schaden zugefügt wird.
2. KI-Anwendungen, die die Schwächen von bestimmten Menschen, zum Beispiel aufgrund ihres Alters, einer Behinderung oder ihrer besonderen sozialen oder wirtschaftlichen Lage, ausnutzen.
3. KI-Systeme, um Menschen über einen bestimmten Zeitraum basierend auf ihrem Sozialverhalten oder bekannten bzw. vorhergesagten persönlichen Eigenschaften zu bewerten oder zu klassifizieren (Social Scoring).
4. Biometrische Fernidentifizierungssysteme an öffentlich zugänglichen Orten durch oder im Namen von Strafverfolgungsbehörden zum Zwecke der Strafverfolgung, es sei denn, eine solche Verwendung ist für die nationale Sicherheit von Bedeutung.

Neben diesen verbotenen KI-Systemen gibt es Hochrisikosysteme. Anwendungsfälle dafür sind: KI-gestützte Vergabe von Lebensversicherungen oder Privatkrediten, sowie Fälle im Recruiting, Zugang zur betrieblichen Weiterbildung und Karrierewegen, sowie die KI-gestützte Beurteilung im Bildungssektor. Für Anbieter und Betreiber solcher Systeme gelten strenge Pflichten: Sie müssen ein Risikomanagementsystem einrichten, eine technische Dokumentation erstellen, eine automatische Ereignisprotokollierung implementieren und das System in einer EU-Datenbank registrieren. Zudem sind sie verpflichtet, schwerwiegende Vorfälle zu melden und spezifische Transparenz- und Robustheitsanforderungen zu erfüllen.

KI-Systeme mit begrenztem Risiko beziehungsweise minimalem oder keinem Risiko bilden die unteren beiden Stufen der Risikokategorien im AI Act. Systeme mit begrenztem Risiko, wie beispielsweise Chatbots oder KI-Anwendungen, die direkt mit Nutzern interagieren, unterliegen hauptsächlich Transparenzpflichten. Anbieter und Betreiber müssen hier sicherstellen, dass Nutzer darüber informiert werden, dass sie mit einem KI-System kommunizieren. Diese Offenlegung soll es Menschen ermöglichen, informierte Entscheidungen über die Nutzung zu treffen. Letztlich werden im AI Act KI-Systeme mit minimalem oder keinem Risiko definiert, zu denen etwa Spam-Filter in E-Mail-Programmen zählen. Für diese Kategorie sieht der AI Act keine spezifischen Vorschriften oder Anforderungen vor. Es wird davon ausgegangen, dass diese beiden letzten Stufen $\frac{3}{4}$ aller KI-Anwendungsfälle betreffen. Für alle Risikostufen gilt allerdings die folgende KI-Kompetenz-Pflicht.

Förderung von KI-Kompetenz: Alle Unternehmen und Behörden sind verpflichtet, ihre Mitarbeitenden und Partner, die mit KI-Systemen in Berührung kommen, gezielt zu schulen. Diese Schulungen sollten einen rechtlichen, technischen, funktionalen und ethisch korrekten Umgang mit Künstlicher Intelligenz vermitteln. Dafür ist es sinnvoll, zunächst zu analysieren, ob bzw. welche KI-Systeme im Unternehmen genutzt werden. Auf dieser Basis können spezifische Nutzergruppen gebildet werden, die kontinuierlich einsatzorientierte Weiterbildungen erhalten. Da die meisten derzeit im Einsatz befindlichen KI-Systeme einem schnellen und starken Wandel unterworfen sind, werden wahrscheinlich nur laufende und aktualisierte Weiterbildungen den gewünschten Effekt bringen. In diesem Zusammenhang sind für Unternehmen auch interne Governance- und Compliance-Leitlinien empfehlenswert, welche den grundsätzlichen Umgang der Mitarbeiter mit KI-Systemen festlegen.

Menschliche Aufsicht

Der AI Act stellt sicher, dass KI-Systeme nicht autonom kritische Entscheidungen treffen, ohne dass Menschen eingreifen können. Besonders bei Hochrisiko-KI ist eine angemessene menschliche Kontrolle erforderlich, um potenzielle Fehler oder unerwartete Entwicklungen zu erkennen und zu korrigieren. Dies bedeutet, dass KI-Anwendungen so gestaltet sein müssen, dass Menschen ihre Funktionsweise verstehen, ihre Entscheidungen hinterfragen und bei Bedarf eingreifen können. In diesem Zusammenhang sei auf Art. 22 Abs. 1 DSGVO verwiesen, wonach die betroffene Person grundsätzlich das Recht hat, nicht einer ausschließlich auf einer automatisierten Verarbeitung – einschließlich Profiling – beruhenden Entscheidung unterworfen zu werden, die ihr gegenüber rechtliche Wirkung entfaltet oder sie in ähnlicher Weise erheblich beeinträchtigt.

Transparenzgebot: Anbieter und Betreiber von KI-Systemen unterliegen verschiedenen Offenlegungspflichten, wenn Künstliche Intelligenz zur Erstellung von Texten, Bildern, Videos oder Audiodateien eingesetzt wird. Besonders bei Firmen-Chatbots oder KI-generierten Marketingmaterialien ist eine klare Kennzeichnung erforderlich, um Transparenz zu gewährleisten. Das Ziel ist, dass Nutzerinnen und Nutzer jederzeit erkennen können, wenn sie mit einem KI-generierten oder manipulierten Medium interagieren.

Die Digitalpolitik und hier im besonderen Fokus der AI Act stellen also den verpflichtenden Rahmen für das digitalhumanistische Management dar. Aus unternehmerischer Sicht ist es sinnvoll, aus der Pflicht eine Tugend zu machen, also aus der ohnehin notwendigen Compliance das betriebswirtschaftlich-ethische Optimum herauszuholen. Regulierungen sind oft ein Kostenfaktor, doch sie lassen sich in einen ethischen Mehrwert für das Unternehmen und seine Stakeholder verwandeln. Für Unternehmensleitungen, die sich im Dschungel der EU-Regulierungen – insbesondere im Hinblick auf den AI Act – zurechtfinden müssen, bedarf es Orientierung.

22.3 AI Leadership Compass als Übersetzung des AI Acts in den Unternehmensalltag

Genau diese Orientierung und Tugend-aus-einer-Pflicht-machen braucht aber aktuelle und führungsebenenspezifische Vorschläge. Aktuell, weil sich durch die sich ständig ändernden technischen Möglichkeiten der wirtschaftliche Einsatz ständig anpassen muss. Führungsebenenspezifisch deshalb, weil das Top-Management andere Handlungsoptionen durch KI und für den ethischen Einsatz hat als eine Teamleiter*in auf operativer Ebene. Die WU Executive Academy unterscheidet in ihrem einschlägigen Leitfaden (AI Leadership Compass; Giesswein, 2025) vier Führungsebenen. In Anlehnung an die vier Himmelsrichtungen versucht dieser Leitfaden wie ein „Leadership Compass“ den folgenden Führungsebenen Inspiration für den ethischen Einsatz von KI zu geben:

1. Strategische Unternehmensführung (AI for Executives)
2. Operative Teamführung (AI for Leaders)
3. HR Management (AI for HR)
4. Selbstführung für Managers (AI and I)

Strategische Unternehmensführung (AI for Executives)

Die Integration von Künstlicher Intelligenz (KI) in die ethische, strategische Unternehmensführung erfordert eine ausgewogene Herangehensweise, die sowohl Investitionsentscheidungen als auch langfristige Entwicklungen berücksichtigt. Ein rein reaktives Abwarten kann ebenso hinderlich sein wie ein übermäßiger KI-Aktivismus mit unausgereiften Technologien. Das Top-Management sollte gezielt Pilotprojekte der Fachabteilungen gemeinsam mit der IT initiieren, um den praktischen Nutzen und die wirtschaftlichen Implikationen von KI-Technologien auf Basis eigener Erfahrungen bewerten zu können. Auf dieser Basis können Anwendungsbereiche für KI-Agenten gefunden und kritisch evaluiert werden. Da sich die technischen Möglichkeiten kontinuierlich weiterentwickeln, sind solche Testphasen nicht einmalige Maßnahmen, sondern müssen regelmäßig wiederholt und angepasst werden.

Für eine fundierte strategische Entscheidungsfindung ist es essentiell, dass Top-Führungskräfte nicht nur über technologische Grundlagen von KI informiert sind, sondern auch ein tiefgehendes Verständnis der dahinterliegenden Geschäftsmodelle der Anbieter und den ethischen Kontext entwickeln. Da klassische Lernsituationen oft schwer in den Alltag des Top-Managements zu integrieren sind, bieten sich speziell konzipierte Formate wie „Executive AI-Thinking“ an. Diese ermöglichen den Austausch über Erfahrungen, die Diskussion innovativer Ideen und die Entwicklung strategischer Perspektiven im direkten Dialog mit Experten und anderen Führungskräften.

Ein strukturiertes Vorgehen in der Implementierung von KI kann durch eine klare Kategorisierung der KI-Anwendungen im Unternehmen unterstützt werden. Eine sinnvolle Einteilung trennt: (1) Office-Produktivitätstools mit KI zur allgemeinen Effizienzsteigerung im Arbeitsalltag, (2) spezialisierte KI-Systeme, die branchenspezifische Herausfor-

derungen adressieren, wie z. B. in der Medizintechnik oder Finanzbranche und (3) KI-gesteuerte Robotik für Produktion, Wartung und kundennahe Prozesse.

Die Unternehmensleitung trägt auch eine wesentliche Verantwortung, die IT so auszurichten, dass ein nachhaltiger, resilienter Betrieb ermöglicht wird. Im Sinne der KI-Souveränität gilt es, eine möglichst hohe Unabhängigkeit von oligopolistischen KI-Anbietern zu erreichen, eigene KI-Kompetenzen und gegebenenfalls Modelle aufzubauen und die Chancen von Open-Source-Angeboten zu nutzen. Die KI-Aufrüstung der operativen Cybersicherheit ist eine wesentliche Grundlage dieses Ansatzes.

Ein weiteres wesentliches positives Potenzial von KI liegt in der effizienten Zusammenführung und Analyse von Unternehmensdaten. Die Harmonisierung von Daten aus unterschiedlichen Systemen und Abteilungen ist traditionell mit hohen Kosten und Aufwand verbunden. Erste KI-Anwendungen zeigen jedoch, dass eine intelligente Vernetzung und Verarbeitung dieser Daten („Synthese AI“) erheblichen Mehrwert schaffen kann. Unternehmen, die diese Entwicklung aktiv beobachten und testen, könnten sich in den kommenden Jahren einen nachhaltigen Wettbewerbsvorteil sichern. Dabei bleibt die Einhaltung rechtlicher und ethischer Standards ein entscheidender Faktor für eine verantwortungsbewusste Nutzung der Technologie.

Operative Teamführung (AI for Leaders)

Die Integration von Künstlicher Intelligenz (KI) in den Arbeitsalltag stellt Team-Führungskräfte vor die Herausforderung, sowohl Akzeptanz bei den Mitarbeitenden zu schaffen als auch einen verantwortungsvollen Umgang mit der Technologie sicherzustellen. Ein zentraler Aspekt ist der Umgang mit Bedenken seitens der Kollegen. Die Sorge um Arbeitsplatzsicherheit und veränderte Aufgabenprofile erfordert eine proaktive und transparente Kommunikation. Führungskräfte sollten die Rolle von KI als unterstützendes Werkzeug klarstellen und betonen, dass technologische Kompetenz ein wesentlicher Produktivitätsfaktor für die Wissensarbeit der Zukunft ist. Mitarbeitende, die sich frühzeitig mit KI auseinandersetzen, können sich durch Weiterbildungen und Verwendung einen Vorteil verschaffen.

Führungskräfte, die selbst aktiv neue KI-Tools erproben und sich weiterbilden, können diese Veränderung authentisch vorleben. Ein offenes Arbeitsumfeld, in dem Fragen gestellt und neue Methoden ausprobiert werden dürfen, stärkt die Akzeptanz neuer Technologien und fördert die Innovationsfähigkeit im Unternehmen.

Führungskräfte müssen daher mit den intern geltenden KI-Richtlinien vertraut sein, deren Einhaltung sicherstellen und eine verantwortungsbewusste Nutzung fördern. Ein strukturiertes Vorgehen in der Einführung und Anwendung von KI gewährleistet, dass die Technologie gezielt zur Effizienzsteigerung beiträgt, ohne die Integrität der Unternehmensprozesse zu gefährden.

Neben der direkten Führungsarbeit lohnt es sich für operative Managers auch, den KI-Kosteneinsparungseffekte im täglichen Arbeiten (z. B. PR und Marketing) zu prüfen. Gerade Large Language Modells (LLMs) arbeiten inzwischen sehr gut mit jeder Form der Sprache. Ob Social Media Posts, Blog-Beiträge oder Newsletter – das Einbeziehen von

LLMs in den Arbeitsalltag kann Zeit sparen bzw. die Ergebnisse und den Lesefluss verbessern.

Mit neuen KI-Tools lassen sich PowerPoint-Präsentationen für z. B. Kick-Off Termine in Sekundenschnelle erstellen. Bildgenerierende KI-Systeme können teure bzw. aufwändige (Stock-)Fotos ablösen bzw. ergänzen. Und immer mehr KI-Tools schaffen mehrere Sekunden lange Bewegungsbilder mit inzwischen täuschend echten Wasseranimationen. Mit dem richtigen Prompt entsteht so ein qualitatives Video, welches früher mehrere Drehtage und viel Budget verbraucht hätte. Der Kreativität sind dabei kaum Grenzen gesetzt. Dabei sollten stets geschulte Mitarbeiter die KI-Anwendungen bedienen, die insbesondere auf Datenschutz und Urheberrecht achten.

HR Management (AI for HR)

Die kontinuierliche Weiterbildung der Mitarbeitenden im Bereich Künstliche Intelligenz (KI) ist essenziell, um sowohl den regulatorischen Anforderungen als auch den strategischen Unternehmenszielen gerecht zu werden. Regelmäßige Schulungen stellen sicher, dass Mitarbeitende nicht nur mit den technologischen Entwicklungen vertraut sind, sondern auch die ethischen und funktionalen Implikationen der KI-Nutzung verstehen. In vielen Unternehmen nimmt KI-Kompetenz eine zunehmend zentrale Rolle für die Weiterentwicklung und Wettbewerbsfähigkeit ein.

Eine weitere wichtige HR-Aufgabe ist die Definition klarer KI-Richtlinien gemeinsam mit dem Top-Management und den Mitarbeitenden. Eine fundierte Governance-Struktur ist erforderlich, um gesetzliche Vorschriften einzuhalten und einen verantwortungsvollen Umgang mit KI zu gewährleisten. In vielen Organisationen nutzen Mitarbeitende bereits eigenständig KI-Tools im Arbeitsalltag („Bring Your Own AI“, BYOAI), was die Gefahren von Datenschutzverletzungen und eines unkontrollierten Abfluss sensibler Unternehmensdaten mit sich bringt.

Die Entwicklung interner, partizipativ gestalteter Richtlinien kann helfen, diese Herausforderungen zu adressieren und eine sichere sowie produktive Integration von KI-Systemen in betriebliche Abläufe zu ermöglichen.

Auch im Personalwesen selbst ergeben sich durch KI-Anwendungen vielfältige Optimierungsmöglichkeiten. Ob im Recruiting, im Payroll-Management oder eben in der Entwicklung von Lernformaten – KI kann Prozesse beschleunigen und vergünstigen, die Qualität der Personalentscheidungen verbessern und hoch-personalisierte Weiterbildungspfade generieren. Durch den gezielten Einsatz von KI in der HR lassen sich somit sowohl die betriebliche Effizienz als auch die Positionierung als attraktiver Arbeitgeber nachhaltig stärken.

Selbstführung für Managers (AI and I)

Die KI bietet auch Chancen für die Arbeit und persönliche Weiterentwicklung der Führungskraft selbst.

Insbesondere bei der Erstellung von Konzepten, Präsentationen und Redebeiträgen kann KI den inhaltlichen Einstieg erleichtern, strukturierte Vorschläge generieren und

kreative Impulse liefern. Darüber hinaus unterstützt sie bei der Textproduktion und Übersetzungen, etwa durch die Verarbeitung von Diktaten oder die Formulierung anspruchsvoller Kommunikation. Auch die Dokumentation von Besprechungen lässt sich durch KI deutlich beschleunigen, Protokolle und Aufgabenverteilungen stehen unmittelbar nach einem Meeting für die Führungskraft zur Verfügung.

Neben der operativen Unterstützung kann KI auch als ethischen Reflexions- und Coaching-Tool dienen. Durch interaktive Dialoge lassen sich Argumentationsstrategien testen, Lösungsansätze für Konfliktsituationen entwickeln und mögliche Auswirkungen von Maßnahmen auf die Mitarbeitenden diskutieren.

Ein fundiertes Verständnis der Funktionsweise und Grenzen von KI-gestützten Tools ist eine essenzielle Voraussetzung für ihre effektive Integration in betriebliche Abläufe. Nur wer die Technologie selbst ausprobiert, kann realistisch einschätzen, welche Mehrwerte sie bietet und in welchen Bereichen menschliche Expertise unverzichtbar bleibt. Diese praktische Erfahrung bildet zudem die Grundlage für einen authentischen und glaubwürdigen Dialog mit Mitarbeitenden über den Einsatz von KI im Unternehmen.

22.4 Conclusio aus dem AI Leadership Compass

Die Effizienzgewinne durch KI-Nutzung können in signifikanten Zeitersparnissen resultieren. Die Frage nach der Nutzung dieser freien Kapazitäten ist daher von zentraler Bedeutung. Eine menschenzentrierte Herangehensweise könnte darin bestehen, die gewonnene Zeit für intensivere Mitarbeitergespräche oder die gezielte Förderung individueller Potenziale zu verwenden.

Über alle Führungsrichtungen hinweg gesehen scheint der KI-Einsatz für Führungskräfte einen Schub für mehr Individualisierung der Führung, Leistungssteigerung des gesamten Unternehmens, neue Karrierewege und auch mehr Menschlichkeit im realen Umgang mit Kollegen und Kolleginnen zu bringen. Ob dies Realität wird, entscheidet nicht die Technologie, sondern die Manager in den Unternehmen, die im digitalhumanistischen Sinne agieren und die Organisation gestalten.

22.5 Fazit: Management und die europäischen Gesetze

Mit den digitalhumanistischen Vorgaben der EU ist die generelle Richtung bereits vorgegeben. Digitaler Humanismus ist der gesellschaftliche Nordstern, an dem sich auch die Unternehmen zu halten haben. Statt nur ein Mindestmaß an Compliance herzustellen und negative rechtliche und öffentliche Effekte zu riskieren, sollte das Management aus der Pflicht eine Tugend machen und die positiven, digitalhumanistischen Effekte der KI-Welt für die eigene Organisation nutzen. Damit können nicht nur mögliche Haftungen vermieden, sondern auch nachhaltige Wettbewerbsvorteile und ein besseres Leadership erzielt werden.

Literatur

- Aiken, M. (2018). *Der Cyber-Effekt: Wie das Internet unser Denken, Fühlen und Handeln verändert*. S. Fischer.
- Europäische Kommission (2022). *Europäische Erklärung über digitale Rechte und Prinzipien*. Brüssel: Europäische Kommission. <https://digital-strategy.ec.europa.eu/en/library/european-declaration-digital-rights-and-principles> (zuletzt abgerufen am 7. Mai 2025)
- Europäische Kommission (2025). *Europäische Kommission veröffentlicht Leitlinien zu verbotenen Praktiken im Bereich der künstlichen Intelligenz gemäß der AI-Verordnung*. Brüssel: Europäische Kommission. <https://digital-strategy.ec.europa.eu/en/library/commission-publishes-guidelines-prohibited-artificial-intelligence-ai-practices-defined-ai-act> (zuletzt abgerufen am 5. Juni 2025)
- Europäische Union Charta der Grundrechte der Europäischen Union. Amtsblatt der Europäischen Union, C 326/391. <https://eur-lex.europa.eu/DE/legal-content/summary/charter-of-fundamental-rights-of-the-european-union.html> (Erstellt: 26. Okt. 2012). Zugriffen: 5. Mai 2025.
- Europäische Union: Verordnung (EU) (2022a) 2022/1925 des Europäischen Parlaments und des Rates vom 14. September 2022 über bestreithbare und faire Märkte im digitalen Sektor und zur Änderung der Richtlinien (EU) 2019/1937 und (EU) 2020/1828 (Digital Markets Act). Amtsblatt der Europäischen Union, L 265/1, 21. September 2022. Online verfügbar unter: <https://eur-lex.europa.eu/eli/reg/2022/1925/oj/eng> (zuletzt abgerufen am 5. Mai 2025).
- Europäische Union: Verordnung (EU) (2022b) 2022/2065 des Europäischen Parlaments und des Rates vom 19. Oktober 2022 über einen Binnenmarkt für digitale Dienste und zur Änderung der Richtlinie 2000/31/EG (Digital Services Act). Amtsblatt der Europäischen Union, L 277/1, 27. Oktober 2022. Online verfügbar unter: <https://eur-lex.europa.eu/eli/reg/2022/2065/oj/eng> (zuletzt abgerufen am 5. Mai 2025).
- Europäische Union: Verordnung (EU) 2016/679 des Europäischen Parlaments und des Rates vom 27. April 2016 zum Schutz natürlicher Personen bei der Verarbeitung personenbezogener Daten und zum freien Verkehr solcher Daten sowie zur Aufhebung der Richtlinie 95/46/EG (Datenschutz-Grundverordnung). Amtsblatt der Europäischen Union, L 119/1, 4. Mai 2016. Online verfügbar unter: <https://eur-lex.europa.eu/eli/reg/2016/679/oj/de> (zuletzt abgerufen am 5. Mai 2025).
- Europäische Union: Verordnung (EU) 2023/2854 des Europäischen Parlaments und des Rates vom 13. Dezember 2023 über harmonisierte Vorschriften für einen fairen Zugang zu und die Nutzung von Daten und zur Änderung der Verordnung (EU) 2017/2394 und der Richtlinie (EU) 2020/1828 (Data Act). Amtsblatt der Europäischen Union, L 322/1, 22. Dezember 2023. Online verfügbar unter: <https://eur-lex.europa.eu/eli/reg/2023/2854/oj/de> (zuletzt abgerufen am 5. Mai 2025).
- Europäische Union: Verordnung (EU) 2024/1689 des Europäischen Parlaments und des Rates vom 13. Juni 2024 über harmonisierte Vorschriften für künstliche Intelligenz und zur Änderung bestimmter Rechtsakte der Union (Künstliche Intelligenz-Verordnung – AI Act). Amtsblatt der Europäischen Union, L 180/1, 12. Juli 2024. Online verfügbar unter: <https://eur-lex.europa.eu/eli/reg/2024/1689/oj/de> (zuletzt abgerufen am 5. Juni 2025).
- Europäischer Gerichtshof Urteil vom 20. Februar 2018 in der Rechtssache C-16/16 P – Königreich Belgien gegen Europäische Kommission. Verfahren zur Hauptsache. <https://curia.europa.eu/juris/liste.jsf?language=de&num=C-16/16%20P>. Zugriffen: 5. Juni 2025.
- Giesswein, M. (2025). Kompass. <https://www.martingiesswein.com/kompass>. Zugriffen: 5. Juni 2025.
- Intergovernmental Panel on Climate Change (IPCC) Offizielle Webseite. <https://www.ipcc.ch>. Zugriffen: 5. Mai 2025.

The Future Society Final Report on Global Consultations for France's 2025 AI Action Summit. <https://thefuturesociety.org/aiactions summitconsultationreport> (Erstellt: 8. Dez. 2024). Zugriffen: 5. Mai 2025.

World Health Organization Regional Office for Europe (2024). *A focus on adolescent social media use and gaming in Europe, central Asia and Canada: Health Behaviour in School-aged Children international report from the 2021/2022 survey*. WHO Regional Office for Europe.

Stefan Tauber ist Rechtsanwalt und Vergabeexperte bei Schiefer Rechtsanwälte. Er absolvierte sein Studium der Rechtswissenschaften an der Universität Wien und sammelte bereits währenddessen wertvolle Praxiserfahrungen als juristischer Mitarbeiter in verschiedenen Anwaltskanzleien. Nach seiner Gerichtspraxis begann er 2018 seine Ausbildung zum Rechtsanwalt unter Martin Schiefer und spezialisiert sich seither auf das Vergaberecht. Seine Schwerpunkte liegen insbesondere auf Digitalisierungsprojekten mit Fokus auf neue Technologien sowie der EU-Regulatorik der digitalen Dekade. Neben der Beratung der öffentlichen Hand im Vergabe- und IT-Recht ist er ein gefragter Vortragender auf Fachseminaren und Autor von Fachpublikationen.

Martin Giesswein ist Fakultätsmitglied der WU Executive Academy mit den Schwerpunkten KI, Digitalökonomie und Leadership. Er ist Autor und Herausgeber mehrerer Bücher zu den Themen Künstliche Intelligenz, Digital Leadership und Zukunftsgestaltung. 2025 wurde er mit einem ersten Preis des Awards „Digitaler Humanismus in der Praxis“ ausgezeichnet. Er ist als AI Ethics Advisor Mitglied des Advisory Committee im Projekt DRAIGON. Martin Giesswein ist einer der Mitinitiatoren der DigitalCity.Wien, einer seit über 10 Jahren bestehenden Kooperation der Digitalwirtschaft mit der Stadt Wien. Zuvor war er CEO und General Manager in der IKT-Branche.

Verankerung des digitalen Humanismus in der öffentlichen Beschaffung

Martin Schiefer und Maria Troger

Zusammenfassung

Der Beitrag beschäftigt sich mit der Verankerung des digitalen Humanismus in der öffentlichen Beschaffung als Hebel für eine verantwortungsvolle Nutzung von Technologie, die den Menschen in den Mittelpunkt stellt. Öffentliche Auftraggeber können ihre Vorbildwirkung bei öffentlichen Ausschreibungen nutzen, um europäische Werte und Rechtsgrundlagen hochzuhalten und ein ethisches und humanes Miteinander zu fördern. Mit Vergaberecht kann der digitale Humanismus in die Gesellschaft gebracht werden und den Spagat zwischen Laissez-faire und Überregulierung durch innovative und gleichzeitig digital ethische Beschaffungen auflösen.

Digitaler Humanismus ist der Hebel für die verantwortungsvolle Nutzung von Technologie, die den Menschen in den Mittelpunkt stellt.

23.1 Einleitung

Durch Digitalisierung und den Einsatz von künstlicher Intelligenz (KI) werden neue Möglichkeiten in der Gesellschaft eröffnet. Gleichzeitig wächst die Unsicherheit bei der Nutzung innovativer Technologien und dem Einsatz von KI. Unsere Gesellschaft steht vor der Herausforderung eines Wandels, der alle Lebensbereiche betrifft. Ob in der Verwaltung,

M. Schiefer (✉) · M. Troger
Schiefer Rechtsanwälte GmbH, Wien, Österreich
E-Mail: office@schiefer.at

im Gesundheitsbereich oder in der Industrie – digitale Technologien prägen unsere Prozesse, Entscheidungen und Interaktionen. Dieser Wandel bietet die Chance, Prozesse zu verbessern, Ressourcen zu schonen und gleichzeitig gesellschaftliche Herausforderungen zu meistern – vorausgesetzt, der Wandel wird verantwortungsvoll und menschenzentriert gestaltet und es wird auch entsprechend gehandelt.

Die Legitimität technologischer Innovationen wird dabei auch zunehmend durch ethische Debatten geprägt. Viele Menschen sorgen sich, dass ihre Arbeitsplätze durch KI ersetzt werden, anstatt sie zu ergänzen. Andere befürchten einen Verlust von Kontrolle, Transparenz und Privatsphäre, wenn Entscheidungen durch Algorithmen getroffen werden. Ein „Datenbias“, die Gefahr von Cyberangriffen oder der Missbrauch von KI zu Kontroll-/Überwachungszwecken tragen zusätzlich zur Skepsis bei. Besonders kritisch wird die Abhängigkeit von Technologien, die mit einer Dehumanisierung des Arbeitsplatzes einhergeht, gesehen.

Trotz dieser Sorgen ist klar, dass Digitalisierung kein Trend ist, sondern eine Notwendigkeit, um in einer zunehmend vernetzten Welt wettbewerbsfähig zu bleiben. Dennoch sollte der Fokus nicht darauf liegen technische Innovation um jeden Preis zu ermöglichen, sondern viel mehr darauf das Leben der Menschen mittels Digitalisierung zu erleichtern. Technologische Innovationen müssen daher so gestaltet werden, dass sie nicht nur effizienter und billiger als herkömmliche Anwendungen sind, sondern auch gesellschaftliche Werte fördern.

Um die mit der Digitalisierung einhergehende „Negativdigitalisierung“ zu bezwingen, muss digitalen Geschäftsmodellen und technischen Innovationen jedenfalls eine positive und ethisch angeleitete Wertorientierung mitgegeben werden, sodass Digitalisierung in Zukunft einen menschengerechten, wertvollen Fortschritt gewährleistet (Spiekermann, 2019). Digitaler Humanismus ist der Hebel für die verantwortungsvolle Nutzung von Technologie, die den Menschen in den Mittelpunkt stellt. Die Prinzipien des digitalen Humanismus betreffen dabei jedes Unternehmen – denn jedes Unternehmen bzw. jede Organisation hat eine gewisse, digitale Wirkung (Giesswein et al., 2022).

Durch die Verankerung des digitalen Humanismus in der öffentlichen Beschaffung können öffentliche Auftraggeber ihre Vorbildwirkung nutzen, um europäische Werte hochzuhalten und ein ethisches und humanes Miteinander zu fördern. Mit Vergaberecht kann der digitale Humanismus in die Gesellschaft gebracht werden und den Spagat zwischen Laissez-faire und Überregulierung durch innovative und gleichzeitig digital ethische Beschaffungen auflösen.

23.2 Vergaberecht als Teil der Lösung

Das Vergaberecht stellt ein wichtiges Steuerungselement in der Gesellschaft dar. Öffentliche Ausschreibungen sind geeignet, neue Maßstäbe in der Gesellschaft zu setzen, indem Unternehmen für ihren Einsatz und die Verankerung und Umsetzung des digitalen Humanismus belohnt werden. Das Vergaberecht kann somit als Belohnungs-Tool den digitalen

Humanismus fördern und in die Gesellschaft bringen. Erfolgen Vergaben an vorbildlich agierende Unternehmen, führt dies zu einer nachhaltigen Stärkung von europäischen Werten. So können sowohl soziale Gerechtigkeit als auch technologische Innovation gefördert werden. Mit der Integration des digitalen Humanismus in Vergabeverfahren kann somit sichergestellt werden, dass technologische Innovationen umgesetzt werden, die im Einklang mit den grundlegenden Werten und Bedürfnissen der Menschen stehen. Vergaberecht wird somit Teil der Lösung.

Das Ziel des Digitalen Humanismus ist Technologie transparent, fair und nachhaltig umzusetzen, sodass auch die ethischen Anforderungen an digitale Systeme eingehalten werden. Dabei kann gerade das Vergaberecht eine große Wirkung auf die am Markt agierenden Unternehmen – und somit wiederum auf die Gesellschaft als Nutzer von digitalen Systemen und innovativen Technologien – haben. Gemäß § 20 Abs. 1 BVergG sind Vergabeverfahren unter Beachtung der unionsrechtlichen Grundsätze wie insbesondere der Gleichbehandlung aller Bewerber und Bieter, der Nichtdiskriminierung, der Verhältnismäßigkeit, der Transparenz sowie des freien und lauten Wettbewerbes und unter Wahrung des Grundsatzes der Wirtschaftlichkeit durchzuführen. Die Vergabe hat an befugte, leistungsfähige und zuverlässige (geeignete) Unternehmer zu angemessenen Preisen zu erfolgen. Werden diese Grundsätze des Vergabeverfahrens um die digital-humanistischen Ziele ergänzt, können Digitalisierung und KI als transformative Kräfte ihr volles Potenzial entfalten – in einem fairen und transparenten Wettbewerb unter Berücksichtigung einer durchdachten Regulierung wird somit der verantwortungsvolle Einsatz von Technologie nachhaltig in einen gesellschaftlichen Mehrwert überführt.

23.3 Digitaler Humanismus im Vergabeverfahren

Ein Vergabeverfahren wird – unabhängig von der Wahl eines einstufigen- oder zweistufigen Vergabeverfahrens – stets in mehreren Phasen abgewickelt. In jeder dieser Phase ergibt sich ein unterschiedlicher „Scope“, also eine unterschiedliche Betrachtungsebene, in der Entscheidungen zu treffen sind, die das gesamte Vergabeverfahren (und somit schließlich die zu beschaffende Leistung bzw. das zu beschaffende Produkt) betreffen. Jeder Scope kann dazu genutzt werden, digitalen Humanismus im Vergabeverfahren zu verankern, sodass die digital-humanistischen Grundsätze bei der Leistungserbringung entsprechend berücksichtigt werden. Nachstehend wird dargestellt, wie die Ziele und Grundsätze des digitalen Humanismus in den unterschiedlichen Ebenen eines Vergabeverfahrens umgesetzt werden können.

Scope 0: Bedarfsanalyse und Zielsetzungen

Die Bedarfsanalyse bildet den Ausgangspunkt für jedes Projekt. Hier werden alle grundlegenden Aspekte und Anforderungen sorgfältig geprüft und definiert, um sicherzustellen, dass das Projekt eine klare Richtung hat. Wird das Ziel einer digital-humanistischen Beschaffung und Projektabwicklung bzw. Leistungserbringung definiert, sind die zum Zeit-

punkt der Einleitung des Projektes gültigen und wesentlichen (insbesondere europäischen) Rechtsakte zu identifizieren und diese bei der Projektkonzeption zugrunde zu legen.

Neben den Möglichkeiten, die das Bundesvergabegesetz zur Berücksichtigung des Digitalen Humanismus bietet, werden derzeit laufend neue Rechtsakte der EU erlassen, die einen Schutz ethischer Prinzipien in der Privatwirtschaft sicherstellen sollen. Diese Regularien ist wesentliche Voraussetzung für ein sicheres, ethisches und humanes Miteinander bei der Implementierung technischer Innovationen. Die europäischen Rechtsgrundlagen wie die DSGVO (EU Verordnung, 2016), NIS 2 (EU Richtlinie, 2022), AI Act (EU-Verordnung, 2024), DORA (EU-Verordnung, 2022), etc., sind Chancen für die Beibehaltung einer funktionierenden Demokratie. Anders als in anderen Wirtschaftsräumen (USA, China) sind gerade diese Regelungen wesentliche Voraussetzungen für ein sicheres, ethisches und humanes Miteinander bei der Verwendung von Technologie. Digital-humanistische Ansätze betreffen dabei sowohl die Beschaffung von Hardware und Software sowie den Umgang mit Daten im Vergabeverfahren und bei der Auftragsabwicklung.

Die Definition von konkreten Zielen für ein Projekt kann die Umwelt, die Arbeitsbedingungen, soziale Verantwortung, Menschenrechte sowie Korruptionsbekämpfung umfassen. Als Leitziele und Hilfestellungen können z. B. die 17 Sustainable Development Goals (Vereinte Nationen, 2025) herangezogen werden. Diese Leitziele sind als allgemeine Wegweiser einer nachhaltigen und humanen Zukunft zu verstehen und können als Inspiration für die Definition konkreter Projektziele dienen. Beschaffungsmethoden, die auf soziale Bedürfnisse ausgelegt sind, wie z. B. Förderung menschenwürdiger Arbeit, Zugänglichkeit und Design für alle, Förderung fairer Beschäftigungsmöglichkeiten und sozialer Inklusion oder Achtung der Menschenrechte und Behandlung von Fragen in Zusammenhang mit ethischem Handel werden auch zum „Socially Responsible Public Procurement“ (SRPP) gezählt (Kranabetter, 2024). Genau diese Beschaffungsmethoden sollen auch mit der Umsetzung des digitalen Humanismus bei der Beschaffung von Technologie (wie insbesondere Hardware und Software) weiter gestärkt werden.

Neben europäischen Vorgaben ist die Berücksichtigung von ethischen und moralischen Grundsätzen auch bereits im Bundesvergabegesetz 2018 (BVergG 2018) verankert. So ist in § 20 Abs. 6 BVergG zur Integration sozialer Kriterien im Vergabeverfahren festgelegt, dass auf die Beschäftigung von Frauen, von Personen im Ausbildungsverhältnis, von Langzeitarbeitslosen, von Menschen mit Behinderung und älteren Arbeitnehmern sowie auf Maßnahmen zur Umsetzung sonstiger sozialpolitischer Belange Bedacht genommen werden. Dies kann insbesondere durch die Berücksichtigung derartiger Aspekte bei der Beschreibung der Leistung, bei der Festlegung der technischen Spezifikationen, durch die Festlegung konkreter Zuschlagskriterien oder durch die Festlegung von Bedingungen im Leistungsvertrag erfolgen. Es handelt sich bei § 20 Abs. 6 BVergG zwar bloß um eine „Kann-Bestimmung“, bei zukunftsorientierten und nachhaltigen Beschaffungsvorgängen wird jedoch die Berücksichtigung sozialer Kriterien (inklusive den digital-humanistischen Grundsätzen) als Zielsetzung im Vergabeverfahren – unabhängig vom konkreten Scope – in Zukunft unabdingbar sein.

Scope 1: Digitaler Humanismus auf Unternehmensebene

Unternehmen müssen in Zukunft echte Partner sein, um nachhaltig gemeinsame Ziele umzusetzen. Der bloße Leistungsaustausch gegen Entgelt wird daher in Zukunft durch Partnerschaften ersetzt werden. Dazu müssen einerseits ethische und ökologische Werte gegenseitig mitgetragen werden, andererseits sind kontinuierliche Anstrengungen und gemeinsame Entwicklungen von Innovation erforderlich, um die Digitalisierung der Zukunft ethisch, moralisch und ökologisch zu gestalten. Um Transparenz sicherzustellen und Verantwortung zu schaffen sind dabei auch alle Stakeholder eines Projektes zu involvieren.

Echte Partnerschaften können zwischen der öffentlichen Hand und privaten Unternehmen insbesondere bei der Umsetzung menschenzentrierter, digitaler Technologie umgesetzt werden. Gemeinsame ethische und moralische Werte können bei Implementierung im Vergabeverfahren bei der Leistungserbringung partnerschaftlich umgesetzt und gelebt werden.

Konkret kann die Umsetzung des digitalen Humanismus im Unternehmen im Vergabeverfahren von öffentlichen Auftraggebern belohnt werden. Einerseits kann diese „Belohnung“ durch die Zulassung zum Vergabeverfahren durch Erfüllung von (unternehmensbezogenen) Eignungskriterien erfolgen, andererseits durch die Bewertung von (angebotsbezogenen) Zuschlagskriterien. Eignungs- und Zuschlagskriterien sind in jedem Vergabeverfahren zu definieren, unabhängig davon, ob es sich um ein einstufiges Vergabeverfahren oder ein 2-stufiges Vergabeverfahren handelt. Zur Besonderheit der Direktvergabe vgl. §§ 46, 47 BVergG (im Sektorenbereich vgl. §§ 213, 214 BVergG).

Auf der Unternehmensebene werden im Scope 1 Eignungskriterien für die Teilnahme am Vergabeverfahren definiert. Nur ein Unternehmen, das i. S. d. BVergG auch geeignet ist, kann in einem Vergabeverfahren als Zuschlagsempfänger ermittelt werden.

Eignungskriterien sind unternehmensbezogene Kriterien. Sie sind gemäß § 2 Z 22 lit. c) BVergG die vom Auftraggeber festgelegten, nichtdiskriminierenden, mit dem Auftragsgegenstand in Verbindung stehenden und zu diesem im Verhältnis stehenden Mindestanforderungen betreffend die Befugnis, Zuverlässigkeit und Leistungsfähigkeit (Eignung) an den Bewerber oder Bieter, die gemäß den Bestimmungen dieses Bundesgesetzes nachzuweisen sind. Bei der Leistungsfähigkeit ist zwischen der finanziellen und wirtschaftlichen sowie der technischen Leistungsfähigkeit zu unterscheiden. Die Eignungskriterien sind in den Teilnahme- bzw. Ausschreibungsunterlagen festzulegen.

Digital-humanistische Anforderungen können z. B. im Rahmen der technischen Leistungsfähigkeit abgefragt werden. So können Unternehmen, die selbst ethische und moralische Verantwortung übernehmen, soziale Gerechtigkeit und digitale Inklusion (Barrierefreiheit) fördern sowie die Zielsetzungen einer wertegeleiteten Technologieentwicklung verfolgen, zum Vergabeverfahren zugelassen werden. Auch unternehmensinterne Verhaltensregeln im Umgang mit KI und dem Einsatz von KI sowie die konkrete Förderung des digitalen Humanismus oder die Einhaltung eines unternehmensinternen Ethik-Kodex (faire Arbeitsbedingungen, Ablehnung von Kinder- und Zwangsarbeit, soziales Engagement) können mitberücksichtigt werden. So kann die digitalhumanistische Unternehmenskultur aktiv im Vergabeprozess belohnt werden.

Bei der Festlegung der konkreten Eignungskriterien kann auch auf das Vorliegen bzw. die Abfrage möglicher Zertifikate abgestellt werden (z. B. für die rechtssichere Verarbeitung von Daten im Unternehmen, z. B. ISO 27001-Zertifizierung). Ob eine ausreichende Anzahl an Unternehmen die entsprechenden Zertifikate bereits besitzt und der Wettbewerb bei Festlegung konkreter Zertifizierungen als Eignungskriterium dennoch hochgehalten werden kann, kann vor Einleitung eines Vergabeverfahrens auch im Zuge einer Markterkundung eruiert werden (§§ 24, 197 BVergG).

Scope 2: Digitaler Humanismus als Belohnung in der Angebotsbewertung

Sollen die Prinzipien und Ziele des digitalen Humanismus und die entsprechende Umsetzung im Auftragsfall in einem Vergabeverfahren bewertet und somit in Punkten „belohnt“ werden, können entsprechende Zuschlagskriterien festgelegt werden. Zuschlagskriterien sind zu Beginn des Vergabeverfahrens bekanntzugeben und im Laufe des Vergabeverfahrens nicht mehr veränderbar.

Zuschlagskriterien sind auftragsbezogene Kriterien. Diese sind gemäß § 2 Z 22 lit. d) sublit. aa) BVergG bei der Wahl des technisch und wirtschaftlich günstigsten Angebotes die niedrigsten Kosten oder die vom Auftraggeber im Verhältnis oder ausnahmsweise in der Reihenfolge ihrer Bedeutung festgelegten, nichtdiskriminierenden und mit dem Auftragsgegenstand in Verbindung stehenden Kriterien, nach welchen das für den Auftraggeber technisch und wirtschaftlich günstigste Angebot ermittelt wird (Bestangebotsprinzip) sowie gemäß § 2 Z 22 lit. d) sublit. bb) BVergG bei der Wahl des Angebots mit dem niedrigsten Preis der Preis (Billigstbieterprinzip).

Neben dem Preis wird im Sinne des Bestangebotsprinzips die Qualität des Angebots im Vergabeverfahren bewertet. Im Rahmen einer „360-Grad Betrachtung“ können dabei zur Sicherstellung der Werte des digitalen Humanismus insbesondere folgende Aspekte positiv bewertet werden:

- Umgang mit ethischen und moralischen Grundsätzen auf der Lieferkette
- Barrierefreie, inklusive und nichtdiskriminierende Technologie
- Datenschutz und Datensicherheit zum Schutz der Privatsphäre
- Sicherheit und Resilienz (Bewertung eingesetzter Sicherheitsmaßnahmen zur Abwehr von Cyberbedrohungen und anderen Risiken, insbesondere zum Schutz kritischer Infrastruktur)
- Ressourcenschonende Technologie (nachhaltiger und sozialer Beitrag bzw. Mehrwert)
- Umgang mit bzw. Beschaffung von Ressourcen (ökologisch und sozial nachhaltige Lieferkette)
- Ethik-Kodex des Unternehmens bzw. Umsetzung eines Ethik-Kodex auch in Zusammenarbeit mit Lieferanten bzw. Subunternehmern (faire Arbeitsbedingungen, Ablehnung von Kinder- und Zwangsarbeit, soziales Engagement auch in der Lieferkette)
- Flexible und innovative Systeme, Umgang mit technologischen Innovationen
- KI-basierte Lösungen, Unterscheidung partielle Ausarbeitung durch KI sowie gänzliche Ausarbeitung durch KI

- Umweltauswirkungen der eingesetzten digitalen Infrastruktur
- Bewertung der im Unternehmen eingesetzten Maßnahmen zur Förderung von Nachhaltigkeit, Energieeffizienz und umweltfreundlichen Praktiken
- Mitgliedschaften zur Demonstration eines verantwortungsvollen Handelns in Bezug auf Menschenrechte, Arbeit, Umwelt und Korruptionsbekämpfung (z. B. UN Global Compact)

Darüber hinaus wird auch in Zukunft bei der Bewertung von Angeboten im Vergabeverfahren der Einsatz von KI zu berücksichtigen sein; geistige Dienstleistungen können teilweise durch den Einsatz von KI-basierten Lösungen sowohl unterstützt als auch gänzlich ersetzt werden. Für den Einsatz von KI müssen daher von öffentlichen Auftraggebern einerseits taugliche Nachweise in der Ausschreibung definiert werden. Andererseits ist der Einsatz von KI bei der Angebotslegung vom Bieter offenzulegen und die genauen Datenquellen sind namhaft zu machen.

Scope 3: Leistungsanforderungen/Leistungsbeschreibung

In der Leistungsbeschreibung sind die „Muss“-Anforderungen an die Leistung selbst zu definieren. Dabei sind neben den gesetzlichen Rahmenbedingungen (siehe Punkt 2.1) auch technische Spezifikationen festzulegen. Durch partnerschaftliche Herangehensweisen und Lösungen wird der Fokus dabei in Zukunft nicht mehr auf „starren“ Leistungsverzeichnissen liegen. Bei der Beschreibung technischer Spezifikationen oder Anforderungen an Hardware und Software wird auch insbesondere die Lieferkette zu berücksichtigen sein.

Darüber hinaus werden neben den „Muss“-Anforderungen aus technischer Sicht in Zukunft auch ethische und moralische Grundsätze – allenfalls in Form eines Ethik-Kodex als Anlage zum Vertrag neben den technischen Leistungsbeschreibungen bzw. -verzeichnissen – zu definieren sein. Im Zusammenhang mit der Betrachtung der Lieferkette werden allenfalls auch die ethischen und moralischen Grundsätze der Subunternehmer und Lieferanten des Auftragnehmers zu hinterfragen bzw. vom Bieter/Auftragnehmer offenzulegen sein („Supplier code of conduct“).

Auch eine ständige (gemeinsame) Überprüfung von Standards und rechtlichen Rahmenbedingungen bei der Auftragsabwicklung sowie eine kontinuierliche Weiterbildung und Weiterentwicklung der Leistungsbeschreibungen wird in Zukunft zwingend erforderlich sein, insbesondere um langfristige, erfolgreiche Partnerschaften zu etablieren. Dabei sind insbesondere für den Fall der technischen Weiterentwicklung entsprechende Änderungs- bzw. Anpassungsklauseln bereits bei der Gestaltung der Leistungsbeschreibung im Zusammenhang mit dem Leistungsvertrag im Vergabeverfahren zu berücksichtigen (vgl. § 365 BVergG). Auf der Ebene der Definition von Leistungsbeschreibung können auch bestimmte Standards für die Leistungserbringung definiert werden, dies insbesondere durch die Festlegung bestimmter Zertifikate (wie z. B. ISO 27001 oder ISO 14001).

Scope 4: Leistungsvertrag

Der „Scope 4“ bildet die abschließende Stufe eines Vergabeverfahrens. In dieser Ebene erfolgt die Zuschlagserteilung und wird darüber hinaus mit dem ausgewählten Bestbieter ein Leistungsvertrag oder eine Rahmenvereinbarung abgeschlossen.

Um echte Partnerschaften zu fördern und die digitalhumanistischen Ziele gemeinsam in der Gesellschaft zu etablieren kommen über die üblichen Leistungsverträge hinaus unterschiedliche Modelle der Zusammenarbeit von öffentlichen Auftraggebern und Unternehmen in Frage. Neben Unternehmenskäufen oder Beteiligungen der öffentlichen Hand an privaten Unternehmen, die innovative Technologien entwickeln, können auch Joint-Ventures oder Innovationspartnerschaften den technischen Fortschritt weiter fördern.

Wesentlich bei der Vertragsgestaltung werden neben der Definition eines eigenen Ethik-Kodex für die Zusammenarbeit des öffentlichen Auftraggebers und Auftragnehmers auch Technologieanpassungsklauseln oder der Umgang mit Qualitäts-, IT-Sicherheit- und Risikomanagement eine immer größere Rolle spielen.

23.4 Fazit

Digitalisierung und technische Innovation bringen große Chancen, stellen die Gesellschaft aber auch vor zahlreiche Herausforderungen. Das Vergaberecht schafft diesen Spagat – zwischen Förderung und Beschaffung neuer Technologien sowie Hochhaltung und Gewährleistung sozialer Standards und europäischer Werte. Öffentliche Auftraggeber können durch ihre Vorbildrolle in der Gesellschaft durch Vergabeverfahren Standards setzen, um Technologie in der Gesellschaft verantwortungsvoll zu nutzen und den Menschen weiterhin in den Mittelpunkt zu stellen. Die unterschiedlichen Scopes eines Vergabeverfahrens schaffen dabei einen weiten und flexiblen Umsetzungsspielraum, sodass Digitalisierung und KI als transformative Kräfte ihr volles Potenzial entfalten können – in einem fairen und transparenten Wettbewerb unter Berücksichtigung einer durchdachten Regulierung.

Literatur

- Bundesgesetz über die Vergabe von Aufträgen (Bundesvergabegesetz 2018 – BVergG 2018), BGBl. I Nr. 65/2018 idF II Nr. 167/2025
- Giesswein, S., Pirkner, M., Rohla, M., & Stöttinger, B. (2022). *The Digital Impact of Your Organisation. Digital Humanism in Action*. Wien. WU Executive Academy.
- Kranabetter, M. (2024). Soziale Kriterien in der öffentlichen Auftragsvergabe. *NR*, 2024, 47 Heft 1 v. 12.4.2024.
- Richtlinie (EU) 2022/2555 des Europäischen Parlaments und des Rates vom 14. Dezember 2022 über Maßnahmen für ein hohes gemeinsames Cybersicherheitsniveau in der Union (NIS-2-Richtlinie).

- Spiekermann, S. (2019). *Digitale Ethik. Ein Wertesystem für das 21. Jahrhundert*. München. Droemer Knaur
- Vereinte Nationen Sustainable Development Goals (SDGs). <https://sdgs.un.org/goals>. Zugriffen: 15. Jan. 2025.
- Verordnung (EU) 2016/679 des Europäischen Parlaments und des Rates vom 27. April 2016 zum Schutz natürlicher Personen bei der Verarbeitung personenbezogener Daten, zum freien Datenverkehr und zur Aufhebung der Richtlinie 95/46/EG (Datenschutz-Grundverordnung, DSGVO).
- Verordnung (EU) 2022/2554 des Europäischen Parlaments und des Rates vom 14. Dezember 2022 über die digitale operationale Resilienz im Finanzsektor (Digital Operational Resilience Act, DORA).
- Verordnung (EU) 2024/1689 des Europäischen Parlaments und des Rates vom 13. Juni 2024 zur Festlegung harmonisierter Vorschriften für künstliche Intelligenz (Verordnung über künstliche Intelligenz, AI Act).

Weiterführende Literatur

- Hattinger, G., & Zimmer, C. (2021). Leitfaden der Europäischen Kommission für die Berücksichtigung sozialer Belange bei der Vergabe öffentlicher Aufträge. *NR*, 2021, 486 Heft 4 v. 30.12.2021.

Martin Schiefer ist Österreichs top Vergaberechtsexperte und Partner der Kanzlei Schiefer Rechtsanwälte mit Hauptsitz in Wien. Mit seinem 50 köpfigen Expert:innen Team hat es sich Martin Schiefer zur Aufgabe gemacht, Vergaberecht neu zu denken: strategisch, innovativ und regional. Zu den Mandant:innen der Kanzlei zählen prominente Auftraggeber sowie Entscheidungsträger:innen aus Politik, Verwaltung und Privatwirtschaft. Schiefer Rechtsanwälte berät und begleitet bei Ausschreibungen mit hohem Reputationsrisiko in den Bereichen Digitalisierung & Innovation, Informations & Kommunikationstechnologie, Gesundheit & Soziales, Bau- & Infrastrukturprojekte, Mobilität, Energieversorgung sowie Compliance. Martin Schiefer findet sich seit Jahren im Spitzenfeld sämtlicher relevanten Ratings und wurde erneut zum Leading Individual in der österreichischen Vergaberechtsszene gewählt.

Maria Troger ist Rechtsanwältin bei Schiefer Rechtsanwälte mit Schwerpunkt auf Vergaberecht. Sie berät und vertritt Auftraggeber und Bieter in allen Phasen von Vergabeverfahren, einschließlich Ausschreibungen, Vertragsverhandlungen und Nachprüfungsverfahren. Ihre Expertise umfasst sowohl nationale als auch europäische Vergaberechtsvorschriften. Darüber hinaus ist sie in der Compliance-Beratung tätig und unterstützt Unternehmen bei der Umsetzung gesetzlicher Vorgaben im Bereich des öffentlichen Auftragswesens.

Digitaler Humanismus jenseits von „Human-in-the-Loop“

24

Partizipative Ansätze zur Förderung menschenzentrierter Technologien

Peter Parycek, Diogo Sasdelli und Noella Edelmann

Zusammenfassung

Als Antwort auf die durch die digitale Transformation herbeigeführten gesellschaftlichen Herausforderungen zielt die Bewegung des Digitalen Humanismus darauf ab, den technologischen Fortschritt mit humanistischen Werten zu kombinieren, um somit eine möglichst menschenzentrierte Entwicklung der digitalen Transformation zu fördern. Viele der in diesem Kontext gestellten Forderungen lassen sich im Zusammenhang mit dem hoch relevanten KI-Themenfeld unter dem Begriff „Human-in-the-Loop“ zusammenfassen, d. h. der Vorstellung, dass Menschen nach wie vor bei (besonders wichtigen oder sensiblen) Entscheidungsfindungsverfahren beteiligt sein müssen. Der vorliegende Beitrag diskutiert die Rolle von Human-in-the-Loop-Ansätzen im Kontext des Digitalen Humanismus. Es wird argumentiert, dass Human-in-the-Loop in vielen Fällen nicht mit einer fortschrittlichen bzw. humanistischen Auffassung zur digitalen Transformation vereinbar ist, sondern eher den Ausdruck eines konservativen bzw. technologiekritischen Standpunkts darstellt. Als ergänzende Alternative zur Verwirklichung der Ziele des digitalen Humanismus wird ein partizipativer, auf anderen Beteiligungsformen (bspw. Co-Creation) basierender Ansatz nahegelegt, der die Teilhabe einschlägiger Stakeholder ermöglicht und somit Menschenzentriertheit erfüllt, ohne in die ausschließlichen Lösungskonzepte der Human-in-the-Loop-Ansätze zu münden.

P. Parycek
Frauenhofer-Institut, Berlin, Deutschland
E-Mail: peter.parycek@donau-uni.ac.at

D. Sasdelli (✉) · N. Edelmann
Danube University Krems, Krems, Österreich
E-Mail: Diogo.Sasdelli@donau-uni.ac.at

N. Edelmann
E-Mail: noella.edelmann@donau-uni.ac.at

Human-in-the-Loop erweist sich als eine zu einfache, geradezu einfallslose Antwort auf eine eigentlich sehr komplexe Frage.

24.1 Einleitung

Das disruptive Potenzial der Digitalisierung, das – wenn überhaupt – nur mit dem der industriellen oder gar der neolithischen Revolution zu vergleichen wäre (Nida-Rümelin & Weidenfeld, 2018), ist nicht nur mit großen Chancen verbunden, sondern birgt in sich ebenso große Risiken. Dieselben technologischen Entwicklungen, die den heutigen beinahe utopischen Zugang zu Wissen und Information ermöglichen und daher als größte Triebfeder aller Zeiten für wirtschaftliche Entwicklung, Demokratie und Aufklärung erscheinen mögen, bedingen zugleich den jüngsten Aufstieg von Verschwörungstheorien, Desinformation und autokratischen Bewegungen. Dieselben Innovationen, die zu Durchbrüchen in den verschiedensten Bereichen, von Medizin, Kommunikation und Industrie bis hin zu Bildung und Kunst führen, können auch für die Herstellung neuer Waffensysteme und Überwachungsmechanismen eingesetzt werden. Die sich aktuell entfaltende, in ihrer wahren Reichweite noch nicht vollkommen zu erfassende digitale Transformation des menschlichen Lebens und der Gesellschaft stellt die Menschheit somit vor beträchtlichen Herausforderungen, zu deren Bewältigung verschiedene technologieethische Paradigmata bzw. digitalisierungspolitische Programme entwickelt werden. Besondere Aufmerksamkeit genießt dabei die Bewegung des *Digitalen Humanismus*, die durch eine Kombination aus technologischem Fortschritt und humanistischen Werten darauf abzielt, eine möglichst *menschenzentrierte* Entwicklung der digitalen Transformation zu fördern.

Unumstritten gilt die Forderung nach Menschenzentriertheit bzw. nach einer menschenzentrierten (technologischen) Gestaltung als wesentlicher Bestandteil des Programms des Digitalen Humanismus (Nida-Rümelin & Weidenfeld, 2018; Geyer, 2023; Werthner, 2020).¹ Den wohl naheliegendsten (wenn auch, wie hier zu argumentieren sein wird, nicht den einzigen) Weg, Menschenzentriertheit zu erfüllen, stellen die durch den Begriff *Human-in-the-Loop* erfassten Ansätze dar, wonach vor allem besonders kritische Entscheidungen trotz prinzipieller Automatisierungsmöglichkeit weiterhin Menschen überlassen werden sollen. In diesem Sinne scheint *Human-in-the-Loop* – als Verwirklichung der Forderung nach Menschenzentriertheit – im Programm des Digitalen Humanismus enthalten zu sein. Dies lässt sich auch durch konkrete Beispiele belegen: Das im Mai 2019 erschienene *Wiener Manifest für Digitalen Humanismus* scheint in

¹ Menschenzentriertheit spielt über den Rahmen des Digitalen Humanismus hinaus auch bei – aus inhaltlicher Sicht doch recht nahen – weiteren theoretischen Paradigmata für Technologieethik, etwa bei Wertebasierten Ansätzen. Vgl. hierfür exemplarisch Spiekermann (2019).

ihrer fünften „Kernforderung“ einem *Human-in-the-Loop*-Ansatz ausdrücklich zu folgen:

„Entscheidungen, deren Folgen die individuellen oder kollektiven Menschenrechte betreffen können, müssen weiterhin vom Menschen getroffen werden. Die Entscheidungsträger*innen müssen für ihre Entscheidungen verantwortlich und haftbar sein. Automatisierte Entscheidungssysteme sollen die Entscheidungsfindung durch den Menschen nur unterstützen und nicht ersetzen.“ (Wiener Manifest für Digitalen Humanismus, 2019)

Doch gerade in ihren strengeren Auslegungen – was im Manifest durch die durchaus vage Beschränkung auf „Entscheidungen, deren Folgen Menschenrechte betreffen *können*“ prinzipiell nicht vermieden wird – kann *Human-in-the-Loop* in stark ausschließliche oder gar technologiefeindliche bzw. innovationsverhindernde Konzepte münden. Dies widerspräche allerdings dem Kerngedanken des Digitalen Humanismus, der durch die Verbindung technologischen Fortschritts mit humanistischen Werten danach strebt, einerseits menschen-, andererseits aber ebenso technologiefreundlich zu sein (Nida-Rümelin & Weidenfeld, 2018; Werthner, 2020). Der vorliegende Beitrag erörtert die Möglichkeit, Menschenzentriertheit durch alternative Ansätze zu erfüllen und somit den Digitalen Humanismus auch jenseits von *Human-in-the-Loop* zu begründen. Hierfür werden zunächst im Abschn. 24.2 einige der aktuell wichtigsten *Human-in-the-Loop*-Ansätze in Ethik und Recht dargestellt. Im Abschn. 24.3 werden die Nachteile dieser Ansätze diskutiert. Im Abschn. 24.4 wird sodann eine alternative Perspektive zu Menschenzentriertheit nahegelegt, die zu Ansätzen wie *Co-Creation* führt, welche das Prinzip der Menschenzentriertheit erfüllen können, ohne in dieselben ausschließlichen Konzepte von *Human-in-the-Loop* zu münden. Abschn. 24.5 enthält die Schlussbemerkungen.

24.2 Human-in-the-Loop und verwandte Ansätze in Ethik und Recht

Neben der bereits oben angeführten Passage des *Wiener Manifests* für Digitalen Humanismus lassen sich *Human-in-the-Loop*-Ansätze auch bei weiteren ethischen und rechtlichen Regulierungsrahmen für neue Technologien (dabei insbesondere für Künstliche Intelligenz) wiedererkennen. Manchmal wird dabei auch zwischen Varianten desselben Grundgedankens unterschieden: Die von der Unabhängigen Hochrangigen Expertengruppe für Künstliche Intelligenz verfassten, im Jahre 2018 veröffentlichten *Ethik-Leitlinien für eine Vertrauenswürdige KI* führen bspw. die größtenteils im sog. *Grundsatz der Achtung der menschlichen Autonomie* gegründete Anforderung des *Vorrangs menschlichen Handelns und menschlicher Aufsicht* an. Erreicht könne diese menschliche Aufsicht durch verschiedene Lenkungs- und Kontrollmechanismen. Davon nennt die Expertengruppe explizit drei (Unabhängige Hochrangige Expertengruppe für Künstliche Intelligenz, 2018):

1. *Human-in-the-Loop*: Der Mensch kann in jedem Entscheidungszyklus eingreifen.
2. *Human-on-the-Loop*: Der Mensch kann im Entwurfszyklus eingreifen und den Systembetrieb überwachen.

3. *Human-in-Command*: Der Mensch kann den Gesamtbetrieb des Systems (inklusive hinsichtlich weiterer wirtschaftlicher, gesellschaftlicher, rechtlicher und ethischer Auswirkungen) beaufsichtigen und entscheiden, wann und wie das System in einer bestimmten Situation eingesetzt werden soll.

Bei näherer Betrachtung erweist sich diese Unterscheidung (zumindest in der von der Expertengruppe angeführten Form) als etwas konfus und somit in der Praxis wenig brauchbar: Der Unterschied zwischen *Human-in-the-Loop* (im von der Expertengruppe verwendeten Sinne) und *Human-in-Command* ist bspw. unklar, da die Möglichkeit des Eingriffs in jedem Entscheidungszyklus die Aufsicht über den gesamten Betrieb voraussetzt und im Effekt mit der Entscheidung gleichfällt, wann und wie das System einzusetzen ist. *Human-on-the-Loop* scheint hingegen die Möglichkeit des Eingriffs nur auf die Entwurfsphase zu beschränken, sieht dabei die Aufsicht über den Systembetrieb vor. Indessen erscheint die menschliche Beteiligung beim Entwurf eines Systems relativ trivial – zumindest, wenn darunter auch die Konzeptionsphase mitgemeint wird –, eine Betriebsüberwachung ohne Eingriffsmöglichkeit erweist sich wiederum als wenig sinnvoll. Insofern, als diese drei (und etwaige weitere) Varianten damit einhergehen, dass (ggf. besonders wichtige) Entscheidungen weiterhin Menschen überlassen werden, fallen sie unter die hier angenommene, etwas breitere Definition von *Human-in-the-Loop*-Ansätzen.²

Der von der Expertengruppe geförderte *Human-in-the-Loop*-Ansatz dient primär dem Zweck, zu verhindern, dass die menschliche Autonomie (der dabei zumindest implizit einen hohen ethischen Wert verliehen wird) durch KI-Systeme untergraben wird, womit insbesondere die Forderung nach Nachvollziehbarkeit und Transparenz von automatisierten Entscheidungen verbunden ist.³ Dieser Schutz der menschlichen Autonomie wird auch in der bereits oben erwähnten fünften Forderung des *Wiener Manifests* formuliert: „Automatisierte Entscheidungssysteme sollen die Entscheidungsfindung durch den Menschen nur unterstützen und nicht ersetzen“ (Wiener Manifest für Digitalen Humanismus, 2019). Autonomieschutz stellt auch die Kernforderung von Art. 22 Abs. 1 DSGVO dar, d. h. des Rechts, nicht einer ausschließlich auf einer automatisierten Verarbeitung beruhenden

² Eine ähnlich breite Definition von *Human-in-the-Loop* wird auch vertreten in Chang (2024, S. 223 f.): „This idea of putting the human in the loop has a variety of Implementations. In autonomous weapons, for instance, humans are ‚in the loop‘ in that they must *initiate* such weaponry; a machine is not allowed to decide by itself whether to bomb a village. Sometimes humans are in the loop when they have the power to abort a machine’s output. There are more interesting ways humans can be in the loop, too. If a machine-learning algorithm is uncertain about its reward function, it can observe or interact with humans and thereby – in principle – learn the human’s reward function. In mixed machine learning rules-based programming, a human might input specific weights to be assigned to factors in decision-making.“

³ Vgl. hierzu ausführlicher Kőszegi, Der autonome Mensch im Zeitalter des Digitalen Wandels, in: Hengtschläger und Rat für Forschung und Technologieentwicklung (2020, S. 62–90). Vgl. auch Koulu (2020, S. 41 ff.).

den Entscheidung mit rechtlich relevanten Folgen unterworfen zu werden.⁴ Der Schutz menschlicher Autonomie ist nur das erste von insgesamt drei Hauptzielen, die mit Menschenzentriertheit assoziiert werden können und somit von *Human-in-the-Loop*-Ansätzen (im hiesigen, weiteren Sinne) verfolgt werden.

Das zweite Hauptziel ist die Gewährleistung der Verantwortlichkeit bzw. der Zurechenbarkeit, bezieht sich also im Grunde auf die Haftung für die Folgen der jeweiligen automatisierten Entscheidung. *Köszegi* diskutiert bspw. die Tendenz, (z. T. selbst den einfachsten) automatisierten Entscheidungssystemen intuitiv *Handlungsmacht* zuzuschreiben: *Der Computer lehnt es ab; der Computer will es nicht*.⁵ Der Handlungsfähigkeit der Maschine sollte indes konsequent auch ihre Zurechnungsfähigkeit folgen, was aus praktischer Sicht wenig Sinn ergibt. Die Zulassung vollautomatisierter Entscheidungsprozessen (d. h. Entscheidungen, die ohne menschliche Beteiligung getroffen werden) führt also dazu, dass kein einschlägiger Verantwortlicher gefunden werden kann, der für etwaige negative Folgen bzw. Schäden dieser Entscheidungen haften würde.⁶ *Human-in-the-Loop*-Ansätze sollen dieses Problem vermeiden, indem am Entscheidungsprozess stets Menschen beteiligt sind, die bei Bedarf zur Rechenschaft gezogen werden können (Buckley et al., 2021). Das Ziel der Gewährleistung der Verantwortlichkeit lässt sich deutlich im Rahmen der Regulierung autonomer Fahrzeuge wiedererkennen, wobei die Verantwortung zwischen Fahrzeugführer, technischer Aufsicht und Hersteller als verschiedenen *Humans in the Loop* verteilt wird. § 1b Abs. 2 des deutschen Straßenverkehrsgesetzes verpflichtet bspw. den Fahrzeugführer bei Nutzung hoch- oder vollautomatisierter Fahrfunktionen, die Fahrzeugsteuerung bei Bedarf oder nach Aufforderung des Systems unverzüglich wieder zu übernehmen. § 1d Abs. 3 leg. cit. führt wiederum den Begriff einer technischen Aufsicht ein, d. h. einer natürlichen Person, die ggf. per Fernsteuerung das Fahrzeug deaktivieren bzw. Fahrmanöver freigeben kann. Fahrzeughersteller müssen zudem nach § 1e Abs. 2 Z. 2 leg. cit. sowie in Österreich nach § 1 Abs. 2 AutomatFahrV dafür sorgen, dass autonome Fahrzeuge in der Lage sind, die jeweils geltenden Straßenverkehrsregeln zu erfüllen (Gstöttner & Lachmayer, 2021). Ähnliche, auf die Gewährleistung von Verantwortlichkeit und Zurechenbarkeit zurückzuführende rechtliche Anforderungen sind auch im Zusammenhang mit der Automatisierung von Prozessen in der öffentlichen Verwaltung festzustellen: Das österreichische Bundesverfassungsgesetz enthält bspw. Vorschrif-

⁴ Nach EG 71 DSGVO sollten entsprechenden Verarbeitungen mit angemessenen Garantien verbunden sein, einschließlich u. a. des Anspruchs auf direktes Eingreifen einer Person. Art. 22 DSGVO wird in diesem Zusammenhang auch von der Expertengruppe ausdrücklich erwähnt. Vgl. Unabhängige Hochrangige Expertengruppe für Künstliche Intelligenz (2018, S. 19).

⁵ Vgl. *Köszegi*, Der autonome Mensch im Zeitalter des Digitalen Wandels, in: *Hengstschläger und Rat für Forschung und Technologieentwicklung* (2020, S. 74 ff.); vgl. auch *Nida-Rümelin & Weidenfeld* (2018, S. 90 ff.). Dabei geht es im Grunde um einen historisch häufig auftretenden *Topos* der Vermenschlichung von Maschinen; vgl. hierfür *Küchen & Sasdelli* (im Druck).

⁶ Aufgrund der fehlenden Verantwortung bzw. Zurechenbarkeit bei vollautomatisierten Entscheidungen lehnt bspw. *Luís Greco* die (rechtliche bzw. moralische) Möglichkeit eines Richterautomaten ab. Vgl. *Greco* (2020).

ten etwa zur Weisungsbildung und Verantwortlichkeit (Art. 20 B-VG) bzw. Amtshaftung (Art. 23 leg. cit.), aus denen auf eine menschliche Beteiligung bzw. auf den Bedarf eines menschlichen behördlichen Willens geschlossen werden kann (Wutscher, 2024).

Das dritte mit *Human-in-the-Loop* verfolgte Hauptziel ist die Verhinderung von (faktisch oder ethisch) fehlerhaften Entscheidungen. Dieses Ziel fußt auf der Annahme, dass automatisierte Entscheidungen qualitativ verschieden (bzw. aus ethischer Sicht schlechter)⁷ oder zumindest in Grenzfällen für besonders gravierende Fehler anfällig sind. Berichte von der unbeabsichtigten Übernahme von Vorurteilen und *Bias* in den jeweils verwendeten *Trainingsdaten* (Favaretto et al., 2019) sowie von sog. *Halluzinationen* (Ji et al., 2023), aber auch theoretische bzw. philosophische Diskussionen über fundamentale Unterschiede zwischen dem schlechterdings menschlichen Denkvermögen einerseits und der maschinellen Datenverarbeitung bzw. Berechnung andererseits⁸ dienen als Begründung für die Skepsis gegenüber vollautomatisierten Entscheidungen und für die damit einhergehende Forderung nach menschlicher Aufsicht. Am deutlichsten lässt sich dies in Art. 14 KI-VO wiedererkennen, in dem die Pflicht einer wirksamen menschlichen Aufsicht während der Dauer der Verwendung sog. *Hochrisiko-KI-Systeme* verankert wird. Gemäß Art. 14 Abs. 2 leg. cit. dient diese Aufsicht der Verhinderung oder Minimierung von Risiken für Gesundheit, Sicherheit oder Grundrechte, die entstehen können, wenn ein Hochrisiko-KI-System im Einklang mit seiner Zweckbestimmung oder im Rahmen einer vernünftigerweise vorhersehbaren Fehlanwendung verwendet wird. Ähnliche Überlegungen begründen auch die Forderung nach menschlicher Aufsicht im Zusammenhang mit dem bereits oben erwähnten Bereich autonomer Fahrzeuge.

Als Zwischenfazit ist festzuhalten: *Human-in-the-Loop*-Ansätze – d. h. im weiteren Sinne die im Rahmen mehrerer ethischen und rechtlichen Regulierungsrahmen formulierte Forderung, insbesondere kritische Entscheidungen trotz prinzipieller Automatisierbarkeit weiterhin Menschen zu überlassen – sollen im Einklang mit dem Digitalen Humanismus das Prinzip der Menschenzentriertheit vor allem im Hinblick auf die drei oben angeführten Ziele erfüllen, d. h. (1) den Schutz menschlicher Autonomie, (2) die Gewährleistung der Verantwortlichkeit bzw. der Zurechenbarkeit und (3) die Verhinderung von fehlerhaften Entscheidungen. Im Folgenden wird argumentiert, dass *Human-in-the-Loop*-Ansätze bei genauerer Betrachtung aufgrund häufig übersehener Nachteile gerade hinsichtlich dieser drei Ziele keinen echten Beitrag zur Erfüllung von Menschenzentriertheit im Sinne des Digitalen Humanismus zu leisten vermögen. Vielmehr bergen diese Ansätze die Gefahr in sich, in konservative, fortschritthindernde bzw. technologiefeindliche Konzepte zu münden.

⁷ Vgl. etwa Nida-Rümelin & Weidenfeld (2018, S. 102 ff.); Spiekermann (2020).

⁸ Vgl. etwa Dreyfus (1989); Fodor (2001); Weizenbaum (1977); Penrose (1994); Lakoff (1987). Für die gegenteilige Position vgl. etwa Minsky (1986) sowie bereits im 18. Jhdt. La Mettrie (2015). Für einen Überblick der Diskussion vgl. auch Russel & Norvig (2022, S. 1032 ff.); vgl. auch Lober & Sasdelli (2025).

24.3 Die Schattenseite von Human-in-the-Loop

Wie oben diskutiert besteht das erste mit *Human-in-the-Loop* verfolgte Ziel darin, menschliche Autonomie zu schützen. Dass dieses Ziel einfach durch die Beteiligung eines Menschen bei sonst automatisierten Entscheidungsverfahren wirklich erreicht werden kann, wird durch neue Erkenntnisse aus dem Bereich *Human-Computer-Interaction* in Frage gestellt. Die Haupteinwände weisen auf die Wirkung der Inklusion menschlicher Akteure in teilautomatisierten Entscheidungssysteme hin und lassen sich in drei Kategorien einteilen:

1. *Automation Bias*. Unter diesem Begriff versteht man die Tendenz des Menschen, automatisierten Entscheidungen übermäßig zu vertrauen und sie ggf. akritisch anzunehmen (vgl. Fischer-Abaigar et al., 2024, S. 6; ausführlicher etwa Gaddard et al., 2012; Alon-Barkat & Busuioc, 2023). Gerade bei sich häufig wiederholenden Verfahren könnte somit der vermeintliche Mehrwert von *Human-in-the-Loop* hinsichtlich der Wahrung der menschlichen Autonomie beträchtlich eingedämmt werden, wenn sich der am Entscheidungsprozess beteiligte Mensch sozusagen daran gewöhnt, auf die automatisierte Entscheidung der Maschine zu verlassen. Damit verwandt ist das Phänomen der sog. *Selective Adherence*, d. h. die Tendenz des Menschen, automatisierte Entscheidungen eher dann zu akzeptieren, wenn sie ohnehin mit den eigenen Erwartungen (bzw. Vorurteilen) im Einklang sind (vgl. Parycek et al., 2024b, S. 53; ausführlicher: Alon-Barkat & Busuioc, 2023).
2. *Algorithm Aversion*. Darunter ist das konverse Phänomen zu *Automation Bias* zu verstehen, d. h. die Abneigung gegenüber automatisierten Entscheidungen und die damit einhergehende Tendenz, menschliche Entscheidungen zu bevorzugen, und zwar auch in Fällen, bei denen automatisierte Entscheidungen menschliche i. d. R. nachweislich übertreffen (vgl. Fischer-Abaigar et al. 2024, S. 6; ausführlicher etwa Dietvorst et al., 2015). Dies hätte erhebliche negative Folgen nicht nur für die Wahrung der menschlichen Autonomie, sondern auch für die allgemeine Effizienz der Entscheidungsfindung.
3. *Langeweile*. Gerade bei sich häufig wiederholenden Prozessen kann Langeweile und Müdigkeit (hierbei wird häufig von *Alert Fatigue* gesprochen) dazu führen, dass die beteiligten Menschen ihre Überwachungstätigkeit nicht mehr wirksam wahrnehmen können.

Diese und andere Einwände (Koulu, 2020) im Zusammenhang mit *Human-Computer-Interaction* zeigen, dass die Beteiligung von Menschen an sonst automatisierten Entscheidungsprozessen mit neuen Herausforderungen gerade hinsichtlich der Wahrung der menschlichen Autonomie verbunden ist.⁹ Generell ist der Mensch für die Überwachung

⁹ Wohl zur Eindämmung dieser Nachteile hat der EU-Gesetzgeber in Art. 14 Abs. 4 KI-VO mehrere Vorgaben verankert, die von den mit der Aufsicht von KI-Systemen betrauten natürlichen Personen erfüllt werden müssen. Art. 14 Abs. 4 lit. b leg. cit. verweist dabei explizit auf *Automation Bias*.

automatisierter Systeme nicht gut geeignet. Die schiere Teilnahme eines Menschen garantiert mithin keinen Mehrwert für den Schutz der menschlichen Autonomie, sondern kann im schlimmsten Fall sogar zu einer vermeintlichen Legitimierung von Fehlentscheidungen führen (Parycek et al., 2024). Hinzu kommen ethisch problematische Folgen für die Menschen, die in solchen teilautomatisierten Prozessen jeweils eingebunden oder in Spiekermanns Worten ‚*eingespannt*‘ werden (Spiekermann, 2019, S. 66 f.). Spiekermann bezieht sich dabei auf die im Allgemeinen eher sehr schlechten Arbeitsbedingungen (niedrige Löhne, hoher Leistungsdruck, kaum Kontakte zu menschlichen Vorgesetzten usw.) denen Menschen ausgesetzt werden, die bei sonst vollautomatisierten Prozessen diejenigen Teilaufgaben übernehmen, die eigentlich aus technischen Gründen (noch) nicht auf effiziente Weise automatisierbar sind, etwa den Liefersdienst von online bestellten Waren. Mit *Human-in-the-Loop* besteht hier das durchaus ernstzunehmende Risiko, ähnliche Arbeitsbedingungen zwar nicht aus technischen, dafür aber aus rechtlichen Gründen zu fördern.

Auch in Bezug auf das zweite Ziel, d. h. die Gewährleistung von Verantwortlichkeit bzw. Zurechenbarkeit weist *Human-in-the-Loop* beträchtliche Probleme auf. Insbesondere besteht hierbei die Gefahr, dass Verantwortlichkeit für Fehlentscheidungen, die in komplexen Zusammenhängen entstehen und somit eigentlich auf mehrere menschliche, organisatorische und technische Faktoren zurückzuführen sind, übermäßig auf den jeweils beteiligten Menschen zugewiesen wird, die indes ggf. nicht über eine ausreichende Kontrollmacht oder gar über den Überblick über dem gesamten Verfahren verfügen könnten (Koulu, 2020). *Elish* führt für diese ethisch problematische Verteilung der Verantwortung für aus Menschen und Maschinen bestehende hybride Entscheidungssysteme den Begriff von sog. *Moral Crumple Zones* ein (Elish, 2019). Dieses Problem kann unter Umständen auch im Zusammenhang mit Art 14 Abs. 4 KI-VO (vgl. dabei auch EG 73 leg. cit.) entstehen, da es von den natürlichen Personen verlangt wird, die KI-Systeme überwachen, dass sie über die erforderliche Kompetenz, Ausbildung und Befugnis verfügen (vgl. dabei auch Constantino, 2022, S. 12 ff.).

Schließlich ist auch der Mehrwert von *Human-in-the-Loop*-Ansätzen bei der Verhinderung von Fehlentscheidungen fragwürdig. Erstens verlieren gerade in Anbetracht jüngster technologischen Entwicklungen die zugrunde liegenden, vermeintlichen qualitativen Unterschiede zwischen menschlichen und automatisierten Entscheidungen immer mehr an Bedeutung.¹⁰ Solange die Maschinen auf einschlägige Datensätze eingesetzt werden, werden automatisierte Entscheidungen *ceteris paribus* den menschlichen stets übertreffen. Liegen entsprechende Bedingungen vor, so mündet die (ggf. gesetzlich bedingte) pauschale Forderung menschlicher Aufsicht im Sinne von *Human-in-the-Loop* entweder in

Es ist indes fraglich, ob die Erfüllung dieser Pflichten realistisch ist. Vgl. dafür ausführlicher (im Zusammenhang mit der öffentlichen Verwaltung) etwa Constantino (2022).

¹⁰ Für einen aktuellen Überblick über Aufgaben, bei denen KI-Modelle das menschliche Leistungsniveau erreichen oder sogar übertreffen vgl. bspw. Stanford University, *Human-Centered Artificial Intelligence (HAI)* (2024, S. 73 ff.).

Ressourcenverschwendung oder in durch Menschen verursachten Fehler. Dabei spielen auch die oben kurz erwähnten Probleme der *Human-Machine-Interaction*, etwa *Automation Bias*, *Algorithm Aversion* usw. eine entscheidende Rolle.

Die obigen Erwägungen zeigen, dass *Human-in-the-Loop*-Ansätze im Hinblick auf die Ziele des Digitalen Humanismus – dabei allen voran die Förderung eines menschenzentrierten technologischen Fortschritts – über ihre relativ niedrigen Umsetzungskosten hinaus wenig anzubieten haben. Bei genauerer Betrachtung lösen die vermeintlichen Vorteile dieser Ansätze hinsichtlich des Schutzes der menschlichen Autonomie, der Gewährleistung von Verantwortlichkeit bzw. Zurechenbarkeit, der Verhinderung von fehlerhaften Entscheidungen und somit auch der Erfüllung des Prinzips der Menschenzentriertheit in Luft auf. Wie eingangs erwähnt stellt die digitale Transformation die wohl größte gesellschaftliche Herausforderung unserer Zeit dar. *Human-in-the-Loop* erweist sich in diesem Kontext als eine zu einfache, geradezu einfallslose Antwort auf eine eigentlich sehr komplexe Frage.¹¹ In Wahrheit bietet *Human-in-the-Loop* kein echtes Konzept zum Umgang mit Digitalisierung und Automatisierung. Vielmehr werden die damit verbundenen Fragen und Herausforderungen überhaupt vermieden, indem pauschal verlangt wird, dass der Mensch trotz zahlreicher Nachteile bei bestenfalls fragwürdigem Mehrwehrt konsequent an Entscheidungsprozessen beteiligt bleibt. Aus dieser Perspektive zeigt sich *Human-in-the-Loop* weniger als humanistische, menschenzentrierte Vorstellung von Mensch-in-der-Schleife, sondern – um *Dario Martires* Wortspiel zu verwenden – als fortschrittsverhindernde, etwas unheimliche Vorstellung von Mensch-in-der-Schlinge (Martire, 2024).

24.4 Eine alternative Perspektive zu Menschenzentriertheit

24.4.1 Menschenzentriertheit jenseits von Human-in-the-Loop

In Anbetracht der oben diskutierten Probleme erweist sich als ratsam, mit *Human-in-the-Loop* auch die mit ihm verfolgten Hauptziele, die eine Verwirklichung von Menschenzentriertheit darstellen sollen, zu hinterfragen; denn eine nähere Betrachtung der Ziele, die Menschen mit der Entwicklung und dem Einsatz neuer Technologien historisch verfolgt haben und immer noch verfolgen, legt eine andere Gestaltung von Menschenzentriertheit nahe, in deren Mittelpunkt keine pauschale Anbindung des Menschen an sonst

¹¹ So etwa *Koulu* (2020, S. 46): „What makes human control over automation such a tempting solution for digital technologies is its relatively easy implementation. Human control comes with the promise of a relatively simple and operational way to address AI-related issues: to operationalize human control both within the legal system and software development requires relatively easy political choices that can be met by establishing a legal right to human oversight and then creating technical design solutions for implementing this right within the technological architecture. Especially if the alternative is to engage in grueling societal debates over the dynamics of technological change and the critical analysis of existing societal power imbalances, such solutions provide attractive and straightforward policy actions.“

automatisierten Prozessen steht, sondern einfach der Mensch selbst, und zwar mit seinen klassischen, seit jeher bestehenden Bedürfnissen.

Menschen sind sozial, sie leben und arbeiten zusammen, spielen, interagieren und wollen ihre sozialen Kontakte pflegen. Sie besitzen mentale Modelle oder Schemata, die sie anleiten, und ihr Verhalten in der Welt führen. Interaktive Systeme müssen den Menschen helfen, ihre Ziele zu erreichen. Sie müssen also in der Lage sein, die Kommunikation *zwischen Menschen* zu unterstützen und zu vermitteln (Sharp et al., 2019). Aus dieser Perspektive sollte im Sinne der Menschenzentriertheit jede Interaktion zwischen Menschen und Maschinen zumindest mittelbar dem Aufbau oder die Aufrechterhaltung zwischenmenschlicher Interaktion dienen (Sypher & Collins, 2001). Dieser Aspekt wird bei der Begründung von *Human-in-the-Loop*-Ansätzen i. d. R. völlig ignoriert. Eine Möglichkeit, Menschen mit digitaler Technologie zu unterstützen, ist durch Ausbildung, Schulung und Bereitstellung von Handbüchern. Indessen wollen Menschen sich oft nicht die Zeit nehmen, zu lesen und zu lernen, was zu den oben diskutierten Problemen führt. Eine effektive Interaktion zwischen Menschen und Maschine (eine erforderliche Voraussetzung für die Effektivität von *Human-in-the-Loop*) setzt also voraus, dass die Maschinen leicht verständlich sind, eine intuitive Interaktion gewährleisten und angemessene Hilfe bieten, d. h. auf die Benutzer:innen und deren Erfahrungsstand abgestimmt sein und als Reaktion auf die Eingaben der Nutzer:innen Hilfe und nützliches Feedback bieten können – eben wie ein Mensch es tun würde.

24.4.2 Intuitive Interaktion und universelle Nutzbarkeit

Sind nun bei digital vermittelter Kommunikation die notwendigen Hinweise, die in einer persönlichen Situation hilfreich wären, nicht verfügbar, kommt es leicht zu Fehlern und Pannen. Solche Fehler oder „*Failures*“ können kostspielig, verschwenderisch, erschreckend und frustrierend sein. Digitale Anwendungen, die schwer zu verstehen sind, führen dazu, dass sich die Benutzer:innen *in einer unverständlichen Dialogbox gefangen fühlen, die Ja/Nein-Antworten verlangt* (Shneiderman, 2002, S. 24), und dass sie Zeit verlieren. Bereits in den 80er Jahren forderte Woodson, dass Produkte so gestaltet werden sollten, dass die Benutzer:innen ihre Aufgabe mit einem Minimum an Stress und einem Maximum an Effizienz erledigen können (Woodson, 1981; Tillman et al., 2016). Weitere Forscher:innen haben gezeigt, dass bei der Produktentwicklung oft das Produkt und nicht die Benutzer:innen im Mittelpunkt stehen: Die Benutzer:innen sollten dabei lernfähig sein und sich flexibel an die Technologie anpassen können (vgl. etwa Rubin, 1994; Norman, 2004; Cooper, 2004). Dies hat Usability-Studien populär und zu einem profitablen Geschäft gemacht (Rubin, 1994). Kelley schlägt demgegenüber vor, dass die Verbesserung über das *Einfachmachen* hinausgehen sollte: Das Produkt oder die Interaktion sollte es den Benutzer:innen ermöglichen, das zu erreichen, was sie zu erreichen versuchen; das Gefühl zu geben, die Kontrolle zu haben; sowie ihre Erfahrungen mit der Technologie neu gestalten (Kelley & Littman, 2001). Bei der „universellen Nutzbarkeit“ („*universal*

usability“) geht es darum, die Kluft zwischen dem, was die Nutzer:innen wissen wollen, und dem, was sie wissen müssen, zu überbrücken und die Nutzer:innen in die Lage zu versetzen, ihre Aufgaben mit Hilfe von digitalen Technologien zu erfüllen.

Diese Erwägungen fokussieren auf die Verbesserung der Interaktion zwischen Mensch und Maschinen und münden theoretisch in einen Zustand, bei dem der Mensch genauso gut mit der Maschinen interagieren kann wie mit anderen Menschen. Im Sinne der Menschenzentriertheit sollten sie einen wichtigen Bestandteil von *Human-in-the-Loop*-Ansätzen ausmachen – dies würde viele der oben diskutierten Probleme zumindest abmildern. Allerdings werden solche Aspekte bei *Human-in-the-Loop*-Ansätzen nur selten berücksichtigt. Dies ergibt im Prinzip auch Sinn: Würde sich die Maschine schon wie ein Mensch verhalten, warum wäre noch ein *Human-in-the-Loop* nötig?

Die Entwicklung digitaler Technologien kann aber sogar noch weiter gehen und darauf abzielen, eine noch höhere Nutzerzufriedenheit zu erreichen, die digitale Beteiligung zu erweitern und digitale Leistungen zu nützen. Dies erfordert ein Umdenken in Bezug auf die Beziehung zwischen Nutzer:innen und Technologie, der Umgang mit der digitalen Kluft, die Vielfalt der NutzerInnen und ihre Bedürfnisse. Die Berücksichtigung der Vielfalt fördert die Qualität, da sie gewährleistet, dass alle Nutzer:innen Zugang zu digitalen Produkten und Leistungen haben. Bei der Barrierefreiheit geht es, wie das Wort schon sagt, darum, sicherzustellen, dass alle Menschen Zugang zu den im Internet verfügbaren Informationen und Leistungen haben. Bereits in den 1950er Jahren argumentierte *Galbraith*, dass die Schwächsten unserer Gesellschaft nicht von der Schaffung von Wohlstand profitieren würden, wenn wir nicht in sie investieren, und dass sich Armut, wenn unkontrolliert, von selbst fortsetzen kann (*Galbraith*, 1958). Im digitalen Kontext wird eine Behinderung in der Regel als ein Zustand verstanden, der es schwierig macht, herkömmliche Eingabe- und Ausgabegeräte so zu nutzen, wie sie gedacht sind. Normen, die für Benutzer:innen mit körperlichen, kognitiven Behinderungen oder finanziellen Schwierigkeiten entwickelt wurden, führen zu Änderungen, die Benutzer:innen mit unterschiedlichen Vorlieben, Aufgaben, Fähigkeiten, Software und Hardware zugutekommen (vgl. Section 508 des US-amerikanischen Rehabilitation Act, 1998). Diese Veränderungen sind für alle Nutzer:innen positiv und können die Beteiligung der gesamten Bevölkerung erhöhen. Dies ist aber mit *Human-in-the-Loop* schwer zu vereinbaren: *In the Loop* sollen nämlich nur diejenigen sein, die über eine besondere entsprechende Qualifikation verfügen (etwa, wie oben diskutiert, i. S. v. Art. 14 Abs. 4 KI-VO). Die Grundsätze der *Web Accessibility Initiative* (W3C Web Accessibility Initiative, 2025) sind zwar wichtig für die Einhaltung von ethischen und rechtlichen Vorschriften, aber es gibt auch sehr gute geschäftliche Gründe, die digitale Welt auch für behinderte Nutzer:innen zugänglich zu machen. Hindernisse können rechtliche und regulatorische Barrieren, Budgets, technologischer Wandel, staatliche Standards sowie nationale und internationale Harmonisierung sein. Die Einführung von E-Government kann risikoreich, teuer und schwierig sein und erfordert ein angemessenes Änderungsmanagement. Oft werden behinderte Nutzer:innen zu sehr treuen Kunden, wenn sie Anbieter:innen finden, die ihren besonderen Bedürfnissen entgegenkommen. Die Gestaltung für Nutzer:innen mit geringeren Fähigkeiten führt

nicht zu einer Verdummung, sondern ist benutzerfreundlicher und zieht mehr noch mehr Nutzer:innen an, und führt zu vertrauenswürdigen und nutzbaren Leistungen, v. a. die, die im öffentlichen Sektor angeboten werden. Das Wissen um Zugänglichkeit für alle und die Umsetzung von Lösungen führen zu informativen, nützlichen und nutzbaren Interaktion zwischen Bürger:innen und Organisationen der öffentlicher Verwaltung.

24.4.3 Digitale Inklusion und nutzerzentrierte öffentliche Dienste

Erfolgreiche, menschenzentrierte digitale Technologien sollen sich auf die menschlichen Ziele, die die Nutzer:innen erreichen wollen, konzentrieren und im Einklang mit den menschlichen Bedürfnissen ästhetische Werte sein (Norman, 2004). Sie können jedoch noch mehr erreichen, indem sie Nutzer:innen dabei unterstützen, Sicherheit, Leistung, Entspannung, Spaß und Entdeckung zu erleben. Die Herausforderung besteht also vor allem darin, zu verstehen, was Benutzer:innen, wollen, und Ihnen zu helfen, es zu erreichen. Um das Prinzip von Menschenzentriertheit zu erfüllen, müssen Technologien daher so entwickelt werden, dass Nutzr:innen ihre Ziele schnell und in einer Atmosphäre des Vertrauens und der Verantwortung erreichen können. Sie sollten in der Lage sein, Informationsquellen und Angebote, die Ihnen gemacht werden, und der Privatsphäre, die Ihnen versprochen wird, zu vertrauen. So können sie die Verantwortung für ihre Entscheidungen und Kommunikation mit anderen übernehmen. Die zugrundeliegenden Systeme sollten eine Infrastruktur bereitstellen, die dem Benutzer ein Gefühl der Zuverlässigkeit und Sicherheit vermittelt, so dass sie sich auf Arbeit und Beziehungen konzentrieren können (Shneiderman, 2002).

Menschenzentriertes Design bedeutet also, dass die Organisation, die Dienstleistungen anbietet, die Art und Weise, wie ihre Dienstleistungen und Produkte bereitgestellt werden, überdenkt, indem sie sich auf die Nutzer:innen und potenziellen Nutzer:innen konzentriert, sie einbezieht und mit ihnen zusammenarbeitet, um ihren Input und ihr Feedback zu berücksichtigen. Dies kann unabhängig von *Human-in-the-Loop* erreicht werden. Die *Tallinn Declaration* (Europäische Kommission, 2017) stellt die Nutzer:innen in den Mittelpunkt der Gestaltung und Erbringung öffentlicher Dienstleistungen, und nutzerzentrierte Prinzipien sollen für die Gestaltung und Erbringung von (digitalen) öffentlichen Dienstleistungen verwendet werden. Die *Berlin Declaration* (Europäische Kommission, 2020) stellt menschenorientierte Systeme in den Mittelpunkt des öffentlichen Sektors, um eine widerstandsfähige und nachhaltige digitale Gesellschaft, Vertrauen und Sicherheit bei digitalen Interaktionen mit öffentlichen Verwaltungen, digitale Souveränität und Interoperabilität zu erreichen. Dies erfordert die Einhaltung von Grundsätzen wie Gültigkeit und Achtung der Grundrechte und demokratischen Werte in der digitalen Sphäre, um soziale Teilhabe, digitale Integration, Befähigung und digitale Kompetenz zu gewährleisten. Die Umsetzung solcher Ideen und Werte durch öffentliche Verwaltungen spiegelt sich in der Nutzung von Ansätzen wie *Co-Creation* für öffentliche Strategien und Dienstleistungen wider (Mureddu & Osimo 2020). *Co-Creation* bezieht ein breites Spektrum

von Akteur:innen ein, um komplexe Probleme zu lösen, innovative Lösungen zu fördern und qualitativ hochwertige Dienstleistungen zu entwickeln, die den Bedürfnissen der Nutzer entsprechen (Ansell & Torfing, 2021; Rodriguez Müller et al., 2021). Um ethische Abweichungen zu minimieren, müssen Interaktionen, Aktivitäten, Verhaltensweisen und Einstellungen zwischen Organisationen des öffentlichen Sektors und ihren Stakeholdern vorgeschrieben werden (Edelmann & Pereira, 2024).

24.5 Schlussbemerkungen

Human-in-the-Loop will Menschen weiterhin bei klassischen Prozessen teilnehmen lassen, übersieht dabei, dass digitale Transformation nicht bloß die Automatisierung dieser klassischen Prozesse, d. h. den Ersatz von Menschen durch Maschinen bedeutet, sondern vielmehr die Weiterentwicklung bzw. den Ersatz der alten klassischen Prozesse durch neue, an das digitale Zeitalter angepasste Prozesse. Bei vielen dieser Prozesse ist im idealen Fall keine menschliche Beteiligung mehr nötig.

Aus Sicht der Interaktion zwischen Menschen und Maschine führt *Human-in-the-Loop* zu einer Art Paradox: Seine vollkommene Umsetzung würde erfordern, dass der Mensch genauso gut mit der Maschine wie mit anderen Menschen interagieren kann. In diesem Fall wäre die menschliche Beteiligung aber zwecklos. Die Erfüllung des Prinzips der Menschenzentriertheit erfordert also keine pauschale Anbindung von Menschen in sonst automatisierten Prozessen. Die menschliche Beteiligung an diesen Verfahren ist kein *Zweck an sich*, sondern dient, wenn überhaupt, den üblichen Zwecken des menschlichen Lebens. Menschenzentriertheit erfordert somit eine Abkehr von traditionellen, oft statischen Konzepten wie *Human-in-the-Loop* hin zu dynamischen, partizipativen Prozessen, die den digitalen Humanismus in seiner ganzen Breite erfassen. *Co-Creation*-Ansätze bieten eine vielversprechende Grundlage, um Menschen – inklusive denjenigen, die von *Human-in-the-Loop* leider eher übersehen bzw. ignoriert werden – als aktive Mitgestalter digitaler Systeme einzubeziehen. Sie fördern nicht nur die Akzeptanz und Legitimität neuer Technologien, sondern ermöglichen auch eine breite Einbindung von Perspektiven, die für eine inklusive und gerechte digitale Gesellschaft essenziell sind. Diese Ansätze setzen jedoch voraus, dass Organisationen die Bereitschaft entwickeln, Strukturen und Denkweisen zu transformieren, um den Anforderungen einer zunehmend vernetzten Welt gerecht zu werden.

Zukunftsweisende digitale Transformation darf nicht nur auf die Lösung bestehender Probleme mit bestehenden Mitteln ausgerichtet sein, sondern sollte aktiv danach streben, neue Möglichkeiten zu schaffen, die das Potenzial des technologischen Fortschritts mit humanistischen Werten vereinen. Hierbei spielen Bildung, kontinuierlicher Dialog und die Förderung digitaler Kompetenz eine zentrale Rolle. Nur durch ein bewusstes Engagement für Innovation und die Zusammenarbeit aller relevanten Stakeholder kann eine digitale Zukunft gestaltet werden, die nicht nur effizient, sondern auch ethisch und nachhaltig ist.

Literatur

- Alon-Barkat, S., & Busuioc, M. (2023). Human-AI Interactions in Public Sector Decision-Making: 'Automation Bias' and 'Selective Adherence' to Algorithmic Advice. *Journal of Public Administration Research and Theory*, 33(1), 153–169.
- Ansell, C., & Torfing, J. (2021). Co-Creation: The New Kid on the Block in Public Governance. *Policy & Politics*, 49(2), 211–230.
- Buckley, R., Zetzsche, D. A., Arner, D. W., & Tang, B. (2021). Regulating Artificial Intelligence in Finance: Putting the Human in the Loop. *Sydney Law Review*, 43(1), 43–81.
- Chang, R. (2024). Human in the Loop! In D. Edmonds (Hrsg.), *AI Morality* (S. 222–234). Oxford: Oxford University Press.
- Constantino, J. (2022). Exploring Article 14 of the EU AI Proposal: Human in the Loop Challenges when Overseeing High-Risk AI Systems in Public Service Organisations. *Amsterdam Law Forum*, 14(3), 1–17.
- Cooper, A. (2004). *The Inmates Are Running the Asylum: Why High-Tech Products Drive Us Crazy and How to Restore the Sanity*. Carmel: Pearson Higher Education.
- Dietvorst, B. J., Simmons, J. P., & Massey, C. (2015). *Journal of Experimental Psychology*, 144(1), 114–126.
- Dreyfus, H. L. (1989). *Was Computer nicht können*. Frankfurt a.M.: Athenäum.
- Edelmann, N., & Pereira, J. V. (2024). Public Administrators' Perception of Ethical Issues in Public Sector Co-Creation Processes and Outcomes: A Meta-Synthesis of Cases. In S. Chun, G. Karuri-Sebina, J. Przybilovicz, L. Barbosa & J. Braga (Hrsg.), *Proceedings of the 17th International Conference on Theory and Practice of Electronic Governance* (S. 14–21). ACM.
- Elish, M. (2019). Moral Crumple Zones: Cautionary Tales in Human-Robot Interaction. *Engaging Science, Technology, and Society*, 5, 40–60.
- Europäische Kommission (2017). Ministerial Declaration on eGovernment – Tallinn Declaration. <https://digital-strategy.ec.europa.eu/en/news/ministerial-declaration-egovernment-tallinn-declaration>. Zugegriffen: 15. Jan. 2025.
- Europäische Kommission (2020). Berlin Declaration on Digital Society and Value-based Digital Government. <https://digital-strategy.ec.europa.eu/en/news/berlin-declaration-digital-society-and-value-based-digital-government>. Zugegriffen: 15. Jan. 2025.
- Favaretto, M., De Clercq, E., & Elger, B. S. (2019). Big Data and discrimination: perils, promises and solutions. A systematic review. *Journal of Big Data*, 6, 1–27.
- Fischer-Abaigar, U., Kern, C., Barda, N., & Kreuter, F. (2024). Bridging the Gap: Towards an Expanded Toolkit for AI-driven Decision-Making in the Public Sector. *Government Information Quarterly*, 41(4), 1–22.
- Fodor, J. (2001). *The Mind Doesn't Work That Way*. Cambridge: MIT Press.
- Gaddard, K., Roudsari, A., & Wyatt, J. (2012). Automation Bias: A Systematic Review of Frequency, Effect Mediators, and Mitigators. *JAMIA*, 19, 121–127.
- Galbraith, J. K. (1958). *The Affluent Society*. Boston: Houghton Mifflin.
- Geyer, C. (2023). Digitaler Humanismus – by Design. In T. Krause (Hrsg.), *Die Praxis des Digitalen Humanismus* (S. 269–282). Wiesbaden: Springer VS.
- Greco, L. (2020). Richterliche Macht ohne richterliche Verantwortung: Warum es den Roboter-Richter nicht geben darf. *Rechtswissenschaft*, 11, 29–62.
- Gstöttner, S., & Lachmayer, K. (2021). Digitalisierung des Straßenverkehrsrechts im Zusammenhang mit automatisiertem Fahren. *ZVR – Zeitschrift für Verkehrsrecht*, 12a, 478–482.
- Hengtschläger, M., & Rat für Forschung und Technologieentwicklung (Hrsg.). (2020). *Digitaler Wandel und Ethik*. Salzburg: EcoWing.

- Ji, Z., Lee, N., Frieske, R., Yu, T., Su, D., Xu, Y., Ishii, E., Bang, Y.J., Madotto, A., & Fung, P. (2023). Survey of Hallucination in Natural Language Generation. *ACM Computing Surveys*, 55(12), 1–38.
- Kelley, T. & Littman, J. (2001). *The Art of Innovation*. New York: Currency.
- Koulu, R. (2020). Human Control over Automation: EU Policy and AI Ethics. *European Journal of Legal Studies*, 12(1), 9–46.
- Küchen, J., & Sasdelli, D. The Interplay Between the Anthropomorphism of Automata and the Mechanisation of Humanity as Represented in Contemporary Cinema. In I. MacDuff & B. Falconer (Hrsg.), *Logic, Modern Literature, and Artificial Intelligence*. Erscheint voraussichtlich 2026.
- de La Mettrie, J. O. (2015). *L'Homme machine*. Stuttgart: Reclam.
- Lakoff, G. (1987). *Women, Fire, and Dangerous Things*. Chicago: University of Chicago Press.
- Lober, J., & Sasdelli, D. (2025). Die Mechanisierung des Geistes in Ghost in the Shell. In G. Bleser & C. Clemens (Hrsg.), *Körperliche Gegenwart: Figurationen durch Status, Macht und Sichtbarkeit* (S. 149–172). Münster: Lit.
- Martire, D. (2024). Human in the Loop. L'essere umano come fattore condizionante della – o condizionato dalla – intelligenza artificiale. *Rivista Italiana di Informatica e Diritto*, 6(2), 467–481.
- Minsky, M. (1986). *The Society of Mind*. New York: Simon & Schuster.
- Mureddu, F., & Osimo, D. (2020). Co-Creation of Public Services – Why and How? *Revue Européenne d'Économie et Management des Services*, 25, 177–200.
- Nida-Rümelin, J., & Weidenfeld, N. (2018). *Digitaler Humanismus*. München: Piper Verlag.
- Norman, D. (2004). *Emotional Design: Why We Love (or Hate) Everyday Things*. New York: Basic.
- Parycek, P., Siegel, J., Hunt, J., Voigt, T., Novak, M., & Schmid, M. (2024). *Von der Digitalisierung zur Automatisierung des Verfahrens*. Berlin: Fraunhofer Fokus.
- Penrose, R. (1994). *Shadows of the Mind*. Oxford: Oxford University Press.
- Rodriguez Müller, A., Flores, C. C., Albrecht, V., Steen, T. & Crompvoets, J. (2021). A Scoping Review of Empirical Evidence on (Digital) Public Services Co-Creation. *Administrative Sciences*, 11(1), 1–21.
- Rubin, J. (1994). *Handbook of Usability Testing: How to Plan, Design, and Conduct Effective Tests*. New York: Wiley.
- Russel, S., & Norvig, P. (2022). *Intelligence. A Modern Approach*. Harlow: Pearson.
- Sharp, H., Rogers, Y., & Preece, J. (2019). *Interaction Design: Beyond Human-Computer Interaction*. New York: Wiley.
- Shneiderman, B. (2002). *Leonardo's Laptop: Human Needs and the New Computing Technologies*. Cambridge: MIT Press.
- Spiekermann, S. (2019). *Digitale Ethik*. München: Droemer HC.
- Spiekermann, S. (2020). Zum Unterschied zwischen künstlicher und menschlicher Intelligenz und den ethischen Implikationen der Verwechslung. In M. Hengstschläger, Rat für Forschung & Technologieentwicklung (Hrsg.), *Digitaler Wandel und Ethik* (S. 94–125). Salzburg: EcoWing.
- Stanford University, & Human-Centered Artificial Intelligence (HAI) AI Index Report 2024. <https://aiindex.stanford.edu/report/>. Zugriffen: 15. Jan. 2025.
- Strauß, S., & Bogner, A. (2020). Demokratische Herausforderungen im Zeitalter des digitalen Wandels. In M. Hengstschläger & Rat für Forschung und Technologieentwicklung (Hrsg.), *Digitaler Wandel und Ethik* (S. 22–37). Salzburg: Ecwin.
- Sypher, B. D., & Collins, K. (2001). Virtual-online communities: how might new technologies be related to community? In G. Shepherd & E. W. Rothenbuhler (Hrsg.), *Communication and Community* (S. 191–200). Mahwah: Routledge.
- Tillman, B., Fitts, D., Woodson, W. E., Rose-Sundholm, R., & Tillman, P. (2016). *Human Factors and Ergonomics Design Handbook* (3. Aufl.). New York: McGraw-Hill Education.

- Unabhängige Hochrangige Expertengruppe für Künstliche Intelligenz (2018). Ethik-Leitlinien für eine vertrauenswürdige KI. <https://digital-strategy.ec.europa.eu/en/library/ethics-guidelines-trustworthy-ai>. Zugegriffen: 31. Dez. 2025.
- W3C Web Accessibility Initiative (WAI) <https://www.w3.org/WAI/>. Zugegriffen: 15. Jan. 2025.
- Weizenbaum, J. (1977). *Die Macht der Computer und die Ohnmacht der Vernunft*. Frankfurt a.M.: Suhrkamp.
- Werthner, H. (2020). Das Wiener Manifest für Digitalen Humanismus. In M. Hengstschläger & Rat für Forschung und Technologieentwicklung (Hrsg.), *Digitaler Wandel und Ethik* (S. 384–413). Salzburg: Edition Rs. Zitat: S. 399.
- Wiener Manifest für Digitalen Humanismus (2019). Vienna Manifesto on Digital Humanism (deutsche Fassung). https://dighum.ec.tuwien.ac.at/wp-content/uploads/2019/07/Vienna_Manifesto_on_Digital_Humanism_DE.pdf. Zugegriffen: 31. Dez. 2024.
- Woodson, W. (1981). *Human Factors Design Handbook*. New York: McGraw-Hill Education.
- Wutscher, C. (2024). Der Einsatz von KI im Verwaltungsverfahren. *ecolex – Fachzeitschrift für Wirtschaftsrecht*, 10, 895–898.

Univ.-Prof. Mag. Dr. Peter Parycek, MAS MSc, ist Vizerektor für Lehre/Wissenschaftliche Weiterbildung und digitale Transformation der Universität für Weiterbildung Krems und leitet als Universitätsprofessor für E-Governance das Department für „E-Governance in Wirtschaft und Verwaltung“. Seit 2017 führt er das Kompetenzzentrum Öffentliche IT (ÖFIT) am Fraunhofer FOKUS Institut Berlin. Seine Forschungsschwerpunkte sind der Einsatz von Technologie in Regierung und Verwaltung und deren Wirkungen auf die bestehenden Rechts- und Organisationssysteme, u.a. Electronic Governance, Digitale Transformation in Staat, Gesellschaft und Wirtschaft sowie Strategien, Politiken & Regulierung für das digitale Zeitalter.

Dr. phil. Diogo Sasdelli, M.A., ist Senior Researcher am Department für E-Governance und Verwaltung an der Universität für Weiterbildung Krems. Seine Forschungsschwerpunkte umfassen Bereiche wie Rechts- und Verwaltungsinformatik, (deontische) Logik, juristische Methodenlehre, Argumentationstheorie und philosophische Rhetorik sowie die vielen Bereiche des Informationsrechts (insbesondere Datenschutzrecht) und der Technologieethik.

Noella Edelmann, PhD MSc MAS, ist Senior Researcher am Department für E-Governance und Verwaltung an der Universität für Weiterbildung Krems und Gastprofessorin am Public Governance Institute der KU Leuven, Belgien. Ihre Forschungsschwerpunkte sind E-Governance, insbesondere der organisatorische Wandel und die digitale Transformation in Organisationen des öffentlichen Sektors, neue Arbeitsweisen, Beteiligung und Nachhaltigkeit. Ein weiteres besonderes Forschungsinteresse ist die digitale Ethik in der öffentlichen Verwaltung.

Digitaler Humanismus in der Wiener Wirtschaftsförderung

25

Roadmaps und Praxistips von Wiener Unternehmen

Kristina Maurer und Eva Czernohorszky

Zusammenfassung

Der Beitrag beleuchtet, wie die Wirtschaftsagentur Wien gemeinsam mit Kollaborationspartnern wie dem Wiener Wissenschafts-, Forschungs- und Technologiefonds (WWTF) den Digitalen Humanismus in der Wirtschaftsförderung vorantreibt. Durch Förderprogramme wie „Roadmaps Digitaler Humanismus“ werden Unternehmen und Forschungseinrichtungen unterstützt, menschenzentrierte digitale Geschäftsmodelle und Arbeitsprozesse zu etablieren. Beispiele erfolgreicher Förderprojekte und Kurzinterviews mit geförderten Unternehmer*innen zeigen, wie digitale Technologien ethisch und nachhaltig implementiert werden können.

Am Ende der Roadmap-Entwicklung und Etablierung des Themas in der strategischen Planung sollte der erste Schritt zum gelebten Digitalen Humanismus in der Unternehmenspraxis stehen.

K. Maurer (✉) · E. Czernohorszky
Wirtschaftsagentur, Wien, Österreich
E-Mail: Maurer@wirtschaftsagentur.at

E. Czernohorszky
E-Mail: czernohorszky@wirtschaftsagentur.at

25.1 Einleitung

Digitalisierung ist mittlerweile untrennbar verflochten mit der ökonomischen, politischen, gesellschaftlichen und persönlichen Lebensrealität des Menschen. Wie wir menschenzentrierte Innovationsprozesse und die menschenzentrierte Implementierung digitaler Technologie vorantreiben können (im Sinne eines Umdenkens von „computer-literate people“ zu „people-literate-technology“) (Strassnig et al., 2019), wird in Wien von einem stetig wachsenden Netzwerk bestehend aus der öffentlichen Verwaltung, Forschungseinrichtungen, Unternehmen und Fördergeber*innen unter dem „Hashtag“ Digitaler Humanismus verhandelt.

25.2 Der Wiener Weg des Digitalen Humanismus

Seit 2019 hat die Stadt die Positionierung als globale Vorreiterin eines menschenzentrierten Zugangs zu Design, Entwicklung und Implementierung digitaler Technologien als zentrales Anliegen definiert.

Der Digitale Humanismus ist essenzieller Bestandteil des Wiener Weges der Digitalisierung, der in der Wiener Wirtschafts- und Innovationsstrategie WIEN 2030 und in der Digitalen Agenda der Stadt Wien für die Aktionsfelder Arbeit, Bildung, Ökonomie, Kunst & Kultur, Gesundheit und Nachhaltigkeit (Stadt Wien, 2022) definiert wurde.

Wien schlägt damit ganz bewusst einen Weg ein, der sich auch auf nationaler und EU-Policy-Ebene in Regulierungsinstrumenten wie dem EU AI Act, dem Digital Services Act, dem Digital Markets Act, der DSVGO, oder auch den Data und Data Governance Acts widerspiegelt, die digitale Innovation und technologischen Fortschritt mit dem Wohlbefinden des Menschen in Einklang bringen sollen.

25.3 Digitaler Humanismus in Unternehmens- und Forschungsförderungsprogrammen

„Transformation braucht Orientierung. Und Unternehmer*innen, die Widersprüche der digitalen Welt produktiv in Wertepreferenzen transformieren. Derartige Innovationen verdienen Unterstützung – damit der technologische Wandel von heute die Basis unseres Wohlstandes von morgen ist.“ (Klemens Himpele, CIO der Stadt Wien)

Nun geht es also darum, diese Werte auch im Aktionsfeld der Wirtschaft und Alltag des unternehmerischen und forschenden Tuns aktiv umzusetzen und zu leben, um wettbewerbsfähig zu bleiben, die hohe Lebensqualität in Wien zu erhalten und die Digitalisierung als Enablerin für eine erfolgreiche Wirtschaft innerhalb der planetaren Grenzen zu nutzen. Erfahrungswerte, an denen sich KMU bei der Umsetzung digital-humanistischer Strategien orientieren können, gibt es von Wiener Unternehmer*innen, die auf erste erfolgreich abgeschlossene Fördervorhaben zurückblicken.

2022 lancierten Wirtschaftsagentur und WWTF den Fördercall Roadmaps Digitaler Humanismus, um Forschungseinrichtungen und Unternehmen bei der Entwicklung und Umsetzung von Strategien des Digitalen Humanismus zu unterstützen. In kollaborativen Projekten arbeiteten die geförderten Organisationen daran, einen menschenzentrierten Umgang mit digitaler Technologie sowie menschenzentrierte digitale Geschäftsmodelle und Arbeitsprozesse zu etablieren.

Die von den Fördernehmer*innen der Wirtschaftsagentur Wien umgesetzten Vorhaben machten dabei vor allem die Vielfalt individueller Ansätze sichtbar: KMU identifizieren von barrierefreien digitalen Kommunikationstools und Bewerbungsprozessen über den ressourcenschonenden Umgang mit digitaler Infrastruktur (sowohl im Hinblick auf Hard- als auch Software) bis hin zur Auseinandersetzung mit Bias-Vermeidung in den KI-Systemen von Chatbot-Produkten unterschiedlichste, erste Anknüpfungspunkte für den digitalen Humanismus in der eigenen Unternehmenspraxis.

25.4 Roadmaps Digitaler Humanismus: Die Förderprojekte der Wirtschaftsagentur Wien im Überblick

Vollpension Generationen Café GmbH und das Institut für Sozialethik Uni Luzern erarbeiteten anhand des Fallbeispiels der internen digitalen Kommunikation bei der Vollpension einen Leitfaden, der u. a. barrierefreie, digitale Bewerbungsprozesse, die Auswahl sicherer, effizienzverbessernder, datenbasierter Systeme und einen Wertekatalog für die Auswahl digitaler Tools in Unternehmen beinhaltet.

leiwand.ai und Plattform Industrie 4.0 arbeiteten gemeinsam an einer Roadmap für Trustworthy AI. Im Rahmen mehrerer Workshops wurden sowohl die relevantesten Märkte für Trustworthy AI-Geschäftsmodelle identifiziert (Arbeitsmarkt, öffentliche Verwaltung, Medien, Gesundheit) als auch konkrete Angebote und Produkte abgeleitet, die den Kapazitätsausbau im Bereich ethische und vertrauenswürdige KI betreffen, und die leiwand.ai für den KI-Sektor in Österreich anbieten wird.

Goodshares und WU Executive Academy entwickelten eine Roadmap mit individuellen, praktischen Schritten zur Umsetzung des Digitalen Humanismus, angepasst an die jeweiligen Produkte und Services. Sie stellen nun anderen Unternehmen einen Management Guide Digitaler Humanismus zur Verfügung, der eine Anleitung zur Erhebung des digitalen Impacts der eigenen Unternehmenspraxis bietet.

Auf Basis der von Radiology Center und Spiritdesign erarbeiteten Roadmap für den Einsatz künstlicher Intelligenz in der Radiologie konnte das Radiology Center ein KI-Betriebssystem identifizieren, das einen sicheren und ethischen Einsatz der Technologie im sensiblen Bereich der medizinischen Bildgebungsverfahren ermöglicht und die Weiterbildung sowie den Know-How Transfer unter den Mitarbeitenden als essentiellen Fahrplanbaustein identifizieren.

Kybernos und Own Your Data entwickelten eine Roadmap für das Projekt Food Hinterland System – dieses soll einen Datenraum für einen Digital Product Passport für den Le-

bensmittelsektor aufbauen, der zentrale Herausforderungen wie Lebensmittelverschwendung, Kreislaufwirtschaft und die effektive Beseitigung von Informationsungleichgewichten zwischen Produzent*innen, Zwischenhändler*innen und Verbraucher*innen adressiert.

Onlim und TU Wien Informatics entwickelten eine Roadmap für die Implementierung von Fairness und Vorurteilsfreiheit in den von Onlim verwendeten Large Language Models, die die technologische Basis ihrer Chatbots bilden.

Vouchercube und AIT Austrian Institute of Technology brachten mit ihrer Roadmap im Bereich der Softwareentwicklung neue Impulse in das Design der E-Commerce-Plattform von Vouchercube für B2C-Kund*innen im Tourismussektor ein: diese soll nun auf digital-humanistischen Werten wie Barrierefreiheit, Datensicherheit, Wahrung der Privatsphäre bei der Datenerfassung oder Fairness der angebotenen Services aufbauen.

Die bee produced GmbH und Human-Computer Interaction Gruppe der TU Wien befassten sich mit Transparenz und Vertrauen im Bereich von B2B-Plattformen. Der Fokus lag auf der Erhebung der Erwartungen von drei Stakeholder*innengruppen (Kund*innen, Produzent*innen, Plattform-Betreiber*innen) in gemeinsamen Workshops. Im Mittelpunkt standen dabei Themen wie Datentransparenz (z. B. die Art der Daten, die bei Bestellprozessen erfasst werden). Die Erkenntnisse werden von bee produced zur Weiterentwicklung ihres Software-Portfolios herangezogen.

Ingenieur Sumetzberger GmbH analysierte in einer Kooperation mit der FH Campus Wien die aktuellen Innovationsprozesse. Gemeinsam wurde eine Möglichkeit gefunden, Themenstellungen rund um den digitalen Humanismus im Kontext von Rohrpost und Robotersystemen für Transportaufgaben im Krankenhaus zu thematisieren, zu diskutieren und in entsprechenden innovativen Produkten einzubringen.

TETRAGON Braille Systems GmbH und die Hilfsgemeinschaft der Blinden und Sehschwachen erarbeiteten einen Plan, um afrikanische Partnerunternehmen für Fertigung, Wartung und Vertrieb von kostenoptimierten Braille-Geräten zu qualifizieren. So soll in den Zielländern ein leistbares Angebot für Menschen mit Sehbehinderung geschaffen werden, das den Zugang zu Bildung und hochwertiger Arbeit erleichtert.

25.5 Die eigene Roadmap entwickeln: Sechs Praxistipps für Unternehmen auf dem Weg zum Digitalen Humanismus

Die vorgestellten Beispiele machen deutlich, dass KMU bei der ersten Implementierung digital-humanistischer Ansätze sowohl bei internen Arbeits-, Entwicklungs- und Organisationsprozessen als auch direkt bei ihren kund*innenorientierten Produkten, Services und Geschäftsmodellen ansetzen. Um ein besseres Verständnis für gemeinsame Learnings, aber auch Herausforderungen zu entwickeln, diskutierten die Wirtschaftsagentur Wien und der WWTF bei einem Co-Creation Workshop im Rahmen der Digital Days 2023 mit den Fördernehmer*innen über ihre Erfahrungswerte zur praktischen Umsetzung des Digitalen Humanismus.

Sechs Tipps, die aus diesem Workshop hervorgegangen sind, bieten Anhaltspunkte für Unternehmer*innen, die sich erstmals mit der Entwicklung einer digital-humanistischen Strategie auseinandersetzen. Sie sollen sowohl eine niederschwellige, praxisorientierte Starthilfe bieten als auch dabei helfen, Einstiegshürden bei der Auseinandersetzung mit wertorientierter, digitaler Unternehmenspraxis abzubauen.

1. Konkrete Use Cases identifizieren

Am Ende einer Roadmap zum Digitalen Humanismus sollte die Menschenzentrierung die Entwicklung einer Organisation in allen Bereichen durchdringen. Es ist aber kontraproduktiv, mit dieser allgemeinen Zielsetzung zu starten.

Es hat sich bewährt, eine konkrete Frage als Ausgangspunkt zu definieren: Wie können wir unsere digitalen Arbeitsprozesse menschenzentriert gestalten? Wie können wir den Menschen bei der Entwicklung digitaler Produkte in den Mittelpunkt stellen? Wie können wir sicherstellen, dass der Digitale Humanismus Teil unserer Unternehmenskultur wird? Wie können wir unser Geschäftsmodell an digital-humanistischen Prinzipien orientieren? So können Potenziale für konkrete Handlungen identifiziert werden.

Diese Koppelung an einen unternehmensrelevanten Use Case hilft dem Projektteam und involvierten Mitarbeitenden bei der Identifikation mit dem Projekt und der Entwicklung eines praxisnahen Verständnisses für den menschenzentrierten Einsatz digitaler Technologie am Beispiel der Produkte, Dienstleistungen und digitalen Arbeitsprozesse, mit denen Sie tagtäglich konfrontiert sind.

2. Kooperation als Katalysator

Für viele Unternehmen und KMU bedeutet die Beschäftigung mit digital-humanistischer Unternehmenspraxis das Betreten von Neuland. Die Zusammenarbeit mit Kooperationspartner*innen, die Expertise im Bereich menschenzentrierter Technologieentwicklung, -anwendung und -implementierung haben, kann gerade bei der Identifizierung von unternehmensrelevanten Use Cases und konkreten, individuellen Ansatzpunkten unterstützen. Neue Perspektiven auf das Vorhaben eröffnen sich auch durch die Einbindung von Sozial- oder Geisteswissenschaftler*innen. Diese Erfahrung machte z. B. die Ing. Sumetzberger GmbH bei ihrer Auseinandersetzung mit der Entwicklung von Rohrpostanlagen durch die Linse des Digitalen Humanismus:

„Die Einbindung einer Professorin für Philosophie war wichtig und hilfreich, um einerseits eine nicht-technische Perspektive auf das Thema zu bekommen und andererseits die philosophischen Grundlagen Ethik, Logik und Ästhetik in die vorerst rein technische Diskussion zu integrieren.“ (Lukas Pinter, Techniker Rohrpostanlagen)

Innerhalb der Stadt Wien, die den Digitalen Humanismus seit 2019 aktiv vorantreibt, gibt es eine große Anzahl von Personen mit Expertise, die KMU während des Umsetzungsprozesses begleiten können. Dabei empfiehlt es sich, diese Kollaborateur*innen bereits von Anfang an in die Projektentwicklung einzubinden.

3. Bewusstseinsbildung & Definition von greifbaren Zielen

Unternehmerisches Handeln nach den Prinzipien des Digitalen Humanismus baut auf gemeinsamen Lernprozessen innerhalb des Unternehmens auf und benötigt die Bereitschaft der Geschäftsführung, in zunächst noch nicht messbare Ziele zu investieren. Förderungen helfen Unternehmen dabei, Freiräume für derartige Investitionen zu schaffen.

Die Definition smarter und klar messbarer Ziele sowie KPIs, die am Ende der Entwicklung einer Roadmap stehen, sind für die Validierung der Strategie innerhalb des Unternehmens ebenfalls hilfreich.

„Soziotechnische, wertebasierte Standards, wie IEEE 7000, sind wichtige Instrumente für die Implementierung eines digital humanistischen Zugangs in der Unternehmenspraxis im Einklang mit kommenden regulatorischen Anforderungen.“ (Dr. Clara Neppel, Senior Director, IEEE European Business Operations)

Orientierung und Anhaltspunkte für die Definition dieser Ziele finden sich in bereits implementierten und in Entwicklung befindlichen Standards und Regulierungsinstrumenten. Neben den eingangs erwähnten, EU-weiten Instrumentarien (u. a. EU AI Act, Digital Services Act, Digital Markets Act, Data Act, Data Governance Act) bietet zum Beispiel der IEEE Standard 7000-2021 auf Basis von Value Based Engineering ein Rahmenwerk für ethisches Systemdesign im allgemeinen und das Design ethischer KI-Systeme im Besonderen.

4. Communication is key

Die Kommunikation der Roadmap-Strategie und die Definition einer gemeinsamen Vision, die sich intern von der Geschäftsführung und dem C-Level über das Projektteam bis hin zu den Mitarbeitenden erschließt, ist essentiell für den Erfolg des Projektes. So schafft man Verständnis auf der Führungsebene und garantiert ein langfristiges Bekenntnis des Unternehmens zur neu entwickelten Strategie. Ebenso wichtig ist, dass Share- und Stakeholder*innen des Unternehmens über die Vision unterrichtet werden und diese teilen.

Die inhaltliche Zugänglichkeit sollte durch unternehmensinterne Übersetzer*innen bzw. Botschafter*innen gewährleistet werden, die das Thema des Digitalen Humanismus besetzen, seine Bedeutung innerhalb des spezifischen Use Case vermitteln und die Theorie des Digitalen Humanismus auf den Boden der Unternehmenspraxis bringen können. Feedback hierzu sollte man sich von Mitarbeitenden außerhalb des Projektteams suchen: Diese können rückmelden, ob sich die entwickelte Roadmap in der Praxis erschließt und angewendet werden kann.

Neben der internen Kommunikation sollte sich auch der externe Kollaborationspartner mit der Roadmap identifizieren.

Investiert man genügend Zeit in die Konsolidierung der gemeinsamen Vision, kann die Kollaboration Rückenwind für neue Innovationsprojekte im Unternehmen mit sich bringen und hilft dabei, Mitarbeitende für das Thema zu begeistern und zu motivieren. Außerdem können sich längerfristige Zusammenarbeit mit neuen Partnerinstitutionen, neue Forschungsfelder, Publikationsmöglichkeiten sowie technologisches Know-How eröffnen.

5. Zeitfenster für die Entwicklung schaffen

Die Entwicklung einer Roadmap für Digitalen Humanismus wird nur dann erfolgreich sein, wenn für die zuvor erwähnten Maßnahmen der Kommunikation und Bewusstseinsbildung wie auch den Prozess der Entwicklung einer neuen, menschenzentrierten Kultur der Digitalisierung genügend zeitliche Ressourcen eingeplant werden. Hierzu zählt beispielsweise die Anpassung der Personalplanung, um das Projektteam vom Tagesgeschäft freizuspüren und Zeitfenster für Workshops und die intensive Auseinandersetzung mit dem Thema zu schaffen. Auch sollten Ressourcen für das Projektmanagement der Zusammenarbeit mit den Kollaborationspartner*innen eingeplant werden.

6. Von der Strategie zum gelebten Digitalen Humanismus

Am Ende der Entwicklung einer Roadmap für Digitalen Humanismus, die das Thema im Unternehmen im Rahmen strategischer Planung etabliert, sollte der erste Schritt zum gelebten Digitalen Humanismus in der Unternehmenspraxis stehen. Dies kann z. B. durch die schrittweise Implementierung konkreter Projekte passieren.

25.6 Digitaler Humanismus als nachhaltige Unternehmensstrategie: Erfahrungswerte von KMU

Einige Unternehmen, die im Rahmen des Fördercalls Roadmaps Digitaler Humanismus Projekte erfolgreich umgesetzt haben, befassen sich bereits mit Möglichkeiten einer digital-humanistischen Unternehmensstrategie. Im Folgenden wird anhand von Kurzinterviews mit Unternehmen aus der technischen Gebäudeausstattung und der Elektronikproduktion veranschaulicht, welche nachhaltige Wirkung und langfristigen Veränderungen durch die Umsetzung von digital-humanistischen Vorhaben diese Unternehmen beobachtet haben.

Use Case 1 – Innovationsprozesse im Kontext von Rohrpostanlagen digital-humanistisch denken (Ing. Sumetzberger GmbH//Lukas Pinter, Techniker Rohrpostsysteme)

Die Ing. Sumetzberger GmbH ist ein unabhängiges Familienunternehmen. Die Tätigkeitsschwerpunkte von Sumetzberger liegen auf nationaler Ebene in der technischen Gebäudeausstattung und auf internationaler Ebene in den Geschäftsbereichen Rohrpostsysteme und Fördertechnik. Das Unternehmen ist heute ein weltweit führender Hersteller von Rohrpostsystemen.

• Digitaler Humanismus und Sumetzberger: Der Anstoß

„Den Anstoß für die Beschäftigung mit Digitalem Humanismus gab ein Newsletter der Wirtschaftsagentur Wien, in dem die Thematik beschrieben wurde. Danach wur-

den erste interne Diskussionen geführt, ob Digitaler Humanismus uns auch betrifft – also Rohrpost für Krankenhäuser, Blutproben und Medikamententransport.

Da wir uns seit vielen Jahren mit der Automatisierung von intralogistischen Transpostprozessen in Spitälern und zum Teil auch der Industrie beschäftigen, war es naheliegend, unsere realisierten Innovationen und unsere zukünftigen Ideen unter dem Aspekt des Digitalen Humanismus zu betrachten und zu bewerten. Als Techniker*in oder Entwickler*in geht man prinzipiell davon aus, dass der entwickelte Prozess bzw. Produkt den funktionalen Zweck erfüllt und alle Beteiligten vollends zufrieden sind. Eben diese Sichtweise unter anderen Aspekten zu hinterleuchten, stand zu Beginn dieses Projektes.“

● Konkrete Schritte bei der Roadmap-Entwicklung

„In Kooperation mit der FH Campus Wien wurden zunächst die aktuellen Innovationsprozesse analysiert. In einem interdisziplinären Umfeld (Projektentwickler*innen, F&E Mechanik/Elektronik, IKT und Philosophie) wurden digital-humanistische Themenstellungen im Kontext von Rohrpostanlagen und Robotersystemen für Transportaufgaben im Krankenhaus diskutiert.

Ein Fragebogen, der im Rahmen eines Workshops einem interdisziplinären Team von Sumetzberger präsentiert wurde, brachte den Mitarbeiter*innen das Thema grundsätzlich näher. Die Erarbeitung der ‚Road of Digitalisation‘ sowie die Auswertung der Fragebögen in Kooperation mit der FH Campus Wien konnten in einem Vorschlag für die prozessuale Implementierung des Digitalen Humanismus zusammengefasst werden.

Erst durch die Schaffung einer gemeinsamen Sichtweise, Diskussions- und Reflexionsgrundlage wurde sichergestellt, dass möglichst viele Aspekte des digitalen Humanismus ihre Berücksichtigung finden. Für die Innovationsprozesse bei Sumetzberger wurde auf diesem Weg eine Checkliste entwickelt, mithilfe derer menschenzentrierte, ethische Technologie, digitale Inklusion und kulturelle Überlegungen sowie die Zusammenarbeit Mensch & Technik abgedeckt und laufend über eine kritische Reflexion betrachtet und bewertet werden.“

● Die Maßnahmen und ihre nachhaltige Wirkung

„Die wesentlichste nachhaltige Wirkung ist eindeutig das geschaffene Bewusstsein des Themas ‚Digitaler Humanismus‘ und der Bezug zu unseren Produkten und Prozessen. Die erarbeiteten Spielregeln für neue Innovationen unter Einbindung eines interdisziplinären Teams und, sofern möglich, Anwender*innen und Stakeholder*innen, sind ein stabiler und wirkungsvoller Mechanismus. Auch der Besuch eines Teams von Sumetzberger im Operationssaal-Innovation Center (OPIC) an der

FH Campus Wien diente der Festigung der Projektergebnisse in der praxisnahen Betrachtung des klinisch-digitalen Patient*innenpfades von der Diagnostik über die Operation bis hin zur intensivmedizinischen Nachsorge.“

- **Erfahrungswerte und Tipps für Unternehmen**

„Zu Beginn stand natürlich die Frage: ‚Was ist das?‘ und ‚Was hat das mit uns zu tun?‘ Hilfreich war die frühzeitige Einbindung von Mitarbeiter*innen aus Technik, Vertrieb, Marketing und Projektentwicklung in die Grundsatzdiskussionen mit dem Team der FH Campus Wien.

Vor allem die Einbindung von Personen mit Expertise der FH (Technik, IT, Philosophie) war ein Erfolgsfaktor für dieses Projekt, da nur so eine neutrale Sicht auf unsere Produkte und Prozesse möglich war. Ein weiterer empfehlenswerter Schritt war der interne Workshop inkl. Fragebogen zu diesem Thema mit fast 60 % des Sumetzberger Rohrpost/Robotik-Teams.

So konnten wertvolle Erfahrungen, Sichtweisen und Aspekte eingeholt und das gesamte Team zu diesem Thema abgeholt werden. Auch die Vorschläge der FH Campus zur Dissemination der Projektergebnisse zur internen und externen Bekanntmachung konnten in die Praxis umgesetzt werden.“

- **Digitaler Humanismus und Sumetzberger: Ein Ausblick in die Zukunft**

„Die Zukunft zeigt die Intensivierung des Bedarfes von Automatisierungslösungen in der Probenlogistik – vor allem in der Individualisierung der Problemlösungen.

Wir sehen uns in fünf Jahren als kompetenter Partner für Systemlösungen der digital gestützten Nutzung von Rohrpostanlagen und Robotersystemen im Gesundheitswesen mit dem Merkmal der digital, humanen, menschenzentrierten Technologie. Vor allem unter dem Aspekt des Fachkräftemangels werden solche Lösungen unter Einbeziehung digitaler Inklusion und Kollaboration die Schlüsselfaktoren zur Erhaltung und Verbesserung der medizinischen Versorgung der Patient*innen.“

Use Case 2 – Ein fairer, transparenter B2B-Marktplatz für Elektronikproduktion (bee produced//Dr. Wilfried Leputschitz, CEO & Mitgründer)

bee produced ist ein in Wien ansässiges Software-Unternehmen, das sich auf die Entwicklung effizienter Lösungen für die lokale Elektronikproduktion konzentriert. Das Unternehmen hat im Vorfeld der Firmengründung viele Jahre in die Forschung

und den Netzwerkaufbau investiert. Die dabei gewonnenen Erkenntnisse und Kontakte fließen in die Produkte ein und kommen heute den Kund*innen zugute.

- **Digitaler Humanismus und bee produced: Der Anstoß**

„Als Anbieter einer Plattform in Europa sehen wir Transparenz und Fairness als wichtige Pfeiler, um Vertrauen zu schaffen. Unser Ziel ist somit eine nachhaltige Datennutzung, die nicht nur gesetzliche Vorgaben erfüllt, sondern die Nutzer*innen begeistert.“

- **Konkrete Schritte bei der Roadmap-Entwicklung**

„Wir haben eine Roadmap erstellt, Workshops zur Datensicherheit durchgeführt und datenschutzfreundliche Features entwickelt. Die Veröffentlichung unserer Erkenntnisse zeigt unser Engagement über das Projekt hinaus.“

- **Die Maßnahmen und ihre nachhaltige Wirkung**

„Durch die Anwendung der Prinzipien des Digitalen Humanismus wird das Vertrauen in unsere Plattform gesteigert. Somit stärken die damit verbundenen transparenten und fairen Praktiken die Bindung zu unseren Kund*innen.“

- **Erfahrungswerte und Tipps für Unternehmen**

„Wichtig ist es, Stakeholder*innen früh einzubinden und Transparenz hinsichtlich Datenverarbeitung und Entscheidungsprozesse zu forcieren, um das Vertrauen zu stärken. Das ‚Warum‘ hinter Plattformprozessen muss klar sein und kann als Asset dienen, um Nutzer*innen anzusprechen.“

- **Digitaler Humanismus und bee produced: Ein Ausblick in die Zukunft**

„Wir sehen die Integration der Prinzipien des Digitalen Humanismus als Kernmerkmal, das uns auch international abhebt. Das ist vor allem wichtig, da wir kontinuierlich unsere Plattform mit neuen KI-basierten Funktionen ausstatten werden und auch hierbei Transparenz und Fairness ermöglichen wollen.“

25.7 Die Wirtschaftsagentur Wien: Bindeglied zu Netzwerken und dem Wissenspool der Stadt Wien

„Für KMU und speziell auch für Startups, bei denen meist eine technologische Innovation im Zentrum steht, ist es wichtig, einen menschenzentrierten Umgang mit digitaler Technologie zu entwickeln, denn dadurch können sie sicherstellen, dass ihre Innovationen direkt auf die Bedürfnisse der Menschen abzielen und einen echten Mehrwert bieten. Dieser Ansatz fördert nicht nur die Akzeptanz ihrer Lösungen am Markt, sondern trägt auch zur Schaffung einer nachhaltigen und ethisch verantwortungsvollen Unternehmenskultur bei.“ (Hannah Wundsam, Co-Managing Director Austrian Startups)

Menschenzentrierte, digitale Innovation ist in Wien Programm: Das Netzwerk von Botschafter*innen in der Stadtverwaltung, den Standortagenturen sowie Unternehmen und Forschungseinrichtungen ist weitreichend. Wie Hannah Wundsam von Austrian Startups anmerkt, unterstützt ein menschenzentrierter Zugang bei der Entwicklung von digitalen Services und Produkten nicht nur bei der Sicherstellung eines Mehrwerts für die Kund*innen, sondern auch langfristig bei der Implementierung einer digital verantwortungsbewussten Unternehmenskultur.

Die Wirtschaftsagentur Wien möchte diesen Ansatz unternehmerischen Tuns vorantreiben und versteht sich in diesem Kontext als Bindeglied für Wiener Unternehmen und Unternehmer*innen. Sie verknüpft diese mit jeglichen Anlaufstellen, Vereinen, Institutionen und Individuen, die menschenzentrierte, digitale Transformationsprozesse in Wien vorantreiben und die bei der Umsetzung von Strategien des Digitalen Humanismus als Sparring Partner*innen, mit praktischem Know-How und durch Förderprogramme unterstützen. Daneben treibt sie den Digitalen Humanismus im Kontext der Wiener Wirtschaft auch mit ihren eigenen Services voran: So können sich Unternehmen und Startups in Beratungen über KMU-spezifische Strategien für die Umsetzung des Digitalen Humanismus informieren, an Fördercalls mit Fokus auf menschenzentrierter, digitaler Technologienentwicklung teilnehmen, oder im Rahmen eines Co-Creation-Workshops der Wirtschaftsagentur gemeinsam mit Personen mit Expertise aus Wirtschaft und Wissenschaft konkrete Use Cases und Herausforderungen bearbeiten, die den Digitalen Humanismus in den Arbeitsprozessen, Geschäftsmodellen oder Produkten des eigenen Unternehmens verankern.

Literatur

- Stadt Wien (2022). *Digitaler Humanismus in Wien*. Wien: Stadt Wien. <https://digital-strategy.ec.europa.eu/en/library/ethics-guidelines-trustworthy-ai> (zuletzt abgerufen am 5. Mai 2025)
- Strassnig, M., et al. (2019). *Akteure, Instrumente und Themen für eine Digital Humanism Initiative in Wien*. Wien: Stadt Wien. <https://repository.fteval.at/id/eprint/734/> (zuletzt abgerufen am 5. Mai 2025)

Kristina Maurer ist Medienwissenschaftlerin und Kulturproduzentin. In der Wirtschaftsagentur Wien ist sie als Technologieexpertin im Team Digitale Technologien der Abteilung Technologie Services tätig. Sie entwickelt Aktivitäten und Programme im Kontext des Digitalen Humanismus und unterstützt Wiener KMU im Bereich der Vernetzungs- und Fördermöglichkeiten. Zuvor war sie Head of European Projects des Ars Electronica Festival Linz und verantwortete als ausführende Produzentin die Neugestaltung des Ars Electronica Centers im Jahr 2019. Ihre Forschungsinteressen umkreisen die menschenzentrierte, wertebasierte Entwicklung digitaler Technologien sowie deren soziopolitischen und gesellschaftlichen Implikationen.

Eva Czernohorszky ist Politikwissenschaftlerin und Expertin für Forschungs- und Innovationspolitik. In der Wirtschaftsagentur Wien leitet sie die Abteilung Technologie Services und unterstützt mit ihrem Team Unternehmen bei ihren Forschungs- und Innovationsaktivitäten. Sie hat die Wiener Wirtschaftsstrategie WIEN 2030 federführend mitgestaltet und treibt die Umsetzung dieser Strategie durch Leitprojekte voran. Darüber engagiert sie sich im Universitätsrat der JKU, im Climate Lab, beim High-tech Inkubator INITs und bei Austrian Standards International für wertebasierte Technologieentwicklung im Sinne eines Digitalen Humanismus und für die Zusammenarbeit in diversen Teams über Organisations- und Disziplinengrenzen hinweg.

Der Wiener Weg zum Digitalen Humanismus

26

Der Beitrag Wiens zu einer wertorientierten Digitalen Transformation – Motivation, Hintergründe und Maßnahmen

Michael Stampfer

Zusammenfassung

Der Beitrag beleuchtet Erfolge und Chancen Wiens in der Verbesserung der Lebensqualität durch die Digitalisierung und entsprechende Policies. Das ist wichtig, denn die digitale Revolution ist tatsächlich eine solche. Sie greift tief in Strukturen ein und fordert unser Verständnis von Freiheit massiv heraus. Unsere Bemühung muss sein, eine Balance zu schaffen zwischen individueller Freiheit und stabilen Institutionen, die soziale Freiheit und Wohlfahrt fördern. Der Wiener Wissenschafts-, Forschungs- und Technologiefonds (WWTF) spielt eine wichtige Rolle bei der Förderung von Spitzenforschung und der Integration menschlicher Werte in der digitalen Transformation. Mit dem Programm Digitaler Humanismus hat der WWTF schon viele Initiativen und Projekte unterstützt.

„Die Digitalisierung hat in den letzten Jahrzehnten immens zur Verbesserung der Lebensqualität auf der ganzen Welt beigetragen. Sie wirkt mehr als nur evolutionär, wir haben es mit einer Umwälzung zu tun.“ (Himpele)

Der Text wurde ohne ML-basierte Unterstützung verfasst. Einige Formulierungen beruhen auf einem WWTF-internen Dokument, der als Impact Story unseren Beitrag zum Digitalen Humanismus skizziert, damit auch Dank an meine Kolleg*innen im WWTF, insbesondere Petra Steinkogler für die Hilfe bei der Vorbereitung.

M. Stampfer (✉)

WWTF, Wien, Österreich

E-Mail: michael.stampfer@wwtf.at

26.1 Warum ein solcher Weg?

Es geht um Freiheit und Struktur in revolutionären Zeiten.

Die Digitalisierung hat in den letzten Jahrzehnten immens zur Verbesserung der Lebensqualität auf der ganzen Welt beigetragen. Sie wirkt mehr als nur evolutionär, wir haben es mit einer Umwälzung zu tun: Wir können über beliebige Entfernungen multimedial in Echtzeit mit Individuen und Gruppen kommunizieren, selbst Akteur*in, Redakteur*in und Produzent*in sein, neue Geschäfts- und Steuerungsmodelle entwickeln und nutzen, die Produktivität mal mehr, mal weniger erhöhen, pausenlos spielen und in virtuelle, fremde Welten eintauchen. Unsere Auswahl-, Informations- und Konsummöglichkeiten sind beinahe endlos geworden. Wir können ganz viele Dinge tun, sagen und gestalten, wie wir sie wollen. Wir können etablierte Ordnungssysteme unterlaufen. Neue Akteur*innen können in sehr kurzer Zeit sehr rasch und global wachsen. Unsere Freiheit hat zugenommen. Hier sind wir schnell beim Begriff der Freiheit als solcher, der auch die Ambivalenz des Fortschreitens zeigt.

Freiheit, sofern sie eine positive ist (also Freiheit „zu“, ... i. d. R. zur Teilhabe), ist entweder eine, die in gegenseitiger Anerkennung ent- und besteht oder eine, die einseitig ausgerufen und konsumiert wird. Letztere scheint mir durch die digitale Revolution in ihrer Dynamik und ihren Wirkungen zwar nicht ursächlich erzeugt, jedoch beschleunigt und vervielfacht. Bestehende institutionelle Bindungen, ob territorial, vertraglich oder gebräuchlich, schwächen sich ab, ohne dass sich stabile neue Strukturen schon etabliert haben. Das erhöht zwar ein subjektives Freiheitsempfinden, zugleich erfährt aber auch das Prekäre und Instabile unserer Existenzen eine Steigerung.

Sofern die Freiheit eine negative ist (also Freiheit „von“, ... keinen Eingriffen ausgesetzt zu sein und im Genuss der klassischen Freiheitsrechte zu sein), genügt es heute nicht mehr, sich vor dem Staat allein zu fürchten. Länger werden die Schatten jener privaten Akteurskonstellationen, die die Kommunikationsplattformen kontrollieren, und eine Freiheit versprechen, deren Gehalt angesichts von Marktmacht, Täuschung oder Vereinsamung mir trügerisch scheint. Wie ein unsterblicher Vampir zieht die libertäre Ideologie der völlig ungebundenen Freiheit (... so wie auch andere Ideologien) neue Kraft aus dem Digitalen, und wer von Ideen einer solidarischen Gesellschaft und des Rechtsstaates durchdrungen ist, der kann das Blut tatsächlich fließen sehen. Die libertäre Ideologie heißt: Ich mache, was ich will, und die Bindung an die Gesellschaft ist mit den negativen Freiheitsrechten schon konsumiert. Das Libertäre hat interessanterweise mit den Frühzeiten der vernetzten Digitalisierung in den 1970er Jahren an der Westküste einen regelrechten Aufschwung genommen (Lepore, 2018, S. 841 ff.; Vardi, 2018).

Die negative Freiheit von Druck und Zwang muss zwar weiterhin die Basis unserer Gesellschaft bleiben und wir sollen sie auch nicht mit Gerechtigkeit oder anderen Werten verwechseln (Berlin, 2022). Angesichts der Macht- und Informationsasymmetrien im digitalen Zeitalter können wir mit ihr allein nicht auskommen. Wir brauchen auch die Erinnerung an die reflexive, wertgebundene Freiheit und an die Verteidigung dessen, was soziale Freiheit genannt wird (Honneth, 2011): Hier geht es um die Verwirklichung von

Freiheit in wechselseitig gebundenen und stabilen Institutionen. Diese Institutionen sind Normen, Gebräuche, Organisationsformen, also der Rechtsstaat, offene Märkte, Diskurs und die Produktion von Wahrheit. Scheint es nur mir so, oder beginnen sie vor unseren Augen zu zerbröseln, auch durch die Form und Dynamik, in der die digitale Revolution voranschreitet? Es ist noch nicht absehbar, wie zukünftig bestehenbleibende, veränderte und neue Institutionen auf eine Weise zusammenspielen, um alle Arten von Freiheit so garantieren zu können, dass es zu einer Steigerung und nicht zu einem Abbau individueller und kollektiver Wohlfahrt kommt. Freiheit und Wohlfahrt stehen für mich in einem engen positiven Verhältnis: Ich will in einem Staat leben, der die Freiheit garantiert und ich will in einem Wohlfahrtsstaat leben, der dadurch vielen Freiheit überhaupt erst ermöglicht.

Warum die Sorgen? Es haben die negativen Seiten der Digitalisierung wie Monopolisierung, Zunahme von Desinformation, Ungleichheiten und der Verlust von Privatsphäre bei gleichzeitiger Privatisierung des Öffentlichen massiv zugenommen. Autor*innen sprechen von „Cloud Empires“ (Lehdonvirta, 2022) außerhalb von Territorien und Jurisdiktionen, von der AI-getriebenen „Coming Wave“ mit disruptiven Folgen (Suleyman, 2023), von der Koevolution von Mensch und Maschine (Lee, 2018) und vom „Überwachungs-kapitalismus“ (Zuboff, 2019). Zugleich können und wollen wir uns der Dynamik nicht entziehen: Wir können nicht *nicht* digitalisieren und wir wollen es auch nicht. Ich lebe in Wien, also im Kaffeehausparadies und nicht in einer der avanciertesten Digitalmetropolen, die einem als erstes einfallen: Dennoch kenne ich niemanden, der oder die nicht intensiv digitale Tools und Prozesse nützt und nicht in Panik verfiel, wenn er oder sie das Leben wieder rein analog organisieren müsste. Ich kenne viele und außerordentlich sympathische Leute, die ihr Geld damit verdienen, die Digitalisierung in Geschäfts-, Regierungs- oder Lebensstilbelangen weiter voranzutreiben und zu optimieren. Wir sind mitten im Digitalen. Die Metapher vom Cyborg einmal beiseitegelassen, allein als „fleischliche“, sprich biologische Menschen sind wir in den letzten 50 Jahren andere geworden – und die meisten von uns finden das überwiegend gar nicht so schlecht.

Das ist wichtig für unser Thema des Digitalen Humanismus: Gestaltung und Entwicklung steht im Vordergrund, nicht Angst oder Sentimentalität. Es steht aber auch der biologische Mensch als Konstante im Vordergrund, und nicht nur als Individuum, sondern in seinen gesellschaftlichen Bindungen. Wir sollen eben nicht in einer Welt aufwachen, in der wir uns dieser Gestaltungsmöglichkeiten begeben haben und in der uns unsere Institutionen hohl und brüchig geworden sind. Meine Sorgen – obwohl ich Innovationsförderer von Beruf bin – sind tatsächlich eher auf der Seite der *Unterregulierung* angesiedelt.

Wir wollen und sollen nicht nur Passagiere sein, auch wenn wir das lange waren, in einem sehr kleinen Land leben und die Stellung Österreichs wie Europas in der digitalen Welt uns nicht gerade den Fahrersitz zuweist. Ein entwickeltes Gesellschaftsmodell haben wir allemal. Dazu kommt der Ehrgeiz, im Digitalen starke Akteur*innen und nicht Getriebene zu sein. Getriebene*r wird man freilich schnell in revolutionären Perioden, und oft sind es Revolutionäre selbst, die besonders intensiv riechen, dass an der Umwälzung etwas grundsätzlich Fäules dran ist (Büchner, 2018). Ein konkreter Weckruf für uns in Wien kam daher nicht von einer Ökonomin, einem Juristen oder einer Poetin, sondern von Tim

Berners-Lee, dem Schöpfer des World Wide Web, der über den Zustand des Internets und darauf blühender Geschäftsmodelle meinte: „The system is failing“ (Berners-Lee, 2017).

26.2 Was also tun? The system should *not* fail

Warum war diese Botschaft für uns so aufrüttelnd? Auf die eine oder andere Weise werden wohl alle in diesem Band versammelten Beiträge auf die Frage eingehen, was uns Sorgen bereitet und wo wir mit dem Digitalen Humanismus neue Wege einschlagen wollen. Für meine Arbeitsstätte, den Wiener Wissenschafts-, Forschungs- und Technologiefonds (WWTF, <https://www.wwtf.at>) möchte ich die Frage so beantworten: Zuerst einmal glauben wir – klar, mit vielen diskursiven Schleifen und, jawohl, mit oft schmerzhaften Ambivalenzen – daran, dass wissensgetriebener Fortschritt nicht nur in uns Menschen ziemlich fest verankert ist, sondern auch überwiegend etwas Positives darstellt. Wir glauben an die Öffnung von Möglichkeiten und an die Gestaltbarkeit der Welt.

Wir tragen materiell und auch ideell in Wien unseren Teil dazu bei und tun das im Rahmen von konkreten Förderprogrammen, die Spitzenforschung ermöglichen und herausragende Talente nach Wien holen. Eines unserer größten Programme heißt „*Information and Communication Technologies*“ und läuft seit 2008. Weit über 70 Mio. € haben wir in etwa 100 Förderungen am Standort für digitale Forschungsprojekte vergeben. Das beinhaltet Forschung zu Themen wie Funknetze, Security, Recommender Systems, Logik und Graphentheorie oder Machine Learning. Viele junge Forscher*innen konnten mit Hilfe der Förderung Schlüsselpublikationen veröffentlichen, wichtige Karriereschritte machen oder mit Unternehmen kooperieren. Unsere Eingriffslegitimation war und ist die beträchtliche Unterfinanzierung der computerwissenschaftlichen Grundlagenforschung in Österreich. Auch die Ergebnisse der großen Zehnjahresevaluierung des WWTF bestätigen die Wirkungen und den Erfolg dieser Interventionslogik, siehe: https://www.wwtf.at/upload/wwtf_impact-eval-2025_panel-report.pdf.

Wir sind also Mitproduzierende bzw. Dienstleister*innen der digitalen Revolution. Zugleich sind wir aber auch Anhänger*innen der repräsentativen Demokratie, der sozialen Marktwirtschaft, einer sozial inklusiven Entwicklung, einer wiederum sozial rückgebundenen Freiheit, der Grund- und Menschenrechte und eines halbwegs stabilen Wahrheitsbegriffs (vom Diskurs über die Produktion bis zur Verbreitung derselben). Damit passen wir gut zu unserem Wirkungsgegenstand Wissenschaft und zu unserem Wirkungsort: Wien war und ist in vielen Politikbereichen ein Ort der Ausbalancierung der sozialen Sicherung und der individuellen Entfaltung: Soziale Rechte und gut ausgebaute öffentliche Dienstleistungen und Daseinsvorsorge sind kein Widerspruch dazu, die Welt mit neuen Ideen und Initiativen zu beglücken, vielleicht sogar neuerlich zu revolutionieren (Cockett, 2024).

Was im Analogen ging und geht – sprich bei Gütern wie Wasser, Gesundheit, Wohnen oder Mobilität – ist auch im Digitalen angesagt: Wiener Ideen und Initiativen für ein besseres Zusammenleben können Kraft entwickeln und für die Welt Impulse aussenden. Es ist auch nichts Neues, dass es große gesellschaftliche, politische und wirtschaftliche

Verwerfungen und Herausforderungen gibt. Heute ist es die Ortlosigkeit der Digitalkonzerne, die sie dem Diskurs, den Regeln, der Sorgfalts- und der Steuerpflicht in einem Land oft entkommen lässt. Es ist die hohe Geschwindigkeit von Innovation in Kommunikationsformen, Geschäftsmodellen und Akteurskonstellationen, die unsere Institutionen ebenfalls zur Weiterentwicklung zwingt bzw. zwingen sollte. Es sind die Größenvorteile und die Trajektorien – sprich die Kumulation von Kapital, Ideen, Marktanteilen und neuen Produkten –, die die wirtschaftliche Basis vor Ort unter Druck setzen. Da sorgen sich die einen um den Wohlstand und die anderen um die Demokratie, und eine gewisse intelligente Schnittmenge zwischen diesen beiden Sorgen-Pools gibt es auch.

26.3 Welche Schritte setzen: Der Hammer ist schon in der Hand

Bei mir selbst ist 2017/18 – also eher spät, aber dafür dann so richtig – aus einer diffusen Sorge ein manifestes Anliegen geworden. Damals habe ich in einer Reihe von Gesprächen mit meinem Freund Harald Katzmaier, den viele durch seine Netzwerkforschung kennen, mehr über die Funktionszusammenhänge gelernt, die zwischen unreguliertem Digitalkapitalismus, Extraktion von Nutzerdaten und umfassenden Beeinflussungsmöglichkeiten bestehen. Dadurch bin ich mit dem Begriff des Digitalen Humanismus als gleichsam zivilisatorischer Notwendigkeit in Berührung gekommen. In die Gänge gebracht hat mich die historisch-gesellschaftlich-politische Dimension der digitalen Revolution, die mich mehr interessiert als ihre technisch-futuristische Seite. Hunderte Lesestunden später kommt dann die Frage: Was ist meine Rolle, wenn ich mich nicht nur wie der abgemagerte Eisbär auf der schmelzenden analogen Eisscholle wiederfinden will? Was kann ich tun, wenn ich meine Geräte und Vernetzungen nicht auf die Müllhalde werfen möchte? Wie könnte ich auch? Was kann ich tun, wenn ich auch kein digitaler *Flagwaver* sein will? Mit wem kann ich welchen Diskurs führen, noch dazu unter guten Voraussetzungen, wo ich doch massenweise herausragende, engagierte und sympathische Informatiker*innen und Geistes- und Sozialwissenschaftler*innen kenne, die doch alle etwas zu sagen haben oder haben sollten?

Die Hammermetapher für naheliegende Arten der Problemlösung darf ich als bekannt voraussetzen und unser Hammer ist die Forschungsförderung. Die zu verbindenden Teile sind die jeweiligen Disziplinen: Die Computerwissenschaften stehen im Zentrum dieser Entwicklungen. Sie tragen einerseits massiv zum wissenschaftlichen Fortschritt bei, gleichzeitig fehlen aber oft die notwendigen Mittel und Voraussetzungen, um die Auswirkungen auf die Gesellschaft zu reflektieren und diese in Forschung und Technologieentwicklung zu berücksichtigen. Während die Technikfolgenabschätzung und die Regulierung brauchbare Instrumente für nachträgliche Eingriffe auf dem Markt bieten, sind Möglichkeiten rar, menschliche und gesellschaftliche Perspektiven in die Forschung selbst zu integrieren, um digitale Technologien so zu gestalten, dass sie von vornherein mit menschlichen Werten in Einklang stehen. Die übliche Art von Forschungsförderung, die Wissensproduktion in so genannten epistemischen communities (Glaeser et al., 2018),

die Publikationsmöglichkeiten und damit Karrierechancen von Forscher*innen sind stark von der Nähe zu einer etablierten Disziplin geprägt. Die Disziplin prägt wiederum das Mindset. In einem Anfall von Respektlosigkeit haben wir im WWTF bei der Vorbereitung unserer Intervention gesagt: Wir tun was gegen die, oft durch ihre Art des Denkens oder durch mangelnde Breite ihrer Ausbildung bei den Computerwissenschaften aufzufindende, jedenfalls vermutete *Naivität*.

Wichtige Perspektiven können von Disziplinen aus dem Bereich der Sozial- und Geisteswissenschaften (GSK) geboten werden. Die Initiative Digitaler Humanismus ist der Versuch des WWTF, diese Perspektiven in die Computerwissenschaften/Informatikforschung zu integrieren und zusätzlich die Zivilgesellschaft und politische Akteure für Digitalisierungstechnologien und -praktiken zu erreichen, die sich an menschlichen und gesellschaftlichen Bedürfnissen orientieren. Der Fächer der wissenschaftlichen Fächer spannt sich vom Recht über die Ökonomie, Soziologie, Politologie, Psychologie bis hin zur Philosophie, mit menschlichen Grundfragen: Wer sind wir, was folgt daraus, und wie organisieren wir uns bei geringstmöglichem Leid? Hier hieß die respektlose Aufgabe: Wie bekämpfen wir die Gebundenheit der GSK-Fächer an die analoge Welt, was tun gegen eine solche vermutete *Ignoranz*?

Das Ergebnis war wenig überraschend der Vorschlag, die beiden Gebiete a priori durch Finanzierungsmöglichkeiten zusammenzubringen. Finanzierung bedeutet hier Freiheit, mit gesicherten Mitteln neue Wege zu beschreiten, und ist in dem Fall höher zu werten als die Freiheit, dass die Forscher*innen bottom-up alleine bestimmen, was das Thema ist. Die Einbeziehung der menschlichen Perspektive und der Interdisziplinarität in Form des Digitalen Humanismus war ein entscheidendes neues Merkmal des Fördercalls 2020, der innerhalb unseres IKT-Programms ausgeschrieben wurde.

Vorab gab es viele wichtige Gespräche und Diskussionen, und eine, die entscheidend war. Zuerst zu den vielen:

Wir haben erstens nicht unbeträchtliche und auch begründete Kritik für den Begriff „Digitaler Humanismus“ bekommen: Es gab -zig Humanismen in der Menschheitsgeschichte und keiner war gänzlich koscher; es gibt schon die Digital Humanities; es gab und gibt philosophische und semantische Einwände. Glauben Sie mir, wir haben viele Alternativen erwogen, keine überzeugendere gefunden oder serviert bekommen. Vielmehr sind wir froh, dass es seit einigen Jahren den Begriff Digitaler Humanismus gibt, mit dem mehr und mehr Menschen aus den verschiedensten Herkunft und Hintergründen arbeiten. Julian Nida-Rümelin und Nathalie Weidenfeld haben mit ihrem gleichnamigen Buch geholfen (Nida-Rümelin & Weidenfeld, 2018).

Zweitens haben wir sehr viel mit Computerwissenschaftler*innen interagiert. Über große Strecken basiert der Digitale Humanismus auf Forschungsplänen und auf der Alarmiertheit dieser Community. Der Bedarf war da, Forschungsfragen aus den Computerwissenschaften auf eine Art und Weise nachzugehen, die bereits von Anfang an die menschlichen, sozialen und politischen Aspekte dieser digitalen Technologien gleichberechtigt berücksichtigt, in der Hoffnung, dass negative Auswirkungen auf die Gesellschaft da-

durch vermieden werden können. Also, da war doch viel weniger *Naivität* als anfangs angenommen.

Drittens ging es um die Bedürfnisse der Geistes- und Sozialwissenschaften, was auch einen wichtigen Inhalt einer umfassenden Bedarfserhebung (Strassnig et al., 2019) ausmachte. Dabei haben wir auch neuerlich festgestellt, dass finanzielle Mittel für die interdisziplinäre Zusammenarbeit zwischen Computerwissenschaften und GSK in Wien, in Österreich und international dünn gesät waren. Fachbereiche und Universitäten sind institutionell getrennt und haben wenig Anreize, disziplinäre und institutionelle Grenzen zu überbrücken. Freilich ist so viel da: Unter anderem gibt es an zwei Universitäten Informatikfakultäten und große GSK-Fakultäten und Departments an mehreren Universitäten und außeruniversitären Einrichtungen – ideale Voraussetzungen für höchst relevante und neuartige Kombinationen der Forschungsbereiche. Wir haben dort sofort höchste Aufmerksamkeit und Bereitschaft vorgefunden, die These von der *Ignoranz* war also ebenfalls falsch.

Viertens ist unseren Wanderungen durch die Forschungslandschaft schließlich die COVID-19 Pandemie in die Quere gekommen, und irgendwie doch nicht. Sie hat uns nicht aufgehalten und sie hat gezeigt, wie wichtig die durchdachte Ausgestaltung der Digitalen Welt ist.

Das sind die Gründe, warum eine Digitaler Humanismus-Initiative sehr vielversprechend schien: es bot sich eine Mission für uns als kleiner Nischenplayer im größeren Förderumfeld und eine Chance für den Forschungsstandort Wien, sich mit einer sehr spezifischen Ausrichtung in einer internationalen Wettbewerbslandschaft zu positionieren. Der vor einigen Jahren erhobene Anspruch der Stadt Wien, ein digitaler Top-Standort in Europa zu werden (Stadt Wien, 2024), konnte so auf spezifische Weise mit Leben gefüllt werden.

26.4 Was wird erforscht: Konzepte und Praktiken erdenken und erproben

Wenn der WWTF als Institution eine programmatische Initiative macht, bedeutet das die Widmung von relativ viel Geld über einen langen Zeitraum. Wenn der WWTF als kleiner, regionaler Wissenschaftsförderer eine Mission (Schot & Steinmueller, 2018) verfolgt und aktiv etwas Neues ohne Scheu vor großen Begriffen anschieben will, dann sollte er auch wirklich wissen, was er tut. Lesen, Reden, Analysieren und Zusammentragen war eine wichtige Voraussetzung.

Dann gab es das entscheidende Gespräch: Im November 2018 hatten wir die Möglichkeit, die Idee einer derartigen Initiative samt Förderprogramm den Professor*innen der Informatikfakultät der TU Wien und den prominenten Mitgliedern ihres internationalen wissenschaftlichen Beirats vorstellen. Beim Hingehen habe ich mir gedacht: Entweder sie zerlegen diese Idee in zehn Minuten, oder es wird etwas Schönes daraus.

Es wurde etwas Schönes daraus: Die Fakultät mit ihrem Dekan Hannes Werthner wurde sofort zu einem zentralen Träger der Initiative, und begann zusammen mit internationalen Expert*innen am nächsten Tag mit der Vorbereitung eines internationalen Workshops in Wien. Im April 2019 trafen sich über hundert nationale und internationale Teilnehmer*innen. Das Ergebnis war das „*Vienna Manifesto on Digital Humanism*“ (Werthner et al., 2019) ein Plädoyer für eine menschenzentrierte, inklusive Sichtweise auf digitale Technologien und ihre Rolle in der und auf die Gesellschaft. Das Manifest, mittlerweile in mehreren Sprachen erschienen und von sehr vielen Entscheidungsträger*innen unterzeichnet, weist auf die Gestaltbarkeit und Gestaltungsbedürftigkeit der digitalen Welt hin, die gerade eben nicht deterministisch ist. Wie sie gestaltet wird, hängt von uns allen ab. Die rasch skalierenden globalen Unternehmen haben zumeist die besseren Karten, aber welches Spiel gespielt wird, sollen und wollen wir – in Europa – mitbestimmen und mitgestalten: „*Wir müssen Technologien nach menschlichen Werten und Bedürfnissen formen, anstatt nur zuzulassen, dass Technologien Menschen formen*“, so das Manifest. Dabei geht es um zentrale Felder wie Demokratie, Inklusion, Privatsphäre, Marktregulierung und um eine Stärkung von Forschung und Bildung, deren Repräsentant*innen eine Verantwortung zukommt, nicht in Silos und Themen der Vergangenheit zu verharren.

Kurz darauf begann die Fördertätigkeit. Zuerst machte die Stadt Wien 2019 einen kleineren Pilot-Call (TU Wien & CAIML, 2019), der bereits das breite Interesse für interdisziplinäre Gestaltung der digitalen Welt zeigte. 2020 kam die erste größere Ausschreibung des WWTF zum Digitalen Humanismus. Die Pandemie brachte uns dazu, die notwendige Arbeit des wechselseitigen Kennenlernens in eine digitale „Matchmaking-Plattform“ zum Kennenlernen anderer Disziplinen zu verlegen. Davon machten hunderte Forscher*innen Gebrauch und in der Folge erhielten wir knapp hundert Förderanträge, von denen wir neun mit rund 3,6 Mio. € auch unterstützen konnten. Diese sind zum Zeitpunkt des Erscheinens dieser Zeilen bereits größtenteils abgeschlossen und über die geleistete Arbeit werde ich gleich berichten.

Der nächste Schritt war die Finanzierung für „Digital Humanism Roadmaps“ im Jahr 2022. Das sind kleinere Förderungen dafür, wie Organisationen (akademische Einrichtungen und Firmen) Prinzipien des Digitalen Humanismus in ihre Strategien implementieren können. Dies war die erste gemeinsame Förderausschreibung der Wirtschaftsagentur Wien und des WWTF, die u. a. zu sechs vom WWTF geförderten Projekten führte. Dieser Schritt war wichtig, um die Brücke zur Praxis stärker zu schlagen und um die Grundgedanken des Digitalen Humanismus auch auf die institutionellen Führungsebenen zu tragen.

Im Jahr 2023 veröffentlichten wir eine zweite Förderausschreibung für größere Projekte zum Digitalen Humanismus. Wir erhielten 60 Einreichungen, von denen im Juni 2024 sechs zur Förderung ausgewählt wurden, wiederum mit 3,6 Mio. €. Alle diese Förderungen werden durch Vernetzungs- und Feedback-Workshops unterstützt. In einem solchen entstand die Forderung, dass die Doktoratsausbildung stärker in den Fokus rücken soll.

Die jüngste Initiative des WWTF betrifft daher ein Doktoratskolleg, welches sich mit TU Wien, Universität Wien und WU Wien über drei Universitäten spannt und neun Fakultäten bzw. Departments umfasst. Damit können etwa 20 PhD-Stellen ermöglicht werden,

mit hohen Eigenmitteln der beteiligten Universitäten und mit ca. 1,8 Mio. € zusätzlicher Förderung durch den WWTF. Von dieser interdisziplinären Klammer über die jeweils disziplinären PhD-Arbeiten erwarten wir uns einen wichtigen Schub für die Initiative. In einem nächsten Schritt sollten auch Module zu entwickeln sein, die für Praktiker*innen stärker anschlussfähig sind und auch mehr Praxisbezug und Erfahrungen in Forschung und Ausbildung einspeisen.

Die Förderkriterien waren und sind neben wissenschaftlicher Exzellenz natürlich Interdisziplinarität, Relevanz und die Befassung mit Forschungsfragen, die sich um die menschen- und gesellschaftsrelevante Bearbeitung der Digitalisierung drehen. Ausgewählt werden die Anträge von internationalen Jurys: In diesen muss ebenso wie bei der Antragserstellung selbst im Detail aus verschiedenen disziplinären Perspektiven Qualität, Passfähigkeit und Tiefe der Anträge diskutiert und beurteilt werden. Diese Aufgaben meistern unsere ebenfalls interdisziplinären Jurys sehr gut.

Wir verstehen den Digitalen Humanismus nicht nur als Prinzip für interdisziplinäre Forschung, sondern auch als Mittel zur Gestaltung der Digitalisierungspolitik, insbesondere auf regionaler und nationaler Ebene. Um das voranzutreiben, haben wir das Konzept in zahlreichen Konferenzen, Vorträgen und Workshops sowohl regional als auch national und international vorgestellt und zur Diskussion gebracht. Das reicht von Arbeiten mit städtischen Einrichtungen, Ministerien, UN-Organisationen oder EU-Programmen bis hin zu großen Veranstaltungen wie den Digital Days und der Digital Humanism Conference. Besonders wichtig ist, dass die Stadt Wien politisch und operativ das Thema massiv aufgenommen hat.

26.5 Was können wir lernen: Die ersten Projekte gehen in die Umsetzung

Zahlreiche Projekte sind bereits abgeschlossen oder gehen in die letzte Phase. Praktiker*innen fragen sich vielleicht zu Recht: so what? Welche Ergebnisse können berichtet werden?

Die am raschesten verwirklichten Vorhaben waren die sog. Digital Humanism Roadmaps (<https://www.wwtf.at/funding/programmes/ei/#RO22>), wo es um Strategiekonzepte ging: Wie können bessere Schritte unternommen werden, damit die Ideen des Digitalen Humanismus eben die ganze Organisation affizieren? Besonders beeindruckt hat mich etwa das Vorhaben des COMET Zentrums VRVis, einer Forschungseinrichtung für Virtual Reality und Visualisierung mit zahlreichen Unternehmenspartnern. Das VRVis-Management hat alle Prozesse und Initiativen des Zentrums darauf überprüft, wo ethische Fragestellungen in der Visualisierung – oft unter Berücksichtigung von KI-Einsatz – auftauchen und wie Forschung und Transfer gestaltet werden können. Durch den starken Unternehmensfokus des VRVis ist ein Spillover auf zahlreiche Firmen zu erwarten. In ähnlicher Weise hat der Complexity Science Hub (CSH) sich eine Organisationsstrategie gegeben. Hier ist ebenfalls durch die zahlreichen Forschungs- und Praxispartner des Hubs

eine rasche Diffusion zu erwarten, sowohl thematisch als auch methodisch. Ein konkretes Ergebnis war die Erstellung eines Konzepts für den verantwortungsvollen Einsatz von KI als komplexes soziotechnisches System und die Möglichkeiten, mit Komplexitätsforschung AI-Risiken und den AI-Act der EU besser analysieren zu können (Kondor et al., 2024).

Ein drittes Vorhaben erlaubt es den Expert*innen des Österreichischen Instituts für angewandte Telekommunikation, ihre Beratungstätigkeit für Anwender*innen in Feldern wie „Konsumentenschutz und Digitalisierung“ oder „Cybercrime“ im Sinne des Digitalen Humanismus zu verbessern und die Angebote zu verbreitern. Stellvertretend für mehrere Vorhaben aus dem Bereich künstlerischer oder kuratorischer Forschung steht die Universität für Angewandte Kunst, die u. a. mit dem Wiener Bildungsserver digitale Spielgaben entwickeln und ausrollen. Damit können Kinder im Elementar- und Primärbereich Wissen über digitale Prozesse spielend erwerben.

Die ersten großen Forschungsprojekte (<https://www.wwtf.at/funding/programmes/ict/index.php?lang=EN#ICT20>), die 2020/21 begonnen haben, haben bereits zahlreiche Ergebnisse vorzuweisen. Die meisten Vorhaben befassen sich mit Fragen von Online-Kommunikation und -Hass, um Bildungsthemen, um Forschung zu Datenkommunikation bzw. -governance, oder um Fragen, wie Arbeits- und Pflegeprozesse optimiert werden sollten – oder eben nicht.

In der Onlinekommunikationswelt versucht ein Vorhaben von Uni Wien und TU Wien (TACo) Methoden der automatisierten Content-Moderation so weiterzuentwickeln, dass ein deutlich höheres Maß an Transparenz und Qualität erzielt werden kann. Parallel dazu wurden und werden die Wünsche der Nutzer*innen sehr ausführlich erhoben. Derzeit erfolgt die Entwicklung neuer automatisierter Modelle zur Bestimmung toxischer Inhalte. Ein weiteres Projekt der Uni Wien und der FH St. Pölten (Hate) entwickelt computergestützte Methoden der digitalen Gegenrede für Jugendliche, um Zivilcourage zu fördern. Dabei sind deutlich effizientere und effektivere Tools entwickelt worden als bisher verfügbar waren. Der nächste Schritt geht in Richtung Erhöhung der Automatisierung der digitalen Gegenrede. Ein Projekt des schon erwähnten CSH zusammen mit der Uni Wien behandelt das Phänomen von *Emotional Misinformation* und den Vertrauensverlust durch Miss- und Desinformation. Hier ist eine Erkenntnis, dass die landläufig angenommene Verbreitung von Fehlinformation durch Emotionalisierung deutlich schwächer ist als bislang gedacht. Ein Deep Learning Modell ist in Entwicklung, welches die Zusammenhänge von Emotionalisierung, vorbestandene Annahmen und Fehlinformation noch besser analysieren kann. Ein anderes Vorhaben des CSH und des ZSI (*Humanized Algorithms*) hat Methoden entwickelt, um algorithmischen Bias zu verringern und um Minderheiten fairer zu behandeln. Projekte wie diese können Medienbetreibern, aber auch anderen kommunikationsintensiven Unternehmen und Institutionen helfen, ihre Strategien und Tools stärker

darauf auszurichten, Werte wie Demokratie, Faktenbasiertheit, zivilisierte Diskurse oder Bildungsaufträge besser zu erfüllen.

Ein Vorhaben von Uni Wien, TU Wien und AIT (*ShapeTech*) versucht dem Phänomen der oft nicht ganz freiwilligen Selbstoptimierung und „*motivierten Selbstgefährdung*“ in hochdigitalisierten Arbeitswelten auf den Grund zu gehen. Dabei zeigen sich unterschiedliche Formen von digital erhobenem Techno-Stress je nach Qualifikationslevel und Freiheitsgrad der Arbeit. Ein Folgeprojekt mit der Allgemeinen Unfallversicherungsanstalt (AUVA) versucht die Ergebnisse zu vertiefen und vor allem in die Breite zu bringen, um spezifische Gesundheitsgefährdungen durch stark digitalisierte Arbeitsprozesse zu verringern. Ein weiteres Projekt von TU Wien, VICESSE Research und der Karl Landsteiner Privatuni mit dem Namen *Algorithmic Governance of Care* geht der Frage nach, wo und wie KI-basierte Unterstützungsleistungen in der Langzeitpflege sinnvoll und hilfreich sind. Dabei zeigt sich ein vielschichtiges Bild mit zahlreichen algorithmischen Stolpersteinen für Pflegepersonal und Patient*innen, aber auch Chancen durch den stärkeren Einsatz von Explainable AI Tools. Projekte wie diese können Unternehmen verschiedener Branchen dabei helfen, das langfristige Wohlbefinden ihrer Belegschaft zu fördern, ohne in der Digitalisierung auf die Bremse zu steigen.

Bei den Projekten zu Datengovernance und -darstellung geht es in einem Forschungsvorhaben der Uni Wien mit dem Titel *Interpretability and Explainability as Drivers to Democracy* um die Entwicklung von erklärbaren algorithmischen Modellen, um deren Rolle in Entscheidungsprozessen offenzulegen und die Kommunikation in der Öffentlichkeit zu verbessern. Dabei zeigt sich, dass die Kontextualisierung und Erklärung von Daten(-grundlagen) wichtiger ist als XAI-Modelle. Derartige Erklärungsansätze wurden im Projekt entwickelt. Ähnlich analysiert ein weiteres Vorhaben der Uni Wien (*Talking Charts*) Grundlagen und Ausformungen der Datenvisualisierung, um den breiten Bogen von Unverständlichkeit, Missverständnissen und Biases besser zu verstehen und auf dieser Basis mithilfe neuer Testverfahren Datenvisualisierung besser und verständlicher zu gestalten. Das Projekt des *Digital Well-Being Index for Vienna* hat ein innovatives Scoreboard für die Messung der Lebensqualität entwickelt, das sich stark auf Wertungen und Emotionen aus benutzergenerierten digitalen Inhalten stützt. Auch diese Forschungsergebnisse sind für zahlreiche Unternehmen und ihre interne wie externe datenbasierte Kommunikation von Interesse.

Alle Forschungsarbeiten und ebenso die 2023 gestarteten neuen Projekte sind genuin interdisziplinär. Die 2023-er Projekte reichen in ihrem Spektrum von Social Media als Instrument in autoritären Systemen über partizipatorische digitale Governancemodelle oder Vertrauensaufbau in der Digitalität bis hin zu juristisch-philosophischen Normen-Modellen für AI (<https://www.wwtf.at/funding/programmes/ict/index.php?lang=EN#ICT23>).

Die Forschenden sind sehr aktiv in der Kommunikation der Ergebnisse und bereit, mit Institutionen aller Art zur Vertiefung und Verbreitung der Arbeit in den Austausch zu treten. Der WWTF ist gerne behilflich dabei.

26.6 Wohin: Die nächsten Schritte

Digitaler Humanismus ist ein langfristiges Förderprogramm des WWTF. Wir möchten dabei praxisrelevante Tools ebenso unterstützen wie Bausteine für eine Gesellschaftstheorie des Digitalen bzw. Governance-Modelle für ein Leben in derselben. Der nächste große Fördercall folgt daher mit Anfang 2026.

Neben zahlreichen Vernetzungs- und Verbreitungsaktivitäten steht ein großes Event in Wien im Vordergrund: Vom 26.–28. Mai 2025 fand die erste Wiener Digital Humanism Conference statt, finanziert v. a. von der Stadt Wien und stark unterstützt vom WWTF. Dort haben Expert*innen aus Wissenschaft, Wirtschaft, Verwaltung, Politik und Zivilgesellschaft über die Herausforderungen diskutiert, die sich aus der AI-getriebenen Vertiefung und Beschleunigung der digitalen Revolution ergeben. Die nächste Digital Humanism Conference ist in Vorbereitung ebenso das Wiener Zentrum für Digitalen Humanismus als zentraler Hub.

Mit all diesen Aktivitäten wollen wir unseren Beitrag dazu leisten, dass die Digitalisierung stark einem Freiheitsbegriff folgt, der eine Gesellschaft mit starken Bindungen ermöglicht und einen sozialen Zusammenhalt fördert, in dem es große Freiheitsräume gibt.

Literatur

- Berlin, I. (2022). *Freiheit. Vier Versuche* (2. Aufl.). Frankfurt a.M.: S. Fischer.
- Berners-Lee, T. (2017). Tim Berners-Lee on the future of the web: 'The system is failing'. The Guardian. <https://www.theguardian.com/technology/2017/nov/15/tim-berners-lee-world-wide-web-net-neutrality> (Erstellt: 16. Nov. 2017).
- Büchner, G. (2018). *Dantons Tod*. Leipzig: Reclam. 1835
- Cockett, R. (2024). *Stadt der Ideen. Als Wien die moderne Welt erfand*. Wien: Molden.
- Glaeser, J., Laudel, G., Grieser, C., & Meyer, U. (2018). Scientific fields as epistemic regimes: new opportunities for comparative science studies. GESIS SSOAR open access repository. https://www.ssoar.info/ssoar/bitstream/handle/document/60196/ssoar-2018-glaser_et_al-Scientific_fields_as_epistemic_regimes.pdf
- Honneth, A. (2011). *Das Recht der Freiheit. Grundriss einer demokratischen Sittlichkeit*. Frankfurt a.M.: Suhrkamp.
- Kondor, D., et al. (2024). Komplexe Systeme zur Bewertung von Risiken in der Künstlichen Intelligenz. *Philosophical Transactions der Royal Society A*, 382(2285), 20240109. <https://doi.org/10.1098/rsta.2024.0109> (zuletzt abgerufen am 5. Juni 2025).
- Lee, E. (2018). *Plato and the Nerd. The Creative Partnership of Humans and Technology*. Cambridge: MIT Press.
- Lehdonvirta, V. (2022). *Cloud Empires: How Digital Platforms are Overtaking the State and How we Can Regain Control*. Cambridge: MIT Press.
- Lepore, J. (2018). *Diese Wahrheiten. Geschichte der Vereinigten Staaten von Amerika*. München: C.H. Beck.
- Nida-Rümelin, J., & Weidenfeld, N. (2018). *Digitaler Humanismus. Eine Ethik für das Zeitalter der Künstlichen Intelligenz*. München: Piper.

- Schot, J., & Steinmueller, W. E. (2018). Three frames for innovation policy: R&D, systems of innovation and transformative change. *Research Policy*, 47(9), 1554–1567.
- Stadt Wien (2024). Digitale Agenda der Stadt Wien 2030. <https://www.wien.gv.at/spezial/digitale-agenda/>
- Strassnig, M., Mayer, K., Stampfer, M., & Zingerle, S. (2019). Akteure, Instrumente und Themen für eine Digital Humanism Initiative in Wien. <https://www.wwtf.at/upload/digital-humanism-wien.pdf>
- Suleyman, M. (2023). *The Coming Wave: Technology, Power and the Twenty-first Century's Greatest Dilemma*. London: Bodley Head.
- TU Wien, & CAIML (2019). *Vienna Workshop on Digital Humanism*. 4.–5. April 2019. Wien: TU Wien, CAIML. <https://caiml.org/dighum/workshops/vienna-workshop-on-digital-humanism-2019-04-04/> (zuletzt abgerufen am 5. Juni 2025)
- Vardi, M. (2018). How the Hippies destroyed the Internet. In Communications of the ACM. <https://cacm.acm.org/opinion/how-the-hippies-destroyed-the-internet/> (Erstellt: 1. Juli 2018).
- Werthner, H., Lee, E. A., Akkermans, H., et al. (2019). *Vienna Manifesto on Digital Humanism*. Wien: CAIML/TU Wien. <https://caiml.org/dighum/dighum-manifesto/> (zuletzt abgerufen am 5. Juni 2025)
- Zuboff, S. (2019). *The Age of Surveillance Capitalism: The Fight for a Human Future at the New Frontier of Power*. London: Profile Books.

Dr. Michael Stampfer ist Geschäftsführer des Wiener Wissenschafts- und Technologiefonds (WWTF), einer gemeinnützigen Förderorganisation für wissenschaftliche Forschung in Wien. Der WWTF fördert große Projekte in Wien in Bereichen wie Life Sciences, ICT, Kognitionswissenschaften oder Umweltsystemforschung und holt Top-Talente aus aller Welt für Spitzenkarrieren nach Wien. Michael ist Absolvent der rechtswissenschaftlichen Fakultät der Universität Wien und verfügt über langjährige Erfahrung im Bereich der österreichischen und internationalen Forschungs- und Technologiepolitik, mit zahlreichen Organfunktionen, internationalen Engagements und Publikationen in diesen Feldern.