

Alexander Huber

**Umfassender Einsatz von KI in
Unternehmen: Spannungsfeld zwischen
Optimierung und Verantwortung**

03/2024

Schriften aus dem Fachbereich I
**Wirtschafts- und
Gesellschaftswissenschaften**

Schriften aus dem Fachbereich I
Wirtschafts- und Gesellschaftswissenschaften

Alexander Huber
Annette Pattloch
Matthias Schmidt
(Hrsg.)

Bericht Nr. 03/2024

**Umfassender Einsatz von KI in Unternehmen:
Spannungsfeld zwischen Optimierung und
Verantwortung**

Alexander Huber

Lizenz



Diese Schrift (z. B. Bericht, Sammelband, Monografie) wird unter Creative Commons Namensnennung-Weitergabe unter gleichen Bedingungen 4.0 International lizenziert, sofern nichts anderes angegeben ist.

Berlin, April 2024
www.bht-berlin.de/i

Berliner Hochschule für Technik
Studiere Zukunft

Inhaltsverzeichnis

Inhaltsverzeichnis	1
1 Einführung	2
2 Künstliche Intelligenz	5
2.1 Ausprägungen.....	6
2.2 Autonomie	7
2.3 Erklärbare KI.....	8
2.4 Abdeckung.....	9
2.5 Chancen, Risiken und Kritikalität	11
3 KI in der Unternehmensführung	15
3.1 KI in der Unternehmensführung bestehender Unternehmen.....	17
3.2 Unternehmensführung in nativen KI-Unternehmen.....	22
4 Verantwortung bei Nutzung KI in der Unternehmensführung	27
4.1 Einsatzspezifische Verantwortungskomponenten	27
4.2 Exemplarische Verantwortungskomponenten nach Abdeckungsgrad	27
4.2.1 Abdeckungsgrad „Support“ und höher	27
4.2.2 Abdeckungsgrade „Run“, „Cover“ und höher	31
4.2.3 Abdeckungsgrad „As“	33
5 Fazit / Zusammenfassung	35
6 Literaturverzeichnis	36

1 Einführung

Die Digitalisierung hielt bereits vor vielen Jahren Einzug in die Praxis. Zunächst mit der digitalen Abbildung bisher analoger Daten, über deren Verarbeitung bis zu Prozessen, die unstrukturierte, sich ändernde und fehlerbehaftete Massendaten für Maschinen verständlich machen [1]. Trotz steigender Komplexität lag den Verarbeitungsvorgängen in der Praxis eine *regelbasierte* Ausführung zugrunde. Das heißt: *Menschen* entscheiden sich für eindeutige Regeln bzw. Wenn-Dann-Beziehungen und programmieren diese. Insofern galt zumindest in der Praxis lange die in den 1960er Jahren aufgestellte Behauptung von Peter Drucker: „Der Computer trifft keine Entscheidungen, er führt nur Befehle aus. Er ist ein totaler Idiot“ [2].

Seit neuerem schreiben Menschen Computerprogramme und Algorithmen in einer Art, dass diese *selbstständig lernend* in der Lage sind, *eigene Regeln aufzustellen und selbst Entscheidungen zu fällen*. Diese Regeln werden nicht vom Menschen kodifiziert, sondern von der IT-Anwendung erlernt. Die IT-Anwendung entscheidet sich also nicht notwendigerweise in einer für den Menschen vorhersehbaren Art und Weise.

Beispielsweise bestellen IT-Anwendungen schon lange „selbstständig“ Material bei Lieferanten – wenn die programmierten Regeln dies fordern: Unterschreitet ein Materialbestand eine festgelegte Grenze, wird automatisch bestellt. Die IT-Anwendung berechnet, wann, wieviel, bei wem, zu welchen Kosten bestellt wird. Trotzdem ist die IT-Anwendung hier nicht der Entscheider (sondern ein „totaler Idiot“ [2]), denn: Auch wenn der Mensch vom einzelnen Bestellvorgang nichts mitbekommt, hat er doch alle entscheidenden Wenn-Dann-Beziehungen vorher programmiert und trägt somit unzweifelhaft die volle Verantwortung für die korrekte Ausführung der Bestellungen.

Blicken wir ein wenig in die Zukunft: An einem warmen Herbsttag steht ein Lieferant völlig unerwartet mit dem 10-fachen der sonst üblichen Bestellmenge von Sensoren vor der Tür eines fiktiven technischen Geräteherstellers. Der Mann im Wareneingang ruft verwundert seine Chefin an, und diese den Zulieferer. Alle rätseln, warum diese übertrieben hohe Menge geliefert wurde. Aber: die IT-Anwendung hat rechtsgültig beim Lieferanten bestellt und dieser nimmt die Sensoren nicht zurück.

Es bleibt völlig unklar, warum diese hohe Menge bestellt wurde, bis drei Wochen später im monatlichen Purchase-Council des Unternehmens vor weltweit auftretenden Lieferengpässen bei Sensoren und den damit verbundenen Preiserhöhungen gewarnt wird. Im Nachgang zeigt sich, dass am Tag vor der Bestellung mehrere Analysten Warnungen vor weltweit auftretenden Lieferengpässen bei Sensoren ausgesprochen hatten. Weitere Analysen zeigen ergänzende Faktoren auf, die die Entscheidung der IT-

Anwendung beeinflussten: Das Demand-Forecasting zeigte, dass sich die Nachfrage nach Produkten, in denen der Sensor verbaut wird, stark erhöht. Gleichzeitig ist der Krankenstand im Unternehmen unerwartet niedrig, sodass mehr Menschen als sonst um diese Jahreszeit in der Produktion verfügbar sind. Eine gute Stimmung, und damit Produktivität, unter der Belegschaft lässt sich auch aufgrund der von der IT-Anwendung ausgewerteten LinkedIn-Einträge und -Vernetzungen der Mitarbeiter sowie deren Beiträge in ihren öffentlich zugänglichen Instagram-Profilen vermuten. Wie auch immer. Am Ende hat das IT-System die optimale Menge bestellt, auch wenn dies anfangs nicht klar war. Alle sind sich einig: Schwein gehabt.

Was wäre passiert, wenn die Einkaufschefin frühzeitig Wind von der ihr unerklärlich großen Bestellung bekommen hätte? Sie hätte die Entscheidung der IT-Anwendung nicht nachvollziehen können und die Bestellung daraufhin manuell auf ein normales Maß reduziert. Das Unternehmen verliert daraufhin Millionen. Gerade als Mitglied des Management-Teams hat sie aber die Pflicht, alle ihr zur Verfügung stehenden Informationen zu nutzen, um optimale Entscheidungen zu treffen. Der Aufsichtsrat des Unternehmens weist daraufhin den Vorstand an, die Einkaufschefin persönlich für die Preiserhöhungen in Haftung zu nehmen.

Zukünftig wird sich die Einkaufschefin aus der Verantwortungslinie nehmen. Sie hat gelernt: Die IT-Anwendung ist eine Autorität, deren Entscheidungen, seien sie noch so unverständlich, sie sich beugen muss.

Möglicherweise treffen IT-Anwendungen in manchen Situationen bessere Entscheidungen als Menschen. Möglicherweise aber auch schlechtere und diskriminierende.

Je intelligenter die IT-Anwendungen werden, desto schwieriger fällt es Menschen, ihre Entscheidungen zu verstehen. Die IT-Anwendungen sind nicht dafür konzipiert, ihre Entscheidungen zu erklären. Gibt es dennoch ein Recht auf Erklärungen, die Menschen verstehen? Sind die Menschen an Entscheidungen gebunden, die sie nicht verstehen? Wer haftet, wenn IT-Anwendungen schlechte Entscheidungen fällen und Menschen verpflichtet sind, diese umzusetzen?

Wenn IT-Anwendungen Entscheidungen auf Ebene von Unternehmensführung und Strategischer Planung treffen, sind Auswirkungen auf Menschen und Umwelt besonders gravierend. Seit einiger Zeit entstehen autonome algorithmische Entitäten, die ausschließlich aus IT bestehen. Wer haftet hier?

Wie werden sich Unternehmen technisch, organisatorisch und dadurch auch kulturell verändern? Welche Auswirkungen gibt es auf Mitarbeiter, Führung, Agilität und Effizienz? Was muss passieren damit sich die digitale Entwicklung verantwortungsvoll vollzieht? Wie lässt sich das Spannungsfeld zwischen Optimierung, Compliance und Integrität in der Balance halten?

2 Künstliche Intelligenz

Intelligenz wird nicht mehr nur als weitgehend angeborene, von äußeren Faktoren und Erfahrungen unbeeinflusste, analytische Fähigkeit betrachtet [3], [4]. Heute verstehen wir Intelligenz als Zusammenfallen von analytischer, kreativer (Fähigkeit zur unkonventionellen Problemlösung) und praktischer Intelligenz (Fähigkeit zur Anpassung an die Umwelt). Intelligenz ist insgesamt also die Fähigkeit, gesetzte Ziele zu erreichen, sich dabei an geänderte Bedingungen anzupassen und dadurch zu lernen [3], [5].

Unter *Künstlicher Intelligenz* (KI) werden Methoden und Anwendungen verstanden, die den kognitiven Apparat des Menschen durch technische Methoden ersetzen. KI bedeutet vor diesem Hintergrund, die Umgebung durch Datenerfassung wahrzunehmen, aus diesen Daten zu lernen, Schlussfolgerungen abzuleiten und sich dabei flexibel an die Gegebenheiten anzupassen [4], [6].

Je nach Fähigkeit der KI, spezifische Schlussfolgerungen zu ziehen, wird KI insbesondere nach ihrer *Stärke* in zwei Klassen unterteilt (andere Klassifikationsschemen unterteilen KI z. B. nach Zielerreichung, Leistungsfähigkeit oder Lernfähigkeit) [4], [7], [8]:

1. *Schwache KI* löst eine spezifische, vordefinierte Aufgabe mittels elementarer Schlussfolgerungen und simuliert dabei menschliches Verhalten. Für Schwache KI gibt es viele Praxis-Beispiele, die in der Regel als „Insellösungen“ abgebildet sind: Von Bilderkennung (beispielsweise bei der Identifikation von Tumoren), Übersetzung, IT-Sicherheit, autonomem Fahren bis zur Unterstützung bei juristischen Gutachten und tlw. Rechtsprechung [1], [9].
2. *Starke KI* greift auf eine breitere Wissensbasis zu und löst Probleme kontextübergreifend durch eigenständige Denkvorgänge. Die KI simuliert nicht nur Denkvorgänge, sondern sie denkt tatsächlich selbstständig und erarbeitet dabei komplexe Lösungen [9]. Als „Stark“ würden KI-Systeme bezeichnet, die schwierige Aufgaben gleich gut oder besser als Menschen erledigen [10]. Eine Starke KI könnte im industriellen Umfeld zum Einsatz kommen und gleichzeitig philosophische Fragen diskutieren und abends auf einfallsreiche Art und Weise mit den Kindern spielen [1].

Für Starke KI gibt es derzeit keine Praxis-Beispiele. Um festzustellen, wann es soweit ist, hat Alan Turing 1950 einen Test vorgeschlagen. Dabei stellt ein Proband beliebige Fragen an ein Terminal, ohne dass ihm bekannt ist, ob ihm ein Mensch oder ein KI-System antwortet. Wenn der Proband das KI-System nicht vom Menschen

unterscheiden kann, ist die KI intelligent [11]. Bisher ist dies nicht gelungen und es ist fraglich, ob dies jemals gelingt, denn dieser über 70 Jahre alte Test wird nicht zu Unrecht kritisiert: So könnten KI-Systeme etwas tun, was dem Konstrukt „Denken“ sehr nahekommt, aber eben von Menschen nicht als solches verstanden wird. Insofern müsste ein KI-System möglicherweise die limitierten Fähigkeiten eines Menschen imitieren, um den Test zu bestehen. Das fängt schon damit an, sich für eine Antwort sehr viel Zeit zu lassen...

2.1 Ausprägungen

Um die angestrebten Eigenschaften der KI zu erreichen, menschenähnliches, intelligentes Verhalten zu zeigen [1], können sich IT-Entwickler verschiedener Methoden bzw. Werkzeuge bedienen.

Eine der zentralen Voraussetzungen für den Einsatz und die Verbreitung von KI ist *Big Data* [4], [12]. Unternehmen und andere Organisationen sammeln massenhaft Daten, gleichzeitig, aus verschiedenen Quellen und auf verschiedenen Wegen, z. B. aus den Bereichen IoT (Internet of Things), smarten Endgeräten oder Social Media. KI hilft nun dabei, aus diesen massenhaft und mit großer Geschwindigkeit wachsenden, tlw. unstrukturierten, vielfältigen, sich ändernden und unzuverlässigen Daten, zuverlässige Informationen zu synthetisieren.

Neben Big Data ist die technische Ausprägung *Maschinelles Lernen* hervorzuheben. Diese bezeichnet Verfahren mithilfe derer KI-Systeme lernen können, ohne für den einzelnen Anwendungsfall programmiert worden zu sein. So kann beispielsweise ein Algorithmus darauf trainiert werden, die Stimmung beliebiger Menschen zu deuten, indem man ihm viele Bilder von Gesichtern zeigt, die jeweils mit „fröhlich“, „traurig“, „wütend“ „überrascht“ usw. ausgezeichnet sind [13], [1]. Maschinelles Lernen wird zumeist in drei Bereiche unterteilt [14]:

- 1) *Überwachtes Lernen*: Wie im Beispiel (Stimmungserkennung) oder in anderen Anwendungsfällen (z. B. Spam-E-Mail-Filter), beurteilen zunächst Menschen bestimmte Sachverhalte („Stimmung“ bzw. „Spam/Ham“) bezogen auf bestimmte Daten (Bilder bzw. Mails), sodass Datenpaare (Input – Output) entstehen. Input: „Bild eines Gesichts mit bestimmter Mimik“ – Output: „Stimmung“. Diese Datenpaare werden dem KI-System zur Verfügung gestellt, bevor sie mit neuen Bildern bzw. Mails gefüttert wird und raten muss. Menschen überwachen die Ergebnisse, berichtigen Fehler und das KI-System wird nach und nach besser. Hinter dem überwachten Lernen stehen Methoden

des Deep Learning bzw. der Neuronalen Netze, die nach dem Vorbild des menschlichen Gehirns und dessen Möglichkeiten, Verbindungen zu schließen agieren. Klassifikation, Regression, Neuronale Netze und Deep Learning kennzeichnen das überwachte Lernen.

- 2) *Nicht-überwachtes Lernen*: Das nicht-überwachte Lernen arbeitet, im Unterschied zum überwachten Lernen, ohne Outputdaten, also ohne Datenpaare. Das KI-System versucht, selbstständig Ähnlichkeiten/Muster in den Attributen der Inputdaten (z. B. Farbe, Größe, Preis, Form) zu erkennen. Und diese dann entsprechend zu clustern. Clustering, Dimensionsreduktion oder Assoziationsanalyse kennzeichnen das nicht-überwachte Lernen
- 3) *Bestärkendes Lernen*: In diesem Bereich lernt das KI-System über ein Belohnungs-/Bestrafungssystem. Ziel ist es beispielsweise, die Anzahl von Schritten, Zügen oder auch monetäre Ausgaben zu minimieren. Jeder ausgeführte Zug wird z. B. mit einem Punkt bestraft. Durch anfänglichen Trial-and-Error lernt das KI-System, welche Strategien unter welchen Umweltbedingungen eher erfolgreich sind als andere.

2.2 Autonomie

Während *Maschinelles Lernen* das Lernen abbildet, bildet *Automatisiertes Schlussfolgern* das Denken ab. Mithilfe von Algorithmen und Methoden des Maschinellen Lernens sollen aus großen Datenmengen Schlüsse gezogen und Lösungen erarbeitet werden. KI-Systeme, die Automatisiertes Schlussfolgern einsetzen, werden auch als Automated (oder auch Algorithmic) Decision Making-Systeme (ADM) bezeichnet. ADM können Entscheidungen vorbereiten oder autonom ausführen. Je nach Grad der Autonomie werden in der Regel drei verschiedene Stufen unterschieden, mit denen Fragen der Haftung und Rechenschaft eng verbunden sind [15], [16]:

- 1) *Assistierend*: Menschen sind die Entscheider. KI-Systeme unterstützen bei Aufgaben, die ein Mensch auch selbst erbringen könnte (z. B. Übersetzung). Vereinfacht wird hier auch vom *Human-in-the-Loop*-Ansatz gesprochen.
- 2) *Erweitert* (augmented, amplified): Menschen sind die Entscheider oder mit Vetorechten ausgestattete Beobachter. Das KI-Systeme bereitet komplexe Entscheidungen vor und führt Aufgaben durch, die vom Menschen schwer vorgenommen werden können (z. B. Vorhersagemodelle auf Basis großer Datenmengen). Das Ergebnis der maschinellen Entscheidungsfindung ist für den Menschen nicht immer nachvollziehbar. Menschen müssen sich ggf. auf die

maschinelle Entscheidung verlassen, sodass dann von einer *gemeinsamen* Entscheidungsfindung gesprochen werden kann. Damit wird der menschenzentrierte Governance-Ansatz in Unternehmen in Frage gestellt. Diese Stufe kommt dem *Human-on-the-loop*-Ansatz nahe: Entscheidungen werden vom Menschen überwacht.

3) *Autonom*: KI-Systeme entscheiden allein (*Human-out-of-the-loop*).

Intelligente, autonome KI-Systeme setzen dort an, wo Menschen an ihre naturgegebenen Grenzen stoßen, und brechen mit ihrer tatsächlichen – aber zumindest zugeschriebenen – Überlegenheit etablierte Machtstrukturen in Unternehmen auf. Bei einer Übertragung von Macht und Verantwortung auf KI-Systeme als Blackbox mit undurchsichtiger Funktionsweise, ist zwangsläufig die Balance zwischen extrinsischer Regeltreue (Compliance) und intrinsischer Selbststeuerung (Integrität) gefährdet [17], [18]. Naturgemäß stehen auf Algorithmen beruhende KI-Systeme den Compliance-Ansätzen deutlich näher und gefährden über Anreize und Sanktionen individuelle Autonomie, Freiheitsgrade und Integrität, die bei menschlicher Führung zwar nicht garantiert, aber doch wahrscheinlicher sind [17], [18]. Es droht eine Rückkehr zu autoritären Führungspraktiken in neuem Gewand [19].

2.3 Erklärbare KI

Intransparenz und mangelnde Nachvollziehbarkeit von KI-Systemen in Verbindung mit der ihnen zugeschriebenen Überlegenheit sind wenigstens so lange problematisch, wie Algorithmen mit Menschen zusammenarbeiten oder von Menschen kontrolliert und gesteuert werden (s. „Human-in-the-Loop“ und „Human-on-the-loop“). Denn beim Einsatz von KI-Systemen in der Unternehmensführung bzw. Strategischen Planung ohne verbindliche Nachvollziehbarkeit, verschiebt sich die Intelligenz in letzter Konsequenz vom Mensch zur Maschine und zwar auf eine besonders kritische Unternehmensebene. Auf der es eben nicht mehr um die Bestellung von Sensoren geht, sondern um die Zukunft des Unternehmens, nun also Maschinenentscheidungen unmittelbaren Einfluss auf das berufliche, aber auch private und familiäre Leben vieler Menschen haben.

Wird also von Menschen erwartet, dass sie Entscheidungen von KI-Systemen vertrauen *und* nicht gleichzeitig zu reinen Erfüllungsgehilfen werden, müssen sie den Kern der getroffenen Entscheidungen verstehen [20]. Die Forschung zur erklärbaren KI (eXplainable AI): XAI steckt noch in den Kinderschuhen [21], wird aber wichtige Voraussetzung zur Erfüllung des Rechts auf Erklärung (s. z. B. Art. 13 DSGVO). Heute

ist das noch nicht der Fall: Hätte die Einkaufschefin aus dem obigen Beispiel ihr System befragt, warum es zur Entscheidung gekommen ist, die 10-fache Menge zu bestellen, hätte das System ihr geantwortet: „Weil es mit 93-prozentiger Wahrscheinlichkeit die beste Entscheidung ist.“

Automatisierte inhaltliche Erklärungen sind aber nicht nur für Vertrauensbildung, Zusammenarbeit und Feststellung von Verantwortlichkeiten relevant, sondern auch, um zu bestätigen, dass eine Entscheidung korrekt getroffen wurde. So lernen KI-Systeme manchmal, dass ihnen „schummeln“ die Arbeit erleichtert und sie dennoch die explizit vorgegebenen Ziele erreichen: 2017 sollte ein KI-System lernen, wie es Pferde auf Bildern in Trainingsdaten erkennt. Stattdessen stellte es fest, dass es nur nach dem Copyright-Tag eines bestimmten Fotografen suchen musste, der zufälligerweise alle in den Trainingsdaten enthaltenen Pferdebilder aufgenommen hatte, statt zu lernen, ob tatsächlich ein Pferd abgebildet war [22]. So lernte die KI zwar aus den Trainingsdaten (Datenpaaren) und erfüllte die expliziten (programmierten) Ziele, kann aber nicht die impliziten, deutlich komplexeren Ziele des menschlichen Systemdesigners interpretieren.

Auf dem Weg zur erklärten KI-Entscheidung sind diverse Hürden zu nehmen. Zum einen wird ein negativer Zusammenhang zwischen Prognosequalität eines KI-Verfahrens und deren Erklärbarkeit postuliert. Je besser die KI, desto schwieriger die für Menschen verständliche Erklärung [23]. Weiterhin ist zu diskutieren, ob eine Trennung von Prognoseverfahren (Blackbox) und parallelen mitlaufenden, dedizierten Erklärverfahren (Whitebox) sinnvoll sein kann. Dies würde die menschen-zentrierte künstliche Begrenzung der KI-Leistungsfähigkeit obsolet werden lassen, allerdings auf Kosten der Erklärungsqualität. Nicht zuletzt stehen KI-Systeme der eigenen Domäne im Austausch mit beliebigen KI-Systemen außerhalb des eigenen Einflussbereichs. Was bedeutet, dass entweder alle extern angebotenen KI-Systeme dieselben Erklärungsanforderungen erfüllen müssten oder die zugelieferte „Zwischenergebnisse“ für das interne KI-System nicht interpretierbar sind und es zwangsläufig zu Erklärungslücken kommt.

2.4 Abdeckung

Während bisherige IT-Anwendungen (z. B. ERP-Systeme) als „totale Idioten“ [2] einfach nicht clever genug waren, aus massenhaft und immer schneller anfallenden, verschiedenartigen, sich ändernden und unzuverlässigen Daten zu lernen und das Gelernte in sinnvolle Entscheidungen zu überführen, gelingt dies KI-Anwendungen nun immer besser. Die drei derzeit wichtigsten bzw. verbreiteten Funktionalitäten von KI-

Anwendungen sind: Daten- und Informationsextraktion, datenbasierte Prognose (Predictive Analytics) sowie die Entscheidungsunterstützung und -automatisierung [24].

KI-Anwendungen können nun in bisher von IT wenig unterstützten Bereichen Anwendung finden (z. B. Strategische Planung), vorhandene IT-Anwendungen übergreifend steuern und so sukzessive immer weitere Unternehmensprozesse und -entscheidungsbereiche abdecken (s. Abbildung 1).

1. *AI-supports-the-organization*: KI unterstützt einzelne Funktionen in Support- und Kernprozessen (wie z. B. Einkauf, Produktion) und deckt einzelne Kern- und Stütz-Prozesse ab.
2. *AI-runs-the-organization*: Perspektivisch wird KI stärker in Management-Prozesse bzw. Unternehmensführung (z. B. Strategische Planung) integriert und zunehmend mit weiteren Entscheidungsrechten ausgestattet. Im Extremfall könnten KI-Systeme Teile der Unternehmensführung übernehmen [16].
3. *AI-covers-the-organization*: Evolutionäre Ausbreitung der KI bis zur Abbildung vieler/aller Unternehmensprozesse auf allen Ebenen (Management-, Kern- und Stützprozessen). Menschen sind hier i. W. für Entscheidung, Erstellung und Parametrisierung von Zielen, Strategien und Algorithmen zuständig, aber nicht mehr dominierend für den operativen Ablauf, Entscheidungsfindung oder -umsetzung. Menschen wirken entweder an den Zielen, Strategien und Algorithmen mit oder werden Erfüllungsgehilfen, z. B. wenn Maschinen nicht zulässig sind.
4. *AI-as-the-organization*: Letztendlich ist die Frage, inwiefern sich neue Geschäftsmodelle entwickeln, die gänzlich ohne menschliches Zutun auskommen. Dabei sind diejenigen Unternehmen im Vorteil, die sich nicht *evolutionär* in Richtung eines solchen Ansatzes transformieren, sondern als „Algorithmic Entities“ [25] bereits *nativ* („von Geburt an“) als solche angelegt sind.

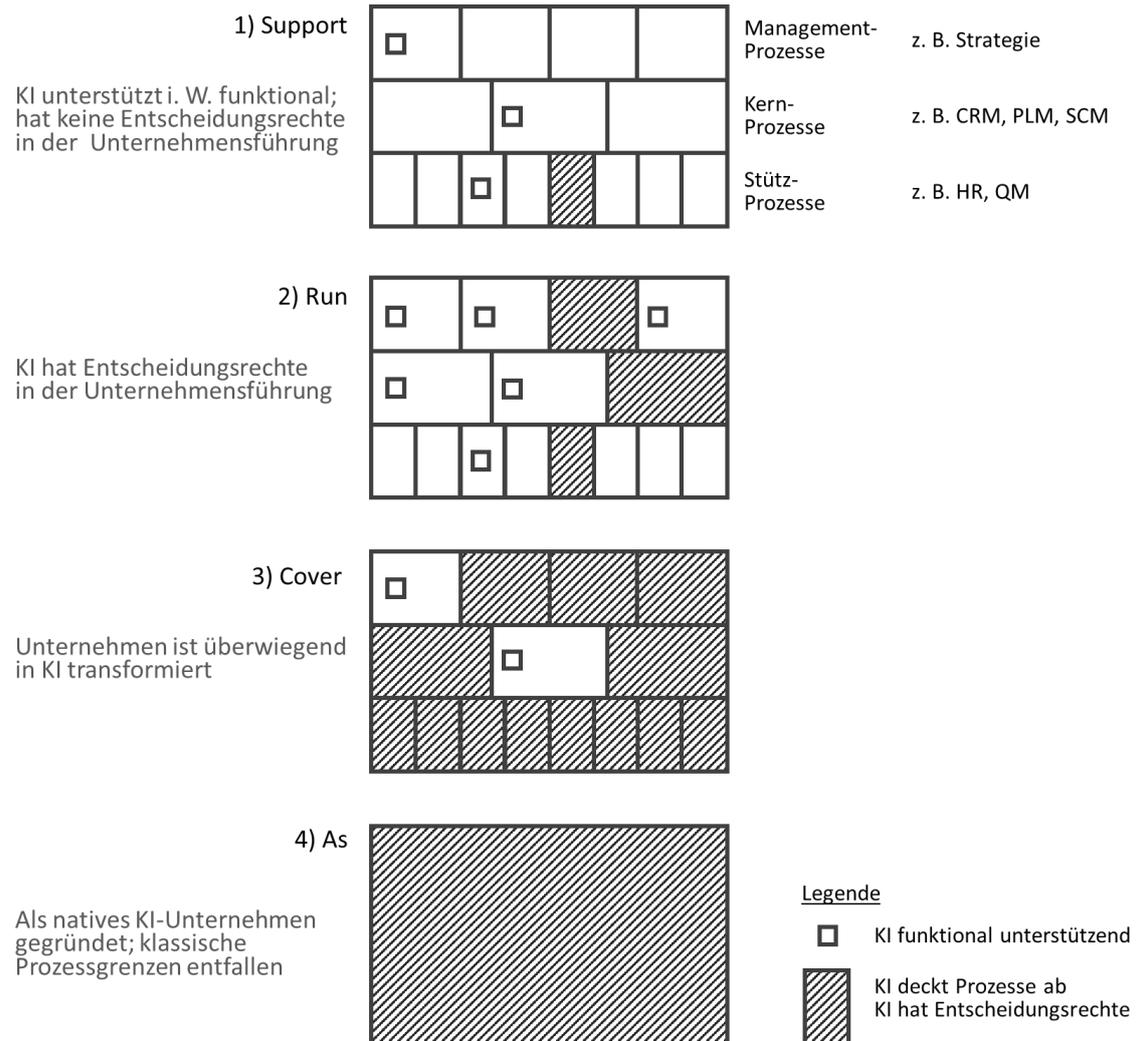


Abbildung 1: Abdeckung KI im Unternehmen (Projizierung auf Prozesshaus)

2.5 Chancen, Risiken und Kritikalität

Eine von der Europäischen Kommission eingesetzte Expertengruppe sieht in KI ein „vielversprechendes Mittel, um das Wohlbefinden von Individuum und Gesellschaft und das Gemeinwohl zu steigern sowie zur Förderung von Fortschritt und Innovation beizutragen. Damit verbunden sind fundamentale, gesamtgesellschaftliche Risiken „im Hinblick auf Demokratie, Rechtsstaatlichkeit, Verteilungsgerechtigkeit oder den menschlichen Geist als solchen“ [6]

Der Einsatz von KI in Unternehmen kann deren Fähigkeiten wie Entscheidungsgeschwindigkeit, Effizienz, Analyse (z. B. Märkte, Kunden, Technologie) aber auch Kreativität positiv beeinflussen und damit Wettbewerbsvorteile in vielen Bereichen (z. B. Innovation, Reputation oder Kostenstruktur) aufbauen. Zudem

beeinflusst KI schon heute die Entwicklung neuer Geschäftsmodelle zahlreicher Branchen [24], [26].

Gleichzeitig ergeben sich rechtliche, ethische und unternehmerische Risiken (Beispiele) [27]:

- *Mangelhafte Entscheidungen* (z. B. durch implizite, für die KI aber unverständliche Ziele)
- *Diskriminierung/Bias* (z. B. durch Trainingsdaten, die aus der Vergangenheit stammen oder nur Ausschnitte abbilden)¹
- *Unklare Haftung und Rechenschaft* (z. B. durch singular unproblematisches – in Gemeinschaft aber risikobehaftetes Wirken)² oder unklare Rechenschaft bei mit externen Anwendungen vernetzten KI-Systemen)
- *Negativer Einfluss auf Unternehmenskultur, Mitarbeitermotivation und Integrität* (z. B. durch angeordnetes Befolgen nicht nachvollziehbarer Entscheidungen)
- *Entstehung einer neuen Art der Unternehmensführung* als „Super-User“, mit „ultimativer“ Entscheidungsbefugnis – über alle KI-Entscheidungen
- *Negativer Einfluss auf Menschen und Umwelt* (z. B. durch KI-Entscheidungen mit direkter Auswirkung auf die berufliche, aber auch private und familiäre Situation von Mitarbeitern)

Die EU-Experten nennen in ihren Ethik-Leitlinien für vertrauensvolle KI bestimmte Voraussetzungen, die KI-Systeme erfüllen sollten: *geltendes Recht* und *ethische Grundsätze* einhalten sowie *robust* sein, um unbeabsichtigten Schaden zu verhindern [6]. Die Datenethikkommission (DEK) der Bundesregierung erweitern und konkretisieren die Forderungen um *menschenzentriertes Design*, *Nachhaltigkeit*, *Qualität*, *Diskriminierungsfreiheit*, *Nachvollziehbarkeit* und *klare Rechenschaftsstrukturen* [28]. Weiterhin empfiehlt DEK, KI-Systeme anhand ihrer Kritikalität einer von fünf Stufen zuzuordnen. Entsprechend sind KI-Systeme ggf.

¹ Beispiele: (1) Medikamente wirken und Spracherkennung funktioniert bei Männern besser als bei Frauen. Grund: zu wenig Trainingsdaten von Frauen (2) Männer werden eher für Führungspositionen ausgewählt. Grund: weniger Frauen auf Führungspositionen in der Vergangenheit (3) Gesichtserkennung arbeitet bei weißen Menschen besser als bei schwarzen. Grund: zu wenig Trainingsdaten von schwarzen Menschen.

² Beispielsweise könnten Auftraggeber, Entwickler, Händler und Nutzer von Algorithmen und Trainingsdaten jeweils plausibel erklären, dass ihr Zutun allein unkritisch ist. Nur in Summe entsteht das Risiko. Ebenso ist die Problematik einer Anbindung externer KI-Systeme mit unzureichender XAI zu sehen.

Zulassungsverfahren unterworfen (Stufe 3), zu beaufsichtigen (Stufe 4) oder gar zu verbieten (Stufe 5). Der Stufe 3 wäre z. B. grundsätzlich der Einsatz staatlicher algorithmischer Systeme, als Träger hoheitlicher Gewalt, zuzuordnen. Für Stufe 4 wird etwa beispielhaft das fiktive KI-System eines Unternehmens mit sehr hoher Marktmacht genannt, welches die Kreditwürdigkeit von Personen oder Unternehmen ermittelt. Systeme zur Rechtsetzung und Rechtsprechung werden von DEK i. d. R. Stufe 5 (aber zumindest Stufe 4) zugeordnet. Dem gegenüber wäre z. B. die in einem Getränkeautomaten verbaute KI der Stufe 1 zuzuordnen („Ohne oder mit geringem Schädigungspotenzial“) [28].

Die Regulierung von Risiken im Unternehmenskontext ist über staatliches Recht allein nicht sicher zu stellen, da multinationale Unternehmen, wie auch z. B. zur Steuervermeidung, in andere Länder ausweichen können und dazu möglicherweise sogar gegenüber ihren Shareholdern verpflichtet sind. KI-freundliche Gesetze in nationalen Rechtsordnungen können zu einem neuen Regulierungswettbewerb nach unten führen [29]. So könnten native KI-Unternehmen (*AI-as-the-organization*) die Arbitrage des regulatorisch optimierten Nomadisierens zum Teil des Geschäftsmodells werden lassen. Eine Regulierung auf nationaler Ebene oder auch im Staatenverbund wird daher zur Einhaltung ethischer Standards nicht genügen: Es ergibt sich die Forderung nach Übernahme unmittelbarer *unternehmerischer* Verantwortung für die Folgen digitalen Handelns.

Die Wahrnehmung einer freiwilligen, über das gesetzlich Vorgeschriebene hinausgehenden, unternehmerischen Verantwortung, für die Folgen der Entwicklung, Verbreitung oder Nutzung digitaler Technologien auf die Gesellschaft insgesamt aber auch die konkrete einzelne Person wird unter dem Begriff *Corporate Digital Responsibility* (CDR) diskutiert [28], [30].

CDR umfasst u. a. die beiden Bereiche Compliance und Integrity:

- 1) *Compliance*: hat einen verbindlichen formal-legalistischen Charakter, der davon ausgeht, dass menschliche Entscheidungen über extrinsische Regeln (Gesetze und sonstige (Selbst-)Verpflichtungen aus Normen, Zertifikaten oder internen Richtlinien), Sanktionen und Anreize gehegt werden.
- 2) *Integrity*: stellt auf die intrinsische Motivation und Fähigkeit ab, Entscheidungen freiwillig und nach individuellen ethischen Überlegungen abzuwägen.

Beides zusammen fokussiert auf die Vermeidung rechtlichen und ethischen Fehlverhaltens [17].

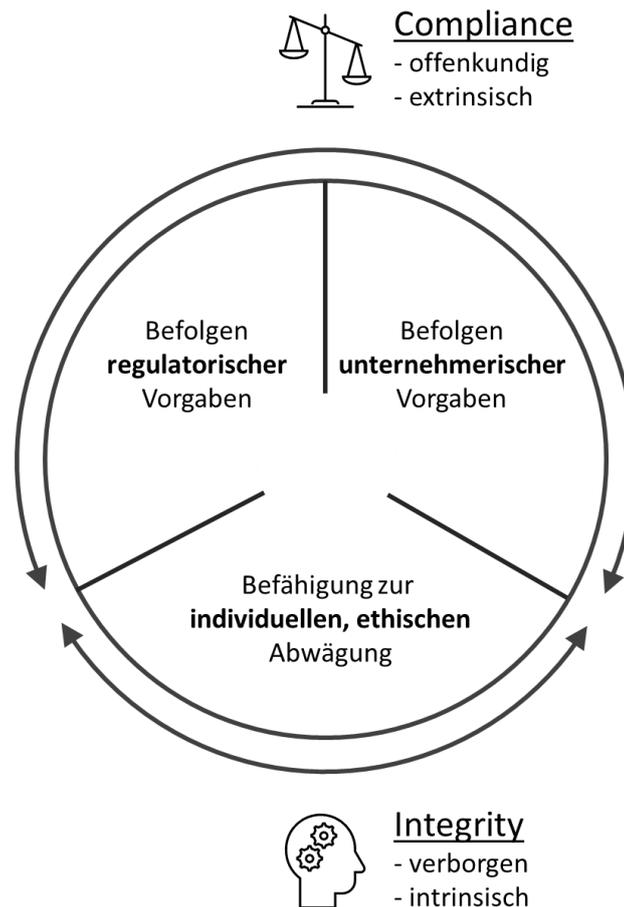


Abbildung 2: Ausgestaltung einer wirksamen CDR (Anlehnung an [17], [31], [32])

CDR kodifiziert Vertrauen und schafft einen Handlungs- und Entscheidungsrahmen, der zeigt, wie KI-Risiken und -Chancen auszubalancieren sind, um der sozialen Verantwortung von Unternehmen hinsichtlich digitaler Fragestellungen gerecht zu werden. In einer wohldurchdachten CDR steckt das Potenzial zur Differenzierung und Schaffung von Wettbewerbsvorteilen durch Erlangung von Vertrauen der Stakeholder gegenüber dem Unternehmen.

3 KI in der Unternehmensführung

Unternehmensführung besteht aus Herbeiführen, Treffen und Umsetzen von Entscheidungen, durch die festgelegt wird, wie die Arbeit von Menschen im Unternehmen koordiniert und die Entwicklung des Unternehmens geprägt wird [33]. Die Unternehmensführung ist daher zweifellos diejenige Unternehmensfunktion, die den größten Einfluss auf das Unternehmen selbst (z. B. Positionierung, Erfolg, Prozesse, Mitarbeiter), seine Share- und Stakeholder (z. B. Eigentümer, Kunden, Lieferanten) und seine politische, rechtliche, gesellschaftliche, ökologische und technische Umwelt hat.

Entscheidungen der Unternehmensführung sind gekennzeichnet von hoher Unsicherheit, Volatilität, Komplexität und Wirkung. Aus diesem Blickwinkel betrachtet, bietet die Unternehmensführung ein vielversprechendes Einsatzgebiet für KI-Systeme.

„Die Wahrheit ist, dass das Geschäft zu komplex geworden ist und sich zu schnell bewegt, als dass Vorstände und CEOs ohne intelligente Systeme gute Entscheidungen treffen könnten“ [34].

In der Literatur wird die Unternehmensführung häufig nach drei *Ebenen* [35], [36]:

1. In der *normativen Unternehmensführung* stehen weniger sachlich begründete Entscheidungen im Vordergrund als vielmehr die von den Trägern eines Unternehmens vorgegebenen Normen (z. B. Vision, Selbstverständnis oder grundlegende Ziele).
2. Die *strategische Unternehmensführung* ist das Feld der strategischen Planung. Nach Analyse von Umwelt und Unternehmen werden Strategien formuliert und umgesetzt. Klassisch wird diese Ebene als zentrale Aufgabe des Top-Managements angesehen.
3. Die dritte Ebene bezieht sich auf die *operative Unternehmensführung*, die innerhalb des von der strategischen Unternehmensführung aufgespannten Rahmens agiert.

Während eine Abbildung von KI-Systemen auf die Ebene der normativen Unternehmensführung nur schwer vorstellbar ist (dies wäre eher ein Gebiet der sog. *Starken KI*), bieten sich die beiden anderen Ebenen für einen KI-Einsatz an.

Eine andere Unterteilung der Unternehmensführung fokussiert auf die *Aufgabenfelder* der Unternehmensführung. Diese umfassen u. a. Strategische Planung, Finanzplanung, Risikomanagement oder Interne Revision (als damit zusammenhängender Bereich gilt auch die externe Wirtschaftsprüfung). In all diesen Aufgabenfeldern werden im KI-Einsatz große Chancen gesehen. Einer der interessantesten Bereiche ist dabei die Strategische Planung.

Strategische Planung ist das Aufgabenfeld der Unternehmensführung, mit dessen Hilfe Strategien formuliert und umgesetzt werden. Eine Strategie beschreibt den Weg zum Erreichen von Unternehmenszielen. Die Qualität der gesetzten Strategien und deren Umsetzung entscheiden über das Erreichen der gesetzten Unternehmensziele [37].

Der Entwicklungsbeginn der Strategischen Planung kann auf die 1960er Jahre zurückverfolgt werden, als sich die Langfristplanung zur Strategischen Planung wandelte. In dieser Zeit erlebte die Strategische Planung eine Phase kaum hinterfragter Popularität. In Stabsabteilungen großer Unternehmen waren teilweise hunderte Menschen mit der mechanistischen Planung und Lenkung des Unternehmens beschäftigt. Strategische Planung als bürokratisches System, getragen von mächtigen Zentralabteilungen, ist heute weitgehend ausgerottet, auch wenn manch einer, beschäftigt mit dem jährlichen Planungsprozess, das ganz anders sieht [38].

Die meisten Modelle der Strategischen Planung folgen dabei im Grundsatz einem einheitlichen Phasenmodell (Abbildung 3).

Phasen 1 & 2	Phase 3	Phase 4	Phase 5
Analyse von Umfeld und Unternehmen	Strategie-Formulierung	Strategie-Umsetzung	Umsetzungs-Controlling
Analysen von: - Markt/Kunden - Wettbewerb - Chancen/Risiken - Stärken/Schwächen - ...	- Erfassung - Validierung - Kaskadierung - Einflussanalyse - Backplanning - ...	- Planung - Budgetierung - Kommunikation - Verantwortungen - Implementierung - ...	Kontrolle - Prämissen - Durchführung - Wirksamkeit Messung - ...

Abbildung 3: Phasen und Aufgaben der Strategischen Planung [37]

Auf Basis der Unternehmensvorgaben (z. B. Vision, Ziele, Positionierung) werden Umfeld und Unternehmen systematisch analysiert (Phasen 1 & 2). Auf dieser Basis werden Strategiealternativen evaluiert und konsolidiert; Strategien werden formuliert (Phase 3). Anschließend wird die Umsetzung der formulierten Strategien geplant und

vorangetrieben (Phase 4). Begleitend zur Umsetzung werden Fortschritt und Ergebnis kontrolliert und gesteuert (Phase 5).

„KI kann dabei helfen, Entscheidungsprozesse zu rationalisieren, große Entscheidungen vom Bauchgefühl in datengesteuertes Wissen umzuwandeln und das zukünftige Ergebnis solcher Entscheidungen besser vorherzusagen.“ [39]

3.1 KI in der Unternehmensführung bestehender Unternehmen

KI-Technologien können Prozesse der strategischen Entscheidungsfindung mit dem Ziel, bessere Entscheidungen zu ermöglichen, unterstützen (z. B. Prognosen, Auswertungen großer Datenmengen). Hier liegen Interpretation und Entscheidungen in den Händen der Menschen, die durch diese KI-Systeme unterstützt werden [40]. Bereits heute ist eine Tendenz zum *Automated-Enterprise* [25] bzw. *Algorithmic-Business* [26], zu beobachten, in denen KI-Systeme eigenständig Entscheidungen ableiten. Dadurch liegt die Steuerung von Organisationen immer mehr in den Händen von KI-Systemen [40].

Bei der Betrachtung möglicher Einsatzgebiete für den konkreten Einsatz von KI in der Strategischen Planung bzw. Unternehmensführung sind vor allem drei Fragestellungen zu evaluieren:

1. Ist KI *tatsächlich notwendig* (Kosten/Nutzen)? Oder führen auch regelbasierte Programme zum gewünschten Ergebnis? Grundsätzlich kann KI immer dann sinnvoll zum Einsatz kommen, wenn Menschen keine Regeln aufstellen können, mithilfe derer Entscheidungen getroffen werden können.
2. Ist die *Organisation bereit* KI zu nutzen? Während KI in Spezialbereichen (z. B. Chatbots oder Kfz-Schaden-Regulierung auf Basis von Unfall-Bildern) schnell auf Zustimmung stößt, wird der (häufig durch Intransparenz gekennzeichnete) Einsatz von KI-Systemen in Bereichen mit komplexen Auswirkungen auf Mitarbeiter, Stakeholder, Umwelt und Unternehmen selbst, eher kritisch gesehen und bedarf ethischer Abwägungen und einer grundsätzlichen Veränderung der Unternehmenskultur.
3. Ist KI *technisch in der Lage* das Problem zu lösen? Einsatzgebiete sind z. B. Analyse/Erkennung von Events aus Bild- und Textdaten oder Prognosen auf Basis der Analyse historischer, unstrukturierter Daten. Voraussetzung ist i. d. R. das Vorliegen ausreichender Trainingsdaten.

Die Strategische Planung umfasst zahlreiche Aufgaben (Abbildung 3), die potenziell für die KI-Unterstützung geeignet sind. Insbesondere scheinen die Analysen der Phasen 1 und 2 (Unternehmen/Umfeld), anhand einer Reflexion der o. a. drei Fragestellungen, für eine KI-Unterstützung grundsätzlich interessant (s. o. *AI-supports-the-organization*). Natürlich werden diese Analysen bereits heute durchgeführt bzw. erworben; KI kann allerdings zusätzliche und aktuelle Einsichten ergänzen:

- Beispiel „Kunden- und Mitarbeiterzufriedenheitsanalyse“: Eine kontinuierliche Analyse von in Sozialen Medien geposteten Texten oder Bildern im Zusammenhang mit dem eigenen Unternehmen kann frühzeitig Signale für Events oder Trendänderungen geben. Auch kann KI in den bestehenden Verfahren der Mitarbeiter- und Kundenzufriedenheitsanalysen zu neuen Erkenntnissen führen.
- Beispiel „Marktanalyse“: Ergänzend zu von vielen Unternehmen einmal jährlich von großen Analystenhäusern erworbenen, auf Gesprächen mit Marktteilnehmern basierenden Reports, können Unternehmen ganz eigene Fragen an die Marktentwicklung stellen und sich tagesaktuelle Veränderungen aufzeigen lassen. Praktisch ein KI-Monitor mit Hinweisen zu Märkten, Patenten, Risiken.³
- Beispiel „Wettbewerbsanalyse“: Auf Basis verfügbarer Daten⁴ bewertet die KI Wettbewerber in Hinblick auf den Zusammenhang zwischen aktueller Situation (z. B. Größe, Preisniveau, Kundenzufriedenheit, Image, Mitarbeiterzufriedenheit, Bonitätsniveau, Liquidität, Innovationen, Arten von Vertriebskanälen, Produktfeatures) und Zielgrößen (z. B. EBIT, Aktienkurs, Umsatz) und kann so Impulse für geeignete Maßnahmen geben.

Während die Strategische Planung früher und auch heute noch häufig ein jährlich-formalisierter Kraftakt war und ist, deren Ende von allen Beteiligten herbeigesehnt wird [37], werden resultierende bis heute eher „starre“ Strategien durch kontinuierliche, transparente Aktualisierung der Rahmenbedingungen (z. B. prognostizierte Marktverschiebungen) kontinuierlich hinterfragt. Die Strategie wird agil und flüssig: zu einer sich stetig an wechselnde Umweltbedingungen annähernden, indifferenten Gleichgewichtslage. Das Konzept „Strategie“ und das Werkzeug „Strategische

³ Ein Unternehmen, das (Micro-)Trendanalysen durchführt, ist das Unternehmen nwo.ai

⁴ Unternehmensinformationen von Dienstleistern (z. B. Reuters, Bloomberg, Handelsregister, Gartner, Nielsen) sowie Produkt-Vergleichsplattformen, Bewertungsplattformen bei denen Mitarbeiter ihren Arbeitgeber bewerten aber auch öffentliche Quellen wie Marken- und Patentämter.

Planung“ werden sich verändern. Zum echten Wettbewerbsvorteil kann die agile Strategie bei Anpassung von Menschen, Kultur, Prozessen und IT-Systemen werden.

Die traditionelle Frage in der Strategischen Planung „Wo wollen wir in einem Jahr sein – und wie kommen wir da hin?“ ändert sich in eine andauernde Suche nach denjenigen Maßnahmen, die (unter Beachtung verfügbarer Ressourcen) zur bestbewerteten Stellung führen. Die jährlichen „Zielbilder“ werden abgelöst von einer kontinuierlichen Stellungsbewertung „Zug um Zug“.

In nachgelagerten Phasen der Strategischen Planung (z. B. Phase 3: „Strategie-Formulierung“) würde die KI-Unterstützung noch stärker in die Unternehmensführung integriert und ggf. mit Entscheidungsrechten ausgestattet (s. o. *AI-runs-the-organization*). Entsprechend sind, bei gleichzeitig unvollständiger Entscheidungstransparenz, Auswirkungen auf Mitarbeiter, Stakeholder und Unternehmen zu erwarten. „Künftig werden wir Systeme haben, die Aufgaben und Tätigkeiten verteilen werden. Dies bedeutet eine tiefgreifende Verschiebung der Rolle von KI-Systemen und erfordert umso mehr Achtsamkeit gegenüber unseren Mitarbeitern.“ [24]

Beispiel: Die KI-gestützte Zusammenführung von Analysen des Unternehmensumfelds und des Unternehmens selbst (Phasen 1 & 2) führt zu Strategieentscheidungen, die eine Anpassung von Vertriebswegen und Portfolio beinhalten. Der prognostizierte Absatz für ein Produkt sinkt um einen Teil, der der jährlichen Produktion verschiedener Werke in Europa entspricht. Die KI empfiehlt/entscheidet die Schließung eines bestimmten Werkes in einem osteuropäischen Land, da dies den geringsten regulatorischen und finanziellen Aufwand bedeutet und gleichzeitig den Image-Schaden für das Unternehmen im deutschen Ursprungs- und Kernabsatzmarkt minimiert.

Dieses Beispiel lässt zahlreiche Auswirkungen erwarten:

- 1) Strategische KI-Entscheidungen haben massive, komplexe und unmittelbare Auswirkungen auf Menschen, die direkt (z. B. Mitarbeiter und Familien) oder indirekt (z. B. Zulieferer, Servicekräfte) betroffen sind.
- 2) Entscheidungen, von einer als objektiv, neutral und überlegen angesehenen KI-Entscheidungsarchitektur getroffen, haben starke Legitimationsfunktion für Management und Eigentümer. Ähnlich wie bisher vielleicht Unternehmensberatungen (mutmaßlich neutrale und objektive Instanz) als Sündenbock unangenehmer Entscheidungen herangezogen werden, wird KI diese Rolle mit noch mehr Gewicht ausfüllen.

- 3) Die Beziehung von Eigentümer und Management (Prinzipal-Agent) verändert sich: Während Manager bisher gerade wegen ihrer besonderen Fähigkeit, Entscheidungen auf Basis ihrer persönlichen Erfahrungen auch „aus dem Bauch heraus“ zu treffen, ausgesucht werden und die Nachvollziehbarkeit von Management-Entscheidungen für Eigentümer nicht immer gegeben ist, können Eigentümer und Dritte (z. B. interner/externer Audit, Insolvenzverwalter) nun sehr einfach verfolgen, bei welchen Entscheidungen sich das Management nicht an KI-Empfehlungen hielt. Bei einer Abweichung wird das Management seine Entscheidungen deutlich stärker als bisher erklären müssen. KI-Entscheidungen nicht oder anders umzusetzen, führt möglicherweise zu rechtlichen Konsequenzen: So gilt in Deutschland seit 1997 die sog. *Business Judgement Rule* (§ 93 Abs. 1 Satz 2 AktG), die besagt, dass Vorstände nur dann nicht für negative Folgen unternehmerischer Entscheidungen haften, wenn die Entscheidung u. a. auf Grundlage angemessener Informationen getroffen wurde. Vorstände müssen dies im Zweifel beweisen (§ 93 Abs. 2 Satz 2 AktG). Können Sie dies nicht (z. B. im Falle einer nicht absolut nachvollziehbar begründeten Abweichung von der KI-Entscheidung), sind sie dem Unternehmen zum Ersatz des entstandenen Schadens aus ihrem Privatvermögen verpflichtet.
- 4) Diese neuartige Beziehung zwischen Eigentümern und Management sowie das zumindest teilweise intransparente Zustandekommen von Entscheidungen, wandelt Führungspraktiken, Rolle und Kompetenzanforderung des Top-Managements vom modernen, partizipativen, risikonehmenden Unternehmensführer und Entscheider zum reinstitutionalisiert-autoritären, risikoaversen Umsetzer und Controller.
- 5) Das Gefühl von Maschinen regiert zu werden, wirkt sich auf Unternehmenskultur, Mitarbeitermotivation und Integrität aus: Während in der Ära der Mechanisierung, Maschinen einfache, repetitive Arbeiten *für Menschen* ausführen, und in der Zeit der Automatisierung Maschinen einfache Arbeiten *selbstständig* erledigen, führen nun Menschen Arbeiten *für Maschinen* aus. In KI geprägten Unternehmen wird insgesamt ein Ungleichgewicht zugunsten von Compliance und zulasten von Integrität entstehen.
- 6) Weiterhin sind viele weitere wichtige Risiken (etwa Diskriminierung oder Haftungsfragen) zu berücksichtigen.

In einer zukünftigen Entwicklungsstufe ist für bestimmte Bereichen denkbar, die gesamte Strategische Planung (von Analyse bis Controlling) in autonome KI-Systeme zu überführen (s. o. *AI-covers-the-organization*). Ein Eingreifen des Managements in

die Entscheidungen ist nur noch im Ausnahmefall vorgesehen („Human-in-Command“). „Human-in-the-Loop-Ansätze in der Unternehmensführung erscheinen wenig sinnvoll, wenn es zu Effizienzgewinnen kommen soll [16]. Nur die Umsetzung wird noch in den Händen von Menschen liegen, die aber eher die Rolle dezentraler Projektleiter haben, als Managementfunktion in der Linie. Dennoch bleibt nach heutiger Rechtsordnung die Rechenschaftspflicht vorerst bei der (menschlichen) Unternehmensleitung, selbst wenn der gesamte Entscheidungsfindungsprozess automatisiert wäre [15].

Über die Entscheidungsbereiche der Strategischen Planung (z. B. Innovation, Zusammenarbeit, Optimierung, Transformation, Diversifikation/Konzentration oder Internationalisierung) hinaus ist der Einsatz in weiteren Gebieten der Unternehmensführung (z. B. Entscheidungen über Zielerreichung, Einhaltung von Standards zur Rechnungslegung, Gesetzen und ethischen Aspekten) [15] denkbar bzw. heute schon etabliert. Dazu zählt beispielsweise das Risikomanagement oder die Bereitstellung zuverlässiger Informationen für Aktionäre und Aufsichtsräte: Im Zentrum dabei steht die Aufhebung der bislang existierenden Informationsasymmetrie zwischen Eigentümern und Management [41]. Die Manipulation von Geschäftsdaten wird schwieriger [42]. Heute sind externe Auditoren mit der jährlichen Prüfung befasst, was verschiedene Probleme verursacht: Zum einen vergeht viel Zeit, bis ein Quartals- oder Jahresabschluss vorliegt und zum anderen fokussieren sich Prüfungsgesellschaften auf die routinemäßige und tlw. manuelle Prüfung ausgewählter Stichproben [43].

Ähnlich, wie bei der Einführung und massenhaften Verbreitung von Standard ERP-Systemen, die zu einer Angleichung von Prozessen über Unternehmen und Branchen hinweg und damit zur Reduktion der Differenzierungsmöglichkeit über besondere Unternehmensprozesse führte, ist nun im Bereich Strategie Ähnliches zu erwarten: Der zukünftige Einsatz von Standard-KI-Systemen (zu Standard-KI-Systemen z. B. [44], [45]) in der Unternehmensführung (wie z. B. KI-Systeme zur Marktanalyse) führt zur Annäherung strategischer Entscheidungen über Unternehmen einer Branche hinweg. Differenzierungspotenziale über Elemente der Unternehmensstrategie zur Erlangung von Wettbewerbsvorteilen (z. B. Innovationen, Image/Positionierung) werden reduziert und damit Effizienz- und Größenvorteile noch stärker als bisher betont. Gleichzeitig werden die Nachteile großer Unternehmen, der „schwerfälligen Tanker“ – gegenüber kleineren, vermeintlich flexibleren Organisationen, durch agile Strategie verringert. So verschiebt sich der Wettbewerb um die beste Strategie zum Wettbewerb um die Implementierung der besten Algorithmen, Datenbasis und KI-Systeme.

Voraussetzung für die Ausschöpfung der Wettbewerbsvorteile ist, dass die Unternehmenskultur dem neuen, agilen Effizienzgedanken folgt und Menschen im

Unternehmen die Entscheidungen eher ausführen als hinterfragen. Die wirtschaftlich erfolgreichsten Unternehmen werden versuchen, sich von den Limitationen der XAI und den mensch-zentrierten Arbeitsweisen und -zeiten zu befreien: den Menschen weitestgehend aus dem Spiel nehmen.

Denkbar ist z. B. die Ausgründung rein digitaler, algorithmischer, autonomer Ein-Zweck-KI-Unternehmen die, von den Zwängen menschlicher Bedürfnisse und einiger ethischer Beschränkungen befreit, KI-Potenziale optimal ausschöpfen.

3.2 Unternehmensführung in nativen KI-Unternehmen

Neben bestehenden Unternehmen, die derzeit dabei sind, KI-Systeme zu testen oder produktiv im Einsatz haben, entsteht derzeit eine weitere Unternehmensform, die sich KI-Anwendungen bedienen kann: Native KI-Unternehmen, die im operativen Handeln ohne Menschen auskommen (s. Abbildung 1). Sie sind „nativ“, weil sie digital, z. B. in der Blockchain, als Programmcode „geboren“ werden.

Solche Unternehmen können in Abhängigkeit der Entscheidungsrechteverteilung zwischen Mensch und Maschine unterschiedliche Formen annehmen.

Eine bereits heute existierende Organisationsform ist die sog. *Decentralized Autonomous Organization* (DAO). Diese Form entstand mit einem anderen Zweck erschaffen, weist aber Merkmale auf, die für autonom am Geschäftsverkehr teilnehmende KI-Unternehmen Voraussetzung sind.

DAOs werden dezentral und unmittelbar durch die Eigentümer verwaltet. Eigentumsanteile an DAOs, die gleichzeitig Stimmrechte beinhalten können, lassen sich mit Kryptowährungen erwerben. DAOs basieren häufig auf dem Technologiekonzept der Ethereum-Blockchain. Über diese Blockchain werden, auf Basis vordefinierter Bedingungen (z. B. Eigentümerentscheidungen) automatisch Handlungen vorgenommen, die in sog. Smart Contracts kodifiziert sind: Das operative Handeln der DAOs wird Programmen überlassen.

DAOs sind keine „KI-Unternehmen“, sondern heute i. W. Kooperations- und Abstimmungsmaschinen, bei denen die Eigentümer bestimmen können, welche und wie Smart Contracts ausgeübt werden (z. B. für Finanzierung, Crowdfunding, Wetten). Die Eigentümer können auf strukturierte Art und Weise und ohne Notwendigkeit vertrauenswürdiger Dritter (z. B. Plattformen, Banken, Versicherungen, Notaren) über die Verwendung von Ressourcen (z. B. Geld) entscheiden. Eine juristische Anmeldung

oder ein physischer Ort, an dem die Organisation beheimatet ist i. d. R. nicht vorgesehen.

Beispiele für DAOs sind z. B. „Infinity Economic“, eine DAO über das Crowdfunding etwa zur Finanzierung von Start-ups oder die Erstellung von Währungen stattfinden kann. Ein anderes Beispiel ist „ConstitutionDAO“, die sich zum Ziel gesetzt hatte, im November 2021 in einer Versteigerung eines von nur noch 13 Exemplaren der ursprünglich 500 in Umlauf gebrachten Originaldrucke der Verfassung der Vereinigten Staaten von Amerika zu erwerben. Auch wenn „ConstitutionDAO“ überboten wurde und nicht zum Zuge kam, äußerten sich die Initiatoren positiv: „Auch wenn das nicht das Ergebnis ist, das wir uns gewünscht hatten, haben wir mit ConstitutionDAO immer noch Geschichte geschrieben.“ [46] Derzeit ist die Landschaft der DAOs geprägt durch das Einsammeln, Verwalten und Abstimmen über die Verteilung von Finanzmitteln – für idealistische Ziele (z. B. „FreeAssangeDAO“) oder kommerzielle Zwecke (z. B. „AugurDAO“, das „Wetten auf alles“ ermöglicht, ebenfalls ohne Dritte also quasi „Pferdewetten ohne Buchmacher“ [47]). Ebenso kann eine DAO die gemeinsame Leistungserbringung oder Ressourcennutzung koordinieren. Dabei steht die Zusammenarbeit der Beteiligten (Eigentümer) im Vordergrund.

Aufgrund des beschriebenen Charakters der DAOs, ist eine Übertragbarkeit des DAO-Gedankens auf klassische Unternehmensgeschäftsmodelle nicht unproblematisch. Denkbar ist z. B. die Zusammenarbeit von Individuen oder Organisationen in einer DAO um eine existierende, zentralisierte Machtbündelung (z. B. Plattform⁵ oder dominantes Unternehmen in einer Kooperationsbeziehung⁶) aufzubrechen. Beim „Emerging Legal Technology Forum 2021“ des Thomson-Reuters-Institute wurde prognostiziert, dass eine DAO in wenigen Jahren ein großes Sport-Team besitzen wird: Die Mitglieder könnten gemeinsam investieren und abstimmen, welche Spieler z. B. das „New York Giants“-Profi-Football-Team erwirbt, wer spielen darf oder auf der Bank sitzt. Eine DAO kann auch in kleinem Maßstab existieren: Als Snack-Automat. Der Nutzer geht mit der Maschine durch Geldeinwurf und Snack-Wahl einen Vertrag ein. Die DAO verwendet das Geld, um selbstständig Nachschub zu bestellen, Reinigungsdienste oder die Miete für den Stellplatz zu bezahlen. Es gibt keine menschlichen Manager, alle Prozesse wurden programmiert [48].

Als Vorteile von DAOs werden gesehen: unmittelbare Vertrauenswürdigkeit durch Dezentralität und Transparenz ohne Notwendigkeit vertrauenswürdiger Dritter, gemeinschaftliche Entscheidung der Eigentümer, Ausschluss der Prinzipal-Agent-

⁵ z. B. Vermittlung von Freelancern

⁶ Z. B. Großkonzern als zentraler Anker in eine Supply Chain

Problematik, garantiertes Handeln auf Basis zuvor gemeinsam festgelegter Regeln (Smart Contracts) unter Ausschluss subjektiver Einflüsse und Machtkonstellationen und Pseudonymisierung der Handelnden. Die Organisation unterliegt zumeist keiner Rechtsform, Gesetzgebung und Regulierung.

Auf der anderen Seite lassen sich Risiken identifizieren, die im Wesentlichen im Zusammenhang mit dem unklaren rechtlichen Status der Organisationen stehen: Allen Voran das Risiko der gemeinschaftlichen Haftung: Jeder einzelne DAO-Eigentümer müsste für die Gesamtschulden einstehen – auch in sein Privatvermögen. Erst seit Juli 2021 können im US-Bundesstaat Wyoming DAOs unter gewissen Voraussetzungen als LLC (ähnlich der deutschen GmbH) angemeldet werden.⁷ Dies stärkt die Rechtsstellung und vermindert die persönliche Haftung. Eine LLC könnte einer DAO als eine Art legaler Verpackung dienen.

Ein weiteres Problem ist die inhärent begründete Nicht-Änderbarkeit der zu Grunde liegenden Smart Contracts, die ihrerseits zur Angreifbarkeit führen kann: Ein 2016 in Deutschland initiiertes Projekt „THE DAO“ zeigte die Verletzlichkeit der neuen Technologie. Anleger (Eigentümer) konnten sich an „THE DAO“ beteiligen und gemeinsam entscheiden, wie bereitgestellte Mittel an Start-ups verteilt werden. Ein Programmierfehler erlaubte es Hackern zunächst, auf einen Großteil der eingezahlten Mittel zuzugreifen. Das Scheitern von „THE DAO“ war ein Rückschlag, den einige als Beweis sehen, dass digitale Technologien die Probleme traditioneller Unternehmensführung nicht lösen können, andere argumentieren, dass solche Entwicklungen bei jeder neuen Technologie zu erwarten sind und gestehen DAOs großes Potenzial zu [49].

Auch wenn die heute existierenden DAOs keine Blaupause für digital-native KI-Unternehmen sind, bieten sie doch Merkmale und Voraussetzungen, die für den Aufbau digitaler algorithmisch-autonomer KI-Entitäten prägend sein können: Dazu gehört z. B. der Einsatz von Smart Contracts oder die vertrauenswürdige, von Zentralen und Plattformen unabhängige dezentrale Steuerungsfähigkeit durch die Eigentümer.

Im Folgenden soll also von einer wie auch immer geartete Parallel- oder Weiterentwicklung von DAOs in Richtung von autonom am Geschäftsverkehr teilnehmenden nativ-digitalen, dezentralen KI-Organisationen (AKIO) ausgegangen werden (s. o.: *AI-as-the-Organization*).

⁷ In einigen Nicht-OECD-Ländern ist dies schon länger möglich: Cayman Islands, Panama, Seychellen. Eine dortige Gründung birgt allerdings andere Risiken

Stellt man sich klassisch-hierarchische Unternehmen schematisch in drei Ebenen aufgeteilt vor:

1. *Eigentümer* mit konstitutiven Entscheidungsrechten (Aktionäre, Gesellschafter),
2. *Management* mit strategischen und taktischen Entscheidungsrechten (Vorstand, Geschäftsführung) und
3. *Operatives Handeln*,

sind AKIOs in drei verschiedenen Varianten, mit zunehmenden Entscheidungsrechten auf der Seite von Maschinen, vorstellbar (s. Abbildung 4).

Var.	Unternehmensform	Entscheidungsrechte		
		konstitutiv	strategisch & taktisch	operativ
	Bisherige/klassische Unternehmen	M	M	M Maschine
A	DAO	M	M	Maschine (SC)
B	AKIO	M	Maschine (KI)	Maschine (SC)
C	Von AKIO gegründetes AKIO	Maschine (AKIO)	Maschine (KI)	Maschine (SC)

M = Mensch
SC = Smart Contract

Abbildung 4: Allokation von Entscheidungsrechten verschiedener AKIO-Varianten

Var. A: DAO

- *Eigentümer* mit konstitutiven, strategischen und taktischen Entscheidungsrechten (Menschen)
- *Operatives Handeln* (Smart Contracts)

Var. B: AKIO

- *Eigentümer* mit konstitutiven Entscheidungsrechten (Menschen),
- *Management* mit strategischen und taktischen Entscheidungsrechten (KI) und

- *Operatives Handeln* (Smart Contracts). Eine AKIO ist, einmal etabliert, eine menschenleere Organisation, die in der Lage ist, sich selbst zu erhalten, zu verwalten und zu wachsen.

Var. C: AKIO-gegründetes AKIO

- Ähnlich Variante B., allerdings gegründet von einer anderen AKIO, die diese Gründung in ihren Smart Contracts vorsieht.

Menschen geben im Laufe dieser Variantenbildung immer mehr Entscheidungsrechte an Maschinen ab.

4 Verantwortung bei Nutzung KI in der Unternehmensführung

4.1 Einsatzspezifische Verantwortungskomponenten

Es ist eine Frage der Zeit, bis KI signifikanten Einzug in die Unternehmensführung hält. Insgesamt lässt sich aber feststellen, dass es dabei keine einheitliche (rechtliche und/oder ethische) Verantwortungslage für den Einsatz von KI im Unternehmen bzw. in der Unternehmensführung geben kann. Vielmehr ist die unternehmerische Verantwortung stark von jeweiligen spezifischen Bedingungen abhängig.

Dazu gehören u. a. die Fragen,

- ob die *Aufgabe* in operativen (z. B. Einkauf) oder strategischen (z. B. Strategische Planung) Unternehmensbereichen liegt,
- welcher *Abdeckungsgrad* (s. Abbildung 1) gegeben ist (Support, Run, Cover, As),
- welcher *Autonomiegrad* vorliegt (assistierend, erweitert, autonom),
- welche *Auswirkungen* anzunehmen sind und
- welcher *Führungsstil* bzw. welche *Unternehmenskultur* vorherrscht.

Unternehmen laufen Gefahr, sich ohne eigene Reflexion an erste (noch nicht wirklich etablierte [50]) und unvollständige digitaletische Leitlinien anzulehnen (s. z. B. „Ethische und rechtliche Grundlagen und Prinzipien“ [28] oder „EU-Leitlinie für vertrauenswürdige KI“ [6]). Wenn das CDR-Konzept eines Unternehmens aber nicht beliebig, austauschbar und ohne Wirkkraft für Management, Mitarbeiter und Stakeholder sein soll, muss es an spezifische Unternehmensbedingungen und -situationen angepasst und auf agile Veränderung vorbereitet sein.

4.2 Exemplarische Verantwortungskomponenten nach Abdeckungsgrad

4.2.1 Abdeckungsgrad „Support“ und höher

Die im Folgenden angegebenen, exemplarischen Verantwortungskomponenten können für alle Unternehmen (Abdeckungsgrad „Support“ und höher), die KI-Systeme einsetzen von Bedeutung sein. Im Abdeckungsgrad „Support“ (*AI-supports-the-organization*) unterstützt die KI einzelne Funktionen (z. B. Einkauf, Produktion),

einzelne Kern- und Stütz-Prozesse und/oder ist bereits vereinzelt mit Entscheidungsrechten ausgestattet. Auswahl und Konkretisierung sind im Einzelfall und je nach den jeweiligen spezifischen Bedingungen (Aufgabe, Grad der Autonomie, Auswirkungen, Führungsstil und Unternehmenskultur) auszugestalten.

Eingeschränkte Nachvollziehbarkeit und Transparenz: KI-Systeme liefern tlw. intransparente Entscheidungen/Vorschläge, die von den beteiligten Personen nicht nachvollzogen werden können. Beteiligte Personen stehen möglicherweise im Spannungsfeld zwischen der Verpflichtung, KI-Entscheidungen/Vorschläge umzusetzen (s. Business Judgement Rule) und den eigenen Zweifeln an der Sinnhaftigkeit der KI-Entscheidungen/Vorschläge. Die Motivation der beteiligten Personen leidet darunter und es entsteht Stress. In artverwandten Bereichen existieren bereits Vorschriften, die das Recht auf Erklärung mitunter sogar gesetzlich festschreiben (z. B. DSGVO Art. 13). Prinzipiell könnten beteiligte Personen von allen Haftungsfragen befreit und zu Ausführenden maschineller Entscheidungen reduziert werden. Andererseits gibt es im Bereich XAI Ansätze, die maschinengetriebene Entscheidungen erklärbar machen. Am Markt existierende Produkte (z. B. AI Explainability 360, AI Adversarial Robustness 360) und diverse Forschungsansätze (s. dazu DARPA [51]) versuchen diese Lücke zu schließen.

Einschränkung des Unternehmertums („Entrepreneurship“): Vorstände haften für negative Folgen ihrer unternehmerischen Entscheidungen, wenn sie die Entscheidung nicht „auf Grundlage angemessener Informationen“ treffen (s. Business Judgement Rule). Werden KI-Systeme zur Entscheidungsfindung herangezogen, können Eigentümer und durch sie beauftragte Dritte (z. B. interne/externe Revision) nachvollziehen, bei welchen Entscheidungen sich das Management nicht an KI-Empfehlungen gehalten hat. Die Beziehung von Eigentümer und Management (Prinzipal-Agent) verändert sich. Hierdurch wird eine als bislang zentrale und geschätzte Fähigkeit des Managements, die datengestützte „Bauchentscheidung“, eine wichtige Facette des „Entrepreneurships“, unterminiert. Rolle und Kompetenzanforderung des Top-Managements wandelt sich von einer partizipativen, risikonehmenden Unternehmer-Kultur zum Umsetzer und Kontrolleur. Wollen sich Eigentümer die unternehmerische Kultur ihrer Führungskräfte erhalten, müssen sie für belastbare, adaptive Regeln sorgen, unter welchen Umständen und in welcher Art sich das Top-Management über KI-Empfehlungen hinwegsetzen darf.

Bias/Diskriminierung: KI-Trainingsdaten können verzerrt sein, aus der Vergangenheit stammen oder nur Ausschnitte abbilden. Wenn Maschinen nun die menschliche Lösungsfindung imitieren, indem sie auf Basis der Trainingsdaten entscheiden, können

nicht-intendierte, diskriminierende Entscheidungen/Vorschläge entstehen. Entsprechend sind die immanenten Datenprobleme zu identifizieren und zu korrigieren. Bestimmte Techniken und Tools (z. B. IBM Toolbox 360 Fairness oder Open-Source-Python-Paket Fairlearn) stehen dabei zur Verfügung, sind aber (insbesondere durch das Problem der eingeschränkten Nachvollziehbarkeit) nicht immer ausreichend.

Mangelhafte KI-Entscheidungen, -Robustheit und -Sicherheit: Nicht immer sind die Auswirkungen von KI-System-Entscheidungen bei Entwicklung, Implementierung oder Anwendung absehbar. Gründe hierfür können Datengrundlage, Programmierung oder Anbindung an externe Systeme sein. Bei der Programmierung ist z. B. darauf zu achten, ob Entwickler implizite Ziele verfolgen, die aber für KI-Systeme unverständlich sind. Ebenso auf die Sicherheit gegen Angriffe von außen. Die Erreichung der verfolgten (legitimen) Ziele ist dann gefährdet. In Abhängigkeit von u. a. Aufgabe, Autonomie und Abdeckung kann der Betrieb von KI-Systemen unbeabsichtigt zu finanziellen, materiellen, immateriellen, psychischen oder physischen Schäden für Unternehmen, Mitarbeiter und Stakeholder führen, gerade wenn es um weitreichende Entscheidungen der Unternehmensstrategie geht. Durch Nutzung von KI entstehen, manchmal als Kuppelprodukt, Kenntnisse und Fähigkeiten, die missbraucht werden können. Nicht nur während Design und Entwicklung, sondern auch bei Implementierung und Betrieb sind KI-System-Risiken auf Stakeholder und Umwelt daher mit geeignetem Mitteln zu identifizieren und zu bewerten. Standardmäßige, risikoadaptierte Schutzbedarfsanalysen müssen auf KI-Systeme ausgedehnt werden. Kontrollgremien und/oder Zertifizierungen [52] können nötig sein.

Träge Zertifizierungssysteme: Um durch KI-Systeme entstehende Risiken zu mitigieren, wird eine Zertifizierung von KI-Systemen (und ggf. auch schon der zugrundeliegenden Trainingsdaten) diskutiert. In Abhängigkeit von der Kritikalität wären KI-Systeme mit einem risikoadaptierten Regulierungsansatz unterschiedlichen Zulassungsverfahren unterworfen [28], [52]. Zertifizierungssysteme die einen traditionellen, eher statischen, Ansatz verfolgen, würden allerdings der dynamischen Entwicklung von KI-Systemen hinterherlaufen und deren Weiterentwicklung begrenzen [28]. Das KI-Zertifizierungswesen bedarf daher der Etablierung von Agilitätskomponenten, mit denen auf technologische, aber auch rechtliche und gesellschaftliche Weiterentwicklungen reagiert werden kann. Ein Zertifikat muss seine Gültigkeit auch bei gewissen Veränderungen des Systems bzw. der zugrundeliegenden Daten behalten. Die dem (Re-)Zertifizierungsregime zugrunde liegenden Anforderungen müssen für Entwicklung und Betrieb ein vertrauenswürdiger Kompass sein [52].

Ungleichverteilung des Reichtums: Die Voraussetzung für den wirkungsvollen Einsatz von KI-Systemen sind bestehende und entstehende Daten, die zur wichtigsten Anlagenklasse der immateriellen Wirtschaft zählen [53]. Besitz und Möglichkeit vom ökonomischen Wert der Daten zu profitieren ist eine der wichtigsten Debatten der Zukunft. Die Gesellschaft erwartet eine gerechte Verteilung der durch KI erlangten Vorteile [54].

Destabilisierung der Integrität: Gemeinsame Werte und starke Unternehmenskultur ermöglichen integriertes, verantwortungsvolles Handeln auch in Situationen, die nicht ex-ante durch Compliance-Vorgaben geregelt sind. Integriertes Verhalten bezeichnet also gerade das, was schwierig in Algorithmen und Programmcode zu übertragen ist. Daher stehen KI-Systeme eher der Compliance als der Integrität nah. Vor- und Nachteil von KI-Systemen ist, dass sie ohne subjektive Einflussnahme entscheiden: Integrität als einzelfallabhängiges, reflektiertes, subjektives Entscheiden und Handeln lässt sich nicht kodifizieren. Eine rein an Compliance orientierte KI-Entscheidung kann, beabsichtigt oder nicht, integriertes Verhalten erschweren oder verunmöglichen. Es existieren Ansätze, die darauf zielen, Ethik in KI-Entscheidungen zu messen und zu labeln (s. z. B. „AI Ethics Impact Group“ unter Leitung des VDE e.V. und der Bertelsmann Stiftung). Weiterentwickelt können solche Ansätze dabei helfen, ex-ante Fehlentwicklungen zu identifizieren. Ein im Einzelfall „nicht-integriertes Verhalten“ von KI-Systemen wird sich aber zumeist erst zum Zeitpunkt der Entscheidung feststellen lassen und Überwachungsmaßnahmen/-gremien erfordern. Menschlichem Handeln ist ggf. Vorrang einzuräumen [28]. Insbesondere können unternehmensweite Integrity-Self-Assessments und -Schulungen wie auch niedrighschwellige, anonyme Meldekanäle (auch für Stakeholder) im Umgang mit diesen Risiken helfen.

Verletzung von Datenschutz und Privatheit durch Wettbewerb der Regulierungsregime: Länder bzw. Staatengemeinschaften nehmen teilweise sehr unterschiedliche Positionen in Bezug auf Datenschutz und KI-Regulierung ein [15]. Dadurch entstehen u. a. zwei Problemfelder: Zum einen könnten KI-Hersteller und -Anwender auf Staaten mit eingeschränkter Regulierung ausweichen und so einem Wettbewerb um die laxesten KI-Regelungen Vorschub leisten. Zum anderen könnten KI anwendende Unternehmen Probleme bekommen, länderübergreifenden Datenfluss und Datenverarbeitung aufrecht zu halten. An dieser Stelle sind integriertes Handeln, technische Maßnahmen aber auch entsprechende Überwachungsmaßnahmen und -gremien gefordert.

Unklare Rechenschaft und Haftung: Es gibt viele Beteiligte an KI-Systemen: Entwickler von Software oder Algorithmen, Berater, Implementierer, Datenzulieferer, Auftraggeber, Anwender (Organisation und Menschen) und zusätzlich die wieder ganz

eigenen Verantwortungsstrukturen extern angebundener KI-Systeme. Die Beteiligten sind in verschiedenen Ländern ansässig und damit unterschiedlichen Rechtsordnungen unterworfen [28]. Während die Rückführung unternehmerischer Entscheidungen auf natürliche Personen schon heute nicht einfach ist, ist dies aufgrund der vielen an algorithmisch getriebenen Entscheidungen Beteiligten hochkomplex bis unmöglich [30]. Forensiker werden es nicht immer schaffen, KI-Entscheidungen zu reproduzieren, wenn sich z. B. Datengrundlagen zwischenzeitlich verändern, sodass auch der Versuch, die Blackbox hinter KI-System(-Verbänden) zu entschlüsseln, ins Leere laufen kann. Dies fordert die Einbringung regulatorischer Korrekture (siehe z. B. Institut der „gemeinsamen Verantwortlichkeit“ im Artikel 26 DSGVO), unterstützt durch unternehmensinterne Regularien.

4.2.2 Abdeckungsgrade „Run“, „Cover“ und höher

Für KI-einsetzende Unternehmen mit höheren Abdeckungsgraden als „Support“ sind weitere Verantwortungskomponenten, je nach Einzelfall und spezifischen Bedingungen (hinsichtlich Aufgabe, Grad der Autonomie, Auswirkungen, Führungsstil und Unternehmenskultur) zu berücksichtigen. KI-Systeme werden tiefer in Management-Prozesse (z. B. Strategische Planung) integriert und mit weiteren Entscheidungsrechten ausgestattet; bis hin zur Übernahme (von Teilen) der Unternehmensführung [16] (*AI-runs-the-organization*). Darauf aufbauend ist weiter denkbar, dass KI-Systeme Schritt für Schritt die Prozesse auf allen Ebenen (Management-, Kern- und Stützprozesse) abdecken (*AI-covers-the-organization*).

Strategieopportunismus: KI-Systeme in der Strategischen Planung sind nicht auf die langfristige Erreichung einmal beschlossener Unternehmensziele geeicht, vielmehr bewerten sie kontinuierlich die aktuelle Stellung des Unternehmens und fällen, von Branchendoktrin und Zielbildern losgelöst, diejenigen Entscheidungen, die die Stellung Zug um Zug maximieren. Geradezu kleinlich und grotesk wird es zurückblickend wirken, wie wir uns heute über Reorganisationen und Unternehmenswandel beklagen. Die Strategie-KI wird jederzeit bereit sein, alles über den Haufen zu werfen, wenn es dem Ziel („maximale Stellungsverbesserung mit dem nächsten Zug/Entscheidung“) dient. Dabei machen die Entscheidungen nicht an gedachten Branchengrenzen halt. Vorwärts-, Rückwärtsintegration und Diversifikationen werden mehr als heute zum Standardrepertoire strategischer Entscheidungsinhalte gehören.

Effizienzkultur: Der Einsatz von KI-Systemen in der Unternehmensführung wird völlig neue Richtungen vorgeben und gleichzeitig durch das Aufkommen von KI-System-Standards zu einer Annäherung strategischer Entscheidungen über Unternehmen einer

Branche hinweg führen. So verschiebt sich der Wettbewerb vom Ringen um die beste Unternehmensstrategie zu einer vom Effizienzgedanken geprägten, volatilen Umsetzungsdogmatik. Die Unternehmenskultur ändert sich: Mitarbeiter arbeiten in unvorhersehbaren Umgebungen und mit blindem Umsetzungsgehorsam. Generell sind Menschen, mit ihrem Bedürfnis nach Stabilität und Anerkennung, hier eher hinderlich: Der Einsatz von Menschen wird besonders „teuer“, wenn die Nichtdurchführung von Entscheidungen eine vom System berechnete Bewertung der neuen Stellung ausbremst.

Veränderung der Aufgabenverteilung von Leitung und Kontrolle: Als „wesentlich“ definierte Geschäfte, also z. B. Entscheidungen und Maßnahmen, die großen Einfluss auf die Vermögens- oder Ertragslage eines Unternehmens haben, sind vom Aufsichtsrat zu genehmigen (§ 111 Abs. 4 Satz 2 AktG). Dazu werden regelmäßig auch KI-getriebene Strategie-Entscheidungen gehören. Da es nicht mehr die „eine große Strategische Planung“ im Jahr gibt, ist die Sitzungskultur des Aufsichtsrats nicht aufrechtzuhalten. Entscheidungen müssen kontinuierlich, bei automatisierten Entscheidungen ggf. nachträglich, bewertet und legitimiert werden. Der Aufsichtsrat wird näher an das Unternehmen heranrücken. Besonders konfliktär ist, dass der Aufsichtsrat nicht zum neuen Vorstand werden darf – er kann nicht die Rolle eines „Super Users“ einnehmen, der zum eigentlichen und ultimativen Entscheider wird, denn damit wäre die vorgesehene Kontrollfunktion und Unabhängigkeit des Aufsichtsrats nicht länger zu halten.

Bisherige Leitlinien helfen nur eingeschränkt: Bisherige Leitlinien (z. B. [28], [6]) sind nicht vollständig, da sie u. a. nur auf die Unternehmensführung fokussiert sind, die eine KI überwacht – nicht aber darauf, dass KI selbst Unternehmen führt [16]. Generell ist auch der Vorrang menschlichen Handelns, der so gerne in den Leitlinien betont wird, nicht zu halten. Die EU-Leitlinie schlägt z. B. vor [6]: „Ihnen [den Benutzern] sollte das notwendige Wissen und die Mittel zur Verfügung gestellt werden, um die KI-Systeme hinreichend zu verstehen [...], und sie sollten möglichst in die Lage versetzt werden, [...] sich ihm entgegenstellen zu können.“ Dieser noble Wunsch wird sich in der Realität sicherlich nicht immer umsetzen lassen. Genau darin liegt aber die eigentliche Aufgabe einer Leitlinie: Unternehmen, Mitarbeiter, Stakeholder und Umwelt für konfliktäre und widersprüchliche Situationen zu wappnen. Die EU-Leitlinie bietet hier keine Antworten. Unternehmensspezifische Leitlinien bzw. CDR sollten genau hier ansetzen und sich an der eigenen Situation orientieren.

Ungleichgewicht zwischen Aufgabe, Kompetenz und Verantwortung: Das Prinzip der Rechenschaftspflicht steht im Mittelpunkt der Unternehmensführung. Der Einfluss automatisierter, strategischer KI auf dieses Prinzip ist von zentraler Bedeutung [55]. Zur

Vermeidung des Missbrauchs des Konstrukts der beschränkten Haftung von Kapitalgesellschaften, wurde die Leitung und Überwachung ausschließlich an natürlichen Personen gebunden. Die Kernaufgaben des Managements und Aufsichtsrats: Leitung und Kontrolle eines Unternehmens, sind nicht delegierbar, auch nicht an Maschinen. Selbst wenn der gesamten Entscheidungsfindungsprozess automatisiert wäre, bliebe das Management nach heutiger Rechtslage rechenschaftspflichtig. Wenn gleichzeitig Maschinen für den Menschen bindende und/oder nicht nachvollziehbare Entscheidungen treffen, erzeugt dies ein unlösbares Ungleichgewicht zwischen Aufgabe, Kompetenz und Verantwortung (AKV-Prinzip).

4.2.3 Abdeckungsgrad „As“

Der höchste Abdeckungsgrad (*AI-as-the-organization*) bezeichnet Unternehmen, die bereits als native KI-Unternehmen gegründet wurden. Es lässt sich einerseits argumentieren, dass dieser Fall weit hergeholt scheint, andererseits gibt es mit den DAOs schon heute Organisationen, die Merkmale und Voraussetzungen aufweisen, die für den Aufbau autonomer, digitaler KI-Organisationen (AKIO) prägend sein können, – selbst einige Jurisdiktionen haben sich schon auf die reguläre Teilnahme dieser Organisationen am Geschäftsverkehr vorbereitet.

Nicht mehr alle bisher aufgeführten Verantwortungskomponenten erscheinen relevant, da z. B. die Komponente „Mitarbeiter“ aus der Verantwortung heraus fällt. Dafür kommen neue Bereiche hinzu: auch native KI-Unternehmen werden mit Menschen zusammenarbeiten, verstärkt vielleicht mit Freelancern oder zumindest Mitarbeitern anderer Unternehmen. Insbesondere an dieser Stelle versagen die bisherigen Leitlinien (z. B. [28], [6]), da sie den Fall, dass KI selbst Unternehmen führt, schlicht nicht berücksichtigen [16].

Regime-Opportunismus: Native KI-Unternehmen können sich, ohne den „Ballast menschlicher Bedürfnisse“, frei bewegen und immer wieder neu erfinden. Starre Branchenzugehörigkeiten gehören genauso der Vergangenheit an, wie länderspezifische Firmensitze. Je nach Regulierungsregime und Aufwand werden sich native KI-Unternehmen optimieren, was den Regulierungswettbewerb nach unten beschleunigen kann.

Keine Menschen im Unternehmen: Die Abwesenheit angestellter Menschen auf Ebene von Mitarbeitern und Management könnte zu Entlastungen hinsichtlich z. B. der eingeschränkten Nachvollziehbarkeit und Transparenz führen und Folgen wie die Einschränkung des Unternehmertums mindern. Allerdings steht das native KI-

Unternehmen immer noch in der Verantwortung gegenüber Externen (z. B. Kunden, freien Mitarbeitern, Lieferanten und anderen Stakeholdern).

Auswirkungen der Teilnahme am Geschäftsleben: Die Teilnahme am Geschäftsleben kann enorme Auswirkungen auf Stakeholder, Umwelt und Gesellschaft haben. Ein menschliches Regulativ fehlt für die strategischen, taktischen und operativen Entscheidungen und Handlungen. Auch wenn die Eigentümer in eine stärkere Kontroll-Verantwortung kommen, sind ggf. weitere Stellen vorzusehen, insbesondere wenn es um kritische Aufgaben geht (z. B. Finanztransaktionen oder Verwaltung von Fonds). Ebenfalls kann, analog zur Zertifizierung einzelner KI-Systeme, die Zertifizierung der digitalen KI-Unternehmensentität notwendig sein. In jedem Falle wird das in der EU-Leitlinie fest verankertes Prinzip der Aufsicht durch Vorstand und weitere Abteilungen, also Menschen, [6] nicht zu halten sein.

Haftung: Für native KI-Unternehmen ist u. a. die Frage der persönlichen Haftungsbegrenzung ungeklärt. Das Risiko der gemeinschaftlichen Haftung aller Eigentümer bleibt, solange KI-Unternehmen nicht als eigene Persönlichkeit gegründet oder in eine Kapitalgesellschaft „verpackt“ werden können. Auch wenn dies in manchen Ländern möglich ist, bietet Deutschland dazu noch keine Rechtsgrundlage.

5 Fazit / Zusammenfassung

KI ist schon heute für Unternehmen unverzichtbar, und wird in absehbarer Zeit auch die Unternehmensführung durchdringen. Wie bereits ERP-Systeme, werden auch KI-Systeme zum Standard werden. Wenn KI-Standard-Systeme in der Strategischen Planung zum Einsatz kommen, ist zu erwarten, dass sich die strategischen Entscheidungen von Unternehmen annähern.

Unternehmen werden in dieser Situation Wettbewerbsvorteile erringen, wenn die Unternehmensorganisation und -kultur zur neuen Entscheidungsarchitektur passt: Befreiung von mensch-zentrierten KI-Limitationen und ausgeprägte Umsetzungseffizienz. Menschliches Eingreifen wird als wenig sinnvoll erachtet, wenn Effizienz im Vordergrund steht und ist nur noch im Ausnahmefall vorgesehen. Für Manager wird technisches Know-how und Umsetzungskompetenz wichtiger als langjährige und fundierte Branchenerfahrung. Unternehmen werden „algorithmische Entitäten“ ausgründen die, mit maximaler Autonomie ausgestattet, den Menschen als limitierenden Faktor aus dem Spiel nehmen.

Grundsätzlich kann der Einsatz von KI in der Unternehmensführung positive Auswirkungen haben. Probleme entstehen vor allem dadurch, dass sich der Mensch potenziell immer weiter von Entscheidungsprozessen entfernt und Unternehmen KI-Systeme heranziehen, um Entscheidungen, die die Interessen der Shareholder über die Interessen der Allgemeinheit stellen zu legitimieren.

Eine Orientierung an Leitlinien kann sinnvoll sein. Allerdings bieten die bisher vorliegenden Leitlinien nur begrenzte Hinweise für den Einsatz von KI in der Unternehmensführung bzw. für autonome, native KI-Unternehmen.

Die verantwortlichen Personen in Unternehmen sind gefordert, auf Basis ihrer eigenen spezifischen Situation, kontinuierlich rechtliche und ethische Fragestellungen zu reflektieren und entsprechende Fähigkeiten zu entwickeln. Dabei kann das Instrument der „Corporate Digital Responsibility“ (CDR) einen Rahmen für die Entwicklung eigener Verantwortungskomponenten bieten. Die zentrale Herausforderung liegt in der Ausbildung und Verankerung von Integrität über die gesamte Kette: vom Design des KI-Systems, über Entwicklung, Verbreitung, Implementierung, Anwendung und Betrieb bis zur Pflege.

6 Literaturverzeichnis

- [1] DFKI und bitkom, „Künstliche Intelligenz. Wirtschaftliche Bedeutung, gesellschaftliche Herausforderungen, menschliche Verantwortung,“ 2017. [Online]. Available: https://www.dfki.de/fileadmin/user_upload/import/9744_171012-KI-Gipfelpapier-online.pdf.
- [2] P. Drucker, „The manager and the moron,“ *McKinsey Quarterly*, 3(4), 1967.
- [3] H. Schlinger, „The Myth of Intelligence,“ *he Psychological Record*, Vol. 53, No. 1, 2003.
- [4] A. Kaplan und M. Haenlein, „Rulers of the World, Unite! The Challenges and Opportunities of Artificial Intelligence,“ *Business Horizons*, Vol. 63, No. 1, 2020.
- [5] R. Sternberg, „The Theory of Successful Intelligence,“ *Interamerican Journal of Psychology*, Vol. 39, No. 2, 2005.
- [6] Hochrangige Expertengruppe für KI, „Ethik-Leitlinien für eine vertrauenswürdige KI,“ Europäische Kommission, Brüssel, 2018.
- [7] A. Ostrom, D. Fotheringham und M. Bitner, „Customer Acceptance of AI in Service Encounters: Understanding Antecedents and Consequences,“ in *Handbook of Service Science, Volume II*, 2019.
- [8] M. Huang und R. Rust, „Artificial Intelligence in Service,“ *Journal of Service Research*, Vol. 21, No. 2, 2018.
- [9] S. Russell und P. Norvig, *Artificial Intelligence – A Modern Approach*, London, 2016.
- [10] N. Nilsson, *The Quest for Artificial Intelligence – A History of Ideas and Achievements*, New York, 2009.
- [11] A. Turing, „Computing Machinery and Intelligence,“ *Mind No. 236*, 1950.
- [12] A. Gandomi und M. Haider, „Beyond the Hype – Big Data Concepts, Methods and Analytics,“ *International Journal of Information Management Vol. 35, No. 2*, 2015.
- [13] A. Munoz, *Machine Learning and Optimization*, New York., 2012.
- [14] T. H. Davenport, E. Brynjolfsson und A. McAfee, „Artificial Intelligence – The Insights You Need,“ *Harvard Business Review*, 2019.
- [15] M. Hilb, „Toward artificial governance? The role of artificial intelligence in shaping the future of corporate governance,“ *Journal of Management and Governance* 24, 2020.
- [16] E. P. M. Hickman, „Trustworthy AI and Corporate Governance: The EU’s Ethics Guidelines for Trustworthy Artificial Intelligence from a Company Law Perspective,“ *European Business Organization Law Review*, 2021.
- [17] C. Schank, „Algorithmen und ihr Einfluss auf Compliance und Integrity,“ *Bertelsmann Stiftung: Unternehmensverantwortung im digitalen Wandel*, 2020.
- [18] K. Elliott, R. Price, P. Shaw, T. Spiliotopoulos, M. Ng, K. Coopamooto und A. Moorsel, „Towards an Equitable Digital Society: Artificial Intelligence (AI) and Corporate Digital Responsibility (CDR),“ *Society* 58, 2021.
- [19] F. Daumann-Habersack, „Autorität, Algorithmen und Konflikte – Die digitalisierte Renaissance autoritärer Führungsprinzipien,“ in *Kooperation in der digitalen Arbeitswelt. Verlässliche Führung in Zeiten virtueller Kommunikation*, O. Geramanis, S. Hutmacher und L. Walser, Hrsg., 2021.

- [20] F. Pasquale, „The black box society: The secret algorithms that control money and information,“ *Harvard University Press*, 2015.
- [21] D. Elton, Self-explaining AI as an Alternative to Interpretable AI, 2020.
- [22] S. Lapuschkin, S. Wäldchen und A. e. a. Binder, „Unmasking Clever Hans predictors and assessing what machines really learn,“ *Nature Communications volume 10*, 2019.
- [23] DARPA, „Explainable Artificial Intelligence (XAI),“ *Broad Agency Announcement*, 2016.
- [24] C. Dukino, M. Friedrich und e. al, „Künstliche Intelligenz in der Unternehmenspraxis,“ W. Bauer, W. Ganz, M. Hämmerle und T. Renner, Hrsg., 2019.
- [25] L. LoPucki, „Algorithmic entities,“ *Wash Univ Law Rev 95*, 2018.
- [26] A. Huber, „Geschäftsmodellinnovation in disruptiven Branchen. Am Beispiel der Berufsfotografie,“ *Wirtschaftsinformatik & Management (13)*, 2021.
- [27] T. Simonite, „Tech firms move to put ethical guard rails around AI,“ *WIRED*, 2018.
- [28] DEK, „Gutachten der Datenethikkommission,“ Datenethikkommission der Bundesregierung, Berlin, 2019.
- [29] Eidenmüller, „The rise of robots and the law of humans,“ *Oxford Legal Studies Research Paper No 27/201*, 2017.
- [30] M. Schmidt, Ethik in der IT-Sicherheit, Berlin, 2021.
- [31] E. Göbel, Unternehmensethik, Konstanz und München, 2016.
- [32] B. Noll, Wirtschafts- und Unternehmensethik in der Marktwirtschaft, Stuttgart, 2002.
- [33] E. Heinen, „Der entscheidungsorientierte Ansatz der Betriebswirtschaftslehre,“ *Zeitschrift für Betriebswirtschaft, 41. Jg.*, 1971.
- [34] B. Libert, M. Beck und M. Bonchek, „<https://sloanreview.mit.edu>,“ 2017. [Online].
- [35] K. Bleicher, Das Konzept Integriertes Management, Frankfurt, 2007.
- [36] H. Hungenberg, Strategisches Management in Unternehmen, Wiesbaden, 2014.
- [37] A. Huber, Praxishandbuch Strategische Planung, Berlin, 2015.
- [38] G. Müller-Stevens und C. Lechner, Strategisches Management, 2011.
- [39] M. Rijmeanam, „How Artificial Intelligence Will Change Corporate Governance,“ <https://www.linkedin.com/pulse/how-artificial-intelligence-change-corporate-mark-van-rijmenam>, 2017.
- [40] P. Gentsch, Künstliche Intelligenz für Sales, Marketing und Service, Wiesbaden, 2018.
- [41] J. M.C. und M. W.H., „Theory of the firm: Managerial behavior, agency costs and ownership structure,“ *Journal of Financial Economics 3(4)*, 1976.
- [42] M. R., E. N., B. P. und H. L., The digital transformation of external audit and its impact on corporate governance. Technological Forecasting and Social Change, 2020.
- [43] B. L.A., M. R. und S. R.Ø, „Audit quality and corporate governance: Evidence from the microfinance industry,“ *International Journal of Auditing*, 2015.
- [44] C. Dukino und e. al, „MARKTSTUDIE: KI-ANWENDUNGEN FÜR DIE SACHBEARBEITUNG,“ *Fraunhofer IAO*, 2021.

- [45] M. Hanussek, „CLOUDBASIERTE KI-PLATTFORMEN,“ *Fraunhofer IAO*, 2021.
- [46] „<https://twitter.com/ConstitutionDAO/status/1461498841820192771>,“ [Online].
- [47] „https://www-forbes-com.translate.goog/sites/michaeldelcastillo/2020/07/28/ethereums-first-ico-blazes-trail-to-a-world-without-bosses/?_x_tr_sl=en&_x_tr_tl=de&_x_tr_hl=de,“ [Online].
- [48] „<https://www.thomsonreuters.com/en-us/posts/legal/daos-business-structure/>,“ [Online].
- [49] M. Fenwick und E. Vermeulen, „Technology and Corporate Governance: Blockchain, Crypto, and Artificial Intelligence,“ *Corporate Law & Economics Working Paper No. 2018-7*, 2018.
- [50] S. Becker, A. Nemat und M. Rebbert, „Der Ruf nach operationalisierbarer Ethik – Unternehmensverantwortung in der digitalen Welt,“ *Bertelsmann Stiftung: Unternehmensverantwortung im digitalen Wandel*, 2020.
- [51] „<https://www.darpa.mil/program/explainable-artificial-intelligence>,“ [Online]. Available: <https://www.darpa.mil/program/explainable-artificial-intelligence>. [Zugriff am 01 02 2022].
- [52] J. Heesen, J. Müller-Quade und e. al., „Zertifizierung von KI-Systemen,“ Whitepaper aus der Plattform Lernende Systeme, 2020.
- [53] J. Haskel und S. Westlake, „Capitalism without capital,“ in *Princeton University Press*, 2018.
- [54] R. Altman, „Distribute AI benefits fairly,“ *Nature*, 521, 2015.
- [55] A. Wooldridge und J. Micklethwai, *The company: a short history of a revolutionary idea*, London, 2003.
- [56] U. Waltinger, „Zitiert,“ *Künstliche Intelligenz in der Unternehmenspraxis*, 2019.

Zum Inhalt

KI wird zukünftig die Unternehmensführung durchdringen und auch den Kern autonomer Unternehmen bilden. Das Verständnis für KI-getriebene Strategieentscheidungen ist jedoch eingeschränkt, denn KI ist selten dafür konzipiert, Entscheidungen zu erklären. Ist ein Veto der Unternehmensleitung gegen KI-Entscheidungen so überhaupt möglich?

Zum Autor

Prof. Dr. Alexander Huber lehrt an der Berliner Hochschule für Technik. Zuvor war er zehn Jahre im Führungskreis von Siemens und Berater bei Accenture. Als in Informatik promovierter Wirtschaftsingenieur forscht er an der Schnittstelle von Strategie, Organisation und IT.

ISBN 978-3-910268-02-9

ISSN 1862-3018 (e-book)

ISSN 1862-1198 (print)

www.bht-berlin.de/i

Berliner Hochschule für Technik
Studiere Zukunft