

Katja Langenbucher

# KI in der Leitungsentscheidung des Vorstands der Aktiengesellschaft

SAFE White Paper No. 96 | August 2023

**Leibniz Institute for Financial Research SAFE**  
Sustainable Architecture for Finance in Europe

[policy\\_center@safe-frankfurt.de](mailto:policy_center@safe-frankfurt.de) | [www.safe-frankfurt.de](http://www.safe-frankfurt.de)

# KI in der Leitungsentscheidung des Vorstands der Aktiengesellschaft \*\*

*Katja Langenbucher, Goethe-Universität Frankfurt und SAFE*

August 2023

## Abstract

*Die Erklärung von Intelligenz fasziniert Menschen seit Jahrtausenden, scheint sich doch mit ihr die menschliche Singularität gegenüber Natur und Tier zu manifestieren. Zugleich betonen nicht nur philosophische Strömungen,<sup>1</sup> sondern auch die Mathematik,<sup>2</sup> die Neuro-<sup>3</sup> und die Computerwissenschaften<sup>4</sup> die Abhängigkeit menschlicher Intelligenz von mechanistischen Prozessen. Ob damit eine Verwandtschaft beider Formen der Informationsverarbeitung verbunden ist oder genau umgekehrt fundamentale Unterschiede bestehen, ist seit knapp hundert Jahren Gegenstand wissenschaftlicher Kontroversen.<sup>5</sup> Fest steht allerdings, dass Maschinen jedenfalls in manchen Bereichen die menschliche Leistungsfähigkeit in Schnelligkeit und Präzision übertreffen können. Nähert man sich dieser Vorstellung, drängt sich die Frage auf, ob es sich empfiehlt, bestimmte Entscheidungen besser von Maschinen treffen, jedenfalls aber unterstützen zu lassen. Neben Ärzten, Rechtsanwälten und Börsenhändlern betrifft das auch Leitungsentscheidungen von Unternehmensführern.*

*Vor diesem Hintergrund wird im Folgenden ein Überblick über Formen künstlicher Intelligenz (KI) gegeben. Im Anschluss fokussiert der Beitrag auf die Rolle von KI im Kontext von Vorstandsentscheidungen. Dazu zählen allgemeine Sorgfaltspflichten, wenn über den Einsatz von KI im Unternehmen zu entscheiden ist (unten II). Geht es um die Unterstützung gerade von Vorstandsentscheidungen stellen sich zusätzlich Fragen der Kooperation von Mensch und Maschine, der Delegation des Kernbestands von Leitungsentscheidungen und der Einstandspflicht für KI (unten III).*

---

\*SAFE Policy Paper repräsentieren die persönlichen Ansichten der Autoren und nicht notwendigerweise die von SAFE oder seiner Mitarbeiter.

\*Erscheint demnächst in Florstedt, Wirtschaftsrecht in Zeiten der Krise.

<sup>1</sup> Rolffs, Kausalität und mentale Verursachung.

<sup>2</sup> Stiehl et al. in: Holm-Hadulla/Funke/Wink, Intelligenz: Theoretische Grundlagen und praktische Anwendungen, 279.

<sup>3</sup> Glimcher, Decisions, Uncertainty, and the Brain.

<sup>4</sup> Larson, The Myth of Artificial Intelligence.

<sup>5</sup> Überblick bei Hawkins, A Thousand Minds; Larson; Nath, Philosophy of Artificial Intelligence.

## I. Überblick: KI

Unter den Begriff „KI“ lassen sich unterschiedliche Teilgebiete fassen, die vom maschinellen Lernen (ML) über neuronale Netze bis zu large language models (LLM) und Robotics reichen.<sup>6</sup> Dem entspricht es, dass im Unternehmen vielfältige Einsatzformen denkbar sind, abhängig davon, welche Aufgabe eine KI übernehmen soll. So spielt bei der Planung und Automatisierung von Prozessen die Logik eine Rolle,<sup>7</sup> in der Fertigung die Robotics<sup>8</sup> und in „Wissensarbeit“, wie der Zusammenfassung von Studien oder der Durchsicht von Verträgen, die LLMs.<sup>9</sup> In vielen Bereichen zählt der Einsatz von KI schon heute zum Tagesgeschäft nachgeordneter Unternehmensebenen.<sup>10</sup> Ob sich das empfiehlt und auf welche Weise KI zum Einsatz kommen soll, ist Bestandteil der allgemeinen Leitungsverantwortung des Vorstands. Denkbar ist der Einsatz von KI weiter bei der originären Vorstandsarbeit selbst. Ziel ist in diesem Fall die präzise Vorhersage künftiger Entwicklungen, auf deren Basis Leitungsentscheidungen getroffen werden.<sup>11</sup> Die Entscheidung, ob und welche KI für eine derartige Aufgabe eingesetzt werden kann, hängt von den verfügbaren Daten, den Erwartungen an den Beitrag der Maschine, aber auch von der Art und Weise ab, wie sich Fehler auswirken können.

Zu den klassischen Werkzeugen der Vorbereitung von Leitungsentscheidungen zählt seit jeher die Statistik. Deren Ziel ist es, auf der Grundlage einer Hypothese zu Schlussfolgerungen<sup>12</sup> über die Beziehung zwischen unterschiedlichen Variablen zu gelangen. Der Vorstand einer Bank mag beispielsweise über den Abbau von Filialen zugunsten des Onlinebankings zu entscheiden haben. Hierfür bieten sich Studien zu Präferenzen von Kunden, möglicherweise auch deren Alter, Beruf oder Wohnort, samt Mobilfunknetzabdeckung und Anzahl von Bankfilialen an. Der Vorstand könnte vor diesem Hintergrund, eine Hypothese testen lassen, etwa: ‚Das Alter der Bankkunden ist der zentrale Faktor für unsere Entscheidung‘. Würde sich hierbei herausstellen, dass der angenommene Zusammenhang nicht besteht, etwa: der Wohnort ist wichtiger, wird die Entscheidung anders ausfallen.

Den Statistiker ergänzend oder ersetzend mag man sich vorstellen, eine KI heranzuziehen. Man trainiert die KI mit Daten, welche die Bank zu den Reaktionen ihrer Kunden auf bisher durchgeführte Schließungen einzelner Filialen vorhält. Befragt über die Folgen weiteren Filialabbaus würden

---

<sup>6</sup> Ertel, Grundkurs Künstliche Intelligenz, 5. Aufl. 2021, S. 3 ff.; Russell et al., Artificial Intelligence, 4<sup>th</sup> ed. 2021, chapter 1.1.

<sup>7</sup> Ertel, S. 29 ff.; Russell et al., chapter 7.1, 11.

<sup>8</sup> Russell et al., chapter 26.1.

<sup>9</sup> Ebd., chapter 24.

<sup>10</sup> Zu KI als „Chefsache“ aber: High Level Expert Group/AI, S. 25; Hickman et al., EBOR 2021, 593, 599 f.; Noack, ZHR 183 (2019), 105, 124; Spindler, ZGR 2018, 17, 40.

<sup>11</sup> Agrawal et al., Prediction Machines, S. 23 ff.; Ertel, S. 201 ff.; Russell et al., chapter 19.-22.

<sup>12</sup> Unter diesen Begriff fallen u.a. die Deduktion, die Induktion und die Abduktion. Letztere ist für das Verständnis von KI besonders wichtig, vgl. Finnis, in: Coleman/Shapiro, The Oxford Handbook of Jurisprudence & Philosophy of Law, S. 31; siehe noch unten I.2.b.

möglicherweise Muster erkennbar, etwa Gruppen von Bankkunden mit ähnlichen Präferenzen und Reaktionen („clustering“).<sup>13</sup> Zusätzlich könnte die KI Vorhersagen über die Wechselbereitschaft von Bankkunden treffen und die gezielte Ansprache derartiger Gruppen ermöglichen.

## 1. KI und Training

Für die Auswahlentscheidung der passenden KI spielt unter anderem die Art und Weise, wie die KI trainiert wird, sowie die Kosten, die hierfür aufzuwenden sind, eine Rolle. In manchen Situationen wird es sich anbieten, die KI zu programmieren, indem möglichst erschöpfend denkbare Fälle antizipiert und die gewünschte Reaktion vorgegeben wird. Das besondere Potential jüngerer KI liegt allerdings in deren Potential, selbständig zu „lernen“. Mit diesem Anthropomorphismus meint man: die Leistung der Maschine wird besser, nachdem sie Beobachtungen über die Welt angestellt hat.<sup>14</sup> Ein viel diskutiertes Beispiel hierfür sind selbstfahrende Autos. Es wäre nicht nur aufwändig, sondern vermutlich praktisch unmöglich, für deren Betrieb alle künftigen Verkehrssituationen zu erfassen und Reaktionen zu programmieren. Viel effizienter ist es, die KI aus Daten lernen zu lassen, wie menschliche Fahrer in vergleichbaren Situationen reagieren.<sup>15</sup> Die KI wird dann Muster in den Daten erkennen, etwa: bei starkem Regen fahren Autos langsamer. Die Programmierung des selbstfahrenden Autos passt die KI daraufhin an dieses Muster an.<sup>16</sup> Abhängig davon, welches Optimierungsziel die KI verfolgt (etwa: schnelles Erreichen des Ziels oder angenehmes Fahrgefühl), wird sie die Programmierung des selbstfahrenden Autos kontinuierlich feinsteuern.

### a) Überwachtes Lernen

Zur Programmierung maschinellen Lernens stehen derzeit drei Grundvarianten zur Verfügung.<sup>17</sup> Bei überwachtem Lernen wird die KI auf das Abbilden von Input-Output Paaren programmiert.<sup>18</sup> Input kann beispielsweise das Bild eines Wolfs und Output die Klassifikation als Wolf sein. Trainiert wird die KI mit einer Datenbank, die in ausreichender Zahl Beispiele für den gesuchten Zusammenhang enthält. Dabei zeichnen sich die Beispieldaten dadurch aus, dass sie „beschriftet“ sind. Input- und Output-Daten enthalten bildlich gesprochen ein Etikett (Label).<sup>19</sup> Das Vorhandensein eines solchen Labels erklärt den Begriff „überwachtes“ Lernen. Er bezieht sich auf den etikettierten Trainingsdatensatz. „Überwacht“

---

<sup>13</sup> Russell et al., S. 671, s unten I.2.b.

<sup>14</sup> Russell et al., S. 669.

<sup>15</sup> Ebd., S. 670, zusätzlich mit dem Beispiel von Software zur Analyse von Galaxien unter dem Gravitationslinseneffekt, deren Leistung sich um das 10 Millionen-fache erhöhte während die Energiekosten zur Kühlung von Rechenzentren um 40% abnahm; Agrawal et al., S. 7 ff., 23 ff.

<sup>16</sup> Russell et al., S. 669 f.

<sup>17</sup> Zum folgenden: Russell et al., S. 671.

<sup>18</sup> Russell et al., S. 827: „Supervised learning algorithms all have essentially the same goal: given a training set of inputs  $\mathbf{x}$  and corresponding outputs  $y = f(\mathbf{x})$ , learn a function  $h$  that approximates  $f$  well“.

<sup>19</sup> Ertel, S. 351.

bedeutet dabei: Der KI wird durch das Labeling vorgegeben, welche Funktion sie finden soll. Überwachtes Lernen setzt deshalb bereits bearbeitete Datensätze voraus, die mit den gewünschten Labels versehen sind. Auf dieser Basis lernt die KI für neue Daten Vorhersagen zu treffen. Nützlich ist eine auf diese Weise trainierte KI aber nur, wenn die Labels kenntnisreich programmiert wurden.

#### b) Unüberwachtes Lernen

Geht es um die Analyse von Daten, für die weder ein Trainingsset mit Labels existiert noch überhaupt klar ist, welche Fragen sinnvoll gestellt werden können, bietet sich unüberwachtes Lernen an.<sup>20</sup> In diesem Fall wird ein Datensatz ohne Labels verwendet. Aufgabe der KI ist es, selbständig Strukturen und Muster aufzufinden.<sup>21</sup> Eine gängige Aufgabe für die KI ist hierbei das Clustering. Dabei sortiert die KI ungeordnete („unlabeled“) Daten auf der Basis von Unterschieden und Gemeinsamkeiten.<sup>22</sup> Auch das lässt sich am Beispiel der Bilderkennung verdeutlichen: “when shown millions of images (...) a computer vision system can identify a large cluster of similar images which an English speaker would call ‘cats’”.<sup>23</sup>

Auf welche Weise die KI diese Identifikationsleistung vornimmt, gibt der Programmierer ihr ebenso wenig vor wie bestimmte Ziele. Es verhält sich also gerade umgekehrt wie beim überwachten Lernen. Bei diesem wird die KI auf ein bestimmtes, vorher bekanntes Ziel trainiert, beispielsweise Wolf/kein Wolf oder Spam/kein Spam. Bei jenem zeigt die KI dem Anwender, auf welche Weise sich die ungeordneten Daten sortieren lassen, beispielsweise Ultraschallbilder mit und ohne Anomalien. Um verwertbare Ergebnisse zu erzielen, sind hierfür sehr große Datensätze und Computer mit hoher Rechenleistung erforderlich. Bisweilen bietet es sich in diesen Situationen an, eine Teilmenge des großen Datensatzes doch zu etikettieren („semi-supervised learning“), etwa Ultraschallbilder mit krankhaften Auffälligkeiten. Auf diese Weise wird die Struktur des Problems verdeutlicht, für welches die KI ein Modell entwerfen soll. Das schließt aber nicht aus, dass die KI andere Muster findet, aus denen ein behandelnder Arzt möglicherweise weitere Schlüsse ziehen kann.

---

<sup>20</sup> Russell et al., S. 827: „Unsupervised learning algorithms (...) take a training set of unlabeled examples  $x$ .“ Sie können für verschiedene Aufgaben eingesetzt werden: „The first is to learn new representations – for example, new features of images that make it easier to identify the object in an image. The second is to learn a generative model – typically in the form of a probability distribution from which new samples can be generated (...) Many algorithms are capable of both representation learning and generative modeling“.

<sup>21</sup> Ertel, S. 260.

<sup>22</sup> Russell et al., S. 671.

<sup>23</sup> Ebd.

### c) Lernen durch Verstärkung

Gleichsam zwischen diesen Methoden befindet sich das Lernen durch Verstärkung.<sup>24</sup> Auch hier arbeitet die KI ohne bereits etikettierte Trainingsdaten. Sie führt bestimmte Abläufe durch, etwa ein Brettspiel<sup>25</sup> oder eine Robotik-Aufgabe.<sup>26</sup> Im Gegensatz zu vollständig unüberwachtem Lernen erhält die KI aber nach Abschluss ihrer Aufgabe eine menschliche Rückmeldung in Form von positivem oder negativem Feedback. Auf dieses Feedback reagiert die KI, indem sie den zurückliegenden Ablauf anpasst, um (weiterhin oder erstmalig) positives Feedback zu erhalten: „Imagine playing a new game whose rules you don't know; after a hundred or so moves, the referee tells you „You lose“. That is reinforcement learning in a nutshell“.<sup>27</sup>

## 2. Die Qualität von Vorhersagen

### a) Daten

Begreift man maschinelles Lernen als Umsetzung von Informationen und als Mustererkennung, tritt die zentrale Bedeutung des verwendeten Datenmaterials hervor. Von statistischen Untersuchungen ist längst bekannt, dass deren Qualität eng mit den ausgewählten Daten zusammenhängt. Verzerrungen können durch die Selektion der Stichprobe, durch ausgelassene Variablen oder die Nichtbeachtung von Störvariablen<sup>28</sup> eintreten und ebenso schädlich sein wie Messfehler.

Beim Einsatz von KI stellen sich viele dieser Fragen auf ähnliche Weise. Abhängig davon welche Daten die KI erhält, wie diese Daten strukturiert und möglicherweise etikettiert sind, wird die KI Zuordnungen lernen, Muster erkennen und Modelle für die Beurteilung künftiger Situationen erstellen.<sup>29</sup> So werden sich beispielsweise Vorhersagen, die auf der Basis eines sorgfältig kuratierten,<sup>30</sup> möglicherweise sogar synthetischen<sup>31</sup> Datensatzes erfolgen, erheblich von Vorhersagen unterscheiden, die von einer KI mit Zugriff auf das gesamte Internet erstellt werden. Ebenso wichtig ist die Auswahl von Daten, anhand derer die KI lernt, mithin: ihr Modell anpasst.<sup>32</sup> Im obigen Beispiel des Bankvorstands, der eine Filialschließung erwägt, kann sich das beispielsweise zeigen, wenn die KI anhand eines Datensatzes bisheriger Filialschließungen in nur einem Bundesland trainiert wird. Das kann zu unpräzisen Vorhersagen führen, wenn die KI ein Modell entwickelt, welches zwar den bestehenden Datensatz

---

<sup>24</sup> Ebd., S. 840 ff.

<sup>25</sup> Ebd., S. 671, S. 840.

<sup>26</sup> Ertel, S. 351 mit einem Beispiel aus der Robotik.

<sup>27</sup> Russell et al., S. 840.

<sup>28</sup> Im Kontext rechtlicher Verwertung: Langenbacher, *Economic Transplants*, S. 107 ff.

<sup>29</sup> Russell et al., S. 669.

<sup>30</sup> Zur Kuratierung von Daten: *The Role of Data in AI*, Report for the Data Governance Working Group of the Global Partnership of AI, <https://gpai.ai/projects/data-governance/role-of-data-in-ai.pdf>.

<sup>31</sup> Jordan et al., *Synthetic data – what, why and how?*.

<sup>32</sup> Russell et al., chapter 19, S. 669.

exzellente abbildet, aber sich zu stark auf dessen Besonderheiten konzentriert, um für die Beurteilung hiervon abweichender, künftiger Entwicklungen hilfreich zu sein („overfitting“).

Auf den ersten Blick scheint deshalb die Größe der Datenmenge ein zentraler Faktor für die Qualität der Vorhersage zu sein. So gestattet der offene Zugriff auf das Internet häufig präzise Vorhersagen, beispielsweise über menschliche Präferenzen. Ein Beispiel aus dem Jahr 2016 belegt allerdings damit einhergehende Reputationsrisiken. Microsofts Chatbot „Tay“, der in menschlicher Alltagskonversation trainiert werden sollte, verwandelte sich binnen 24 Stunden in einen Rassist und Antisemiten.<sup>33</sup> Es sind nicht nur derartige Reputationsrisiken geringer, wenn die KI nur auf der Grundlage eines kuratierten oder jedenfalls „gesäuberten“ Datensatzes lernen darf. Kontextabhängig kann auch die Qualität der Vorhersage besser sein, wenn die KI mit kuratierten Daten trainiert wird.<sup>34</sup>

#### b) Methode: Hypothesen versus Muster

Statistik und ML verfolgen unterschiedliche methodische Herangehensweisen.<sup>35</sup> Die statistische Studie beginnt üblicherweise mit einer Hypothese, sammelt geeignete Daten und liefert Zusammenhänge zwischen den untersuchten Variablen. Das Ziel einer auf ML trainierten KI ist es hingegen, möglichst treffsichere Vorhersagen auf der Basis erkannter Muster in den ihr zur Verfügung gestellten Daten zu produzieren.

Auf den ersten Blick unterscheiden sich Statistiker und KI im praktischen Ergebnis nicht grundlegend. Die auf Vorhersagen trainierte KI kann in manchen Fällen, sozusagen als Nebenprodukt, auch Zusammenhänge zwischen Variablen erhellen. Umgekehrt mag man Vorhersagen künftigen Verhaltens als Nebenprodukt der Statistik verstehen. Das lässt sich am Beispiel der Einschätzung der Kreditwürdigkeit eines Bankkunden verdeutlichen.<sup>36</sup> Eine Bank könnte sich für den Zusammenhang zwischen Kreditausfallrisiko und bestimmten Datenpunkten interessieren, die sie von potenziellen Bewerbern schnell und kostengünstig erhalten kann. Für die Auswahl der Datenpunkte stellt ein Statistiker Hypothesen an: Relevant könnte das Einkommen, die Vermögenssituation oder das Rückzahlungsverhalten in der Vergangenheit sein. Die statistische Studie liefert dann Aussagen über eine betrachtete Population, etwa einen Datensatz über Kreditnehmer, mit denen die Bank in der Vergangenheit zu tun hatte. Auf diese Weise können Hypothesen getestet und Wahrscheinlichkeiten quantifiziert werden. Der Statistiker mag beispielsweise belegen, mit welcher Wahrscheinlichkeit

---

<sup>33</sup> Siehe zB Vincent, Twitter taught Microsoft’s AI chatbot to be a racist asshole in less than a day, The Verge vom 24.5.2016, <https://www.theverge.com/2016/3/24/11297050/tay-microsoft-chatbot-racist>.

<sup>34</sup> Anschaulich Bender et al., FAccT 2021, 610: „On the Danger of Stochastic Parrots: Can Language Models Be Too Big?“

<sup>35</sup> Bzdok et al., 15 Nature (2018) 233: „statistics draws population inferences from a sample, and machine learning finds generalizable predictive patterns“; Langenbucher, S. 97 ff.

<sup>36</sup> Eingehend Langenbucher, BKR 2023, 205 ff.; dies., Consumer Credit in the Age of AI.

hohes Einkommen negativ mit Kreditausfall korrelierte. Variiert man Versionen einer KI mit und ohne die betreffende Variable (etwa: hohes Einkommen), erhält man in der Sache eine ähnliche Information. Zwar erkennt die KI wieder Muster. Arbeitet sie aber auf der Basis einer verhältnismäßig geringen Datenmenge, lässt sich durch Hinzufügen und Weglassen von Variablen überprüfen, welche das Muster besonders deutlich beeinflussen.

Je größer allerdings die Datenmenge und je komplexer das maschinelle Lernverfahren, desto deutlicher prägt die unterschiedliche Methode das Ergebnis. Bei großem, aber noch überschaubarem Datenmaterial kann das fruchtbar sein, etwa wenn die KI unerwartete Korrelationen aufweist, die ein Statistiker bei seiner Hypothesenbildung nicht berücksichtigt hätte. Das lässt sich erneut am Beispiel der Kreditwürdigkeitsprüfung verdeutlichen.<sup>37</sup> Das deutsche Fintech Unternehmen Kreditech hatte KI zu Scoringzwecken eingesetzt. Zu den besonders wichtigen Variablen für die Beurteilung des Kreditausfallrisikos zählte das Vorhandensein einer bestimmten (sehr untypischen) Schrifttype auf dem Computer von Bewerbern – zweifellos eine Variable, die wenige Statistiker bei ihrer Hypothesenbildung in den Blick genommen hätten. Es stellte sich später heraus, dass die betreffende Schrifttype von einem Online-Casino verwendet worden war. Der Zusammenhang zwischen Glücksspiel und Kreditwürdigkeit hätte einen Statistiker interessiert, nicht zuletzt, weil er intuitiv einleuchtet. Für die KI des Fintech Unternehmens war der Nachweis dieser Korrelation dagegen im Grunde gleichgültig, solange nur die Vorhersagequalität hoch war. Um noch einen Schritt weiterzugehen: Möglicherweise hatte die KI anhand von Daten gelernt, in denen Personen, die online Glücksspiele mochten, überrepräsentiert waren. Ein Schriftsetzer, auf dessen elektronischen Geräten sich dieselbe Schrifttype findet, würde aus diesem Grund zu Unrecht keinen Kredit erhalten.

### c) Neuronale Netze

Werden der KI sehr große Datenmengen zur Verfügung gestellt und arbeitet sie mit besonders komplexen Lernverfahren, gelingt es nicht mehr ohne weiteres, einzelne Variablen zu identifizieren, ohne dabei Abstriche an der erreichten Präzision des Vorhersageergebnisses hinnehmen zu müssen. Deutlich wird das, wenn statt der bislang beschriebenen ML-Verfahren neuronale Netze eingesetzt werden.

Bei neuronalen Netzen handelt es sich um eine vom menschlichen Gehirn inspirierte KI-Architektur.<sup>38</sup> Vorbild ist unter anderem die Verteilung auf eine Vielzahl kleiner „Rechenzentren“, den Neuronen, und deren Vernetzung. Außerordentlich komplex sind die Datenverarbeitung und Mustererkennung nicht nur, weil jedes einzelne Neuron, abhängig von der Art und Stärke des Inputs, unterschiedlichen Output

---

<sup>37</sup> Zu diesem Beispiel: Langenbucher et al., in: Avgouleas/Marjosola, Digital Finance in Europe: Law, Regulation, and Governance, 141.

<sup>38</sup> Mehlig, Machine Learning with Neural Networks, S. 5 ff; Russell et al., S. 801.



produziert. Zusätzlich erlaubt diese Art der KI, Verbindungen zwischen Neuronen herzustellen. Reagierend auf Feedback entstehen immer neue Verbindungen, das Netzwerk wird also zunehmend verzweigter und für den menschlichen Programmierer „undurchsichtiger“.<sup>39</sup>

Häufig wird bei neuronalen Netzen analog zum Gehirn<sup>40</sup> von drei Schichten („layers“) gesprochen. So unterscheidet man die der KI anfänglich zur Verfügung gestellten Daten („input layer“) von den durch die KI produzierten Ergebnissen („output layer“) und den Zwischenbereich, in welchem die KI die Rechenoperationen durchführt („hidden layer“).<sup>41</sup> Mit Blick auf diese Schichten (oder auch: die vielen Schritte, die ein Rechenweg durchläuft) spricht man von Deep Networks oder Deep Learning.<sup>42</sup>

Auch neuronale Netze können mit überwachtem, unüberwachtem und Lernen mit Verstärkung betrieben werden. Klassische Einsatzbereiche eines neuronalen Netzes mit überwachtem Lernen sind die Objekterkennung oder Übersetzungen. In diesen Fällen wird die KI trainiert, ein bestimmtes Label mit einzelnen Input-Elementen zu verbinden – wie im obigen Beispiel ein Bild mit dem Label „Wolf“.<sup>43</sup> Soll sie mit nicht etikettierten Daten arbeiten, lernt die als neuronales Netz organisierte KI ohne Überwachung und kann selbst Labels vergeben.<sup>44</sup> Wiederum nimmt das Lernen mit Verstärkung eine Art Zwischenrolle ein, bei welchem die KI zeitversetzt (nach Durchführung ihrer Aufgabe) eine Rückmeldung erhält.<sup>45</sup> Ein Einsatzbereich sind Navigationssysteme. Lernen mit Verstärkung bietet die Möglichkeit, der KI eine Rückmeldung zu geben, beispielsweise wenn ein schneller oder ein eher ressourcenschonender Weg gewünscht wird. Die KI optimiert dann die erstellten Vorschläge als Ergebnis eines Lernvorgangs auf positives oder negatives Feedback.<sup>46</sup>

#### d) Erklärbarkeit

Die Wahl zwischen herkömmlichen statistischen Verfahren, einer ML oder einem neuronalen Netz<sup>47</sup>

---

<sup>39</sup> Mehlig, S. 1.

<sup>40</sup> Ebd., S. 5 ff.

<sup>41</sup> Ertel, S. 321. Die Werte, die in den einzelnen „hidden layers“ errechnet werden, stellen jeweils unterschiedliche Abbildungen eines einzelnen inputs dar, Russell et al., S. 810 „In the process of forming all these internal transformations, deep networks often discover meaningful intermediate representations of the data. For example, a network learning to recognize complex objects in images may form internal layers that detect useful subunits: edges, corners, ellipses, eyes, faces—cats. Or it may not—deep networks may form internal layers whose meaning is opaque to humans, even though the output is still correct“.

<sup>42</sup> Russell et al., S. 801.

<sup>43</sup> Mehlig, S. 2.

<sup>44</sup> Ebd., S. 3.

<sup>45</sup> Zur Bedeutung der zeitversetzten Rückmeldung: Ebd., S. 4.

<sup>46</sup> Ebd., S. 2, 4.

<sup>47</sup> Zur Auswahlentscheidung Russell et al., S. 819 f.; Beispiele bei Mehlig, S. 3 ff.

hängt von einer Vielzahl von Faktoren ab. Dazu zählen die Art der verfügbaren Trainingsdaten,<sup>48</sup> die Kosten für deren Beschaffung, Zusatzkosten für Labeling, die Möglichkeit, klare Zielvorgaben für die KI zu formulieren,<sup>49</sup> aber auch das Interesse an Nachvollziehbarkeit des von einer KI gelieferten Ergebnisses.

Worin sich Statistik und KI mit Blick auf die Nachvollziehbarkeit – und damit auch auf die Erklärbarkeit – von Ergebnissen unterscheiden, wurde bereits deutlich. Die statistische Studie nimmt ihren Ausgang bei einer Hypothese und interessiert sich für den Einfluss jeweils konkret benannter Variablen. Die KI sucht stattdessen nach wiederkehrenden Datenmustern. Je komplexer und netzwerkbasierter die KI programmiert ist, desto undeutlicher wird der Einfluss einzelner Variablen und desto geringer die Erklärbarkeit.

Die Frage, welche Bedeutung dem nachvollziehenden und verstehenden Erklären zukommt, welches über Vorhersagen auf der Basis von Mustern hinausreicht, wird in den Naturwissenschaften mit Blick auf KI und Big Data seit längerem diskutiert.<sup>50</sup> Mit der Zunahme von KI als Grundlage von Entscheidungen, die unmittelbare Folgen für Verbraucher oder Staatsbürger zeitigen, sind Varianten desselben Grundproblems deutlich geworden. Geht es in den Naturwissenschaften um die theoriegeleitete Erklärung von Phänomenen, welche bloße Datenmuster nicht befriedigend liefern können,<sup>51</sup> steht im sozialwissenschaftlichen Kontext die Einordnung dieser Phänomene in einen Bedeutungszusammenhang in Rede. Werden Folgerungen aus Datenmustern in unternehmerische Entscheidungen eingebaut, handelt es sich hierbei unter anderem um menschliches Vertrauen auf die maschinell erzeugte Vorhersage.

Zwischen dem Entscheidenden und dem hiervon Betroffenen, beispielsweise der Bank und dem Bewerber um einen Kredit, kann die Erklärbarkeit des KI-Outputs bedeutsam sein. Sie hilft, die fremde

---

<sup>48</sup> Neuronale Netze erzielen insbesondere bei hochdimensionalen Daten Erfolge, Russell et al., S. 801. Diese definiert man als Datensätze, bei denen die Zahl der Merkmale höher ist als die Anzahl der Beobachtungen/Stichproben. Das verhindert den deterministischen Schluss von input auf output, siehe Narisetty in: Srinivasa Rao/Rao, Handbook of Statistics, 207 mit einem Beispiel zu genetischen Untersuchungen. Ein Datensatz mag nur einige hundert Patientenstichproben enthalten, zugleich aber über 100.000 Geninformationen pro Stichprobe. Die Abbildung der Stichprobe auf die Geninformation bereitet dann Probleme sowohl für KI, die auf linearer und logistischer Regression beruht, Russell et al., S. 801, als auch für KI, die mit Entscheidungsbäumen arbeitet, dies., S. 692 f.

<sup>49</sup> Hierzu Ertel, S. 321.

<sup>50</sup> Plakatv Anderson, The End of Theory: The Data Deluge Makes the Scientific Method Obsolete, Wired vom 23.6.2000, <https://www.wired.com/2008/06/pb-theory/>; am Beispiel der Neurowissenschaften kritisch: Krakauer et al., 93 Neuron (2017), 480; Larson, chapter 16.

<sup>51</sup> Krakauer et al., 93 Neuron (2017), 480, 481.

Entscheidung besser verstehen und infolgedessen beurteilen zu können, möglicherweise auch künftiges Verhalten anzupassen.<sup>52</sup>

Geht es um KI-generierte Vorhersagen zur Vorbereitung der eigenen Entscheidung, wie im Fall des Vorstandsmitglieds, der über die Schließung von Bankfilialen zu befinden hat, hängt die Bereitschaft zu einer Vertrauensinvestition in die KI nicht auf dieselbe Weise mit der Erklärbarkeit des KI-Outputs zusammen. Manch einer mag diese einfordern und nur unter dieser Voraussetzung bereit sein, die KI-Vorhersage bei seinen eigenen Überlegungen zu berücksichtigen. Manch anderer mag hierauf aber verzichten, etwa wenn er um die hohe Präzision von Black Box Vorhersagen weiß oder wenn Wettbewerber ähnliche Programme verwenden. Vertraut dieser auch ohne Nachvollziehbarkeit auf die KI, knüpft jener die Kooperation mit einer Maschine an die Erklärbarkeit einer Rechenoperation. Ob rechtliche Parameter diesen Entscheidungskontext prägen, soll im folgenden Abschnitt gefragt werden. Dabei wird unterschieden zwischen allgemeinen Sorgfaltspflichten, welche die Entscheidung über den Einsatz von KI insgesamt betreffen (unten II) und besonderen Pflichten mit Blick auf die Verwendung von KI bei der Leitungsentscheidung des Vorstands (unten III).

## **II. Allgemeine Sorgfaltspflichten: Der Einsatz von KI im Unternehmen**

Zu den selbstverständlichen Bestandteilen einer wohlinformierten Vorstandsentscheidung gehört es, sich über Vor- und Nachteile des KI-Einsatzes im Unternehmen zu informieren. Mit der Entscheidung für deren Einsatz zur Unterstützung von Mitarbeitern oder zur Automatisierung von Prozessen gehen eine Reihe von Organisationspflichten mit Blick auf Produkt- und Datenkontrolle, Ausbildung und fortlaufendes Training einher.

### **1. Art und Ort des Einsatzes von KI**

a) Soll KI eingesetzt werden („Ob“)?

Das betrifft zunächst die Frage, ob überhaupt eine KI verwendet werden soll. Hierfür sind die unterschiedlichen Unternehmensprozesse sorgfältig zu analysieren und zu entscheiden, an welcher Stelle ein Mehrwert geschaffen werden könnte.<sup>53</sup> Das beinhaltet eine Reihe komplexer Abwägungsfragen. Selbst wenn nämlich auf den ersten Blick ein geeigneter Einsatzort gefunden ist, ist der erhoffte Zugewinn an Produktivität gegenüber den weiter oben angedeuteten Limitierungen ins

---

<sup>52</sup> EBA Report on Big Data and Advanced Analytics 2020/01, S. 35 ff., A governance framework for algorithmic accountability and transparency, 2019, S. 31 ff.; HLEG on AI, Ethics Guidelines for Trustworthy AI, 2018, S. 14 ff.; Vredenburg, 30 Political Philosophy (2021), 209. Häufig wird “explainable AI” gemeinsam mit datenschutzrechtlicher Verantwortlichkeit erörtert: Kaminski, 92 Southern California Law Review (2019), 1529; Wachter et al., 2 Columbia Business Law Review (2019), 494; Wachter et al., 31 Harvard Journal of Law & Technology (2018), 842; Wachter et al., 7 International Data Privacy Law (2017), 76.

<sup>53</sup> Agrawal et al., S. 123 ff.

Verhältnis zu setzen. Steht hingegen nicht Produktivität, sondern Kontrolle im Fokus, können sich die Entscheidungsparameter umkehren und die Verwendung einer KI sich geradezu aufdrängen. Beispielhaft hierfür steht deren Einsatz in Compliance, Monitoring und Risikomanagementsystemen, § 91 Abs. 2, 3 AktG.

Auf die Bedeutung von Erklärbarkeit im konkreten Kontext und auf die Datenabhängigkeit wurde bereits hingewiesen. Nicht zu Unrecht hat man KI als „Induktionsmaschine“<sup>54</sup> bezeichnet. Sie lernt auf der Grundlage von Korrelationen, welche aus vorhandenem Datenmaterial erschlossen werden. Damit einher gehen Probleme, einen existierenden Wissensbestand („prior knowledge“) zu integrieren.<sup>55</sup> Eine für Menschen einfache Abstraktion lernt die KI erst nach zahlreichen Lernprozessen, die unter Umständen enormes Datenmaterial und feine Präzisierungen innerhalb der Daten erfordern.<sup>56</sup> Tatsächliche Veränderungen, neuartige, ungewöhnliche oder seltene Situationen, technische Innovationen oder Veränderungen menschlicher Präferenzen kommen aus demselben Grund bei einer KI gleichsam zeitversetzt an, möglicherweise auch überhaupt nicht.<sup>57</sup> Auch wenn dem Datenbestand eine kontrafaktische Entwicklung fehlt, leidet die Vorhersagekraft einer KI.<sup>58</sup>

Schließlich sind die Fehlerkosten zu evaluieren, die beim Einsatz von KI entstehen können. Dabei geht es beispielsweise um die Frage, wie hoch die Wahrscheinlichkeit ist, dass man sich vernünftigerweise auf deren Vorhersagen verlassen kann. Eine KI, die hierüber transparent Auskunft geben kann, mag man eher geneigt sein einzusetzen als eine Black Box KI. Das gilt erst recht für LLMs, etwa zur Zusammenfassung von Studien oder zur Beantwortung von Fragen. Deren Verwender geht mitunter hohe Risiken ein, wenn die KI sozusagen „im Brustton der Überzeugung“ eine unzutreffende Auskunft gibt.<sup>59</sup>

b) Welche Art KI soll eingesetzt werden („Wie“)?

Ist ein Potential für Mehrwert identifiziert, reichen mögliche Einsatzformen von vollständig antizipierten Entscheidungen über eine Kooperation von KI und Mensch bis zur Delegation von Entscheidungen unter Unsicherheit an die KI.

Vollständig antizipierte Handlungen sind im Bereich Robotics vorstellbar, wenn Fertigungsstraßen mit KI-gesteuerten Maschinen laufen, aber auch wenn der in einer Kunden-Hotline eingesetzte Chatbot

---

<sup>54</sup> Larson, chapter 10 (“induction engines”).

<sup>55</sup> Marcus, Deep Learning: A Critical Appraisal, S. 11 f.

<sup>56</sup> Ebd., S. 7. Beispiel für die Notwendigkeit präziser Daten bei ders., S. 10.

<sup>57</sup> Marcus, S. 9: deep learning language models repräsentieren Sätze als Abfolge von Worten, während Linguisten (grundlegend: Chomsky) die hierarchische Struktur von Sprache (zB Haupt- und Nebensätze) betonen; griffig: Lake et al., Still not systematic after all these years.

<sup>58</sup> Beispiel bei Langenbucher, Consumer Credit in the Age of AI, S. 12, 15.

<sup>59</sup> Ein Grund für derartige „Halluzinationen“ kann sein, dass die KI keinen Abgleich mit kuratierten Daten vornimmt oder bei ihrer Programmierung Kohärenz der gegebenen Antworten besonders hoch bewertet wird.

bekannte Fragen beantwortet. Ebenso gut lässt sich vorstellen, dass die KI unbekannte Fragen an einen menschlichen Mitarbeiter weiterleitet und in diesem (limitierten) Sinne mit ihm kooperiert. Um eine Kooperation handelt es sich auch, wenn die KI eine Kreditentscheidung vorbereitet, indem sie verfügbare Daten über den Bewerber auswertet, eventuell auch einen Entscheidungsvorschlag unterbreitet, den der menschliche Mitarbeiter weiter überprüft. Noch eigenständiger agiert die KI, wenn ihr Entscheidungen unter Unsicherheit überantwortet werden, etwa ein Kleinkredit ohne menschliche Kontrolle vergeben oder abgelehnt wird.

## 2. Begleitende Kontrolle

Wie bei jedem Einsatz einer Maschine oder Technik ist die Art und Weise der Verwendung von KI ein Teil der Organisationsverantwortung des Vorstands. Zu deren sorgfältiger Wahrnehmung zählt die Eignung des ausgewählten Produkts, dessen fortlaufende Kontrolle und die nachträgliche Produktbeobachtung. Kauft ein Unternehmen die eingesetzte KI bei Dritten ein, ist die Anpassung an die konkrete Umgebung im Unternehmen zu überprüfen.<sup>60</sup> Typische Defizite von KI, beispielsweise Probleme der Codierung bekannten Wissens<sup>61</sup> oder abduktive Schlussfolgerungen<sup>62</sup> sind durch wiederkehrende Überprüfung im Blick zu behalten.

Ergänzend zur Produktkontrolle tritt beim Einsatz von KI die Datenkontrolle. Dabei kann es sich um eigene Daten des Unternehmens handeln, welche zum Training der KI verwendet oder dieser zur Verfügung gestellt werden, damit sie anhand der konkreten Umstände lernt. Für unternehmenseigene Daten ist die Eignung für den vorgesehenen Zweck kritisch zu prüfen. Das umfasst die Kontrolle der notwendigen Quantität von Daten, die Entscheidung über deren Aufbereitung und Kuratierung samt der Frage, ob ein synthetischer Datensatz besser geeignet sein könnte. Weitere Kontroll- und Überwachungspflichten können hinzutreten, wenn die KI auf offene oder halb offene Daten, etwa aus dem Internet, aus Netzwerken mit Zugangsschranken und ähnlichem zugreifen soll. Dabei kann es um die inhaltliche Korrektheit erfasster Daten gehen, mitunter werden aber auch Fehler in der Codierung oder die Korrektur von Ergebnissen erforderlich sein. Letztere kann faktische Fehler betreffen, etwa wenn Umweltveränderungen nicht (oder nicht rasch genug) berücksichtigt werden. Normative Korrektur kann erforderlich werden, wenn die KI rechtlich unzulässige oder unerwünschte Handlungsempfehlungen gibt, beispielsweise diskriminierende Vorschläge im Personalbereich.

Wie bereits die Entscheidung über den Einsatz ist auch die Intensität von Kontrollpflichten davon abhängig, in welchem Bereich die KI eingesetzt wird und welches Risiko damit einhergeht. Für die Produkt- und Datenkontrolle enthält der Entwurf für eine EU-Verordnung zu KI ein fein differenziertes,

---

<sup>60</sup> Hacker, NJW 2020, 2142, 2145.

<sup>61</sup> Marcus, S. 11 f.

<sup>62</sup> Larson, chapter 12.

an Risikogesichtspunkten orientiertes Pflichtenprogramm. Die DSGVO tritt mit Blick auf die Speicherung und Verarbeitung von Daten hinzu.

### 3. Kalibrierung von Vertrauen<sup>63</sup> und Training

Zu den situativen, an den konkreten Einsatz der KI angepassten Sorgfaltspflichten des Vorstands zählt weiter die Koordination und Überwachung von Trainingsmaßnahmen mit Blick auf Mitarbeiter, von denen der Einsatz oder die Kontrolle von KI bzw. die Kooperation mit dieser erwartet wird. Dabei geht es um Basistraining, wie dies jede neue Maschine oder Technik erfordert, um deren Funktionsweise und Risiken korrekt einordnen zu können. Eigene Herausforderungen gerade des Einsatzes von KI, insbesondere wenn diese in menschliche Entscheidungsfindung inkorporiert werden soll, treten hinzu.

Das betrifft insbesondere das viel diskutierte verhaltenspsychologische Phänomen der „algorithm aversion“ und sein Gegenstück, die „algorithm appreciation“. Wiederum erhellt der Vergleich mit dem Einsatz einer neuen Maschine, worum es hierbei geht. Manch einer mag der KI kritisch gegenüberstehen und deren Einsatz ablehnen, andere mögen deren Beitrag besonders positiv einschätzen und persönliche Fragen oder Zweifel zurückstellen. Für den Vorstand übersetzt sich dieses Phänomen in die Pflicht, geeignete Strukturen für die Zusammenarbeit von Mensch und Maschine zu schaffen, um das Potential der KI heben zu können.<sup>64</sup> Die richtige Balance zwischen Vertrauen und Skepsis zu finden ist eine Schulungs- und Organisationsaufgabe, die von der konkreten Aufgabe und der Art der Einbindung der KI ebenso abhängt wie von der Expertise des Mitarbeiters.<sup>65</sup> In einer Studie hielten etwa erfahrene Radiologen einen (inhaltlich identischen) Vorschlag für qualitativ weniger überzeugend, wenn er als Ergebnis einer KI präsentiert wurde, ordneten ihn hingegen als korrekt ein, wenn er als Aussage eines Kollegen erschien.<sup>66</sup> Die Berücksichtigung solcher Effekte als Aufgabe des Verwenders einer KI verlangt Art. 14 Abs. 4 lit. b der KI-Verordnung ausdrücklich.

Derartigen Phänomenen mag die Verwendung einer erklärbaren KI entgegenwirken. Wiederum ist allerdings das konkrete Entscheidungsumfeld zu bedenken. Neuere Studien heben etwa hervor, dass das Angebot von Erklärbarkeit keineswegs stets genutzt wird. Dem liegt nicht nur Bequemlichkeit der Anwender zugrunde. Eine nähere Auseinandersetzung mit der Erklärung KI-generierter Ergebnisse kann mitunter als Einschränkung eigener Entscheidungsmacht verstanden oder als nicht kompatibel mit den kognitiven Prozessen des Entscheidenden erlebt werden.<sup>67</sup>

---

<sup>63</sup> Zu diesem Begriff: Miller, Explainable AI is Dead, Long Live Explainable AI, S. 2; Muir, 27 International Journal of Man-Machine Studies (1987), 527.

<sup>64</sup> Zum Beispiel aus der Radiologie: Gaube et al., 31 Digital Medicine (2021), 1.

<sup>65</sup> Ebd., S. 1, 4 ff. zu beispielhaften Maßnahmen.

<sup>66</sup> Im Gesamtergebnis differenzierend Gaube et al, S. 4.

<sup>67</sup> Miller, S. 4 empfiehlt die Entwicklung eines Empfehlungs- statt eines Erklärungssystems.

Auch der Einsatz eines menschlichen Mitarbeiters („human oversight“)<sup>68</sup> mag hilfreich für die erfolgreiche Integration von KI in Unternehmensprozess sein. Auf diese Weise können Mitarbeiter beim unternehmensinternen Einsatz von KI begleitet, zusätzlich auch bestimmte Prozesse oder Entscheidungen gegenüber Unternehmenskunden erläutert werden. Beispielhaft sei erneut die Entscheidung einer Bank über die Kreditwürdigkeit eines Kunden auf der Basis einer KI-generierten Beurteilung herangezogen. Ob und warum dieser im konkreten Fall zu folgen ist, muss abschließend ein Bankmitarbeiter beurteilen (sofern nicht vollständig automatisiert wurde). Die Akzeptanz insbesondere einer ablehnenden Entscheidung durch den Kreditbewerber wird häufig höher sein, wenn er eine Erklärung erhält, warum der Kredit nicht gewährt werden konnte. Ob und in welchem Umfang Art. 22 DSGVO eine solche Information verlangt, ist Gegenstand eines derzeit beim EuGH anhängigen Verfahrens.<sup>69</sup> Zusätzlich verlangt Art. 13 Abs. 2 lit. d der KI-VO für die Gebrauchsanweisungen von Hochrisikosystemen den Hinweis auf menschliche Aufsicht. Art. 14 Abs. 1 Abs. 1 KI-VO erfordert eine Konzeption, welche die wirksame Beaufsichtigung durch menschliche Personen erlaubt, um mit dieser einhergehende Fehler und Risiken zu minimieren, Art. 14 Abs. 2.

### **III. KI in der strategischen Leitungsentscheidung des Vorstands**

Bislang ging es um unterschiedliche Aspekte der Organisationsverantwortung des Vorstands beim Einsatz von KI im Tagesgeschäft des Unternehmens. Im folgenden Abschnitt steht die strategische Leitungsentscheidung<sup>70</sup> des Vorstands selbst im Fokus.<sup>71</sup> Dabei wiederholt sich in mancherlei Hinsicht das bereits erörterte Grundgerüst, wenn über „ob“ und „wie“ eines Einsatzes von KI zu entscheiden ist. Für die Entscheidungsfindung des Vorstands sehen allerdings die §§ 76 Abs. 1, 77 Abs. 1 S. 1, 93 Abs. 1 S. 1, 2 AktG jedenfalls in Andeutungen einen eigenen normativen Rahmen vor.

#### **1. Delegationsverbot**

Ob der Einsatz von KI im Kernbereich der Leitungsentscheidung des Vorstands zulässig ist, wird in der Literatur vor dem Hintergrund des Gebots eigenverantwortlicher Entscheidung samt des

---

<sup>68</sup> Zum Einsatz im Verhältnis von Staat und Bürger: Green, 45 Computer Law & Security Review (2022), 105681.

<sup>69</sup> EuGH, Rs. C-634/21.

<sup>70</sup> Unter den Begriff der Leitungsentscheidung fasst man für gewöhnlich die Unternehmensplanung, Steuerung, Organisation, Risikomanagement und Finanzierung, darüber hinaus aber auch einzelne, aktienrechtlich dem Vorstand als Leitungsaufgabe zugewiesene Aufgaben, vgl. §§ 124 Abs. 3 S. 1, 83, 90, 91, 92, 121 Abs. 2, 119 Abs. 2, 170 AktG, hierzu Fleischer, ZIP 2003, 1, 6.

<sup>71</sup> Zur Abgrenzung von Leitungs- und Geschäftsführungsaufgaben siehe Fleischer, ZIP 2003, 1; Freund, NZG 2015, 1419; Koch, 50 Jahre AktG, 2016, 65; Kuntz, AG 2020, 801; Seibt, in: FS K. Schmidt, 2009, 1463.

Delegationsverbots erörtert.<sup>72</sup> Mit Blick auf den Prozess der Entscheidungsfindung enthält § 76 Abs. 1 AktG den Grundsatz der Weisungsunabhängigkeit,<sup>73</sup> gemeinsam mit § 77 Abs. 1 S. 1 AktG die Gemeinschaftlichkeit als Normalfall<sup>74</sup> und die Betonung eines Kreises nicht delegierbarer Entscheidungen.<sup>75</sup> Zugleich ist unbestritten, dass Vorbereitung und Unterstützung der Entscheidungsfindung zulässig sind, gedacht wird dabei vor allem an nachgeordnete Mitarbeitererebenen, ggf. auch an von außen eingekauften Expertenrat.<sup>76</sup>

Daraus folgen einige Selbstverständlichkeiten. Leitungsentscheidungen lassen sich nicht vollständig antizipieren, sondern zeichnen sich durch kontext- und situationsabhängige Reaktion aus. Einer KI können sie folglich nicht zur vollautomatisierten Bearbeitung übertragen werden.<sup>77</sup> Ebenso deutlich ist, dass die vollständige Delegation von Leitungsentscheidungen an eine KI, die selbstständig Entscheidungen unter Unsicherheit trifft, das Verbot, Kernbereiche der Vorstandstätigkeit auszulagern verletzt.<sup>78</sup>

In Betracht kommt somit für Leitungsentscheidungen nur eine Kooperation von Mensch und Maschine. Damit ist allerdings noch nicht viel gewonnen, denn gerade in der Art und Weise kognitiver Kooperation liegt die eigentliche Herausforderung.

#### a) Die Illusion der Letztentscheidung

Zweifelhaft ist zunächst, ob sich trennscharf zwischen der (zulässigen) Delegation von Aufgaben, Informationsbeschaffung oder anderen Vorbereitungshandlungen und der (unzulässigen) Delegation von Entscheidungen differenzieren lässt.<sup>79</sup> Auf den ersten Blick erscheint es ohne weiteres möglich, eine KI mit der Vorhersage künftiger Entwicklungen, beispielsweise was Verbraucherpräferenzen oder geopolitische Entwicklungen betrifft, zu beauftragen und hiervon die Reaktion des Vorstands, die „eigentliche Entscheidung“, zu trennen, so dass der Vorstand „Herr des Verfahrens“<sup>80</sup> bleibt. Bei näherem Hinsehen ist diese Trennung allerdings keineswegs evident. Je klarer nämlich aus der

---

<sup>72</sup> Einschränkend Dubovitskaya et al., ZIP 2023, 63, 67 ff; eingehend Uwer, ZHR 183 (2019), 154, der Orientierung an geltenden gesellschaftsrechtlichen Grundsätzen zum outsourcing für ausreichend hält; abwägend: Möslin in: Barfield/Pagallo, Research Handbook Artificial Intelligence, S. 649, 658 ff.; zum US-amerikanischen Pendant in Delaware: Bruner, WP University of Georgia 2021-23, S. 9; Petrin, Columbia Business Law Review 2019, 965, 1014 ff.

<sup>73</sup> Fleischer in: BeckOGKActG, Stand: 01.04.2023, § 76 AktG Rn. 66 f.

<sup>74</sup> Ebd., Rn. 8 („Gesamtleitung“), 71; ders., ZIP 2003, 1, 2.

<sup>75</sup> Fleischer in: BeckOGKActG, Stand: 01.04.2023, § 76 AktG Rn. 9, 65 ff.

<sup>76</sup> Ebd., Rn. 20, 74.

<sup>77</sup> Möslin, ZIP 2018, 204, 209.

<sup>78</sup> Fleischer in: BeckOGKActG, Stand: 01.04.2023, § 76 AktG Rn. 78.

<sup>79</sup> In diese Richtung ebd., Rn. 74, 77; Grigoleit in: Grigoleit AktG § 76 AktG Rn. 87; Möslin, ZIP 2018, 204, 208; Noack, ZHR 183 (2019), 105, 119.

<sup>80</sup> So Möslin, ZIP 2018, 204, 208.



Vorbereitung der eigentliche Inhalt der Entscheidung folgt, desto eher wird die Leitungsentscheidung reiner Nachvollzug des von der KI Vorgeschlagenen.<sup>81</sup>

Der Überlagerung der eigenverantwortlich entwickelten Entscheidung wirkt der Aufruf an Vorstandsmitglieder, Technikverständnis aufzubauen und die Eigenlogik von Algorithmen zu begreifen, entgegen.<sup>82</sup> Ob sich auf diese Weise dem in zahlreichen Studien nachgewiesenen Phänomen des „algorithm appreciation“ wirksam gegensteuern lässt, steht allerdings nicht fest. Hinzu kommt, dass weder ein Nachfrageritual noch hausgemachte Intuition ohne weiteres geeignete Gegenmaßnahmen darstellen. Gerade in der Beurteilung statistischer Wahrscheinlichkeiten (und damit einem Kernelement von KI) unterliegen Menschen bekanntlich vielfältigen Entscheidungsanomalien,<sup>83</sup> das gilt auch für Experten.<sup>84</sup> Damit entsteht die Gefahr, dass eine Trennung zwischen Vorbereitung und Letztentscheidung entweder überflüssig ist oder mitunter zu qualitativ schlechteren Ergebnissen führen würde.<sup>85</sup>

#### b) Kooperationsverantwortung

Nähern sollte man sich der Einbindung von KI in Vorstandsentscheidungen vor dem Hintergrund des mit dem aktienrechtlichen Delegationsverbot verfolgten Sinn und Zweck. Einschlägig ist dabei zunächst das von Fleischer als „Pflichtrecht“ bezeichnete Element:<sup>86</sup> Die Obliegenheit, im Interesse der Eigentümer (und ggf. weiterer Stakeholder) die dem Vorstand in seiner Gesamtheit übertragene Leitungsaufgabe eigenverantwortlich auszufüllen. Anders gewendet: es darf nicht die „Flucht“ aus der Verantwortung angetreten werden. Besonders intensiv soll diese Pflicht in einem Kernbereich eigentlicher Leitungsaufgaben ausfallen, dessen Definition allerdings noch nicht trennscharf gelungen ist.<sup>87</sup> Dem Gesetzgeber scheint mithin daran gelegen zu sein, Verantwortlichkeit festzuschreiben, mithin ein Organ zu identifizieren, welches gleichsam „den Kopfinhält“, wenn es schief geht.

Erblickt man hierin den eigentlichen Grund des Delegationsverbots, erscheinen weder das Einarbeiten von Vorstandsmitgliedern in die Codierung der AI noch deren Erklärbarkeit (im oben angedeuteten, technischen Sinne) unverzichtbar.<sup>88</sup> Zentral ist stattdessen die Übernahme der Verantwortung für die

---

<sup>81</sup> Zur „foundational nature“ von Vorhersagen für Entscheidungen: Agrawal et al., S. 29.

<sup>82</sup> Fleischer, in: BeckOGKActG, Stand: 01.04.2023, § 76 AktG Rn. 78; Möslein, ZIP 2018, 204, 208.

<sup>83</sup> Kahneman, Thinking, fast and slow, S. 411, mit Hinweis auf de-biasing im Unternehmenskontext auf S. 417; Tversky/Kahneman, 90 Psychological Review (1983), 293, 299; dies. Science 1974, 185 (4157), 1124.

<sup>84</sup> Beispiele bei Agrawal et al., S. 55 ff.

<sup>85</sup> Eine Verletzung des Delegationsverbots halten manche Autoren nur dann für ausgeschlossen, wenn ausschließlich erklärbare KI eingesetzt wird, siehe Dubovitskaya et al., ZIP 2023, 63, 67 ff.; aA. Kremer et al., VGR-Sonderheft 2024 (im Erscheinen).

<sup>86</sup> Fleischer, ZIP 2003, 1.

<sup>87</sup> Ebd., 1, 2.

<sup>88</sup> Ähnlich Zetsche, AG 2019, 1, 7.

gelingende Kooperation zwischen Vorstand und KI.<sup>89</sup> Das Delegationsverbot überschneidet sich deshalb mit den an Vorstandsmitglieder gestellten Haftungserwartungen. Es ist nicht verletzt, wenn der sorgfaltsgemäß handelnde Vorstand die Verantwortung für die in Kooperation mit der KI getroffenen Entscheidung übernimmt.

#### aa) Datenqualität

Die Ausprägung dieser Kooperationsverantwortung ist kontextabhängig und die Entscheidung über die Art und Weise, sie zu erfüllen, liegt im Ermessen des Vorstands. Das betrifft die bereits angesprochenen Sorgfaltspflichten unter Einschluss der oben vermerkten Produkt- und Datenkontrollpflichten.<sup>90</sup> Nicht anders als beim Einsatz von KI im Tagesgeschäft des Unternehmens ist für die Leitungsentscheidung zu fragen, ob Bereiche identifiziert wurden, bei denen KI einen Mehrwert verspricht, die Relevanz von Erklärbarkeit, Datenabhängigkeit<sup>91</sup> und Fehlerkosten ist einzuschätzen.

Nicht hinreichend kann die Qualität der Vorhersage außerdem ausfallen, wenn sie von unbekanntem Daten („unknowns“) abhängt, und zwar unabhängig davon, ob deren Existenz bekannt oder unbekannt ist.<sup>92</sup> In manchen Fällen führt sogar das Vorhandensein bekannter Informationen („knowns“), welche die Maschine inkorrekt verarbeitet, zu einer irrtumsbehafteten Prognose.<sup>93</sup>

#### bb) Stärken des menschlichen Vorstandsmitglieds

Gelingende Kooperation setzt neben einem (generalisierten) Bewusstsein von Schwächen des „KI-Gesprächspartners“ eine klare Vorstellung der Stärken des menschlichen Vorstandsmitglieds voraus. Häufig handelt es sich dabei um die Behandlung derjenigen Situationen, in welchen das vorhandene Datenmaterial das Risiko unzutreffender Vorhersagen birgt. So ist beispielsweise im Angesicht eines Kriegs oder einer Pandemie das menschliche Vorstandsmitglied aufgerufen, Analogien zu suchen oder Modellentwürfe zu konzipieren.<sup>94</sup> Umbrüche, die von technischen Innovationen über neue Kundenpräferenzen bis zu veränderten Rahmenbedingungen gesellschaftlichen Zusammenlebens reichen, sind zu adressieren. Unerwartete Muster, welche die KI identifiziert, müssen beurteilt werden. Mitunter mag der Mehrwert der menschlichen Entscheidung auch darin liegen, gerade in der statistisch unwahrscheinlichen Entwicklung eine unternehmerische Chance zu erkennen.

Hinzu treten Fallgestaltungen, in welchen die Vorhersage zwar zutrifft, die von der KI vorgeschlagene Entscheidung aber Gesetzesnormen verletzen oder Reputationsrisiken bergen würde. Ein bestimmtes

---

<sup>89</sup> Zu „kooperativer Intelligenz“: Herberger, NJW 2018, 2825, 2828 (allerdings in anderem Kontext).

<sup>90</sup> S. o. unter II.

<sup>91</sup> Agrawal et al., S. 98 ff. zur Verwertbarkeit von „prediction with little data“.

<sup>92</sup> Agrawal et al., S. 59 ff.

<sup>93</sup> Hierzu Agrawal et al., S. 61 mit der Erklärung von reverse causality und omitted variables Effekten.

<sup>94</sup> Agrawal et al., S. 98 ff. zu den komparativen Vorteilen menschlicher Entscheidungsfindung mit Blick auf Modellbildung und Analogien; Larson, chapter 12-14 zur bislang unmöglichen Codierung abduktiver Schlussfolgerungen.

Scoring-Verfahren mag mit Diskriminierungs- oder Datenschutzrisiken einhergehen, welche die „normativ neutrale“ KI nicht berücksichtigt.

#### cc) Kognitive Kompatibilität<sup>95</sup>

Eigene Herausforderungen für die erfolgreiche Zusammenarbeit von Mensch und Maschine entstehen, weil die menschliche Kognition anderen Mustern folgt als die maschinelle. Dem Risiko einer Überschätzung der vermeintlich unfehlbaren Maschine<sup>96</sup> wirken die bereits vorgestellten Pflichten entgegen. Ebenso bedeutsam ist das gegenläufige Phänomen einer Ablehnung KI-generierter Prognosen.<sup>97</sup> Der Grund hierfür können menschliche Präferenzen für soziale Interaktion statt maschineller Beratung sein,<sup>98</sup> für ein diskursives „Vor und Zurück“ statt eines fertig präsentierten Vorschlags<sup>99</sup> oder die Überschätzung der eigenen Marktkenntnis und Erfahrung („overconfidence“).<sup>100</sup> Ein Vorstandsmitglied mag sich in seiner Autonomie eingeschränkt fühlen oder in einer wichtigen Strategieentscheidung einen schrittweise ablaufenden Gedankenprozess einfordern.<sup>101</sup>

Noch intrikater ist der vernünftige „Einbau“ des maschinellen Vorschlags in den menschlichen Entscheidungsprozess. Setzt etwa das Vorstandsmitglied dem KI-generierten Ergebnis die eigene Intuition entgegen, ersetzt zunächst einmal Konfrontation die Kooperation. Nimmt man hinzu, dass „overconfidence“ und „conservatism“ hartnäckige Verhaltensanomalien sind<sup>102</sup>, und gerade der Intuition bisweilen in § 93 Abs. 1 S. 2 AktG eine besondere Rolle zugestanden wird,<sup>103</sup> zählt die Auseinandersetzung mit diesem Spannungsverhältnis zu den zentralen Bausteinen der Kooperationsverantwortung.<sup>104</sup>

---

<sup>95</sup> Zu „cognitive compatibility“: Burton et al., 33 Journal of Behavioral Decision Making (2020), 220, 225.

<sup>96</sup> Zu algorithm appreciation siehe oben II. 3.

<sup>97</sup> Mit den Gründen hierfür beschäftigt sich eine eigene Forschungsrichtung innerhalb der verhaltenswissenschaftlichen Entscheidungstheorie, Metastudie bei Burton et al., 33 Journal of Behavioral Decision Making (2020), 220.

<sup>98</sup> Ebd., 220, 223, 224 für soziale Strukturen, die das fördern („existing research points out that organizational and social structures favor the expert intuiter over a cold algorithmic decision maker“).

<sup>99</sup> Miller, Explainable AI is Dead, Long Live Explainable AI, schlägt die Übernahme eines derartigen „vor und zurück“ für die Beratung durch KI vor, zur Kalibrierung von Vertrauen siehe bereits oben II. 3.

<sup>100</sup> Burton et al., 33 Journal of Behavioral Decision Making (2020), 220.

<sup>101</sup> Ebd., 220, 224: „human decision-makers often expect deliberation, a slow and effortful consideration of evidence in high-stake scenarios despite empirical findings suggesting that deliberation does not necessarily equate to better decision-making“.

<sup>102</sup> Ebd., 220, 225 f. mWN.

<sup>103</sup> Vetter, in: FS Bergmann, 2018, S. 827, 829, 836 f.

<sup>104</sup> Hierzu gehört auch die Frage, auf welcher Entscheidungstheorie die KI aufsetzt, siehe Burton et al., 33 Journal of Behavioral Decision Making (2020), 220, 226 zur Vorstellung von KI als „cognitive fix“ für klassische Verhaltensanomalien sowie der fehlenden Berücksichtigung alternativer Entscheidungstheorien (satisficing, ecological rationality) und dem Vorschlag „ecological rationality rather than probabilistic optimality“ anzustreben, hierzu im corporate governance Kontext Langenbacher, ZGR 2019, 717.

## 2. KI und Ision

Die Verschränkung zwischen dem Delegationsverbot des § 76 Abs. 1 AktG und den einer Kooperationsverantwortung inhärenten Sorgfaltspflichten wurde im vorangegangenen Abschnitt bereits deutlich. Eng mit der Delegation von Entscheidungsmacht hängt deshalb die Frage zusammen, inwieweit der Vorstand sich in der Sache auf die von einer KI erstellte Prognose verlassen darf. Im Gefolge insbesondere der Ision-Rechtsprechung des BGH ist Vertrauen auf sorgfältig ausgewählte externe Ratgeber im Grundsatz zu billigen, Auswahlermessen und Inhalt sind aber über eine Plausibilitätskontrolle des Vorstands abzusichern.<sup>105</sup>

Die in Ision aufgeworfene Frage ist verwandt, wenn auch nicht identisch, mit der Verwendung von KI. In Ision-Fällen hat der Vorstand keine eigene Sachkompetenz und ist folglich auf Expertenrat angewiesen. Bei KI-Input zur Unterstützung von Leitungsentscheidungen handelt es sich nicht notwendig um die Deckung eines Bedarfs an fachfremder Information, sondern zunächst einmal um die Verbesserung der Entscheidungsqualität beispielsweise durch schnellere und umfassendere Bearbeitung von Datenmaterial. Möchte der Vorstand hierauf zurückgreifen, sollte man ihm nicht entgegenhalten, dass bislang keine KI-Zertifizierungen existieren, die eine Gleichsetzung von Ision-Experten und KI erlauben.<sup>106</sup> Umgekehrt gilt: Sobald sich – insbesondere im Gefolge der KI-Verordnung – derartige Zertifizierungen etabliert haben, sind sie zu beachten.

Weil es sich um eine Verbesserung der originären Vorstandsentscheidung, nicht um die Beurteilung einer fachfremden Auskunft handelt, wird man beim Einsatz von KI gleichsam „erst recht“ eine Plausibilisierung fordern müssen.<sup>107</sup> Die weiter oben angestellten Überlegungen zur Kooperationsverantwortung legen dabei nahe, dass es weder um den inhaltlichen Nachvollzug der Datenarbeit noch um die Überprüfung der Codierung gehen kann. Stattdessen ist nach Einbruchstellen für die bekannten Schwächen KI-generierter Auskunft zu fragen<sup>108</sup> und diese mit den Stärken menschlicher Entscheidungsfindung abzugleichen. Einfangen lässt sich auf diese Weise auch ein Nebeneffekt der Plausibilitätskontrolle, der mit der haftungsrechtlichen Anreizsituation zu tun hat. Weder darf die Einholung von Ision-Expertenrat strategisch erfolgen,<sup>109</sup> um die präferierte Marschrichtung des Vorstands zu bestätigen, noch entlastet der Verweis auf ein gegriffenes KI-

---

<sup>105</sup> BGH NZG 2011, 1271; Fleischer, in: BeckOGKActG, Stand 01.04.2023, § 93 Rn. 44, 252, ders., NJW 2009, 2337, 2339; ders., ZIP 2009, 1397; ders., KSzW 2013, 3.

<sup>106</sup> Noack, ZHR 183 (2019), 105, 119; Zetsche, AG 2019, 1, 8 zur Frage, ob IT (nicht KI) als „Experte“ anerkannt werden könnte (beide zweifelnd).

<sup>107</sup> Zur Übertragbarkeit der Grundsätze Fleischer, Der Aufsichtsrat 2018, 121 (mit Blick auf Entscheidungen des Aufsichtsrats); Zetsche, AG 2019, 1, 8; zurückhaltend Noack, ZHR 183 (2019), 105, 119.

<sup>108</sup> So auch Zetsche, AG 2019, 1, 8.

<sup>109</sup> Fleischer in: BeckOGKActG, Stand: 01.04.2023, § 93 AktG Rn. 44, 252 („opinion shopping“).

Abfrageergebnis, wenn nicht feststeht, dass es sich hierbei um eine (so oder ähnlich) wiederholbare Prognose handelt.<sup>110</sup>

### 3. Pflicht zum Einsatz einer KI?

Aus der Analyse der Kooperationsverantwortung folgt zwanglos, dass es im Ermessen des Vorstands liegt, Situationen und auch ganze Bereiche zu definieren, in welchen der Einsatz einer KI ausscheidet. So liegt es, wenn für die konkrete Entscheidung kein Mehrwert identifizierbar, die Prognose ohne Erklärbarkeit nicht verwertbar, das Datenmaterial zu dünn oder die Relevanz von Fehlerkosten zu hoch ist. Auch die Reaktion auf bereits eingetretene oder nur befürchtete Reputationsverluste kann dazu führen, etwa wenn nicht beherrschbare Diskriminierungs- oder Datenschutzrisiken im Raum stehen oder die Verletzung gesetzlicher Normen droht.

Komplexer ist die Beurteilung einer etwaigen Pflicht zum Einsatz von KI. Angesichts der vielfältigen Erscheinungsformen dessen, was sich als „KI“ bezeichnen lässt, kann derzeit nicht sinnvoll von einer generalisierten Einsatzpflicht gesprochen werden. Für Ausnahmefälle, in welchen das Geschäftsmodell selbst auf KI gründet, ergibt sich deren Einsatz von selbst, aufsichtsrechtliche Anforderungen im Finanzbereich verlangen ggf. detaillierte Vorkehrungen.<sup>111</sup>

Außerhalb derartiger Fälle stellen KI-Kooperationspflichten gleichsam das Gegenstück zur KI-Kooperationsverantwortung dar. Bei den Überlegungen zur Kooperationsverantwortung ging es um die Frage, wie weit der Vorstand seine Entscheidungsobliegenheit auf die KI auslagern darf. Bei KI-Kooperationspflichten geht es umgekehrt darum, ob die Verbesserung der Entscheidungsqualität durch eine KI so deutlich ist, dass ein Vorstand, der sie nicht konsultiert, pflichtwidrig handelt.

Steht fest, dass eine sachlich einschlägige KI im Unternehmen vorgehalten wird, sind deshalb die bei der Kooperationsverantwortung vorgestellten Fragen zu überprüfen: Lässt sich ein Mehrwert eindeutig identifizieren, ist Erklärbarkeit gegeben (oder nicht relevant), lässt die Datenlage auf verwertbare Ergebnisse hoffen und fallen Fehlerkosten in der Gesamtabwägung nicht ins Gewicht? Können diese Fragen bejaht werden, besteht grundsätzlich die Pflicht zur Konsultation der KI.<sup>112</sup> Fällt die Antwort

---

<sup>110</sup> Dasselbe dürfte für den Verweis auf alternative Kausalität gelten, etwa wenn anzunehmen ist, dass Menschen eine noch schlechtere Entscheidung als die KI getroffen hätten.

<sup>111</sup> Diff. Zetsche, AG 2019, 1, 8: keine Analogie zu § 80 Abs. 2 WpHG; hierfür aber Möslein, ZIP 2018, 204, 211; abschwächend ders., in: Wendland/Eisenberger/Niemann, Smart Regulation: Theorie- und evidenzbasierte Politik, S. 177, 182.

<sup>112</sup> Für eine Pflicht, KI einzusetzen: Mertens, WP Gent 2023-01 S. 17 ff.; jedenfalls nur bei sehr guter Begründung gegen Konsultation einer KI: Fleischer, Der Aufsichtsrat 2018, 121 (mit Blick auf Entscheidungen des Aufsichtsrats); ähnlich Noack, ZHR 183 (2019), 105, 121 f.; diff Zetsche, AG 2019, 1, 8: mit Blick auf Organisationspflicht: grds. frei, aber uU. wg. sektorieller Gesetzgebung erforderlich; tendenziell für eine Pflicht: Möslein, ZIP 2018, 204, 211; abschwächend aber jedenfalls derzeit: ders., in: Barfield/Pagallo, S. 649, 660 ff.; und ders., in: Wendland/Eisenberger/Niemann, S. 177, 182; unter US-amerikanischem Recht jedenfalls derzeit für plausibel, künftig ggf. für erforderlich gehalten von Bruner, WP University of Georgia 2021-23 S. 11.

uneindeutig aus, scheidet die Kooperationspflicht nicht ohne weiteres aus. Möglicherweise können etwa Erklärbarkeit erreicht, die Datenlage verbessert oder Fehlerkosten auf andere Weise reduziert werden. Damit sind bei KI-Kooperationspflichten Überlegungen einschlägig, die in anderem Zusammenhang zu Pflichten führen, auf dem Stand der Technik zu bleiben, Softwareaktualisierungen durchzuführen oder Marktstandards zu beachten. Wer von Selbstverständlichkeiten abweicht, trägt die Begründungslast. Eine Pflicht, KI routiniert zu befragen, um auf der sicheren Seite zu sein, besteht hingegen ebenso wenig wie eine Pflicht, aufzurüsten, ohne auf einen verlässlichen Erfahrungsschatz bauen zu können.

Handelt es sich bei der Leitungsentscheidung um eine unternehmerische Entscheidung, sind die Anforderungen an KI-Kooperationspflichten abgesenkt. Sie kommen nur in Betracht, wenn ohne den Einsatz einer KI das Niveau der „angemessenen Information“, § 93 Abs. 1 S. 2 AktG, nicht erreicht würde. Erneut verbietet sich eine schematische Antwort. Ob eine Kooperationspflicht besteht, hängt vom Kontext der unternehmerischen Entscheidung, der Art der in Betracht kommenden KI und den für die Kooperationsverantwortung ausgeführten Überlegungen ab. Je verlässlicher und schneller die von der KI zu erwartende Antwort für die konkrete Situation, desto eher besteht auch bei unternehmerischen Entscheidungen eine Pflicht zur Vorab-Abfrage einer KI.<sup>113</sup> Die Absenkung des Sorgfaltsstandards im Rahmen der „business judgment rule“ betrifft die menschliche Entscheidung. Eine Fortentwicklung zu einer „AI-judgment rule“, etwa für unerkannte und nach dem Stand der Technik unerkennbare Fehler der KI, kommt nach alledem nicht in Betracht.

#### **IV. Zusammenfassung und Ergebnis**

Der Beitrag stellt zunächst einige Grundformen von KI vor. Fokussiert wird auf ML und neuronale Netze, erläutert werden überwachtes, unüberwachtes und Verstärkungslernen.

Sodann kommen Sorgfaltspflichten des Vorstands, der über den Einsatz von KI entscheidet, zur Sprache. Unterschieden werden allgemeine Sorgfaltspflichten bei der Verwendung von KI und besondere Pflichten beim Einsatz von KI in der Leitungsentscheidung.

Allgemeine Sorgfaltspflichten betreffen das „ob“ und das „wie“ des KI-Einsatzes im Unternehmen. Ob sich dieser empfiehlt, hängt von der Identifikation von Mehrwert, von vorhandenem Datenmaterial, der Relevanz von Erklärbarkeit und von Fehlerkosten ab. Das Spektrum der Art und Weise des KI-Einsatzes reicht von vollständig antizipierten Handlungen bis zur eigenverantwortlichen Entscheidung der KI unter Unsicherheit. Der Schwerpunkt in der Praxis dürfte derzeit bei der Kooperation von Mensch

---

<sup>113</sup> Mit Blick auf Eröffnung des sicheren Hafens der business judgment rule ablehnend Zetzsche, AG 2019, 1, 9; ebenso jedenfalls derzeit: Möslein, in: Barfield/Pagallo, S. 649, 660 ff.

und KI liegen. Für den Vorstand entstehen dabei Pflichten zur Daten- und Produktkontrolle und zum Mitarbeitertraining, insbesondere was „algorithm aversion“ und „algorithm appreciation“ betrifft.

Kommt KI in der Leitungsentscheidung des Vorstands zum Einsatz, rückt die Beachtung des Delegationsverbots des § 76 Abs. 1 AktG in den Mittelpunkt des Interesses. Der Beitrag schlägt vor, den Kern des Verbots als Übernahme von Kooperationsverantwortung zu begreifen. Dazu zählt die Wahrung der allgemeinen Sorgfaltspflichten. Weiter erfordert die Leitungsentscheidung eine kritische Auseinandersetzung des Vorstands mit den komparativen Vorteilen menschlicher Intelligenz. Das betrifft beispielsweise Analogien und abduktive Schlüsse aber auch die Behandlung von Reputationsrisiken oder der Legalitätspflicht. Damit entsteht eine der Ision-Rechtsprechung ähnliche Plausibilisierungspflicht. Eine „AI judgment rule“ ist hingegen abzulehnen. Über Kooperationspflichten mit KI, etwa im Kontext des § 93 Abs. 1 S. 2 AktG, lässt sich nur einzelfallabhängig entscheiden.

## Literaturverzeichnis

*Agrawal, Ajay/Gans, Joshua/Goldfarb, Avi*: Prediction Machines. The Simple Economics of Artificial Intelligence, Boston 2018.

*Bender, Emily M./Gebru, Timnit/McMillan-Major, Angelina/Shmitchell, Shmargaret*: On the Dangers of Stochastic Parrots: Can Language Models Be Too Big?, FAccT 21, S. 610 ff.

*Bruner, Christophe M*, Artificially Intelligent Boards and the Future of Delaware Corporate Law, University of Georgia School of Law, Legal Studies Research Paper No. 2021-23; erhältlich unter: [https://papers.ssrn.com/sol3/papers.cfm?abstract\\_id=3928237](https://papers.ssrn.com/sol3/papers.cfm?abstract_id=3928237).

*Burton, Jason W./Stein, Mari Klara/Jensen, Tina Blegind*: A Systematic Review of Algorithm Aversion in Augmented Decision Making, 33 Journal of Behavioral Decision Making (2020), S. 220 ff.

*Bzdok, Danilo/Altman, Naomi/Krzywinski, Martin*: statistics versus machine learning, Nature Methods 15 (2018), S. 233 f.

*Dubovitskaya, Elena/Buchholz, Annika*: Die Geschäftsleitung und der Rat des Algorithmus, ZIP 2023, S. 63 ff.

*Ertel, Wolfgang*: Grundkurs Künstliche Intelligenz. Eine praxisorientierte Einführung, 5. Auflage, Wiesbaden 2021.

*Finnis, John*: Natural Law: The Classical Tradition, in: Coleman, Jules/Shapiro Scott (Hrsg.), The Oxford Handbook of Jurisprudence and Philosophy of Law. Oxford 2002, S 1 ff.

*Fleischer, Holger*: Zur Leistungsaufgabe des Vorstands im Aktienrecht, ZIP 2003, S. 1 ff.

*Fleischer, Holger*: Aktuelle Entwicklungen der Managerhaftung, NJW 2009, S. 2337 ff.

*Fleischer, Holger*: Vertrauen von Geschäftsleitern und Aufsichtsratsmitgliedern auf Informationen Dritter, ZIP 2009, S. 1397 ff.

*Fleischer, Holger*: Expertenrat und Organhaftung, KSzW 2013, S. 1 ff.

*Fleischer, Holger*: Algorithmen im Aufsichtsrat, Der Aufsichtsrat 2018, S. 121.

*Freund, Stefan*: Brennpunkte der Organhaftung. Anmerkungen der Praxis zur organrechtlichen Innenhaftung, NZG 2015, S. 1419 ff.

*Gaube, Susanne/Suresh, Harini/Raue, Martina/Merritt, Alexander/Berkowitz, Seth J./Lermer, Eva/Coughlin, Joseph/Guttag, John V./Colak, Errol/Ghassemi, Marzyeh*: Do as AI say: susceptibility in deployment of clinical decision-aids, Digital Medicine 31 (2021), S. 1 ff.

*Glimcher, Paul W.*: Decisions, Uncertainty, and the Brain. The Science of Neuroeconomics, Cambridge 2004.

*Green, Ben*: The flaws of policies requiring human oversight of government algorithms, Computer Law & Security Review 45 (2022) 105681, S. 1 ff.

*Grigoleit, Hans Christoph*: Aktiengesetz, 2. Auflage, München 2020.

*Hacker, Philipp*: Europäische und nationale Regulierung von Künstlicher Intelligenz, NJW 2020, S. 2142 ff.

*Hawkins, Jeff*: A Thousand Brains. A New Theory of Intelligence, New York 2021.



*Herberger, Maximilian*: „Künstliche Intelligenz“ und Recht – Ein Orientierungsversuch –, NJW 2018, S. 2825 ff.

*Hickmann, Eleanore/Petrin, Martin*: Trustworthy AI and Corporate Governance: The EU's Guidelines for Trustworthy Artificial Intelligence from a Company Law Perspective, EBOR 2021, S. 593 ff.

*Jordan, James/Houssiau, Florimond/Cherubin, Giovanni/Cohen, Samuel N./Szpruch, Lukasz/Bottarelli, Mirko/Maple, Carsten/Weller, Adrian*: Synthetic Data – what, why and how?, [https://royalsociety.org/-/media/policy/projects/privacy-enhancing-technologies/Synthetic\\_Data\\_Survey-24.pdf](https://royalsociety.org/-/media/policy/projects/privacy-enhancing-technologies/Synthetic_Data_Survey-24.pdf), 2022.

*Kahneman, Daniel*: Thinking, fast and slow, New York 2011.

*Kaminski, Margot E.*: Binary Governance: Lessons From the GDPR's Approach to Algorithmic Accountability, Southern California Law Review 92 (2019), S. 1529 ff.

*Koch, Jens*: Der Vorstand im Kompetenzgefüge der Aktiengesellschaft, in: Fleischer, Holger/Koch, Jens/Lutter, Marcus (Hrsg.), 50 Jahre Aktiengesetz, Berlin/Boston 2016, S. 65 ff.

*Krakauer, John W./Ghazanfar, Asif A./Gomez-Marin, Alex/Maclver, Malcom A./Poeppel, David*: Neuroscience Needs Behavior: Correcting a Reductionist Bias, Neuron 93 (2017), S. 480 ff.

*Kremer, Thomas/Langenbucher, Katja*: Einfluss von Digitalisierung und KI auf das Aktienrecht, VGR-Sonderheft 2023 (im Erscheinen).

*Kuntz, Thilo*: Leitungsverantwortung des Vorstands und Delegation, AG 2020, S. 801 ff.

*Lake, Brenden/Baroni, Marco*: Still not systemnetic after all these years: On the compositional skills of sequence-to-sequence recurrent networks, 2018, <https://openreview.net/forum?id=H18WqugAb>.

*Langenbucher, Katja*: Economic Transplants. On lawmaking for Corporations and Capital Markets, Cambridge 2017.

*Langenbucher, Katja*: Interdisziplinäre Forschung im Unternehmensrecht – auf dem Weg zu einer cognitive corporate governance?, ZGR 2019, S. 717 ff.

*Langenbucher, Katja*: Consumer Credit in the Age of AI, 2023, <https://www.ecgi.global/working-paper/consumer-credit-age-ai—beyond-anti-discrimination-law>.

*Langenbucher, Katja*: Diskriminierung bei der Vergabe von Verbraucherkrediten? Eine Anmerkung zu den Plänen des europäischen Gesetzgebers, BKR 2023, S. 205 ff.

*Langenbucher, Katja/Corcoran, Patrick*: Responsible AR Credit Scoring – A Lesson from Upstart.com, in: Avgouleas, Emiliós/Marjosola, Heikki (Hrsg.), Digital Finance in Europe: Law, Regulation, and Governance, Berlin/Boston 2022, S. 141 ff.

*Larson, Erik, J.*: The Myth of Artificial Intelligence. Why Computers Can't Think the Way We Do, Cambridge 2021.

*Marcus, Gary*: Deep Learning: A Critical Appraisal, 2018, <https://arxiv.org/pdf/1801.00631.pdf>.

*Mehlig, Bernhard*: Machine Learning with Neural Networks. An Introduction for Scientists and Engineers, Cambridge 2021.

- Mertens, Floris*: The use of artificial intelligence in corporate decision-making at board level: A preliminary legal analysis, WP Gent 2023-01, <https://financiallawinstitute.ugent.be/wp-content/uploads/2023/03/2023-01.pdf>.
- Miller, Tim*: Explainable AI is Dead, Long Live Explainable AI! Hypothesis-driven decision support, 2023, <https://arxiv.org/abs/2302.12389>.
- Möslein, Florian*: Robots in the Boardroom: Artificial Intelligence and Corporate Law, in: Barfield, Woodrow/Pagallo, Ugo (Hrsg.), Research Handbook on the Law of Artificial Intelligence, Northampton 2018, S. 649 ff.
- Möslein, Florian*: Digitalisierung im Gesellschaftsrecht: Unternehmensleitung durch Algorithmen und künstliche Intelligenz?, ZIP 2018, S. 204 ff.
- Möslein, Florian*: Künstliche Intelligenz in der aktienrechtlichen Organtrias, in: Wendland, Matthias/Eisenberger, Iris/Niemann, Rainer (Hrsg.), Smart Regulation: Theorie- und evidenzbasierte Politik, Tübingen 2023, S. 177 ff.
- Muir, Bonnie M.*: Trust between humans and machines, and the design of decision aids, 27 International Journal of Man-Machine Studies (1987), S. 527 ff.
- Narisetty, Naveen Naidu*: Bayesian model selection for high-dimensional data, in: Srinivasa Rao, Arni/Rao, C.R. (Hrsg.), Principles and Methods for Data Science, Handbook of Statistics Volume 43, Oxford 2020, S. 207 ff.
- Nath, Rajakishore*: Philosophy of Artificial Intelligence. A Critique of the Mechanistic Theory of Mind, Florida 2009.
- Noack, Ulrich*: Organisationspflichten und –strukturen kraft Digitalisierung, ZHR 183 (2019), S. 105 ff.
- Petrin, Martin*: Corporate Management in the Age of AI, Columbia Business Law Review 3 (2019), S. 965-1030.
- Rolffs, Mathias*: Kausalität und mentale Verursachung. Eine Verteidigung des nicht-reduktiven Physikalismus, in: Abhandlungen zur Philosophie, Berlin/Heidelberg 2023.
- Russel, Stuart/Norvig, Peter*: Artificial Intelligence: A Modern Approach, 4. Auflage, London 2021 (Global Edition).
- Seibt, Christoph H.*: Dekonstruktion des Delegationsverbots bei Unternehmensleitung, in: Bitter, Georg/Lutter, Marcus/Priester, Hans-Joachim/Schön, Wolfgang/Ulmer, Peter (Hrsg.), Festschrift für Karsten Schmidt zum 70. Geburtstag, Berlin/Boston 2009, S. 1463 ff.
- Spindler, Gerald*: Gesellschaftsrecht und Digitalisierung, ZGR 2018, S. 17 ff.
- Spindler, Gerald/Stilz, Eberhard*: BeckOGKAktG, Stand: 01.04.2023.
- Stiehl, Thomas/Marciniak-Czochra, Anna*: Intelligente Algorithmen und Gleichungen? – Eine Annäherung an die Intelligenz mathematischer Konzepte, in: Holm Hadulla, Rainer M./Funke, Joachim/Wink, Michael (Hrsg.), Intelligenz: Theoretische Grundlagen und praktische Anwendung, Band 6, Heidelberg 2021.
- Tversky, Amos/Kahneman, Daniel*: Extensional versus intuitive reasoning: The conjunction fallacy in probability judgment, 90 Psychological Review (1983), S. 293 ff.

*Tversky, Amos/Kahneman, Daniel: Judgment under Uncertainty: Heuristics and Biases: Biases in judgments reveal some heuristics of thinking under uncertainty, Science 1974, 185 (4157), S. 1124 ff.*The Role of Data in AI, Report for the Data Governance Working Group of the Global Partnership of AI, Edinburgh 2020, <https://gpai.ai/projects/data-governance/role-of-data-in-ai.pdf>.

*Uwer, Dirk: Outsourcing digitaler Funktionen, ZHR 183 (2019), S. 154 ff.*

*Vetter, Jochen: Intuition und Business Judgment, in: Dreher, Meinrad/Drescher, Ingo/Mülbert, Peter O./Verse, Dirk A. (Hrsg.), Festschrift für Alfred Bergmann zum 65. Geburtstag am 13. Juli 2018, Berlin/Boston 2018, S. 827 ff.*

*Vredenburg, Kate: The Right to Explanation, The Journal of Political Philosophy 30 (2021), S. 209 ff.*

*Wachter, Sandra/Mittelstadt, Brent: A Right to Reasonable Inferences. Re-Thinking Data Protection Law in the Age of Big Data and AI, Columbia Business Law Review 2 (2019), S. 494 ff.*

*Wachter, Sandra/Mittelstadt, Brent/Russel, Chris: Counterfactual Explanations Without Opening the Black Box: Automated Decisions and the GDPR, Harvard Journal of Law & Technology 31 (2018), S. 841 ff.*

*Wachter, Sandra/Mittelstadt, Brent/Floridi, Luciano: Why a Right to Explanation of Automated Decision Making Does Not Exist in the General Data Protection Regulation, International Data Privacy Law 7 (2017), S. 76 ff.*

*Zetzsche, Dirk: Corporate Technologies – Zur Digitalisierung im Aktienrecht, AG 2019, S. 1 ff.*