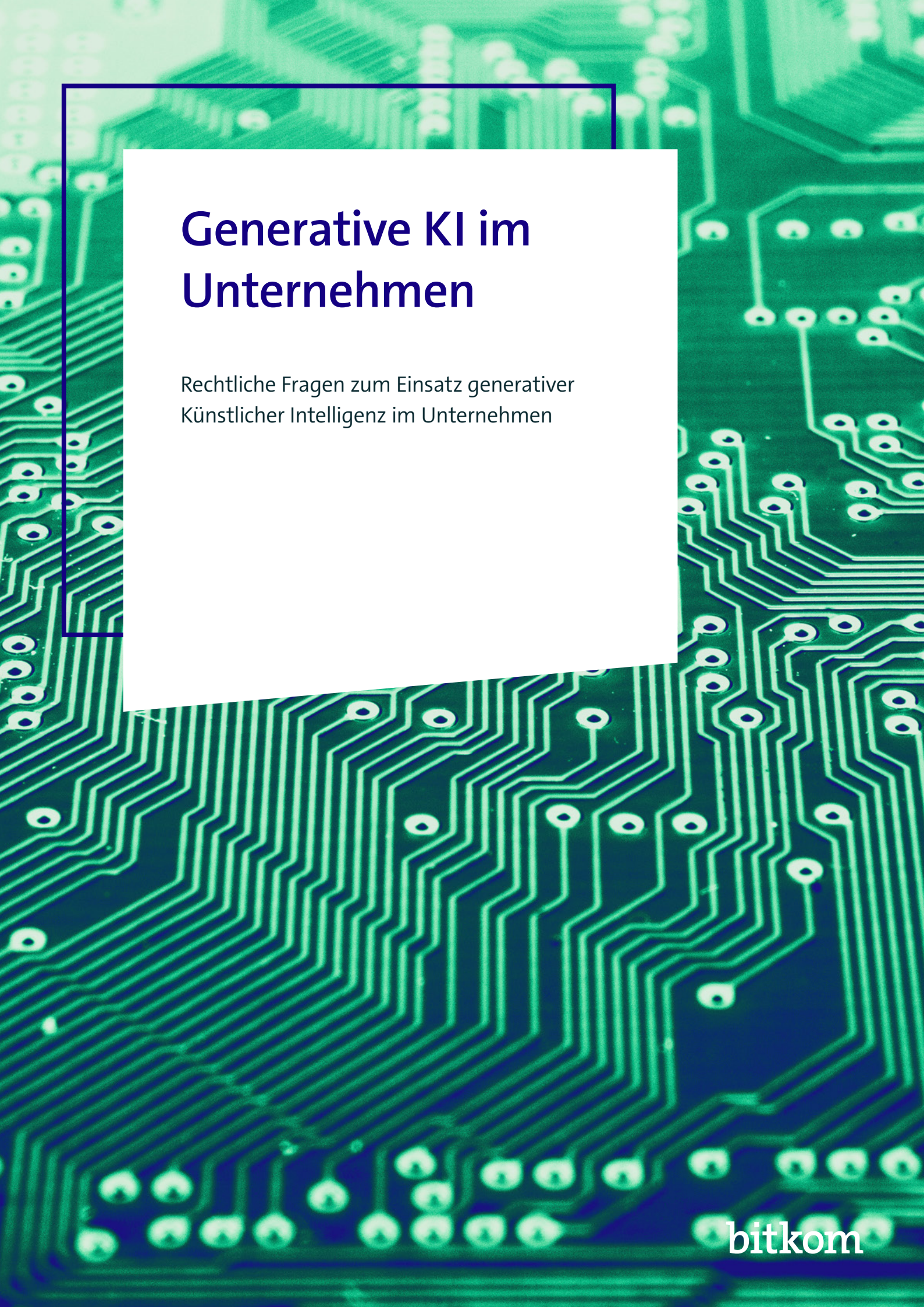

Generative KI im Unternehmen

Rechtliche Fragen zum Einsatz generativer Künstlicher Intelligenz im Unternehmen

bitkom

Herausgeber
Bitkom e. V.
Albrechtstraße 10
10117 Berlin
T 030 27576-0
bitkom@bitkom.org
www.bitkom.org

Ansprechpartner
Markus Scheufele | Leiter Recht
T 0 30 27576-154 | m.scheufele@bitkom.org

Verantwortliches Bitkom-Gremium
AK Recht im Unternehmen & Compliance

Layout
Jennifer Albrecht | Bitkom

Titelbild
© Fotolia_65509694_Subscription_XXL

Copyright
Bitkom 2025

Diese Publikation stellt eine allgemeine unverbindliche Information dar. Die Inhalte spiegeln die Auffassungen im Bitkom zum Zeitpunkt der Veröffentlichung wider. Obwohl die Informationen mit größtmöglicher Sorgfalt erstellt wurden, besteht kein Anspruch auf sachliche Richtigkeit, Vollständigkeit und/oder Aktualität, insbesondere kann diese Publikation nicht den besonderen Umständen des Einzelfalles Rechnung tragen. Eine Verwendung liegt daher in der eigenen Verantwortung der Leserin bzw. des Lesers. Die Haftung des Bitkom für Verletzungen von Leben, Körper und Gesundheit, für Schäden aus dem Produkthaftungsgesetz sowie für Schäden, die auf Vorsatz, grober Fahrlässigkeit oder aufgrund einer Garantie beruhen, ist unbeschränkt. Im Übrigen ist die Haftung des Bitkom ausgeschlossen. Alle Rechte, auch der auszugsweisen Vervielfältigung, liegen beim Bitkom.

Inhalt

Geleitwort 6

1 Einleitung 8

Ziel des Leitfadens 9

Definition: Generative KI 10

Technische Grundlagen generativer KI 12

Anwendungsbeispiele für generative KI in Unternehmen 13
Bildgenerierung 14
Audiogenerierung (Sprache und Musik) 14

2 Rechtliche Aspekte bei der Beschaffung von KI 15

Verteilung der Verantwortlichkeiten entlang der Wertschöpfungskette nach AI Act 16

Checkliste und Vorüberlegungen vor der Beschaffung von KI (IT Procurement) 19

3 Rechtliche Aspekte in der Einsatzphase 24

Datenverarbeitung 25
Pflichten nach der DSGVO 25

Cybersicherheit 35

Schutzrechte 37
Urheberrecht 37
Know-how-Schutz und Geschäftsgeheimnisschutz 46
Weitere Schutzrechte 51

Haftungsrechtliche Aspekte 53
Typische Risiken beim KI-Einsatz 53
Haftung des Anwenders nach derzeitigen Haftungsregelungen 53
Maßnahmen gegen Haftungsrisiken 58
Geplante Haftungsregelungen für KI-Anwender 58

Arbeitsrechtliche Aspekte 62
Individualarbeitsrecht 62
Betriebsverfassungsgesetz 66
Betriebssicherheitsverordnung 68
Arbeitsschutz bzw. Anwendung von KI und Datenschutz 68
Erstellung einer unternehmensinternen Richtlinie zur Nutzung generativer KI 70

Verträge und Willenserklärungen 72
Vertragsschluss und Loslösung von ‚KI'-Verträgen 72
Verträge (über den Einsatz von) KI 74

4 **Ethische Aspekte beim Einsatz von generativer KI** 76

Einleitung: KI-Ethik und die Verantwortung in der Knowledge Society 77
Begriffsbestimmung und Abgrenzung: KI-Ethik 78

Konkreter Kontext des Einsatzes generativer KI und Unterscheidung ethischer Einsatzszenarien für generative KI 80
Kundenumgang und Kundenerlebnis (UX, CX, HCX) 80
Mitarbeiterhilfen (EX) und Prozessunterstützung 81
Administrative Verwaltungsaufgaben mit GenAI:
Technische Organisation, HR, Betriebsrat, AI Agents (»AI Agents Ethics«) 81

Organisatorische Anforderungen (»Governance«) 83

Anwendungsregeln für generative KI 84
Klare Funktionen und Entscheidungsprozesse 84
Vertrauen schaffen 84
Nachvollziehbarkeit ermöglichen 85
Berücksichtigung unterschiedlicher Einsatzszenarien 85
Transparente Kommunikation 86
Umgang mit Komplexität und Verantwortung 86

Inhaltliche Anforderungen an KI-Ethik und ethische Bereiche 88
Ethische Kernprinzipien für GenAI 88
Regelungsgrundlagen 88
Andere Quellen 90

Exkurs: KI-Ethik in der EU und global: Der Kampf gegen Algorithmic Bias 92
Die EU: Regulatorische Vorreiterrolle 92
Globale Initiativen: Fragmentierte Ansätze mit gemeinsamen Zielen 93
Herausforderungen und kritische Perspektiven 93
Fazit: Von Kriterien zu verbindlicher Praxis 93

Exkurs: Ethische Verantwortung für psychologische Zusammenhänge 94
Nudging 94
Was ist Micro-Nudging durch GenAI? 94
Wie nutzt GenAI Opportunismus und Toleranzen aus? 95
Psychologische Manipulation: 95
Veränderung der Entscheidungsräume 95
Einschränkung der Autonomie 95
Personalisierte Empfehlungen 96
Transparenz und Kontrolle 96

Psychologische Hintergründe für Mitarbeiter 96
Kundenbetroffenheit 97
Lösungsansätze und ethische Gestaltung 97
Risiken und ethische Herausforderungen 97
Fazit Nudging 98

Konkrete Umsetzung in Unternehmen 99

Ausblick: KI-Ethik im Zeitalter exponentieller Dynamik 101

Geleitwort

Der vorliegende Leitfaden entstand in einer Taskforce, bestehend aus Expertinnen und Experten aus folgenden Arbeitskreisen des Bitkom:

- ↗Personal & Arbeitsrecht
- ↗Intellectual Property
- ↗Artificial Intelligence
- ↗Legal Tech
- ↗Sicherheitspolitik
- ↗Arbeit 4.0
- ↗Datenpolitik & Datenräume
- ↗Datenschutz
- ↗Vertrags- & Rechtsgestaltung
- ↗Recht im Unternehmen & Compliance
- ↗Arbeitssicherheit

Besonderer Dank gilt folgenden Personen, die mit ihrer Expertise und wertvollen praktischen Erfahrung die Publikation erstellt haben:

- Martin Agethen, Deutsche Telekom AG
- Lennart Bartl, Deutsche Bahn AG
- Prof. Dr. Christoph Bauer, ePrivacy GmbH
- Thomas Bauer, Atruvia AG
- Stefanie Bauer, ePrivacy GmbH
- Dr. Malte Baumann, Nordemann Czychowski & Partner Rechtsanwältinnen und Rechtsanwälte mbB
- Dr. Nadja Christe, Bayer AG
- Susan Bischoff, Morrison & Foerster LLP
- Dr. Christoph Cordes, YPOG Partnerschaft von Rechtsanwälten und Steuerberatern mbB Schnittker + Partner
- Jonas von Dall'Armi, Giesecke+Devrient GmbH
- Leon Feldmann, YPOG Partnerschaft von Rechtsanwälten und Steuerberatern mbB Schnittker + Partner
- Dr. Benedikt Flöter, YPOG Partnerschaft von Rechtsanwälten und Steuerberatern mbB Schnittker + Partner
- Dr. Anne Förster, Taylor Wessing Partnerschaftsgesellschaft von Rechtsanwälten und Steuerberatern mbB
- Dr. Alexander Fritz, OmegaLambdaTec GmbH
- Mareike Gehrmann, Taylor Wessing Partnerschaftsgesellschaft von Rechtsanwälten und Steuerberatern mbB
- Alexander Höcht, Fujitsu Technology Solutions GmbH
- Dr. Inka Knappertsbusch, CMS Hasche Sigle Partnerschaft von Rechtsanwälten und Steuerberatern mbB
- Dr. Benedikt Kohn, Taylor Wessing Partnerschaftsgesellschaft von Rechtsanwälten und Steuerberatern mbB
- Tobias Kutscheidt, Deutsche Telekom AG
- Dr.-Ing. Ingo Marko Wolf, Robert Bosch GmbH
- Andrea Martin, IBM Deutschland GmbH
- Dilan Mienert, GÖRG Partnerschaft von Rechtsanwälten mbB
- Dr. Dominik Rabe, Rewe Zentralfinanz e.G.
- Filipp Revinzon, Vay Technology GmbH
- Michael Schemel, UnternehmerTUM GmbH
- Dr. Eberhard Schnebel, Commerzbank AG
- Rene Schneider, DataGuard GmbH
- Tanja Schuchmann, Software AG
- Marie Slowioczek, ELEMENT Insurance AG
- Daniel Sohn, Giesecke + Devrient GmbH
- Dr. Gerald Spiegel, EnBW Energie Baden-Württemberg AG
- Christiane Stützle, Morrison& Foerster LLP
- Joerg Twellmeyer, Dell GmbH
- Dr. Maximilian Vonthien, LLM, CMS Hasche Sigle Partnerschaft von Rechtsanwälten und Steuerberatern mbB
- Dr. Nicolai Wiegand, Taylor Wessing Partnerschaftsgesellschaft von Rechtsanwälten und Steuerberatern mbB

1 Einleitung

1.1 Ziel des Leitfadens

Der vorliegende Praxisleitfaden »Generative KI im Unternehmen – rechtliche Fragen zum Einsatz generativer Künstlicher Intelligenz im Unternehmen« setzt sich mit den rechtlichen Herausforderungen und Aspekten bei der geschäftlichen Nutzung von generativer Künstlicher Intelligenz (KI) auseinander. Bevor wir tiefer in das Thema eintauchen, möchten wir nachfolgend klären, was genau unter »generativer KI« zu verstehen ist.

Generative KI ist ein Unterbereich der Künstlichen Intelligenz, der darauf ausgerichtet ist, neue Daten oder Inhalte zu generieren. Diese Technologie findet Anwendung in einer Vielzahl von Bereichen, einschließlich der Generierung von Texten, der Erstellung künstlicher Bilder oder Videos, automatischer Übersetzungen, und sogar in der Musikkomposition. Prominente Beispiele hierfür sind GANs (Generative Adversarial Networks) zur Erzeugung realistischer Bilder oder Textmodelle wie GPT-4 für die automatische Textgenerierung.

Bitte beachten Sie, dass sich dieser Leitfaden grundsätzlich auf generative KI konzentriert und nicht die allgemeinen rechtlichen Aspekte der Künstlichen Intelligenz behandelt. Die mit der Nutzung generativer KI verbundenen spezifischen Herausforderungen und Risiken sind komplex und verdienen eine eigenständige Betrachtung. Dennoch sind die Grenzen zu anderen KI-Technologien stellenweise fließend und insbesondere machen rechtliche Fragestellungen nicht an den Grenzen der Technologie Halt. Deshalb lässt es sich nicht vermeiden, rechtliche Fragen aufzuwerfen und zu beantworten, die nicht (nur) primär generative KI, sondern KI generell betreffen.

Die Anwendung von generativer KI bringt eine Reihe von rechtlichen Fragestellungen mit sich, die für Unternehmen und Nutzende von Bedeutung sind. Dies umfasst sowohl die Daten, die in die KI-Modelle eingegeben werden, als auch die Verwendung der von der KI generierten Ergebnisse. Relevante Themenbereiche sind dabei unter anderem die EU-KI-Verordnung und insbesondere deren Kapitel V Artikel 51 ff. (gelten ab 2. August 2025), geistiges Eigentum, insbesondere Urheberrecht, Datenschutz, Haftungsfragen und Arbeitsrecht.

Der vorliegende Praxisleitfaden hat das Ziel, Ihnen eine klare Übersicht über die rechtlichen Aspekte im Kontext generativer KI zu bieten. Zudem werden wir praktische Handlungsempfehlungen darlegen, wie mit diesen rechtlichen Herausforderungen umzugehen ist. Ein Exkurs zu ethischen Aspekten bei der Beschaffung und dem Einsatz von generativer KI rundet den Leitfaden ab.

Da der Bereich der generativen KI äußerst dynamisch ist, ist es unser Anspruch, diesen Leitfaden regelmäßig zu aktualisieren und an neue Erkenntnisse sowie Entwicklungen anzupassen. Der Stand des KI-Leitfadens ist Februar 2025. Insbesondere in Bezug auf die EU-KI-Verordnung (AI Act) bezieht sich dieser Leitfaden auf die finale Fassung der ↗KI-Verordnung vom 13.Juni 2024, Verordnung (EU) 2024/1689.Der Leitfaden richtet sich insbesondere an Unternehmen als Anwender von generativer KI. Der Fokus liegt also auf einer Anwenderperspektive in der Einsatzphase von generativer KI.

1.2 Definition: Generative KI

Künstliche Intelligenz ist ein Forschungsgebiet der Informatik. In dem Sinne existiert auch (noch) keine abschließende Definition von KI. Künstliche Intelligenz lässt sich dennoch als Simulation menschenähnlicher kognitiver Prozesse durch Computerprogramme bezeichnen. Anders als herkömmliche algorithmenbasierte IT-Systeme, die für sehr spezifische Aufgaben programmiert werden und nur auf festgelegten Regeln basieren, kann KI aus Daten lernen und ihre Leistung im Laufe der Zeit verbessern. Zum Beispiel kann ein herkömmliches algorithmenbasiertes System für die Lagerverwaltung zwar Bestandslisten aktualisieren, aber es kann nicht vorhersagen, welche Produkte in der nächsten Saison gefragt sein werden. Ein KI-System hingegen könnte durch die Analyse von zusätzlichen Verkaufsdaten, Wetterbedingungen und anderen Faktoren ziemlich genaue Vorhersagen treffen.

Der AI Act lehnt sich an die OECD-Definition an und definiert in Artikel 3 Nr. 1 ein KI-System als ein »maschinengestütztes System, das für einen in unterschiedlichem Grade autonomen Betrieb ausgelegt ist und das nach seiner Betriebsaufnahme anpassungsfähig sein kann und das aus den erhaltenen Eingaben für explizite oder implizite Ziele ableitet, wie etwa Vorhersagen, Inhalte, Empfehlungen oder Entscheidungen erstellt werden, die physische oder virtuelle Umgebungen beeinflussen können«.[1]

»Generative KI«, manchmal auch als GenAI bezeichnet, ist künstliche Intelligenz (KI), die als Antwort auf den Prompt (auch »Input«), also die Anfrage einer Nutzerin oder eines Nutzers, Inhalte wie Text, Bilder, Video, Audio oder Softwarecode generiert (auch als »Output« bezeichnet). Generative KI ist damit ein Teilbereich der KI, der darauf ausgerichtet ist, neue Inhalte oder Daten zu schaffen. Diese Art von KI unterscheidet sich von anderen Formen wie der »diskriminativen KI«, die hauptsächlich darauf trainiert ist, Unterschiede zwischen verschiedenen Arten von Daten zu erkennen. Ein Einsatzbereich ist zum Beispiel, Spam-E-Mails von legitimen E-Mails zu unterscheiden. Während diskriminative KI für Klassifizierungsaufgaben eingesetzt wird, ist generative KI darauf spezialisiert, originelle Daten zu erzeugen, sei es Text, Bilder oder sogar Musik.

Generative KI kann als ein spezialisierter Zweig innerhalb der General Purpose AI (GPAI) betrachtet werden. Während GPAI auf vielfältige Weise eingesetzt werden kann und oft als »Allzweck-KI« bezeichnet wird, fokussiert sich generative KI auf die Schaffung neuer Inhalte. Basismodelle sind oft die Grundlage für beide Typen und können entweder für generative oder diskriminative Aufgaben spezialisiert werden, je nachdem, wie sie trainiert wurden.

1 Die offizielle englische Definition lautet: »An AI system is a machine-based system designed to operate with varying levels of autonomy and that may exhibit adaptiveness after deployment and that, for explicit or implicit objectives, infers, from the input it receives, how to generate outputs such as predictions, content, recommendations, or decisions that can influence physical or virtual environments.«

Schließlich ist es wichtig, zwischen KI-Basistechnologie und KI-Systemen zu unterscheiden. KI-Technologie bezieht sich auf die Algorithmen und Modelle (z. B. Large Language Models), die die Grundlage für KI-Anwendungen (z. B. ChatGPT) bilden. Ein KI-System hingegen ist eine vollständige Lösung, die KI-Technologie enthält, aber auch andere Elemente wie Benutzeroberflächen, Datenmanagement-Tools und Anwendungslogik, also zusätzlich noch eine Applikationskomponente umfasst. Ein KI-System kann beispielsweise eine generative KI-Technologie nutzen, um personalisierte Inhalte für eine Marketingkampagne zu erzeugen, aber es würde auch Tools für die Verwaltung dieser Kampagne und die Analyse der Ergebnisse enthalten.

In der KI-Verordnung steht die Definition von »KI-Modell mit allgemeinem Verwendungszweck« in Artikel 3 Nr. 63 und legt fest: »ein KI-Modell — einschließlich der Fälle, in denen ein solches KI-Modell mit einer großen Datenmenge unter umfassender Selbstüberwachung trainiert wird —, das eine erhebliche allgemeine Verwendbarkeit aufweist und in der Lage ist, unabhängig von der Art und Weise seines Inverkehrbringens ein breites Spektrum unterschiedlicher Aufgaben kompetent zu erfüllen, und das in eine Vielzahl nachgelagerter Systeme oder Anwendungen integriert werden kann, ausgenommen KI-Modelle, die vor ihrem Inverkehrbringen für Forschungs- und Entwicklungstätigkeiten oder die Konzipierung von Prototypen eingesetzt werden«. Der Erwägungsgrund 99 der KI-Verordnung nennt konkrete Beispiele: »Große generative KI-Modelle sind ein typisches Beispiel für ein KI-Modell mit allgemeinem Verwendungszweck, da sie eine flexible Erzeugung von Inhalten ermöglichen, etwa in Form von Text-, Audio-, Bild- oder Videoinhalten, die leicht ein breites Spektrum unterschiedlicher Aufgaben umfassen können.«

Ein solches KI-Modell kann nicht direkt durch Nutzerinnen und Nutzer verwendet werden, sondern wird in aller Regel in ein KI-System eingebunden oder eingebaut (vgl. ↗Erwägungsgrund 97 der KI-Verordnung: »KI-Modelle sind in der Regel in KI-Systeme integriert und Teil davon.«).

Zur Abgrenzung des KI-Modells zum KI-System sagt Erwägungsgrund 97 der KI-Verordnung weiter: »Damit KI-Modelle zu KI-Systemen werden, ist die Hinzufügung weiterer Komponenten, zum Beispiel einer Nutzerschnittstelle, erforderlich.« Wenn ein solches KI-System dann nicht auf einen spezifischen Anwendungsfall ausgerichtet ist – und damit mehrere Anwendungsfälle haben kann, spricht die KI-Verordnung in Artikel 3 Nr. 66 von einem »KI-System mit allgemeinem Verwendungszweck«, »ein KI-System, das auf einem KI-Modell mit allgemeinem Verwendungszweck beruht und in der Lage ist, einer Vielzahl von Zwecken sowohl für die direkte Verwendung als auch für die Integration in andere KI-Systeme zu dienen«. Die Erwägungsgründe 100ff der KI-Verordnung erläutern das noch ergänzend, denn diese gewissermaßen mittlere Kategorie zwischen KI-Modell und KI-System war bis zuletzt sehr umstritten im Rahmen des finalen Trilogs zur Finalisierung der KI-Verordnung.
Generative KI umfasst damit regulatorisch KI-Modelle mit allgemeinem Verwendungszweck und KI-Systeme mit allgemeinem Verwendungszweck. Wenn diese noch mit einer Nutzeroberfläche versehen werden, handelt es sich dann bereits um ein KI-System, gegebenenfalls eines mit »allgemeinem Verwendungszweck«, das auch als »generatives KI-System« bezeichnet wird.

1.3 Technische Grundlagen generativer KI

Die Grundlage für jedes KI-System ist die Phase des Anlernens des integrierten KI-Modells mit Daten, was als »Training« bezeichnet wird. In dieser Phase wird ein KI-Modell mit einer großen Menge an Daten gefüttert, um bestimmte Muster oder Beziehungen in den Daten zu erkennen. Die Qualität dieser Datenbasis hat einen entscheidenden Einfluss auf die Qualität der Ergebnisse, die das KI-Modell später erzeugt. Obwohl sich dieser Leitfaden auf die Nutzungsphase konzentriert, ist es wichtig zu beachten, dass KI-Modelle auch in der Lage sein können, aus den Daten der Nutzungsphase zu lernen, abhängig davon, ob sie dynamisch oder statisch eingestellt sind – das ist abhängig von der das KI-Modell umgebenden Applikationssoftware, die eine solche Einstellung vorsieht. Im letzten Fall, der statischen Einstellung, lernen sie nichts mehr durch die Nutzung dazu. Der Speicherort dieser Datenbasis sowie der Kreis der Zugriffsberechtigten können rechtliche Implikationen haben, die in späteren Abschnitten dieses Leitfadens behandelt werden.

In der Nutzungsphase benötigt ein generatives KI-System Eingabeinformationen, sogenannte Prompts, um spezifische Aufgaben zu erfüllen. Diese Prompts können in Form von Textanweisungen erscheinen, aber auch durch Bilder (dann »multimodal«), Texte oder Audiodaten ergänzt werden. Als Ergebnis generiert das System entsprechende Texte, Bilder oder Audiodaten.

Generative KI-Systeme haben die Fähigkeit, in Sitzungen iterativ vorzugehen. Das bedeutet, dass sie in der Lage sind, auf der Grundlage von Rückmeldungen oder zusätzlichen Eingaben ihre Ergebnisse kontinuierlich zu verbessern und zu verfeinern.

Der Begriff der »Explainable AI« (auch »XAI«) beschreibt den Ansatz, KI-Entscheidungen nachvollziehbar und erklärbar zu gestalten. Dies ist besonders im Kontext des AI Acts für Hochrisikosysteme von Bedeutung. Allerdings stellt XAI für generative KI generell eine besondere Herausforderung dar, da generative KI nicht nach einem festgelegten Algorithmus arbeitet, sondern auf statistischen Methoden basiert. Oft ist es deshalb schwierig, genau zu erläutern, warum und wie ein bestimmtes Ergebnis aus den Eingabeinformationen entstanden ist. Die mehrmalige Eingabe desselben Prompts kann daher zu verschiedenen Ergebnissen, dem sog. Output, führen.

1.4 Anwendungsbeispiele für generative KI in Unternehmen

Generative KI bietet eine Vielzahl von Anwendungsmöglichkeiten, die von einfachen Anwendungsfällen bis hin zu komplexeren reicht, bei denen die Transparenz des Prozesses, die Erklärbarkeit des Outputs sowie Zukunftssicherheit eine Rolle spielen können. Beispiele sind hier nicht nur die Generierung von (Software-)Codes oder Algorithmen, sondern auch die Automatisierung von Prozessen oder die Generierung kreativer Inhalte.

1.4.1 Textgenerierung

- Content-Erstellung: Automatische Erstellung von Artikeln, Blogposts und Social-Media-Inhalten.
- Automatische Übersetzungen: Schnelle und präzise Übersetzungen von Texten in verschiedene Sprachen.
- E-Mail-Automatisierung: Generierung von personalisierten E-Mails für Marketingkampagnen.
- Chatbots: Kundenservice-Chatbots, die natürliche Sprache verstehen und generieren können.
- SEO-Optimierung: Automatische Erstellung von SEO-optimierten Texten.
- Finanzberichterstattung: Automatische Generierung von Finanzberichten und -analysen.
- Vertragsanalyse: Automatische Zusammenfassungen und Analysen von rechtlichen Dokumenten.
- FAQs: Automatische Erstellung und Aktualisierung von häufig gestellten Fragen.
- Wissenschaftliche Forschung: Literaturrecherche und Generierung von Forschungsanträgen.
- E-Learning: Erstellung von Quizfragen und Lernmaterialien.
- Verständliche Formulierung von Texten, bspw. mit dem Prompt: »Formuliere den Text so um, dass ihn ein 12-Jähriger versteht.«
- Einheitliche Formulierung von Texten nach Unternehmensvorgaben, bspw. mit dem Prompt: »Erzeuge aus dem mitgelieferten Text einen Text, welcher unseren Unternehmensvorgaben entspricht.«
- Prüfung von Eingangsmails und Ergänzung von Lösungsvorschlägen für die nachgelagerte Weiterbearbeitung durch Mitarbeitende (die vollautomatische Beantwortung durch KI ist im Bankenumfeld kritisch und wahrscheinlich ihn den meisten Fällen nicht zulässig)
- Grundsätzliche Muster bei der Textgenerierung:
  - Frage (Prompt) an Modell senden. Modell antwortet auf Basis dessen, was es gelernt hat (Problem z. B. bei bestimmten Anwendungen hier: Quelle des Wissens kann aufgrund der Menge nicht einfach genannt werden.)
  - Frage (Prompt) an Modell senden und auszuwertendes Know-how bzw. Kontext mitliefern. Das Modell ist dabei für das Verstehen der natürlichsprachigen Anfrage und für die Formulierung einer Antwort in natürlicher Sprache zuständig. Vorteil: Die Wissensquelle ist bekannt. Nachteil: Die Abfragekosten steigen mit Größe des mitgelieferten Kontexts.

1.4.2 Bildgenerierung

- Design-Prototypen: Schnelle Erstellung von Design-Entwürfen für Websites oder Produkte.
- Bildbearbeitung: Automatische Retusche oder Anpassung von Bildern.
- Kunst: Erstellung künstlerischer Werke für digitale oder physische Medien.
- Architektur: Generierung von architektonischen Entwürfen und Modellen.
- Spieleentwicklung: Erstellung von Charakteren, Szenarien und Texturen.
- Retail: Automatische Generierung von Produktbildern in verschiedenen Umgebungen.
- Augmented Reality: Erstellung von AR-Inhalten.
- 3D-Druck: Entwurf von 3D-Modellen für den Druck.
- Landkarten und Geodaten: Generierung topografischer oder thematischer Karten.

1.4.3 Audiogenerierung (Sprache und Musik)

- Sprachassistenten: Generierung natürlicher Sprachausgabe für virtuelle Assistenten.
- Musikproduktion: Automatische Komposition von Musikstücken.
- Hörbucherstellung: Generierung einer menschenähnlichen Stimme für Hörbücher.
- Sounddesign: Erstellung von Soundeffekten für Filme oder Spiele.
- Audio-Ads: Automatische Erstellung von Audio-Werbung.
- Automatische Transkription: Umwandlung gesprochener Sprache in Text.
- Sprachanalyse: Automatische Erkennung und Analyse von Stimmungen oder Emotionen in Sprachaufnahmen.
- Telekommunikation: Verbesserung der Audioqualität in Kommunikationssystemen.
- Podcast-Erstellung: Automatische Schnitte, Zusammenfassungen und Highlights für Podcasts.
- Sprachbildung: Erstellung von Übungs- und Trainingsmaterial für Sprachkurse.

2 Rechtliche Aspekte bei der Beschaffung von KI

2.1 Verteilung der Verantwortlichkeiten entlang der Wertschöpfungskette nach AI Act

An der Konzeption, Entwicklung und Bereitstellung von KI-Systemen sind unterschiedliche Akteure beteiligt. Die Verantwortlichkeiten sind demnach unter den Akteuren der Lieferkette aufzuteilen.
Der AI Act unterscheidet im Bereich generativer KI zwischen KI-Modellen (vormals auch »foundation model«, jetzt »General Purpose AI model« in der englischen Fassung der KI-Verordnung) und KI-Systemen mit allgemeinem Verwendungszweck (»General Purpose AI system«):
Ein Basismodell ist ein KI-Modell, das auf einer breiten Datenbasis in großem Umfang trainiert wurde, auf eine allgemeine Ausgabe ausgelegt ist und an ein breites Spektrum unterschiedlicher Aufgaben angepasst werden kann.
Ein KI-System mit allgemeinem Verwendungszweck ist ein KI-System, das in einem breiten Spektrum von Anwendungen eingesetzt und an diese angepasst werden kann, für die es nicht absichtlich und speziell entwickelt wurde und das direkt verwendet werden kann, also eine Nutzeroberfläche hat. Ein solches GPAI-System kann eine Implementierung eines Basismodells sein.
Eine vertiefte Darstellung der Abgrenzung entlang der verschiedenen Stufen der Wertschöpfungskette von KI-Technologie befindet sich oben im ersten Kapitel.

Verpflichtungen für KI-Basismodelle

Basismodelle sollen nach dem Vorschlag des Europäischen Parlaments unabhängig vom Risiko reguliert werden, was damit nicht der ursprünglichen risikobasierten Ausrichtung des AI Acts entspricht. Basismodelle sollen vor dem Markteintritt in der EU eine Reihe von Transparenzanforderungen erfüllen müssen, zum Beispiel mit Bezug auf Verwendung, Architektur, Trainingsdaten und die weitere technische Dokumentation.

Alle diese Anforderungen befinden sich in Kapitel V des AI Acts mit den Artikeln 51 ff. KI-Verordnung. Hier reguliert der AI Act ausnahmsweise die Technologie und nicht den konkreten Anwendungsfall. Diese Pflichten treffen im Wesentlichen aber nur den Anbieter eines KI-Modells im Sinne von Artikel 3 Nr. 3 der KI-Verordnung.
Ergänzend enthält der AI Act umfangreiche Anforderungen für KI-Systeme, abhängig vom jeweiligen Einsatz, wenn diese als ein Hochrisiko-KI-System eingestuft werden und abhängig von der Rolle des Handelnden, nämlich ob dieser Einführer, Händler, Anbieter oder Betreiber ist. Auch generative KI in Form eines »KI-Systems mit allgemeinem Verwendungszweck« im Sinne von Artikel 3 Nr. 66 KI-Verordnung kann ein Hochrisiko-KI-System sein, wenn es die Voraussetzungen des Artikels 6 KI-Verordnung erfüllt. Zu beachten ist allerdings, dass gemäß Artikel 113 KI-Verordnung die Regelungen des Artikels 6 Absatz 2 erst ab dem 2. August 2026 und die des Artikels 6 Absatz 1 erst ab dem 2. August 2027 gelten.
Im Folgenden sollen kurz die verschiedenen Rollen des Handelnden betrachtet werden:

Verpflichtungen für Einführer von Hochrisiko-KI-Systemen

Bevor ein Hochrisiko-KI-System in den Verkehr gebracht wird, stellen Einführer dieser Systeme unter anderem sicher, dass ein Konformitätsbewertungsverfahren durchgeführt und die technische Dokumentation erstellt werden und ihre Namen, ihre eingetragenen Handelsnamen oder ihre eingetragene Handelsmarke und Kontaktanschrift beigefügt sind.
Verpflichtungen für Händler von Hochrisiko-KI-Systemen
Bevor Händler ein Hochrisiko-KI-System auf dem Markt bereitstellen, überprüfen sie unter anderem, ob das Hochrisiko-KI-System mit der erforderlichen CE-Konformitätskennzeichnung versehen ist, ob die erforderliche Dokumentation und Gebrauchsanweisung beigefügt sind und ob der Anbieter bzw. gegebenenfalls der Einführer des Systems die Verpflichtungen für Hochrisiko-KI-Systeme erfüllt hat.

Verpflichtungen für Händler von Hochrisiko-KI-Systemen

Bevor Händler ein Hochrisiko-KI-System auf dem Markt bereitstellen, überprüfen sie unter anderem, ob das Hochrisiko-KI-System mit der erforderlichen CE-Konformitätskennzeichnung versehen ist, ob die erforderliche Dokumentation und Gebrauchsanweisung beigefügt sind und ob der Anbieter bzw. gegebenenfalls der Einführer des Systems die Verpflichtungen für Hochrisiko-KI-Systeme erfüllt hat.

Verpflichtungen für Anbieter von Hochrisiko-KI-Systemen

Anbieter von Hochrisiko-KI-Systemen unterliegen umfangreichen Verpflichtungen nach dem AI Act, vgl. unter anderem Art. 16 AI Act. Händler, Einführer, Betreiber oder sonstige Dritte gelten nach Artikel 25 Abs. 1 des AI Acts als Anbieter eines Hochrisiko-KI-Systems, wenn sie:

1. ihren Namen oder ihr Markenzeichen auf ein Hochrisiko-KI-System setzen, das bereits in Verkehr gebracht oder in Betrieb genommen wurde,

2. eine wesentliche Änderung (im Sinne von Artikel 3 Nr. 23 KI-Verordnung) an einem Hochrisiko-KI-System vornehmen, das bereits in Verkehr gebracht oder in Betrieb genommen wurde (und es weiterhin ein Hochrisiko-KI-System bleibt),

3. die Zweckbestimmung eines KI-Systems, einschließlich eines KI-Systems mit allgemeinem Verwendungszweck, das nicht als Hochrisiko-KI-System eingestuft wurde und bereits in Verkehr gebracht oder in Betrieb genommen wurde, so verändern, dass das KI-System zu einem Hochrisiko-KI-System wird.

In den Punkten 1–3 gilt der Anbieter, der das KI-System ursprünglich in Verkehr gebracht oder in Betrieb genommen hat, nicht mehr als Anbieter. Der ursprüngliche Anbieter ist somit nicht mehr verantwortlich und soll dem neuen Anbieter lediglich die technische Dokumentation und alle relevanten und vernünftigerweise zu erwartenden Informationen und Fähigkeiten des KI-Systems, den technischen Zugang oder sonstige Unterstützung auf der Grundlage des allgemein anerkannten Stands der Technik zur Verfügung, die für die Erfüllung der in dieser Verordnung festgelegten Verpflichtungen erforderlich sind.

Dies gilt auch für Anbieter von KI-Systemen mit allgemeinem Verwendungszweck (»General Purpose AI System«), wenn diese direkt in ein Hochrisiko-KI-System integriert sind.

Verpflichtungen für Betreiber von Hochrisiko-KI-Systemen

Der Betreiber im Sinne von Artikel 3 Nr. 4 der KI-Verordnung hat vor allem die Pflichten aus den Artikeln 26 und 27 der KI-Verordnung zu erfüllen, die verschiedene Pflichten im Zusammenhang mit dem Betrieb und eine vorherige Grundrechte-Folgenabschätzung umfassen.
Demnach muss vor der Beschaffung von KI-Technologie unbedingt geklärt werden, worum es sich technisch genau handelt und welche Rolle die verwendende Stelle haben wird.

2.2 Checkliste und Vorüberlegungen vor der Beschaffung von KI (IT Procurement)

Vor der Beschaffung von KI-Lösungen sollten vielfältige technische, kommerzielle und rechtliche Bewertungen und Festlegungen der internen und externen Anforderungen an die zu beschaffende KI erfolgen. Hierbei ist irrelevant, ob es sich um eine kostenpflichtige Lösung handelt, oder um ein kostenfreies Open Source KI-Modell, das direkt beim Anbieter oder auf einer Plattform wie beispielsweise Hugging Face heruntergeladen, installiert und dann betrieben wird.

1. Empfehlenswert ist zunächst die **Erstellung einer generellen KI-Strategie in einem partizipativen Abstimmungsprozess mit allen relevanten Stakeholdern** (d. h. Management, Leitungen von Fachabteilungen, KI-Fachleuten, Datenschutz, Informationssicherheit, Rechtsabteilung, Compliance, Betriebsrat, betroffene Beschäftigte etc.) zur Dokumentation, in welcher Weise die Nutzung von KI-Anwendungen zur Erreichung der betrieblichen Ziele beitragen soll. Das schließt ein, ob eventuell generative KI in Form von KI-Modellen beschafft und vielleicht sogar selbst betrieben werden soll, oder doch lieber ein komplettes KI-System, was natürlich vom konkreten Beschaffungsbedarf abhängt. Wenn das Unternehmen selbst Anwendungsfälle von KI erstellt, beispielsweise unter Verwendung eigener, proprietärer Daten des Unternehmens mittels Finetunings, also Nachtrainieren eines KI-Modells, oder sogenannter Retrieval Augmented Generation (»RAG«) Techniken, dann ist meistens die Beschaffung generativer KI, also von KI-Basistechnologie, erforderlich.

2. Unerlässlich ist die **Festlegung der wesentlichen technischen, kommerziellen und rechtlichen Anforderungen** an die zu beschaffende KI als »IT Procurement-Projekt« unter Beteiligung der relevanten Abteilungen und Stakeholder. Hinsichtlich der konkreten Anforderungen sind diese generell vom Risikoprofil und der Risikobereitschaft« des Unternehmens abhängig, ebenso wie vom Kontext, ob das Unternehmen in einer regulierten Industrie tätig ist, wie beispielsweise dem Finanzbereich, der bereits sektorspezifisch stark reguliert ist.

3. Die **Entscheidung, ob** die Leistung **als AI as a Service** (»AIaaS« – vergleichbar mit Software as a Service/SaaS) bezogen werden soll, **oder als »on premise«-Installation** (auch oft einfach nur als »on-prem« bezeichnet) ist vorab zu treffen.

Die Entscheidung zwischen AIaaS und on-premise-Installation, also einer Installation im Rechenzentrum des Unternehmens oder bei einem Dienstleister des Unternehmens, ist von entscheidender Bedeutung. Sie hat direkte Auswirkungen auf die Skalierbarkeit und Flexibilität von KI-Lösungen. AIaaS ermöglicht Unternehmen den Zugang zu leistungsstarken KI-Algorithmen und Ressourcen ohne die Notwendigkeit eigener Infrastruktur, insbesondere spezieller Serverprozessoren in Form von Grafikprozessoren, sog. »GPUs«. Dies ist besonders vorteilhaft für Unternehmen mit begrenzten Ressourcen oder einem kurzfristigen Bedarf an KI-Funktionalität. Auf der anderen Seite kann die Entscheidung für ein on-premise-installiertes KI-Produkt eine bessere

Anpassung an spezifische geschäftliche Anforderungen ermöglichen, was zu einer höheren Effektivität und Präzision führen kann. Außerdem bieten on-premise KI-Lösungen den Unternehmen eine lokale Lösung für den Schutz sensibler Daten. Edge Computing ermöglicht gegebenenfalls zudem die lokale Verarbeitung von Daten auf dem Gerät und beseitigt das Risiko, das mit dem Senden von Daten an einen Cloud-Anbieter und deren Speicherung in der Cloud verbunden ist. Diese Vorteile kommen vor allem Unternehmen zugute, die in sensiblen Bereichen wie dem Gesundheitswesen, dem Bildungswesen und der Verteidigung tätig sind. Weiterhin können Investitionen in die Infrastruktur sich als vorteilhaft erweisen, da sie insbesondere KMU flexible und kosteneffiziente eigene Rechenkapazitäten bieten und sie in die Lage versetzen, ihre eigenen KI-Modelle zu entwickeln und auszubauen. Diese Anpassung von Basis-KI-Modellen insbesondere mit eigenen Daten begründet häufig einen Wettbewerbsvorteil, da ein KI-Modell einzigartig wird, beispielsweise in Form von Fähigkeiten oder Kenntnissen.

Daher ist die Prüfung von strategischen Erwägungen wie der Nutzung (eventuell schon vorhandener) eigener Infrastruktur oder der eines Drittanbieters (z. B. eines IT-Outsourcingpartners oder eines Hyperscalers und dessen Cloud) und einer hierdurch gegebenenfalls entstehenden Abhängigkeit von diesen externen Dienstleistern z. B. durch Offenlegung von Know-how oder Daten, sowie Folgekosten (z. B. Updates, Wartung, Wechsel des externen Dienstleisters zu späteren Zeitpunkten), entscheidend. Darüber hinaus spielen hier auch Aspekte des Schutzes personenbezogener Daten sowie der Betriebs- und Geschäftsgeheimnisschutz eine wesentliche Rolle.

4. Die Festlegung der wichtigsten rechtlichen »Must-have«- oder »No-Go«-Konditionen, ist von entscheidender Bedeutung für den sachgerechten Vergleich und für die Risiken der ausgewählten KI, insbesondere hinsichtlich Haftung, Datenschutz, Datennutzung, Gewährleistung Einhaltung regulatorischer Anforderungen, Rechten an geistigem Eigentum und der Nutzung der von KI verwendeten und generierten Daten sowie Regelungen für Implementierung, Verfügbarkeit, Skalierbarkeit, Support und Wartung samt Service Level Agreement für Support und gegebenenfalls Verfügbarkeit (»SLAs«), ist von entscheidender Bedeutung für den sachgerechten Vergleich und für die Risiken der ausgewählten KI.

5. Außerdem ist insbesondere die Festlegung des erforderlichen sachlichen und räumlichen Umfangs der Lizenzierung bzw. Nutzungsrechte an der KI, der von KI verwendeten internen oder externen Daten (»Input«) und den dadurch generierten Daten (»Output«) ebenfalls wichtig. Hier ist auch die Nutzung der eingegebenen (Trainings-) Daten durch den KI-Provider zu berücksichtigen. Diese sollte bzw. kann ggf. durch eine vertragliche Vereinbarung oder die Wahl einer bestimmten Lizenz ausgeschlossen werden, wo nötig. Weiteres besonderes Augenmerk sollte auf die Nutzung von R&D- bzw. Entwicklungsdaten durch die KI sowie auf die Nutzung in produktiver Umgebung (Software und Hardware) gelegt werden.

6. Vor allem aus technischer und kommerzieller Sicht ist frühzeitig zu bewerten, ob die Implementation bzw. Integration der KI intern erfolgt oder ob dafür ein externer IT-Servicepartner nötig oder sinnvoll ist.

7. Gleiches gilt für die Bewertung, ob der Einkauf der KI über externe IT-Servicepartner (Systemintegrator, Reseller oder Kooperationspartner der KI-Anbieter) vorteilhaft ist.

8. Rechtlich sind insbesondere folgende Punkte zu prüfen:

- Die Möglichkeit und Sinnhaftigkeit der (teilweisen) Nutzung von eigenen Einkaufsbedingungen oder Vertragsbedingungen, die meist allerdings nicht dem komplexen Vertragsgegenstand eines KI-Systems gerecht werden,
- Anbieter-AGB bzw. Vertragsbedingungen, sowie
- Bewertung und Auswahl des bevorzugten anwendbaren Rechts für Beschaffung von KI, um ggf. für Besteller vorteilhaftere Jurisdiktionen zu nutzen (die z. B. weniger Haftungsausschlüsse erlauben).
- In jedem Fall muss der Anbieter eines neuen KI-Modells mit allgemeinem Verwendungszweck ab dem 2. August 2025 die Erfüllung seine Pflichten nach Artikeln 51 ff. KI-Verordnung nachweisen und eine umfangreiche Dokumentation nach Artikel 53 Absatz 1 b) liefern, damit das kaufende Unternehmen als Kunde gegebenenfalls als Anbieter eines KI-Systems, in dem das KI-Modell mit allgemeinem Verwendungszweck integriert wird, die Fähigkeiten des gelieferten KI-Modells beurteilen sowie den eigenen Pflichten als Anbieter des KI-Systems nachkommen kann. Diese Anforderungen gelten nach Artikel 113 b) zwar erst ab dem 2. August 2025. Sie sollten davor aber auch schon für Investitionsentscheidungen berücksichtigt werden, auch wenn Artikel 111 Absatz 3 eine großzügige Übergangsfrist für die Anbieter von KI-Modellen vorsieht: »Anbieter von KI-Modellen mit allgemeinem Verwendungszweck, die vor dem 2. August 2025 in Verkehr gebracht wurden, treffen die erforderlichen Maßnahmen für die Erfüllung der in dieser Verordnung festgelegten Pflichten bis zum 2. August 2027.«

  Ergänzend möchten wir noch darauf hinweisen, dass die vorgenannten Anforderungen für KI-Modelle durch den Anbieter gegebenenfalls auch durch das Stützen auf einen »Code of Practice« der EU-Kommission nachgewiesen werden können, vgl. Artikel 53 Absatz 4, 56 KI-Verordnung, den es voraussichtlich ab Mai 2025 geben wird. Ob das Unternehmen schließlich eventuell auch formal Anbieter des KI-Modells wird, hängt vom konkreten Einzelfall ab, beispielsweise wenn das Modell wesentlich verändert wird (siehe oben) oder das Modell hierdurch erstmalig an den Markt kommt, ein sogenannten »Inverkehrbringen« im Sinne von ↗Artikel 3 Nr. 9 der KI-Verordnung.

9. Risikoabwägung
   Elementar für die Auswahl ist eine »State of the Art«-Risikoabwägung. Im Rahmen der Risikoabwägung gilt es bei KI, neue Herausforderungen zu beachten, sodass bisherige Prozesse des Risikoassessments ggf. überarbeitet werden müssen. Das BSI nennt in seinem Bericht zur Lage der IT-Sicherheit in Deutschland 2023 vor allem die Nutzung der KI für Cyberangriffe, das Ausnutzen von Schwachstellen in der KI, Herausforderung im Training und vor allem das umfassende Wissen der KI bei einer weiten Einbindung in bestehende Strukturen und Datenbanken als neue Bedrohungen. [2]

2 https://www.bsi.bund.de/SharedDocs/Downloads/DE/BSI/Publikationen/Lageberichte/Lagebericht2023.pdf?__blob=publicationFile&v=6 S. 40ff Guidelines for secure AI system development (ncsc.gov.uk)

Die Risikobewertung sollte u. a. die vom Fraunhofer IAO empfohlenen Punkte berücksichtigen:

- Welche Datensätze werden für Entwicklung bzw. das Training der KI genutzt? Können oder sollten Trainingsdaten ggf. gesondert auf dem Markt für Trainingsdatensätze beschafft werden?
- Welche Datensätze werden im anschließenden Praxisbetrieb verarbeitet?
- Welche rechtlichen Rahmenbedingungen müssen dabei beachtet werden?
- Ist es für die Zwecke der jeweiligen Anwendung ausreichend, anonyme Daten zu verwenden?
- Wer erhält Zugriff auf die Daten? Wie sind betriebliche Datenbestände gegen Missbrauch oder Betriebsspionage geschützt? Müssen die Risiken für personenbezogene Daten vorab in einer Datenschutz-Folgenabschätzung geprüft und dokumentiert werden? Wer führt diese durch? Wie werden Mitbestimmungsrechte hinsichtlich des Schutzes personenbezogener Daten wahrgenommen? Wie werden die Kriterien der Datensparsamkeit und Zweckbindung der Datennutzung umgesetzt?
- Vor allem sind aber die vertraglichen und gesetzlichen Haftungsrisiken aus der konkret geplanten Nutzung der KI für interne Zwecke und/oder Leistungen für Kundinnen und Kunden im Detail zu bewerten und soweit möglich angemessen durch vertragliche Gewährleistungs-, Haftungs- und Freistellungsregeln abzusichern (siehe Ziffer 4.3). Dabei dürfen mögliche direkte und indirekte (Vermögens-)Schäden beim Kunden oder Dritten aufgrund der Nutzung durch von der KI generierten Daten nicht außer gelassen werden.
- Nutzung von bereitgestellten Filtern.
- KI-Governance.
- Es ist weiterhin zu beachten, dass die in folgendem Kapitel 3 genannten rechtlichen Aspekte bereits im Rahmen der Beschaffung berücksichtigt werden sollten, da im eingekauften Produkt enthaltene Rechtsmängel sich auch auf die Einsatzphase auswirken würden.

Da KI-Technologie fast immer in Form von Software implementiert wird, gelten die Überlegungen und Anforderungen, die bei der Beschaffung von Software zu beachten sind, grundsätzlich entsprechend. Wird die Software »as a Service« (AIaaS, s. oben) angeboten, gilt es zu beachten, dass der Anbieter des generativen KI-Systems neben der »reinen« Software auch eine Infrastruktur bereitstellt, die für die Nutzung des konkreten KI-Systems erforderlich ist. Dabei bedarf es einer ausreichenden Kontrolle über die genutzten KI-Systeme der Anbieter durch die Kundinnen und Kunden, um etwa sicherzustellen, dass gesetzliche oder interne Anforderungen eingehalten werden. Volle technische Kontrolle und insbesondere auch Souveränität über die eigenen Daten, die in die KI fließen sollen, hat der Kunde nur bei einer on-premise-Lösung.

Aufgrund der Vergleichbarkeit zu SaaS-Lösungen kann für die vertragliche Gestaltung und Prüfung der Beschaffung von KI-Systemen grundsätzlich auf die entsprechenden Anforderungen zu SaaS-Verträgen zurückgegriffen werden. Insbesondere sind im Rahmen der vertraglichen Prüfung ausreichende Nutzungsrechte und Service-Level für die Verfügbarkeit zu vereinbaren, die dem Zweck der beabsichtigten Anwendungsbereiche des jeweiligen KI-Systems im Unternehmen

3 Fraunhofer IAO, Leitfaden zu Strategie und Wandel für den KI-Einsatz, Leitfaden zur Durchführung von KI-Projekten (fraunhofer.de) https://www.digital.iao.fraunhofer.de/de/publikationen/Leitfaden_DurchfuehrungVonKI-Projekten.html, zuletzt besucht am 12. Februar 2024

entsprechen. In Verträgen über KI-Systeme sollte zudem die Verantwortlichkeit für Herkunft und Qualität von Trainingsdaten der KI-Modelle eindeutig festgelegt sein.

Darüber hinaus sollte es klare Regelungen dafür geben, wenn die vertragsgemäße Nutzung des KI-Systems durch den Kunden gegen regulatorische Vorgaben, Gesetze und/oder Rechte Dritter verstößt. Hier sollten etwa Klauseln zur Freistellung von solchen Ansprüchen Dritter durch den Anbieter des KI-Systems enthalten sein. Schließlich sollte ausdrücklich vereinbart werden, in welchem Umfang und in welcher Form die in dem KI-System verarbeiteten Kundendaten vom Anbieter des KI-Systems (ggf. zu eigenen Zwecken) genutzt werden dürfen. Je nachdem, wie sensibel und kritisch diese Kundendaten sind, sollten sie durch ausdrückliche vertragliche Regelungen vor einer vertragswidrigen Nutzung geschützt werden.

Die spezifischen Risiken, die mit der Nutzung eines generativen KI-Systems einhergehen können, sollten deshalb unbedingt schon im Rahmen des Beschaffungsprozesses ausreichend geprüft werden. Daher sollte die beabsichtige Nutzung des jeweiligen KI-Systems, insbesondere von Expertinnen und Experten aus dem Datenschutz, der Informationssicherheit, der Rechtsabteilung und Compliance, dahingehend geprüft werden, ob ein rechtskonformer Einsatz möglich ist und welche Risiken ggf. bestehen. Bei dieser Prüfung sind die von den Anbietern offerierten unterschiedlichen Lizenzmodelle (z. B. Versionen speziell für Unternehmen anstatt von allgemein offen zugänglichen Versionen) sowie unter Umständen bereitgestellte Sicherheitsmechanismen, wie etwa Filter oder der Ausschluss bestimmter Datenverarbeitungen, zu berücksichtigen.

Wegen der Komplexität einer solchen interdisziplinären Prüfung und auch wegen der zunehmenden gesetzlichen Anforderungen, z. B. durch den AI Act mit einem ggf. einzurichtenden eigenen Risikomanagementsystem, empfiehlt sich die Implementierung einer KI-Governance, die in den Beschaffungsprozess einzubinden ist. Diese Funktion kann dann sowohl rechtliche als auch technische Anforderungen an die Nutzung von KI-Systemen im Unternehmen bündeln, deren Einhaltung überwachen und darüber hinaus als Kompetenzstelle für interne und externe Ansprechpartner dienen.

3 Rechtliche Aspekte in der Einsatzphase

3.1 Datenverarbeitung

3.1.1 Pflichten nach der DSGVO

3.1.1.1 Einleitung, Allgemeines und Anwendungsbereich

Die Datenschutz-Grundverordnung (DSGVO) und das Bundesdatenschutzgesetz (BDSG) sind zwei wesentliche rechtliche Rahmenwerke, die den Einsatz von Künstlicher Intelligenz in Unternehmen in der Europäischen Union und in Deutschland regeln, sofern die KI personenbezogene Daten verarbeitet. Sie sind anwendbar, wenn Unternehmen als Verantwortliche oder Auftragsverarbeiter personenbezogene Daten im Rahmen der KI-Nutzung verarbeiten. Dabei ist irrelevant, ob die KI selbstständig Entscheidungen trifft oder als unterstützendes Werkzeug dient. Im Gegensatz zu KI-Systemen für Profiling, Diagnostik etc., die grundsätzlich (sensible) personenbezogene Daten verarbeiten und in der Regel keine generative KI darstellen, werden beim Einsatz generativer KI personenbezogene Daten weniger häufig unmittelbar verarbeitet. Ein Anwendungsbeispiel für eine solche unmittelbare Verwendung ist das Account-Management.[4]

Im folgenden Abschnitt werden auch Bezüge zu Anwendungsfällen jenseits von generativer KI – bspw. diskriminativer KI – hergestellt. Dies soll den Blick trotz des gesetzten Rahmens dieses Leitfadens weiten und für Problembereiche innerhalb des Datenschutzrechts sensibilisieren, die sich bei KI außerhalb des Anwendungsbereiches der generativen KI ergeben. Dies betrifft vor allem Bereiche des Profilings und generell Bereiche, wo Entscheidungen autonom bzw. maschinengestützt getroffen werden.

Beispiele für eine breitere Verarbeitung personenbezogener Daten im Rahmen von generativer KI sind die Erstellung von Texten mit personenbezogenem Inhalt oder das Generieren von Bildern und Videos auf Basis von Originalmaterial mit Personenbezug. In diesen Fällen gelten die Regelungen der DSGVO und des BDSG.

Ein Unternehmen, das mittels KI personenbezogene Daten verarbeitet, gilt als »Verantwortlicher« im Sinne der DSGVO, wenn es Entscheidungen über die Zwecke und Mittel der Datenverarbeitung trifft. Die Verarbeitung umfasst jede Operation, die mit personenbezogenen Daten durchgeführt wird – von der Erhebung über die Speicherung und Analyse bis hin zur Löschung.

4 Zwar sind in gängigen Modellen generativer KI selbst personenbezogene Informationen nicht als Klardatum enthalten. Allerdings können solche personenbezogenen Informationen in Form von Wahrscheinlichkeiten bzw. Gewichtungen, auf deren Basis diese Modelle arbeiten, repräsentiert sein. Diese Wahrscheinlichkeiten repräsentieren letztlich die Trainingsdaten, aus denen sie errechnet wurden, und somit potenziell auch jeden Personenbezug, der in diesen Trainingsdaten enthalten ist – vergleichbar mit einer Pseudonymisierung, bei der allerdings die KI selbst die Mittel hat, den Personenbezug wiederherzustellen. Zumindest bei allen KI-Modellen, die mit CommonCrawl oder vergleichbaren Datensätzen trainiert wurden, dürfte ein Personenbezug in den Trainingsdaten nicht auszuschließen sein. In Einzelfällen »erinnert« sich das Modell an die personenbezogene Information. Zu vertiefteren Datenschutzfraen konsultieren Sie bitte unseren Praxisleitfaden KI & Datenschutz, im Erscheinen.

Bei der Nutzung von KI kann es zu Konstellationen kommen, in denen das Unternehmen entweder als Auftragsverarbeiter oder in gemeinsamer Verantwortung mit dem KI-Anbieter agiert:

Bei der Auftragsverarbeitung verarbeitet das Unternehmen Daten im Auftrag eines anderen Unternehmens (dem Auftraggeber). Die KI dient als Werkzeug zur Erledigung spezifischer Aufgaben. Die Verantwortlichkeit für die Einhaltung der Datenschutzvorschriften liegt primär beim Auftraggeber.
Wenn das Unternehmen und der KI-Hersteller gemeinsam Entscheidungen über die Verarbeitung personenbezogener Daten treffen, spricht man von gemeinsamer Verantwortlichkeit. Beide Parteien müssen in diesem Fall die Einhaltung der DSGVO sicherstellen und dies in einer Vereinbarung dokumentieren.

In einigen Fällen kann es zur getrennten Verantwortung kommen, z. B. wenn der Anbieter der KI-Daten für Zwecke, die vom Auftraggeber weder bestimmt noch mit ihm abgesprochen sind, verarbeitet.

Eine Herausforderung kann im Falle der gemeinsamen Verantwortlichkeit darin liegen, die Rechte und Pflichten und damit auch die Haftung zwischen dem KI-Hersteller und dem einsetzenden Unternehmen für beide Parteien akzeptabel aufzuteilen.

Die wichtigsten Rechtsgrundlagen für die Verarbeitung personenbezogener Daten durch KI im Unternehmenskontext sind:

- Einwilligung (Art. 6 Abs. 1 lit. a DSGVO): Die betroffene Person muss ihre Einwilligung zur Datenverarbeitung geben. Diese muss freiwillig, bestimmt und informiert abgegeben werden.
- Vertragserfüllung (Art. 6 Abs. 1 lit. b DSGVO): Die Datenverarbeitung ist notwendig, um einen Vertrag zu erfüllen, in dem die betroffene Person Partei ist oder um einen Vertrag auf Anfrage der betroffenen Person zu schließen.
- Rechtliche Verpflichtung (Art. 6 Abs. 1 lit. c DSGVO): Die Verarbeitung ist notwendig, um einer rechtlichen Verpflichtung nachzukommen. [5]
- Berechtigte Interessen (Art. 6 Abs. 1 lit. f DSGVO): Die Verarbeitung ist zur Wahrung berechtigter Interessen des Unternehmens oder eines Dritten notwendig, sofern nicht die Interessen oder Grundrechte und Grundfreiheiten der betroffenen Person überwiegen.

Dabei ist es wichtig zu beachten, dass es für die Erhebung und Nutzung personenbezogener Daten zum Training einer KI und der Verarbeitung dieser Daten durch die KI jeweils einer – gegebenenfalls unterschiedlichen – Rechtsgrundlage bedarf. Es sollte geprüft werden, ob die Datennutzung sachgerechter Weise von vorneherein auf aggregierte Daten ohne Personenbezug beschränkt werden kann, oder eine Verwendung anonymer Daten ausreicht. Dies gilt erst recht mit Blick auf die sich teilweise widersprechenden Vorgaben der DSGVO und des Data Acts bzgl. der Nutzung von durch verbundene Produkte oder Dienstleistungen generierten Daten, inklusive von Betriebsgeheimnissen. Hier dürfte eine Vermischung von personenbezogenen Daten nach der DSGVO, IoT-Daten nach dem Data Act und vertraulicher interner Daten (R&D, Produktionsdaten etc.) wohl zu vermeiden sein, um komplexe Folgeprobleme zu vermeiden (siehe dazu auch Ziffer 3.1.2.).

5 Bislang besteht jedoch noch keine rechtliche Pflicht zum Einsatz von KI.

Wenn personenbezogene Daten in Staaten außerhalb des Geltungsbereiches der DSGVO übermittelt werden, muss sichergestellt sein, dass trotzdem ein den Anforderungen der DSGVO entsprechendes Schutzniveau gewährleistet ist, und eine der Grundlagen aus Art. 45 ff. DSGVO, z. B. ein Angemessenheitsbeschluss der EU-Kommission oder zwischen Datenexporteur und -importeur abgeschlossene Standarddatenschutzklausel, vorliegt.

Neben den Bestimmungen der DSGVO sind im Daten- und KI-Kontext auch die Bestimmungen des Data Acts sowie des Data Governance Acts zu beachten, auf die im Rahmen dieses Leitfadens jedoch nicht vertieft eingegangen werden soll.

Weitere Einzelheiten zu Fragen der datenschutzkonformen Nutzung und Implementierung von KI-Technologien in Unternehmen finden sich in dem Bitkom Praxisleitfaden ↗»Künstliche Intelligenz & Datenschutz».

3.1.1.2 Pflichten im Einzelnen

Zweckbindung

Gemäß der DSGVO müssen Daten für festgelegte, eindeutige und legitime Zwecke erhoben werden. Das bedeutet, dass bei der Nutzung von KI-Systemen die Daten nur für den spezifischen Zweck verarbeitet werden dürfen, für den sie gesammelt wurden.

Wenn KI-Systeme Daten verarbeiten und Ergebnisse liefern, müssen diese Ergebnisse im Einklang mit dem ursprünglichen Zweck stehen. Nach Erreichen dieser Ergebnisse müssen die eingegebenen Daten gelöscht werden. Eine Weiterverarbeitung zu einem anderen Zweck erfordert in der Regel eine neue Rechtsgrundlage.

Beim Training von KI-Modellen mit personenbezogenen Daten muss sichergestellt werden, dass die Datenverarbeitung den Zwecken entspricht, für die die Daten ursprünglich erhoben wurden. Darüber hinaus müssen Maßnahmen wie Pseudonymisierung oder Anonymisierung ergriffen werden, um die Privatsphäre der Betroffenen zu schützen. Gegebenenfalls können Daten zu Trainingszwecken auf Basis einer Einwilligung als Rechtsgrundlage verarbeitet werden. Dies ist zu empfehlen, falls keine ausreichende Pseudonymisierung möglich ist und Anonymisierung dem Zweck entgegensteht. Eine Pseudonymisierung ist nur dann möglich, wenn die Zuordnung von Pseudonymen und Klarnamen nach der Verarbeitung durch und mittels KI noch verlässlich erhalten bleibt.

Aktuell etablierte KI-Modelle verwenden häufig Datensets als Teil der Trainingsdaten, für die pauschal enorme Mengen im Internet zugänglicher Daten aus zahlreichen Quellen gesammelt werden (z. B. CommonCrawl und WebText). Für alle diese Daten muss gewährleistet sein, dass eine Weiterverarbeitung als Trainingsdaten für die KI entweder durch den ursprünglichen Verarbeitungszweck abgedeckt ist oder eine neue Rechtsgrundlage vorliegt.

Vermeidung der Entstehung besonderer Kategorien personenbezogener Daten i. S. d. Art. 9 DSGVO

Wenn besondere Kategorien personenbezogener Daten im Sinne des Art. 9 Abs. 1 DSGVO Gegenstand der Verarbeitung sind, hat der verantwortliche Datenverarbeiter erhöhte Schutzanforderungen umzusetzen, damit die Risiken für den Schutz der Rechte und Freiheiten natürlicher Personen minimal bleiben.
Die im Bereich der Verfahren der Künstlichen Intelligenz durchgeführten umfangreichen Datenverarbeitungen sind dahingehend zu beurteilen, ob aus den personenbezogenen Daten im Laufe der Anwendung und Weiterentwicklung der KI Informationen abgeleitet werden können, die besondere Kategorien personenbezogener Daten nach Art. 9 Abs. 1 DSGVO sein können.
Dies kann z. B. der Fall sein, wenn aus personenbezogenen Daten Schlüsse auf besondere personenbezogene Daten, auch indirekt, abgeleitet werden können. Siehe dazu EuGH:
»Art. 9 Abs. 1 DSGVO ist dahingehend auszulegen, dass die Veröffentlichung personenbezogener Daten, die geeignet sind, die sexuelle Orientierung einer natürlichen Person indirekt zu offenbaren, auf der Website der Behörde, die für die Entgegennahme und die inhaltliche Kontrolle von

Erklärungen über private Interessen zuständig ist, eine Verarbeitung besonderer Kategorien personenbezogener Daten im Sinne dieser Bestimmung darstellt.« (EuGH v. 1.8.2022 - C-184/20)

Ferner sind nach Art. 10 des AI Acts bei sogenannten Hochrisiko-KI-Systemen die Risiken bei der Nutzung von besonderen Kategorien personenbezogener Daten »angemessene Vorkehrungen für den Schutz der Grundrechte und Grundfreiheiten natürlicher Personen« zu treffen, »wozu auch technische Beschränkungen einer Weiterverwendung und modernste Sicherheits- und Datenschutzmaßnahmen wie Pseudonymisierung oder Verschlüsselung gehören« (Art. 10 Abs. 5 des Entwurfs der EU KI-Verordnung).
Daraus lässt sich schließen, dass bei den umfangreichen Verarbeitungen von Daten im Rahmen des Trainings und der Anwendung von KI-Systemen zu prüfen ist, ob durch die Anwendung besondere Kategorien personenbezogener Daten (Art. 9 DSGVO) entstehen können. Dann wären für das Training und für die Anwendung eines KI-Systems die genannten erhöhten Anforderungen zum Schutz dieser besonderen personenbezogenen Daten umzusetzen. Zum aktuellen Stand der Technik dürfte dies bei Anwendungen von generativer KI eher weniger der Fall sein, dafür aber bei anderen KI-Technologien, insb. der diskriminativen. [6]

Informationspflichten (Transparenz)

Einer der Hauptgrundsätze der DSGVO ist die Transparenz. Unternehmen müssen Personen, deren Daten sie verarbeiten, umfassend darüber informieren. Dies schließt Informationen über den Zweck der Datenverarbeitung, die Rechtsgrundlage, die Dauer der Speicherung und die Rechte der betroffenen Personen ein. Besonders wichtig sind dabei die Artikel 13 Abs. 2 lit. f und 14 Abs. 2 lit. g, die sich auf automatisierte Entscheidungen, einschließlich Profiling, konzentrieren. Sie verlangen, dass Unternehmen die Betroffenen über die Existenz solcher Verarbeitungsmechanismen, deren Logik sowie die Bedeutung und die beabsichtigten Auswirkungen informieren.

Die Integration von KI in Geschäftsprozesse bringt allerdings Herausforderungen mit sich. Eine der größten ist dabei die Transparenz bei selbstlernenden Systemen. Diese Systeme sind dynamisch. Ihre Entscheidungsfindungsprozesse können komplex und schwer nachvollziehbar sein. Dies erschwert die Einhaltung der Transparenzvorschriften der DSGVO.

Ein weiteres Problem ergibt sich, wenn Unternehmen KI-Lösungen einsetzen, die sie nicht selbst entwickelt haben. In solchen Fällen sind sie von den Datenschutzpraktiken Dritter abhängig, was zusätzliche Risiken in Bezug auf die Einhaltung der DSGVO birgt. Häufig fehlt es auch an Transparenz, sodass der Verpflichtung zur Information der Betroffenen über die Verarbeitung ihrer personenbezogenen Daten möglicherweise nicht vollständig nachgekommen wird.

Ein proaktiver Ansatz zur Einhaltung der DSGVO-Vorschriften ist das Konzept »Privacy by Design«. Es bedeutet, dass Datenschutzmaßnahmen von Anfang an in die Entwicklung und Implementierung von KI-Systemen integriert werden. Dies umfasst die Minimierung der Datenerfassung, die Sicherstellung der Transparenz und das Einbetten von Kontrollmöglichkeiten für die Nutzerinnen und Nutzer.

6 Siehe zu den technologieunabhängigen vertieften Datenschutzausführungen den datenschutzspezifischen Praxisleitfaden KI & Datenschutz, im Erscheinen.

Die Rechenschaftspflicht ist ein weiterer wesentlicher Aspekt der DSGVO, der eng mit dem Einsatz von Künstlicher Intelligenz in Unternehmen verknüpft ist. Diese Pflicht erfordert von Unternehmen nicht nur die Einhaltung der Datenschutzvorschriften, sondern auch den Nachweis, dass geeignete Maßnahmen und Verfahren implementiert wurden, um die Compliance mit der DSGVO zu gewährleisten.

Im Kontext der KI bedeutet dies, dass Unternehmen nicht nur transparent über ihre Datenverarbeitungspraktiken informieren und diese an die DSGVO anpassen müssen, sondern auch belegen müssen, dass ihre KI-Systeme den Datenschutzvorschriften entsprechen. Dies beinhaltet die Dokumentation der Datenverarbeitungsprozesse, die Durchführung von Datenschutz-Folgenabschätzungen, insbesondere bei risikoreichen KI-Anwendungen, sowie die Implementierung und Überwachung von Datenschutzmaßnahmen.
Die Rechenschaftspflicht ist besonders herausfordernd im Umgang mit KI, da die Technologie oft komplexe und dynamische Datenverarbeitungsprozesse umfasst. Unternehmen müssen daher sicherstellen, dass ihre KI-Systeme nicht nur zum Zeitpunkt der Implementierung, sondern auch über ihren gesamten Lebenszyklus hinweg den Datenschutzanforderungen entsprechen. Dies erfordert eine kontinuierliche Überwachung und Anpassung der Systeme sowie eine klare Dokumentation aller Änderungen und Entscheidungen.

Eine weitere Herausforderung liegt bei der Verwendung von KI in der Frage, ob und inwieweit es als Voraussetzung für einen ausreichenden Grad an Transparenz erforderlich ist, den Algorithmus und/oder die Trainingsdaten offenzulegen. Im Falle des Einsatzes eines proprietären KI-Systems eines externen Anbieters stünde dem typischerweise das Interesse des Herstellers entgegen, diese für die Leistungsfähigkeit der KI entscheidenden Faktoren gerade nicht offenzulegen. Es entsteht somit ein »Blackbox«-Problem, da für die Verarbeitung wesentliche Aspekte der KI gegebenenfalls nicht ausreichend transparent dargestellt werden können.

Recht auf Berichtigung & Löschen

Das Recht auf Berichtigung (Art. 16 DSGVO) erlaubt Individuen, die Korrektur unrichtiger oder unvollständiger sie betreffender personenbezogener Daten zu fordern. Dies bedeutet für KI-Systeme, dass sie über Mechanismen verfügen müssen, die eine schnelle und präzise Anpassung von Daten ermöglichen, sobald ein Fehler festgestellt wird.

Das Recht auf Löschung, auch bekannt als »Recht auf Vergessenwerden« (Art. 17 DSGVO), verpflichtet zur Löschung personenbezogener Daten unter bestimmten Umständen, beispielsweise wenn die Daten für die ursprünglichen Zwecke nicht mehr notwendig sind oder die betroffene Person ihre Einwilligung widerruft. In Bezug auf KI stellt dies sicher, dass Systeme in der Lage sind, Daten effizient und vollständig zu entfernen, wenn dies erforderlich ist.

Ein weiterer wichtiger Aspekt im Kontext von KI und Datenschutz ist das Datamanagement. Vor der Nutzung von Daten in der Produktion müssen diese validiert und verifiziert werden. Dies gewährleistet, dass KI-Systeme mit zuverlässigen und akkuraten Informationen arbeiten. Dies gilt

nicht nur für die Eingangsdaten, sondern auch für Daten, die während des Produktionsprozesses generiert werden.

Eine besondere Herausforderung im Umgang mit KI und Datenschutz ist das Konzept des »Unlearning«. Es geht darum, wie KI-Systeme Daten »vergessen« oder aus ihren Modellen entfernen können, insbesondere im Hinblick auf die Anforderungen des Rechts auf Löschung. Dies ist ein komplexer Prozess, da KI-Modelle dazu neigen, Informationen tief in ihren Strukturen zu verarbeiten und zu speichern. Es erfordert daher fortschrittliche Techniken, um sicherzustellen, dass die betreffenden Daten vollständig und effektiv aus den Modellen entfernt werden, ohne die Integrität oder Leistungsfähigkeit der KI zu beeinträchtigen.

Sonstige Betroffenenrechte

Die DSGVO legt einen besonderen Fokus auf den Schutz personenbezogener Daten und räumt Betroffenen spezifische Rechte ein. Diese Rechte sind auch im Fall der Verarbeitung personenbezogener Daten durch KI von großer Bedeutung, wobei ihre Umsetzung besondere Herausforderungen mit sich bringt.

Eine der Hauptforderungen der DSGVO ist die Transparenz in der Verarbeitung personenbezogener Daten. Im KI-Kontext wird dies durch die Komplexität oder unvollständige Informationen zur Datenverarbeitung bei proprietären Systemen (Blackbox) und die oft intransparenten Entscheidungsprozesse erschwert. KI-Systeme können so programmiert sein, dass sie selbstlernend sind und Entscheidungen auf der Grundlage von Algorithmen treffen, die für Außenstehende schwer nachvollziehbar sind. Dies wirft Fragen hinsichtlich der Erklärbarkeit und Nachvollziehbarkeit von KI-Entscheidungen auf, die für die Wahrung der Transparenzpflicht essenziell sind.

Ein weiteres Recht ist das Auskunftsrecht der betroffenen Person. Gemäß der DSGVO haben Individuen das Recht zu erfahren, ob und wie ihre persönlichen Daten verarbeitet werden. Bei KI-Anwendungen kann es schwierig sein, präzise Auskünfte zu geben, insbesondere wenn es um komplexe Datenverarbeitungsvorgänge geht. Unternehmen müssen daher sicherstellen, dass sie auch bei komplexen KI-Operationen in der Lage sind, klare und verständliche Informationen über die Datenverarbeitung zu liefern.

Ein zentrales Element der DSGVO ist auch das Recht auf Widerruf einer Einwilligung und der Widerspruch gegen die Datenverarbeitung (Art. 21). Sofern ein Betroffener seine Einwilligung widerruft oder der Datenverarbeitung widerspricht, müssen Unternehmen in der Lage sein, die Verarbeitung dieser Daten unverzüglich zu stoppen. Dies kann insbesondere bei fortgeschrittenen KI-Systemen, die auf großen Datenmengen trainiert wurden, eine Herausforderung darstellen.

Datenminimierung & Speicherbegrenzung

Der Grundsatz der Datenminimierung besagt, dass nur so viele personenbezogene Daten verarbeitet werden sollen, wie unbedingt nötig sind, um den festgelegten Zweck zu erfüllen. Im Kontext generativer KI kann dies herausfordernd sein. Die Systeme sind oft auf große Datenmengen angewiesen, um effektiv zu lernen und präzise zu sein. Hier muss ein Gleichgewicht gefun-

den werden zwischen dem Bedarf an umfangreichen Daten für das Training von KI-Modellen und dem Schutz der Privatsphäre.

Der Grundsatz der Speicherbegrenzung verlangt, dass personenbezogene Daten nur so lange gespeichert werden, wie sie für den Zweck ihrer Verarbeitung benötigt werden. Im KI-Bereich ist es oft schwierig, den genauen Zeitraum zu bestimmen, für welchen Daten benötigt werden, insbesondere bei lernenden Systemen, die kontinuierlich auf Daten zugreifen. Es ist wichtig, klare Kriterien für die Aufbewahrungsdauer zu definieren und die Daten zu löschen, sobald sie ihren Zweck erfüllt haben.

Datensicherheit

Die DSGVO stellt hohe Anforderungen an die Datensicherheit. Artikel 32 verlangt von Unternehmen, Schutzmaßnahmen nach dem aktuellen Stand der Technik umzusetzen. Dies beinhaltet eine fortlaufende Überwachung und Anpassung der Sicherheits- und Datenschutzpraktiken von KI-Systemen. Zudem fordert die DSGVO Transparenz und Rechenschaft, insbesondere im Hinblick auf die Rechtfertigung der Datenverarbeitung. Ein besonderes Problem ist die »Richtigkeit der Daten«, da KI-Systeme auf Basis fehlerhafter Daten irreführende Ergebnisse liefern können. Um die DSGVO-Anforderungen zu erfüllen, müssen Unternehmen ihre KI-Systeme kontinuierlich überwachen und validieren.

Vorbehalt menschlicher Entscheidung

Die DSGVO setzt mit Art. 22 eine wichtige Begrenzung für automatisierte Entscheidungsfindungssysteme, einschließlich Profiling, fest.[7] Dieser Artikel beschränkt die Verwendung solcher Systeme, wenn deren Entscheidungen rechtliche Wirkungen haben oder die betroffene Person erheblich beeinträchtigen. Es wird klargestellt, dass Entscheidungen, die signifikante Auswirkungen auf Einzelpersonen haben, nicht ausschließlich auf automatisierten Verarbeitungsprozessen basieren dürfen. Die Definition von »automatisierten Entscheidungen« umfasst Entscheidungen, die ohne menschliches Eingreifen getroffen werden und auf algorithmischer Datenverarbeitung basieren. Für das Verbot gelten jedoch bestimmte Voraussetzungen: Die Entscheidung muss rechtliche Wirkungen entfalten oder die Person erheblich beeinträchtigen. Es existieren aber Ausnahmen von diesem Verbot, wenn beispielsweise die automatisierte Entscheidung für den Abschluss oder die Erfüllung eines Vertrags notwendig ist, die betroffene Person ausdrücklich eingewilligt hat oder die automatisierte Entscheidung durch das Recht der EU oder der Mitgliedstaaten erlaubt ist, sofern dieses Recht angemessene Maßnahmen zum Schutz der Rechte und Freiheiten sowie der berechtigten Interessen der betroffenen Person vorsieht. Praktisch relevant dürfte wohl nur die Einwilligung sein.

Des Weiteren ist eine Unterscheidung zwischen entscheidungsunterstützenden Systemen und selbstentscheidenden Systemen wichtig. Während entscheidungsunterstützende Systeme lediglich Daten, Analysen oder Empfehlungen für menschliche Entscheidungsträger liefern, treffen selbstentscheidende Systeme Entscheidungen ganz ohne menschliches Eingreifen. In der

7 KI gestützte Entscheidungsfindungssysteme stellen in der Regel keine generative KI dar, dennoch gilt es die Bestimmungen von Art. 22 DSGVO zu beachten, insbesondere, wenn Ergebnisse generativer KI für automatisierte Entscheidungen genutzt werden.

DSGVO ist insbesondere die Nutzung der selbstentscheidenden Systeme reglementiert, da diese ein hohes Potenzial haben, erhebliche Auswirkungen auf Einzelpersonen auszuüben. Ziel des Artikels 22 der DSGVO ist es, die potenziellen Risiken der Automatisierung für die Grundrechte und Freiheiten von Einzelpersonen zu begrenzen und sicherzustellen, dass Entscheidungen, die erhebliche Auswirkungen haben, eine menschliche Überprüfung beinhalten.

Datenschutz-Folgenabschätzung

Die Durchführung einer Datenschutz-Folgenabschätzung ist gemäß Art. 35 DSGVO erforderlich, wenn eine Verarbeitung personenbezogener Daten voraussichtlich zu einem hohen Risiko für die Rechte und Freiheiten der betroffenen Personen führt. Dies kann insbesondere beim Einsatz neuer Technologien, wozu KI-Anwendungen im Regelfall zählen dürften, der Fall sein.

In der Datenschutz-Folgenabschätzung sind die Verarbeitungsvorgänge und die Risiken der Verarbeitung für Betroffene sowie die getroffenen technisch-organisatorischen sowie andere risikominimierende Maßnahmen zu beschreiben und zu bewerten. Wie schon bei den Anforderungen bezüglich Informationspflichten und Transparenz liegt auch hier eine besondere Hürde in der Komplexität und Dynamik von KI-Systemen, die bereits eine Beschreibung der Verarbeitung, aber insbesondere auch eine Bewertung der Risiken, erschweren können.

Grundsätzlich nehmen die deutschen Aufsichtsbehörden an, dass der Einsatz von generativer KI mit einem hohen Risiko verbunden ist. Verbleibt unter Berücksichtigung aller risikominimierenden Maßnahmen in der Bewertung ein hohes Risiko, muss gemäß Art. 36 DSGVO die zuständige Aufsichtsbehörde konsultiert werden.

3.1.1.3 Praxishinweise

Anonymisierung im Kontext der KI

Anonymität liegt gemäß Erwägungsgrund 26 DSGVO vor, wenn die Identifizierbarkeit einer natürlichen Person nicht oder nur unter dem Einsatz von nach allgemeinem Ermessen unwahrscheinlichen Mitteln möglich ist. Dabei sind alle Mittel zu berücksichtigen, die vernünftigerweise zur Identifikation genutzt werden könnten, unabhängig davon, ob sie technischer, rechtlicher oder sonstiger Natur sind. Als Maßstab, ob es wahrscheinlich ist, dass bestimmte Mittel zur Identifizierung genutzt werden oder nicht, sind insbesondere die erforderlichen Kosten, Fähigkeiten sowie der benötigte Zeit- und Arbeitsaufwand relevant.

Es ist nicht erforderlich, dass sich alle zur Identifizierung einer Person erforderlichen Informationen in den Händen der jeweils betrachteten Stelle befinden, sondern es reicht aus, dass ein realistischerweise nutzbarer Weg vorhanden ist, die Informationen von einer anderen Stelle zu erlangen.

Der Status der Anonymität muss regelmäßig überprüft werden, da zum Beispiel aufgrund neuer technologischer Entwicklungen eine Identifikation gegebenenfalls zu einem späteren Zeitpunkt leichter möglich sein kann als zum Zeitpunkt der ursprünglichen Bewertung.

Unter dem Begriff Anonymisierung versteht man Verfahren, mit denen personenbezogene Daten dahingehend verändert werden, dass sie die Identifikation einer natürlichen Person nicht mehr oder nur noch mit dem oben beschriebenen unverhältnismäßigen Aufwand ermöglichen. Anony-

misierungsverfahren arbeiten entweder auf Basis von Randomisierung oder Generalisierung. Die eingesetzten Verfahren sollten stets dem Stand der Technik entsprechen.
Auch der Vorgang der Anonymisierung stellt eine Verarbeitung personenbezogener Daten im Sinne der DSGVO dar und erfordert somit eine Rechtsgrundlage.
Für KI liegt im Kontext der Anonymisierung eine große Herausforderung im Umfang des ihr verfügbaren »Hintergrundwissens«, aus dem sie Rückschlüsse ziehen könnte, die zu einer Identifizierbarkeit führen. Ein weiteres Problem liegt im sogenannten »Prompt Engineering«, über das die KI, insbesondere in Anwendungsfällen mit Eingabe in Form von Freitext, durch Anwenderinnen und Anwender gezielt dazu angeleitet werden könnte, eine eigentlich nicht vorgesehene De-Anonymisierung vorzunehmen oder Nutzerinnen und Nutzer die Möglichkeit haben, zusätzliches zu einer Identifizierbarkeit führendes Hintergrundwissen einzugeben.

Zusammenarbeit mit Dienstleistern

Die Zusammenarbeit mit Dienstleistern beim Einsatz von KI kann sich insbesondere dahingehend beschwerlich gestalten, dass die gesetzlich geforderten Kontrollmöglichkeiten gegenüber Dienstleistern häufig ihre Grenzen darin finden dürften, dass es oft nicht den Interessen der Hersteller von KI-Modellen entspricht, für die Leistungsfähigkeit der KI entscheidende Faktoren wie den Algorithmus oder die Trainingsdaten- und Methoden offenzulegen. Zertifizierungen, Gütesiegel oder andere Regularien können dabei helfen, bei den Anbietern bzw. Dienstleistern die Offenlegung wesentlicher Punkte, z. B. aus dem EU-AI-Act, anzufragen. Lässt sich der Anbieter nicht darauf ein, sind die anwendenden Unternehmen in der Pflicht, ggf. den Anbieter zu wechseln. Denn die Verantwortung für die Erfüllung von Vorgaben wie Anti-Diskriminierung oder Transparenz liegt nicht allein beim Anbieter bzw. Dienstleister: Auch die anwendenden Unternehmen müssen ihrer Verantwortung hinsichtlich rechtlicher Vorgaben gerecht werden.

3.2 Cybersicherheit

Effektive Cybersicherheit (= IT-Sicherheit) ist insbesondere für Hochrisiko-KI-Systeme und für die zugrundeliegenden KI-Modelle essenziell und erfordert u. a. auch durch den Gesetzgeber (vgl. AI Act, Art. 15.5 und Art. 55.1) sowohl effektive Schutzmaßnahmen als auch hinreichende Nutzerinformationen über das jeweils realisierte Cybersicherheitsschutzlevel (vgl. AI Act, Art. 13.3).
Dabei unterscheiden sich die meisten KI-Systeme bzgl. Cybersicherheitsrisiken und den dafür anwendbaren Schutzmaßnahmen nicht wesentlich von klassischen IT-Systemen bzw. klassischen IT-Produkten. Folglich betrachtet v. a. die im Dezember 2024 verabschiedete horizontale Regulierung zur Produkt-Cybersicherheit in der EU (»Cyber Resilience Act«, CRA, EU 2024/2847) die Anforderungen zur Cybersicherheit von Hochrisiko-KI-Systemen aus Artikel 15.5 des AI Acts als erfüllt, sofern das KI-System die grundlegenden Cybersicherheitsanforderungen des CRA erfüllt (vgl. CRA, Art. 12).

Die notwendigen, grundlegenden Cybersicherheitsanforderungen des CRA ergeben sich wiederum aus der individuellen Risikoanalyse des jeweiligen KI-Systems gemäß dessen »bestimmungsgemäßen Zweck oder [dessen] vernünftigerweise vorhersehbare Verwendung« (CRA, Art. 2.1) über die gesamte Planungs-, Entwurfs-, Entwicklungs-, Produktions-, Liefer- und Wartungsphase, einschließlich aller Komponenten und Daten von Drittherstellern oder eingesetzter OSS-Quellen.
Die mit dem CRA verbundenen Obligationen umfassen neben der bereits genannten verpflichtenden Risikoanalyse (vgl. CRA. Art. 13 Absatz (2)) im Wesentlichen:

- Technische Schutzmaßnahmen (vgl. CRA, Anhang I, Teil I)
- Regelmäßiges Testen und verpflichtendes Monitoring (vgl. CRA, Anhang I Teil II Absatz (1) und (3))
- Meldung und Behandlung von Schwachstellen und Sicherheitsvorfällen (vgl. CRA, Art. 14 und Anhang I Teil II)
- Dokumentation & Kennzeichnung (vgl. CRA, Art. 31 und Anhang VII)

Der Nachweis zu den o.g. Anforderungen erfolgt über eine Herstellerselbsterklärung oder durch entsprechend notifizierte Dritte (vgl. CRA, Art. 32).

Dennoch gibt es einige zusätzliche KI-spezifische Angriffspunkte (»AI specific vulnerabilities«), wenngleich diese auch den Bereichen Genauigkeit und Robustheit zugeordnet werden können. Im Folgenden ein kurzer Überblick der bekanntesten KI-spezifischen Angriffspunkte.

Datenvergiftung (»data poisoning«): Bei diesem Angriff werden schädliche, falsche oder irreführende Daten ins Trainingsset eingeschleust, um die darauf resultierenden Ausgaben der KI zu manipulieren. Ziel des Angreifers ist die Erzeugung falscher oder tendenziöser Ausgaben.
Da sich dieser Angriff letztlich auf einen unzureichenden Schutz der Zugriffsberechtigungen bzw. die Integrität und/oder Authentizität der Trainingsdaten zurückzuführen lässt, kann und sollte man diesen Bedrohungen mit klassischen Schutzmaßnahmen der Cybersicherheit (z. B. Zugriffskontrolle, digitale Signaturen etc.) effektiv begegnen. Zusätzlich lässt sich durch »robust training«, d. h. durch die Verwendung von Techniken wie »out-of-distribution detection« oder »adversarial training«, das Modell resistenter gegen Datenvergiftung machen.

Eingabeangriffe (»input attacks« auch »evasion attacks« oder »adversarial examples«): Angreifer versuchen, durch speziell präparierte Eingaben (»prompt injections«) dem KI-System Antworten zu entlocken, die vom Anbieter des KI-Systems nicht vorgesehen oder erlaubt sind (z. B. geschützte Personendaten).

Da diese Angriffspunkte v.a. in der (unzureichenden) Funktionalität des KI-Systems begründet sind, sind hier v.a. Maßnahmen zur Verbesserung der Genauigkeit, Qualität und Robustheit angebracht. Durch die Überprüfung aller Eingabedaten auf Anomalien oder manipulierte Inhalte (»input sanitization«) können zum Beispiel viele Arten von Eingabeangriffen abgewehrt werden. Für diese Überprüfung kann wiederum KI eingesetzt werden. Mit Techniken der »differential privacy« können zudem die Identifizierbarkeit von Einzelpersonen in den Daten minimiert und Reidentifikationsangriffe erheblich erschwert werden.

Modell-Extraktion (»model extraction attack«): Angreifer extrahieren Informationen über die Architektur oder die Parameter eines KI-Modells, um ein ähnliches Modell zu erzeugen. Damit könnten sie dann ein eigenes KI-System trainieren oder das Originalmodell gezielt kompromittieren.

Bei diesem KI-spezifischen Angriffspunkt handelt es sich im Wesentlichen um (i. d. R. legales; vgl. § 3 Abs. 1, GeschGehG) Reverse-Engineering und ist damit nur unzureichend via Cybersicherheit direkt adressierbar, sondern eher über funktionale Einschränkungen (z. B. begrenzte Abrufe pro Zeiteinheit). Darüber hinaus können strenge Zugriffskontrollen und Verschlüsselungsmechanismen während der Verarbeitung dabei helfen, unbefugten Zugriff auf das Modell und seine Parameter einzuschränken, was wiederum die Gefahr einer Modell-Extraktion mindert.

Denial-of-Service (DoS): Bei diesem Angriff geht es darum, das KI-System durch eine Flut von Anfragen zu überlasten, sodass es für legitime Nutzer unzugänglich wird.

Eine einfache, aber effektive Methode zur Abwehr von DoS-Angriffen ist die Begrenzung der Anzahl der Anfragen an das System pro Zeiteinheit (»rate limiting«).

Grundsätzlich sind für alle KI-Systeme und KI-Modelle wie für jedes IT-System auch regelmäßige Sicherheitsüberprüfungen (inkl. Penetrationstests, Fuzzing etc.) sowie regelmäßige Aktualisierungen unerlässlich, um kontinuierlich auf neue Bedrohungsvektoren reagieren zu können. Nachweis über Herstellerselbsterklärung oder durch entsprechend notifizierte Dritte (TBD).

Social Engineering: Die Kombination von Künstlicher Intelligenz und Social-Engineering-Angriffen, wie personalisierte Phishing-Mails, Deepfakes oder gezielte Täuschungsversuche, stellt eine wachsende Bedrohung dar. Viele herkömmliche Cybersicherheitsprodukte sind nicht in der Lage, diese Bedrohungen zu bewältigen. Es ist deswegen wichtig, sich an aktuelle BSI-Mindeststandards und -Richtlinien für Unterstützung zu wenden.

3.3 Schutzrechte

3.3.1 Urheberrecht

3.3.1.1 Schutzfähigkeit des Outputs

Schutzfähigkeit als Werk

Gemäß § 2 Abs. 2 UrhG sind lediglich persönliche geistige Schöpfungen urheberrechtlich als Werke geschützt. Das aktuelle Verständnis des Urheberrechts setzt einen direkten Bezug zwischen Mensch und geistiger Schöpfung voraus. Daher sind rein maschinell geschaffene Inhalte vom Urheberrechtsschutz ausgenommen. Dies basiert auf der Annahme, dass nur ein Mensch eine persönlich-geistige Schöpfung hervorbringen kann. In diesem Sinne sind Inhalte, die von Künstlicher Intelligenz generiert werden, keine schutzfähigen »Werke« der KI.

In aller Regel ist der Output auch kein schutzfähiges Werk des Nutzers, der den Prompt eingibt. Dem Output fehlt ein hinreichender Bezug zu einem menschlichen Urheber, da die konkrete Ausgestaltung durch die KI vorgenommen wird. Ausnahmsweise können KI-Erzeugnisse dann urheberrechtlichen Schutz genießen, wenn ein Mensch ein KI-System lediglich als Werkzeug verwendet und alle gestalterischen Entscheidungen selbst trifft (sog. »AI-assisted works«). Ausschlaggebend ist der Einfluss des Nutzers auf das Erzeugnis. Ein urheberrechtlicher Schutz kann dann in Erwägung gezogen werden, wenn der menschliche Nutzer dem KI-System derart präzise Anweisungen gibt, dass die endgültige Gestaltung des Werkes bereits festgelegt ist und die KI lediglich diese Gestaltung umsetzt. Denn in diesem Fall nutzt der Mensch den kreativen Gestaltungsspielraum beim Schaffen des Werkes selbst aus und verwendet die KI nur als technisches Hilfsmittel. In Anbetracht der Vielfalt möglicher Ausgestaltungen und der Unvorhersehbarkeit von KI-Output dürfte dies aber nur in wenigen Ausnahmefällen überhaupt in Betracht kommen. Die Eingabeinformationen oder Prompts liefern in der Regel eher eine Idee oder ein Konzept, auf deren Basis die KI dann arbeitet, was allein nicht für die Annahme eines maßgeblichen Einflusses ausreicht und für sich genommen nicht schutzfähig ist. Abzustellen ist auf den konkreten Einzelfall.

In eine ähnliche Richtung argumentiert das United States Copyright Office: nur Werke, die das Produkt hinreichender menschlicher Kreativität sind, können urheberrechtlich geschützt sein.[8] Das Office lehnt die Registrierung und damit implizit die Schutzfähigkeit von Inhalten ab, die ausschließlich von einer KI-Technologie ohne menschlichen Beitrag erzeugt wurden. Es akzeptiert jedoch die Registrierung von Werken, die sowohl menschliche als auch KI-generierte Elemente enthalten, sofern ein Mensch die schöpferische Kontrolle über den kreativen Ausdruck des Werks hatte und gerade diejenigen Werkbestandteile verantwortet, die klassischerweise eine menschliche Kreativität ausmachen. Das Office geht davon aus, dass über Prompts bei den gegenwärtig

8 US Copyright Office, Copyright and Artificial Intelligence - Part 2: Copyrightability, 29.1.2025; Copyright Registration Guidance: Works Containing Material Generated by Artificial Intelligence, 16.3.2023, jeweils abrufbar unter: https://www.copyright.gov/ai.

verfügbaren KI-Anwendungen keine solche hinreichende menschliche Kontrolle über die wesentlichen Elemente kreativen Ausdrucks ausgeübt wird. Der Einsatz von KI, die die menschliche Kreativität nicht ersetzt, sondern lediglich unterstützt, steht einer Schutzfähigkeit aber nicht per se entgegen. Erforderlich ist stets eine Einzelfallbetrachtung. Bei der Stellung eines Antrags auf Registrierung eines Werks muss der Antragsteller daher die Einbeziehung von KI-generierten Inhalten detailliert offenlegen und die menschlichen Beiträge erläutern.

Demgegenüber gibt es einige Länder, die die Grundlage dafür schaffen, dass unter Umständen auch der Programmierer als Urheber gelten kann. Das sind insbesondere, aber nicht abschließend, Großbritannien (Section 9 (3) Copyright, Designs and Patents Act 1988), Neuseeland (Section 5 (2) lit. a Copyright Act 1994) und Irland (Section 21 lit. f Copyright and Related Rights Act 2000). In allen drei zitierten Gesetzen findet sich folgende Formulierung:

- »the author shall be taken to be the person by whom the arrangements necessary for the creation of the work are undertaken«. Unter »arrangements necessary for the creation« ist die Erstellung der Künstlichen Intelligenz, also des Algorithmus, zu verstehen.

Insgesamt kann man also festhalten, dass an dem von KI generierten Output in der Regel kein Urheberrecht entsteht. Ein urheberrechtlicher Schutz kann jedoch nachfolgend durch eine hinreichend kreative Bearbeitung des Outputs erreicht werden. Dies bedeutet, dass der Output einer KI zur urheberrechtlichen Schutzfähigkeit gelangen kann, wenn er von einer natürlichen Person so bearbeitet wird, dass ihm ein eigener schöpferischer Gehalt zukommt. Für den Verwender ist es ratsam, den eigenen Schaffensprozess hinreichend zu dokumentieren. Denn nach allgemeinen Grundsätzen hat derjenige den urheberrechtlichen Schutz und die eigene Rechtsstellung am Werk darzulegen und zu beweisen, der sich darauf beruft. Soweit darauf abgestellt wird, dass Künstliche Intelligenz als Werkzeug eingesetzt wurde, muss also der Verwender seinen eigenen Beitrag, der die Schutzfähigkeit begründet, darlegen und beweisen. Da bisher menschliche Schöpfungen der Regelfall waren, wurde meist ungeprüft unterstellt, dass eine menschliche kreative Leistung das Werk geschaffen hat. Dieser Nachweis für einen eigenen menschlichen schöpferischen Beitrag unter Nutzung einer KI als Werkzeug dürfte künftig deutlich schwieriger werden, spätestens wenn KI-Erzeugnisse für bestimmte Ergebnisse eher Regel statt Ausnahme sind. Dann bedarf es neuer bzw. veränderter Lösungen zum Nachweis der persönlichen geistigen Schöpfungen.

Schutzfähigkeit über Leistungsschutzrechte

Das Urheberrecht bietet neben dem Werkschutz auch den Schutz verwandter Schutzrechte, den sogenannten Leistungsschutzrechten gemäß den §§ 70ff. UrhG.

Die Voraussetzungen für den Schutz von Leistungsschutzrechten nach UrhG variieren stark. Gemeinsam ist ihnen aber, dass sie – anders als das Schutzrecht für Werke nach § 2 Abs. 2 UrhG – nicht an eine persönliche geistige Schöpfung anknüpfen, sondern auf »Leistungen anderer Art« abzielen, welche »der schöpferischen Leistung des Urhebers ähnlich sind oder in Zusammenhang mit den Werken der Urheber erbracht werden« (siehe amtliche Begründung zum Entwurf des Urheberrechtsgesetzes, BT-Drucksache 4/270 vom 23. März 1962, S. 33 f.). Die verwandten Schutz-

rechte begründen daher Rechtspositionen sowohl für kreative Leistungen, die unterhalb der Werkschöpfungsschwelle bleiben, als auch für organisatorische und wirtschaftliche Leistungen. Welches Leistungsschutzrecht im konkreten Fall in Betracht kommt, hängt von der Art des generierten Outputs und seiner Anschlussverwendung ab. Denkbar ist mit Blick auf KI generierten Output das Leistungsschutzrecht des Tonträgerherstellers nach § 85 f. UrhG, der Laufbildschutz nach § 95 UrhG, sowie in einem gewissen Rahmen das Recht des Datenbankherstellers nach § 87a – § 87e UrhG.

Soweit der Output einer Künstlichen Intelligenz in Tonfolgen besteht, kann die Erstfixierung diesen Outputs auf einem Tonträger potenziell unter das Tonträgerherstellerrecht aus § 85 f. UrhG fallen. Dieses Recht dient einem Investitionsschutz, geschützt wird die wirtschaftliche, technische und organisatorische Herstellerleistung der Aufzeichnung von Tonmaterial auf einen Tonträger (BGH, Urteil vom 20.11.2008 - I ZR 112/06, GRUR 2009, 403, 404). Eine persönliche geistige Schöpfung ist dagegen nicht erforderlich. Daher können auch KI-generierte Tonfolgen unter den Schutzbereich fallen, sobald diese erstmals fixiert werden.

Vergleichbar verhält es sich beim Recht des Laufbildschutzes nach § 95 UrhG. Auch dieses Recht dient dem Schutz der wirtschaftlichen und organisatorischen Verantwortung (BGH, Urteil vom 6.2.2014 - I ZR 86/12, GRUR 2014, 363), hier des Filmherstellers i. S. d. § 94 UrhG. Geschützt wird eine solche Leistung im Hinblick auf die Fixierung von Bildfolgen (Stummfilm) oder Bild- und Tonfolgen (Tonfilm) auf einen Bild- oder Bild-Ton-Träger, durch die der Eindruck eines Bewegtbildes entsteht, ohne dass eine persönliche geistige Schöpfung im Sinne eines Filmwerks vorliegen muss. Da für den Laufbildschutz der künstlerische »Wert« des Films also unerheblich ist, kommt der Leistungsschutz für den Filmhersteller auch dann in Betracht, wenn der fixierte Film von einer KI generiert wurde.

Sowohl das Recht des Tonträgerherstellers als auch das Schutzrecht für Laufbilder stünden dann demjenigen zu, der die wirtschaftliche und organisatorische Verantwortung für die Fixierung des Outputs in tatbestandlicher Form trägt. Dies kann denkbar auch der Programmierer der Künstlichen Intelligenz bzw. das Unternehmen sein, das die KI betreibt, ebenso aber auch eine dritte Person.

Das Datenbankherstellerrecht der §§ 87a ff. UrhG schützt nach Art und Umfang wesentliche Investitionen in die Herstellung, Überprüfung oder Darstellung einer Datenbank. Gemäß § 87a Abs. 2 UrhG ist derjenige Datenbankhersteller und damit Rechtsinhaber, der diese Investitionen getätigt hat. Wichtig ist, dass es sich bei der Datenbank im urheberrechtlichen Sinne um eine Sammlung von Werken, Daten oder anderen unabhängigen Elementen handeln muss, die systematisch oder methodisch angeordnet und einzeln mithilfe elektronischer Mittel oder auf andere Weise zugänglich sind. Die Elemente gelten als unabhängig, wenn sie einen eigenständigen Informationsgehalt aufweisen. Es muss möglich sein, die Elemente zu trennen, ohne dass sie durch die Trennung ihren Informationsgehalt verlieren. Stellt das KI-generierte Ergebnis also eine Datenbank im Sinne der §§ 87a ff. UrhG dar, kommt das Leistungsschutzrecht des Datenbankherstellers in Betracht. Entscheidend ist dabei, dass dem Rechtsinhaber nach § 87b UrhG nur das ausschließliche Recht zur Vervielfältigung, Verbreitung und öffentlichen Wiedergabe der gesamten Datenbank oder eines nach Art oder Umfang wesentlichen Teils der Datenbank zusteht. Einzelne

oder eine Vielzahl von Datenbankelementen unterhalb der Wesentlichkeitsschwelle sind vom Ausschließlichkeitsrecht nicht erfasst.

Regelmäßig nicht in Betracht dürfte hingegen das Schutzrecht für Lichtbildhersteller kommen. Dafür dürfte einerseits bereits der Schutzbereich nicht eröffnet sein. Der Lichtbildschutz nach § 72 UrhG steht nur bei Abbildern offen, die mittels strahlender Technik[9] erstellt werden und die Wirklichkeit naturgetreu wiedergeben. Dies ist bei computergenerierten Abbildern gerade nicht der Fall (vgl. auch KG Berlin, Urteil vom 16.1.2020, U 12/16 Kart), weshalb der entsprechende Schutzbereich auch bei KI-generierten Bildnissen nicht eröffnet sein dürfte. Andererseits mangelt es regelmäßig auch an dem Erfordernis der menschlichen geistigen Leistung. Die Rechtsprechung geht nämlich davon aus, dass das Leistungsschutzrecht des Lichtbildherstellers nach § 72 UrhG ein Mindestmaß an persönlicher (wenn auch nicht schöpferischer) geistiger Leistung erfordert. Bei einem von Künstlicher Intelligenz geschaffenen und auf Algorithmen basierenden Output dürfte es hieran fehlen. Etwas anderes könnte nur dann in Betracht kommen, wenn ein Mensch einen hinreichend bestimmenden Einfluss auf einen durch KI durchgeführten Prozess zur Erstellung von Lichtbildern hat.

Die Frage, ob von KI generierte Texte unter das neuere Leistungsschutzrecht des Presseverlegers fallen, ist aktuell unklar und in der juristischen Literatur umstritten. Teilweise wird darauf abgestellt, dass nur ein aus journalistischen und damit menschlichen Beiträgen zusammengestelltes Online-Angebot geschützt wird und keine Zusammenstellung computergenerierter Produkte, wobei freilich Mischformen denkbar sind. Teilweise wird darauf verwiesen, dass das Leistungsschutzrecht wiederum den wirtschaftlichen, organisatorischen und technischen Aufwand der Veröffentlichung schützt, unabhängig davon, wie die veröffentlichten Inhalte zustande gekommen sind. Hier wird man die weiteren Entwicklungen abwarten und auf Entscheidungen der Gerichte warten müssen.

9 So bspw. beim Analogfilm aber auch beim Sensor in der Digitalfotografie.

Wegen des lückenhaften Schutzes sollte je nach Einzelfall über alternative Schutzkonzepte wie vertragliche Regelungen oder Kopierschutzmechanismen nachgedacht werden.

3.3.1.2 Haftung durch Eingabeinformation

Neben den selbst formulierten Handlungsanweisungen können Eingabeinformationen und Prompts auch geschützte Inhalte Dritter enthalten. Besonders relevant wird dies, wenn in den Eingabeinformationen oder Prompts ein geschütztes Werk direkt wiedergegeben wird, zum Beispiel in der Form: »Fasse den [geschützten Text XYZ] zusammen«, wobei der Text in die Eingabeinformationen bzw. den Prompt kopiert wird oder (audio-)visuelle Inhalte zur weiteren Bearbeitung durch die KI hochgeladen werden. Hierdurch könnte jedenfalls in das Vervielfältigungsrecht aus § 16 UrhG an den eingegebenen urheberrechtlich geschützten Inhalten eingegriffen werden.

Es ist wichtig zu beachten, dass die zustimmungsfreie Nutzung urheberrechtlich geschützter Werke eine urheberrechtliche Verletzungshandlung darstellt, soweit keine urheberrechtliche Schranke vorliegt. Betreibt man ein neuronales Netzwerk nicht auf eigenständiger Hardware, die vom Internet isoliert ist, muss man davon ausgehen, dass man eine Vervielfältigung auf den Servern des KI-Anbieters veranlasst. Diese Server löschen nicht in jedem Fall die Eingabeinformationen bzw. Prompts, sondern können sie zur Optimierung der KI verwenden, was urheberrechtlich relevant sein kann.

Die urheberrechtlichen Schranken des § 44a UrhG (vorübergehende Vervielfältigung) und § 53 UrhG (Privatkopie) dürften in diesem Kontext nicht anwendbar sein. § 44a UrhG erlaubt die vorübergehende Vervielfältigung nur dann, wenn sie keine eigenständige wirtschaftliche Bedeutung hat. Die Vervielfältigung in KI-Servern ist jedoch nicht nur vorübergehend, sondern kann für Trainings- und Optimierungszwecke der KI längerfristig genutzt werden. § 53 UrhG erlaubt die Nutzung nur im privaten Kreis und für den eigenen Gebrauch. Jedoch verlässt mit der Eingabe urheberrechtlich geschützter Werke in ein KI-System die Vervielfältigung und damit Nutzung wohl die Privatsphäre, was über den privaten Gebrauch hinausgehen dürfte.

Die faktischen Haftungsrisiken sind in diesem Zusammenhang jedoch zunächst als gering einzuschätzen. Das liegt daran, dass Informationen darüber, wer Vervielfältigungen auf den Servern der KI-Anbieter veranlasst, in der Regel nicht nach außen dringen. Allerdings ist besondere Vorsicht geboten, da durch solche Handlungsanweisungen die Wahrscheinlichkeit von Rechtsverletzungen durch den Output der KI signifikant erhöht wird und diesbezüglich auch eine Verantwortlichkeit begründet werden kann (siehe dazu weiter unten).

3.3.1.3 Relevanz der Urheberrechte an KI-Trainingsdaten auf Anbieter-, Betreiber- und Nutzerseite

Die Betrachtung des Anbieterhandelns im Kontext der Künstlichen Intelligenz ist essenziell, da Handlungen der Anbieter direkt Einfluss auf die Anwendungsphase der KI und die Verwendung des Outputs haben können.

Vorbereitung des KI-Trainings durch Zusammenstellen von Trainingsinhalten

Als Grundlage für das Trainieren eines KI-Modells wird in einem ersten Schritt häufig eine Sammlung von urheberrechtlich geschützten Werken und Inhalten erstellt, was die Vervielfältigungsrechte gemäß § 16 UrhG berühren kann. Gegenwärtig ist es noch die Ausnahme, dass KI-Modelle ausschließlich mit unternehmenseigenen und/oder lizenzierten oder solchen Inhalten trainiert werden, die frei von urheberrechtlichen Rechtspositionen sind. Jedenfalls bei den frei zugänglichen generativen KI-Anwendungen werden die zugrundeliegenden Modelle in der Regel mit riesigen Datenmengen trainiert, bei denen es sich um durch sog. Text und Data Mining-Verfahren zusammengesammelte Inhalte aus dem Internet handelt.
Eine zentrale Rolle spielt hierfür die Schranke des § 44b UrhG, die Vervielfältigungen bei der Erstellung von Datensammlungen durch Text und Data Mining auch für kommerzielle Zwecke unter den folgenden Voraussetzungen erlaubt:

1. Rechtmäßige Zugänglichkeit des Trainingsmaterials: Nach § 44b Abs. 2 S. 1 UrhG müssen die verwendeten Werke rechtmäßig zugänglich sein, was der Fall ist, wenn der Anbieter, der sie durch Text und Data Mining »einsammeln« will, diese ohne Rechtsverstoß abrufen kann. Das erfasst alle frei oder über einen individuellen Zugang im Internet verfügbaren Inhalte, unabhängig davon, ob sie dort mit oder ohne Zustimmung des Rechteinhabers verfügbar gemacht wurden.

2. Löschpflicht: Gemäß § 44b Abs. 2 S. 2 UrhG müssen die Vervielfältigungen wieder gelöscht werden, wenn sie für das Text und Data Mining nicht mehr erforderlich sind. Derzeit ist noch nicht rechtssicher geklärt, wann genau die Löschpflicht einsetzt, also ob der Trainingskorpus etwa auch für weiteres oder neues Trainieren des KI-Modells vorgehalten werden kann.

3. Opt-out des Rechtsinhabers: § 44b Abs. 3 UrhG ermöglicht Rechtsinhabern, die Nutzung ihrer Werke für Text und Data Mining durch einen Opt-out (Nutzungsvorbehalt) für die Zukunft auszuschließen. Dies bedeutet, dass die Schranke des § 44b UrhG nur dann eingreift, wenn der Rechtsinhaber sich die Nutzung nicht vorbehalten hat. In Bezug auf im Internet veröffentlichte Werke verlangt § 44b Abs. 3 S. 2 UrhG, dass der Opt-out in maschinenlesbarer Form erklärt wird. Welche Erklärungsformen diese Voraussetzung der Maschinenlesbarkeit erfüllen und damit zulässiges Text und Data Mining ausschließen, wird derzeit eruiert. Die juristische und technische Fachliteratur spricht sich überwiegend für Lösungen wie robots.txt aus, was auch den Entwürfen des vom EU-KI-Büro verantworteten Verhaltenskodex (Codes of Practice) für die KI-Verordnung entspricht. Dagegen tendiert das Landgericht Hamburg in einer ersten Entscheidung zum Text und Data Mining im Kontext des Trainings von KI – indes in einem nicht entscheidungserheblichen obiter dictum – dazu, selbst einen in natürlicher, menschlicher Sprache formulierten allgemeinen Vorbehalt in den Nutzungsbedingungen einer Website genügen zu lassen (LG Hamburg, Urteil vom 27.9.2024 - 310 O 227/23, NJW 2024, 3383).

Auf eine praktisch wirksame Ausgestaltung der Opt-Out-Möglichkeit zielt auch die in der KI-Verordnung festgelegte Pflicht von Anbietern von KI-Modellen mit allgemeinem Verwendungszweck (»General Purpose AI«, »GPAI«-Modelle) ab, eine hinreichend detaillierte Zusammenfassung ihrer Trainingsinhalte zu veröffentlichen (Art. 53 Abs. 1 lit. c). Auch wenn es genügen kann, hier die wichtigsten Datenerhebungen und Datensätze aufzuführen, soll dies doch gerade auch den Urhebern die Ausübung und Durchsetzung ihrer Rechte erleichtern (Erwägungsgrund 107 KI-Verordnung).

Soweit die vorstehenden Voraussetzungen erfüllt sind, ist die Vervielfältigung von Werken und urheberrechtlichen Schutzgegenständen durch Text und Data Mining zulässig. Ganz überwiegend wird davon ausgegangen, dass die Schranke auch für Vervielfältigungen greift, die dann für das Training von (generativer) KI verwendet werden (so auch die Tendenz des LG Hamburg, a. a. O.). In diesem Sinne muss auch der Verweis auf die unionsrechtliche Vorlage des § 44b UrhG, auf Art. 4 der DSM-Richtlinie ((EU) 2019/790) in der KI-Verordnung im Zusammenhang mit GPAI-Modellen und damit generativer KI (Art. 53 Abs. 1 lit. c, Erwägungsgrund 105) verstanden werden. Auch der vom EU-KI-Büro verantwortete Verhaltenskodex basiert auf der Annahme einer Anwendbarkeit der Schranke. Schließlich hat auch die deutsche Bundesregierung in einer parlamentarischen Anfrage die Anwendbarkeit des § 44b UrhG bestätigt.[10]

Es existieren dennoch Meinungen, dass die gegenwärtige Rechtslage die Text und Data Mining-Schranke des § 44b UrhG als Rechtfertigungsbasis für die Zusammenstellung von Trainingsinhalten für KI eventuell nicht unterstützt. Diese Ansicht basiert auf der Argumentation, dass die Verwendung reiner KI-Trainingsdaten lediglich als vorbereitende Maßnahme für spätere Analysen fungiert und somit nur indirekt zur Informationsgewinnung beiträgt. Es wird betont, dass, solange der primäre Zweck nicht die unmittelbare Informationsgewinnung ist – wie es bei der Datenanalyse der Fall sein soll –, diese Vorgehensweise nicht von der genannten Norm abgedeckt ist.

Jedenfalls in Anbetracht der noch offenen Fragen zu den Voraussetzungen und der Aufbewahrungsreichweite des § 44b UrhG sollte der rechtssichere Weg über die Lizenzierung der zum Training genutzten Werke erwogen werden.

Das Schutzlandprinzip ermöglicht es, neuronale Netzwerke in Ländern mit geringeren Urheberrechtsstandards zu trainieren und anschließend als fertiges Produkt in Deutschland anzubieten, wodurch die Opt-Out-Rückausnahme des § 44b UrhG (bzw. unionsweit durch Art. 4 Abs. 3 DSM-Richtlinie) praktisch umgangen werden können. Zu denken wäre dabei etwa an die potenziell günstigere Rechtslage unter dem »fair use« in den USA, wobei die Anwendbarkeit dieser Doktrin auf KI-Trainingsmaterial derzeit in einer Vielzahl von Gerichtsverfahren erst noch geklärt wird. Solchen lokalen Ausflüchten vor der europäischen Opt-Out-Regelung schiebt nun aber die KI-Verordnung faktisch einen Riegel vor. Denn Erwägungsgrund 106 S. 3 KI-Verordnung stellt klar, dass die Pflicht von Anbietern von GPAI-Modellen (also gerade auch generativen KI-Modellen) zur Ermittlung und Einhaltung der Text und Data Mining Rechtsvorbehalte (Art. 53 Abs. 1 lit. c KI-Verordnung) unabhängig davon gilt, in welchem Hoheitsgebiet die urheberrechtlich relevanten Handlungen für das Training dieser Modelle stattfinden. Wegen des internationalrechtlich verankerten Territorialitätsprinzips des Urheberrechts dürfte es sich dabei zwar nicht um eine Ausweitung der Schranke und ihrer Opt-Out-Regelung über die Grenzen der Mitgliedsstaaten und damit der EU hinaus handeln; Vervielfältigungen bei der Zusammenstellung von Trainingsinhalten, die außerhalb der EU entstehen, sind nur nach dem lokalen Urheberrecht zu bewerten und daher nicht an der europäischen Schranke und ihrem Rechtsvorbehalt zu messen. Indem die KI-Verordnung es nun aber zur Anforderung an Anbieter von GPAI-Modellen macht, dass diese Opt-Outs überall beachtet werden, wird gleichsam »durch die Hintertür« einer Produktanforde-

10 Antwort der Bundesregierung auf eine Kleine Anfrage der Fraktion der CDU/CSU, BT-Drs. 20/9714, 12.12.2023, S. 3.

rung die faktische Wirksamkeit der Opt-Out-Regelung über das Gebiet der EU hinaus ausgeweitet.

Neben der Text und Data Mining Schranke für beliebige, auch kommerzielle, Zwecke kann im Einzelfall auch die Schranke des Text und Data Mining für wissenschaftliche Zwecke aus § 60d UrhG (Art. 3 DSM-Richtlinie) eingreifen. Zur Vervielfältigung berechtigt sind Forschungsorganisationen. Eine erste Entscheidung hierzu hat die Bedeutung dieser Schranke gestärkt (LG Hamburg, Urteil vom 27.9.2924 - 310 O 227/23, NJW 2024, 3383). Im Rechtsstreit des Fotografen Kneschke gegen den gemeinnützigen Verein LAION e.V., der Open-Source KI-Trainingsdatenbanken bereitstellt, lässt das Gericht die Schranke zugunsten von LAION gelten, obgleich die Datensätze (Milliarden von Verlinkungen auf im Internet frei verfügbare Bilder mit Kurzbeschreibungen zu diesen) letztlich auch von kommerziellen KI-Akteuren zum Training genutzt werden, mit denen wohl auch personelle Verflechtungen bestehen. Im Wesentlichen stellte das Gericht heraus, dass die Erstellung eines solchen Datensatzes eine hinreichende (Vorbereitungs-)Handlung im Sinne wissenschaftlicher Forschung ist, deren Nutzung durch Dritte genügt und es allein auf den nicht kommerziellen Charakter der konkreten Handlung von LAION, nämlich der Erstellung der Datensätze, ankommt.

Die Schranke für vorübergehende Vervielfältigungen des § 44a UrhG (Art. 5 Abs. 1 InfoSoc-Richtlinie (2001/29/EC)) dürfte dagegen für die Freistellung von Vervielfältigungen bei der Zusammenstellung eines KI-Trainingsdatensatzes in der Regel keine entscheidende Rolle einnehmen. Auch das Landgericht Hamburg hat in den hierfür erfolgten Downloads der Bilddateien durch LAION Vervielfältigungen erkannt, die weder nur flüchtig noch nur begleitend sind.

Das eigentliche Trainieren des neuronalen Netzes

Ob das eigentliche Trainieren eines künstlichen neuronalen Netzes urheberrechtlich relevante Vervielfältigungen i. S. d. § 16 UrhG des Trainingsmaterials bedingt, ist technisch wie rechtlich noch nicht abschließend geklärt. Dagegen kann vorgebracht werden, dass ähnlich dem Lesen eines Buchs durch einen Menschen eine reine Rezeption der Inhalte durch das neuronale Netz stattfindet, das nur abstrakte Informationen entnimmt.

Bei der Fortbildung der KI im Live-Betrieb muss zwischen Training mit eigenen und fremden Daten unterschieden werden. Bei fremden Daten sind die Grenzen des § 44b UrhG zu beachten (s. o.). Bei eigenen Daten müssen diese Grenzen des Urheberrechts nicht berücksichtigt werden,[11] sofern die entsprechenden Rechte für das Training vorliegen, was in der Regel vertragliche Regelungen mit Mitarbeitenden beinhaltet. Das Training stellt wahrscheinlich eine eigenständige Nutzungsart dar, die wirtschaftlich selbstständig verwertbar ist und im Zweifel vertraglich explizit eingeräumt werden muss.

11 Andere Grenzen, die sich bspw. aus Persönlichkeitsrechten, dem Recht am eigenen Bild oder der eigenen Stimme sowie sonstigen personenbezogenen Daten ergeben, gelten jedoch weiterhin und sich zu beachten.

Das KI-Modell als solches

Zudem besteht die Frage, ob Vervielfältigungen der Trainingsinhalte oder wesentlicher Teile dieser im neuronalen Netzwerk vorliegen, was zu einer »Infizierung« des KI-Modells führen könnte. Die herrschende Meinung sieht die Informationsgewinnung aus geschützten Werken und die Anpassung der Gewichtungswerte des neuronalen Netzwerks nicht als eine solche Vervielfältigung an. Es gibt jedoch auch die Ansicht, dass die in das neuronale Netzwerk eingeflossenen Parameter eine Vervielfältigung darstellen könnten, was erhebliche Folgen hätte, da ein gezieltes Entfernen dieser Parameter effektiv nicht möglich ist.

Risiken für Betreiber und Nutzer

Für Betreiber und Nutzer des KI-Systems ergeben sich aus einem rechtswidrigen Training nur dann Probleme, wenn im Output urheberrechtlich geschützte Inhalte aus dem Trainingskorpus wiedererkennbar sind, also Vervielfältigungen oder Bearbeitungen dieser ausgegeben werden. Dies kann vorkommen, wenn die KI Trainingsdaten ganz oder teilweise im Output reproduziert, ein Phänomen, das als »Erinnern« bezeichnet wird. Die Forschung ist sich uneinig, wie häufig dies der Fall sein kann, aber das Risiko bleibt bestehen.

3.3.1.4 Rechte der KI-Anbieter am generierten Output

Die Entwickler der eingesetzten KI erlangen ganz regelmäßig keine urheberrechtlichen Rechtspositionen, wenn Dritte mit der KI Erzeugnisse generieren. Allenfalls und ganz ausnahmsweise, wenn der Programmierer bereits bei der Entwicklung des Programms substanzielle Gestaltungsentscheidungen zu dem konkreten Erzeugnis getroffen hat, kommt ein urheberrechtlicher Schutz zugunsten des Programmierers in Betracht. Das wird bei generativen KI-Systemen regelmäßig nicht der Fall sein.

Es kann sein, dass die KI-Anbieter Rechte am Output durch die Nutzungsbedingungen zu erlangen suchen, oder zumindest den Einsatz der Erzeugnisse einschränken. Die Wirksamkeit dieser Regelungen ist dann im Einzelfall zu prüfen (etwa im Licht des AGB-Rechts) und nur schuldrechtlicher Natur.

3.3.1.5 Urheberrechtliche Haftung auf Verwenderseite

Ein relevantes Beispiel für die immaterialgüterrechtliche Haftung auf Verwenderseite ist der Einsatz von KI-generierten Erzeugnissen, wie Fotos, für Werbemaßnahmen. Dabei ist zunächst festzuhalten, dass der künstlerische Stil, ähnlich wie abstrakte Ideen, nicht unter den Schutz des Urheberrechts fällt. Jedoch können Urheberrechtsverletzungen entstehen, wenn der von der KI generierte Output wesentliche Teile eines urheberrechtlich geschützten Werkes enthält, das für das Training der KI verwendet wurde. In diesem Zusammenhang ist zu beachten, dass die Haftung für solche Vervielfältigungen unabhängig davon besteht, ob das Training der KI mit dem urheberrechtlich geschützten Material rechtmäßig erfolgte.

Ein wesentlicher Maßstab für die Feststellung einer Urheberrechtsverletzung ist die Wiedererkennbarkeit des vorbestehenden, geschützten Werkes im KI-Output im Rahmen einer Gesamtbetrachtung. Hierbei ist entscheidend, dass nur jene Teile des vorbestehenden Werkes berücksichtigt werden, die tatsächlich urheberrechtlich geschützt sind. Eine Einzelfallbetrachtung ist daher unerlässlich.

Das Risiko von Urheberrechtsverletzungen kann durch bestimmte Maßnahmen reduziert werden. Eine Möglichkeit besteht darin, dass Betreiber und Nutzer in ihren Prompts nicht explizit geschützte Werke oder konkrete Künstler nennen. Weiterhin könnte in einigen Fällen die Pastiche-Schranke des § 51a UrhG anwendbar sein, allerdings ist der genaue Inhalt dieses Begriffs und die Reichweite der Schranke derzeit noch nicht abschließend geklärt. Eine eigene Prüfung des KI-generierten Erzeugnisses auf mögliche Drittinhalte ist in jedem Fall zu empfehlen, auch wenn dadurch keine abschließende Rechtssicherheit erlangt werden kann.
Im Falle einer festgestellten Urheberrechtsverletzung stehen dem Rechteinhaber jedenfalls Unterlassungsansprüche zu. Zudem können, bei Nachweis von Verschulden, Schadensersatzansprüche geltend gemacht werden. Dabei ist zu beachten, dass im Urheberrecht bereits eine niedrigschwellige Fahrlässigkeit für die Begründung der Haftung ausreichend sein kann.

3.3.2 Know-how-Schutz und Geschäftsgeheimnisschutz

Bei der Anwendung von KI im Unternehmen sollte der Schutz von Geschäftsgeheimnissen i. S. d. Geschäftsgeheimnisschutzgesetzes (GeschGehG) beachtet werden. Hierbei bestehen sowohl Chancen als auch Risiken: Zum einen kann der Geschäftsgeheimnisschutz den Urheberschutz an KI-Modellen, Trainingsdaten und Prompts sowie an KI-generiertem Output ergänzen bzw. etwaige Schutzlücken schließen. Zum anderen dürfen bestehende Geschäftsgeheimnisse nicht durch einen unbedarften Umgang bei der Eingabe in KI-Anwendungen gefährdet werden.
An Geschäftsgeheimnissen bestehen verkehrsfähige Schutzrechte, an denen der Berechtigte Lizenzen einräumen kann und die (unter Beachtung der Schutzvoraussetzungen) grundsätzlich eine umfassende Verwertung erlauben. Daraus resultiert ein mitunter erheblicher wirtschaftlicher Wert. Davon abzugrenzen ist der uneinheitlich verwendete Begriff des Know-hows. Dieser bezeichnet kein gewerbliches Schutzrecht, sondern beschreibt lediglich den faktischen Vorteil einer geheimen Information, mittels derer ein Unternehmen sich von Wettbewerbern abheben kann.

Der Geschäftsgeheimnisschutz wird in Deutschland durch das auf einer europäischen Richtlinie beruhende Gesetz zum Schutz von Geschäftsgeheimnissen gewährleistet. Eine Information wird nach dessen § 2 zum Geschäftsgeheimnis, wenn sie (i) geheim und deshalb von wirtschaftlichem Wert ist, (ii) sie Gegenstand von angemessenen Geheimhaltungsmaßnahmen ist und (iii) soweit ein berechtigtes Interesse an ihrer Geheimhaltung besteht. Damit hat das GeschGehG in inhaltlicher Hinsicht einen sehr breiten Anwendungsbereich.

Dies hat zur Folge, dass jede Kategorie von Informationen dem Geheimnisschutz zugänglich ist, wenn und soweit sie die vorgenannten Voraussetzungen erfüllt. Hierunter können mit Blick auf KI-Anwendungen beispielsweise trainierte und untrainierte KI-Modelle, Algorithmen, Prompts, Trainingsdaten sowie durch KI-Anwendung generierter und anderweitig nicht schutzfähiger Output von KI-Anwendungen (grafische Darstellungen, Video- und Musiksequenzen, Software und Computerprogramme sowie technische (Produkt-)Gestaltungen wie CAD oder Halbleiterdesigns) fallen. Die inhaltlichen Beschränkungen anderer Rechte des geistigen Eigentums, wie z. B. die fehlende persönliche, geistige Schöpfung i. S. d. Urheberrechts, bestehen hier nicht. Denn anders als das Urheberrecht verlangt das GeschGehG gerade keinen menschlichen Beitrag bei der Entstehung des Geschäftsgeheimnisses, sodass rechtlicher Schutz auch für vollautomatisch generierte Informationen in Betracht kommt. Rechtsinhaber ist dann gem. § 2 Abs. 2 GeschGehG jede natürliche oder juristische Person, welche die rechtmäßige Kontrolle über die entsprechende Information hat.

Geheim ist eine Information, wenn sie weder insgesamt noch in der genauen Zusammensetzung ihrer Bestandteile Dritten allgemein bekannt oder ohne Weiteres zugänglich ist. Es liegt in der Verantwortung des Rechtsinhabers, dies sicherzustellen und mit angemessenen Geheimhaltungsmaßnahmen (wie z. B. dem Abschluss von Verschwiegenheitsvereinbarungen insbesondere mit Kunden, Geschäftspartnern und Angestellten, der Beschränkung des Zugriffs auf die geschützte Information nach dem Need-to-know-Prinzip, IT-Sicherheitsmaßnahmen und -schulungen) dauerhaft aufrechtzuerhalten. Die Umsetzung geeigneter Geheimhaltungsmaßnahmen ist Teil der Organisationspflicht der Geschäftsführung. So kann ein Versäumnis und der daraus resultierende Verlust des Geheimnisschutzes zu einer Haftung des GmbH-Geschäftsführers gem. § 43 Abs. 1 GmbHG führen.

3.3.2.1 Schutz von KI-Modellen nach dem GeschGehG

Der immaterialgüterrechtliche Schutz von KI-Modellen weist in mehrfacher Hinsicht Lücken auf. Das können Geschäftsgeheimnisse im Sinne des GeschGehG unter Umständen abfedern.

a. Schutzlücken von KI-Modellen im Immaterialgüterrecht

KI-Modelle sind nach herrschender Auffassung wohl nur beschränkt dem Schutz als Computerprogramm nach § 69a UrhG zugänglich, nämlich soweit sie das Ergebnis einer eigenen geistigen Schöpfung ihres Urhebers sind. Hierbei ist zu unterscheiden. Während der dem KI-Modell zugrundeliegende Algorithmus bei hinreichender Schöpfungshöhe sowie die Einbettung des KI-Modells in einen Application-Layer als Computerprogramm angesehen werden können, ist das KI-Modell hinsichtlich der im Wege des Training-Prozesses ausgeprägten Gewichtungen der neuronalen Verknüpfungen kein Computerprogramm. Denn diese Gewichtungen sind automatisiert durch den Trainings-Prozess entstanden. Sie sind gewissermaßen durch die KI selbst programmiert und daher dem urheberrechtlichen Computerprogrammschutz nicht zugänglich.

Die Gewichtungen sind, selbst wenn sie isoliert von dem Quellcode eines künstlichen neuronalen Netzes in einer Datei abgespeichert werden, auch nicht als Datenbank i. S. v. § 87a I UrhG anzuse-

hen. Denn das Datenbankherstellerrecht erfordert die derartige Unabhängigkeit der einzelnen Elemente einer Datenbank, dass sich diese voneinander trennen lassen, ohne den Wert ihres informativen oder sonstigen Inhalts zu verlieren. In diesem Sinne sind die einzelnen Gewichtungen nicht unabhängig. Denn anders als z. B. eine einzelne Telefonnummer aus einem Telefonbuch ist eine einzelne Gewichtung für sich genommen nutzlos und nur in der Gesamtheit mit allen anderen Gewichtungen in dem künstlichen neuronalen Netz nutzbar.
Auch ein Patentschutz des KI-Modells scheidet grundsätzlich aus, da abstrakte Algorithmen sowie ganze Programme für Datenverarbeitungsanlagen mangels Technizität gem. § 1 Abs. 3 Patentgesetz keine schutzfähigen Erfindungen sind. Nur in Ausnahmefällen ist ein Schutz als computerimplementierte Erfindung denkbar, wenn und soweit das künstliche neuronale Netz einen spezifischen technischen Beitrag leistet. Das ist etwa der Fall, wenn das KI-Modell die Steuerung eines technischen Geräts oder die Optimierung eines industriellen Prozesses übernimmt.

b. Ergänzender Schutz von KI-Modellen nach dem GeschGehG

Da die Gewichtungen der neuronalen Verknüpfungen für die Arbeitsweise des KI-Modells entscheidend sind, stellen sie oftmals den wirtschaftlichen Wert des Modells dar. Umso wichtiger ist es, ihren Schutz (zumindest) als Geschäftsgeheimnis sicherzustellen. Dafür muss das künstliche neuronale Netz inkl. der Gewichtungen Gegenstand angemessener Geheimhaltungsmaßnahmen i. S. d. § 2 GeschGehG sein (siehe oben).

So sollte dem Schutz von Geschäftsgeheimnissen bei der Ausgestaltung von Nutzungs- oder Lizenzverträgen für KI-Systeme besondere Beachtung geschenkt werden. Es sind bereits Fälle bekannt geworden, in denen durch entsprechendes Prompt-Engineering die Gewichtungen eines KI-Modells rekonstruiert werden konnten. Um die derartige Erlangung der Gewichtungen zu unterbinden, finden sich in den Nutzungsbedingungen von KI-Anbietern oftmals Klauseln, die das gezielte Reverse Engineering des KI-Modells i. S. d. § 3 Abs. 1 Nr. 2 GeschGehG sowie sonstige Handlungen, die auf dessen Rekonstruktion abzielen, untersagen. Hierdurch soll die Extraktion der Gewichtungen zum Zweck des Nachbaus des KI-Modells unterbunden werden. Wenn und soweit Gewichtungen Vertrags- oder Entwicklungspartner offengelegt werden, sollten sie durch angemessene Geheimhaltungsmaßnahmen und insbesondere Geheimhaltungsverpflichtungen geschützt werden, um einen effektiven Geheimnisschutz zu gewährleisten. Darüber hinaus sollten klare Regelungen darüber getroffen werden, wer Inhaber von Rechten an Verbesserungen und Weiterentwicklungen des künstlichen neuronalen Netzes sein soll. Das gilt insbesondere im Kontext von F&E-Kooperationen.

Bei dem Betrieb eines KI-Modells auf den Servern eines Cloud-Anbieters (zum Beispiel im Kontext von AIaaS oder auch in einer Private-Cloud-Umgebung) sollten auch gegenüber dem Cloud-Anbieter Geheimhaltungsmaßnahmen ergriffen werden. Dies kann auf technischer Ebene und/oder durch den Abschluss von Geheimhaltungsvereinbarungen geschehen.

3.3.2.2 Schutz von KI-generiertem Output nach GeschGehG

Ob und inwieweit KI-generierter Output dem Geschäftsgeheimnisschutz zugänglich ist, hängt von der Art und Weise des Einsatzes der KI ab. Entscheidend für die Bewertung als »geheim« ist, ob ein Dritter auf den Output zugreift oder zugreifen kann. Nutzt der Anwender eine selbst entwickelte (proprietäre) KI – sind also KI-Anwender- und Anbieter personenidentisch – oder nutzt der Anwender zwar die KI eines Drittanbieters, aber betreibt diese lokal auf seiner Infrastruktur (on-premise, s. oben), dann kann der Output geheim sein. Entscheidend hierfür ist dann lediglich das Informationsmanagement im Unternehmen des KI-Anwenders. Dass der Inhalt ursprünglich durch eine KI erzeugt wurde, ist für dessen Schutzfähigkeit nach dem GeschGehG unerheblich.

Nutzt der Anwender die KI dagegen als Cloud-Produkt des Drittanbieters (AI as a Service – AIaaS, s. oben), ist jedenfalls faktisch eine Zugangsmöglichkeit des Anbieters eröffnet. In diesem Fall kommt es auf die Nutzungsbedingungen des Anbieters an: Verpflichtet sich dieser zu Vertraulichkeit oder gestatten sie ihm den Zugriff auf den Output etwa zu Zwecken der Produktentwicklung und Qualitätskontrolle? Nur im ersten Fall besteht Geschäftsgeheimnisschutz am Output. Im sogenannten freemium-Vertrieb machen KI-Anbieter die Geheimhaltung teilweise schlicht davon abhängig, ob der Anwender eine kostenfreie oder kostenpflichtige Version der KI verwendet. Aus Sicht des Anwenders empfiehlt sich hier unbedingt die genaue Prüfung der Nutzungsbedingungen. Sichert der Anbieter die Vertraulichkeit des Outputs zu, sodass dieser nur »für die Augen des Nutzers« bestimmt ist, kann der generierte Output als Geschäftsgeheimnis des Anwenders betrachtet werden. Sofern die Geheimhaltungskette über den gesamten Produktvertrieb aufrechterhalten werden kann, können auf Grundlage des Geschäftsgeheimnisschutzes auch tragfähige Lizenzmodelle entwickelt werden.

3.3.2.3 Prompts als Geschäftsgeheimnisse

Auch Prompts können Geschäftsgeheimnisse darstellen oder enthalten. Da die Entwicklung und Gestaltung von Prompts mitunter ganz erheblichen zeitlichen und personellen Aufwand erfordern können, erscheint die Schutzfähigkeit auch interessengerecht. Aus der Praxis sind zudem Fälle bekannt, dass Prompts exklusiv für Kunden entwickelt werden und daher auch ein schutzwürdiges Wirtschaftsgut darstellen.

Wie schon bei den Ausführungen zum Schutz von KI-generiertem Output nach dem GeschGehG ist auch hier die Art und Weise des KI-Betriebs (proprietär, on-premise oder AIaaS, s. dazu genauer oben) entscheidend. Daher sind unbedingt die einschlägigen Nutzungsbedingungen genau in den Blick zu nehmen. Behält sich der Anbieter darin das Recht vor, seine KI mit den Prompts und Input-Daten der Anwender (weiter) zu trainieren, resultiert daraus ein erhebliches Risiko für die enthaltenen Geschäftsgeheimnisse. Teilweise finden sich in den Nutzungsbedingungen von KI-Anbietern sogar Regelungen, nach denen sich diese die einfachen, weltweiten Nutzungsrechte an Prompts einräumen lassen, um diese auch über die Trainingszwecke hinaus weiterzuverwerten.

Das Risiko des Geheimnisverlusts gilt aber nicht nur im Verhältnis zum KI-Anbieter. Darüber hinaus kann nicht ausgeschlossen werden, dass die KI die zum Training verwendeten Informationen oder Teile davon gegenüber Dritten als Output wieder offenlegt. Hierdurch riskiert der Anwender den Verlust bestehender Geschäftsgeheimnisse, wenn diese zunächst als (Teil eines) Prompt in eine cloudbasierte KI (AIaaS) eingegeben und damit z. B. zu Trainingszwecken in das KI-Modell aufgenommen werden. So wurde z. B. der Fall bekannt, in dem die Softwareentwickler eines IT-Unternehmens Quellcode zur Kontrolle von programmiertechnischen Fehlern in eine KI-Anwendung eingespeist haben. Hierdurch wurde das Risiko begründet, dass dieser Quellcode – je nach konkreter Gestaltung der KI-Anwendung – bei entsprechender Ansteuerung des KI-Modells von diesem auch gegenüber unbefugten Dritten reproduziert wird. Technische Maßnahmen, wie Confidential Computing, können hier in Betracht gezogen werden, um dieses Risiko gegenüber den involvierten Anbietern zu minimieren.

3.3.2.4 Geheimnisschutz und KI im Unternehmen

Dem Geheimnisschutz kommt im Hinblick auf den sonst lückenhaften Schutz im Kontext der Nutzung generativer KI-Systeme eine besondere Bedeutung zu, die von der Geheimhaltung der betreffenden Informationen abhängt. Dem sollte zum einen durch ein internes Informationsmanagement und zum anderen im Umgang mit Vertragspartnern Rechnung getragen werden. Spiegelbildlich zum Schutz von KI-Systemen selbst ist auch der Schutz eigener Geschäftsgeheimnisse vor dem Zugriff durch KI-Systeme zu behandeln. Um ein hinreichendes Schutzniveau erzielen zu können, sollten Unternehmen präventive interne Maßnahmen ergreifen. Zunächst ist es notwendig, einen Überblick über bestehende Geschäftsgeheimnisse und deren interne Handhabung zu gewinnen. Auf dieser Basis kann sodann eine Risikobewertung vorgenommen werden, anhand derer sich die Geheimnisschutzmaßnahmen orientieren und gezielt angepasst werden können. So sollten KI-Anwendungen nach Möglichkeit lokal (on-premise) oder in einer gesicherten Unternehmensumgebung betrieben werden, um eine größtmögliche Kontrolle über die Daten zu gewährleisten.

Zudem ist die Sensibilisierung der Mitarbeiter durch klare Anweisungen und Schulungen unerlässlich. Mitarbeiter müssen wissen, welche Daten für die Verarbeitung mittels KI-Anwendungen geeignet sind und welche nicht. Art. 4 der neuen EU-KI-Verordnung fordert zudem ausdrücklich, dass Mitarbeiter in KI-Kompetenzen geschult werden. Eine ausreichende KI-Kompetenz soll dabei »nach besten Kräften« sichergestellt werden. Der konkrete Umfang der Schulungen sollte dabei an die technischen Kenntnisse und den spezifischen Einsatzkontext der Mitarbeiter angepasst werden. Dies kann z. B. durch interne Schulungen oder durch externe E-Learning-Angebote umgesetzt werden. Die Dokumentation dieser Maßnahmen ist sowohl für die interne Kontrolle als auch aus Beweisgründen im Streitfall wichtig.

Besonders wertvolle proprietäre Datensätze sollten durch technische Schutzmaßnahmen vor automatisiertem Zugriff geschützt werden und nicht in Gänze ungesichert KI-Systemen zur Analyse überlassen werden. Vor der Offenlegung geheimer Informationen sollten Unternehmen klare Regelungen zur Geheimhaltung und zur Nutzung für Trainingszwecke festlegen. Neben vertraglichen Regelungen (u. a. mit dem Anbieter der KI) sollten auch technisch-organisatorische

Maßnahmen ergriffen werden. Diese können z. B. in Form von kryptografischen Maßnahmen direkt an den betroffenen Informationen oder durch technische Zugriffsbeschränkungen auch allgemein umgesetzt werden.

Im Fall des Vertriebs eigener KI-Systeme sollten sich vertragliche Regelungen zudem an dem Ziel orientieren, eine Offenbarung oder rechtmäßige Erlangung der in der KI-Anwendung oder ihren Datensätzen enthaltenen Geschäftsgeheimnisse zu vermeiden. Dafür sollten Anbieter darauf achten, den Zugriff auf ihre KI-Anwendungen möglichst zu limitieren (z. B. als Software as a Service ohne unmittelbaren Zugriff auf die zugrundeliegenden Modelle nebst Gewichtungen) und für den Umgang mit ihr strikte vertragliche Nutzungsvorgaben zu machen.

3.3.3 Weitere Schutzrechte

3.3.3.1 Marken

Aus der Nutzerperspektive sind Marken für die Eingabe (Prompt) in eine KI-Anwendung und den Output relevant. Für das Training einer KI-Anwendung mit Daten, die Marken enthalten, gelten die Ausführungen zu Ziffer 3.3.1. Im Trainingsprozess wird die digitale Repräsentation der Marke, aber nicht die Marke in ihrer Markenfunktion genutzt.

Die Markenverwendung für einen Prompt ist zunächst eine maschinenlesbare Bestimmung der Marke. Eine Markennutzung liegt vor, wenn durch die Bestimmung der Marke ein Output geschaffen wird, der diese Marke wiedergibt. Allerdings: anders als z. B. bei der Buchung eines Keywords führt die Wiedergabe der Marke im KI-Output noch nicht zu einer Nutzung im geschäftlichen Verkehr. Im Fall der Keywords wird die Anzeige einer elektronischen Werbung bewirkt (Google Ad). Der KI-Output bedarf erst einer weiteren Verwendung, anhand der sich die Rechtmäßigkeit der Markennutzung beurteilt.

Zu berücksichtigen ist, dass die KI auch ohne Angabe einer Marke im Prompt eine Marke in einem Output wiedergeben kann. Dies ist insbesondere bei bekannten Marken der Fall, die z. B. nahezu repräsentativ für eine bestimmte Warengattung sind. Viele dieser Marken sind nicht nur unter einem Bildzeichen oder Wort eingetragen, sondern sie können auch z. B. in einer Produktform geschützt sein. Zudem gibt es Farben oder auch Muster, die dem Markenschutz unterstehen. Der Katalog der Markenformen ist nicht abgeschlossen. Häufig sind diese »besonderen« Marken auch bekannte Marken, die insbesondere gegen Rufausbeutung geschützt sind.

Vor der Nutzung eines KI-Outputs muss daher geklärt werden, ob sich darin fremde Marken finden. KI-Output, der erkennbar eine fremde geschützte Marke wiedergibt, kann im geschäftlichen Verkehr nicht ohne Zustimmung des Inhabers für die unter der Marke geschützten Waren und Dienste verwendet werden. Sollte es sich um eine bekannte Marke handeln, ist diese vor rufausbeutenden oder rufschädigenden Handlungen geschützt. Zudem kann sich der Inhaber einer bekannten Marke gegen Verwässerung wenden.

KI-Output, der eine humorvolle, kritische oder kreative Auseinandersetzung mit der Marke vermittelt, kann nach Abwägung im Einzelfall eine zulässige Form der Meinungsäußerung oder der künstlerischen Freiheit sein, sodass die Rechte des Markeninhabers zurücktreten. Die Verwendung von Outputs, die die Marke trivialisieren oder schlicht entwerten, bleibt rechtswidrig.
Der Output, den KI-Systeme generieren, kann insbesondere als Wort- oder Bildmarke markenrechtlich geschützt werden. Der Schutz einer Marke entsteht in der Regel durch deren Eintragung in das zuständige Markenregister, wobei es auch den Sonderfall des Schutzes durch Verkehrsgeltung gibt. Bei der Überlegung, KI-generierte Inhalte als Marke anzumelden, sollte im Vorfeld gründlich geprüft werden, ob bereits ältere Markenrechte bestehen, um mögliche Konflikte mit Rechteinhabern zu vermeiden.

3.3.3.2 Designs

KI-Output kann, ob ausdrücklich in einem Prompt vorgegeben oder nicht, ein geschütztes Design wiedergeben. Ein Design wird dann unerlaubt nachgeahmt, wenn es bei dem informierten Benutzer keinen anderen Gesamteindruck erweckt als das geschützte Design.

Eine Abbildung eines geschützten Designs kann auch eine unerlaubte Nachahmung sein, wenn sie die wesentlichen Merkmale des geschützten Designs übernimmt oder nur unwesentliche Abweichungen aufweist. Dies gilt auch für eigenständige Gestaltungen, sofern diese an den charakteristischen Elementen des Designs orientiert wurden.

KI-Output, der ein geschütztes Design wiedergibt, kann nicht ohne Zustimmung des Inhabers verwendet werden. Allerdings kann eine Wiedergabe zulässig sein, sollte die Verwendung des KI-Outputs in einem bestimmten Kontext erfolgen, wie z. B. für Lehrzwecke oder als Bildzitat.

3.3.3.3 Persönlichkeitsrechte

KI-Output kann, ob ausdrücklich über einen Prompt eingegeben oder aufgrund der Datensätze des KI-Tools generiert, Personen oder auch Merkmale von Personen, z. B. die Stimme, erkennbar machen.

Die Verbreitung des KI-Outputs, der eine Person identifizierend beinhaltet, kann das Recht auf informationelle Selbstbestimmung, das die Privatsphäre und die Ehre einer Person schützt, verletzen, sofern die Verbreitung gegen ihren Willen erfolgt. Das Allgemeine Persönlichkeitsrecht schützt z. B. auch die stimmliche Identität einer Person. Darüber können Rechte am eigenen Bild betroffen sein, wenn es an der Einwilligung der Person fehlt.

Sollte eine Person über ihre Darbietungen Eingang in einen Prompt oder KI-Output finden, gelten die Schutzinstrumente des Urheber- und Leistungsschutzrechts (s. o. Ziffer 3.3.1) sowie des Datenschutzrechts (s. o. Ziffer 3.1.1.) und anderer die Persönlichkeitsrechte betreffender Gesetze.
Im Einzelfall kann, unter Abwägung widerstreitender grundgesetzlicher sonstiger Interessen, die Verwendung des KI-Outputs gestattet sein.

3.4 Haftungsrechtliche Aspekte

3.4.1 Typische Risiken beim KI-Einsatz

Der Einsatz generativer KI in Unternehmen birgt eine Reihe von Risiken, die sowohl operationelle als auch rechtliche Implikationen haben können. Zu den häufigsten Herausforderungen zählen Fehler und Mängel in den Trainingsdaten, die zu unzuverlässigen oder irreführenden Ergebnissen in der Einsatzphase führen können, wie Vorurteile, Ungenauigkeiten, ungenaue, unethische und schädliche Inhalte wie unbeabsichtigte Diskriminierung, die durch verzerrte Daten oder Algorithmen entstehen kann. Darüber hinaus können KI-Systeme fehlerhafte Schlussfolgerungen ziehen oder aufgrund ungeeigneter Algorithmen und Entscheidungslogiken in spezifischen Anwendungskontexten versagen.

In Produkten und Dienstleistungen eingebettete bzw. verwendete fehlerhafte Ergebnisse können zu Unzufriedenheit der Kundinnen und Kunden, Vertragsbruch, (unverhältnismäßiger) Haftung und Vertragsstrafen, Geschäfts- und Umsatzverlusten, Reputationsschäden, Bußgeldern, Problemen in der Lieferkette usw. führen.

Sicherheitslücken und Manipulationsanfälligkeit stellen weitere kritische Risiken dar, ebenso wie potenzielle Hardwarefehler, die die Funktionalität der KI beeinträchtigen können. Zudem müssen Unternehmen die möglichen Rechte Dritter am Output der KI berücksichtigen, um Urheberrechtsverletzungen oder andere rechtliche Konflikte zu vermeiden.
Zuletzt bestehen Compliance-relevante Risiken durch Nichteinhaltung der gesetzlichen und/oder vertraglichen Verpflichtungen, Verstoß gegen Anforderungen an Dokumentation, Nachvollziehbarkeit, Transparenz und Rechenschaftspflicht und Risiken des Third-Party-Managements (bei Nutzung von fremder KI).

3.4.2 Haftung des Anwenders nach derzeitigen Haftungsregelungen

3.4.2.1 Vertragliche Haftung

Gegenüber dem KI-Hersteller

Bei der vertraglichen Haftung gegenüber dem Hersteller von KI-Systemen spielen mögliche Pflichtverletzungen eine zentrale Rolle. Unternehmen, die KI-Lösungen einsetzen, müssen die Vertragsbedingungen genau prüfen, insbesondere im Hinblick auf Garantien, Leistungszusicherungen und Verantwortlichkeiten. Eine Pflichtverletzung kann vorliegen, wenn die KI nicht wie vereinbart funktioniert, beispielsweise aufgrund von Fehlern in der Software oder unzureichender Leistungsfähigkeit.

In vielen Fällen versuchen Hersteller, ihre Haftung durch vertragliche Begrenzungen zu minimieren. Dies kann Haftungsausschlüsse für bestimmte Arten von Schäden oder Haftungsobergrenzen umfassen. Unternehmen sollten diese Bedingungen sorgfältig bewerten und bei Bedarf Verhandlungen führen, um einen angemessenen Schutz ihrer Interessen zu gewährleisten.

Gegenüber den Endnutzern bzw. Kundinnen und Kunden (z. B. bei Angebot eines Chatbots, den Kundinnen und Kunden »bedienen«)

Im Verhältnis zum Endnutzer oder Kunden, beispielsweise beim Einsatz eines Chatbots, können ebenfalls vertragliche Haftungsfragen aufkommen. Mögliche Pflichtverletzungen in diesem Kontext umfassen die Nichterfüllung von Leistungszusagen, wie die Genauigkeit oder Zuverlässigkeit der durch den Chatbot generierten Antworten.

Haftungsbegrenzungen gegenüber Kundinnen und Kunden gestalten sich oft komplex, insbesondere angesichts der strengen Vorschriften des Verbraucherschutzrechts. Unternehmen, die eine Begrenzung ihrer Haftung anstreben, müssen gewährleisten, dass ihre Haftungsbeschränkungen nicht nur rechtlich durchsetzbar sind, sondern auch den geltenden rechtlichen Anforderungen entsprechen. Dies schließt eine sorgfältige Überprüfung und Gestaltung der Allgemeinen Geschäftsbedingungen (AGB) ein, um sicherzustellen, dass diese klar, verständlich und nicht unangemessen benachteiligend für den Verbraucher sind.

Ein konkretes Beispiel für eine Haftungssituation könnte die Richtigkeit der generierten Inhalte sein. Falsche oder irreführende Informationen, die durch einen Chatbot bereitgestellt werden, könnten zu Schäden oder Verlusten für den Nutzer führen. Zudem können Rechtsverletzungen entstehen, wenn der Chatbot geschützte Inhalte Dritter ohne entsprechende Lizenz verwendet. In solchen Fällen könnten Unternehmen sowohl gegenüber ihren Kundinnen und Kunden als auch gegenüber den Rechteinhabern haftbar gemacht werden.

3.4.2.2 Deliktsrechtliche Haftung

Verschuldensabhängige Haftung

Wesentliche Anspruchsgrundlage der deliktsrechtlichen verschuldensabhängigen Haftung ist § 823 Abs. 1 BGB. Hiernach ist der zurechenbare Verursacher nur verantwortlich, wenn er sich pflichtwidrig verhalten hat, was sich insbesondere aus der Verletzung von Verkehrssicherungspflichten ergibt. In Bezug auf KI-Systeme können nicht nur die Hersteller oder Entwickler, also die sog. Produzenten, sondern auch der Anbieter oder Anwender als zurechenbarer Verursacher in Betracht kommen. In der Praxis bedeutet dies, dass je nach Rolle und Einfluss auf das Produkt oder die Dienstleistung unterschiedliche Verkehrssicherungspflichten gelten können. Diese Pflichten zielen darauf ab, Schäden für Personen oder Eigentum zu vermeiden und die Sicherheit zu gewährleisten.

Für den **Hersteller** sind die Verkehrssicherungspflichten besonders umfassend. Sie umfassen Konstruktions-, Fabrikations-, Instruktions-, Produktbeobachtungs- und Gefahrabwendungspflichten. Der Hersteller ist verantwortlich dafür, dass das Produkt sicher entworfen, hergestellt und mit den notwendigen Informationen und Warnungen versehen wird.

Anbieter eines Produktes oder einer Dienstleistung, die beispielsweise KI-Systeme nutzen oder integrieren, können ebenfalls Verkehrssicherungspflichten unterliegen. Diese umfassen üblicherweise die Pflicht, sicherzustellen, dass die angebotenen Produkte oder Dienstleistungen keine Gefahr für die Nutzer oder Dritte darstellen. Auch müssen Anbieter über eventuelle Risiken informieren und bei bekannt gewordenen Gefahren reagieren.

Anwender oder **Nutzer** eines Produktes, besonders im Bereich der KI, können ebenfalls bestimmte Pflichten haben, insbesondere wenn sie die KI-Systeme in einem Kontext nutzen, der Risiken für andere mit sich bringt. Hier können Pflichten wie die sachgemäße Anwendung gemäß den Anweisungen des Herstellers oder Anbieters und die Berücksichtigung von Sicherheitshinweisen relevant sein. Bei Anwendern, die das Produkt in einem professionellen oder kommerziellen Kontext nutzen, können die Verkehrssicherungspflichten umfangreicher sein, insbesondere wenn sie das Produkt modifizieren oder in einer Weise verwenden, die über den ursprünglichen Anwendungsbereich hinausgeht.

Eine Haftung gemäß § 823 Abs. 1 BGB kommt nur dann in Betracht, wenn dem Anspruchsgegner ein schuldhafter Verstoß gegen die Verkehrssicherungspflichten nachgewiesen werden kann. Der relevante Sorgfaltsmaßstab wird durch die spezifischen Umstände des Einzelfalls bestimmt. Grundsätzlich hat der Geschädigte sämtliche anspruchsbegründende Tatsachen, also objektiven Tatbestand, Verschulden, Schaden und Kausalität, zu beweisen. Diese Beweislastverteilung gilt gegenüber dem Anbieter und Anwender der KI.

Sonderfall: Hersteller und Entwickler (Produzenten)

Für verschuldensabhängige Ansprüche aus § 823 I BGB gegenüber Herstellern und Entwicklern aus § 823 I BGB, die sog. Produzentenhaftung, sieht die Rechtsprechung eine Beweislastumkehr vor. Da der Geschädigte selbst wenig Einsicht in den Betrieb des Schädigers hat und er daher schwierig seinen Anspruch darlegen und beweisen kann, muss der Schädiger hier beweisen, dass ihn kein Verschulden trifft, er also keine Verkehrspflicht verletzt hat. Der Geschädigte muss lediglich nachweisen, dass sein Schaden durch einen Produktfehler verursacht wurde, der im Organisations- und Gefahrenbereich des Herstellers entstanden ist.

Neben den bereits dargestellten Verkehrspflichten wie Konstruktions-, Fabrikations- und Instruktionspflichten obliegt den Herstellern und Entwicklern als Produzenten insbesondere eine fortlaufende Produktbeobachtungspflicht. Diese Pflicht basiert auf der Annahme, dass der Produzent aufgrund seiner Nähe zum Produkt und seiner technischen Expertise am besten geeignet ist, relevante Informationen über sicherheitsrelevante Eigenschaften seiner Produkte zu sammeln und zu analysieren.

Die aus der Produktbeobachtungspflicht resultierenden Reaktionspflichten können variieren, sind aber häufig durch Warnpflichten gekennzeichnet. Der Umfang und Inhalt dieser Reaktionspflichten hängen maßgeblich von der Größe der potenziellen Gefahr und der Effektivität möglicher Gegenmaßnahmen ab. Die genaue Ausgestaltung der Produktbeobachtungspflicht ist jedoch nicht immer eindeutig und Gegenstand fortlaufender juristischer Diskussionen.

Verschuldensunabhängige Haftung

Produkthaftung nach dem Produkthaftungsgesetz (ProdHaftG) eine besondere Form der verschuldensunabhängigen Gefährdungshaftung. Im Gegensatz zur verschuldensabhängigen Haftung ist ein vorwerfbarer Verstoß gegen Sorgfaltspflichten keine Voraussetzung für die Haftung nach dem ProdHaftG.

Haftungsgrundlage

Die Produkthaftung kommt in Betracht, wenn die folgenden Voraussetzungen erfüllt sind: Verletzung eines geschützten Rechtsguts (Tötung einer Person, Verletzung von Körper oder Gesundheit, Beschädigung einer Sache)

- durch ein fehlerhaftes Produkt
- mit daraus resultierendem (finanziellen) Schaden
- sowie kein Vorliegen einer gesetzlichen Ausnahme aus § 1 Abs. 2, Abs. 3 ProdHaftG.

Schäden am fehlerhaften Produkt selbst oder reine Vermögensschäden, die nicht unmittelbare Folge der Rechtsgutsverletzung sind, fallen nicht unter die Produkthaftung. Potenzielle Haftungsgründe können Konstruktions-, Fabrikations- oder Instruktionsfehler sein.

Anwendung auf KI-Systeme

Nach dem Produkthaftungsgesetz (ProdHaftG) wird unter einem Produkt traditionell eine »bewegliche«, also eine körperliche Sache verstanden. In der rechtlichen Diskussion ist allerdings die Einordnung von Software als Produkt im Sinne des § 2 ProdHaftG, insbesondere wenn sie als eigenständige (»standalone«) Software oder getrennt von der Hardware vertrieben wird, umstritten. Die herrschende Meinung tendiert jedoch dazu, auch Software als »Produkt« anzusehen. Weniger Schwierigkeiten bereitet die Beurteilung bei Hardware-Software-Systemen. In diesem Fall ist es gleichgültig, ob ein Fehler der Hardware vorliegt, oder ein Fehler mit dieser verbundenen Steuerungssoftware (»embedded software«). Das System wird als ein Gesamtprodukt betrachtet.

Fehlerhaftigkeit eines KI-Systems

Nach § 3 ProdHaftG ist ein Produkt fehlerhaft, wenn es nicht die Sicherheit bietet, die berechtigterweise zu erwarten ist. Das einzuhaltende Sicherheitsniveau richtet sich nach dem Stand von Wissenschaft und Technik zu dem Zeitpunkt des Inverkehrbringens. Haftungsrelevante Fehler in den Ausgabeergebnissen von KI-Systemen können sich nicht nur aus einer unzureichenden

Programmierung der verwendeten Algorithmen ergeben, sondern auch aus unvollständigen oder für den angestrebten Lernzweck ungeeigneten Trainingsdaten. Aufgrund der Komplexität der Entscheidungsprozesse innerhalb eines KI-Systems ist es in der Regel schwierig, die Fehlerhaftigkeit nachzuweisen.

Beschränkung der Herstellerhaftung

Die Haftung des Herstellers ist auf Fehler beschränkt, die zum Zeitpunkt des Inverkehrbringens des Produkts vorlagen. Sollte ein Produkt nach seiner Markteinführung weiterentwickelt werden, so obliegt dem Hersteller keine Haftung für Mängel, die erst im Rahmen dieser Weiterentwicklungen entstehen. Dies gilt ebenso für Änderungen in den Sicherheitsstandards: Der Hersteller ist verpflichtet, die zum Zeitpunkt des Inverkehrbringens eines neuen Produkts geltenden Sicherheitsstandards einzuhalten, jedoch nicht für Produkte, die bereits auf dem Markt sind.
Im Kontext von KI-Systemen ist der Hersteller nach dem Produkthaftungsrecht nicht haftbar für Fehlfunktionen, die durch nachträgliche Veränderungen im Einsatz des Systems entstehen. Solche Veränderungen können beispielsweise eine Umprogrammierung durch den Nutzer, den Einsatz des Systems außerhalb des vorgesehenen Anwendungsbereichs oder die Verwendung ungeeigneter Datenbestände umfassen. Andererseits bleibt der Hersteller haftbar, wenn ein Fehler auf eine unzureichende Programmierung der Lernalgorithmen zurückgeführt werden kann und sich dieser Fehler erst nach dem Inverkehrbringen des Produkts manifestiert.
Ferner kann ein Fehlgebrauch eines KI-Systems oder eine Falscheingabe von Informationen durch den Anspruchsteller zu einer Anspruchsreduzierung wegen Mitverschuldens führen.

Beweislast

In Fällen der Produkthaftung nach dem ProdHaftG liegt die Beweislast für das Vorliegen eines Produktfehlers, den entstandenen Schaden und die Kausalität zwischen dem Fehler und dem Schaden beim geschädigten Anspruchsteller. Dies bedeutet konkret, dass der Geschädigte nachweisen muss, dass der Schaden durch einen Mangel des Produkts verursacht wurde.
Wichtig zu beachten ist, dass für die Haftung im Rahmen des ProdHaftG kein Verschulden des Herstellers erforderlich ist. Das heißt, ein vorwerfbarer Verstoß gegen eine Sorgfaltspflicht muss nicht vorliegen, um eine Haftung zu begründen. Es findet auch keine Beweislastumkehr statt, wie sie die Rechtsprechung bei der Produzentenhaftung vorsieht.

Allerdings hat der Hersteller hier ebenfalls die Möglichkeit, sich von einer Produkthaftung zu entlasten. Dies kann erfolgen, indem er belegt, dass kein haftungsrelevanter Fehler innerhalb seiner Verantwortungssphäre begangen wurde. Zudem ist nach § 1 Abs. 2 Ziff. 5 ProdHaftG eine Haftung ausgeschlossen, wenn der Fehler zum Zeitpunkt des Inverkehrbringens des Produkts nach dem damaligen Stand von Wissenschaft und Technik nicht erkennbar war (sog. Einwand des Entwicklungsfehlers).

3.4.3 Maßnahmen gegen Haftungsrisiken

Um Haftungsrisiken im Zusammenhang mit dem Einsatz von KI-Systemen zu minimieren, können verschiedene Maßnahmen ergriffen werden.

In einem ersten Schritt sollten die relevanten Haftungsrisiken, die sich aus dem geplanten Einsatz von KI für interne Zwecke und/oder Dienstleistungen für Kunden ergeben, detailliert evaluiert werden. Hierbei sollten mögliche direkte und indirekte (Vermögens-)Schäden beim Kunden oder Dritten durch die Nutzung der von der KI generierten Daten nicht außer Acht gelassen werden. Eine Absicherung ist durch vertragliche Gewährleistungs-, Haftungs- und Freistellungsregelungen möglich und empfehlenswert.

Eine weitere Möglichkeit besteht in der gesellschaftsrechtlichen Auslagerung des KI-Systems in eine Betreibergesellschaft mit haftungsbeschränkter Rechtsform, z. B. eine GmbH. Darüber hinaus empfiehlt sich der Abschluss entsprechender Versicherungen.

Eine weitere präventive Maßnahme ist die vertragliche Haftungsbeschränkung im gesetzlich zulässigen Umfang – auch im Arbeitsverhältnis. Dies schafft einen klaren Haftungsrahmen. Die Konkretisierung der Leistungen und des Leistungsumfangs in Verträgen trägt ebenfalls zur Risikominimierung bei.

3.4.4 Geplante Haftungsregelungen für KI-Anwender

Die oben dargelegten Ausführungen verdeutlichen, dass auf nationaler Ebene bereits umfassende gesetzliche Haftungsregelungen für den Einsatz von KI existieren. Parallel dazu werden auf europäischer Ebene zusätzliche Regelungen entwickelt, um den Umgang mit KI weiter zu normieren und zu präzisieren. Im Fokus der aktuellen Diskussionen stehen dabei nicht nur der AI Act, sondern auch die geplante KI-Haftungsrichtlinie und die neue Produkthaftungsrichtlinie (EU) 2024/2853, die im Dezember 2024 verabschiedet wurde.

3.4.4.1 NLF & AI Act: »Produktsicherheitsrecht«

Das New Legislative Framework (NLF) stellt einen zentralen Pfeiler im Bereich des europäischen Produktsicherheitsrechts dar und spielt auch im Kontext des AI Acts eine entscheidende Rolle. Es definiert grundlegende Anforderungen an KI-Systeme, die Aspekte wie Risikomanagement, Transparenz, Robustheit, IT-Sicherheit und die menschliche Überwachung der Künstlichen Intelligenz umfassen. Entsprechend dem risikobasierten Ansatz des NLF werden KI-Systeme in vier Kategorien eingeteilt, angefangen bei risikofreien bis hin zu Hochrisiko-KI-Systemen, wobei für jede Kategorie spezifische Regulierungsmaßnahmen vorgesehen sind.

Innerhalb des AI Acts (siehe dazu im Detail weiter oben) werden verschiedene Rollen im Umgang mit KI-Systemen definiert. Dazu gehören die Provider, also die Entwickler oder Hersteller von

KI-Systemen, die für die Einhaltung der gesetzlichen Vorgaben verantwortlich sind, sowie die User, also die Anwender von KI-Systemen, die je nach Risikokategorie bestimmten Auflagen folgen müssen.

Neue Pflichten, die mit dem AI Act einhergehen, sind insbesondere für Hersteller von Hochrisiko-KI-Systemen relevant. Sie müssen nun unter anderem die Konformität mit harmonisierten europäischen Normen sicherstellen und eine CE-Kennzeichnung anbringen, die die Übereinstimmung mit den relevanten rechtlichen Anforderungen signalisiert. Außerdem müssen sämtliche Sicherheitskomponenten, die in Produkte integriert werden, die unter das NLF fallen, den Anforderungen des AI Acts entsprechen.

Die Haftungsregelungen im AI Act zielen darauf ab, die Verantwortlichkeit für Schäden, die durch KI-Systeme verursacht werden, klar zu regeln. Dies ist besonders bei Hochrisiko-KI-Systemen von Bedeutung, wo strenge Anforderungen sicherstellen sollen, dass diese Systeme keine Schäden verursachen, die die Sicherheit von Verbrauchern oder die öffentliche Ordnung gefährden.
Der risikobasierte Ansatz des AI Acts ist für Anwender generativer KI von besonderer Bedeutung, da er gewährleistet, dass die von diesen Systemen ausgehenden Risiken adäquat eingeschätzt und reguliert werden. Anwender müssen die Risikokategorie ihres KI-Systems verstehen und entsprechend den gesetzlichen Anforderungen handeln, was insbesondere für Hochrisiko-Systeme bedeutende Implikationen hat.

3.4.4.2 Die geplante KI-Haftungsrichtlinie

Die ursprünglich geplante KI-Haftungsrichtlinie wird nach Mitteilung der EU-Kommission von Mitte Februar 2025 nicht weiterverfolgt. Die folgenden Ausführungen erfolgen daher lediglich der Vollständigkeit halber.

Die geplante KI-Haftungsrichtlinie soll nur auf außervertragliche, verschuldensabhängige Schadensersatzansprüche anwendbar sein und wird daher vor allem im bestehenden Deliktsrecht zu Veränderungen führen. Eine Ausstrahlung auf vertragliche Ansprüche ist explizit nicht vorgesehen. In der Praxis könnten jedoch Spannungen entstehen, da Anspruchsteller möglicherweise versuchen, Beweise, die sie unter Art. 3 (Offenlegung von Beweismitteln) der KI-Haftungsrichtlinie gewonnen haben, auch für vertragliche Ansprüche zu verwenden. Denn das deutsche Zivilprozessrecht kennt im Grundsatz kein Institut, das eine Verwertung von Beweismitteln für vertragliche Ansprüche verwehrt.

Nach der geplanten KI-Haftungsrichtlinie soll der Anspruchsteller (beispielsweise ein KI-Anwender) das Recht haben, Informationen über ein Hochrisiko-KI-System zu erhalten, welches verdächtigt wird, einen Schaden verursacht zu haben. Diese Richtlinie zielt also auf die Normierung eines Auskunftsanspruchs gegen Anbieter und Betreiber von Hochrisiko-KI-Systemen ab. Das Hauptziel ist es, Anspruchsstellern effektive Werkzeuge zur Verfügung zu stellen, um potenziell haftende Personen zu ermitteln und einschlägige Beweismittel für einen Anspruch zu erhalten. Darüber hinaus soll die Richtlinie einen zusätzlichen Anreiz bieten, die Anforderungen des AI Acts bezüglich der Dokumentation und Aufzeichnung von Informationen zu erfüllen. Wird die Vorlage der geforderten Informationen versäumt, so wird eine Verletzung der Sorgfaltspflicht widerleglich

vermutet. Daher ist eine sorgfältige und umfassende Dokumentation notwendig, um sich effektiv gegen Ansprüche zu schützen. Zudem wird nach der KI-Haftungsrichtlinie ein Kausalzusammenhang zwischen dem Sorgfaltspflichtverstoß und der Ausgabe des KI-Systems widerleglich vermutet. Für Schadensersatzansprüche gegen Anbieter von Hochrisiko-KI-Systemen erfordert diese Kausalitätsvermutung, dass der Anspruchsteller nachweisen muss, dass der Anbieter bestimmte Vorgaben der KI-VO nicht erfüllt hat. Diese Vermutung setzt voraus, dass ein möglicher ursächlicher Zusammenhang zwischen dem Sorgfaltspflichtverstoß und dem Ausgabeverhalten des KI-Systems besteht und dass ebendieses Ausgabeverhalten ursächlich für den entstandenen Schaden war.

3.4.4.3 Die neue Produkthaftungsrichtlinie (EU) 2024/2853

Seit dem 8. Dezember 2024 ist die neue Produkthaftungsrichtlinie (EU) 2024/2853 in Kraft, die die älteren Regelungen aus dem Jahr 1985 ablöst. Die Mitgliedstaaten – darunter Deutschland – haben bis Dezember 2026 Zeit, die Richtlinie in ihr nationales Recht zu überführen. Für bereits zuvor in Verkehr gebrachte Produkte gilt das bisherige Rechtsregime aber fort.
Erweiterter Anwendungsbereich und neue Haftungsregeln

Die neue Richtlinie modernisiert die verschuldensunabhängige Haftung für fehlerhafte Produkte, indem sie den Produktbegriff ausdrücklich auf Software und KI-Systeme ausweitet. Unternehmen können somit auch für Schäden einstehen müssen, die durch fehlende oder mangelhafte Software-Updates oder unzureichende Cybersicherheitsmaßnahmen entstehen. Darüber hinaus umfasst der ersatzfähige Schaden nun nicht mehr nur Sach- und Personenschäden, sondern kann auch Datenverluste einschließen, die nicht für berufliche Zwecke verwendet werden.
Ein weiteres Merkmal ist die Einbeziehung zusätzlicher Akteure in den Kreis der möglichen Haftungspflichtigen: Neben Herstellern und Importeuren werden, unter bestimmten Voraussetzungen, beispielsweise auch Fulfillment-Dienstleister oder Online-Plattformen erfasst. Wer ein Produkt später »wesentlich verändert«, haftet unter Umständen wie ein Hersteller.
Leichtere Anspruchsdurchsetzung für Geschädigte

Um geschädigten Personen den Zugang zu Beweisen zu erleichtern, sieht die neue Produkthaftungsrichtlinie ein besonderes Verfahren zur Offenlegung relevanter Informationen vor. Wenn die geschädigte Person (Kläger) den Gerichtsgang beschreitet und Tatsachen vorlegt, die seinen Schadensersatzanspruch plausibel erscheinen lassen, kann das Gericht den Beklagten anweisen, bestimmte Beweismittel offenzulegen. Kommt der Beklagte dieser Anordnung nicht nach, greifen Vermutungsregeln zugunsten des Klägers.

Darüber hinaus enthält die Richtlinie neue Bestimmungen zu Verjährungs- und Ausschlussfristen: So beginnt die (kurze) Verjährungsfrist von drei Jahren erst, wenn der Geschädigte von Schaden, Fehlerhaftigkeit und verantwortlichem Wirtschaftsakteur Kenntnis erlangt hat. Zusätzlich gibt es eine (lange) Ausschlussfrist von zehn Jahren ab Inverkehrbringen oder Inbetriebnahme des Produkts, die bei verzögert zutage tretenden Gesundheitsschäden auf bis zu 25 Jahre anwachsen kann. Frühere finanzielle Haftungsobergrenzen oder Selbstbehalte sind in der neuen Richtlinie nicht mehr vorgesehen. Für Unternehmen erhöht dies potenziell das finanzielle Risiko im Produkthaftungsfall.

Zusammenspiel mit der KI-Haftungsrichtlinie

Obwohl die neue Produkthaftungsrichtlinie und der Entwurf der KI-Haftungsrichtlinie dieselbe Zielrichtung haben – nämlich eine faire und transparent geregelte Haftung im KI- und Softwarebereich –, unterscheiden sie sich in einigen Punkten. Die Produkthaftungsrichtlinie begründet eine verschuldensunabhängige Haftung von Herstellern oder Anbietern von fehlerhaften KI-Systemen, die physische Schäden, Sachschäden oder bestimmte Datenverluste verursachen, während die KI-Haftungsrichtlinie auf außervertragliche, verschuldensabhängige Haftung abzielt, und zwar für jede Art von Schaden, also auch immaterielle oder reine Vermögensschäden erfassen kann. Sobald die KI-Haftungsrichtlinie endgültig verabschiedet ist, dürften beide Normen zusammen ein verzahntes Haftungskonzept bilden.

3.5 Arbeitsrechtliche Aspekte

Einer der wichtigsten und zugleich komplexesten Bereiche, in denen KI eingesetzt wird, ist die Arbeitswelt. KI-Technologien revolutionieren nicht nur die Art und Weise, wie Arbeit ausgeführt wird, sondern werfen auch grundlegende Fragen zum Arbeitsrecht auf. Der folgende Abschnitt zielt darauf ab, die arbeitsrechtlichen Aspekte im Zusammenhang mit der Anwendung von KI zu beleuchten und einen umfassenden Überblick über die Herausforderungen und Chancen zu geben, die sich aus dem Einsatz von KI am Arbeitsplatz ergeben.

3.5.1 Individualarbeitsrecht

Die Integration von KI in den Arbeitsalltag bringt neue Herausforderungen für das Individualarbeitsrecht mit sich.

3.5.1.1 Risikostufen

Der AI Act verfolgt einen risikobasierten Ansatz. Der Arbeitgeber muss daher ab dem 2. August 2025 KI-Systeme schon im Planungsstand einordnen und entsprechend behandeln, ob diese ein unannehmbares Risiko, ein hohes Risiko oder geringes bzw. kein Risiko mit sich bringen. KI-Systeme mit einem unannehmbaren Risiko sind verboten. Bei KI-Systemen mit einem hohen Risiko sind Auflagen zu beachten, wobei diese bei KI-Systemen mit einem geringen oder ohne Risiko gering ausfallen können.

3.5.1.2 Rechtsgrundlage für die Verarbeitung personenbezogener Daten

Einwilligung von Bewerbenden / Mitarbeitenden

Die Verarbeitung personenbezogener Daten von Mitarbeitende bedarf nach der DSGVO einer Rechtsgrundlage. Eine Einwilligung der Mitarbeitenden könnte eine solche sein. Aber die Anforderungen an die Einwilligung sind hoch. Es ist eine unmissverständliche Erklärung bzw. eine eindeutig bestätigende Handlung des Mitarbeitenden erforderlich, dass er mit der Datenverarbeitung einverstanden ist. Die Einwilligung muss in informierter Weise erfolgen. Problematisch ist weiterhin, ob die Einwilligung freiwillig erfolgt. Mitarbeitende müssen eine echte, freie Wahl haben. Durch eine Verweigerung dürfen dem Mitarbeitenden keine beruflichen Nachteile entstehen, so dürfte er z. B. eine gleichwertige Alternative zu der KI-Verwendung erhalten.Bei Bewerbenden wird man in der Regel immer von einem Nachteil ausgehen können, da bei einer Verweigerung der Einwilligung mit einer Nichtberücksichtigung der Bewerbung zu rechnen ist.

Durchführung des Arbeitsvertrages bzw. Bewerbungsverfahrens

Die Verarbeitung der personenbezogenen Daten muss für die Erfüllung des Arbeitsvertrages oder für das Bewerbungsverfahren erforderlich sein. Es darf keine mildere, gleich geeignete Maßnahme möglich sein.

3.5.1.3 Arbeitnehmerüberwachung

Durch die Nutzung von KI-Systemen durch Mitarbeitende werden Daten gespeichert, die ggf. potenziell auch für eine Überwachung und eine Analyse der Mitarbeitenden verwendet werden könnten. Ein solcher Einsatz der KI wird in der Regel verboten sein. Daraus gewonnene Erkenntnisse dürften dann einem (Sach-/Beweis-)Verwertungsverbot in einem arbeitsgerichtlichen Verfahren unterliegen, sodass darauf z. B. keine Kündigung des Mitarbeitenden gestützt werden könnte. Für eine Verwertung ist ein berechtigtes Arbeitgeberinteresse erforderlich. Es darf kein milderes Mittel verfügbar sein. Außerdem müssen Schutzmaßnahmen zugunsten von Mitarbeitenden bestehen.

3.5.1.4 Weisungsrecht

Das Direktionsrecht des Arbeitgebers berechtigt ihn, seine Arbeitnehmerinnen und Arbeitnehmer zu verpflichten, KI oder bestimmte KI-Systeme als Arbeitsmittel zu verwenden, wenn dies vom vertraglichen Tätigkeitsbereich umfasst ist. Das Weisungsrecht beinhaltet nicht nur das Ob der Nutzung, sondern auch das Wie, also welche Verhaltensregeln und -verbote diesbezüglich zu beachten sind. Wenn Arbeitnehmerinnen und Arbeitnehmer gegen die Weisungen verstoßen, stellt dies eine Verletzung arbeitsvertraglicher Pflichten dar. In solchen Fällen ist der Nachweis des Verstoßes durch den Arbeitgeber entscheidend. Dieser ist gerade bei externen KI-Systemen, auf die der Arbeitgeber keinen Zugriff hat, sehr schwierig zu erbringen. Die Reaktion auf eine Pflichtverletzung kann eine Abmahnung sein, und in schwerwiegenderen Fällen oder bei wiederholten Verstößen ist sogar eine verhaltensbedingte Kündigung möglich. Wenn der Arbeitgeber die Installation und Nutzung von KI nicht will (z. B. zur Vermeidung von AGG-Verstößen, Schutz von Geschäftsgeheimnissen, Vermeidung von Urheberrechtsverletzungen), kann und muss er dies dem Mitarbeitenden untersagen. Unternehmen können ihren Beschäftigten die Nutzung von KI-Tools einseitig verbieten. Eine solche Weisung unterliegt nicht dem Mitbestimmungsrecht des Betriebsrats.

3.5.1.5 KI-Nutzung als Arbeitsmittel

Ein Mitarbeitender muss seine Arbeitsleitung höchstpersönlich erbringen. Somit stellt sich die Frage, ob der Mitarbeitende seine Arbeit noch höchstpersönlich erbringt, wenn er eine KI nutzt, also ob der Mitarbeitende ohne die Weisungen des Arbeitgebers KI nutzen darf. Höchstpersönlich bedeutet, dass der Mitarbeitende im Zweifel persönlich leisten muss und dies nicht auf Dritte übertragen darf, außer es ist vertraglich vereinbart. Daher könnte man meinen, dass die

Nutzung einer KI eine Übertragung auf Dritte ist. Ein KI-Tool ist aber kein »Dritter, sondern ein Arbeitsmittel wie ein PC, ein Werkzeug etc. Es ist zu empfehlen, dass der Arbeitgeber Nutzungsrichtlinien bzw. Betriebsvereinbarungen aufstellt« (siehe 3.5.5 Erstellung einer unternehmensinternen Richtlinie zur Nutzung generativer KI). Mitarbeitende sollten zudem im richtigen Umgang mit KI geschult werden (siehe 3.5.1.10 Vermittlung von KI-Kompetenzen (Artikel 4 KI Act).

Die heimliche Nutzung von KI-Tools von Mitarbeitenden zur Erledigung von Aufgaben ist nur bei ausdrücklicher Nutzungsuntersagung durch den Arbeitgeber ein Verstoß gegen die arbeitsvertraglichen Pflichten (siehe 3.5.1.4 Weisungsrecht).
Der Mitarbeitende muss die Nutzung von KI dem Arbeitgeber offenbaren (vertragliche Nebenpflicht).

3.5.1.6 Arbeitsfehler bei der Nutzung einer KI

Arbeitsfehler eines Mitarbeitenden sind grundsätzlich Arbeitspflichtverstöße, die zu arbeitsrechtlichen Konsequenzen für den Mitarbeitenden führen können, so insbesondere Schadensersatz (je nach Verschulden), Abmahnung oder (fristlose) Kündigung. Dies gilt auch bei Fehlern bei der Nutzung einer KI. Der Mitarbeitende kann sich nicht damit herausreden, dass er ja eine »Künstliche Intelligenz« genutzt habe, sich daher auf diese verlassen konnte, und nichts weiter machen musste. Ein Verschulden könnte aber ggf. dann ausscheiden, wenn die KI als Arbeitsmittel vom Arbeitgeber gestellt wurde, der Fehler unvermeidbar war und der Mitarbeitende keine Kontrollmöglichkeit hatte. Die Rechtsprechung wird sich damit zu beschäftigen haben, ob nicht der Arbeitgeber die Verantwortung für Arbeitsfehler trägt, die aus der angewiesenen Verwendung eines fehlerhaften KI-Tools entstehen, und wie weit die Kontrollpflichten des Mitarbeitenden gehen.

3.5.1.7 Rationalisierung durch KI

Können durch den Einsatz von KI Rationalisierungsmaßnahmen umgesetzt werden? Dies kann nicht mit einem einfachen Ja oder Nein beantwortet werden. Die Entscheidung des Arbeitgebers, KI einzusetzen, muss zu einem spezifischen und berechenbaren Überhang an Arbeitskräften und damit mithin zum Wegfall von Beschäftigungsmöglichkeiten führen. Der Arbeitgeber muss diesen bei Bestreiten rechnerisch nachvollziehbar darstellen und nachweisen. Mit der reinen Verlagerung von Arbeitsschwerpunkten durch den Einsatz von KI (Verlagerung einer Tätigkeit) spart sich der Arbeitgeber nicht automatisch Arbeitskräfte, sodass es nicht unbedingt zu einem Wegfall eines Beschäftigungsbedarfs kommt. Zu denken ist dabei z. B. an einen u. U. höheren Kontrollaufwand.

3.5.1.8 Geschäftsgeheimnisse

Ein weiterer wichtiger Aspekt ist der Schutz von Geschäftsgeheimnissen. KI-Systeme, die große Mengen sensibler Daten verarbeiten, erfordern besondere Sicherheitsmaßnahmen. Dies gilt umso mehr, wenn die KI die Eingaben zu Trainingszwecken verarbeitet und die Eingaben als Output bei unternehmensfremden Nutzerinnen und Nutzern der KI erscheinen können. Arbeitnehmerinnen und Arbeitnehmer sind verpflichtet, Geschäftsgeheimnisse zu schützen, was den verantwortungsvollen Umgang mit KI-Systemen einschließt, die Zugang zu diesen Informationen haben.

3.5.1.9 Schutz vor Diskriminierungen durch KI (AGG)

- Arbeitgeber können beim Einsatz von KI gegen die Vorgaben des Allgemeinen Gleichbehandlungsgesetzes (AGG) verstoßen und somit verpflichtet sein, betroffenen Beschäftigten eine angemessene Entschädigung und/oder Schadensersatz zu zahlen, vgl. § 15 AGG.
- Gem. § 1 AGG ist Ziel dieses Gesetzes, Benachteiligungen aus Gründen der Rasse oder wegen der ethnischen Herkunft, des Geschlechts, der Religion oder Weltanschauung, einer Behinderung, des Alters oder der sexuellen Identität zu verhindern oder zu beseitigen. Auch wenn der Anwendungsbereich des AGG nicht auf das Arbeitsrecht beschränkt ist, so spielen dennoch arbeitsrechtliche Streitigkeiten im Zusammenhang mit dem AGG eine wesentliche Rolle.
- Voraussetzung für eine Haftung des Arbeitgebers nach den Vorgaben des AGG ist, dass er Beschäftigte (vgl. § 6 AGG) ist, wegen eines der in § 1 AGG genannten Merkmalen mittelbar oder unmittelbar benachteiligt oder (sexuell) belästigt. Hierbei gilt die Anweisung zur Benachteiligung einer Person aus einem in § 1 AGG genannten Grundes ebenfalls als Benachteiligung (vgl. § 3 Abs.5 S.1 AGG).
- Eine unmittelbare Benachteiligung liegt vor, wenn eine (konkrete) Person wegen eines in § 1 genannten Grundes eine weniger günstige Behandlung erfährt, als eine andere Person in einer vergleichbaren Situation erfährt, erfahren hat oder erfahren würde.
- Werden KI-Systeme unzureichend programmiert oder haben sie eine ungenügende oder fehlerhafte Datengrundlage (»Dirty Data«), aus der sie dann logische Schlüsse im Rahmen des Selbstlernprozesses ziehen, kann eine unmittelbare oder mittelbare Diskriminierung eines Bewerbers oder eines Arbeitnehmers wegen eines in § 1 AGG genannten Merkmals zu vermuten sein, auch wenn dies vom Betreiber gar nicht intendiert war. Weiterhin neigt eine selbstlernende KI zur Verstärkung bestehender Strukturen. Auch eine ungenaue Kenntnis des (fortgeschriebenen) KI-Algorithmus ist hierbei zu bedenken (sog. Blackbox-Phänomen). Die Betreiber von KI-Systemen, also hier die Arbeitgeber, müssen deshalb darauf achten, wenn sie solche Systeme im Personalmanagement und im Bewerbungsprozess einsetzen, dass Entscheidungen der KI-Systeme nicht gegen das Benachteiligungsverbot verstoßen.

3.5.1.10 Vermittlung von KI-Kompetenzen (Artikel 4 KI Act)

Bereits am 2. Februar 2025 trat Artikel 4 des AI Act in Kraft. Unternehmen müssen sicherstellen, dass Mitarbeitende, die mit solchen Systemen arbeiten, über die erforderliche KI-Kompetenz verfügen. Dazu zählen insbesondere Kenntnisse über die Funktionsweise von KI, rechtliche und ethische Rahmenbedingungen sowie das Bewusstsein für Chancen und Risiken und mögliche Schäden (Definition von KI-Kompetenzen in Art. 3 Nr. 56 AI Act). Unternehmen müssen daher geeignete Maßnahmen ergreifen, um ihren Mitarbeitenden die notwendigen Fähigkeiten und Kenntnisse zu vermitteln. Diese Verpflichtung gilt unabhängig von der Unternehmensgröße und betrifft alle Unternehmen, die KI-Systeme entwickeln, betreiben oder einsetzen. Mit den Schulungen sollten Verständnis und Akzeptanz geschaffen werden. Sie sind aber auch ein Compliance-Instrument.

Unternehmen können diese Verpflichtung durch folgende Maßnahmen umsetzen: Erstellung einer Bedarfsanalyse (ausgerichtet auf jeweilige Mitarbeitendengruppen), Entwicklung von Schulungsmaßnahmen (z. B. Basisschulungen für alle Mitarbeitenden und gezielte Schulungen je nach Einsatz und Funktion der jeweiligen Mitarbeitendengruppe), Etablierung von internen Richtlinien (siehe 3.5.5), Einsatz eines KI-Beauftragten.

3.5.2 Betriebsverfassungsgesetz

Bei der Einführung und Anwendung von KI-Systemen sind die Mitbestimmungs- und Beteiligungsrechte des Betriebsrats zu beachten.

§ 87 Abs. 1 Nr. 1 BetrVG bezieht sich auf das Mitbestimmungsrecht bei »Ordnung des Betriebs und des Verhaltens der Arbeitnehmer im Betrieb«. Dies kann auch die Einführung und Anwendung von KI-Systemen umfassen, sofern diese Systeme einen Einfluss auf das Ordnungsverhalten der Mitarbeiterinnen und Mitarbeiter haben. Beispielsweise kann der Betriebsrat mitbestimmen, wenn eine KI-Lösung eingesetzt wird, um zu kontrollieren, ob ein vom Arbeitgeber herausgegebener Dresscode am Arbeitsplatz eingehalten wird. Allerdings umfasst § 87 I Nr. 1 BetrVG nicht das Arbeitsverhalten im Betrieb, welches dem Direktionsrecht des Arbeitgebers unterliegt. Wenn also eine KI zur Steuerung von Arbeitsprozessen, zur Verteilung von Arbeitsaufgaben oder zur Überwachung und Steuerung von Arbeitsabläufen verwendet wird, betrifft dies das Arbeitsverhalten und § 87 I Nr. 1 BetrVG findet keine Anwendung. Auch die Weisung, KI-Tools zu nutzen oder das Untersagen der KI-Nutzung betrifft das Arbeitsverhalten und unterliegt daher nicht der Mitbestimmung nach § 87 Abs. 1 Nr. 1 BetrVG.

Dennoch können andere Mitbestimmungs- und Beteiligungsrechte des Betriebsrates einschlägig sein, insbesondere § 87 I Nr. 6 BetrVG.

Ein Schlüsselelement ist das Mitbestimmungsrecht nach § 87 Abs. 1 Nr. 6 BetrVG. Dieses gewährt dem Betriebsrat ein Mitbestimmungsrecht bei der Einführung und Anwendung von Technologien, die potenziell zur Überwachung von Verhalten oder Leistung der Arbeitnehmer dienen können. KI-Systeme fallen oft unter diese Kategorie, da sie regelmäßig Daten sammeln und

analysieren, die Rückschlüsse auf die Leistung oder das Verhalten der Arbeitnehmer zulassen. Entscheidend ist hierbei die objektive Eignung der Technologie zur Überwachung, unabhängig von der Absicht des Arbeitgebers oder dem Verwendungszweck des KI-Systems.

Des Weiteren sind die Unterrichtungs- und Beratungsrechte nach § 90 Abs. 1 Nr. 3 BetrVG von Bedeutung. Diese verpflichten den Arbeitgeber zur frühzeitigen Einbindung des Betriebsrats in die Planungsprozesse für den Einsatz von KI, sodass dieser seine Sichtweisen und Bedenken einbringen kann. § 90 I Nr. 3 BetrVG nennt ausdrücklich den Einsatz künstlicher Intelligenz als Anwendungsfall.

Das Recht zur Hinzuziehung eines Sachverständigen nach § 80 Abs. 3 S. 2 BetrVG ist ebenfalls von großer Wichtigkeit. Aufgrund der Komplexität von KI-Technologien kann der Betriebsrat Expertinnen und Experten hinzuziehen, um ein umfassendes Verständnis der Implikationen und Potenziale der Technologie zu gewinnen. Während der Gesetzgeber das Ob der Hinzuziehung anerkennt, bleibt der Umfang der Hinzuziehung offen. Hierbei ist zu berücksichtigen, dass auch der Betriebsrat dazu angehalten werden sollte, im angemessenen Umfang unternehmensinterne Expertise zu nutzen.

Schließlich regelt § 95 Abs. 2a BetrVG eine Zustimmungsbedürftigkeit bei der Anwendung von KI-generierten Mitarbeitenden-Auswahlrichtlinien. Dies bedeutet, dass der Einsatz von KI-Systemen zur Generierung von Richtlinien, beispielsweise für Beförderungen oder Versetzungen, der Zustimmung des Betriebsrats bedarf.

Wenn ein Arbeitgeber plant, Maßnahmen durchzuführen, die den Betrieb wesentlich verändern, wie in § 111 BetrVG beschrieben, kann es sich im Einzelfall um eine Betriebsänderung handeln und der Arbeitgeber muss dies in dem Fall frühzeitig mit dem Betriebsrat im Rahmen eines Interessenausgleichs besprechen und verhandeln. Die Einführung von Künstlicher Intelligenz kann unter bestimmten Umständen als solch eine Betriebsänderung angesehen werden.
Die Implementierung von KI kann zu Effizienzsteigerungen führen, die möglicherweise einen geringeren Bedarf an Arbeitskräften nach sich ziehen. Dies kann von der Reduzierung der Arbeitsplätze bis hin zur Schließung ganzer Betriebe oder wesentlicher Teile davon reichen. Ein solcher Personalabbau könnte schon eine betriebsändernde Maßnahme nach § 111 Abs. 1 S. 3 Nr. 1 BetrVG sein. Allerdings werden solche Rationalisierungsmaßnahmen oft erst nach der Entscheidung zur Einführung von KI getroffen, welche selbst eine eigenständige Betriebsänderung sein kann.
Die Einführung von KI in einem Unternehmen kann wesentliche Veränderungen in den Betriebsanlagen und Arbeitsmethoden mit sich bringen. Diese Veränderungen können Betriebsänderungen nach § 111 S. 3 Nr. 4 und 5 BetrVG darstellen. Es ist wichtig zu untersuchen, ob der Arbeitgeber grundlegend neue Arbeitsmethoden oder Produktionsverfahren einführt, indem er entweder die Art und Weise, wie menschliche Arbeitskraft eingesetzt wird, oder das technische Verfahren ändert. Jedoch stellt nicht jede Verbesserung der Methoden und Verfahren eine Betriebsänderung dar. Es muss eine wesentliche Änderung sein, bei der eine neue Methode die bisherige ersetzt. Ob eine Änderung durch KI-Einsatz als wesentlich angesehen wird, ist umstritten. Einige Fachmeinungen betonen qualitative Aspekte und verlangen einen deutlichen technologischen Fortschritt. Bei Unsicherheiten kann auch die Anzahl der betroffenen Mitarbeitenden als Indiz herangezogen werden.

Zusammengefasst ist es entscheidend, die spezifischen Umstände jedes Einzelfalls zu berücksichtigen und die Auswirkungen der Maßnahme auf Arbeitsplätze und Arbeitsbedingungen im jeweiligen Unternehmen genau zu analysieren.

Falls Arbeitsplatzverluste trotz des Interessenausgleichs unvermeidlich sind, kommt der Sozialplan gemäß § 112 Abs. 1 S. 2 BetrVG zum Einsatz. Dieser regelt den Ausgleich oder die Milderung der wirtschaftlichen Nachteile für die betroffenen Beschäftigten, die durch die Betriebsänderung entstehen, beispielsweise durch Abfindungen, Umschulungen oder andere Unterstützungsformen.

3.5.3 Betriebssicherheitsverordnung

Die Betriebssicherheitsverordnung (BetrSichV) in Deutschland spielt eine zentrale Rolle bei der Gewährleistung von Sicherheit und Gesundheitsschutz bei der Verwendung von Arbeitsmitteln und -anlagen. Im Kontext der fortschreitenden Integration von Künstlicher Intelligenz in industrielle und gewerbliche Prozesse ergeben sich neue Herausforderungen und Chancen, die es im Rahmen der BetrSichV zu betrachten gilt.

Bei der Implementierung von KI in Arbeitsprozessen ist es wichtig, potenzielle Gefahren zu identifizieren und geeignete Maßnahmen zur Minimierung dieser Risiken zu ergreifen.

Zum einen muss die physische Sicherheit der Mitarbeiterinnen und Mitarbeiter berücksichtigt werden. Dies bezieht sich insbesondere auf Arbeitsplätze, an denen Menschen direkt mit KI-gesteuerten Maschinen interagieren. Es muss sichergestellt werden, dass diese Systeme sicher sind und nicht zu Verletzungen führen. Dazu gehört die Überprüfung der Maschinensicherheit, die Einhaltung von Sicherheitsabständen und die Implementierung von Notfallstoppsystemen.

Daneben sind die psychologischen Auswirkungen der KI-Nutzung zu beachten. Die Angst vor Arbeitsplatzverlust, der Umgang mit überwachenden KI-Systemen oder die ständige Interaktion mit KI können Stress und psychische Belastungen für die Mitarbeitenden bedeuten. Hier ist es wichtig, Schulungen anzubieten, Transparenz zu schaffen und Unterstützung für die Beschäftigten zu gewährleisten.

3.5.4 Arbeitsschutz bzw. Anwendung von KI und Datenschutz

Der Einsatz von KI-Technologien zur Bearbeitung personenbezogener Beschäftigtendaten erfordert ein Gleichgewicht zwischen technologischen Innovationen und dem Schutz der Privatsphäre der beteiligten Mitarbeitenden. Dieser Abschnitt hebt die wesentlichen Herausforderungen im Datenschutz hervor, die mit der Verwendung von KI einhergehen (siehe zu den allgemeinen datenschutzrechtlich relevanten Bestimmungen oben ausführlich). Sobald ein Arbeitgeber personenbezogene Beschäftigtendaten mittels KI verarbeitet, werden sowohl die Datenschutzgrundverordnung als auch das Bundesdatenschutzgesetz relevant.

Rechtsgrundlage für die Verarbeitung

Die Nutzung von KI zur Datenverarbeitung von Mitarbeitenden, einschließlich Bewerberinnen und Bewerbern gemäß § 26 Abs. 8 Nr. 2 BDSG, muss auf eine Verarbeitungsgrundlage gestützt werden. Alle Grundlagen in Art. 6, 9 DSGVO und § 26 BDSG, mit Ausnahme der Einwilligung, verlangen die Erforderlichkeit der Datenverarbeitung. Insbesondere in den Bereichen der Persönlichkeitsanalyse und Emotionserkennung ist diese Erforderlichkeit sorgfältig zu prüfen, was eine Abwägung zwischen den Persönlichkeitsrechten der Mitarbeiterin bzw. des Mitarbeiters und den Verarbeitungsinteressen des Arbeitgebers voraussetzt.

Sofern eine Betriebsvereinbarung die Verarbeitungsgrundlage darstellt, ist sorgfältig darauf zu achten, dass durch diese nicht das Datenschutzniveau der DSGVO abgesenkt wird. Insofern bietet es sich an, auch hier einen Erforderlichkeitsmaßstab für die Datenverarbeitung anzulegen.

Eine Datenverarbeitung kann auch mit der Einwilligung des oder der Beschäftigten gemäß § 26 Abs. 2 S. 1 BDSG erfolgen, vorausgesetzt, sie ist insbesondere freiwillig erteilt. Besonders in Bewerbungsverfahren kann die Freiwilligkeit herausfordernd sein, wenn Bewerberinnen und Bewerber Nachteile bei der Ablehnung der Einwilligung fürchten. Dem kann entgegengewirkt werden, wenn traditionelle Bewerbungsverfahren ohne KI-Einsatz eine gleichwertige Option sind. Wirtschaftliche Vorteile für Beschäftigte, wie etwa das kostenlose Erhalten eines durch KI erstellten Persönlichkeitsprofils, können auch eine Rolle spielen.

Rechte der Betroffenen

Gemäß Art. 12 ff. DSGVO haben betroffene Mitarbeitende bestimmte Rechte, wenn ihre Daten mittels KI verarbeitet werden. Arbeitgeber müssen Mitarbeitende über die Datenverarbeitung umfassend informieren, nicht benötigte Daten löschen und unrichtige Daten korrigieren. Bei automatisierten Entscheidungen, einschließlich Profiling, gemäß Art. 13 Abs. 2 lit. f und Art. 14 Abs. 2 lit. g DSGVO, bestehen besondere Informationspflichten. Arbeitgeber müssen über die Logik, Tragweite und die Auswirkungen der Datenverarbeitung informieren, was bei komplexen KI-Systemen eine Herausforderung darstellen kann.

Verbot automatisierter Entscheidungen

Art. 22 Abs. 1 DSGVO verbietet automatisierte Entscheidungen, die rechtliche Wirkungen haben oder die betroffene Person erheblich beeinträchtigen. KI darf demnach keine finalen Entscheidungen über Einstellungen, Beförderungen, Kündigungen oder Abmahnungen treffen. Ausnahmen von diesem Verbot sind in Art. 22 Abs. 2 DSGVO geregelt, allerdings nur unter strengen Voraussetzungen.

Datenschutz-Folgenabschätzung

Der Einsatz von KI unter der Verarbeitung von Beschäftigtendaten kann eine Datenschutz-Folgenabschätzung gemäß Art. 35 DSGVO erforderlich machen. Hierbei sind die Verarbeitungsprozesse, Risiken für betroffene Personen und risikomindernde Maßnahmen zu analysieren und zu bewerten. Die Komplexität und Dynamik von KI-Systemen stellen dabei eine besondere Herausforderung dar, sowohl in Bezug auf die Erfüllung von Informationspflichten und Transparenz als auch bei der Bewertung der Verarbeitungsvorgänge und Risiken.

3.5.5 Erstellung einer unternehmensinternen Richtlinie zur Nutzung generativer KI

Die Einführung einer internen Richtlinie zum Umgang mit generativen KI-Tools in einem Unternehmen ist aus mehreren Gründen sinnvoll. Zum Ersten bietet sie die Möglichkeit, die Anwenderinnen und Anwender innerhalb des Unternehmens über Möglichkeiten und Anwendung dieser Tools zu informieren und ihnen ein Basisverständnis über deren Funktionsweise zu verschaffen. Hierbei sollte sie den Anwendenden auch ein Gefühl dafür geben, was ein solches Tool zu leisten vermag und wo andererseits dessen Grenzen liegen. Zudem sollte in der Richtlinie erwähnt werden, welche Risiken für das Unternehmen mit der Nutzung von KI-Systemen einhergehen können, z. B. Urheberrechts- oder Datenschutzverletzungen. Ein weiterer Zweck ist die Festlegung und innerbetriebliche Umsetzung von verbindlichen Regeln und Verhaltensvorgaben. Die Richtlinie sollte dabei nicht nur die einschlägigen, rechtlichen Anforderungen berücksichtigen, sondern idealerweise auch auf die konkreten, unternehmensrelevanten Anwendungsfälle eingehen. Entscheidend dabei ist, dass die Anwenderinnen und Anwender als Adressaten der Richtlinie in die Lage versetzt werden, die Regelungen in der täglichen Praxis auch umsetzen zu können. Auf diese Weise lassen sich rechtliche Risiken für das Unternehmen effektiv minimieren.

Nicht zu unterschätzen ist auch die vertrauensbildende Funktion einer solchen Richtlinie. So kann sie Vertrauen fördern, indem sie zur Transparenz in Hinblick auf die Nutzung von KI-Tools beiträgt und signalisiert, dass das Unternehmen nicht nur vertrauensvoll mit solchen Tools umgeht, sondern sich auch der damit verbundenen Risiken bewusst ist. Dieses Vertrauen geht einher mit ethischen Überlegungen. Je mehr sich das Unternehmen mit den potenziellen Auswirkungen der Nutzung von KI-Tools sowohl auf die Gesellschaft als auch einzelne Stakeholder auseinandersetzt, desto besser kann es den Risiken und Herausforderungen von KI (wie zum Beispiel Missbrauch, Manipulation und Diskriminierung) begegnen. Auch auf ESG-Ziele eines Unternehmens kann die Richtlinie einzahlen, indem sie die Anwenderinnen und Anwender dazu anleitet und befähigt, Auswirkungen z. B. auf die Umwelt oder die Gesellschaft bei der Nutzung von KI-Tools zu berücksichtigen.

Vor diesem Hintergrund sollte die Richtlinie inhaltlich sowohl auf das Unternehmen als auch auf die Anwenderinnen und Anwender der KI-Tools ausgerichtet sein. Vor allem folgende Fragestellungen können dabei adressiert werden:

- Welche KI-Tools und Applikationen sind vom Unternehmen zu welchem Zweck zur Nutzung durch Mitarbeiterinnen und Mitarbeiter freigegeben?
- Unter welchen Bedingungen dürfen darüber hinaus andere KI-Tools und Applikationen genutzt werden?
- Beschränkung auf bestimmte Anbieter von KI
- Beschränkung des Einsatzes von KI auf bestimmte Unternehmensbereiche
- Ist die Privatnutzung gestattet oder ausgeschlossen?
- Dürfen Mitarbeitende eigene, private KI-Accounts nutzen?
- Was ist bei der Erstellung der Eingabe bzw. der »Prompts« zu beachten?
- Voraussetzung für Kennzeichnungspflicht von KI-generierten Inhalten
- Weitergabe von KI-Ergebnissen an Dritte bzw. Entgegennahme von Dritten
- Umgang mit Geschäftsgeheimnissen
- Datenschutzregelungen bei Einsatz von KI
- Welche Schulungsangebote gibt es?
- Gibt es Vorgaben zur Dokumentation? Zum Beispiel, welche Eingaben für den generierten Content genutzt wurden bzw. wie der Prozess war, um zur endgültigen Eingabe zu kommen bzw. welchen menschlichen Beitrag zum generierten Content geleistet wurde.
- Welche Vertraulichkeitsklassen sind für die Eingabe von Informationen zugelassen?
- In welchen Grenzen dürfen ggf. personenbezogene Daten für die Eingabe genutzt werden?
- Wie ist die Ausgabe zu prüfen, ggf. zu überarbeiten und wie darf sie weiterverwendet werden?
- Wie ist die Ausgabe auf Copyright und IP-Verletzungen zu prüfen?
- Wie ist die Ausgabe für die Weiterverwendung als KI-generierter Content zu kennzeichnen?
- Welche Rechte werden an der Ausgabe erworben und können diese an Dritte übertragen werden?
- Ansprechpartner für Fragen und Regelungen von Sanktionen

Angesichts des rapiden technischen Fortschrittes sowie der ebenso zu erwartenden regulatorischen und gerichtlichen Entwicklungen zu rechtlichen Fragen ist es essenziell, dass diese Entwicklungen fortlaufend verfolgt werden und relevante Entwicklungen in der Richtlinie verarbeitet werden.

Eine Möglichkeit, die Richtlinie den jeweiligen Anwenderinnen und Anwendern näherzubringen, ist, entsprechende Schulungen (ggf. elektronisch) zum Inhalt der Richtlinie anzubieten. Zudem kann die Teilnahme an einer solchen Schulung auch zur Voraussetzung gemacht werden, dass den Mitarbeitenden entsprechende generative KI-Tools zur Verfügung gestellt werden.

3.6 Verträge und Willenserklärungen

In einer Welt, in der KI zunehmend in verschiedenen Bereichen eingesetzt wird, ist es wichtig, ihre rechtliche Stellung und die mit ihr zusammenhängenden Auswirkungen und Regelungsmöglichkeiten zu verstehen. Hierbei sind in erster Linie die nachfolgenden Themenkomplexe gedanklich voneinander zu trennen: der Vertragsschluss und die Loslösung von Verträgen, die unter Verwendung von KI zustande kommen einerseits, sowie Verträge über (den Einsatz von) KI andererseits.

3.6.1 Vertragsschluss und Loslösung von ‚KI'-Verträgen

Auch im Zeitalter von KI ist das Zustandekommen von Verträgen weiterhin nach allgemeinen zivilrechtlichen Regeln zu beurteilen. Wird KI im Zusammenhang mit einem Vertragsschluss eingesetzt, so stellt sich in dieser Hinsicht jedoch die Frage nach der rechtlichen Qualität des konkret durch KI erbrachten Beitrags. Nach den im Zivilrecht geltenden Regelungen ist im konkreten Einzelfall die Frage aufzuwerfen, wer oder was bzw. wie eine zum Vertragsschluss führende Willenserklärung abgegeben hat bzw. wurde.

3.6.1.1 Vertragsschluss mittels KI

Bereits vor Anstieg der Aktualität beim Thema Generativer KI waren nicht unmittelbar von bzw. durch Menschen abgegebene Willenserklärungen relevant und wurden rechtlich thematisiert. Der Geschäftsverkehr im E-Commerce steht exemplarisch für die Funktionsweise und die Durchsetzung automatisierter Erklärungen im Geschäftsverkehr, die täglich millionenfach und zu unterschiedlichsten Uhrzeiten versendet werden.

Neu und hiervon abzugrenzen sind jedoch Erklärungen, die durch (vermeintlich) autonome Systeme abgegeben werden. In diesem Kontext sind solche Erklärungen relevant, die auf einer Art eigener Willenserklärung der Systeme basieren. Unter den Stichwörtern der Reaktivität, d. h. der Fähigkeit, Schlussfolgerungen aus Informationen zu ziehen und Funktionalitäten entsprechend anzupassen, sowie der Proaktivität, d. h. der Möglichkeit aus sich selbst heraus tätig zu werden um vorgegebene Ziele zu erreichen, können die Besonderheiten generativer KI dargelegt werden. Für Willenserklärungen bedeutet dies, dass diese über den vorab vorgegebenen inhaltlichen Umfang hinausgehen, mithin nicht lediglich den bereits vorprogrammierten menschlichen Willen wiedergeben bzw. konkretisieren können.

Problematisch, und zugleich zentraler Aspekt ist hierbei, dass KI sowohl nach bislang geltendem Recht als auch am Maßstab der neuen KI VO (auch) weiterhin keine eigene Rechts- und Geschäftsfähigkeit besitzt bzw. besitzen wird. Lediglich de lege ferenda besteht die Möglichkeit,

‚vollautonomer' KI eine eigene Rechtspersönlichkeit respektive (beschränkte) Geschäftsfähigkeit zuzusprechen.

Folgt man diesem geltenden Gedanken der objektiven KI-Instrumentalisierung, so kann KI heute nicht als eigenständige rechtliche Einheit betrachtet werden und folglich Verträge grundsätzlich weder für sich, noch im eigenen Namen Dritte abschließen respektive rechtlich verbindliche Entscheidungen treffen. Daher muss die zum Vertragsschluss führende Willenserklärung trotz des Einsatzes von KI entweder durch eine natürliche Person abgegeben werden, oder dieser zumindest zurechenbar sein.

Vor diesem Hintergrund – der fehlenden Rechts- bzw. Teilgeschäftsfähigkeit – kann KI jedoch vergleichbar zu der Rolle eines Vertreters, Boten oder Erfüllungsgehilfen agieren, und hierbei für Personen oder Unternehmen Informationen übermitteln, spezifische Anweisungen ausführen, oder in bestimmten Prozessen assistieren, wie etwa bei der Datenauswertung oder der Automatisierung von Routineaufgaben.

Es ist entscheidend zu verstehen, dass trotz der fortschreitenden Entwicklung und der zunehmenden Autonomie von KI-Systemen die rechtliche Verantwortung immer bei Menschen oder juristischen Personen verbleibt. Dies stellt sicher, dass trotz des Einsatzes von KI die menschliche Verantwortung und Kontrolle gewahrt bleiben, was insbesondere in rechtlichen und ethischen Fragen von großer Bedeutung ist.

Lediglich aus dogmatischer Sicht stellt sich die Frage, wie genau die Zurechnung zu erfolgen hat. Hierbei existieren verschiedene juristische Argumentationen mit jeweils eigenen Stärken und Schwächen. Diese bemühen sich beispielsweise um die Bejahung zivilrechtlicher Stellvertretungsgrundsätze qua objektiver und subjektiver Zurechnung der beispielsweise elementaren Bestandteile einer Willenserklärung – Handlungs-, Erklärungs -, und Geschäftswille – gemäß oder analog §§ 164 ff. BGB. Teilweise werden unter Verweis auf den Gesetzgeberwillen, konkret § 165 BGB, stattdessen die Grundsätze der Botenschaft angewandt, was unter Berücksichtigung des Abgrenzungsmerkmals der Übermittlungstätigkeit bei autonomer KI ebenfalls unpassend erscheint. Letztlich kann aber dahingestellt bleiben, ob nicht die Zurechnung nach allgemeinen Rechtsscheingrundsätzen die dogmatisch sauberste Lösung wäre, da Einigkeit besteht, dass eine Zurechnung im Ergebnis zu erfolgen hat, solange KI aus rechtlichen Erwägungen heraus kein selbstständiger Vertragspartner sein kann.

3.6.1.2 Loslösung von mittels KI geschlossener Verträge

Logisches Gegenstück zum Vertragsschluss mittels KI ist die Loslösung von auf diese Art und Weise geschlossenen Verträgen. Wie zuvor ausgeführt, besteht zum aktuellen Zeitpunkt keine eindeutige bzw. rechtsdogmatisch einwandfreie Herleitung, wie genau Verträge durch bzw. mittels KI geschlossen werden können.

Vor dem Hintergrund des mit autonomen Systemen einhergehenden besonderen Autonomierisikos, d. h. der Gefahr, dass KI eigene Fehler unterlaufen, müssen die soeben ausgeführten Erwä-

gungen zum Vertragsschluss mittels KI jedenfalls entsprechend auch für die Loslösung von solchen Verträgen, insbesondere auf Irrtumsbasis, gelten.

Im Ergebnis folgerichtig und konsequent wird eine abstrakte Anfechtbarkeit gemäß §§ 119 ff BGB analog nie vollumfänglich verneint werden dürfen. Hierfür spricht neben dem Bedürfnis nach praxisnahen Ergebnissen auch der Gesetzgeberwille, der technologieunabhängig in § 122 BGB seinen Niederschlag im BGB findet, und auch heute im Zeitalter von KI gültige Wertungen und Rechtsfolgen wiedergibt. Auf dieser Grundlage wird die abstrakte Anfechtungsmöglichkeit jedenfalls nicht pauschal verneint werden können, soll doch gerade der Ersatz des negativen Interesses den Vertragspartner ausreichend schützen. Wie genau etwaigen Beweisschwierigkeiten im Bestreitensfall vor Gericht zu begegnen ist, wird auf Grundlage von § 286 ZPO dem Einzelfall vorbehalten bleiben.

3.6.2 Verträge (über den Einsatz von) KI

3.6.2.1 Gegenständliche Abgrenzung

Vom Vertragsschluss mittels KI bzw. der Loslösung von (irrtumsbedingt) mittels KI geschlossener Verträge sind Verträge mit inhaltlichem KI-Bezug abzugrenzen, die an dieser Stelle anhand von drei kurzen Gedanken exemplarisch thematisiert werden sollen.
Unter diesem Aspekt können Verträge über die Erstellung von KI grundsätzlich von Verträgen über die Verwendung von KI abgegrenzt werden. Bei Letzteren erscheint eine Abgrenzung zwischen Verträgen, die die bloße Zurverfügungstellung bzw. Verwendung eines KI-Tools regeln, von gewöhnlichen schuldrechtlichen Verträgen wie z. B. Dienst- und Werkverträgen tunlich, bei denen die Tätigkeit bzw. ein etwaig geschuldeter Erfolg unter Zuhilfenahme von KI-Komponenten erbracht bzw. erreicht wird.

3.6.2.2 Exemplarische Gedanken

Einen besonders praxisrelevanten Themenkomplex stellen potenzielle Haftungsrisiken bei der Zurverfügungstellung von KI-Tools dar. Diese kann jedoch hinreichend auf vertraglicher Basis begegnet werden, beispielsweise durch Formulierungen, die auf rechtlicher Grundlage ein ausreichendes Erwartungsmanagement gewährleisten. Hilfreich ist dabei die (mittlerweile standardmäßige) Verwendung von Disclaimern zum Output Generativer KI, insbesondere mit Blick auf die (weder gewährleistete noch garantierte) Fehlerfreiheit des Outputs; gepaart mit eigenen Überprüfungsverpflichtungen aufseiten der KI-Tool-Endnutzer.

Weiter relevant bleiben in diesem Kontext auch die Regelungen zum Vertragsgegenstand, insbesondere im Kontext anhängender Fragen zur rechtssicheren Lizenzierung. Nach übereinstimmender Auffassung können gewerbliche Schutzrechte am KI-Tool-Output nur durch aufwendige rechtliche Konstruktionen entstehen. Insbesondere mit Blick auf das Urheberrecht wird sich deshalb künftig zeigen müssen, ob und inwieweit die BGH-Rechtsprechung zu Nutzungs- und Lizenzverträgen – BGH I ZR 162/09 (Delcantos Hits) – übertragen können wird. Diesem Umstand

kann mit entsprechenden Klauseln begegnet werden, die eine etwaige Rechteeinräumung am KI-Tool-Output entweder pauschal ausschließen, oder nur soweit einräumen, wie Rechte auch tatsächlich entstehen – beispielsweise durch die Zurverfügungstellung des KI-Tool-Outputs »as is« oder der vertragsgegenständlichen Ermöglichung lediglich des faktischen Zugangs zum KI-Tool-Output ohne eigenständige Lizenzierung.

Werden bei gewöhnlichen schuldrechtlichen Verträgen KI-Tools lediglich verwendet, um z. B. Auswertungen zu erledigen, Prognosen aufzustellen oder Recherchen durchzuführen, so stellt sich die Frage, ob und inwieweit eine solche Verwendung unter den jeweiligen Abreden gestattet ist, etwaige Transparenzobliegenheiten bestehen, und letztlich welche rechtlichen Konsequenzen ein Unterlassen nach sich zieht.

Grundsätzlich gilt in Bezug auf die Arbeitserbringung das Prinzip der Höchstpersönlichkeit. Dies bedeutet, dass bestimmte Aufgaben und Dienstleistungen, die eine persönliche Leistung erfordern, nicht an eine KI delegiert werden können. Beispielsweise können medizinische oder rechtliche Beratungen, die individuelles Fachwissen und menschliches Urteilsvermögen erfordern, nicht ausschließlich bzw. im Schwerpunkt durch KI erfolgen. In diesen Bereichen kann KI zwar unterstützen; die endgültige Kontrolle, Entscheidung und Verantwortung müssen jedoch beim Menschen verbleiben.

Werden also beispielsweise auf werkvertraglicher Basis künstlerische Elemente in Form von Werken geschuldet, und diese unter Verwendung von KI-Tools dergestalt geschaffen, dass Arbeitsergebnissen am Maßstab des geltenden (Urheber-)Rechts die Schutzfähigkeit fehlt, so stehen dem Auftraggeber entweder Rückzahlungs- und/oder Schadenersatzansprüche zu. Kautelarjuristisch bedeutet dies, dass ein besonderer Fokus auf die exakte und ausführlich Beschreibung des Vertragsgegenstandes bzw. die Definition des Leistungsgegenstandes gelegt werden müssen; optional auch, dass etwaige Rechteübertragungen unter Inkaufnahme geringerer Erlöse vertraglich ausgeschlossen werden können.

4 Ethische Aspekte beim Einsatz von generativer KI

4.1 Einleitung: KI-Ethik und die Verantwortung in der Knowledge Society

In einer zunehmend datengetriebenen Welt entwickeln wir uns rasant zu einer Knowledge Society, in der Künstliche Intelligenz (KI) Schlüsselentscheidungen prägt – von der Personalauswahl bis zur Ressourcenverteilung. Der Deutsche Ethikrat betont in seiner Stellungnahme zur KI (2020), dass dieser Wandel ethische Leitplanken braucht, um Freiheit, Gerechtigkeit und gesellschaftliches Vertrauen zu wahren. Denn die Akzeptanz von KI hängt maßgeblich davon ab, ob Menschen ihren Systemen vertrauen können.

Dieses Vertrauen entsteht nicht von selbst, sondern durch Transparenz und Qualität: Nutzer:innen müssen nachvollziehen können, wie KI-Entscheidungen zustande kommen, und darauf vertrauen, dass Algorithmen robust, fair und zuverlässig arbeiten. Nur so lässt sich Responsible AI realisieren – ein Kernziel der KI-Ethik. Doch noch immer zeigen Negativbeispiele, was geschieht, wenn diese Prinzipien vernachlässigt werden.

So scheiterte beispielsweise KI-gestützte Recruitingtools stellenweise, weil sie Bewerberinnen systematisch benachteiligten, indem sie historisch männlich dominierte Daten reflektierten. In den Niederlanden führte ein fehlerhafter Algorithmus zur fälschlichen Kürzung von Sozialhilfen, wodurch Tausende Familien in existenzielle Not gerieten. Diese Fälle verdeutlichen: Ohne ethische Kontrolle riskiert KI nicht nur Diskriminierung, sondern auch gravierenden gesellschaftlichen Schaden.

Um KI verantwortungsvoll zu gestalten, braucht es daher mehr als technische Expertise – es erfordert einen ethischen Kompass, der Transparenz, Qualität und menschliche Werte ins Zentrum stellt. Nur so wird KI zum Instrument des Fortschritts, nicht der Spaltung.
Diese Entwicklung braucht insbesondere ein neues digitales Mindset in den Organisationen, mit dem in der Verbindung von Anwendungsregeln für GenAI und individuellen Fähigkeiten im Umgang mit GenAI eine veränderte Kultur der Verantwortung entsteht.

Welche Relevanz hat KI-Ethik für Unternehmen?
Im Folgenden stellen wir Aspekte der KI-Ethik dar und erläutern, dass es sich hierbei nicht um eine zwingende, gesetzliche Anforderung der KI-Verordnung handelt, sondern um eine freiwillige. Relevanz haben diese Aspekte bei der Auswahl von Lieferanten für KI-Technologien, beim Einkauf, bei der Anwendung und auch bei der Entwicklung von KI-Technologien. Gegebenenfalls ist es sinnvoll, eigene, auf das jeweilige Unternehmen und Geschäftsmodell angepasste Aspekte der KI-Ethik in ein Regelwerk zu gießen, beispielsweise als Leitlinien oder Richtlinien für KI-Ethik oder übergreifende für digitale Ethik.
Doch was genau bedeutet »KI-Ethik«?

4.2 Begriffsbestimmung und Abgrenzung: KI-Ethik

KI-Ethik bezeichnet die ethische Reflexion und Gestaltung von Künstlicher Intelligenz. Sie umfasst:

a. die philosophische Disziplin, die moralische Prinzipien, Werte und gesellschaftliche Auswirkungen von KI-Systemen analysiert (analog zur klassischen Ethik als »Sittenlehre«),

b. die konkreten normativen Standards, die verantwortungsvolle Entwicklung, Anwendung und Regulierung von KI-Systemen leiten.

c. KI-Ethik zielt darauf ab, technologische Innovation mit sozialen Wertvorstellungen in Einklang zu bringen, wie Gerechtigkeit, Autonomie und Transparenz. Sie adressiert Fragen wie: Welche Entscheidungen dürfen Algorithmen treffen? Wie lassen sich Diskriminierung oder Schäden verhindern? Was verändert sich durch die Verwendung von KI in meinen eigenen Entscheidungen?

4.2.1 Abgrenzung KI-System vs. »Intelligenz«

Der Begriff »Künstliche Intelligenz« ist irreführend, da es sich nicht um menschenähnliche Denkfähigkeiten handelt, sondern um mathematische Methoden (z. B. Backpropagation in neuronalen Netzen), die auf Wahrscheinlichkeitsberechnungen und Datenmustern basieren. Die EU-KI-Verordnung (Art. 3 Nr. 1) definiert daher pragmatisch ein KI-System als:
»maschinengestütztes System [...], das für explizite oder implizite Ziele Ausgaben wie Vorhersagen oder Entscheidungen generiert, die physische oder virtuelle Umgebungen beeinflussen können.«

Diese Definition – angelehnt an die OECD – betont die Funktionalität (Entscheidungsautonomie, Adaptionsfähigkeit) und nicht abstrakte »Intelligenz«. KI-Ethik fokussiert sich somit auf die Wirkung dieser Systeme, nicht auf deren metaphorische Bezeichnung.

4.2.2 Abgrenzung zu verwandten Konzepten

Responsible AI (Verantwortungsvolle KI): Bezeichnet die praktische Umsetzung ethischer Prinzipien in KI-Projekten, z. B. durch risikobasierte Regulierung, Qualitätssicherung oder Rechenschaftspflichten. Während KI-Ethik das theoretische Fundament liefert, ist Responsible AI die angewandte Praxis.

Trustworthy AI (Vertrauenswürdige KI): Beschreibt Eigenschaften wie Transparenz, Robustheit und Fairness, die Nutzer:innen Vertrauen in KI-Systeme ermöglichen. Dies ist ein Teilaspekt von KI-Ethik, der jedoch stärker technische Zuverlässigkeit betont.

KI-Ethik vs. technikzentrierte Perspektiven: KI-Ethik geht über rein technische Lösungen hinaus: Während Algorithmen oft als »neutrale Tools« wahrgenommen werden, zeigt die Praxis (z. B. in Teilen diskriminierende Recruiting-Tools, fehlerhafte Sozialhilfe-Algorithmen in den Niederlanden), dass KI-Systeme gesellschaftliche Pfadabhängigkeiten, Machtstrukturen und Vorurteile reproduzieren können. KI-Ethik fordert daher eine ganzheitliche Betrachtung – von der Datenauswahl über Modelltraining bis zur Implementierung –, um verantwortungsvolle Technikgestaltung zu gewährleisten.

KI-Ethik ist die ethische Richtschnur für den Umgang mit KI-Systemen. Sie verbindet philosophische Reflexion mit konkreten Handlungsmaximen, um Technik in den Dienst menschlicher Werte zu stellen – und Risiken wie Bias oder Intransparenz aktiv zu begrenzen.

4.3 Konkreter Kontext des Einsatzes generativer KI und Unterscheidung ethischer Einsatzszenarien für generative KI

4.3.1 Kundenumgang und Kundenerlebnis (UX, CX, HCX)

Digitale Ethik in GenAI wirkt als integrierter Code, der die Entwicklung digitaler Produkte und Dienstleistungen mit den Erwartungen und Wertvorstellungen der Kunden verknüpft. Ethische Aspekte werden so bereits im Designprozess berücksichtigt, anstatt sie nachträglich aufzusetzen.

a. Digitale Ethik ermöglicht es Unternehmen, transparente und vertrauenswürdige Interaktionen mit Kunden zu gestalten, indem sie sicherstellt, dass digitale Produkte von Anfang an ethische Standards erfüllen (z. B. Datenschutz, Fairness, Vermeidung von Manipulation).

b. Sie schafft ein positives Kundenerlebnis, indem sie Erwartungen der Kunden an Privatsphäre, Sicherheit und Nutzerfreundlichkeit in die Produktentwicklung integriert.

c. Ethics on the Engine: Digitale Ethik in der UX wird als Teil des agilen Entwicklungsprozesses verstanden, der kontinuierlich weiterentwickelt wird.

d. Transparenz und Kommunikation: Unternehmen nutzen digitale Ethik, um Kunden klar zu vermitteln, wie ihre Daten verwendet werden und welche Werte hinter den Produkten stehen.

e. Werte-Kompass: Unternehmen entwickeln einen eigenen Wertekompass, der ihre Produkte und Dienstleistungen prägt und so das Vertrauen der Kunden stärkt und Grundlage ihrer zukünftigen Produktivität wird.

4.3.2 Mitarbeiterhilfen (EX) und Prozessunterstützung

Digitale Ethik unterstützt Mitarbeiter dabei, ethische Aspekte in ihre tägliche Arbeit zu integrieren und fördert so ein wertorientiertes Mindset.

a. Digitale Ethik hilft Mitarbeitenden, ethische Dilemmata in der digitalen Transformation zu erkennen und zu lösen, indem sie klare Handlungsroutinen und Werte vorgibt.

b. Sie stärkt die digitale Awareness der Mitarbeitenden, sodass sie ethische Fragen in ihre Arbeit einbeziehen können (z. B. bei der Entwicklung neuer Features oder Prozesse).

c. Ethics on the Engine: Ethik wird in den täglichen Arbeitsablauf integriert, z. B. durch regelmäßige Updates in Scrum-Meetings oder die Visualisierung ethischer Fragestellungen auf Kanban-Boards.

d. Interne Kommunikation: Mitarbeiter tauschen sich kontinuierlich über ethische Aspekte aus, z. B. zu Geschwindigkeit, Transparenz oder Komplexität, und entwickeln so ein gemeinsames Verständnis.

e. Werteorientierte Handlungsroutinen: Unternehmen entwickeln interne Prozesse, die ethische Überlegungen direkt in die Entscheidungsfindung einbeziehen.

4.3.3 Administrative Verwaltungsaufgaben mit GenAI: Technische Organisation, HR, Betriebsrat, AI Agents (»AI Agents Ethics«)

Digitale Ethik wird hier als strategisches Element eingesetzt, um administrative Prozesse ethisch zu gestalten und die Verantwortung für KI-Systeme zu übernehmen.

a. Sie sorgt dafür, dass KI-Systeme und KI-Agenten ethisch vertretbar agieren, z. B. durch die Vermeidung von Bias, die Sicherstellung von Datenschutz und die Einhaltung von Compliance-Vorgaben.

b. Sie unterstützt HR und Betriebsrat dabei, ethische Standards in der Personalarbeit und Mitbestimmung zu verankern, z. B. bei der Einführung neuer Technologien oder der Gestaltung von Arbeitsbedingungen.

c. Ethics on the Engine: Ethik wird in die Entwicklung und den Betrieb von KI-Systemen integriert, sodass ethische Aspekte von Anfang an berücksichtigt werden.

d. Strukturierte Kommunikation: Unternehmen schaffen klare Kommunikationswege, um ethische Fragen im Zusammenhang mit KI und Datenverarbeitung zu diskutieren und zu lösen.

e. Leitbilder und Standards: Unternehmen entwickeln interne Leitlinien, die ethische Prinzipien in der technischen Organisation und im Umgang mit KI-Systemen verankern.

Digitale Ethik steuert in allen drei Bereichen den integrativen und dynamischen Prozess, der ethische Aspekte direkt in die Geschäftsabläufe einbindet. Sie nutzt ein ethisches Mindset (Ethics on the Engine) als zentrales Modell, um Ethik nicht als nachträgliche Korrektur, sondern als fundamentalen Bestandteil der digitalen Transformation zu verstehen. Durch transparente Kommunikation, werteorientierte Handlungsroutinen und die Integration ethischer Standards in die Produktentwicklung und Prozessgestaltung schafft sie Vertrauen, fördert Innovation und sichert die langfristige Akzeptanz digitaler Lösungen.

4.4 Organisatorische Anforderungen (»Governance«)

Mit der KI-Verordnung stellt sich die grundsätzliche Frage, wer im Unternehmen die Einhaltung dieser Verordnung organisieren und sicherstellen soll.
Zwei Grundsätze haben sich hierbei als Best Practices etabliert:

a. Die Fachabteilungen im Unternehmen, die KI einsetzen, sorgen selbst für die Einhaltung der Anforderungen und werden manchmal durch Spezialisten hierbei unterstützt, z. B. auch durch die Nutzung eines unternehmensweiten Tools zur Erfassung und Risikoklassifizierung von KI und deren Anwendungsfälle; oder

b. Ein bestehendes Team der Governance-Organisation kümmert sich um die Einhaltung der gesetzlichen Anforderungen, beispielsweise das Datenschutz- oder Daten-Team.

Auch der Bereich KI-Ethik wirft die Frage auf, wer im Unternehmen für diese Ethik verantwortlich sein soll und welche Funktionen hierbei welche Rollen haben.
In Betracht kommt, die KI-Ethik im Bereich der oben dargestellten »AI Governance« anzusiedeln oder eine separate, interdisziplinäre Gruppe von Mitarbeitenden hiermit zu betrauen, beispielsweise als »AI Ethics Board« oder »Ausschuss für KI-Ethik«.
Die konkrete Ausgestaltung hängt von den Gegebenheiten des jeweiligen Unternehmens ab.

4.5 Anwendungsregeln für generative KI

Die Funktionen und Entscheidungsprozesse von generativer künstlicher Intelligenz (»GenAI«) sollten klar und verständlich sein, um Vertrauen zu schaffen und Nachvollziehbarkeit zu ermöglichen. Hier ist es sehr wichtig, die unterschiedlichen Einsatzszenarien zu beachten und transparent zu kommunizieren.
Der Umgang mit Transparenz und Nachvollziehbarkeit bei der Entwicklung und dem Einsatz von GenAI erfordert, dass GenAI-Systeme nicht nur effektiv, sondern auch ethisch und verständlich gestaltet werden, um Akzeptanz und Vertrauen in der Gesellschaft zu fördern.

4.5.1 Klare Funktionen und Entscheidungsprozesse

Die Art und Weise, wie GenAI-Systeme funktionieren und Entscheidungen treffen, sollte für Nutzer und Stakeholder verständlich sein. Das bedeutet, dass die zugrunde liegenden Mechanismen, Algorithmen und Datenquellen offengelegt werden sollten, um Misstrauen zu vermeiden.

GDPR (DSGVO):

- Art. 13 und 14: Informationspflichten bei der Erhebung personenbezogener Daten, einschließlich der Zwecke der Verarbeitung und der Logik automatisierter Entscheidungen.
- Art. 22: Recht auf Erklärung bei automatisierten Einzelentscheidungen, die erhebliche Auswirkungen haben.

AI Act (KI-Verordnung):

- Art. 13: Transparenzpflichten für KI-Systeme, insbesondere bei Interaktion mit Menschen (z. B. Chatbots).
- Art. 14: Dokumentations- und Informationspflichten für Hochrisiko-KI-Systeme.

Data Act:

- Art. 13: Pflicht zur Bereitstellung von Informationen über die Funktionsweise von Datenverarbeitungsdiensten.

4.5.2 Vertrauen schaffen

Transparenz ist entscheidend, um das Vertrauen der Nutzer in die Technologie zu stärken. Wenn Menschen verstehen, wie und warum eine Entscheidung getroffen wurde, sind sie eher bereit, die Technologie zu akzeptieren und zu nutzen.

GDPR (DSGVO):

- Art. 5: Grundsätze der Datenverarbeitung, darunter Transparenz, Zweckbindung und Rechenschaftspflicht.

Digital Services Act (DSA):

- Art. 14: Transparenzpflichten für Online-Plattformen, um Vertrauen der Nutzerinnen und Nutzer zu stärken.
- Art. 24: Risikomanagement und Transparenz bei algorithmischen Systemen.

AI Act (KI-Verordnung):

- Art. 29: Verpflichtung zur Risikobewertung und Sicherstellung der Vertrauenswürdigkeit von KI-Systemen.

4.5.3 Nachvollziehbarkeit ermöglichen

Entscheidungen, die von GenAI getroffen werden, sollten nachvollziehbar sein. Das bedeutet, dass es möglich sein sollte, den Prozess zu überprüfen und zu erklären, wie ein bestimmtes Ergebnis zustande gekommen ist. Dies ist besonders wichtig in sensiblen Bereichen wie Medizin, Recht oder Finanzen.

GDPR (DSGVO):

- Art. 15: Recht auf Auskunft über die Verarbeitung personenbezogener Daten.
- Art. 22(3): Recht auf menschliches Eingreifen und Erklärung bei automatisierten Entscheidungen.

AI Act (KI-Verordnung):

- Art. 13(1): Pflicht zur Bereitstellung von Informationen, die es Nutzern ermöglichen, die Entscheidungen von KI-Systemen zu verstehen.
- Art. 14(4): Dokumentation der Entscheidungsprozesse bei Hochrisiko-KI-Systemen.

Data Act:

- Art. 8: Recht auf Zugang zu Daten, die von einem Produkt oder einer Dienstleistung generiert wurden, um Transparenz zu gewährleisten.

4.5.4 Berücksichtigung unterschiedlicher Einsatzszenarien

GenAI wird in verschiedenen Kontexten eingesetzt, z. B. in der Bildung, im Gesundheitswesen oder in der Unterhaltungsindustrie. Jedes Szenario hat spezifische Anforderungen und ethische Implikationen, die berücksichtigt werden müssen. Die Technologie sollte an diese unterschiedlichen Bedürfnisse angepasst werden.

AI Act (KI-Verordnung):

- Anhang III: Liste der Hochrisiko-KI-Systeme, die je nach Einsatzgebiet (z. B. Medizin, Justiz, kritische Infrastruktur) unterschiedliche Anforderungen erfüllen müssen.
- Art. 5: Verbot bestimmter KI-Praktiken, die inakzeptable Risiken darstellen, abhängig vom Anwendungsfall.

Digital Markets Act (DMA):

- Art. 6 und 7: Spezifische Pflichten für Gatekeeper-Plattformen, um faire Wettbewerbsbedingungen in verschiedenen Märkten zu gewährleisten.

Digital Operational Resilience Act (DORA):

- Art. 5: Anforderungen an die Widerstandsfähigkeit von Finanzinstituten, abhängig von ihrer Größe und ihrem Geschäftsmodell.

4.5.5 Transparente Kommunikation

Es ist wichtig, offen und klar über die Fähigkeiten, Grenzen und möglichen Risiken von GenAI zu kommunizieren. Dies hilft, unrealistische Erwartungen zu vermeiden und fördert eine verantwortungsvolle Nutzung.

GDPR (DSGVO):

- Art. 12: Pflicht zur klaren und verständlichen Kommunikation mit Betroffenen.
- Digital Services Act (DSA):
- Art. 14: Transparenzberichte für Online-Plattformen.
- Art. 34: Pflicht zur Offenlegung von Algorithmen, die für Empfehlungssysteme ver-wendet werden.

AI Act (KI-Verordnung):

- Art. 52: Pflicht zur Kennzeichnung von KI-generierten Inhalten (z. B. Deepfakes).
- Art. 13(2): Verpflichtung zur Offenlegung, wenn Nutzer mit einem KI-System interagieren.

Data Governance Act (DGA):

- Art. 5: Transparenzpflichten für Datenvermittlungsdienste.

4.5.6 Umgang mit Komplexität und Verantwortung

Ein neuer Umgang mit Komplexität muss in den Organisationen geregelt werden, da sich durch den Einsatz von GenAI die Möglichkeiten verschieben, Prozesse rational zu gestalten und in sie einzugreifen.

Die menschliche Verantwortlichkeit ist als Grundprinzip ein wichtiger Aspekt. Entscheidungen, die von KI-Systemen getroffen werden, müssen stets von Menschen überwacht und verantwortet werden (»human in the loop« oder »human oversight«). Dies ist bei generativer KI nicht immer möglich. Hier muss die Verantwortung zu einer transparenten, prozessorientierten Verantwortung übergehen, die nicht mehr den Einzelfall überprüfen muss.

Mitarbeiter sollten die Funktionen und Entscheidungen von KI-Systemen verstehen und kritisch hinterfragen können. Schulungen und Weiterbildungen sind hierfür unerlässlich. Aber auch die Einführung von Eingriffsmöglichkeiten und deren Regelungen.

4.6 Inhaltliche Anforderungen an KI-Ethik und ethische Bereiche

Der Einsatz von Generativer Künstlicher Intelligenz (»GenAI«) erfordert eine klare ethische Richtschnur, die technologische Möglichkeiten mit menschlichen Werten, rechtlichen Grenzen und gesellschaftlichen Ansprüchen in Einklang bringt. Im Spannungsfeld zwischen Innovation und Verantwortung sind folgende inhaltliche Anforderungen zentral:

4.6.1 Ethische Kernprinzipien für GenAI

Transparenz & Erklärbarkeit:

GenAI-Systeme (z. B. ChatGPT, DALL-E) müssen nachvollziehbar machen, wie Inhalte entstehen. Nutzerinnen und Nutzer haben ein Recht zu wissen, ob sie mit KI-generierten Texten, Bildern oder Entscheidungen interagieren.
Beispiel: Kennzeichnungspflicht für KI-generierte Medien, um Desinformation vorzubeugen.

Fairness & Nichtdiskriminierung:

Algorithmen dürfen keine historischen oder kulturellen Vorurteile reproduzieren (z. B. genderspezifische Sprachmuster in Recruiting-Tools).
Beispiel: Diversitätschecks bei Trainingsdaten für GenAI-Modelle.

Menschliche Autonomie:

GenAI darf keine Entscheidungsmacht über Menschen ausüben, besonders in sensiblen Bereichen wie Medizin, Recht oder Bildung.
Beispiel: KI-gestützte Diagnosetools als unterstützende, nicht ersetzende Systeme.

Datenschutz & Privatsphäre:

GenAI muss die Privatsphäre wahren – besonders, wenn persönliche Daten verarbeitet werden (DSGVO-Konformität).
Beispiel: Anonymisierungstechniken bei der Nutzung von Kundenfeedback für KI-Training.

4.6.2 Regelungsgrundlagen

Die KI-Verordnung selbst enthält keine Anforderungen an KI-Ethik, verweist aber in Artikel 95 Absatz 2 a) und den Erwägungsgründen 7, 27 und 165 auf die »Ethikleitlinien für eine vertrauenswürdige künstliche Intelligenz«, die am 8. April 2019 eine hochrangige Expertengruppe der Europäischen Kommission vorgelegt hat.

Danach zeichnet sich eine vertrauenswürdige KI durch drei Komponenten aus, die während des gesamten Lebenszyklus des Systems erfüllt sein sollten:

a) Sie sollte rechtmäßig sein und somit alle anwendbaren Gesetze und Bestimmungen einhalten, b) sie sollte ethisch sein und somit die Einhaltung ethischer Grundsätze und Werte garantieren und c) sie sollte robust sein, und zwar sowohl in technischer als auch sozialer Hinsicht, da KI-Systeme selbst bei guten Absichten unbeabsichtigten Schaden anrichten können.

Die Leitlinien erläutern dann sieben Anforderungen an KI-Systeme, wie sich eine vertrauenswürdige KI realisieren lässt.

Es muss gewährleistet sein, dass die Entwicklung, Einführung und Nutzung von KI-Systemen die Anforderungen an vertrauenswürdige KI erfüllen: 1) Vorrang menschlichen Handelns und menschlicher Aufsicht, 2) technische Robustheit und Sicherheit, 3) Schutz der Privatsphäre und Datenqualitätsmanagement, 4) Transparenz, 5) Vielfalt, Nichtdiskriminierung und Fairness, 6) gesellschaftliches und ökologisches Wohlergehen sowie 7) Rechenschaftspflicht.

Quellen:

Überblick: Ethics guidelines for trustworthy AI | Shaping Europe's digital future, ↗https://digital-strategy.ec.europa.eu/en/library/ethics-guidelines-trustworthy-ai

Dokument: https://ec.europa.eu/newsroom/dae/document.cfm?doc_id=60425

Erwägungsgrund 27 der KI-Verordnung greift diese Anforderungen auf und sagt:

»Während der risikobasierte Ansatz die Grundlage für ein verhältnismäßiges und wirksames verbindliches Regelwerk bildet, muss auch auf die Ethikleitlinien für vertrauenswürdige KI von 2019 der von der Kommission eingesetzten unabhängigen hochrangigen Expertengruppe für künstliche Intelligenz verwiesen werden. In diesen Leitlinien hat die hochrangige Expertengruppe sieben unverbindliche ethische Grundsätze für KI entwickelt, die dazu beitragen sollten, dass KI vertrauenswürdig und ethisch vertretbar ist. Zu den sieben Grundsätzen gehören: menschliches Handeln und menschliche Aufsicht, technische Robustheit und Sicherheit, Privatsphäre und Daten-Governance, Transparenz, Vielfalt, Nichtdiskriminierung und Fairness, soziales und ökologisches Wohlergehen sowie Rechenschaftspflicht. Unbeschadet der rechtsverbindlichen Anforderungen dieser Verordnung und anderer geltender Rechtsvorschriften der Union tragen diese Leitlinien zur Gestaltung kohärenter, vertrauenswürdiger und menschenzentrierter KI bei, im Einklang mit der Charta und den Werten, auf die sich die Union gründet.«

Quelle: ↗https://eur-lex.europa.eu/legal-content/DE/TXT/HTML/?uri=OJ:L_202401689#rct_27

Erwägungsgrund 165 der KI-Verordnung ergänzt hierzu:

»Darüber hinaus sollten die Anbieter und gegebenenfalls die Betreiber aller KI-Systeme, ob mit hohem Risiko oder nicht, und aller KI-Modelle auch ermutigt werden, freiwillig zusätzliche Anfor-

derungen anzuwenden, z. B. in Bezug auf die Elemente der Ethikleitlinien der Union für vertrauenswürdige KI, die ökologische Nachhaltigkeit, Maßnahmen für KI-Kompetenz, die inklusive und vielfältige Gestaltung und Entwicklung von KI-Systemen, unter anderem mit Schwerpunkt auf schutzbedürftige Personen und die Barrierefreiheit für Menschen mit Behinderungen, die Beteiligung der Interessenträger, gegebenenfalls mit Einbindung einschlägiger Interessenträger wie Unternehmensverbänden und Organisationen der Zivilgesellschaft, Wissenschaft, Forschungsorganisationen, Gewerkschaften und Verbraucherschutzorganisationen an der Konzeption und Entwicklung von KI-Systemen und die Vielfalt der Entwicklungsteams, einschließlich einer ausgewogenen Vertretung der Geschlechter.«

Quelle: ↗https://eur-lex.europa.eu/legal-content/DE/TXT/HTML/?uri=OJ:L_202401689#rct_165

Damit stellt die KI-Verordnung klar, dass es sich bei ethischen Aspekten um freiwillige Anforderungen handelt, die das Unternehmen gegebenenfalls aufstellen kann, und nicht um eine formale gesetzliche Anforderung. Insbesondere dann, wenn ein Unternehmen umfangreich KI-Technologien einsetzt oder KI vielleicht sogar selbst entwickelt, wird empfohlen, sich mit Aspekten der KI-Ethik zu beschäftigen und ein Dokument hierzu zu erstellen. Weitere Details hierzu weiter unten in diesem Kapitel im Abschnitt »Konkrete Umsetzung in Unternehmen«.

Der Deutsche Ethikrat hat am 20.03.2023 eine Stellungnahme mit dem Titel »Mensch und Maschine – Herausforderungen durch Künstliche Intelligenz« vorgestellt: KI und Robotik | Deutscher Ethikrat, ↗https://www.ethikrat.org/themen/forschung-und-technik/ki-robotik/.

Der Deutsche Ethikrat betrachtet hierbei verschiedene Anwendungsbereiche von KI und spricht ethische Querschnittsthemen an. In seiner Stellungnahme untersucht der Deutsche Ethikrat umfassend die Auswirkungen digitaler Technologien auf das menschliche Selbstverständnis und Miteinander. Die philosophisch grundierten Überlegungen und Empfehlungen geben Orientierung zu den rasanten Innovationen im Bereich der Künstlichen Intelligenz. Eine Kurzfassung in Form eines Flyers ist hier abrufbar: ↗https://www.ethikrat.org/fileadmin/Publikationen/Stellungnahmen/deutsch/flyer-mensch-und-maschine.pdf

4.6.3 Andere Quellen

Auch weitere Institutionen und supranationale Organisationen beschäftigen sich mit dem Thema, unter anderem:
IEEE, IEEE SA – IEEE Introduces New Program for Free Access to AI Ethics and Governance Standards, ↗https://standards.ieee.org/news/get-program-ai-ethics/

Responsible AI Institute, Home – Responsible AI, ↗https://www.responsible.ai/
Die OECD hat bereits 2019 »AI Principles« aufgestellt, die im Mai 2024 überarbeitet wurden und auf die sich einige Länder verpflichtet haben: AI Principles Overview – OECD.AI, ↗https://oecd.ai/en/ai-principles

Die UNO hat sich »Principles for the Ethical Use of Artificial Intelligence in the United Nations System« gegeben, Principles for the Ethical Use of Artificial Intelligence in the United Nations System | United Nations – CEB, ↗https://unsceb.org/principles-ethical-use-artificial-intelligence-united-nations-system, und die UNESCO hat bereits 2022 Empfehlungen erarbeitet, die 4 »core values« und 10 »principles« enthalten: Recommendation on the Ethics of Artificial Intelligence | UNESCO, ↗https://www.unesco.org/en/articles/recommendation-ethics-artificial-intelligence.

4.7 Exkurs: KI-Ethik in der EU und global: Der Kampf gegen Algorithmic Bias

Algorithmischer Bias – die systematische Verzerrung von KI-Ergebnissen aufgrund unausgewogener Daten, voreingenommener Modelle oder mangelnder Diversität in Entwicklungsprozessen – ist eine der größten ethischen Herausforderungen im Umgang mit KI. Sowohl die EU als auch globale Akteure setzen zunehmend auf regulatorische und technische Kriterien, um Diskriminierung durch KI-Systeme zu bekämpfen.

4.7.1 Die EU: Regulatorische Vorreiterrolle

Die EU adressiert Algorithmic Bias durch die KI-Verordnung, den weltweit ersten umfassenden Rechtsrahmen für KI. Zentrale Kriterien zur Vermeidung von Bias umfassen:

a. Risikobasierte Klassifizierung: Hochrisiko-KI-Systeme (z. B. in Personalwesen, Justiz oder Sozialleistungen) müssen strenge Anforderungen erfüllen,

b. Datenqualität: Repräsentative Trainingsdaten, um historische Diskriminierung (z. B. Gender-Bias in Recruiting-Algorithmen zu vermeiden.

c. Impact Assessments: Vorabprüfungen zu möglichen Verzerrungen und deren gesellschaftlichen Folgen.

d. Transparenzpflichten: Dokumentation der Entscheidungslogik für Aufsichtsbehörden (»Explainable AI«).

e. Verbote diskriminierender Anwendungen: KI-Systeme, die sozialen Schaden verursachen (wie der niederländische SyRI-Algorithmus, der fälschlich Sozialbetrug unterstellte), werden untersagt.

f. Ethikgremien und Zertifizierung: Der EU-Ansatz setzt auf unabhängige Prüfstellen und standardisierte Zertifizierungen, um Bias zu identifizieren – ein Modell, das global als Benchmark gilt.

4.7.2 Globale Initiativen: Fragmentierte Ansätze mit gemeinsamen Zielen

Während die EU auf verbindliche Regeln setzt, dominieren global eher soft-law-Initiativen:

a. OECD-Prinzipien für vertrauenswürdige KI (2019): Sie fordern »Inklusivität« und »Fairness« als Kernkriterien, allerdings ohne rechtliche Bindung.

b. UNESCO-Empfehlungen zu KI-Ethik (2021): Betonen die Pflicht von Staaten, »algorithmische Diskriminierung« zu überwachen, besonders im Globalen Süden, wo mangelnde Infrastruktur Bias verstärkt (z. B. fehlende Sprachdaten für afrikanische Dialekte).

c. US-Initiativen: Die »Blueprint for an AI Bill of Rights« (2022) definiert ähnliche Kriterien wie die EU, bleibt aber unverbindlich. Ausgewählte Tech-Konzerne setzen stattdessen auf interne Ethikboards – kritisiert als »Selbstregulierung ohne Zähne«.

4.7.3 Herausforderungen und kritische Perspektiven

Trotz Fortschritte bleiben zentrale Probleme:

a. Datenkolonialismus: Globale KI-Systeme nutzen oft Daten aus dem Globalen Norden, was zu Bias im Globalen Süden führt (z. B. Gesichtserkennung, die dunkle Hauttypen schlechter erkennt).

b. Kulturrelativismus: Was in der EU als »Bias« gilt (z. B. Gender-Kategorien), ist in anderen Ländern tabu (z. B. KI-basierte LGBTQ+-Identifikation in repressiven Staaten).

c. Technische Limits: Selbst »bias-korrigierte« Modelle scheitern oft an komplexen Realitäten (z. B. KI in der US-Gesundheitsversorgung, die dunkelhäutige Patientinnen und Patienten systematisch unterversorgte).

4.7.4 Fazit: Von Kriterien zu verbindlicher Praxis

Algorithmic Bias ist kein rein technisches Problem, sondern ein Spiegel gesellschaftlicher Machtverhältnisse. Die EU zeigt, wie regulatorische Kriterien Wirkung entfalten können – doch global braucht es stärkere Kooperation, um ethische Mindeststandards durchzusetzen. Initiativen wie die Global Partnership on AI (GPAI) oder die UN-Tech-Beratungsgremien könnten hier Brücken bauen. Letztlich entscheidet sich die Zukunft der KI-Ethik daran, ob »Bias-Kriterien« nicht nur auf dem Papier stehen, sondern Macht- und Ressourcenungleichheiten aktiv abbauen.

4.8 Exkurs: Ethische Verantwortung für psychologische Zusammenhänge

Im Rahmen des Einsatzes generativer KI im Umgang mit Kunden und Mitarbeitern (UX, CX, HCX, EX) werden immer mehr emotionale und verhaltensorientierte Elemente eine Rolle spielen, die das Gewicht der Handlungen eindeutig auf die Seite der generativen KI schieben.

4.8.1 Nudging

Das unumgängliche Nudging bildet dabei eine große Herausforderung, um sich darin zu orientieren. Den involvierten Personengruppen muss es ermöglicht werden, innerhalb dieser technischen Umstände zu navigieren und eigenverantwortlich zu handeln.
Der Einsatz von Generativer KI (GenAI) in der Kommunikation mit Kunden und Mitarbeitenden führt zu einer neuen Form des »Micro-Nudging«. Dieses Phänomen beschreibt, wie GenAI die Opportunitäten und Toleranzen des Gegenübers – sei es eine Kundin bzw. ein Kunde oder eine Mitarbeiterin bzw. ein Mitarbeiter – nutzen kann, um bestimmte Verhaltensweisen oder Entscheidungen zu beeinflussen. Dies geschieht oft subtil und in Echtzeit, was sowohl Chancen als auch Risiken birgt. Hier eine detaillierte Betrachtung:

4.8.2 Was ist Micro-Nudging durch GenAI?

Micro-Nudging bezieht sich auf die Fähigkeit von GenAI, kleine, gezielte Impulse zu setzen, die das Verhalten oder die Entscheidungen des Nutzers beeinflussen. Diese Impulse können auf der Grundlage von Echtzeitdaten und Verhaltensanalysen erfolgen, wodurch die KI in der Lage ist, die individuellen Präferenzen, Schwächen und Toleranzen der Nutzenden zu erkennen und auszunutzen.

a. Beispiel im Kundenumgang:
   Ein Chatbot könnte durch gezielte Formulierungen oder Angebote den Kunden dazu bewegen, ein teureres Produkt zu wählen oder eine Entscheidung schneller zu treffen, als er es ursprünglich geplant hatte.

b. Beispiel in der Mitarbeiterkommunikation:
   Ein KI-gestütztes Tool könnte Mitarbeiter dazu anregen, bestimmte Aufgaben schneller zu erledigen, indem es Belohnungen oder soziale Anerkennung in Aussicht stellt (z. B. durch Gamification-Elemente).

4.8.3 Wie nutzt GenAI Opportunismus und Toleranzen aus?

GenAI ist in der Lage, individuelle Schwächen und Toleranzen zu identifizieren und gezielt auszunutzen. Dies geschieht durch:

a. Echtzeitanalyse:
   Die KI analysiert das Verhalten des Nutzers in Echtzeit (z. B. durch Klickverhalten, Antwortzeiten oder Sprachmuster) und passt ihre Kommunikation dynamisch an.

b. Personalisierung:
   GenAI nutzt gesammelte Daten, um maßgeschneiderte Botschaften zu erstellen, die auf die individuellen Bedürfnisse und Schwächen des Nutzers zugeschnitten sind.

4.8.4 Psychologische Manipulation:

Die KI setzt gezielt psychologische Hebel ein, wie z. B. soziale Bewährtheit (»Andere haben das auch getan«), Autorität (»Experten empfehlen …«) oder Dringlichkeit (»Angebot endet bald«).

4.8.5 Veränderung der Entscheidungsräume

Objektiv führt der Einsatz von GenAI zu einer statistischen Verbesserung der Entscheidungen, subjektiv jedoch verändern sich die Entscheidungsräume bei den beteiligten Menschen und sie müssen neu lernen, mit GenAI zu navigieren und darüber zu kommunizieren.

a. Unterstützung bei komplexen Entscheidungen: GenAI kann große Datenmengen analysieren und Handlungsempfehlungen geben, was Mitarbeitern hilft, fundiertere Entscheidungen zu treffen. Dies ist besonders in Bereichen wie Datenanalyse, Prognose und Prozessoptimierung relevant.

b. Automatisierung von Routineaufgaben: Durch die Übernahme repetitiver Aufgaben schafft GenAI Freiräume für Mitarbeiter, sich auf kreativere und strategischere Tätigkeiten zu konzentrieren.

4.8.6 Einschränkung der Autonomie

a. Abhängigkeit von KI-Empfehlungen: Mitarbeiter könnten sich zunehmend auf die Vorschläge der KI verlassen, was zu einer Reduzierung der eigenen Urteilsfähigkeit führen kann. Dies beeinträchtigt die kritische Reflexion und die eigenständige Entscheidungsfindung.

b. Verlust von Kontrolle: Wenn GenAI Entscheidungen teilweise oder vollständig übernimmt, erzeugt dies bei Mitarbeitern das Gefühl, die Kontrolle über ihre Arbeit zu verlieren, was zu Unsicherheit oder Frustration führen kann.

4.8.7 Personalisierte Empfehlungen

a. Micro-Nudging: GenAI kann durch Echtzeitdatenanalyse und Verhaltensvorhersagen personalisierte Empfehlungen geben, die Kunden in ihren Entscheidungen beeinflussen. Dies führt zu einer höheren Zufriedenheit, da Kunden Angebote erhalten, die besser auf ihre Bedürfnisse zugeschnitten sind.

b. Manipulationsrisiko: Gleichzeitig besteht die Gefahr, dass Kunden durch gezielte psychologische Manipulation (z. B. durch Betonung von Verlustaversion oder sozialem Druck) in ihrer Entscheidungsfreiheit eingeschränkt werden. Dies führt zu einem Gefühl der Fremdbestimmung.

4.8.8 Transparenz und Kontrolle

a. Transparenzdefizite: Kunden sind sich oft nicht bewusst, wie ihre Daten genutzt werden, um Entscheidungen zu beeinflussen. Dies führt zu Misstrauen gegenüber Unternehmen, die GenAI einsetzen.

b. Verlust der Autonomie: Wenn Kunden das Gefühl haben, dass ihre Entscheidungen von der KI gesteuert werden, untergräbt dies ihre Autonomie und ihr Vertrauen in die Marke.

4.8.9 Psychologische Hintergründe für Mitarbeiter

a. Verlust von Sinnhaftigkeit: Wenn Mitarbeiter das Gefühl haben, dass ihre Arbeit zunehmend von KI übernommen wird, kann dies zu einem Verlust des Sinns in ihrer Tätigkeit führen. Dies beeinträchtigt die Motivation und Zufriedenheit.

b. Ethischer Druck: Mitarbeiter können unter ethischem Druck stehen, wenn sie gezwungen sind, KI-Entscheidungen umzusetzen, die sie moralisch infrage stellen.

c. Vertrauen vs. Misstrauen: Mitarbeiter müssen der KI vertrauen, um ihre Empfehlungen anzunehmen. Dieses Vertrauen kann jedoch durch mangelnde Transparenz oder fehlende Nachvollziehbarkeit der KI-Entscheidungen untergraben werden.

d. Angst vor Fehlern: Die Sorge, dass die KI falsche Entscheidungen treffen könnte, kann bei Mitarbeitern zu Stress oder Zweifeln führen, insbesondere wenn die Konsequenzen dieser Entscheidungen gravierend sind.

4.8.10 Kundenbetroffenheit

a. Manipulation und Autonomie: Die gezielte Beeinflussung von Kundenentscheidungen durch GenAI wirft ethische Fragen auf, insbesondere in Bezug auf Manipulation und Respektierung der Autonomie.

b. Datenschutz und Privatsphäre: Die Nutzung persönlicher Daten zur Beeinflussung von Entscheidungen kann zu Bedenken hinsichtlich des Datenschutzes und der Privatsphäre führen.

c. Kognitive Dissonanz: Kunden könnten sich in einem Konflikt zwischen ihren eigenen Präferenzen und den Empfehlungen der KI befinden, was zu Unbehagen oder Zweifeln führen kann.

d. Abhängigkeit von KI: Die zunehmende Nutzung von KI-gestützten Empfehlungen kann dazu führen, dass Kunden ihre eigenen Entscheidungsfähigkeiten infrage stellen oder verlieren.

4.8.11 Lösungsansätze und ethische Gestaltung

a. Unternehmen sollten sicherstellen, dass die Entscheidungsprozesse der KI transparent und für Mitarbeiter und Kunden nachvollziehbar sind. Dies stärkt das Vertrauen und die Akzeptanz der Technologie.

b. Ethik als integraler Bestandteil: Ethik sollte von Anfang an in die Entwicklung und den Einsatz von GenAI integriert werden, um sicherzustellen, dass die Technologie verantwortungsvoll, verständlich und im Einklang mit gesellschaftlichen Werten genutzt wird.

c. Mitarbeiter sollten geschult werden, um die Grenzen und Möglichkeiten von GenAI zu verstehen und kritisch mit den Empfehlungen der KI umzugehen.

d. Kunden müssen über die Nutzung ihrer Daten und die Funktionsweise von KI-Systemen aufgeklärt werden, um ihre Autonomie und Entscheidungsfreiheit zu wahren.

4.8.12 Risiken und ethische Herausforderungen

Das Nudging durch GenAI birgt erhebliche ethische Risiken, insbesondere wenn es darum geht, die Autonomie und Entscheidungsfreiheit der Nutzer zu beeinträchtigen.

Die gezielte Ausnutzung von Schwächen und Toleranzen kann als manipulativ empfunden werden, insbesondere wenn die Nutzer nicht transparent über die Mechanismen informiert werden.
Wenn Nutzer das Gefühl haben, dass ihre Privatsphäre oder Autonomie verletzt wird, kann dies das Vertrauen in das Unternehmen oder die Technologie untergraben.

Digitale Ethik kann dazu beitragen, die Risiken des Micro-Nudging zu minimieren und den Einsatz von GenAI verantwortungsvoll zu gestalten, indem Unternehmen offenlegen, wie GenAI eingesetzt wird und welche Mechanismen (z. B. Micro-Nudging) zum Tragen kommen. Nutzer sollten die Kontrolle darüber behalten, ob und wie sie durch KI beeinflusst werden möchten.

4.8.13 Fazit Nudging

Nudging durch GenAI stellt eine mächtige Technologie dar, die sowohl Chancen (z. B. verbesserte Kundenerlebnisse, effizientere Mitarbeiterkommunikation) als auch Risiken (z. B. Manipulation, Verlust von Vertrauen) birgt. Digitale Ethik spielt eine zentrale Rolle, um sicherzustellen, dass diese Technologie verantwortungsvoll eingesetzt wird und die Autonomie sowie Würde der Nutzer respektiert werden. Durch Transparenz, Nutzerkontrolle und ethische Leitlinien kann ein fairer und vertrauenswürdiger Umgang mit GenAI gewährleistet werden.

4.9 Konkrete Umsetzung in Unternehmen

Wie kann ein Unternehmen nun die hier dargestellten Aspekte der KI-Ethik konkret umsetzen?

Ausgehend von den Unternehmenswerten oder vom Leitbild des jeweiligen Unternehmens empfiehlt sich eine Sammlung von ethischen Aspekten, die in Zusammenhang mit den Unternehmenswerten oder dem Leitbild und den konkreten Unternehmensaktivitäten mit KI stehen. Diese Aspekte können dann entweder in ein Dokument überführt werden, das sich mit »KI-Ethik« beschäftigt, oder auch umfassender mit »Digitalethik«, also umfassender mit allen digitalen Technologien im Unternehmen.

Ob diese Inhalte im Bereich Ethik dann mit den Aussagen des Unternehmens zu Corporate Social Responsibility (»CSR«) oder Environmental, Social und Governance (»ESG«) verknüpft oder verbunden werden, ist eine Frage der Branche und des Kontextes des jeweiligen Unternehmens.

Grundsätzlich ist eine Beschäftigung und Dokumentation von Aspekten der KI-Ethik immer dann zu empfehlen, wenn ein Unternehmen umfangreich KI-Technologien nutzt oder diese selbst weiterentwickelt oder komplett selbst entwickelt.

Ob KI-Ethik gegebenenfalls auch Eingang in Berichtspflichten finden muss, wie die aus der in der EU in nationales Recht umzusetzende Corporate Sustainability Reporting Directive, muss im Einzelfall geklärt werden – grundsätzlich wird »Digitalethik« dort als Risiko eingestuft.

Hier sind ein paar konkrete Beispiele für einzelne Aspekte und Bereiche, die in ein KI-Ethik-Dokument im Unternehmen aufgenommen werden könnten, in alphabetischer und nicht wertender Reihenfolge:

Fairness und Nicht-Diskriminierung (Fairness and Non-Discrimination)

»Datengetriebene Technologien können Muster widerspiegeln oder sogar verstärken, die entweder in der Gesellschaft bestehen oder schlichtweg in den verwendeten Daten veranlagt sind, zum Beispiel Vorurteile und Diskriminierungen. Durch eine transparente und ausgewogene Zusammenstellung von Daten können diese Effekte offengelegt oder vermindert werden.«

Mensch im Mittelpunkt (Human-centric AI)

»Wir wollen KI nutzen und entwickeln, bei der der Mensch mit seinen Fähigkeiten im Mittelpunkt steht und die KI den Menschen bestmöglich unterstützt.«

Nachhaltigkeit (Sustainability)

»KI sollte die ökologische, wirtschaftliche und soziale Nachhaltigkeit fördern. Negative soziale, wirtschaftliche und ökologische Auswirkungen müssen vermieden werden.«
Insbesondere mit Blick auf den bislang großen Bedarf an Ressourcen bei der Entwicklung und dem Betrieb von KI, beispielsweise von großen Sprachmodellen als generative KI, ist dieser

Aspekt nicht nur ein ökologischer, sondern auch ein wirtschaftlicher. Eine Beschäftigung mit diesem Aspekt erscheint daher aus heutiger Sicht unerlässlich.
Recht auf Privatsphäre und Datenschutz (Right to Privacy and Data Protection)
»Bei der Entwicklung und Nutzung von KI sollten die Gesetze zum Schutz der Privatsphäre und des Datenschutzes beachtet werden.«

Schutz und Sicherheit (Safety & Security)

»Sicherheitsmaßnahmen müssen während des gesamten Lebenszyklus des KI-Systems gewährleistet sein, um den Schutz vor Angriffen und unerwünschten Schäden sicherzustellen.«

Verantwortung und Transparenz (Responsibility and Transparency)

»Der Einsatz von KI soll verantwortungsvoll erfolgen und wir wollen KI transparent für Mitarbeitende und Kunden machen.«

4.10 Ausblick: KI-Ethik im Zeitalter exponentieller Dynamik

Die Entwicklung künstlicher Intelligenz ist geprägt von einer technischen Dynamik: Neuronale Netze, generative Modelle und autonome Systeme evolvieren schneller als Regulierung und gesellschaftliche Debatten mithalten können. Doch genau diese Geschwindigkeit macht eine kontinuierliche, adaptive Weiterentwicklung ethischer Rahmenwerke unabdingbar.

Das EU-AI-Office, das als zentrale Instanz für die Umsetzung des KI-Acts geschaffen wurde, wird hier zur Schlüsselinstitution. Es muss nicht nur die Einhaltung von Bias-Kriterien überwachen, sondern auch als Katalysator für Dialog agieren – zwischen Politik, Tech-Unternehmen, Zivilgesellschaft und Wissenschaft. Denn die Frage, wie KI-Systeme fair, transparent und verantwortungsvoll gestaltet werden können, lässt sich nur kooperativ beantworten.

Die Herausforderungen sind enorm:

a. Technologische Sprünge (z. B. KI-basierte Echtzeit-Entscheidungen in kritischer Infrastruktur) erfordern agile Regulierung, die Risiken antizipiert, ohne Innovation zu ersticken.

b. Globale Fragmentierung: Während die EU auf strenge Compliance setzt, fehlen in vielen Ländern grundlegende Schutzmechanismen. Das EU-AI-Office könnte hier eine Brückenfunktion einnehmen – etwa durch globale Zertifizierungssysteme oder Wissensaustausch.

c. Machtbalance: Wer kontrolliert KI? Der Einfluss weniger Tech-Giganten steht im Spannungsfeld zu öffentlicher Aufsicht. Das EU-AI-Office muss hier klare Grenzen setzen, etwa durch verbindliche Audits für Hochrisiko-KI.

Die Zukunft der KI-Ethik wird davon abhängen, ob es gelingt, Technologie demokratisch zu gestalten. Das EU-AI-Office steht symbolisch für diesen Anspruch – doch sein Erfolg wird sich daran messen, ob es nicht nur Regeln durchsetzt, sondern auch gesellschaftliche Teilhabe sichert. Letztlich geht es um mehr als Algorithmen: Es geht um die Frage, wer in der Knowledge Society das Sagen hat – Maschinen, Konzerne oder die Menschen, deren Leben sie prägen.

»Innovation braucht Ethik – nicht als Bremsklotz, sondern als Kompass und als Antrieb für bahnbrechende Entwicklungen.«

Bitkom vertritt mehr als 2.200 Mitgliedsunternehmen aus der digitalen Wirtschaft. Sie generieren in Deutschland gut 200 Milliarden Euro Umsatz mit digitalen Technologien und Lösungen und beschäftigen mehr als 2 Millionen Menschen. Zu den Mitgliedern zählen mehr als 1.000 Mittelständler, über 500 Startups und nahezu alle Global Player. Sie bieten Software, IT-Services, Telekommunikations- oder Internetdienste an, stellen Geräte und Bauteile her, sind im Bereich der digitalen Medien tätig, kreieren Content, bieten Plattformen an oder sind in anderer Weise Teil der digitalen Wirtschaft. 82 Prozent der im Bitkom engagierten Unternehmen haben ihren Hauptsitz in Deutschland, weitere 8 Prozent kommen aus dem restlichen Europa und 7 Prozent aus den USA. 3 Prozent stammen aus anderen Regionen der Welt. Bitkom fördert und treibt die digitale Transformation der deutschen Wirtschaft und setzt sich für eine breite gesellschaftliche Teilhabe an den digitalen Entwicklungen ein. Ziel ist es, Deutschland zu einem leistungsfähigen und souveränen Digitalstandort zu machen.

Bitkom e.V.

Albrechtstraße 10
10117 Berlin
T 030 27576-0
bitkom@bitkom.org

bitkom.org

bitkom