Künstliche Intelligenz: praktische Haftungsfragen*

Mauro Quadroni†

I. I. Einführung

A. Eine Arbeitsdefinition von künstlicher Intelligenz

Der Begriff der künstlichen Intelligenz (KI) ist ein alter Begriff, der einem ständigen Wandel unterliegt. Auf kommerzieller Ebene wurden sehr einfache Systeme als KI bezeichnet, während der Technologie heute meist als Synonym für Systeme gilt, die auf *«Deep Neural Network»* basieren. Je nach Land oder Institution enthält die Definition unterschiedliche Kernelemente.[1] Für die Zwecke dieses Artikels wird KI daher allgemein als ein Software-basiertes System bezeichnet, das in der Lage ist, komplexe Probleme zu lösen, die früher dem Menschen vorbehalten waren.[2] Zu diesem Zweck gibt es mindestens zwei Elemente, die, ohne sie allgemein zu definieren, in der aktuellen juristischen Literatur oft grosse Bedeutung haben: die Fähigkeit zu lernen,[3] und die Fähigkeit, mit einem gewissen Grad an Autonomie zu handeln.[4] Um relevante praktische Aspekte nicht ausser Acht zu lassen, wird keine engere Definition von KI verwendet. Dieser Artikel konzentriert sich auf bestimmte rechtliche Aspekte des Einsatzes von KI, die für die Klärung von Haftungsfragen von Interesse sind.

B. Praktische Probleme für Haftungsfragen

Unabhängig von einer Definition von KI gibt es m.E. mehrere Merkmale, die dieser Technologie üblicherweise zugeschrieben werden und die für Haftungsfälle von Bedeutung sind. Diese lassen sich in vier Kategorien zusammenfassen: (i) die Fähigkeit, flexibel eingesetzt zu werden; (ii) die Fähigkeit, autonom zu handeln; (iii) die Fähigkeit, kontinuierlich zu lernen; und (iv) die Fähigkeit, Schlussfolgerungen aus einer grossen Datenmenge zu ziehen.

1. Die Fähigkeit, flexibel angewendet zu werden

Klassische Computerprogramme sind sehr präskriptiv: Jede mögliche Option muss in der Software ausdrücklich programmiert werden, damit sie funktionieren kann. Wenn das Computerprogramm auf eine unerwartete Situation stösst, funktioniert es, grob gesagt, nicht mehr. Bei KI-basierten Lösungen hingegen lernt die KI anhand von Beispielen, eine Aufgabe zu generalisieren, so dass sie in der Lage ist, unendlich viele unterschiedliche Situationen flexibel zu bewältigen, ohne jede einzelne Möglichkeit vorhersehen und vorprogrammieren zu müssen.[5] Bei der Entwicklung von KI-Modellen[6] wird ein System mehrmals trainiert und getestet, bis eine zufriedenstellende Richtigkeitsquote erreicht ist. Wenn jedoch beim Training Perfektion erreicht wird, ist dies fast immer ein Zeichen für einen grundlegenden Fehler, nämlich das *«Overfitting»*. Das heisst, dass es die Beispiele «auswendig» gelernt hat oder einen sehr starren Ansatz erlernt hat und daher zukünftige Anwendungsfälle, die sich auch nur geringfügig von den beim Training verwendeten Beispielen unterscheiden, falsch behandeln könnte.[7] Da diese Verallgemeinerungsfähigkeit letztlich als ein sehr komplexer statistischer Ansatz angesehen werden kann, besteht der Nachteil darin, dass man weniger Kontrolle darüber hat, wie Grenzfälle, neue oder aussergewöhnliche Fälle behandelt werden. Während man also bei klassischen Programmen im Voraus mit Sicherheit sagen kann (und muss), wie jeder Fall behandelt wird, gibt es bei KI-basierten Lösungen immer eine Fehlermarge, die berücksichtigt werden sollte. Um die Tatsache abzumildern, dass reine KI von Natur aus fehleranfällig bzw. potenziell unvorhersehbar ist (obwohl sie zugegebenermassen oft weniger fehleranfällig ist als der durchschnittliche Mensch), wird sie daher mit klassischer Codierung gekoppelt, deren Funktionalität unterstützt.

2. Die Fähigkeit, autonom zu handeln

Eine der grundlegenden Funktionen der Informatisierung und der Computerprogramme bestand schon immer darin, die menschliche Arbeit zu beschleunigen und zu automatisieren. Dank der bereits erwähnten Fähigkeit von KI-basierten Systemen, flexibel mit Eingaben umzugehen und somit sehr komplexe Aufgaben auszuführen, die den Fähigkeiten eines Menschen ähneln oder sie sogar übertreffen, wird Maschinen zunehmend zugetraut, autonom über die beste Vorgehensweise zur Erledigung der zugewiesenen Aufgaben zu entscheiden. Aufgrund dieser Vorstellung, dass das Computerprogramm autonom ist, wird auch die Kausalkette zwischen den Handlungen des KI-basierten Systems und dessen Nutzer als schwächer

---

* Artikel veröffentlicht in der Zeitschrift HAVE 04/2021. Der Artikel wurde in einigen Passagen gegenüber seiner ursprünglichen Fassung verändert, um die Lesbarkeit zu verbessern.

† RA, MLaw, Founding Partner der AI Legal & Strategy Consulting AG in Zürich.

[1] ANDREA BERTOLINI, Artificial Intelligence and Civil Liability, Brüssel 2020, 15 ff.

[2] Bericht der interdepartementalen Arbeitsgruppe «Künstliche Intelligenz » an den Bundesrat, Bern 2019, 7. Diese Definition nähert sich an den EU-Vorschlag für ein Gesetz über künstliche Intelligenz.

[3] MELINDA F. LOHMANN, Ein Zukunftsfähiger Haftungsrahmen für Künstliche Intelligenz, HAVE 2021, 111 ff., 111 f.

[4] ERDEM BÜYÜKSAGIS, Responsabilité pour les systèmes d'intelligence artificielle, HAVE 2021, 12 ff., 12; SILVIO HÄNSENBERGER, Die Haftung für Produkte mit lernfähigen Algorithmen, Jusletter 2018, Rz. 7 f.

[5] ARUN DAS/PAUL RAD, Opportunities and Challenges in Explainable Artificial Intelligence (XAI): A Survey, arXiv 2020, 1.

[6] Als KI-Modell gilt eine Datei, die darauf trainiert wurde, bestimmte Arten von Mustern zu erkennen, <https://docs.microsoft.com/de-ch/windows/ai/windows-ml/what-is-a-machine-learning-model>, besucht am 22.9.2021.

[7] <www.ibm.com/cloud/learn/overfitting>, besucht am 22.9.2021.

empfunden.[8]

3. Die Fähigkeit, fortlaufend zu lernen

Bei traditionell programmierten Computerprogrammen muss jede Anpassung im Voraus durchdacht und explizit einprogrammiert werden. Ändern sich die Umstände, unter denen die Software eingesetzt wird, muss das Computerprogramm manuell angepasst werden. KI-gestützte Systeme können jedoch aus Erfahrung lernen, so dass die bei der Erfüllung der ihnen zugewiesenen Aufgaben gesammelten Informationen genutzt werden können, um ihnen beizubringen, diese besser auszuführen und sich an veränderte Gegebenheiten anzupassen. Dies schwächt die wahrgenommene Verbindung zwischen dem Programmierer und dem Ergebnis des KI-basierten Systems weiter ab.[9]

4. Die Fähigkeit, aus einem grossen Datenvolumen Rückschlüsse zu ziehen

Eine der transformativen Fähigkeiten der KI-Technologie ist die Fähigkeit, automatisch Schlüsse und Zusammenhänge aus einer grossen Datenmenge zu ziehen und dieses Wissen zur Erfüllung einer Aufgabe zu nutzen. Dies ermöglicht Probleme anzugehen, die bisher aufgrund ihrer Komplexität als zu schwierig erachtet wurden.[10] KI ist in der Lage, selbstständig Lösungen für bestimmte Probleme zu finden oder aus Daten Prognosen über die Zukunft mit höherer Sicherheit zu treffen. Da sich die Vorgehensweise eines KI-Systems von der eines Menschen unterscheidet, kann ein und dasselbe System zwar eine hohe Korrektheit erreichen, aber dennoch Fehler machen, die selbst für einen Laien offensichtlich sind. Bald stellte sich heraus, dass die von der KI zu diesem Zweck verwendeten Daten nicht nur das Potenzial haben, grundlegende kausale Zusammenhänge und wissenschaftliche Regeln des zu lösenden Problems zu erkennen, sondern auch eine grosse Menge unbedeutender Korrelationen, eingebetteter Voreingenommenheit und falscher Vorstellungen sowie schlechter Gewohnheiten enthalten. Diese wirken sich auf die Funktionsweise von KI-Systemen aus und führen zu unerwünschten Ergebnissen.[11] Diese Erkenntnis verändert die Bedeutung von Daten grundlegend. Im Gegensatz zu klassischen Computerprogrammen, bei denen die Kodierung das Grundproblem ist, erfordert die Entwicklung von KI-basierten Systemen auch eine bisher kaum bekannte Pflege der zugrunde liegenden Daten.

I. Haftungsfragen

In der Praxis kann KI auf sehr unterschiedliche Weise eingesetzt werden und berührt die meisten Bereiche unseres Lebens. Verursacht sie einen Schaden, besteht in den meisten Fällen ein Vertragsverhältnis zwischen dem Geschädigten und dem Verantwortlichen, oder der Sachverhalt wird durch spezielle gesetzliche Haftungsnormen geregelt, so dass die speziellen Gegebenheiten des Einzelfalls zu berücksichtigen sind. Dennoch lohnt es sich, die speziellen Haftungsfragen der Deliktshaftung zu untersuchen, da diese die meisten rechtlichen Grundfragen des Haftpflichtrechts abdecken.

A. Deliktshaftung

Die Deliktshaftung nach Art. 41 OR gilt als Auffangnorm für die zivilrechtliche Haftung im Schadensrecht. Um die Haftung zu begründen, müssen folgende Punkte nachgewiesen werden: ein Schaden, die Widerrechtlichkeit der Schadenszufügung, der (adäquate) Kausalzusammenhang, das unrechtmässige Verhalten des Schädigers sowie dessen Verschulden.

1. Schaden und Widerrechtlichkeit

Als relevanter Schaden wird im Rahmen der Deliktshaftung in der Regel die Schädigung eines absolut geschützten Rechtsgutes, d.h. von Personen und Sachen, definiert. Relevant sind dabei auch reine Vermögensschäden, für die es eine Schutznorm gibt.[12] Diese beiden haftungsrechtlichen Aspekte stellen bei der Schadensverursachung durch KI keine grosse Besonderheit dar und werden daher nicht weiter erörtert.

2. Kausalzusammenhang

a. Autonomie eines KI-Systems

Bei KI-Anwendungen ist der Aspekt der Identifizierung der schuldigen Person und damit der Bestimmung eines Schädigers besonders schwierig.

Ein erster Aspekt, den es zu berücksichtigen gilt, ist der Aspekt der Autonomie von Systemen, die eigenständig handeln können (d.h. in rechtlich relevanter Weise gemäss ihrer eigenen Programmierung handeln). In diesen Fällen stellt sich die Frage, ob eine KI-Anwendung ein solches Mass an Autonomie erreichen kann, dass niemand für ihre Handlungen verantwortlich gemacht werden kann. Die Antwort ist grundsätzlich negativ, da sich immer mindestens eine Person identifizieren lässt, die für die Erstellung eines KI-basierten Systems und seiner Fähigkeiten verantwortlich ist, und eine, die für seinen Einsatz verantwortlich ist.[13] Es lassen sich verschiedene Szenarien beschreiben, die den Grad der Kontrolle des Menschen über autonome KI-Systeme bei deren Einsatz beschreiben und bei der Zurechnung der Haftung unterstützen können:[14]

– **Human in the loop**: Hier ist ein Mensch als Verantwortlicher und Entscheidungsträger in das System

[8] CAMPOLO ALEXANDER/CRAWFORD KATE, «Enchanted Determinism: Power without Responsibility in Artificial Intelligence», Engaging Science, Technology, and Society, January 2020, 1 ff., 12 f.

[9] CAMPOLO/CRAWFORD (Fn. 8), 12 f.

[10] Eine der letzten grossen Erfolge der KI ist die Fähigkeit, aus ihrer Zusammensetzung die dreidimensionale Form eines Proteins vorherzusagen, <www.nature.com/articles/d41586-020-03348-4>, besucht am 22.9.2021.

[11] DAS/RAD (Fn. 5), 1.

[12] BGE 141 III 527 E. 3.2.

[13] LOHMANN (Fn. 3), 112.

[14] JEROME GURTNER, Les nouvelles technologies et la responsabilité des avocats, in: Responsabilité civile et nouvelles technologies, in: Responsabilité civile et nouvelles technologies, Genève/Zurich 2019, 45 ff., 73 f.; JERMY PRENIO/JEFFERY YONG, Humans keeping AI in check – emerging regulatory expectations in the financial sector, BIS 2021, verfügbar unter <www.bis.org/fsi/publ/insights35.htm>, besucht am 22.9.2021.

eingebunden. Das bedeutet, dass das KI-System in zwei Teilen unterteilt werden kann: Der erste Teil, der von der KI gestaltet wird, ist ein Entscheidungshilfesystem, das die beste Vorgehensweise bzw. Handlungsoptionen anbietet. Die Person trifft dann die Entscheidung, eine bestimmte Handlung zu ergreifen oder nicht zu handeln. Der zweite Teil des Systems ist die anschliessende Ausführung der beschlossenen Handlung, die berechenbar sein sollte und damit der Zurechenbarkeitsfrage eines klassischen Computerprogramms entspricht. Wenn trotz der Einbeziehung einer Person die Folgen der Entscheidung dieser Person nicht bestimmt werden können, weil wieder ein KI-System beteiligt ist, entspricht dieser zweite Teil des Falles einer der nächsten Kategorien.

– **Human on the loop**: Hier ist ein Mensch sowohl an der Implementierung als auch an der Überwachung der Aktivitäten des KI-Systems beteiligt. Das KI-System kann die ihm zugewiesene Aufgabe grundsätzlich selbstständig ausführen. Eine Person muss nicht jede einzelne Aktion genehmigen oder bestimmen, sondern überwacht regelmässig die Aktivität des KI-basierten Systems und kann gegebenenfalls eingreifen, wenn Probleme festgestellt werden. Der grundlegende Unterschied zwischen «in the loop» und «on the loop» besteht also darin, dass der Mensch im ersten Fall eine Entscheidungsfunktion hat, während er im zweiten Fall eine reaktive Überwachungsfunktion hat.

– **Human off the loop**: In diesem Fall läuft das KISystem bei jedem Vorgang völlig autonom, so dass die Möglichkeit des Eingreifens einer Person begrenzt ist. Im Prinzip wirkt der Mensch bei der Ausführung der zugewiesenen Tätigkeit nicht mit. Dies ist der Fall, wenn das System auf der Grundlage einer vortrainierten KI agiert, aber auch, wenn die KI während des Einsatzes eigenständig aus der gesammelten Erfahrung trainiert wird.

Bei «human in the loop» bestimmt eine identifizierbare Person die Handlung des KI-basierten Systems. Dies ist in der Regel dann der Fall, wenn das KI-basierte System die Vorarbeit für den Menschen leistet. Diese Art der Nutzung wird in der Praxis gewählt, wenn die KI nur einfache mechanische Vorarbeiten erfüllen soll, wie z.B. die automatische Extraktion von Informationen aus einem Dokument, in Situationen, in denen das Risiko eines Schadens hoch ist, wie z.B. bei der Erstellung einer Diagnose, oder in komplexen Situationen, in denen die Lösung eine Mischung aus maschinellen und menschlichen Fähigkeiten erfordert, wie z.B. bei der Analyse einer komplexen Marktsituation, um eine Anlageentscheidung zu treffen. Die Haftung des Nutzers kann gemildert oder ausgeschlossen werden, wenn das Programm fehlerhaft ist[15] oder wenn der Nutzer nicht in der Lage ist, die Folgen ihrer Entscheidung zu erkennen oder die von der KI vorgeschlagene Lösung kritisch genug zu prüfen.[16] Die Gründe für die letztgenannte Möglichkeit liegen häufig nicht in der Natur der verwendeten KI-Technologie (da diese bereits vor der Entscheidung der Person einen Beitrag geleistet hat), sondern vielmehr in der mangelnden Information, der Ausbildung der Person, die *«in the loop»* ist, oder in der Selbstverschuldung der Person, die trotz ausreichender Schulung oder Ausbildung nicht über die notwendigen Fähigkeiten verfügt, um ihre Rolle zu erfüllen.

Beim *«human on the loop»* ist das System grundsätzlich fähig, selbstständig zu handeln, wird aber von einer Person beaufsichtigt. Dies geschieht in der Regel, um fortbestehende Risiken auszuschalten, die trotz des bestimmungsgemässen Einsatzes des KI-Systems oder im Rahmen eines vernünftigerweise vorhersehbaren Missbrauchs entstehen können.[17] Die Zurechenbarkeit eines Schadens hängt von den Einflussmöglichkeiten der überwachenden Person ab. Die Aufsichtsperson kann ein Verschulden treffen, wenn sie es trotz Pflicht und Gelegenheit zum Handeln unterlässt, das Fehlverhalten des KI-basierten Systems zu unterbrechen oder zu verhindern und dessen Schaden zu mindern oder gar unverzüglich rückgängig zu machen.[18] Kommt es jedoch trotz pflichtgemässen Verhaltens der Aufsichtsperson zu einem Schaden durch das KI-Fehlverhalten, so ist das Verschulden und damit ein Verantwortlicher an anderer Stelle zu suchen. Dies ist in der Regel der Anbieter der KI-Lösung. Gibt es keine Anhaltspunkte dafür, dass der Anbieter aufgrund eines Fehlers der KI verantwortlich ist, kann die Lehre vom allgemeinen Gefährdungsrecht zu Hilfe kommen,[19] so dass der Nutzer oder der Inhaber dieses Systems, je nachdem, wer die beste Möglichkeit gehabt hätte, den Schaden zu vermeiden, dennoch als verantwortliche Person angesehen werden kann.

Bei *«human off the loop»* wird das System grundsätzlich sich selbst überlassen, um seine Aufgabe auszuführen. Diese Konstellation wird häufig im Zusammenhang mit autonomen Waffen und den damit verbundenen Verbotsbestrebungen diskutiert,[20] ist aber besser bekannt für den Einsatz von autonomen Fahrzeugen sowie die Automatisierung bestimmter Prozesse. Hier ist ein menschliches Eingreifen während des Betriebs des KI-basierten Systems im Prinzip nicht vorgesehen, es sei denn, dies geschieht, um das Ziel der Maschine anzupassen (man denke an ein autonomes Auto, das eine Anpassung des Ziels vor dem Ende der Fahrt erlaubt). Hier kann den Nutzer des KI-gestützten Systems, also die Person, die die Aktivität des KI-gestützten Systems in Gang gesetzt hat, im Prinzip nur dann ein Verschulden treffen, wenn das

[15] Siehe II.A.2.b.

[16] Diese entsprechen namentlich einem Fall der Kausalitäts-unterbrechung aufgrund Drittverschuldens, oder des Mangels an dem subjektiven Element des Verschuldens.

[17] Vgl. der EU-Vorschlag für ein Gesetz über künstliche Intelligenz, verfügbar unter <https://eur-lex.europa.eu/legal-content/DE/TXT/HTML/?uri=CELEX:52021PC0206&from=EN>, besucht am 22.9.2021.

[18] Fehlt eine ausdrückliche Schutznorm, welche eine Garantenstellung der überwachenden Person begründet, kann man bei absolut geschützten Gütern auf die Doktrin des Gefahrensatzes zurückgreifen: HARDY LANDOLT, Haftung für rechtmässige Schadenverursachung, HAVE 2014, S. 3 ff., S. 6; Urteil des BGer 4A_104/2012 vom 3. August 2012 E. 2.1.

[19] Siehe II.A.2.b.

[20] Siehe z.B. die UN-Arbeit diesbezüglich: <www.un.org/disarmament/the-convention-on-certain-conventional-weapons/background-on-laws- in-the-ccw/>, besucht am 22.9.2021.

angestrebte Ziel oder die Nutzung des vollautonomen KI-Systems selbst pflichtwidrig ist. Der Einsatz eines solchen Systems ist z.B. dann pflichtwidrig, wenn das KI-basierte System nicht sicher eingesetzt werden kann (z.B. ein Fahrzeug ist nicht in der Lage, Menschen zuverlässig zu erkennen und zu umfahren),[21] wenn notwendige Sicherheitsmassnahmen nicht vorgesehen sind (z.B. es gibt keine Begrenzung der Transaktionsfähigkeit eines automatisierten Handelssystems) oder der Einsatz des KI-Systems durch Personen ermöglicht wird, die nicht in der Lage sind, die Gefahren zu erkennen und angemessen zu beseitigen (z.B. Bereitstellung eines Diagnosesystems an nicht fachlich geschulte Personen).[22] Meines Erachtens ist die Zulässigkeit des Einsatzes eines vollständig autonomen, auf KI basierenden Systems nicht einfach anzunehmen, wenn keine angemessene Sorgfalt im Hinblick auf das Schadenspotenzial gesetzt wird.

Es mag relativ einfach sein, mit Hilfe der erwähnten Kategorisierung den Kausalzusammenhang zwischen der KI-Anwendung und ihrem Nutzer oder Anbieter herzustellen.[23] Ist das nicht genug, es wird jedoch schwieriger, diesen Kausalzusammenhang von dort aus im Einzelfall zu entwickeln, um eine mögliche technisch fehlerhafte Ursache zu identifizieren. Insbesondere muss analysiert werden, was in der Kausalkette passieren kann, um weitere Personen zu identifizieren, die potenziell für einen Schaden verantwortlich gemacht werden können.

b. Transparenz einer KI-Software

Generell kann gesagt werden, dass die schadensverursachende Person entweder als der Benutzer oder der Ersteller des Programms identifiziert werden kann, wenn das Funktionieren eines klassischen Computerprogramms direkt von seiner Kodierung abhängt.[24] Eine KI-Anwendung hingegen ist das Ergebnis der Kombination aus einer Software (die der klassischen Kodierung entspricht) und den für ihr Training verwendeten Daten. Beide Faktoren spielen eine entscheidende Rolle, und sie können oft mindestens zwei verschiedenen Personen zugeschrieben werden. Das bedeutet, dass sich die Zahl der möglichen Täter in relevanter Weise erweitert, was wiederum bedeutet, dass der zu klärende Sachverhalt komplexer wird.[25]

Auf der Datenseite ist es aufgrund der Tatsache, dass die einzelnen Daten nur einen Bruchteil der Funktionsweise der KI-Anwendung beeinflussen, schwierig festzustellen, welcher Datensatz in welchem Umfang das System dazu veranlasst hat, das rechtsverletzende Ergebnis zu produzieren.[26] Diese sind schwer zu lokalisieren und zu identifizieren. Die Grundlage eines KI-Systems kann ein (oder mehrere) von einem Dritten bereitgestelltes, vorab trainiertes allgemeines Modell sein, wofür die zum Training verwendeten Datensätze nicht zur Verfügung gestellt werden. Dann können proprietäre Datensätze verwendet werden, um das Modell an die eigenen Zwecke anzupassen. Schliesslich wird das Modell häufig anhand neu erhobener Daten trainiert und aktualisiert. Aus Datenschutzgründen sind die Daten für das kontinuierliche Training des Systems oft nicht oder nur in verschlüsselter Form verfügbar.[27]

Was die Kodierung betrifft, so stützt sich die KI-Entwicklung heute, wie bei jeder modernen Software, meist auf viele verschiedene öffentlich verfügbare Ressourcen, die sehr komplex sind.[28] Auch der herumliegende Code, der eine funktionierende Anwendung ermöglicht, kann sich auf Millionen von Codezeilen aus verschiedenen Quellen belaufen, so dass im Einzelfall die Rückverfolgbarkeit eines Fehlers nahezu unmöglich sein kann.[29]

In jüngster Zeit hat sich die Diskussion um die Erklärbarkeit und Nachvollziehbarkeit von KI-Anwendungen so entwickelt, dass derzeit viele Good Practices und Services entwickelt werden, die versuchen, die Entscheidungsfaktoren einer KI transparent zu machen und damit zu erklären, wie sie funktionieren (sog. *«explainable AI»* und *«interpretable AI»*).[30]

Trotz dieser begrüssenswerten Entwicklung wird es für den Geschädigten fast unmöglich, die tatsächliche Ursache des Schadensereignisses nachzuweisen, da dies einen uneingeschränkten Zugang zu allen vorhandenen Logdateien, zur Codierung der Software sowie, wenn nichts anderes verfügbar ist, dass die Transparenz der KI gewährleistet, zu den als Training verwendeten Daten erfordert.[31]

Bei fehlender Transparenz des Systems kommt die Doktrin des allgemeinen Gefahrensatzes zu Hilfe, da nicht genau gesagt werden kann, was die Ursache des Schadens ist. Nach dieser Doktrin muss derjenige, der einen gefährlichen Zustand schafft, aufrechterhält oder anderweitig rechtsverbindlich zu vertreten hat, alle erforderlichen und zumutbaren Schutzmassnahmen ergreifen, die geeign[32]et

---

[21] Siehe II.A.2.b. bei der Besprechung der Doktrin des Gefahrensatzes.

[22] Zu diesem Schluss kommen namentlich die diskutierten Grundsätze des Gefahrensatzes, der Produkthaftung sowie der vertraglichen Haftungspflichten.

[23] Auf die Problematik der heutigen Realität, wo Software oftmals als eine Dienstleistung angeboten wird, und wo diese oft auch eine Kombination von verschiedenen von sich unabhängigen «Microservices» darstellen kann, wird hier nicht eingegangen. Die Fragmentierung der Angebote, welche eine Folge der zu willkommenen Verfügbarkeit der eigenen Daten ist, stellt eine Herausforderung bei der Bestimmung des Tatbestands und der involvierten Parteien.

[24] Der Ersteller des Programms kann heutzutage ein komplexes Netzwerk von unabhängigen mitwirkenden Personen sein. Siehe diesbezüglich CHRISTIANE WENDEHORST, Safety and Liability Related Aspects of Software, Luxembourg 2021, 16 ff.

[25] Bezüglich der Diskussion über einen Anpassungsbedarf der Regulierung, um diese Komplexität aus Gerechtigkeits- und Effizienzgründen zu reduzieren, siehe z.B. LOHMANN (Fn. 3) und WENDEHORST (Fn. 22).

[26] DAS/RAD (Fn. 5), 9 ff.

[27] WENDEHORST (Fn. 22), 82.

[28] WENDEHORST (Fn. 22), 22 ff.

[29] Obwohl es relativ einfach ist, ein neurales Netzwerk zu erstellen (GPT-2 Source Code beträgt 174 Kodierungslinien, siehe <https://github.com/openai/gpt-2/blob/master/src/model.py>, besucht am 22.9.2021), die zugrunde liegende Libraries, welches die einfache Erstellung von neuralen Networks ermöglicht, besteht aus Millionen Kodierungslinien. Tensorflow, z.B., fast 3 Millionen Kodierungslinien. <www.openhub.net/p/tensorflow>, besucht am 22.9.2021.

[30] DAS/RAD (Fn. 5), 1 f.

[31] Vgl. LOHMANN (Fn. 3), 17.

[32] LANDOLT (Fn. 17), 5; BGE 124 III 297 E. 5b: «Der Gefahrensatz ist […] heranzuziehen, wenn der Kausal- bzw. der Rechtswidrigkeitszusammenhang zwischen einer Unterlassung und dem eingetretenen Schaden zu beurteilen ist.»

sind, die Beeinträchtigung absolut geschützten Rechtsgüter zu vermeiden.

Folgt man der Argumentation, dass KI naturgemäss das Potenzial hat, unerwartet zu wirken und dies negative Auswirkungen auf rechtlich absolut geschützte Güter (wie Eigentum, Leben oder Gesundheit) haben kann, so trifft den Nutzer bzw. den Anbieter eine Pflicht zur Schadensminderung. Der Geschädigte ist deswegen meines Erachtens nicht verpflichtet, die Kausalkette weiter als bis zum Anbieter zu beweisen, um seinen Anspruch nach Art. 41 OR geltend zu machen. Er muss nur nachweisen können, dass der Anbieter nicht die erforderlichen Sorgfaltsmassnahmen zur Schadensminderung getroffen hat.

Dieser Ansatz setzt voraus, dass man trotz des grossen Erfolgs eines KI-basierten Systems Sicherheitseinschränkungen treffen muss. Diese Haltung wird durch die Diskussion um den Einsatz von KI bestätigt. Die Grenze dieser Haftung ist der Grundsatz des zulässigen Risikos. Demnach *«lässt sich eine Gefährdung fremder Rechtsgüter, die über das allgemeine Lebensrisiko nicht hinausgeht, nicht verbieten, sondern gefordert werden kann nur die Einhaltung eines bestimmten Mindestmasses an Sorgfalt und Rücksichtnahme. Beim erlaubten Risiko tritt an die Stelle des Verbots jeglicher Gefährdung das Gebot, die Gefahr auf dasjenige Minimum einzuschränken, das gar nicht oder nur mit unverhältnismässigem Aufwand ausgeschlossen werden kann, wenn man die entsprechende Tätigkeit überhaupt zulassen will. Dabei geht es um die Frage, welche Risiken allgemein in Kauf zu nehmen sind, und nicht um eine Ermässigung der Sorgfaltsanforderungen. Eine Sorgfaltspflichtverletzung ist nur anzunehmen, wenn der Täter eine Gefährdung der Rechtsgüter des Opfers hätte voraussehen bzw. erkennen können und müssen. Für die Beantwortung dieser Frage gilt der Massstab der Adäquanz. Danach muss das Verhalten des Täters geeignet sein, nach dem gewöhnlichen Lauf der Dinge und den Erfahrungen des Lebens einen Erfolg wie den eingetretenen herbeizuführen oder mindestens zu begünstigen. Damit der Eintritt des Erfolgs dem Täter zuzurechnen ist, genügt seine blosse Vorhersehbarkeit nicht. Vielmehr stellt sich die weitere Frage, ob er auch vermeidbar war. Dazu wird ein hypothetischer Kausalverlauf untersucht und geprüft, ob der Erfolg bei pflichtgemässem Verhalten des Täters ausgeblieben wäre. Dabei genügt es für die Zurechnung des Erfolgs, wenn das Verhalten des Täters mindestens mit einem hohen Grad der Wahrscheinlichkeit oder mit an Sicherheit grenzender Wahrscheinlichkeit die Ursache des Erfolgs bildete.»*[33]

3. Verschulden

a. Verschulden des Anwenders

Um ein Verschulden festzustellen, muss eine Sorgfaltspflicht verletzt worden sein.[34] In diesem Zusammenhang stellt sich die Frage, wer eigentlich zwischen dem Nutzer und dem Anbieter einer KI-Lösung haftet, wenn es sich bei der Sorgfaltspflichtverletzung um das Unterlassen notwendiger Massnahmen zur Schadensvermeidung handelt. Insbesondere stellt sich die Frage, ob eine Sorgfaltspflicht begründet werden kann und wie der Geschädigte die Verletzung dieser beweisen kann.

1. human in the loop

Bei KI-basierten Systemen, bei denen ein Mensch die Entscheidungsfunktion behält, müssen sowohl die objektive als auch die subjektive Seite des Verschuldens erfüllt sein, damit die Handlungen des Systems diesem Menschen zugerechnet werden können.[35]]

Auf der subjektiven Seite muss die Person in der Lage sein, die Folgen ihres Handelns zu verstehen und anders zu handeln.[36] Dieser Aspekt kann von Bedeutung sein, wenn eine Person ein KI-Produkt verwendet, ohne in der Lage zu sein, die vom KI-System vorgeschlagenen Optionen kritisch zu bewerten. Ist dies der Fall, ist die Person dann nicht in der Lage, den Schadenspotenzial zu erkennen. Dies gilt jedoch nur, solange diese Situation nicht selbst verschuldet ist (sogenanntes Übernahmeverschulden), indem die Person eine Tätigkeit übernimmt, für die sie nicht ausreichend qualifiziert ist.[37]

Auf der objektiven Seite muss der Nutzer die Sorgfalt anwenden, die ein vernünftiger Dritter in der gleichen Situation aufgebracht hätte, um diesen vorhersehbaren Schaden zu vermeiden. Der Nutzer müsste also bei der Auswahl der Lösung und des Anbieters sowie bei der Anwendung der KI-basierten Lösung angemessene Sorgfalt anwenden. Dazu gehören unter anderem eine Risikoanalyse der Nutzung des KI-basierten Systems und die Umsetzung der sich daraus ergebenden angemessenen Massnahmen, die sehr unterschiedlich sein können, wenn es sich bei dem Nutzer um einen Verbraucher oder ein grosses Unternehmen handelt.[38] Wichtig ist, dass der Nutzer kein KI- oder IT-Spezialist sein muss, wenn er auf externe Lösungen zurückgreift. Er muss jedoch in der Lage sein, dem KI-Anbieter seine Anforderungen so weit zu erläutern, dass dieser die erforderlichen technischen Schritte einleiten und eine angemessene Schulung gewährleisten kann.[39] Da der Nutzer in der Entscheidungsposition sein muss, muss er auch sicherstellen, dass er in der Lage sei, entweder die Korrektheit der Funktionsweise des KI-Systems zu bewerten oder über einen geeigneten Sicherheitsrahmen zu verfügen, um eventuelle Schäden zu verhindern. Ist das nicht gegeben, trifft ihm möglicherweise ein Übernahmeverschulden.[40]

Diese Anforderungen können schwer zu erfüllen sein, wenn ein KI-System implementiert wird, das besser als

[33] BGE 134 IV 193 E. 7.2 f.
[34] LANDOLT (Fn. 17), 5.
[35] LANDOLT (Fn. 17), S. 11.
[36] Idem.
[37] Urteil des BGer 6B_1341/2015 vom 25.02.2016 E. 4.3.3.
[38] Die hier angesprochene Risikoanalyse ist eine Ausprägung der objektiven Sorgfaltspflicht, die eine den Umständen angemessene Abwägung der Folgen des eigenen Handelns erfordert (siehe Urteil des BGer 6B_1341/2015 vom 25.02.2016 E. 4.3.1.).
[39] Idem.
[40] Urteil des BGer 6B_1341/2015 vom 25.02.2016 E. 4.3.3.

ein Mensch die beste Lösung finden soll, für das aber keine verständliche Erklärung vorliegt. In diesen Fällen sollte der Mensch als Entscheider eingesetzt werden, um Fehler im KI-System zu korrigieren, die für einen Menschen offensichtlich sind, aber nicht für die Maschine. In diesen Fällen entscheidet sie dann, ob die vorgeschlagene Lösung akzeptabel ist. Die Frage, inwieweit das Vertrauen in eine Maschine entlastend wirkt, hängt meines Erachtens sowohl von der nachgewiesenen Effektivität des Systems als auch von den angewandten Sicherheitsmassnahmen, einschliesslich der Qualifikation des Entscheidungsträgers, ab. Je grösser das Risiko einer Fehlentscheidung ist, desto höher sind die Anforderungen an die überwachende Person.

2. b. human on the loop

Eine ähnliche Situation ergibt sich, wenn eine Person nur eine Überwachungsfunktion hat, aber hier wird die Person oft nicht eingesetzt, um alle Einzelentscheidungen des Systems zu überprüfen, sondern um eine von der beabsichtigten abweichenden Situation zu erkennen und einzugreifen, um möglichen Schaden durch eine Fehlentscheidung zu vermeiden. Die Person muss also nicht im gleichen Bereich wie das KI-System kompetent sein, sondern nur in der Lage sein, schädliches Verhalten oder Situationen mit Schadenspotenzial zu erkennen und entsprechend einzugreifen.

3. c. human off the loop

Ist das KI-gestützte System völlig autonom, muss der Nutzer objektiv alle angemessenen Sicherheitsmassnahmen ergriffen haben, um Schaden zu vermeiden. Wie diese Massnahmen konkret aussehen, hängt von den Umständen der Nutzung ab.

Auf der subjektiven Seite muss die Person in der Lage sein, die Folgen ihres Handelns zu verstehen und anders zu handeln. Dieser Aspekt kann relevant sein, wenn eine Person ein KI-Produkt nutzt und nicht erwartet werden kann, dass diese Person eine Fehlfunktion des KI-gestützten Systems in Kauf nimmt. Dies könnte zum Beispiel der Fall sein, wenn eine Drohne mit einer KI-Steuerungsfunktion ausgestattet ist und gegen eine Person fliegt, weil die Drohne diese nicht als Hindernis erkennt. Ist dies der Fall, kommt oft entweder die Produkthaftung oder ein anderes spezielles Haftungsrecht ins Spiel, das das Risiko eines körperlichen Schadens anders als die normale Deliktshaftung zuweist.

d. Verschulden des Anbieters

Natürlich kann nicht nur der Nutzer schuld sein, sondern auch der Anbieter, der die Verantwortung für die Entwicklung der KI-Lösung übernimmt. Der Anbieter muss sicherstellen, dass bei der Entwicklung der KI-Lösung mit angemessener Sorgfalt vorgegangen wurde.

Der Grad der objektiv anzuwendenden Sorgfalt hängt primär von dem vorgesehenen Anwendungszweck sowie den versprochenen Eigenschaften ab.

Da es sich bei der KI noch um eine neue und sich schnell entwickelnde Wissenschaft handelt, ist es schwierig, Best Practices zu ermitteln, die in einem bestimmten Fall angewendet werden können und als Grundlage für die Bestimmung einer ausreichenden Sorgfaltspflicht bei der Entwicklung dienen.[41] Da es sich bei KI um Software handelt, können etablierte Standards wie die ISO-Normen als Hilfsmittel verwendet werden.[42] Derzeit wird unter anderem in der EU eine umfangreiche Diskussion über die Regulierung von KI geführt. Vor allem wurde eine Regulierung auf europäischer Ebene vorgeschlagen, die Grundsätze für die Anforderungen an eine ausreichende Sorgfalt festlegen soll. Flankiert wurde dies durch ethische Leitlinien, die ethische Anforderungen an die Entwicklung von KI stellen. Es gibt bereits Normen für KI-Technologien und andere Normungsinitiativen, wie die ISO-Initiative, die sich mit KI befassen.[43] Bei all diesen Standards stellt sich dann die Frage, da diese oft keine Gesetzeskraft haben, wie schnell und in welchem Umfang ein KI-Anbieter oder -Entwickler diese Anforderungen umsetzen sollte, da deren Verfolgung und Umsetzung mit einem hohen Aufwand verbunden ist, welcher wiederum die Entwicklung und Anwendung von KI-basierten Systemen behindern kann. Auch hier handelt es sich um eine Werturteilsfrage, die von den jeweiligen Umständen abhängt.

Generell müsste der Anbieter der KI-Lösung die gebotene Sorgfalt walten lassen, um sicherzustellen, dass er bei der Entwicklung seines Tools die damals geltenden (und heute geltenden, wenn er für die Wartung verantwortlich ist, was heutzutage eher die Regel als die Ausnahme ist) Best Practices befolgt hat, so dass die KI-Anwendung nicht in gefährlicher Weise fehlerhaft ist und dass er alle angemessenen Sicherheitsmassnahmen ergriffen hat, um das Schadenspotenzial eines Fehlers zu vermindern.

Es sei darauf hingewiesen, dass bei der Entwicklung von KI-basierten Lösungen zwei Annahmen mit diametral entgegengesetzten Auswirkungen miteinander konkurrieren: Einerseits ist der reine KI-Algorithmus, wie gesehen, von Natur aus fehleranfällig, was sehr oft anerkannt wird, und daher werden Sicherheitsmassnahmen wie stark programmierte oder physische Grenzen eingesetzt. Eine Anwendung des Gefahrensminderungsprinzips, das darauf abzielt, alle Gefahren zu vermeiden, stösst jedoch auf die Tatsache, dass es unmöglich ist, alle möglichen Gefahren zu vermeiden, um praktikable und nützliche Lösungen zu bieten. Daher muss der Grundsatz des zulässigen Risikos berücksichtigt werden. Entscheidend ist in diesen Fällen eine Wertfrage, die nicht im Voraus beantwortet werden kann.

Die Frage nach der schuldhaften Fehlerhaftigkeit einer KI-Anwendung ist mit verschiedenen Problemen verbunden. Zum einen hängt das zu beweisende Verschulden

[41] LOHMANN (Fn. 3), 117.

[42] Z.B. ISO/IEC/IEEE 12207:2017 «Systems and software engineering – Software life cycle processes».

[43] 43 Siehe diesbezüglich NATIVI STEFANO, AI Watch: AI Standardisation Landscape, Luxembourg, 2021, 20 ff.

vom Schaden ab. So ist in der Deliktshaftung aufgrund der Begrenzung der gedeckten Schäden ein Mangel nur dann relevant, wenn er kausal für den Schaden ist, was auch eine bestimmungsgemässe Einsetzung voraussetzt. [44] Zum anderen, wie bereits erwähnt, liegt der Erfolg von KI-Anwendungen darin, dass sie aus den Lerndaten eine allgemeingültige Lösung lernen und diese flexibel einsetzen. Die Erfahrung lehrt jedoch, dass es immer wieder Ausnahmefälle geben wird, so dass sich Fehler nie ganz vermeiden lassen. Eine KI-Anwendung kann unter verschiedenen Aspekten als fehlerhaft angesehen werden.

Ein erster Aspekt, der für die Bestimmung der Fehlerhaftigkeit eines KI-Systems benutzt werden kann, ist die allgemeine Erfolgsquote bei der Erledigung der zugewiesenen Aufgabe: Diese genaue Erfolgsquote, um eine eventuelle Fehlerhaftigkeit zu bestimmen, hängt von den Umständen ab und kann nicht objektiv festgelegt, sondern muss relativ betrachtet werden. In der Diskussion um KI wird oft der Mensch als Massstab genommen. Doch welcher Mensch oder welche Gruppe von Menschen soll als Massstab genommen werden? Und sollte die Messlatte für KI-Anwendungen höher angesetzt werden als die für Menschen? Und wenn man bedenkt, dass kein Mensch perfekt ist und jeder Fehler macht,[45] ist dann der Mensch der richtige Massstab für diese Messung? Diese hochsensiblen ethischen Fragen hängen vom praktischen Anwendungsbereich, vom Entwicklungsstand der Technologie sowie von den auf dem Spiel stehenden rechtlichen Interessen ab. Kann man also bei der Fähigkeit eines KI-basierten Bewässerungssystems für Pflanzen in einer Wohnung weniger streng sein, so wird man bei lebenswichtigen medizinischen Geräten eine nahezu absolute Sicherheit erreichen wollen.[46]

Angenommen, dieser Benchmark wurde identifiziert und definiert, und dass man anhand verschiedener Ereignisse nachweisen kann, dass die tatsächliche Erfolgsquote niedriger ist als der festgelegte Benchmark, kann man von einer fehlerhaften KI-Lösung sprechen. Wird das System jedoch generell als nicht mangelhaft angesehen, weil jeder Fehler innerhalb der akzeptierte Fehlermarge liegt, bedeutet dies noch nicht, dass der einzelne Fehler nicht auf eine zurechenbare Pflichtverletzung zurückzuführen ist.

Ein zweiter Aspekt ist daher die spezifische Erfolgsquote: Es ist schwierig, einzelne Ereignisse innerhalb oder ausserhalb einer akzeptierten Genauigkeit einzuordnen. [47] Um die Erfolgsquote eines KI-basierten Systems zu testen, muss man über die notwendigen statistischen Daten verfügen, und wenn diese fehlen, muss man eine genauere Pflichtverletzung nachweisen. Wie bereits erwähnt, hat sich in den letzten Jahren das Thema *«explainable AI»* verbreitet, so dass Techniken und Dienste entwickelt werden, die versuchen, die wichtigsten Faktoren für die Entscheidung einer KI und deren Funktionsweise transparent zu machen.[48] Diese müssen jedoch oft im Voraus umgesetzt werden, da ein nachträglicher Einsatz nicht immer möglich ist. Ein einzelner Fehler kann dann das Ergebnis eines Problems in den Lerndaten sein, aus denen das KI-System eine falsche Regel gelernt hat (dies wird als *«bias»* bezeichnet). Wenn nachgewiesen wird, dass das schädliche Ergebnis die Folge der Einspeisung schlechter Daten ist (z.B. weil bei der Auswahl und Aufbereitung dieser Daten nicht mit der erforderlichen Sorgfalt vorgegangen wurde oder weil jemand verfälschende Daten eingespeist hat, ohne dass angemessene Massnahmen dagegen ergriefen wurde), kann die Verantwortung auf der Grundlage dieser Fehler zugewiesen werden. Es sei darauf hingewiesen, dass der KI-Anbieter dafür verantwortlich gemacht werden kann, dass er es versäumt hat, die erforderlichen Sorgfaltsmassnahmen zur Vermeidung fehlerhafter Daten zu ergreifen, wenn das schädigende Ereignis aus der Anhäufung verschiedener Datensätze aus unterschiedlichen Quellen resultiert, die für sich genommen das Modell nicht ausreichend beeinflusst haben können.[49]

Das Gleiche gilt, wenn ein Fehler im Code gefunden wird. Hier ist zu prüfen, ob Mängel in der Entwicklung der KI-Architektur (z.B. werden falsche Signale herausgefiltert oder falsch gewichtet) oder in der Codierung in einer Software oder einem Prozess, in den die KI-Lösung eingebettet ist (z.B. in Ermangelung zumutbarer Sicherheitsgrenzen), vorliegen.

Um die genauen Kriterien für die Bestimmung der praktischen Anforderungen an das KI-System und die Personen, die mit ihm interagieren, herauszuarbeiten, ist es wichtig, sich daran zu erinnern, dass der Einsatz dieser KI-basierten Lösungen nicht in einem rechtsfreien Raum erfolgt, so dass es je nach Anwendungsbereich häufig bereits entwickelte technologieneutrale Schutzmassnahmen und Sorgfaltspflichten gibt, die die Zufügung von Schaden verhindern sollen. Die Einzigartigkeit einer KI-Lösung in dieser Hinsicht besteht darin, dass die anwendbaren Sorgfaltspflichten oft schon vor dem Einsatz der KI-Lösung, d.h. während der Entwicklung und Planung, und nicht nur erst während des Betriebs des Systems berücksichtigt werden müssen, da seine Auswirkungen weit verbreitet und daher auffäliger sind als bei früheren Methoden.

B. Produkthaftung

Schäden an physischen Rechtsgütern durch ein KI-gestütztes System fallen häufig unter die Produkthaftung, da die KI verkörpert sein muss, um ein solches Ergebnis zu erzielen. Charakteristisch für diese Haftung ist, dass sie im Rahmen des Produkthaftungsrechts nicht beschränkt werden kann.[50]

Anders als bei der Deliktshaftung muss jedoch statt eines Verschuldens nur ein Fehler des Produkts nach Art. 4

---

[44] Vgl. GIANNI FRÖHLICH-BLEUER, Softwareverträge, Bern 2014, Rz. 2256.

[45] BGE 105 II 284 E. 1.

[46] Vgl. WENDEHORST (Fn. 22), 35 ff.

[47] WENDEHORST (Fn. 22), 82 f. z.B., ob eine einzelne Fehlentscheidung Teil der 1%igen Fehlerquote ist oder ob sie auf etwas anderes zurückzuführen ist.

[48] Für einen technischen Überblick des aktuellen Stands: DAS/RAD (Fn. 5).

[49] Urteil des BGer 4A_606/2017 vom 30. April 2018 E. 3.1.1.

[50] Art. 8 PrHG.

Produkthaftungsgesetz nachgewiesen werden. Der Schöpfer einer KI-Lösung haftet daher, wie von einem grossen Teil der Lehre vertreten, oft als Schöpfer nach dem Produkthaftungsgesetz für Sach- und Personenschäden.[51] Dies ist sicherlich der Fall, wenn die KI ein Bestandteil einer beweglichen Sache ist. Schwieriger wird es, wenn die KI-basierte Software und die Sache, auf der die KI läuft, voneinander getrennt sind, so dass der Hersteller der Sache keine Verantwortung für die Softwarekomponente tragen kann. Das ist zunehmend bei dem aktuellen Trend zu beobachten, physische Plattformen anzubieten, für die Drittanbieter ihre Softwarekomponenten entwickeln und anbieten können.[52]Obwohl umstritten und trotz des Verbesserungspotenzials bei der Regulierung, sieht man in der Lehre eine Befürwortung, auch reine Software als Produkt zu betrachten.[53]

Zu bedenken ist, dass das Produkthaftungsgesetz die Haftung für Sachschäden auf privat genutzte Gegenstände beschränkt, so dass im Falle von KI-basierten Systemen, die für geschäftliche Zwecke genutzt werden, nur der Tod oder die Verletzung einer Person durch das Produkthaftungsgesetz abgedeckt ist.[54] Diese Fragen werden jedoch in der Regel im Vertrag zwischen dem Anbieter und dem Nutzer geregelt.

C. Vertragliche Haftung

Wie bereits erwähnt, wird ein KI-System in den meisten Fällen auch zur Erfüllung vertraglicher Verpflichtungen eingesetzt oder ist sogar Gegenstand eines Vertrages. Die Erscheinungsformen eines solchen Vertrages sind sehr unterschiedlich, und jede von ihnen kann besondere Fragen bei der Nutzung einer KI aufwerfen.

Wir werden hier nur eine besondere Erscheinungsform des Einsatzes von KI bei der Erfüllung einer vertraglichen Verpflichtung untersuchen, die meines Erachtens aufgrund ihrer Aktualität und der Art der aufgeworfenen Fragen ein gutes Beispiel für die Haftungsfragen beim Einsatz von KI bei der Erfüllung eines Vertrags ist.

Nehmen wir konkret an, dass ein Facharzt eine falsche Diagnose stellt, weil er sich in einem zweifelhaften Fall auf die Empfehlung einer KI-Software verlassen hat. Hier ist der Arzt dem Patienten gegenüber vertraglich verantwortlich. Der Vertrag zwischen Patienten und Arzt ist ein Mandat, so dass die ordnungsgemässe Erfüllung des Vertrags die Anwendung der geschuldeten Sorgfaltspflicht erfordert.[55]

1. Als KI nutzender Dienstleister

Die erste Frage, die sich stellen lässt, ist, ob die Verwendung einer KI-Software bei der Erfüllung des Mandats gegen die erforderliche Sorgfaltspflicht verstösst. Es ist inzwischen allgemein anerkannt, dass KI auch im medizinischen Bereich grosse Chancen bietet, und die Krankenhäuser setzen diese Technologie zunehmend ein, sowohl bei der Diagnose als auch bei der Ermittlung der besten Therapie.[56] Deontologisch gesehen ist ein Arzt verpflichtet, eine andere Meinung einzuholen, wenn er der Meinung ist, dass er die «Grenzen der ärztlichen Leistungsfähigkeit» erreicht hat und das Wohl des Patienten dies erfordert.[57] KI-gestützte Diagnosesysteme werden inzwischen häufig als praktische Zweitmeinung angesehen, so dass ihr Einsatz – vorausgesetzt, das System wurde sorgfältig konzipiert und der Nutzer ist sich seiner Grenzen bewusst – zumindest nicht gegen die Sorgfaltspflicht verstösst.

Die diametrale Frage, ob der Einsatz eines solchen Instruments erforderlich ist, ist schwieriger zu beantworten und kann sich aufgrund der Geschwindigkeit des Fortschritts in diesem Bereich rasch ändern. Das erforderliche Mass an Sorgfalt hängt insbesondere von den Umständen ab, wie z.B. der Art der Behandlung, den damit verbundenen Risiken, dem Grad der Ermessensfreiheit, den Mitteln und der Zeit, die dem Arzt im Einzelfall zur Verfügung stehen, sowie von der Ausbildung und den Fähigkeiten des Arztes. Aufgrund der Fortbildungspflicht für Ärzte,[58] des gestiegenen Bedarfs an Unterstützung bei der Analyse von Patientendaten sowie des grossen Interesses an KI-Technologien werden wir uns meines Erachtens zunehmend in die Richtung bewegen, die effizientesten und effektivste dieser Anwendungen als Grundausstattung für den Facharzt zu verlangen. Man kann davon ausgehen, dass der Arzt über ein sorgfältig ausgewähltes KI-gestütztes Werkzeug verfügt, und wenn es nicht genutzt wird, könnte das ein Hinweis auf eine Sorgfaltspflichtverletzung sein. Je nach Verlässlichkeit des Systems wäre der Substantiierungsbedarf bei der Diagnose höher, wenn der Arzt gegen den Rat des Systems gehandelt hat. Das aber nur, wenn der Rat des KI-Systems als Gutachten betrachtet werden kann. Nach der Rechtsprechung haben Berichte und Gutachten Beweiskraft, wenn sie schlüssig, nachvollziehbar begründet und widerspruchsfrei erscheinen und keine Anhaltspunkte gegen ihre Zuverlässigkeit vorliegen.[59] Aufgrund dieser Substantiierungspflicht könnte der Einsatz eines KI-basierten Systems die Anforderungen an die Substantiierung einer Diagnose nur insoweit erhöhen, als das System sein Ergebnis nachvollziehbar erklären kann, da diese Erklärung dann in die Substantiierung der Diagnose einfliessen sollte.[60] Dies ist heute oft nicht der Fall,

---

[51] Art. 1 PrHG.

[52] WENDEHORST (Fn. 22), 65; LOHMANN (Fn. 3), 115.

[53] LOHMANN (Fn. 3), 115; FRÖHLICH-BLEUER (Fn. 43), Rz. 2251 ff.; WENDEHORST (Fn. 22), 65 f.

[54] Art. 1 PrHG.

[55] Urteil des BGer 4A_432/2020 vom 16. Dezember 2020 E. 6.2.

[56] «*Ein wichtiger Bereich ist die Bildgebung, also Röntgen-, CT- oder MRI-Bilder. Hier geht es oft darum, in vielen Schnittbildern eines Organs die «Nadel im Heuhaufen» zu sehen. Heute kann man eine Software darauf trainieren, dass sie ähnlich gut wie erfahrene Radiologinnen solche Bilder einschätzt und mögliche Problemstellen erkennt. Weiter kann KI bei chronischen Erkrankungen eingesetzt werden, also bei Krankheiten, die nicht heilbar sind, aber bei denen mit der richtigen Therapie eine rasche Verschlimmerung vermieden werden kann.*», <www.caim.unibe.ch/unibe/portal/fak_medizin/dept_zentren/inst_caim/content/e998130/e998135/e1025836/e1097065/Gluckspost_KIundArztalsPower-Duo_RaphaelSznitman_Juni2021_eng.pdf>, besucht am 22.9.2021.

[57] Art. 10 Standesordnung der FMH.

[58] Art. 40 MedBG und Art. 9 Fortbildungsordnung.

[59] Urteil des BGer 8C_608/2015 vom 17. Dezember 2015 E. 3.3.3.

[60] Urteil des BGer 9C_195/2015 vom 24. November 2015 E. 3.3.1.

was auch die Verzögerung bei der Anwendung von KI-basierten Diagnosesystemen erklärt. Fehlt eine Erklärung seitens des KI-Systems, ist der Arzt meines Erachtens nur verpflichtet, sich der Richtigkeit seiner Schlussfolgerung zu vergewissern (ein sogenannter Double Check), um eine ausreichende Sorgfalt nachweisen zu können.

Diese Überlegung stösst an ihre Grenze, wenn man den Fall der Unerklärbarkeit des Ergebnisses eines KI-Systems erreicht. Damit ist die Situation gemeint, in der das KI-System die Richtigkeit seiner Schlussfolgerung zwar prinzipiell erklären kann, diese Erklärung aber so kompliziert ist, dass der durchschnittliche Fachmann sie nicht verstehen kann. Der Anwender wäre dann nicht mehr in der Lage, dem KI-System zu widersprechen oder mögliche Fehler zu erkennen. Meines Erachtens ist eine solche Konstellation mit einem Medizinprodukt vergleichbar, für dessen Anwendung ein Zulassungsverfahren erforderlich ist. Dieses Verfahren kann zusätzliche, von der traditionellen ärztlichen Kunst abweichende Sicherheitsmassnahmen erfordern, die aber erfüllt werden müssen, um fehlerhafte Ergebnisse seitens des KI-Systems zu vermeiden.

Obwohl also der Einsatz von KI heute in der Regel weder einen Beweis für eine angemessene Sorgfalt per se noch eine Sorgfaltspflichtverletzung darstellt, wäre im Einzelfall zu prüfen, ob die Fehldiagnose auf einer Sorgfaltspflichtverletzung beruht, weil im Einzelfall den Einsatz eines KI-basierten Unterstützungssystems seitens des Arztes erfordert sein könnte und damit auch erhöhte Rechtfertigungsanforderungen an die Diagnose stellen könnte.

Wie der Anbieter eines KI-gestützten Produkts oder Dienstes für das Ergebnis und etwaige Schäden haftet, ist dann eine Frage der Auslegung des zugrunde liegenden Vertrags zwischen Anbieter und Nutzer.

2. Als KI-Anbieter

Wird die KI-basierte Lösung dem Dienstleister im Rahmen eines Kauf- oder Dienstleistungsvertrags zur Verfügung gestellt, haftet der Dienstleister oder Nutzer in erster Linie für die Ergebnisse der Anwendung gegenüber seinen eigenen Vertragspartnern.

Wird nur die Software verkauft, so haftet der Verkäufer sowohl für das Vorhandensein zugesicherter Eigenschaften als auch dafür, dass der Kaufgegenstand keine seine Brauchbarkeit beeinträchtigenden Mängel aufweist. Da aber, wie erwähnt, KI-Systeme aus Daten lernen, wird man, wenn man ein lernfähiges KISystem kauft, grundsätzlich für die Eingabe der Daten und deren Folgen verantwortlich. Eine Verantwortung des KI-Anbieters für erlerntes Verhalten wird dann nur schwer zu begründen sein, es sei denn, dies wurde im Vertrag ausdrücklich vereinbart oder war nach dem Stand der Technik zum Zeitpunkt der Entwicklung des Systems anders zu erwarten.

Aufgrund dieser Probleme sowie anderer Besonderheiten der modernen Softwareentwicklungspraxis im Allgemeinen wird eine KI-Lösung häufig als Dienstleistung angeboten (sog. Software as a Service, oder *«SaaS»*).[61] Der KI-Anbieter ist dann in der Lage, die erstellte Lösung ständig weiterzuentwickeln und das Training der KI auf der Grundlage einer grösseren Datenmenge zu gewährleisten. Die Verantwortung für eine Fehlerhaftigkeit des KI-Systems liegt dann beim Anbieter, sofern nichts anderes vereinbart wurde.

III. Fazit

Die Zurechnung der Haftung im Schadensfall bei einem KI-basierten System wirft aufgrund der besonderen Ausgestaltung dieser Technologie verschiedene Probleme bei der Ermittlung der anwendbaren Sorgfaltspflichten auf. Aufgrund der flexibleren Anwendungsmöglichkeit, die mit einem inhärenten Risiko verbunden ist, sind beim Einsatz einer KI-basierten Lösung neue besondere Sorgfaltsmassnahmen seitens des Nutzers erforderlich, da dieser andernfalls nach der Doktrin des Gefahrensatzes haftbar gemacht werden könnte. Hat der Nutzer die zumutbare Sorgfalt angewendet, kann das Verschulden nicht nur in der Kodierung der KI-basierten Lösung, sondern auch in der Auswahl und Aufbereitung der für ihr Training verwendeten Daten liegen. Trotz der Fortschritte in diesem Bereich ist es sehr schwierig, in diesen Fällen ein Verschulden nachzuweisen.

[61] WENDEHORST (Fn. 22), 81.