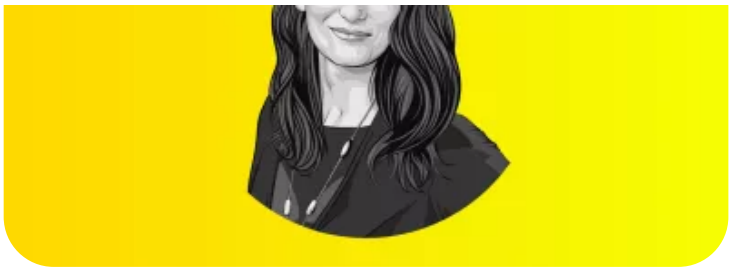
neuen Regulierungen und Gesetzen in der Schweiz und der EU münden dürften. Isabelle Wildhaber, Rechtsprofessorin an der Universität St. Gallen, klärt über die aktuellen Gegebenheiten auf und erläutert, was dies für Unternehmen bedeutet.



MIT ISABELLE WILDHABER SPRACH OLIVER BOSSE

Frau Wildhaber, zunächst mal ganz grundsätzlich gefragt: Warum ist es gerade zum jetzigen Zeitpunkt wichtig, sich mit rechtlichen Fragestellungen in Bezug auf künstliche Intelligenz auseinanderzusetzen?

Die US-Regierungsbehörde OSTP schlägt eine Bill of Rights für das KI-Zeitalter vor, die EU-Kommission macht Vorschläge für europäische KI-Regelungen. Es ist deshalb wichtig zu diskutieren, welche neuen Regelungen in Bezug auf KI notwendig und sinnvoll sind und wie diese ausgestaltet sein sollen. Einerseits wollen wir als Gesellschaft Risiken möglichst ausschliessen oder zumindest vermindern, andererseits die Innovationskraft in Unternehmen nicht schon im Keim ersticken. Da neue KI-Produkte und KI-Dienstleistungen gerade täglich wie Pilze aus dem Boden schiessen, ist spätestens jetzt zweifelsohne der richtige Moment, um sich mit KI-Regulierungen kritisch auseinanderzusetzen.

Können Sie ein Beispiel einer rechtlichen Frage- und Problemstellung nennen, die durch KI entsteht?

Eine relevante und ernstzunehmende Problematik ist **die algorithmische**

Diskriminierung (https://www.trivadis.com/de/magazine/bias-bei-k%C3%BCnstlicher-intelligenz-verhindern). Der Hype um das Potential von KI führt bei vielen Menschen und Unternehmen zur Überzeugung, dass Daten stets die objektive Wahrheit reflektieren würden. Das stimmt aber keinesfalls. Im Gegenteil: Ein unkritischer Datenglaube ist gefährlich, denn Daten werden von Menschen erhoben und interpretiert – und daraus können sich Diskriminierungen ergeben. Schon die Definition des Gewünschten, z.B. was unter einer guten Arbeitsleistung zu verstehen ist, wird vom Menschen bestimmt.

Es gibt viele Ursachen für algorithmische Diskriminierungen. Beispielsweise können die Daten schlecht ausgewählt, falsch oder veraltet sein. Wenn z.B. der Gesichtserkennungsalgorithmus von Google fälschlicherweise schwarze Menschen als Gorillas bezeichnete, lag dies an einem unzureichenden Training mit schwarzen Gesichtern. Häufig fliessen Diskriminierungen unbeabsichtigt, als Produkte unbewusster menschlicher Verzerrungen oder als einfache Fehler, in den algorithmischen Prozess ein.

Die KI einer Versicherung könnte z.B. Übergewichtige aussortieren und so diskriminieren. Rechtlich gesehen bilden Übergewichtige aber keine geschützte Gruppe.

Kann man solche Diskriminierungen nicht mit Diskriminierungsschutzgesetzen verhindern?

Rechtlich gesehen ist es leider sehr schwierig, algorithmische Diskriminierungen zu erfassen. Das hat verschiedene Gründe. Eine Diskriminierung ist schwierig zu beweisen und es gibt kaum abschreckende Sanktionen gegen diskriminierendes Verhalten. Ausserdem werden in den meisten Rechtsordnungen nur punktuell bestimmte Gruppenzugehörigkeiten geschützt, also z.B. werden Individuen aufgrund des Geschlechts oder einer Behinderung geschützt. Die KI identifiziert und kategorisiert Personen aber anhand von ganz anderen Kriterien. So könnten z.B. Übergewichtige von einer Versicherung aussortiert werden. Übergewichtige gehören aber nicht zu einer Gruppe, die gesetzlich geschützt ist.

Und warum ist die Frage der Haftung in aller Munde?

Im Hinblick auf die Haftung müssen wir uns fragen, ob die KI gewisse Risiken schafft, die mit unseren derzeitigen Haftungsregeln nicht angemessen behandelt werden können. Diese Frage stellt sich bei vielen neuen Technologien, so in der Vergangenheit z.B. schon in den 90er-Jahren bei der Biotechnologie oder in den 70er-Jahren bei der Atomenergie. Gross ist jeweils die Sorge vor Kontrollverlust und unabsehbaren, nicht zurechenbaren Schäden.

Bei der KI-Haftungsdiskussion ist der Ausgangspunkt die Auseinandersetzung mit den typischen Risiken der KI. Besonders die drei Risiken der Blackbox, der Komplexität und der Autonomie von KI-Systemen sind Anlass zur Beunruhigung. Denn sie können bei der Durchsetzung haftpflichtrechtlicher Ansprüche Probleme bereiten.

WIE GEHT KI-RISIKOMANAGEMENT?

Wer künstliche Intelligenz nutzen möchte, sollte sich auch ihrer Risiken bewusst sein beziehungsweise wissen, wie man mit diesen umzugehen hat. Mehr darüber erfährst du in folgenden Blog-Beiträgen:

- **Nutzen und Risiken von KI (https://financialservicesblog.accenture.com/the-benefits-and-risks-of-ai)**

- **KI-Regulierungsreise (https://financialservicesblog.accenture.com/ai-regulatory-journey)**

- **Die Herausforderungen der KI meistern (https://financialservicesblog.accenture.com/navigating-the-challenges-of-ai)**

Ist die «Blackbox» bezüglich Haftung das Hauptproblem?
Ja, mit Blick auf die Haftung ist das Blackboxrisiko ein grosses Problem. Das Blackboxrisiko kann daraus entstehen, dass KI-Systeme ihre Leistung durch das Lernen aus Erfahrung verbessern können. Nur Input und Output sind bekannt, während das Modell undurchsichtig bleibt. Das kann dazu führen, dass die KI eine Opazität bzw. einen Mangel an Transparenz aufweist.

Der Blackbox-Effekt bewirkt eine eingeschränkte Erklärbarkeit und Nachvollziehbarkeit der Entscheidungsfindung von selbstlernenden Systemen. Menschen haben Schwierigkeiten, genau zu verstehen, wie die Eingaben zu den Ausgaben von KI-Systemen führen. Warum kriegt man keinen Kredit, obwohl man alle Voraussetzungen erfüllt? Geschädigte sind möglicherweise nicht in der Lage, zu erkennen, was genau den Schaden verursacht hat. Für den Nachweis von Fehlfunktionen in KI-Systemen und für die Durchsetzung daraus resultierender Haftungsansprüche sind die Transparenz und Nachvollziehbarkeit von Kausalverläufen aber erforderlich. Die Eigenschaft als Blackbox stellt deshalb eine Herausforderung für die Haftung dar.

Würde es reichen, eine Explainable Artificial Intelligence (XAI) zu haben, um

das Problem der «Blackbox» zu beheben?

Die Machine-Learning-Community arbeitet seit Längerem daran, das Problem der mangelnden Nachvollziehbarkeit algorithmischer Entscheidungen zu lösen, etwa im Rahmen des 2016 lancierten Projekts zu Explainable AI der amerikanischen Defense Advanced Research Projects Agency (DARPA). Unter dem Titel XAI wird diskutiert, wie undurchsichtige Machine-Learning-Modelle erklärbar gemacht werden können. Dem Nutzer sollen konkret einzelne Entscheidungsschritte erläutert, die allgemeinen Stärken und Schwächen des Modells vermittelt und ein Verständnis dafür erweckt werden, wie sich das System in Zukunft verhalten wird. Aber nicht jede Form der Erklärung ist technisch umsetzbar oder zielführend.

Um die Probleme intransparenter KI für die Haftung zu verringern, fordern auch Wissenschaft und politische Entscheidungsträger seit ein paar Jahren, Algorithmen erklärbar zu machen. Dafür werden Transparenzanforderungen eingeführt, z.B. in der europäischen Datenschutz-Grundverordnung. Transparente KI-Systeme könnten die Probleme für Kläger in Haftungsfällen verringern. Es ist jedoch nicht offensichtlich, dass Transparenz die Probleme lösen kann, mit denen sich Klagende in Zivilprozessen konfrontiert sehen werden. Sie ist teilweise nur unter grossen Mühen erreichbar und sehr teuer. Deshalb kann sie das Problem der «Blackbox» nicht einfach lösen.

Transparente KI-Systeme könnten die Probleme von Klägern in Haftungsfällen verringern. In manchen Fällen ist es jedoch sehr schwierig und teuer, Transparenz zu erreichen.

Und wo genau ist entsprechend Handlungsbedarf angezeigt? Welches sind die relevanten haftpflichtrechtlichen Problemkreise?

Es gibt in meinen Augen schwerwiegende Probleme beim Nachweis des Verschuldens und der Kausalität. Diese Problemkreise führen dazu, dass die Haftung für KI sehr breit diskutiert wird und auf europäischer Ebene die Forderung nach einer harmonisierten Haftungsregelung für KI laut geworden ist.

Schauen wir uns z.B. die Probleme rund um den Verschuldensnachweis an. Das Verschulden besteht in einer Sorgfaltspflichtverletzung beim Einsatz, beim Trainieren, bei der Wartung oder bei der Überwachung der Funktionsweise des KI-Systems. Der Nachweis eines solchen Verschuldens ist für Geschädigte eine riesige Hürde. Sie wissen ja in der Regel viel zu wenig, um diesen Beweis zu führen.

Es kommt hinzu, dass bis anhin die meisten Systeme lediglich entscheid-unterstützend sind, sie ersetzen die menschliche Entscheidfindung noch nicht. Wenn der Mensch das System noch überwachen muss, muss die Schnittstelle zwischen Menschen und Maschine gut funktionieren. Was dürfen aber Menschen von einem KI-System überhaupt erwarten? Unter welchen Umständen darf sich der Betreiber auf das KI-System verlassen und an welchem Punkt sollte er dessen Vorschläge überstimmen? Wir haben in unseren empirischen Studien zu KI-Systemen am Arbeitsplatz festgestellt, dass Betreiber das algorithmische System nur selten übersteuern würden.

Soll Tesla als Hersteller haften oder soll der Fahrer haften, wenn er das Auto nicht richtig gewartet hat?

Derzeit prüfen verschiedene Staaten aber auch die EU (Artificial Intelligence Act), inwiefern im Zusammenhang mit KI neue Vorschriften nötig sein könnten. Was ist diesbezüglich der aktuelle Stand der Dinge in der EU?
Anders als in der Schweiz wird in der EU die Haftung für KI schon seit mehreren Jahre hoch und runter diskutiert. Die EU strebt grundsätzlich ein spezielles, sog. horizontales Haftungsregime (siehe Box) für definierte hochriskante KI-Anwendungen an.

Am 28. September 2022 hat die Europäische Kommission ihre Richtlinienvorschläge zur KI-Haftung publiziert. Es gibt zwei Säulen der Haftung. Es gibt erstens die Herstellerhaftung. Hier macht die Europäische Kommission einen Vorschlag einer neuen Produkthaftungsrichtlinie, welche die heutige Richtlinie 85/374/EWG ersetzen soll. Und zweitens gibt es die Betreiberhaftung, also die Haftung von Nutzern, Eigentümern oder Haltern von KI. Hier sieht die Europäische Kommission einen Vorschlag für eine Richtlinie zur Anpassung der Vorschriften über ausservertragliche zivilrechtliche KI-Betreiberhaftung vor. Es ist wichtig, die Haftung von Hersteller und Betreiber angemessen zu verteilen und die jeweiligen Risikosphären abzugrenzen. Soll Tesla als Hersteller haften oder soll der Fahrer haften, wenn er das Auto nicht richtig gewartet hat? Wenn auch Betreiber haften, so müssen sie sich mit der genutzten KI auseinandersetzen. Ich finde es gut, wenn die Menschen sich kritisch mit der Nutzung von KI beschäftigen und diese hinterfragen. Eine Betreiberhaftung könnte so die Digital Literacy verbessern.

HORIZONTALER & SEKTORSPEZIFISCHER REGULIERUNGSANSATZ

Horizontaler Regulierungsansatz: Die Technologie als solche wird geregelt, egal um welchen Sektor / Branche es sich handelt. Ein KI-Medizinprodukt unterliegt dann nicht mehr den gleichen Regeln wie ein herkömmliches Medizinprodukt.

Sektorspezifischer Regulierungsansatz: Jeder Sektor / Branche wird separat geregelt. Es wird unterschieden danach, ob es sich um Luftfahrt, Strassenverkehr oder den Medizinalbereich handelt. Dieser Ansatz ist technologieneutral.

Und was ist der aktuelle Stand der Dinge in der Schweiz?

Ende 2019 erschien in der Schweiz ein Bericht der interdepartementalen Expertengruppe zu KI. Der Bericht verortete keinen regulatorischen Handlungsbedarf und erachtete das bestehende Haftungsregime für derzeit ausreichend. Diese Einschätzung unterscheidet sich damit fundamental von derjenigen auf europäischer Ebene. Es wird – wo notwendig – eine sektorspezifische Ergänzung der bestehenden Regulierung gefordert, ohne horizontalen Ansatz.

Seit 2020 sind nun aber in der Schweiz basierend auf den europäischen Diskussionen zur KI-Regulierung vermehrt Stimmen laut geworden, die diskutieren, ob vielleicht doch regulatorischer Handlungsbedarf mit Blick auf KI besteht.

In einem Bericht an den Bundesrat vom April 2022 hat das Eidg. Departement für Auswärtiges (EDA) festgestellt, dass die Schweiz zwar in ihrer nationalen Gesetzgebung eigene Akzente zum Umgang mit KI setzen könne. Ein zu grosser Unterschied zwischen internationaler und nationaler Rechtslage sei jedoch nicht im Interesse der Anschlussfähigkeit der Schweiz an internationale Märkte und Lieferketten zu KI. Man kann also gespannt sein, wie sich die Schweiz positioniert. Sie wird regulatorisch Zurückhaltung üben, will sie doch ihren Spitzenplatz in unzähligen Innovations-Rankings gern beibehalten.

Welche konkreten Lösungen werden in der EU diskutiert – und welche haben Ihrer Meinung nach das grösste Potenzial und den grössten Konsens, um einst tatsächlich eingeführt zu werden?

Die EU strebt grundsätzlich ein horizontales Haftungsregime für hochriskante KI an. In der Schweiz spricht man sich eher für einen sektorspezifischen Ansatz aus. Am horizontalen EU-Ansatz scheint aber nicht mehr viel zu rütteln sein.

Wählt man einen horizontalen Ansatz, so sind verschiedene Optionen denkbar: Man könnte die Verschuldenshaftung verschärfen. Man könnte auch eine verschuldensunabhängige Haftung für Betreiber einer ausgewählten Gruppe von KI-Systemen einführen.

In den Vorschlägen der Europäischen Kommission zu einer KI-Haftung vom 28. September 2022 wird in Art. 5 die Idee einer verschuldensunabhängigen Haftung in die Zukunft verschoben. Erst nach ein paar Jahren soll bewertet werden, ob eine solche Haftung angemessen ist, und ob ein Versicherungsschutz erforderlich ist. Da ich einer verschuldensunabhängigen Haftung kritisch gegenüberstehe, finde ich das gut.

Hingegen möchte die Europäische Kommission zwei widerlegbare Vermutungen für den Verschuldens- und Kausalitätsnachweis einführen. Das rechtliche Verfahren wird also für Opfer punkto Nachweis vereinfacht. Sie müssen nicht detailliert erklären, wie der Schaden durch ein bestimmtes Verschulden oder eine bestimmte Unterlassung verursacht wurde. Ich finde diese zwei Vermutungen begrüssenswert. Sie greifen nicht zu sehr in die nationalen Rechtsordnungen ein. Sie können helfen, die Herausforderung der Blackbox der KI bei der Haftung zu überwinden. Ich denke, es könnte sich Konsens finden, dass für Betreiber eine Haftung mit vermutetem Verschulden und Kausalitätsvermutung eingeführt werden soll, jeweils mit der Möglichkeit, die Vermutung zu widerlegen. Wie die Vermutungen genau formuliert sein sollen, kann man noch diskutieren.

KI-Systeme können nicht allgemein als besonders risikogeneigt oder gefahrerhöhend angesehen werden. Im Gegenteil: KI-Systeme werden genutzt, um menschliche Unzulänglichkeiten auszugleichen und damit Risiken und Schäden zu reduzieren.

Sind Sie mit dem horizontalen Regulierungsansatz der EU im Einklang?

Ein Kardinalproblem eines horizontalen Regulierungsansatzes besteht in der Definition des Anwendungsbereichs, das heisst in der Definition von KI einerseits und in der Definition der Risikokategorien andererseits. Genau das erweist sich als äusserst schwierig. Was genau ist denn überhaupt KI und welche KI ist tatsächlich hochriskant? Es kommt hinzu, dass ein horizontales Regime zu wenig nach Kontext differenziert, in dem eine Anwendung eingesetzt wird. Selbst wenn es mit der Definition des hohen Risikos versucht zu differenzieren, kann nicht genügend auf den Kontext abgestellt werden. Kommt das Produkt z.B. mit nicht-spezialisierten Nutzer*innen in Berührung? Wie genau wird es genutzt und wie entwickelt es sich weiter? Eine horizontale Regulierung kann ausserdem innovationshemmend wirken.

Das war z.B. beim Gentechnikgesetz in der Schweiz für die Landwirtschaft der Fall.

Gibt es auch Beispiele für rechtlich völlig unbedenkliche Einsatzgebiete?
Es gibt zahlreiche KI-Systeme, die nicht risikoreich sind, z.B. KI zur Verwaltung des Arbeitsspeichers eines Computers, bei der Terminplanung, bei der Planung von Lagerbeständen und beim Wareneinkauf. KI-Systeme können nicht allgemein als besonders risikogeneigt oder gefahrerhöhend angesehen werden. Im Gegenteil: KI-Systeme werden genutzt, um menschliche Unzulänglichkeiten (verspätete Reaktionszeiten, Übersehen wesentlicher Daten oder Informationen sowie Emotionalität) auszugleichen und damit Risiken und Schäden zu reduzieren.

Wir haben in unseren empirischen Studien zu KI am Arbeitsplatz festgestellt, dass die Einführung von KI-Tools in Schweizer Unternehmen teilweise gar nicht juristisch begleitet wird.

Was würden die aktuellen Entwicklungen in Sachen Haftung bedeuten – wer wird realistischer Weise in die Pflicht genommen?
Wichtig ist und bleibt, dass die Haftung der Kontrolle über die haftungsauslösenden Eigenschaften der jeweiligen Sache folgt. Wenn demnach der Betreiber das «Verhalten» der Sache steuert, so sollte er haften. Wenn der Schaden hingegen auf den sicherheitsrelevanten Eigenschaften der Sache beruht, sollte die Herstellerin haften. Denn sie definiert und kontrolliert diese Eigenschaften. Der zentrale Unterschied zwischen herkömmlichen Produkten und KI-Produkten besteht darin, dass bei KI-Produkten die Kontrolle über das «Verhalten» des KI-Systems in der Tendenz zur Herstellerin hin verlagert wird. Es ist die Herstellerin, welche die KI-Systeme programmiert, trainiert und Instruktionen verteilt. Rund um KI-Systeme wird deshalb in Zukunft die Herstellerin die Zentralfigur des Haftungsgeschehens sein.

Der Betreiber des KI-Systems wird aber in zahlreichen Fällen die erste sichtbare Ansprechstelle für Geschädigte sein. Ein Geschädigter wird sich zuerst an den Betreiber – der häufig auch sein Vertragspartner sein wird – und nicht an die originale Herstellerin wenden. Ein Arbeitnehmer wendet sich also zuerst an seine Arbeitgeberin. Am Arbeitsplatz soll dies auch so bleiben, da es spezielle arbeitsrechtliche Verfahrensregeln und spezielle Arbeitsgerichte gibt, die den betroffenen Arbeitnehmern zur Verfügung stehen. Ausserdem finde ich es richtig, dass sich Betreiber damit beschäftigen müssen, welche Art von KI sie

warum und wie einsetzen.

Was ist Ihre Empfehlung an Unternehmen, die heute bereits KI-Tools im Einsatz haben oder solche in Zukunft nutzen möchten – mit welchen rechtlichen Fragen sollten sie sich beschäftigen?

Wir haben in unseren empirischen Studien zu KI am Arbeitsplatz festgestellt, dass die Einführung von KI-Tools in Schweizer Unternehmen teilweise gar nicht juristisch begleitet wird. Das Tool wird in einem sog. Pilot eingeführt und dabei ist nicht zwingend, dass ein hauseigener oder externer Jurist ein Auge darauf geworfen hat. Meine Empfehlung wäre deshalb, bei der Einführung eines KI-Tools die rechtlichen Fragen – insbesondere solche datenschutzrechtlicher Natur – abklären zu lassen.

Ich würde mit dem Einsatz von KI-Tools nicht zuwarten, denn die Mühlen des Fortschritts drehen sich schnell.

Als Unternehmen mit dem Einsatz von KI zuzuwarten, könnte aus wirtschaftlicher Sicht einen Wettbewerbsnachteil bedeuten. Spricht aus rechtlicher Sicht etwas dafür, zuzuwarten?

Nein, ich würde nicht zuwarten, denn die Mühlen des Fortschritts drehen sich schnell. Ich würde aber wie gesagt empfehlen, dass vor der Einführung eines KI-Tools mögliche rechtliche Problemfelder durch eine/n Juristen/in abgeklärt werden.

Von den Risiken zu den Chancen: Wo sehen Sie diese im Bezug auf künstliche Intelligenz?

Die rasanten Fortschritte in der Künstlichen Intelligenz bieten unzählige tolle Möglichkeiten für die Gesellschaft. KI-Systeme werden in den verschiedensten Branchen eingesetzt, in der industriellen Produktion (Industrie 4.0), im Energiesektor (Smart Grid), Gesundheitssektor (MedTech), Finanzsektor (FinTech), Versicherungssektor (InsureTech) oder Rechtssektor (LegalTech) oder aber in der öffentlichen Verwaltung (E-Government), im Umwelt- und Klimaschutz, in der Landwirtschaft oder in der Bildung. Viele KI-Systeme sind nicht mehr wegzudenken und voller Chancen! Diese Chancen müssen unbedingt genutzt werden, rechtliche Probleme hin oder her.

ZUR PERSON

Prof. Dr. Isabelle Wildhaber, LL.M., ist seit 2015 ordentliche Professorin für Privat- und Wirtschaftsrecht unter besonderer Berücksichtigung des Arbeitsrechts an der Universität St. Gallen sowie geschäftsführende Direktorin am Forschungsinstitut für Arbeit (FAA-HSG). Sie ist Delegierte des Rektors für Gleichstellung, Diversität und Inklusion und Präsidentin der Gleichstellungskommission der Universität St. Gallen. Sie forscht, lehrt und berät zu Themen des Vertrags-, Haftpflicht- und Arbeitsrechts, mit Bezug zu neuen Technologien und Entwicklungen.

AUCH UNSERE EXPERT*INNEN VERMITTELN KI-WISSEN:

DATA ANALYTICS

28 FEB. 2023

Mit DAX Studio Leistungsdaten aus Power BI ziehen (/de/magazine/dax-studio-leistungsdaten-aus-power-bi)

KI-ETHIK

KI IM BUSINESS

21 FEB. 2023

TechTalk Audio: Responsible AI & ChatGPT (/de/magazine/techtalk-audio-responsible-ai-chatgpt)

DATA ANALYTICS

9 FEB. 2023

Schreib sauberen Code mit den 5 SOLID-Prinzipien (/de/magazine/schreib-sauberen-code-mit-den-5-solid-prinzipien)