

Zusammenfassung (Summary in German)

Bei Straftaten unter Beteiligung KI-gesteuerter autonomer Systeme ergeben sich erhebliche Schwierigkeiten bei der Bestimmung der Verantwortlichkeit der hinter der Maschine stehenden Person. Anstatt sich auf eine bestimmte KI-Anwendung zu konzentrieren, zielt diese Studie darauf ab, einen allgemeinen Rahmen zu schaffen, der insbesondere bei Fahrlässigkeit den Umfang und die Grenzen der Haftung festlegt und konkretisiert. In diesem Zusammenhang lassen sich folgende Beobachtungen und Erkenntnisse formulieren:

Bei der Prüfung der Haftung sollte der Schwerpunkt eher auf „Autonomie“ als auf „künstliche Intelligenz“ gelegt werden, da das wesentliche Problem in der technischen Autonomie dieser Systeme, der verringerten menschlichen Kontrolle und der Unvorhersehbarkeit ihrer Ergebnisse liegt. Neben diesen *ex ante*-Herausforderungen entstehen *ex post*-Schwierigkeiten bei der Ermittlung des Kausalzusammenhangs für die Haftung. Diese beruhen auf der Intransparenz solcher Systeme, die sich aus der Geheimhaltung von Algorithmen, dem begrenzten technischen Fachwissen der Allgemeinheit und der Komplexität bei der Handhabung umfangreicher Datensätze und Parameter ergeben kann.

Um potenzielle Haftungslücken zu schließen, wird in der Rechtsliteratur intensiv über die Zuerkennung einer eigenen Rechtspersönlichkeit für Roboter und die Zuweisung eigener Verantwortlichkeiten diskutiert. Dieser in einer anthropomorphen Sichtweise verankerte Ansatz erkennt jedoch, dass KI-Systeme weder ein echtes moralisches Urteilsvermögen noch einen eigenen Willen und die Fähigkeit besitzen, ihre Verhalten zu verstehen. Im Gegensatz zur strafrechtlichen Verantwortlichkeit von Unternehmen stößt dieser Ansatz auch auf zahlreiche technische Hürden, etwa die Unfähigkeit von KI-Systemen, strafrechtlich relevante Handlungen vorzunehmen. Folglich lässt sich eine solche Form der Haftung nicht durch Analogien begründen, sondern müsste über weitgehende juristische Fiktionen aus pragmatischen Gründen konstruiert werden. Ein solches Vorgehen ist in absehbarer Zeit, insbesondere innerhalb bestehender Rechtsrahmen, kaum realisierbar.

Anders als im Strafrecht können im Zivilrecht Mechanismen wie die Gefährdungshaftung die Bestimmung der Haftung zumindest teilweise vereinfachen. Zwar können einige Funktionen des Strafrechts zur Sicherung von

Gerechtigkeit und sozialer Ordnung teilweise durch zivilrechtliche Regelungen abgedeckt werden, aber beide Rechtsgebiete verfolgen grundlegend unterschiedliche Zwecke. Daher lassen sich zivilrechtliche Haftungsmodelle nicht auf das Strafrecht übertragen. Infolgedessen entsteht eher eine „Vergeltungslücke“ („*retribution gap*“) als eine Lücke bei der strafrechtlichen Haftung. Mit der wachsenden Verbreitung solcher Systeme wird es künftig notwendig sein, zu beurteilen, ob ein Leben mit einer solchen Lücke den Erwartungen der Gesellschaft an die Rechtsordnung entspricht.

Strafrechtliche Produkthaftung kann für Hersteller von KI-gesteuerten autonomen Systemen anwendbar sein. Allerdings ergeben sich drei zentrale Herausforderungen: die Einstufung dieser Systeme als Produkte, die Bestimmung dessen, was in diesem Kontext einen Mangel darstellt, und die Bewältigung der Beweislastprobleme, die sich aus ihrer Intransparenz ergeben.

Bei vorsätzlichen Straftaten zeigen KI-gesteuerte autonome Systeme keine erheblichen Unterschiede; eine Haftung ist gegeben, solange der Kausalzusammenhang festgestellt werden kann. Hingegen kann – entgegen einem Teil der Literatur – das Modell der mittelbaren Täterschaft nicht auf Straftaten angewendet werden, bei denen solche Systeme eingesetzt werden; da ihnen ein eigener Wille fehlt, stellen ihre Verhalten keine strafrechtlich relevante Handlung dar, und sie können nicht als „ein anderer“ im menschlichen Sinne betrachtet werden.

Im Rahmen der fahrlässigen Haftung ergibt sich die Sorgfaltspflicht aus einem vielschichtigen Gefüge, das gesetzliche Vorschriften, Verhaltenskodizes, Verhaltensstandards, berufliche Richtlinien, behördliche und betriebliche Anweisungen, Nutzungsprotokolle sowie ungeschriebene Normen umfasst. Die Einhaltung dieser Standards dient jedoch nur als Indiz; allgemeine Grundsätze, wie die Pflicht, Schaden zu verhindern, behalten in jedem Fall Gültigkeit. Letztlich obliegt die Entscheidung über das Vorliegen von Fahrlässigkeit dem Gericht, das alle konkreten Umstände des Einzelfalls berücksichtigt.

Bei Fahrlässigkeit werden das Sonderwissen und die Sonderfähigkeiten der jeweiligen Person oder Organisation berücksichtigt. Wenn ein Unternehmen beispielsweise eine Methode zur angemessenen Risikominderung entwickelt hat, muss diese umgesetzt werden, auch wenn sie sich (noch) nicht als Branchenstandard etabliert hat. Zudem haften Personen oder Organisationen, die sich auf riskante Aktivitäten einlassen, ohne über die nötigen Fähigkeiten oder das erforderliche Fachwissen zu verfügen, für schädliche Folgen im Rahmen einer Übernahmefahrlässigkeit.

Für die fahrlässige Haftung muss das schädliche Ergebnis zumindest im Allgemeinen vorhersehbar und vermeidbar sein. Die mit KI-gesteuerten autonomen Systemen verbundenen Risiken sind jedoch für sich genommen erkennbar. Daher ist die Haftung von Personen zu prüfen, die Aufgaben an solche Systeme delegieren, anstatt sie auf herkömmliche Weise selbst auszuführen. Dies bedeutet nicht, dass die jeweilige Person in jedem Fall haftet. Vielmehr bedarf es einer eingehenden Prüfung, indem die Delegation der Aufgabe als ein relevanter Akt im Sinne des Strafrechts anerkannt wird. Folglich ist die in der Literatur weit verbreitete Auffassung, diese Personen seien lediglich passiv und damit nicht haftbar, kritisch zu hinterfragen. Gleichwohl muss in solchen Fällen zwischen typischen und atypischen Risiken unterschieden werden. Darüber hinaus wird der Umfang der Sorgfaltspflicht durch Erkenntnisse aus vergangenen Vorfällen und die neuen, durch die Technologie ermöglichten Potenziale weiter geprägt.

Die vollständige Ausschaltung von Risiken, die mit KI-gesteuerten autonomen Systemen verbunden sind, ist nicht realisierbar, und die Lehre vom erlaubten Risiko dient bei der Bestimmung der Sorgfaltspflicht als Leitlinie. In Ermangelung gefestigter Erfahrungen und Standards kann es erforderlich sein, den Stand von Wissenschaft und Technik heranzuziehen, um Risiken auf ein zulässiges Maß zu reduzieren. Dieser Ansatz trägt dem dynamischen Charakter des Fachgebiets Rechnung. Eine *ex-ante*-Bestimmung, welche Aktivitäten ein erlaubtes Risiko darstellen, ist schwierig. Zwar können Standards die Belastung für Handelnde mindern, jedoch keine vollständige Entlastung bieten, da sie lediglich als Indikatoren fungieren. Eine bloße Formalkonformität durch „Abhaken von Kästchen“ begründet keine umfassende Freistellung von Verantwortung. Letztlich steht die Frage im Vordergrund, ob die notwendigen Maßnahmen zur Risikoreduzierung angemessen umgesetzt wurden.

Die mit KI-gesteuerten autonomen Systemen verbundenen Risiken können als erlaubt gelten, wenn alle erforderlichen Maßnahmen ergriffen werden, um diese Risiken auf ein akzeptables Niveau zu senken, und wenn sie sich im Rahmen der gesellschaftlichen Toleranz bewegen. Dabei ist der Nutzen für die Gesellschaft und potenzielle Risiken sorgfältig abzuwägen. Wenn, wie in der Literatur vorgeschlagen, allgemeine Aspekte (z. B. ökonomische oder ökologische Beiträge), die nicht unmittelbar mit der konkreten Aufgabe in Verbindung stehen, berücksichtigt werden, ist es ebenso erforderlich, die gesamtgesellschaftlichen negativen Auswirkungen in die Analyse einzubeziehen. Zudem lässt sich keine allgemeingültige Lösung für alle KI-Anwendungen formulieren. Vielmehr sollte ein Kalibrierungsmodell

zur Anwendung kommen, das die Risiken (Schadensschwere und Eintrittswahrscheinlichkeit) der jeweiligen Tätigkeit sowie deren gesellschaftliche Funktionen berücksichtigt, um sie auf ein zulässiges Maß zu reduzieren. Riskante Tätigkeiten, die lediglich bestimmten Segmenten der Gesellschaft zugutekommen, und solche, die für die Gesellschaft unverzichtbar sind, sollten daher nicht gleichermaßen bewertet werden. Auf diese Weise kann eine abgestufte Sorgfaltspflicht gewährleistet werden, die der Natur der jeweiligen Tätigkeit angemessen ist.

Ob die Delegation einer Aufgabe an KI-gesteuerte autonome Systeme im Vergleich zu herkömmlichen Methoden die Risiken erhöht, ist zu untersuchen. Allerdings lassen sich Risiken und Gefahren nicht allein als quantitativ steigende oder fallende Variablen verstehen; vielmehr handelt es sich um eine Substitution verschiedener Gefährdungen. In bestimmten Fällen können sich bestimmte Gefahren verstärken, während andere abnehmen. Eine Bewertung kann in jedem Fall mithilfe der Risikoerhöhungstheorie erfolgen. Dieses Vorgehen zielt letztlich darauf ab zu verhindern, dass Personen, die die Risiken und Verantwortlichkeiten einer Tätigkeit auf autonome Systeme übertragen und sich damit in eine „passive“ Position begeben, das Konzept des erlaubten Risikos für sich ausnutzen.

Wenn die Ausführung einer Aufgabe mithilfe KI-gesteuerter autonomer Systeme die Risiken erheblich verringert, keine neuen oder unzumutbaren Risiken hervorruft und gesellschaftlich akzeptiert ist, kann das Unterlassen des Einsatzes solcher Systeme in Zukunft eine Verletzung der Sorgfaltspflicht darstellen.

Obwohl der „EU AI Act“ die strafrechtliche Verantwortlichkeit nicht unmittelbar regelt, werden den beteiligten Akteuren je nach Risikostufe der KI bestimmte Anforderungen und Pflichten auferlegt. Diese Vorschriften können als Referenz für die Ausgestaltung der Sorgfaltspflicht im nationalen Recht dienen.

Die Entwicklung, Bereitstellung und Nutzung von KI-gesteuerten Systemen erfolgt häufig durch mehrere Beteiligte, und das strafrechtliche Problem kann entweder auf die Handlung einer Einzelperson oder auf eine Summe von Handlungen mehrerer Personen zurückzuführen sein. In dieser Hinsicht unterscheidet sich die Thematik nicht wesentlich von klassischen Fällen im Strafrecht (z. B. Produkthaftung), wobei auch hier der Vertrauensgrundsatz mit seinen Einschränkungen zur Anwendung kommt.

Die Ausweitung des Vertrauensgrundsatzes auf KI-gesteuerte autonome Systeme wirft jedoch einige Probleme auf. Zum einen unterliegen Menschen in der Regel Überwachungspflichten, um die korrekte Funktion

dieser Systeme sicherzustellen. Zum anderen müssen Maschinen so konzipiert sein, dass sie mit vorhersehbaren und häufig typischen menschlichen Fehlern umgehen können. Darüber hinaus ist der Vertrauensgrundsatz ein für den Menschen entwickeltes Konzept, das auf seinen biologischen Fähigkeiten beruht. Maschinen hingegen können durch ihre Sensorik und Datenverarbeitung eine kontinuierliche Überwachung durchführen. Folglich muss der Vertrauensgrundsatz nicht in seiner ursprünglichen Form auf Maschinen angewendet werden.

Entgegen der vorherrschenden Meinung sind selbstfahrende Fahrzeuge kaum mit rein typischen Dilemmasituationen konfrontiert. In diesem Zusammenhang ist der Einsatz von *State-of-the-Art*-Kollisionsvermeidungssystemen unter dem Gesichtspunkt des erlaubten Risikos im Rahmen der Sorgfaltspflicht zu beurteilen. In den seltenen Fällen, in denen sich tatsächlich ein reines Dilemma ergibt, bieten der rechtfertigende oder entschuldigende Notstand oder die Pflichtenkollision keine zufriedenstellenden Lösungen. Auch die Anwendung des übergesetzlichen Notstands wird in der Literatur unterschiedlich kritisiert. Nichtsdestotrotz ist das Prinzip der Unantastbarkeit des menschlichen Lebens aufrechtzuerhalten.

Wie in der Literatur *de lege ferenda* vorgeschlagen, könnte die Einführung gefährlicher Produkte in den Verkehr ohne angemessene Sicherheitsvorkehrungen als abstraktes Gefährdungsdelikt – bei dem der Schadenseintritt eine objektive Bedingung der Strafbarkeit darstellt – ein praktisches Modell zur Abschreckung darstellen. Dieser Ansatz greift zwar viele Schwierigkeiten bei der Bestimmung der strafrechtlichen Verantwortlichkeit auf, stößt jedoch aufgrund der besonderen Eigenschaften von KI auf verschiedene Herausforderungen und Bedenken.

