

Milan Schäfer: Artificial Intelligence und Strafrecht. Zur Leistungsfähigkeit des geltenden Strafrechts im Hinblick auf die Herstellerverantwortlichkeit bei Schädigungen durch tiefe neuronale Netze

von Prof. Dr. Anja Schiemann

2024, Duncker & Humblot GmbH, ISBN: 978-3-428-19095-9, S. 593, Euro 109,90

Die Dissertation, die erst 2024 bei Duncker&Humblot erschienen ist, wurde bereits im Sommersemester 2023 an der Johannes Gutenberg-Universität Mainz von der Juristischen Fakultät angenommen. Ausgewählte Literatur wurde noch bis Oktober 2023 berücksichtigt und aktualisiert. In Zeiten schnelllebiger Digitalisierung ist dies wichtig zu wissen. Allerdings sind die de lege ferenda gegebenen Ausblicke des Schlusskapitels durchaus nicht überholt und durch die Grundlagenarbeit der Dissertationsschrift können auch weitere Monografien darauf aufbauen.

Teil 1 der Dissertation führt in den Untersuchungsgegenstand ein und steckt zunächst das Ziel der Arbeit ab. Es soll „auf den Prüfstand gestellt werden, wie leistungsfähig das geltende deutsche Strafrecht tatsächlich ist, Fälle sachgerecht zu erfassen, in denen es zu Rechtsgutsverletzungen durch Artificial Intelligence kommt“ (S. 25). Danach wird Artificial Intelligence definiert, wobei dies in Nuancen durchaus umstritten ist. Daher geht die Arbeit zunächst der Geschichte der Artificial Intelligence nach, um dann verschiedene AI-Systeme voneinander abzugrenzen. Anwendungsbeispiele sensibilisieren für die komplexe Thematik, die in vielen Bereichen eine Rolle spielt, so dass gegen Ende dieses Teilkapitels festgestellt wird, dass es wohl deutlich einfacher sei, Branchen zu benennen, bei denen AI nicht zum Fortschritt beitragen kann, als umgekehrt (S. 66).

Das strafrechtliche „Potential“ von AI wird in dem nächsten Unterkapitel erwogen und der Untersuchungsgegenstand der Arbeit umrissen und begrenzt (S. 78 ff.).

Teil 2 der Arbeit beschäftigt sich mit dem fahrlässigen AI-Erfolgsdelikt. Dieser bildet mit knapp 350 Seiten einen wesentlichen Schwerpunkt der Arbeit. Sehr dezidiert und differenziert wird zu den einzelnen Voraussetzungen der Fahrlässigkeit Stellung bezogen. Hier überzeugen die stringente Gliederung und die facettenreiche Darstellung. Durch Zwischenergebnisse der einzelnen Unterkapitel gelingt auch die Zusammenfassung der wesentlichen Erkenntnisse sehr gut, die in einem Schlusskapitel nochmals zusammengeführt werden. Dabei dürfte die Prüfung, ob ein Verhalten des potenziellen Täters kausal für den Eintritt des schädlichen Erfolgs geworden ist, keine Probleme bereiten. Zwischen Vertrieb bzw. Einsatz des Systems und Eintritt des Erfolges bestehe eine lückenlose Verbindung und beides könne nicht hinweggedacht werden, ohne dass der tatbestandsmäßige Erfolg entfielen. Dies gelte zu-

meist auch für einzelne Beiträge zur Entwicklung des Systems, insbesondere wo der Täter das System bzw. seinen Beitrag zur Systementwicklung noch für die weitere Entwicklung, den Vertrieb oder Einsatz freigäbe. Zwar sei das Zustandekommen der schädigenden Entscheidungsfindung eines AI-Systems nicht im Einzelnen nachvollziehbar. Dies sei aber für die Kausalitätsprüfung unschädlich, weil hierdurch allenfalls das Wie des Kausalzusammenhangs fraglich bleibe. Technikbedingte Beweisprobleme könnten sich lediglich in seltenen Ausnahmefällen ergeben, in denen eine Unterlassungsstrafbarkeit im Raum stehe.

Ob denjenigen ein Fahrlässigkeitsvorwurf treffe, der ein AI-System entwickelt und im Anschluss vertreibt oder einsetzt oder bei der Entwicklung eines solchen Systems unterstützt, sei mit überzeugenden Stimmen in der Literatur eine Abwägungsfrage. Dabei sei die Vorausschbarkeit des Erfolgesintritts hierbei ein bloßes Indiz für fahrlässiges Verhalten. In einzelnen Teilbereichen werde diese Abwägung durch einschlägige außerstrafrechtliche Verbote indiziert. Beispielsweise begründe der Einsatz oder Vertrieb eines vollautonomen Fahrzeugs der Stufe fünf derzeit stets einen Fahrlässigkeitsvorwurf, ebenso wie die Unterstützung des Einsatzes oder Vertriebs im Rahmen der Entwicklung solcher Fahrzeuge.

Bestünden keine außerstrafrechtlichen Verbote, so seien der Einsatz oder Vertrieb eines AI-Systems sowie die Unterstützung bei der Entwicklung eines solchen Systems in Abwägung der widerstreitenden Interessen nicht per se als fahrlässig anzusehen, obwohl das generell voraussehbare Risiko einer schädigenden Fehlentscheidung des Systems nicht sicher ausgeschlossen werden könne. Andernfalls würden selbst solche AI-Systeme verboten, die eine Aufgabe fehlerfrei oder gleich gut oder besser als ein Mensch ausführen würden und bei denen auch durch eine menschliche Überprüfung und Freigabe jeder Entscheidung des Systems kein Mehr an Sicherheit schaffen würde. Dies würde den Rechtsgüterschutz in unzulässiger Weise konterkarieren oder jedenfalls nicht fördern.

Ein Fahrlässigkeitsvorwurf scheide also aus, solange das System im Rahmen umfangreicher geeigneter Tests eine geringere – oder ausnahmsweise identische – Fehlerquote im Vergleich zum Menschen erziele und diese Quote auch geringer oder identisch im Vergleich zu technischen Alternativen ist, die eine geringere Fehlerquote als der Mensch aufweisen. Als zusätzliche Maßnahme sei dann aber zu verlangen, dass bei der Systementwicklung alle zumutbaren Risikominimierungsmaßnahmen ergriffen wurden. Etwas anderes gelte, sofern eine menschliche

Überprüfung und Freigabe jeder Systemscheidung durch eine hierzu befähigte Person sinnvoll möglich und zumutbar wäre und hierdurch ein Mehr an Sicherheit erzielt würde.

Der Verfasser ist der Auffassung, dass zur weiteren Konkretisierung die Grundsätze der strafrechtlichen Produzentenhaftung herangezogen werden könnten. Deren Anwendungsbereich sei immer dort eröffnet, wo ein AI-System in Verkehr gebracht werde. Ob es sich um ein verkörpertes System oder unkörperliche Software handle, sei unbeachtlich. Aus den Grundsätzen der strafrechtlichen Produzentenhaftung ergebe sich ein Höchstmaß an Anforderungen. Für die Beurteilung, ob die Konstruktion den Grundsätzen strafrechtlicher Produzentenhaftung genüge, sei der Zeitpunkt des Inverkehrbringens maßgeblich. Nach dem Inverkehrbringen träfe den Hersteller eine Beobachtungspflicht.

Bringe der Hersteller ein AI-System in Verkehr, ohne zuvor ausreichende Maßnahmen zur Vermeidung von Fehlentscheidungen ergriffen zu haben und träfe dieses System dann solche Fehlentscheidungen, durch die Leib oder Leben einer Person geschädigt werden, so werde ein Fahrlässigkeitsvorwurf auch im Hinblick auf die konkrete Verursachung eines Todes- oder Körperverletzungserfolgs in der Regel zu bejahen sein. Es fehle allerdings am Schutzzweckzusammenhang, soweit der Erfolg bei fehlerfreiem Verhalten des AI-Systems ebenso eingetreten wäre.

Wichtiger Punkt ist dann noch die Frage der eigenverantwortlichen Selbstgefährdung. Voraussetzung hierfür ist, dass der Nutzer selbst die letzte menschliche Handlung vor dem Erfolgseintritt tätigt und der Hersteller sichergestellt hat, dass der Nutzer hinreichende Informationen über die generell bestehenden Risiken des Systemeinsatzes erhält oder die Risiken anderweitig kennt. In diesen Fällen könne der Nutzer nämlich autonom über seine Selbstgefährdung entscheiden, sodass der Hersteller straffrei ausgehen müsse.

Etwas anderes könne indes gelten, wo der Hersteller die sich aus den Grundsätzen der strafrechtlichen Produzentenhaftung ergebenden Anforderungen missachtet. Dann werde auch eine eigenverantwortliche Selbstgefährdung des Nutzers regelmäßig ausscheiden.

Dagegen bereite der Nachweis des Pflichtwidrigkeitszusammenhangs zumeist keine grundlegenden AI-technologiespezifischen Schwierigkeiten, wo es zu Rechtsgutsverletzungen durch Fehlentscheidungen eines AI-Systems gekommen ist. Hier unterscheidet der Verfasser verschiedenen Fälle. Zunächst sei ein Pflichtwidrigkeitszusammenhang zu bejahen, wenn der Hersteller ein AI-System vertreibt oder einsetzt, das entweder generell nicht vertrieben oder eingesetzt werden darf oder zumindest nicht den zuvor in der Arbeit ausformulierten strafrechtlichen Anforderungen genügt. Daneben sei ein Pflichtwidrigkeitszusammenhang im Rahmen arbeitsteiliger Organisationen zu bejahen, wenn die Strafbarkeit desjenigen in Frage stehe, der an der Entwicklung des Systems mitwirkt, hier-

bei die Anforderungen an die Systementwicklung missachtet und anschließend dennoch das System für weitere Entwicklung, Vertrieb und Einsatz freigibt. Zuletzt wird herausgestellt, dass AI-Systeme selbst nicht als eigenverantwortliche Dritte gesehen werden können, deren Dazwischentreten die Zurechnung ausschließt.

Für das vorsätzliche Erfolgsdelikt können dann die erarbeiteten Voraussetzungen für den objektiven Tatbestand übertragen werden, die sich auf Kausalität und objektive Zurechnung beziehen. Bei gezielter oder wissentlicher Programmierung und Ausrichtung des Systems auf die Schädigung von Rechtsgütern sei der Tatbestand in der Regel unproblematisch gegeben. Unklar bleibt hier zunächst, warum es sich bei dieser Intention um ein Element des objektiven Tatbestands handeln soll. In Hervorhebung des Blackbox-Charakters der AI-Systeme wird dann aber klar, dass eine Programmierung an sich immer noch mit Ungewissheiten verbunden ist. Dies sei aber für die Verwirklichung des Vorsatzes unschädlich, da es laut Verfasser nicht darauf ankommen kann, dass der Hersteller nicht voraussieht, bei welchem konkreten Tatopfer wann, wo und durch welches konkrete AI-System ein schädlicher Erfolg eintritt oder nicht antizipieren kann, wie die schädigende Fehlentscheidung im System im Einzelnen zustande gekommen ist. Hier könne man einen systematischen Vergleich zu anderen Fallgestaltungen bspw. im Medizinrecht oder bei Tötungen bei Distanzdelikten herstellen. Wenn allerdings hier die Unbeherrschbarkeit im Raum steht, so muss die Frage gestattet sein, ob man dann nicht stets im Bereich der Vorsatzdelinquenz ist oder jedenfalls öfter, als dies angesichts der dem Unrechtsgehalt viel naheliegender Fahrlässigkeitsstrafbarkeit geboten ist.

Schließlich stellt der Verfasser fest, dass eine Teilnahme an „Taten“ eines AI-Systems sowie mittelbare Täterschaft „durch“ oder Mittäterschaft „mit“ einem solchen AI-System de lege lata ausgeschlossen sind, da das System selbst keine rechtswidrigen Taten begehen oder Täter sein kann. Setze der Täter ein AI-System rechtsgutsverletzend ein, so sei er unmittelbarer Täter. Mittelbare Täterschaft sei aber insofern möglich, als er ein AI-System an einen ahnungslosen Dritten vertreibe, der dieses dann mit entsprechender Fehlfunktion einsetzt.

Insgesamt kommt *Schäfer* zu dem Ergebnis, dass die neue Technik „im Großen und Ganzen durchaus mit den Mitteln des geltenden Strafrechts zu bewältigen ist“ (S. 508). Allerdings drohten im Rahmen arbeitsteiliger Organisationen aufgrund der AI-Technologie ausnahmsweise grundlegende Beweisschwierigkeiten. Dies gelte für die Strafbarkeit einzelnen Personen, die zwar an der Entwicklung eines AI-Systems mitwirken und denen dabei ein Fehlverhalten anzulasten ist, die dann allerdings das System weder selbst einsetzen oder vertreiben noch ihren eigenen Entwicklungsbeitrag für weitere Schritte freigeben. Es bleibt die Frage, ob gerade in letzteren Fallkonstellationen überhaupt ein strafwürdiges Verhalten vorliegt. So weist auch der Verfasser auf den fragmentarischen Charakter des Strafrechts hin, so dass die Straflosigkeit in einigen wenigen Fällen hinzunehmen sei.

Obwohl dem geltenden Strafrecht die Tauglichkeit im Hinblick auf die technologischen Entwicklungen im AI-Bereich bescheinigt wird, wird dennoch ein Ausblick de lege ferenda gegeben. Dies betrifft zunächst die Frage der Sanktionierung eines AI-Systems an sich. Hier werden sehr umfangreich gegenteilige Auffassungen zitiert, um dann knapp aber überzeugend eine Sanktionierung des Systems abzulehnen.

Anschließend wird diskutiert, ob die wenigen bestehenden Strafbarkeitslücken ggf. durch die Einführung eines abstrakten Gefährdungsdelikts geschlossen werden sollten. Ein solcher Straftatbestand könnte, so der Verfasser, Fehlverhalten von an der Herstellung eines AI-Systems beteiligten Personen unabhängig vom zurechenbaren Eintritt des schädigenden Erfolgs unter Strafe stellen. Auch hier ist *Schäfer* skeptisch und ich schließe mich seiner Auffassung angesichts dieser weiten Vorfeldkriminalisierung an.

Daneben erwägt der Verfasser die Anpassung einzelner bereits existierender Gefährdungstatbestände im Bereich des autonomen Fahrens. Hier könnten Straßenverkehrsdelikte wie §§ 315c, 316 StGB, § 21 StVG modifiziert werden. Die vertiefende Diskussion hierzu verantwortet *Schäfer* aber nach 526 Seiten völlig zu Recht weiteren Ausarbeitungen.

Letztlich wird in der Arbeit noch diskutiert, ob die Grenzen des Erlaubten und somit die Anforderungen gesetzlich

ausdrücklich festgeschrieben werden sollen, die die Hersteller von AI-Systemen erfüllen müssen, um den Vorwurf fahrlässigen Verhaltens bzw. einer rechtlich missbilligten Gefahrschaffung zu vermeiden. Zwingend erscheine das zwar nicht, aber man habe die außerstrafrechtliche Entwicklung in den Blick zu nehmen, so dass auch der AI-Act zu berücksichtigen sei. Hier wird aufgrund des Bearbeitungsstands der Arbeit der Entwurfsinhalt wiedergeben. Die KI-Verordnung ist allerdings mittlerweile schon im Europäischen Amtsblatt veröffentlicht. Insofern ist man an dieser Stelle gut damit beraten, in diese Veröffentlichung zu schauen und das Unterkapitel der Arbeit zu überspringen.

Dennoch kann am Ende dieses Ausblicks durchaus der Ratschlag des Verfassers beherzigt werden, aus Sicht der Strafrechtswissenschaften die außerstrafrechtliche Entwicklung im Blick zu behalten. So sollten außerstrafrechtliche Gesetzesvorhaben zu AI nicht zuletzt auf ihre strafrechtlichen Auswirkungen überprüft werden, um ggf. notwendigen Anpassungsbedarf aufzuzeigen. Allerdings bleibt resümierend das Plädoyer *Schäfers*, Anpassungen nur moderat vorzunehmen und insbesondere im Allgemeinen Teil des geltenden Strafrechts zu unterlassen. Dem kann man sich voll und ganz anschließen. Inwiefern es Nachjustierungen im Besonderen Teil – bspw. beim autonomen Fahren der 5. Stufe – geben sollte, bleibt weiteren Arbeiten überlassen. Hier darf man sicher sein, dass bzgl. Detailfragen eingesetzter AI-Systeme sicher noch zahlreiche Dissertationen erscheinen werden.