Künstliche Intelligenz und Kriminalität

von Hauke Bock und
*Prof. Dr. Katrin Höffler**

Abstract
*Kriminalität ist letztlich eine Spielart menschlichen Verhaltens. Da das menschliche Verhalten von der Implementierung Künstlicher Intelligenz (KI) verändert wird, geschieht dies auch mit diesem besonderen Bereich, der Kriminalität. Während dies für einzelne Phänomene – so z.B. beim automatisierten Fahren – teilweise schon breit auch von den hiesigen Strafrechtswissenschaften aufgegriffen wurde, ist ein grundsätzlicher Blick auf die verschiedenen betroffenen Ebenen (Täter*innen, Taten, Opfer und Verbrechenskontrolle) vor allem in der englischsprachigen Literatur zu finden.[1] Daher soll durch diesen Beitrag eine kriminologische Perspektive, modelliert durch das hiesige Strafrecht als Bezugspunkt, ergänzt werden.*

Crime is essentially a variation of human behaviour. As human behaviour is changed by the implementation of Artificial Intelligence (AI), this also applies to the particular area of crime. While this has already been widely addressed by German criminal law for specific phenomena, such as automated driving, a fundamental view of the different aspects involved (offenders, offences, victims, and crime control) can be found mainly in the anglophone literature. Therefore, this paper aims to add a criminological perspective, modelled by the German criminal law as a point of reference.

I. Arbeitsdefinition „Künstliche Intelligenz“

Will man sich dem Thema „KI und Kriminalität“ nähern, so bedarf es zunächst einer Definition des konturierenden Analysegegenstandes, hier also insbesondere des schillernden Begriffs der „Künstlichen Intelligenz“.[2] Die Disziplinen sind sich über die Fächergrenzen hinweg einig, dass eine einheitliche Definition für die Künstliche Intelligenz bisher jedoch nicht gefunden werden konnte.[3] So versuchte etwa die EU-Kommission in der Vergangenheit, sie als „Systeme mit einem ‚intelligenten‘ Verhalten, die ihre Umgebung analysieren und mit einem gewissen Grad an Autonomie handeln, um bestimmte Ziele zu erreichen“[4] zu definieren. Derartige Definitionsansätze sind nicht nur zirkulär,[5] sie scheitern vor allem daran, dass bereits eine einheitliche Beantwortung der Frage, was Intelligenz ist, die Wissenschaft vor erhebliche Herausforderungen stellt.[6]

Die unterschiedlichsten Disziplinen versuchen trotzdem, den so unscharfen Begriff der Künstlichen Intelligenz mit Leben zu füllen, zu konturieren.[7] Auch für die Zwecke der kriminologischen Analyse ist eine Arbeitsdefinition der Künstlichen Intelligenz notwendig, um den Gegenstand der folgenden Betrachtungen nicht uferlos werden zu lassen. Eine Möglichkeit, dies zu erreichen, ist eine möglichst technologieneutrale und zukunftssichere Definition durch eine erweiterbare Aufzählung der Technologien, die mit dem Begriff der Künstlichen Intelligenz gemeint sein sollen. Diesen Weg wählt nun offenbar auch die EU-Kommission. In ihrem Vorschlag zu einer Regulierung der Künstlichen Intelligenz durch einen sog. Artificial Intelligence Act wird ebenfalls auf folgende Aufzählung von Techniken und Ansätzen verwiesen:[8]

ANNEX I

ARTIFICIAL INTELLIGENCE TECHNIQUES AND APPROACHES referred to in Article 3, point 1
(a) Machine learning approaches, including supervised, unsupervised and reinforcement learning, using a wide variety of methods including deep learning;

---

* *Prof. Dr. Katrin Höffler* ist Inhaberin des Lehrstuhls für Strafrecht, Strafprozessrecht, Kriminologie und Rechtssoziologie an der Universität Leipzig; *Hauke Bock* ist dort als Wissenschaftlicher Mitarbeiter tätig. Der vorliegende Beitrag geht zurück auf einen Vortrag im Rahmen der Ringvorlesung „Künstliche Intelligenz, Data Science und Gesellschaft“ an der Georg-August-Universität Göttingen, veranstaltet vom Campus Institute for Data Science (CIDAS) im Wintersemester 2021/22.

1 *King et al.,* Artificial Intelligence Crime, in: Science and Engineering Ethics 26 (2020), 89; *Pagallo,* AI and bad robots. The criminology of automation, in: McGuire/Holt, The Routledge Handbook of Technology, Crime and Justice, 2017, 643; *Hayward/Maas*, Artificial Intelligence and Crime, in: Crime Media Culture 17 (2021), 209; *Broadhurst et al.,* Artificial Intelligence and Crime. A Research Report for the Korean Institute of Criminology (2019).

2 Auf eine Definition von Kriminalität soll an dieser Stelle verzichtet werden und stattdessen auf die Diskussion um den Verbrechensbegriff als Rahmen des Auftrags der kriminologischen Forschung verwiesen werden (s. nur *Eisenberg/Kölbel,* Kriminologie, 7. Auflage [2017], § 1 Rn. 30 ff. m.w.N.; *Meier,* Kriminologie, 6. Auflage [2021], § 1 Rn. 5 m.w.N.; *Singelnstein/Kunz,* Kriminologie, 8. Auflage [2021], § 1 Rn. 27 ff.).

3 *Kaulartz/Braegelmann,* Rechtshandbuch Artificial Intelligence und Machine Learning (2020), Kap. 1 Rn. 2; *Humm/Buxmann/Schmidt,* in: Gethmann et al., Künstliche Intelligenz in der Forschung, 2022, S. 13 (18); *Krafft et al.,* Defining AI in Policy vs. Practice, 2020, online abrufbar unter: https://arxiv.org/pdf/1912.11095.pdf (zuletzt abgerufen am 18.7.2022).

4 EU-Kommission, COM (2018) 237 final, 25.4.2018, S. 1.

5 *Herberger,* NJW 2018, 2825 (2826).

6 Siehe nur *Legg/Hutter,* A Collection of Definitions of Intelligence 2007, online abrufbar unter: https://arxiv.org/pdf/0706.3639.pdf (zuletzt abgerufen am 18.7.2022) für mehr als 70 Definitionsversuche verschiedener Disziplinen.

7 *Krafft et al.,* Defining AI in Policy vs. Practice, 2020, online abrufbar unter: https://arxiv.org/pdf/1912.11095.pdf (zuletzt abgerufen am 18.7.2022); *Grewal*, A Critical Conceptual Analysis of Definitions of Artificial Intelligence as Applicable to Computer Engineering, in: IOSR Journal of Computer Engineering 2014, Vol. 16, Iss. 2, 9 (13); *Oshida,* Artificial Intelligence for Medicine (2021), Kap. 1.1; *Söbbing,* Rethinking Law 1/2019, 33; *Russel/Norvig,* Artificial Intelligence – A Modern Approach, 4. Auflage (2021), S. 1 ff.

8 EU-Kommission, COM (2021) 206 final, 21.4.2021.

(b) Logic- and knowledge-based approaches, including knowledge representation, inductive (logic) programming, knowledge bases, inference and deductive engines, (symbolic) reasoning and expert systems;
(c) Statistical approaches, Bayesian estimation, search and optimization methods.

An diesem eher weit gefassten Katalog[9] soll sich auch das diesem Beitrag zugrunde liegende Verständnis des Begriffs der Künstlichen Intelligenz orientieren.

II. Potenzielle Berührungspunkte von KI und Kriminalität

Nachdem geklärt ist, was mit Künstlicher Intelligenz bezeichnet sein soll, stellt sich nun die Frage, was für Berührungspunkte es zwischen KI und Kriminalität geben kann. Die Kriminologie unterteilt ihren Untersuchungsgegenstand klassischerweise in die Bereiche Verbrechen, Verbrecher*innen, Verbrechenskontrolle und Viktimologie[10] bzw. Tat, Täter*innen, Opfer und soziale Kontrolle[11]. Es liegt daher nahe, auch die Untersuchung der Auswirkungen des Fortschritts im Bereich der Künstlichen Intelligenz auf die Kriminalität anhand dieser Einteilung zu gliedern. Um Redundanzen zu vermeiden und Bezüge besser verdeutlichen zu können, sollen die Auswirkungen auf die Tat und den*die Täter*in gemeinsam dargestellt werden. Die potenziellen Berührungspunkte sind somit die Auswirkungen der KI auf Täter*innenseite (III.), die Auswirkungen der KI auf Opferseite (IV.) und die KI als Mittel der sozialen Kontrolle (V.).[12]

III. KI auf Täter*innenseite

Zunächst sollen die bereits eingetretenen oder jedenfalls absehbaren Auswirkungen der KI auf Täter*innenseite betrachtet werden. Dass technologische Fortschritte auch immer neue kriminelle Möglichkeiten bedeuten, ist dabei keine Besonderheit der Künstlichen Intelligenz. Wann immer Menschen neue Technologien entwickelt haben, sind diese auch für kriminelle Zwecke missbraucht worden. Das Automobil wurde als Mordwerkzeug und Fluchtwagen genutzt, Telefonanschlüsse für Erpresseranrufe und „Enkeltricks“ und das Internet dient jeden Tag als Schauplatz für eine Vielzahl anonymer Bedrohungen und Beleidigungen. Die Liste ließe sich fortsetzen.

1. KI als Tatmittel

Es liegt somit nahe, dass auch die als Künstliche Intelligenz bezeichneten Technologien als Tatmittel verwendet werden können, was auch bereits geschieht.[13] Teilweise handelt es sich dabei lediglich um neue Spielarten bekannter Tatmuster, teilweise aber auch um eine neue Qualität krimineller Phänomene.

a) Deepfakes

Ein erstes Phänomen ist die Verwendung sog. Deepfakes. Allgemein sind diese der Mehrheit der Bevölkerung wohl eher aus unterhaltsamen Apps oder einer viralen Fake-Weihnachtsansprache der Queen des britischen Senders Channel 4[14] bekannt. Es handelt sich dabei um Bild- und Tonaufnahmen von Personen, die mithilfe Künstlicher Intelligenz hergestellt oder verändert werden und den Eindruck der Echtheit erwecken sollen.[15] Man kann unter Zuhilfenahme von Deepfakes die abgebildeten Personen Handlungen vornehmen oder Äußerungen tätigen lassen, die nie stattgefunden haben. Unterschieden wird dabei zwischen „Face Swap“, dem Austausch der Gesichtszüge, „Lip Sync“, der Anpassung der Mundbewegungen an eine andere Audiodatei, und „Puppetmaster“, bei der das äußere Erscheinungsbild der Person von einem*einer Schauspieler*in bewegt wird.[16] Diese Technologie lässt sich nicht nur zu Unterhaltungszwecken oder für sinnvolle Verwendungen wie die Unterstützung von Menschen mit Sprachbehinderungen einsetzen,[17] sondern bietet auch erhebliches Kriminalitätspotenzial.[18]

Zu Beginn des Missbrauchs der Deepfakes wurden diese vor allem für die Fälschung pornografischen Materials genutzt, indem die Gesichter berühmter Schauspielerinnen mit den Gesichtern der Darstellerinnen vertauscht wurden.[19] Mit zunehmender Verbreitung und Verbesserung der Technologie, die zudem häufig als Open Source verfügbar ist,[20] wachsen auch die Möglichkeiten für ihren Einsatz im Rahmen von Betrugsdelikten, Nötigungen, Erpressungen und Bedrohungen. Zudem sind mit Deep Fakes untermauerte Desinformationskampagnen denkbar, die als neue Qualität der Fake-News-Problematik betrachtet werden können.[21] Einige erachten die mit dem Missbrauch von Deepfakes einhergehenden Gefahren als die in absehbarer Zeit größte kriminelle Bedrohung im Zusammenhang mit KI.[22] Besonders perfide ist dabei, dass die dringend nötige Aufklärung über die Möglichkeiten der Deepfake-Technologien ihrerseits zur Gefahr für die demokratische Gesellschaft werden kann, wenn sie zu einem pauschalen Misstrauen auch gegenüber verifizierten Quellen, statt zu einem reflektierten und kritischen Um-

---

9 *Spindler,* CR 2021, 361 (363).
10 Grundlegend *Kaiser*, Kriminologie, 3. Auflage (1996), S. 1, 207 ff.
11 Ähnlich auch *Göppinger/Bock,* Kriminologie, 6. Auflage (2008), § 1 Rn. 1; *Singelnstein/Kunz* (Fn. 2), § 2 Rn. 1; *Meier* (Fn. 2), § 1 Rn. 5, 29, 31. Zur „Offenheit“ dieses Gegenstandes *Eisenberg/Kölbel* (Fn. 2), § 1 Rn. 1 ff.
12 *Hayward/Maas* (Fn. 1), S. 214 ff. klassifizieren ebenfalls bedenkenswert nach „Crimes *with* AI“, „Crimes *on* AI“ und „Crimes *by* AI“.
13 Vgl. *Brundage et al.,* The Malicious Use of Artificial Intelligence (2018), online abrufbar unter: https://arxiv.org/ftp/arxiv/papers/1802/1802.07228.pdf.
14 https://www.youtube.com/watch?v=IvY-Abd2FfM (zuletzt abgerufen am 7.7.2022).
15 *Thiel,* ZRP 2021, 202.
16 *Farid/Schindler*, Deep Fakes. On the Threat of Deep Fakes to Democracy and Society, 2020, S. 15 f.
17 *Lantwin*, MMR 2019, 574 (575).
18 *Thiel,* ZRP 2021, 202 (203).
19 *Thiel,* ZRP 2021, 202 (203); *Lantwin*, MMR 2019, 574 (575).
20 *Farid/Schindler* (Fn. 16), S. 16.
21 *Holznagel*, MMR 2018, 18; *Thiel,* ZRP 2021, 202 (203).
22 *Caldwell/Andrews/Tanay/Griffin,* AI-enabled future crime, in: Crime Science 9:14 (2020), S. 6 f.

gang (auch) mit bewegten Bildern führt (sog. „liar's dividend").[23]

b) Cyberangriffe

Eine zweite phänomenologische Kategorie lässt sich unter den Oberbegriff „Cyberangriffe" fassen. Hierbei ist zu beachten, dass diese nicht zwingend mit der Verwendung der als Künstliche Intelligenz bezeichneten Technologien einher gehen müssen. Trotzdem zeigt sich auch hier deutlich die Tendenz, dass bereits bekannte kriminelle Begehungsweisen im digitalen Raum durch KI verändert, verstärkt oder ausgeweitet werden können.

Bisher sind erfolgreiche Cyberangriffe meist entweder von großer Präzision geprägt und auf ein konkretes Zielobjekt zugeschnitten[24] oder sie setzen auf ein hohes Maß an Automatisierung und die pure Wucht der Masse.[25]

Ein prominentes Beispiel für die erste Fallgruppe sind Angriffe mit sog. Ransomware.[26] Dabei werden Systeme v.a. von Unternehmen, aber auch von Privatpersonen von häufig hochprofessionell vorgehenden kriminellen Hacker*innen infiltriert und verschlüsselt. Anschließend erfolgt eine Lösegeldforderung, ohne deren Zahlung die Systeme nicht wieder freigeschaltet werden. Im Gegensatz dazu wird beispielsweise bei DDoS-Attacken (Distributed Denial-of-Service) eine Vielzahl gleichzeitiger Anfragen von zuvor infiltrierten Rechnern automatisiert an die angegriffene Webressource gesendet, um ihre Verfügbarkeit durch die Überlastung der Kapazitäten zu stören. Die Durchschlagskraft ergibt sich aus der schieren Menge der Anfragen.[27]

Durch die missbräuchliche Verwendung künstlich intelligenter Systeme ist nun zu befürchten, dass sich die Cyberangriffe spezifisch *und* massenhaft durchführen lassen, indem beispielweise die Schwachstellen in einer Vielzahl von Netzwerken mit Methoden des reinforcement-learnings automatisiert ausfindig gemacht werden, ehe diese gleichzeitig in einer konzertierten und für jedes System spezifischen Weise attackiert werden.[28]

c) Phishing

Ganz ähnlich verhält es sich mit anderen bereits bekannten Kriminalitätsphänomenen wie dem sog. Phishing. Dabei wird versucht, eine der häufigsten Schwachstellen technischer Systeme auszunutzen: den*die Benutzer*in, was auch als „social engineering" bezeichnet wird.[29] Diese*r soll zur Preisgabe von Informationen wie Zugangsdaten oder Passwörtern oder zur unbeabsichtigten Installation von Malware, z.B. durch dubiose Links, gebracht werden, indem unter falscher Identität einer dem Opfer bekannten Person oder Institution Kontakt aufgenommen wird.[30] Beinahe jede*r hat dies wohl schon am eigenen (Computer-)Leib erfahren: solche Nachrichten etwa per E-Mail sind enorm häufig, aber meist auch sehr leicht zu erkennen. Die Erfolgsquote ist dementsprechend niedrig. Der Profit für die Kriminellen ergibt sich erst aus dem automatisierten Versand an eine große Zahl potenzieller Opfer.[31]

Auf einzelne Opfer zugeschnittenes und dadurch schwerer erkennbares Vorgehen, sog. „spear-phishing", ist die Ausnahme und – wie man in der legalen Wirtschaft wohl sagen würde – nur schlecht skalierbar.[32] Eben dies könnte sich ändern, wenn die große Zahl der Adressat*innen beibehalten werden könnte, die Nachrichten aber durch Anpassungen mittels lernender Systeme auf eine höhere Erfolgsquote trainiert werden könnten.[33] Die erhöhte Erfolgsquote in Kombination mit der Vielzahl an Versuchen ließe den Schaden dementsprechend in die Höhe schnellen.

d) Hate Speech durch Social Bots

Die Kommunikation in sozialen Netzwerken wird immer häufiger durch sog. „Social Bots" infiltriert. Dabei handelt es sich um autonome Computerprogramme, die zunehmend besser in der Lage sind, textbasierte menschliche Kommunikation täuschend echt zu imitieren.[34] Diese kommen vor allem in überwiegend textbasierten sozialen Netzwerken wie Twitter zum Einsatz und machen dort inzwischen einen erheblichen Teil der Kommunikation aus.[35] Während sog. „benevolent Bots"[36] vor allem von Kundendiensten oder zu Werbezwecken eingesetzt werden, können mittels „malicious Bots"[37] Falschinformationen verbreitet oder angebliche Mehrheitsmeinungen vorgetäuscht werden, um die öffentliche Meinungsbildung zu beeinflussen. Strafrechtlich relevant werden die so verbreiteten Inhalte insbesondere dann, wenn diese als sog. „Hate Speech" einzuordnen sind, insbesondere in Form von Beleidigungsdelikten oder Volksverhetzung. Dabei ist zu unterscheiden: Der Social Bot kann von vornherein darauf trainiert sein, Hassreden zu verbreiten, sei es durch Reposting oder durch das Erstellen eigener Beiträge.[38] Der lernende Algorithmus kann aber auch von ihrerseits

---

23 *Chesney/Citron,* California Law Review 2019, 1753 (1785 f.); *Thiel,* ZRP 2021, 202 (203).

24 Siehe für ein prominentes Beispiel nur *Kushner,* The real story of Stuxnet, online abrufbar unter: https://spectrum.ieee.org/the-real-story-of-stuxnet (zuletzt abgerufen am 13.7.22).

25 *Caldwell/Andrews/Tanay/Griffin* (Fn. 22), S. 9.

26 *Kochheim,* Cybercrime und Strafrecht in der Informations- und Kommunikationstechnik, 2. Auflage (2018), Kap. 3 Rn. 441 ff.

27 Vgl. etwa *LG Düsseldorf*, MMR 2011, 624 zur Strafbarkeit der DDoS-Attacken als Erpressung (§ 253 StGB) und Computersabotage (§ 303b StGB).

28 *Caldwell/Andrews/Tanay/Griffin* (Fn. 22), S. 9.

29 *Caldwell/Andrews/Tanay/Griffin* (Fn. 22), S. 7 f.

30 *Boddy,* Phishing 2.0: the new evolution in Cybercrime, in: Computer Fraud & Security Bulletin 11/2018, 8.

31 *Vergelis/Shcherbakova/Sidorina,* Spam and Phishing in 2018, online abrufbar unter: https://securelist.com/spam-and-phishing-in-2018/89701/ (zuletzt abgerufen am 13.7.22).

32 *Caldwell/Andrews/Tanay/Griffin* (Fn. 22), S. 8.

33 *Bahnsen/Torroledo/Camacho/Villegas,* DeepPhish: Simulating Malicious AI, 2018, online abrufbar unter: https://albahnsen.files.wordpress.com/2018/05/deepphish-simulating-malicious-ai_submitted.pdf (zuletzt abgerufen am 18.7.2022).

34 *Volkmann,* MMR 2018, 58 (59).

35 *Kochheim* (Fn. 26), Rn. 683 ff.

36 *Freitas/Benevenuto/Veloso/Ghosh,* Social Network Analysis and Mining (2016), S. 1, 3.

37 *Volkmann,* MMR 2018, 58 (59).

38 *Freitas/Benevenuto/Veloso/Ghosh* (Fn. 36), S. 1, 5.

böswilligen anderen Nutzer*innen gezielt dazu gebracht werden, da er von den Äußerungen seiner Umgebung lernt. Das wohl berühmteste Beispiel ist der Chat-Bot „Tay" der Firma Microsoft, der am 23. März 2016 bei Twitter online ging. Bereits nach kurzer Zeit begann dieser, rassistische, sexistische und antisemitische Tweets abzusetzen, sodass Microsoft gezwungen war, das Experiment bereits nach 16 Stunden abzubrechen und den aus dem Ruder gelaufenen Social Bot offline zu nehmen. Einige Internettrolle hatten den lernenden Algorithmus gezielt mit Fragen und Aussagen attackiert, die dieser schnell lernte und übernahm.[39] Auch wenn diese Begebenheit bereits über sechs Jahre zurückliegt, zeigt sie doch das Missbrauchspotenzial, das selbst in nicht von vornherein „böswilligen" Social Bots liegt.

e) Autonome Transportmittel als Tatwerkzeuge

Keine neue Erscheinung, sondern eine technische Erweiterung bereits bekannter Tatmuster würde der absehbare kriminelle oder auch terroristische Gebrauch autonomer Transportmittel als Tatwerkzeuge darstellen.

Fahrzeuge dienen schon lange als Tatmittel für Gewaltdelikte, sei es mittelbar als mit Sprengstoff beladene „Auto-Bombe" oder selbst als gegen Personen geführte Waffe. Kraftfahrzeuge sind leicht verfügbar, aufgrund ihrer Allgegenwart bis zum Tatort unauffällig und insbesondere in terroristischen Attacken gegen Menschenmengen von erschreckender Gefährlichkeit. Es ist deshalb davon auszugehen, dass auch selbstfahrende Fahrzeuge in dieser Weise genutzt werden, sobald sie verfügbar sein sollten. Das Fahrzeug müsste dann nicht einmal mehr von dem*der Täter*in selbst gesteuert werden. Dabei ist allerdings zu beachten, dass die verbaute Sicherheitstechnik einen zweckwidrigen Einsatz im Vergleich zur heutigen Situation bei Anwesenheit im Fahrzeug deutlich erschweren dürfte und durch die Täter*innen erst umgangen werden müsste. Es scheint allerdings nicht vorstellbar, dass die gezielte Zweckentfremdung durch den*die Verwender*in in jeder Situation durch Sicherheitsvorkehrungen verhindert werden kann.[40]

Ähnliches gilt für den Einsatz von Drohnen. Auch diese werden in nicht autonomen Varianten bereits heute durch Kriminelle eingesetzt, etwa für Drogen- oder Waffenschmuggel in Gefängnisse.[41] Auch hier sind Angriffe mit autonomen Varianten, ggf. organisiert in großer Zahl oder mit Sprengstoff oder Waffen bestückt, denkbar. Eine Steuerung durch die Täter*innen in Reichweite wäre dann nicht mehr notwendig, was Gegenmaßnahmen erschweren würde.[42]

2. Strafrechtliche Verantwortlichkeit

Gerade im Zusammenhang mit dem autonomen Fahren stellen sich zudem schwierige Fragen bezüglich der strafrechtlichen Verantwortlichkeit.[43] Während bisher vor allem der gezielte Missbrauch künstlich intelligenter Systeme für kriminelle Zwecke im Fokus stand, also die Verwirklichung von Vorsatzdelikten, geht es hierbei insbesondere um eine denkbare Fahrlässigkeitsstrafbarkeit. Ob und wer überhaupt strafrechtlich zur Rechenschaft gezogen werden kann, wenn ein selbstfahrendes Kraftfahrzeug einen Unfall verursacht, bei dem andere Menschen verletzt oder getötet werden,[44] steht seit einiger Zeit im Zentrum einer lebhaft geführten Debatte in der Strafrechtswissenschaft,[45] erwähnt werden sollen daher nur einige Punkte, die für die kriminologische und viktimologische Rezeption dieser strafrechtlichen Aufarbeitung von Bedeutung sein können.

Zunächst kämen für eine strafrechtliche Verantwortlichkeit weiterhin die „Fahrer*innen" in Betracht, wenn man die Insassen autonomer Fahrzeuge überhaupt noch so bezeichnen kann. Diesen kann eine Sorgfaltspflichtverletzung aber nur dann vorgeworfen werden, wenn sie das Fahrzeug noch beherrschen und kontrollieren können, was im Fall vollständig autonomer Fahrzeuge gerade nicht mehr gegeben wäre.[46] Etwas anderes kann nur gelten, wenn sich schon vor Inbetriebnahme Fehler zeigen, das Fahrzeug trotz solcher Hinweise nicht gestoppt wird oder Benutzungsregeln missachtet werden.[47]

Es ist daher zu erwarten, dass die Hersteller der Fahrzeuge stärker in den Fokus des Strafrechts rücken könnten, wenn Fahrzeuge nicht dem Stand von Wissenschaft und Technik entsprechen oder Produktbeobachtungs- und betreuungspflichten verletzt wurden.[48] Auch hier bedürfte es für eine Strafbarkeit jedoch einer individuellen Sorgfaltspflichtverletzung, deren Nachweis in Anbetracht der arbeitsteiligen Entwicklung, Programmierung und Produktion des autonomen Fahrzeugs nur schwer gelingen dürfte.[49] Ganz ähnlich verhält es sich mit der ebenfalls in Betracht kommenden strafrechtlichen Verantwortlichkeit von Zulieferern oder Zulassungsverantwortlichen.[50]

---

39 *Beuth,* Twitter-Nutzer machen Chatbot zur Rassistin, online abrufbar unter: https://www.zeit.de/digital/internet/2016-03/microsoft-tay-chatbot-twitter-rassistisch (zuletzt abgerufen am 13.7.22).

40 *Caldwell/Andrews/Tanay/Griffin* (Fn. 22), S. 7.

41 *Hans,* Der Himmel über Stadelheim, online abrufbar unter: https://www.sueddeutsche.de/muenchen/muenchen-jva-drohnen-abschuss-1.5068621 (zuletzt abgerufen am 13.7.22).

42 *Caldwell/Andrews/Tanay/Griffin* (Fn. 22), S. 9.

43 *Hilgendorf,* in: FS Roßnagel, 2020, S. 545 (554 f.); *Lohmann,* Strafrecht im Zeitalter von KI, 2021; *Schuster,* in: Beck/Kusche/Valerius, Digitalisierung, Automatisierung, KI und Recht, 2020, S. 387.

44 Für den zivilrechtlichen Schadensersatz greift die Gefährdungshaftung des Halters (§ 7 StVG). Die strafrechtliche Verantwortlichkeit basiert jedoch auf der individuellen Schuld des Täters, hier ist ein solcher Weg daher aus guten Gründen versperrt, dazu *Hilgendorf* (Fn. 43), S. 545 (554); *ders.,* JA 2018, 801 (803).

45 Siehe für einen Überblick *Hilgendorf,* JA 2018, 801; allgemein zu den strafrechtlichen Fragen *Chibanguza/Kuß/Steege/Lutz,* Künstliche Intelligenz, 2022, S. 389 ff.; *Steinert,* SVR 2019, 5; zur strafrechtlichen Verantwortlichkeit *Sander/Hollering,* NStZ 2017, 193; *Schuster,* DAR 2019, 6; *Sandherr,* NZV 2019, 1; zu den Sorgfaltspflichten *Valerius,* in: Hilgendorf, Autonome Systeme und neue Mobilität, 2017, S. 9; zur Perspektive der Strafverteidigung *Staub,* NZV 2019, 392; zu Dilemmasituationen *Beck,* in: Hilgendorf , 2017, S. 117; *Joerden,* in: Hilgendorf, 2017, S. 73; *Hilgendorf,* in: Hilgendorf, 2017, S. 143; *Weber,* NZV 2016, 249.

46 *Schuster,* DAR 2019, 6 (11).

47 *Hilgendorf,* JA 2018, 801 (803); *Schuster* (Fn. 43), S. 395.

48 *Sander/Hollering,* NStZ 2017, 193 (197 f.); *Schuster,* DAR 2019, 6 (8 f.); *Schuster* (Fn. 43), S. 396 ff.; vgl. die bereits heute in § 1f Abs. 3 StVG geregelten Pflichten.

49 *Schuster,* DAR 2019, 6 (9).

50 *Sander/Hollering,* NStZ 2017, 193 (199).

Zu den nicht nur philosophisch, sondern auch juristisch meistdiskutierten Fragen in diesem Zusammenhang gehört das „richtige" Verhalten autonomer Fahrzeuge in Dilemmasituationen.[51] Auch auf diese soll hier nicht vertieft eingegangen werden, den meisten dürften die Problemstellungen bekannt sein: [52] Wie soll sich das Fahrzeug verhalten, wenn trotz aller Sicherheitsfunktionen eines von mehreren Schadensszenarien nicht mehr zu vermeiden ist? Soll eine Abwägung der zu erwartenden Schäden erfolgen und wenn ja, nach welchen Kriterien? Nach Alter, Anzahl oder Vorverhalten der betroffenen Personen? Nach dem ursprünglich eingeschlagenen Fahrweg? Nach dem bestmöglichen Schutz für die Fahrzeuginsassen? Die besondere Härte dieser Entscheidung für die Programmierung autonomer Fahrzeuge folgt daraus, dass sie anders als bei heutigen Unfallsituationen nicht mehr in Bruchteilen von Sekunden durch die Fahrer*innen erfolgt, die – wenn überhaupt – nur noch intuitiv reagieren können, sondern das Ergebnis einer abstrakt generalisierenden Wertungsentscheidung darstellt.[53] Die Leitlinien der Abwägung müssen auf Basis einer breiten gesellschaftlichen Diskussion durch den Gesetzgeber getroffen werden. Die Entscheidung darf nicht auf die einzelnen Hersteller oder gar die einzelnen Programmierer*innen abgewälzt werden, die sich dann ihrerseits der Gefahr einer strafrechtlichen Verfolgung aussetzen würden.

Abgesehen von diesen Einzelsituationen soll an dieser Stelle auf ein weiteres Dilemma hingewiesen werden, das in den bisherigen Ausführungen deutlich geworden ist. Es zeichnet sich ab, dass der Einsatz autonomer Fahrzeuge in Zukunft insgesamt zu einer Reduzierung von Verkehrsunfällen führen wird. Solange diese aber vorkommen, wird es auch ein Bedürfnis der Geschädigten oder ihrer Hinterbliebenen nach strafrechtlicher Verantwortlichkeit für das Geschehene geben. Eben diese individuelle Vorwerfbarkeit wird mit zunehmender Automatisierung jedoch seltener oder jedenfalls schwieriger nachzuweisen und zu verfolgen. Die gesellschaftlich wünschenswerte Entwicklung hin zu weniger Unfällen wird also in Widerspruch zu der individuell unbefriedigenden Situation treten, dass eine strafrechtliche Verantwortung für den Einzelfall häufig nicht geklärt werden kann.

IV. KI auf Opferseite

Die Fortschritte auf dem Gebiet der künstlich intelligenten Systeme können aber auch auf der Seite der Kriminalitätsopfer Folgen haben. Einerseits kann ihre Verwendung zum Einfallstor für Täter*innen werden (1.), andererseits bieten sich aber auch Chancen, den Zugang zum Recht zu erleichtern (2.).

*1. KI als Einfallstor für Täter*innen*

Die KI kann unter anderem im Rahmen sog. „Adversarial Attacks" zum Ziel krimineller Angriffe werden. Dabei werden Input-Daten gezielt so manipuliert, dass sie von dem maschinell lernenden System falsch klassifiziert werden. Die Manipulation kann dabei u.U. so subtil sein, dass sie für menschliche Betrachter*innen nicht als solche zu erkennen ist.[54] Für den Bereich der Bilderkennung lässt sich auch dies am Beispiel des autonomen Fahrens illustrieren. So konnte gezeigt werden, dass sich die automatisierte Erkennung eines Stoppschildes manipulieren lässt, indem man das Verkehrsschild gezielt mit schwarzen und weißen Stickern beklebt.[55] Für das menschliche Auge war das Stoppschild weiterhin als solches zu erkennen und man hätte die schwarzen und weißen Punkte wahrscheinlich bei flüchtigem Hinsehen als Graffiti o.ä. abgetan. Die KI lag bei dieser Manipulation jedoch in manchen Konstellationen zu 100 % falsch und interpretierte das Stoppschild etwa als Anweisung, 45 km/h zu fahren. Welche Gefahren sich daraus sowohl für gezielte Attacken gegen einzelne Fahrzeuginsass*innen als auch für terroristische oder kriminelle Angriffe gegen willkürliche Opfer ergeben können, dürfte offensichtlich sein.

Ein anderes Beispiel für KI auf Opferseite als Einfallstor für kriminelle Taten kann in Angriffen auf sog. Smart Speaker, also intelligente, sprachbasierte Assistenzsysteme gesehen werden, die in mehr und mehr Wohnungen zu finden sind.[56] Durch diese Installation tief in der Privatsphäre und die Vielzahl an Konten, die für eine möglichst bequeme Nutzung hinterlegt werden, stellen Amazon Echo und Co. sehr attraktive Ziele für Hacker*innen dar.[57] Das verstärkt sich noch, je mehr „Smart Home"-Anwendungen – bis hin zum Haustürschloss – verbunden werden.

2. KI zur Opferunterstützung

Abgesehen von diesen Gefahren können KI-Systeme aber andererseits auch genutzt werden, um Opfer von Straftaten zu unterstützen, indem ihnen insbesondere der Zugang zum Recht erleichtert wird.

Im Bereich des niedrigschwelligen Zugangs zum Recht haben nicht nur KI-gestützte Systeme, sondern Legal Tech-Anwendungen allgemein bisher im juristischen Bereich ihre große Stärke. Dies spielt bisher im Privatrecht eine deutlich größere Rolle, man denke nur an standardisierte Tools für Massenverfahren wie „Flightright" oder „Conny". Doch auch für Opfer von Straftaten können automatisierte Systeme ihren Beitrag leisten, um etwa die Hemmschwelle zu einer Anzeige bei den Strafverfolgungsbehörden zu senken.

---

51 Statt vieler *Weber,* NZV 2016, 249 m.w.N.

52 Zumeist illustriert am berühmten Weichensteller-Problem, vgl. *Welzel,* ZStW 1951, 47 (51 ff.).

53 *Sander/Hollering,* NStZ 2017, 193 (201 f.).

54 *Kurakin et al.*, Adversarial Attacks and Defences Competition (2017), S. 2, online abrufbar unter: https://bit.ly/3ztjBzB (zuletzt abgerufen am 18.7.2022).

55 *Eykholt et al.,* Robust Physical-World Attacks on Deep Learning Visual Classification (2018), 1625 ff., online abrufbar unter: https://openaccess.thecvf.com/content_cvpr_2018/papers/Eykholt_Robust_Physical-World_Attacks_CVPR_2018_paper.pdf (zuletzt abgerufen am 18.7.2022).

56 https://de.statista.com/prognosen/999788/deutschland-besitz-von-smart-home-geraeten (zuletzt abgerufen am 13.7.22).

57 Beispiele für Vorgehensweisen vgl. https://www.srlabs.de/bites/smart-spies (zuletzt abgerufen am 13.7.22).

Ein Beispiel ist das Projekt zur Entwicklung eines Chat-Bots zur Aufbereitung von Anzeigen im digitalen Raum durch die Zentral- und Ansprechstelle Cybercrime (ZAC NRW).[58] Dieser soll insbesondere dabei helfen, Anzeigen bei hate speech im Internet zu strukturieren und Polizei und Staatsanwaltschaft so die Möglichkeit geben, die Taten mit einer größeren Wahrscheinlichkeit aufklären zu können. Der Schritt aus der digitalen Welt, in der die Beleidigung stattfindet, zu einer analogen Anzeige bei der Polizei wird selten gegangen, was dazu beiträgt, dass sich Täter*innen von hate speech allzu sicher fühlen können, mit ihren Taten davonzukommen. Hier setzen derartige Projekte an und sollen unkomplizierte, digitale Alternativen ermöglichen.[59]

V. KI als potenzielles Mittel der Sozialkontrolle

Die Einsatzmöglichkeiten von KI durch die Sicherheitsbehörden gehen jedoch noch weit über diese Verwendung hinaus. Im Sinne eines Einsatzes zur sozialen Kontrolle bieten sich Potenziale sowohl für die Gefahrenabwehr (1.), als auch für die Ermittlungsarbeit der Strafverfolgungsbehörden (2.) und den Strafprozess (3.).

1. Gefahrenabwehr

a) Predictive Policing

Lernende Systeme können in der präventiven Polizeiarbeit zum Beispiel dabei helfen, Prognosen zu erstellen. Diese als „Predictive Policing" bezeichneten Prognosen werden gemeinhin in personenbezogenes und ortsbezogenes Predictive Policing unterschieden.[60]

Bei personenbezogenem Predictive Policing geht es vor allem um die algorithmenbasierte Erstellung einer Legalprognose.[61] Dies kann für die Prävention relevant sein, aber auch für die Sanktionsentscheidung, beispielsweise ob (der Rest) eine(r) Freiheitsstrafe zur Bewährung ausgesetzt werden kann.[62] Beim ortsbezogenen Predictive Policing hingegen werden lokale Kriminalitätsschwerpunkte bestimmt, an denen mit der größten Wahrscheinlichkeit Straftaten begangen werden und an denen daher eine erhöhte Polizeipräsenz notwendig erscheint.[63]

Die meistdiskutierten Erfahrungen mit derartigen Systemen sind in den Vereinigten Staaten gemacht worden.[64] Dabei konnten auf den ersten Blick „Erfolge" verzeichnet werden. Einige Prognosetools erzielten tatsächlich höhere Trefferquoten als menschliche Expert*innen.[65] Gleichzeitig wurden aber auch Probleme deutlich, die sich in ähnlicher Weise auch für andere Verwendungen künstlich intelligenter Systeme durch die Sicherheitsbehörden stellen: Zunächst ist die bekannte „Black Box"-Problematik zu nennen, der sich alle lernenden, induktiven Systeme gegenübersehen.[66] Es kann zum jetzigen Zeitpunkt nicht im Einzelnen erklärt werden, warum ein bestimmter Risikoscore durch das Tool ausgegeben wird. In so grundrechtssensiblen Bereichen wie der Gefahrenabwehr und der Strafverfolgung stellt sich dann die Frage, ob dieser Score überhaupt herangezogen werden darf, wenn er nicht im Einzelnen begründet werden kann. Hier ließe sich einwenden, dass auch die Prognoseentscheidung durch menschliche Richter*innen insofern intransparent sein kann, als nicht alle bewussten wie unbewussten Einflussfaktoren offengelegt werden (können). Mit der Gleichsetzung der Intransparenz menschlicher und algorithmenbasierter Entscheidungen sollte aber überaus vorsichtig umgegangen werden.[67] Umso notwendiger ist weitere Forschung im vollkommen zurecht im Trend liegenden Feld der „Explainable AI".[68]

Neben der Intransparenz der Entscheidungen stellt sich vor allem das Problem der Diskriminierung durch das personenbezogene Predictive Policing. Am prominentesten kritisiert wurde dies bei dem US-amerikanischen Risk-Assessment-Tool COMPAS, dem einige Untersuchungen höhere false-positive Raten bei PoC im Gegensatz zu weißen Testpersonen attestierten,[69] auch wenn die Studienlage diesbezüglich nicht eindeutig ist.[70] Es besteht stets die Gefahr, dass in den Trainingsdaten – und damit in den menschlichen Entscheidungen der Vergangenheit – angelegte Diskriminierungen oder jedenfalls Ungleichbehandlungen von der KI reproduziert werden und sich so verfestigen.

b) Intelligente Videoüberwachung

Unter dem Begriff der „intelligenten Videoüberwachung" wird zudem die Nutzung von Bilderkennungssoftware für die automatisierte Auswertung von Videoaufnahmen durch die Polizei diskutiert, die in den Polizeigesetzen ei-

---

58 https://www.justiz.nrw.de/JM/schwerpunkte/zac/index.php (zuletzt abgerufen am 13.7.22).

59 Vgl. zum Ganzen die Ausführungen von *OStA Markus Hartmann* im Podcast „Talking Legal Tech" des Legal Tech Lab Cologne, Folge 45 vom 19.10.21, online abrufbar unter: https://anchor.fm/legaltech/episodes/45-Wie-Legal-Tech-der-Staatsanwaltschaft-bei-der-Aufdeckung-von-Cybercrime-hilft-mit-Oberstaatsanwalt-Markus-Hartmann-e17smcc/a-a6np0sn (zuletzt abgerufen am 18.7.2022).

60 *Singelnstein,* NStZ 2018, 1.

61 *Sommerer,* Personenbezogenes Predictive Policing, 2020, S. 37.

62 Butz/Christoph/Sommerer/Harrendorf/Kaspar/Höffler, BewHi 2021, S. 241.

63 *Sommerer* (Fn. 61), S. 36 f.

64 *Singelnstein,* NStZ 2018, 1 (2).

65 *Holsinger et al.,* A Rejoinder to Dressel and Farid, in: Federal Probation 82 (2018), 51 (54); *Lin et al.*, The limits of human predictions of recidivism, in: Science Advances 6 (2020), S. 5; a.A. zum gleichen Tool *Dressel/Farid,* Accuracy, Fairness, and Limits of Predicting Recidivism, in: Science Advances 4 (2018), S. 3 f.

66 Statt vieler *Martini,* Blackbox Algorithmus - Grundfragen einer Regulierung künstlicher Intelligenz, 2019.

67 *Sommerer* (Fn. 61), S. 226 ff.

68 *Samek et al.*, Explainable AI (2019); *Bhatt et al.,* Explainable Machine Learning in Deployment (2020).

69 *Angwin et al.*, Machine Bias (2016), online abrufbar unter: https://www.propublica.org/article/machine-bias-risk-assessments-in-criminal-sentencing.

70 Zu einem anderen Ergebnis anhand des gleichen Datensatzes kamen etwa *Dieterich/Mendoza/Brennan,* COMPAS Risk Scales (2016), S. 1 ff. und *Flores/Bechtel/Lowenkamp,* False Positives, False Negatives, and False Analyses, 2016, S. 43.

niger Bundesländer bereits eine rechtliche Regelung erfahren hat.[71] Zu unterscheiden ist zwischen der biometrischen Gesichtserkennung, Verhaltens- und Situationsanalysen sowie der Objekterkennung. So sollen, insbesondere an Kriminalitätsschwerpunkten wie Bahnhöfen, Videoaufnahmen in Echtzeit ausgewertet werden, um gefährliche Situationen frühzeitig erkennen und darauf polizeilich reagieren zu können.[72] Durch den Effizienzgewinn versprechen sich die Befürworter*innen einer solchen Technik vor allem einen geringeren Personalaufwand, als er zur menschlichen Überwachung von Videoaufnahmen erforderlich ist.[73]

Der in einer solchen dauerhaften Auswertung liegende Eingriff in die informationelle Selbstbestimmung der von der Videoüberwachung erfassten Personen wiegt umso schwerer, als die Streubreite der Maßnahme enorm ist. Um einige gefährliche Situationen erkennen zu können, wird eine Vielzahl unbeteiligter Bürger*innen anlasslos einem Datenbankabgleich unterzogen.[74] Neben diesen tatsächlichen Grundrechtseingriffen ist auch der darüberhinausgehende Einschüchterungseffekt nicht zu unterschätzen, der aus dem subjektiven Empfinden einer Dauerüberwachung[75] folgen kann, wenn die intelligente Videoüberwachung vermehrt zum Einsatz kommen sollte.[76]

Aus der kriminologischen Forschung ist zudem bekannt, dass durch reine Überwachungsmaßnahmen Kriminalität lediglich verdrängt wird – an andere Orte, auf andere Zeitfenster, in Variationen der Begehungsweisen[77], und deshalb eine intelligentere und sozialere präventive Kriminalpolitik angestrebt werden sollte.

c) Sonstiges Data Mining

Die Mustererkennung in großen Datenmengen ist die Paradedisziplin der KI. Es verwundert daher nicht, dass sich Polizei und Geheimdienste dieses Potenzial – insbesondere zur Terrorabwehr – auch über die intelligente Videoüberwachung hinaus zunutze machen möchten. Diesem sog. „Data Mining" durch Sicherheitsbehörden hat das *BVerfG* mit seinem 2020 ergangenen Beschluss zum Antiterrordateigesetz (ATDG) jedoch Grenzen gesetzt.[78] Die Eingriffsintensität steigt dabei, je mehr Behörden Zugriff auf die analysierbaren Daten haben[79] und je mehr Möglichkeiten bestehen, komplexe Verknüpfungen herzustellen, aus denen sich neue Aussagekraft auch aus einzeln betrachtet uninteressanten Daten ergeben kann.[80] Gleichzeitig lässt sich die Durchschlagskraft erhöhen, je mehr Daten sowohl zu Trainingszwecken der lernenden Algorithmen als auch zur späteren Auswertung zur Verfügung stehen. Wie groß das Interesse insbesondere von Geheimdiensten an einem Zugriff auf möglichst große Datensätze ist, lässt sich beispielhaft an der kürzlich verkündeten Kooperation des britischen Geheimdienstes GCHQ mit dem zum Amazon-Konzern gehörenden Clouddienstleister AWS erahnen.[81]

2. Strafverfolgung

Das Potenzial der Bild- (a.) und Texterkennung (b.) durch künstlich intelligente Systeme ist aber nicht nur präventiv zur Gefahrenabwehr nutzbar. Auch von Strafverfolgungsbehörden können die Anwendungen zur Beweismittelauswertung verwendet werden.

a) Bilderkennung zur Beweismittelauswertung

Die bereits beschriebene Bilderkennung in Videoaufnahmen öffentlicher Räume kann dabei natürlich auch für die repressive Ermittlungsarbeit nützlich sein, um große Datenmengen auszuwerten.[82]

Großen Nutzen verspricht die Bilderkennung zurzeit allerdings vor allem für die strafrechtliche Verfolgung von Kinderpornografie.[83] Wird bei Ermittlungen in diesem Kriminalitätsbereich potenzielles Beweismaterial beschlagnahmt, stehen die Ermittlungsbehörden vor der Aufgabe, immer größere Datenmengen auswerten zu müssen.[84] Das potenziell kinderpornografische Material zu sichten und strafbare von nicht strafbaren Inhalten zu unterscheiden, stellt auch für erfahrene und psychologisch begleitete Ermittler*innen eine Belastung dar. Hier könnten KI-Systeme eine große Hilfe sein, um das Beweismaterial (vor) zu sortieren. So würden nicht nur die Ermittlungspersonen psychisch entlastet, sondern es könnten auch mehr Verfahren mit den zur Verfügung stehenden personellen Ressourcen betrieben werden.[85]

Die künstliche Erstellung von Bildmaterial kann den Ermittlungsbehörden zudem bei der Ermittlung und Bekämpfung von Tauschplattformen für kinderpornografisches Material im Darknet helfen. Dort werden häufig sogenannte „Keuschheitsproben" verlangt, um Zugang zu den einschlägigen Foren zu erhalten. Diese können nur

71 Siehe z.B. § 59 SächsPVDG für grenzüberschreitende Kriminalität.
72 *Kulick,* NVwZ 2020, 1622.
73 LT-Drs. BW 16/2741, S. 28.
74 *Kulick,* NVwZ 2020, 1624.
75 Angelegentlich dieser Stelle sei die Lektüre des Romans „Every" von Dave Eggers empfohlen, in dem eine dystopischen Welt gezeichnet wird, in der die totale Überwachung mit dem Ziel der totalen Straftatvermeidung bzw. allgemeinen Verhaltensmodifikationen herrscht. Entgegen der Annahme, dass sich die Menschen gegen die immer stärker zunehmende Überwachung wehren, wird dem Leser eine erschreckende Zukunftsvision vor Augen geführt.
76 BVerfGE 150, 244, 268 = NJW 2019, 827 (Kfz-Kennzeichen II).
77 *Eisenberg/Kölbel* (Fn. 2), § 52 Rn. 22 m.w.N.; *Göppinger/Schneider* (Fn. 11), § 30 Rn. 22; *Meier* (Fn. 2), § 10 Rn. 26 m.w.N.; *Singelnstein/Kunz* (Fn. 2), § 22 Rn. 35 m.w.N.
78 *BVerfG*, NVwZ 2021, 226 = NJW 2021, 690; *Golla,* NJW 2021, 667.
79 *Golla,* NJW 2021, 668; zum „informationellen Trennungsprinzip" zwischen Geheimdienst und Polizei siehe nur *BVerfG*, Urt. v. 26.4.2022 – 1 BvR 1619/17, Rn. 170 ff.
80 Vgl. BVerfGE 115, 320 (350); *Golla,* NJW 2021, 669.
81 *Warrel/Flides*, Amazon strikes deal with UK, Financial Times, online abrufbar unterhttps://www.ft.com/content/74782def-1046-4ea5-b796-0802cfb90260 (zuletzt abgerufen am 13.7.22).
82 *Kulick,* NVwZ 2020, 1622 (1625 f.); *Burkhardt,* Kriminalistik 2020, 336 (337).
83 *Kaulartz/Braegelmann/Peters* (Fn. 3), Kap. 12 Rn. 71; Justizministerium NRW, Pressemitteilung v. 5.8.2019, online abrufbar unter: https://www.land.nrw/pressemitteilung/kuenstliche-intelligenz-im-kampf-gegen-kinderpornographie (zuletzt abgerufen am 13.7.22).
84 *Burkhardt,* Kriminalistik 2020, 336 (337).
85 Vgl. die Ausführungen von *OStA Markus Hartmann* (Fn. 59).

„bestanden“ werden, indem kinderpornografisches Material hochgeladen wird.[86] Durch die darin liegende Strafbarkeit soll die Infiltrierung durch Ermittlungsbehörden verhindert werden.[87] Künstlich erstelltes Material kann den Ermittler*innen die Möglichkeit geben, Zugang zu derartigen Foren zu erhalten, diese zu schließen und die beteiligten Personen zu ermitteln, um weitere Taten zu verhindern. Ob den Ermittlungsbehörden ein solches Vorgehen trotz des erstrebenswerten Ziels erlaubt sein soll, ist jedoch höchst umstritten.[88]

b) Texterkennung zur Beweismittelauswertung

Auch Texterkennungssysteme können zur Strukturierung großer Datensätze herangezogen werden, die Ermittlungsbehörden sonst vor nicht zu bewältigende Herausforderungen stellen würden. Dies kann insbesondere im Bereich der Wirtschaft- und Steuerkriminalität von Bedeutung werden. Als Anwendungsbeispiel können die sogenannten „Panama Papers“ oder auch die „Paradise Papers“ dienen.[89] Die geleakten Dokumente über Offshore-Konten bieten Anlass für eine Vielzahl von Ermittlungen, vor allem im Bereich der Steuerhinterziehung und der Geldwäsche.[90] Es kann dabei nötig werden, Wiederholungen bekannter Muster in den komplexen Verschleierungsstrukturen zu erkennen und die Ermittler auf solche hinzuweisen – die große Stärke künstlich intelligenter Systeme.[91] Wenn es gelingt, lernende Systeme auf typische Hinterziehungsstrukturen zu trainieren, kann das noch lange nicht ausgeschöpfte Potenzial der Leaks als Informationsquelle genutzt werden. Man hätte endlich einen Metalldetektor für die Nadel im Heuhaufen.

3. Strafprozess

Im Strafprozess ist es denkbar, dass KI-Systeme vor allem unterstützend eingesetzt werden können.

a) Unterschied zum Zivilprozess

Viele Legal Tech-Anbieter, von denen freilich nicht alle, aber doch einige auf KI-gestützte Systeme setzen, sehen das Potenzial ihrer Anwendungen vor allem in einem erleichterten Zugang zum Recht für ihre Kund*innen.[92] Dieses Versprechen bezieht sich dabei vor allem auf das Zivilrecht, in dem Legal-Tech-Anwendungen noch sehr viel mehr diskutiert werden und in einigen Bereichen schon durchaus verbreitet sind.[93] Dies gilt für den Strafprozess aus mehreren Gründen nicht in gleicher Weise.[94]

Erstens gibt es im Strafprozess keinen vergleichbaren Bedarf nach einer Erleichterung des Zugangs zum Verfahren. Jedenfalls der*die potenzielle Täter*in möchte ja gerade nicht Teil des Verfahrens werden, das vielmehr der Staat anstrengt. Zweitens weist der Strafprozess regelmäßig eine höhere Grundrechtsrelevanz auf als das Zivilverfahren, so dass mehr Vorbehalte gegen die Verwendung neuer technischer Lösungen bestehen. Drittens können sich die Beteiligten nicht in gleicher Weise auf andere Wege der Konfliktlösung als das gerichtliche Verfahren im Rahmen ihrer Privatautonomie, ggf. unter Zuhilfenahme von KI-Lösungen, einigen. Und viertens gibt es im Bereich des Strafrechts weniger Innovation durch privatwirtschaftliche Unternehmen, weil die Gewinnaussichten nicht mit denen im zivilrechtlichen Bereich vergleichbar sind.

b) Urteilsvorhersage

In den Vereinigten Staaten ist es bereits vereinzelt gelungen, lernende System auf die Vorhersage von Gerichtsentscheidungen zu trainieren. So konnten etwa 70,2 % der Supreme Court Entscheidungen mittels einer Random Forest Methode korrekt vorhergesagt werden.[95] Ähnliche Systeme erscheinen vor allem aus Sicht der Verteidigung wünschenswert, um eine bestmögliche Prozessstrategie entwerfen zu können.[96] Zum jetzigen Zeitpunkt scheitern solche Systeme in Deutschland jedoch neben den systematischen Unterschieden zum anglo-amerikanischen Case Law[97] vor allem an der Verfügbarkeit einer ausreichenden Anzahl von Urteilen als Trainingsdaten. In Deutschland wird nicht einmal 1 % der ergangenen Urteile veröffentlicht,[98] darunter überwiegend obergerichtliche Rechtsprechung. Die Veröffentlichung von Instanzenrechtsprechung, die für die Schulung von Vorhersagesystemen von entscheidender Bedeutung wäre, stellt bislang die absolute Ausnahme dar,[99] ob sich dies bald ändert – wie im Koalitionsvertrag angekündigt[100] – bleibt abzuwarten.

*c) Entscheidungsunterstützung für Richter*innen*

KI-Systeme bieten jedoch auch für die richterliche Tätigkeit selbst Potenzial. Dabei geht es weniger um die – je nach Sichtweise utopische oder dystopische – Vorstellung eines Roboter-Richters,[101] sondern vielmehr um Ansätze für Unterstützungssysteme, die den Gerichten die alltägliche Arbeit erleichtern können.

---

86 *Gercke,* CR 2018, 480.
87 *Wittmer/Steinebach,* MMR 2019, 650.
88 *Wittmer/Steinebach,* MMR 2019, 650 (653); *Safferling,* DRiZ 2019, 206 (207); *Gercke,* CR 2018, 480.
89 *Kaulartz/Braegelmann/Peters* (Fn. 3), Kap. 12 Rn. 71.
90 *Schuhr,* NZWiSt 2017, 265; *Papathanasiou*, JA 2017, 88.
91 *Baur,* ZIS 2020, 275 (276); *Brüning,* in: Rotsch, Criminal Compliance – Status quo und Status futurus, 2021, S. 63, 75; *Kaulartz/Braegelmann/Peters* (Fn. 3), Kap. 12 Rn. 71.
92 Vgl. etwa prominente Beispiele wie Flightright, Conny etc., s.o.
93 Statt vieler *Hähnchen/Schrader/Weiler/Wischmeyer,* JuS 2020, 625 sowie aus anwaltlicher Sicht *Hellwig/Ewer,* NJW 2020, 1783.
94 Zu Legal Tech im Strafrecht *Valerius,* ZStW 2021, 152.
95 *Katz/Bommarito/Blackman,* A General Approach for Predicting the Behavior of the Supreme Court of the United States, 2017, online abrufbar unter: https://ssrn.com/abstract=2463244 (zuletzt abgerufen am 18.7.2022)
96 *Kuhlmann,* LTO v. 14.6.2019, online abrufbar unter: https://www.lto.de/recht/legal-tech/l/frankreich-legal-tech-beschraenkung-predictive-analysis-verbotene-erkenntnis/ (zuletzt abgerufen am 13.7.22).
97 *Hoch,* MMR 2020, 295 (297).
98 *Hamann,* JZ 2021, 656 (658).
99 *Hamann,* JZ 2021, 656.
100 „Mehr Fortschritt wagen”, Koalitionsvertrag zwischen SPD, Bündnis 90/Die Grünen und FDP, S. 106.
101 *Greco,* RW 2020, 29; *Dreyer/Schmees,* CR 2019, 758.

In Betracht kommt zum Beispiel eine Unterstützung bei der idealen Prozessplanung oder Terminansetzung, aber auch eine Unterstützung bei der Entscheidung selbst durch sogenannte Decision Support Systeme (DSS) ist denkbar.[102] Diese können etwa darauf ausgerichtet sein, die Effizienz der Entscheidung zu steigern, und sei es „nur" durch intelligente Textbausteine, die sich den bevorzugten Formulierungen der Richter*innen bei der Entscheidungsbegründung anpassen.[103] Aber auch qualitative Verbesserungen der gerichtlichen Entscheidungen sind im Grundsatz vorstellbar. Erhebliches Verbesserungspotenzial besteht etwa im Bereich der Strafzumessung, bei der die großen regionalen Unterschiede in der Strafhöhe für an sich vergleichbare Delikte seit Jahrzehnten bekannt sind.[104] Hier könnten Entscheidungs-Unterstützungssysteme genutzt werden, die üblicherweise für eine Tat verhängte Sanktion transparent zu machen und so zu mehr relativer Sanktionsgerechtigkeit beizutragen,[105] bzw. um jedenfalls eine technik-unterstützte Reflexions-„Schleife" einzuführen, die die Richter*innen zwingt, die von ihnen gewählte Strafe mit der in vergleichbaren Fällen „üblichen" Strafe zu vergleichen.[106] Abweichungen können so vielleicht bewusster gemacht werden, was idealerweise dazu führt, dass die Richter*innen sich verstärkt mit den spezifischen Besonderheiten des Falles auseinandersetzen. Freilich ist das auch ein wenig blauäugig – denn auf der anderen Seite weiß man aus psychologischen Studien, dass es dem Menschen schwerer fällt, sich über von „Technik", scheinbar objektiv, generierte Werte hinwegzusetzen (Automation Bias)[107], so dass die Gefahr besteht, dass Besonderheiten (auch denkbar unbewusst) „unter den Tisch gekehrt" werden, insbesondere weil sie zusätzliche Arbeit machen (Begründungslast!), und letztlich doch einfach der vom Computer ausgeworfene Wert übernommen wird. Gut vorstellbar ist, dass der (Königs-)Weg irgendwo dazwischen liegt; der imperfekte Mensch könnte durch Schulungen zum Umgang mit Vorschlägen eines Decision Support Systems besser darin werden, solche Ergebnisse zu verarbeiten, diesen etwas Kritisches entgegenzusetzen, und so könnte vielleicht in der Zukunft doch ein Zusatznutzen möglich sein.

VI. Schlussbemerkungen

Deutlich wurde, dass die KI enorme Gefahren durch Möglichkeiten schafft, mittels ihres missbräuchlichen Einsatzes Straftaten zu begehen. Es zeigt sich, dass innovative Methoden der Strafverfolgung und Prävention zum Teil ein notwendiges Spiegelbild der neuen Erscheinungsformen der Kriminalität darstellen, um mit den technologischen Entwicklungen Schritt zu halten und eine effektive Kriminalitätsbekämpfung weiterhin gewährleisten zu können.

Darüber hinaus stellen neue technische Möglichkeiten aber auch stets eine große Versuchung für den Sicherheitsapparat dar, diese für mehr Überwachung und Sicherheit zulasten der gesellschaftlichen Freiheit zu nutzen,[108] auch über ein der Bedrohungslage angemessenes Maß hinaus. Es bedarf daher klarer gesetzlicher Voraussetzungen und Anlassschwellen für den Einsatz künstlich intelligenter Systeme in der Strafverfolgung und der Gefahrenabwehr, um die Eingriffsintensität und die Streubreite der Maßnahmen zu regeln.

Dabei sind die mit einem Einsatz von KI im grundrechtssensiblen Bereich der Kriminalitätsbekämpfung einhergehenden Gefahren stets im Blick zu behalten: Die Probleme der Intransparenz lernender Systeme und der potenziellen Diskriminierung durch sie müssen für das hiesige Anwendungsfeld gelöst werden, sofern sie überhaupt lösbar sind, bevor ein flächendeckender Einsatz erfolgen kann.[109] Es klingt recht trocken: die Vorgaben an den Datenschutz, die Datensicherheit und die Datenübertragungssicherheit müssen gewährleistet werden. Darin verbirgt sich aber eine besondere Brisanz, denn was auf dem Spiel steht, ist letztlich die Freiheit selbst: Wie weit wollen wir gehen? Die totale Überwachung, wie im Roman „Every" von *Dave Eggers*[110]? Das „Überall Sichtbar Sein" schafft vielleicht ein solches Maß an Sozialkontrolle, dass kaum noch Straftaten begangen werden können, aber zu welchem Preis für all diejenigen, die auch ohne Überwachung rechtlich gesonnen sind. Denn wenn man diesen Gedanken zu Ende denkt: Wenn du immer beobachtet bist, machst du nicht mehr, was du willst, sondern was von dir erwartet wird. Der Mensch in der totalen Beobachtung ist nicht frei.

Strebt man also die totale bzw. jedenfalls scheinbare Gewissheit an, so zerstört dies letztlich die Freiheit.

Shoshanna Zuboff schreibt in ihrem vielgelesenen Buch „Das Zeitalter des Überwachungskapitalismus": „Freiheit ist ohne Ungewissheit undenkbar. Ungewissheit ist das Medium, in dem der menschliche Wille sich in Form von Versprechungen ausdrückt."[111]

---

102 *Nink,* Justiz und Algorithmen, 2021, S. 359 ff.

103 Vgl. für derartige Systeme im Journalismus *Gräfe/Kahl,* MMR 2021, 121.

104 *Grundies,* in: Neubacher/Bögelein, Krise-Kriminalität-Kriminologie, 2016, S. 511 (518 f.).

105 *Giannoulis,* Studien zur Strafzumessung, 2014; *Kohn,* KI und Strafzumessung, 2021.

106 *Kaspar/Höffler/Harrendorf,* NK 2020, 35 (50 ff.); *Butz/Christoph/Sommerer/Harrendorf/Kaspar/Höffler,* BewHi 2021, 241 (242).

107 *Mosier et al.*, Automation Bias, in: International Journal of Aviation Psychology, 8 (1998), 47; *Skitka et al.,* Does Automation Bias Decision-Making?, in: International Journal of Human-Computer Studies, 51 (1999), 991; *Skitka et al.,* Accountability and Automation Bias, in: International Journal of Human-Computer Studies, 52 (2000), 701; *Cummings*, Automation and Accountability in Decision Support System Interface Design, in: Journal of Technology Studies, 32 (2006), 23 (25); dazu auch *Sommerer* (Fn. 61), S. 71 ff.; *Butz/Christoph/Sommerer/Harrendorf/Kaspar/Höffler,* BewHI 2021, 241 (254).

108 *Singelnstein/Stolle,* Die Sicherheitsgesellschaft, 3. Auflage (2012), S. 32 f.; *Puschke,* in: FS Eisenberg, 2019, S. 695.

109 Dies eindrücklich anmahnend *Sommerer* (Fn. 61), S. 344 f.; *Singelnstein*, NStZ 2018, 1, 7.

110 Vgl. Fn. 75.

111 *Zuboff*, Das Zeitalter des Überwachungskapitalismus, 2018, S. 389.

Der nächste wichtige Punkt, nach dem Datenschutz, ist die Qualitätskontrolle: Es muss eine effektive und stetige Qualitätskontrolle sowohl technischer als auch juristischer Natur erfolgen.[112] Dies ist eine zentrale Grundvoraussetzung, damit künstlich intelligente Systeme überhaupt einen wertvollen Beitrag zu einer effektiven und verhältnismäßigen Kriminalitätsbekämpfung leisten können.

Zum Ende soll zudem noch eine Grundlinie i.S.e. Grenze gezeichnet werden: Jedenfalls die Letztentscheidung über die gerechte Strafe für menschliche Verfehlungen muss i.E. in menschlicher Hand bleiben.[113] Doch auch die Vorstufen sind schon umstritten. Teilweise wird vertreten, dass es keine richterliche Macht ohne richterliche Verantwortung geben soll.[114] Angeführt wird hierfür, stark zusammengefasst: Der Mensch ist verletzlich, und wie, in welcher Intensität, er von etwas (z.B. Strafe) getroffen wird, kann der Computer nie wirklich „kennen“,[115] das kann nur simuliert werden,[116] und daher – kurz gefasst – müssen so einschneidende Entscheidungen wie das Verhängen einer Sanktion von einem Menschen getroffen werden, einem Richter, der „das Leben kennt“.[117] Nur andere Entscheidungen, also Verwaltungsentscheidungen, sollen nach dieser Ansicht an die KI, den Roboter-Richter, überantwortet werden dürfen.[118] Andere sind freilich, was Teile der richterlichen Tätigkeit angeht, nicht so streng[119] und lassen mehr zu.

Da das Forschungsfeld äußerst dynamisch ist, mit Blick auf die Neuerungen aus den Computerwissenschaften, aber auch mit Blick auf die psychologischen Erkenntnisse zur Entscheidungsfindung, soll hier der Streit nicht entschieden werden, doch eben diese äußerste Grenze markiert werden: die Letztentscheidung muss in menschliche Hand, selbst wenn dieser Hand auch Fehler unterlaufen können. Dafür spricht nicht zuletzt auch ein spezialpräventiver Gedanke: Der Mensch als Gegenüber, als Straftäter*in vor dem*der Richter*in, braucht bei einem so schweren Eingriff, der in der Strafe liegt, auch den Diskurs, der im menschlichen Urteil liegt.

---

112 CEPEJ, European ethical Charter on the use of Artificial Intelligence in judicial systems and their environment, 2018, S. 10.

113 So auch CEPEJ (Fn. 112), S. 12; *Nink* (Fn. 102), S. 356 f.; *Greco,* RW 2020, 29.

114 *Greco,* RW 2020, 29.

115 Vgl. *Nida-Rümelin/Weidenfeld,* Digitaler Humanismus, 3. Aufl. (2018), S. 110.

116 *Turkle,* Alone Together. Why We Expect More from Technology and Less from Each Other, 3. Auflage (2017), S. 124; *Nida-Rümelin/Weidenfeld* (Fn. 115), S. 41.

117 *Greco,* RW 2020, 29 (49 f.).

118 *Greco,* RW 2020, 29 (53).

119 *Benning,* in: FS Herberger, 2016, S. 61 (67); *Chen*, Artificial Intelligence and Law 27 (2019), 15 (16); insb. zu Unterstützungssystemen *Nink* (Fn. 102), S. 452 ff.; *Kohn* (Fn. 105), S. 372 f., 376 (These 7).