

Möglichkeiten und Grenzen des Einsatzes von Künstlicher Intelligenz im Jugendstrafverfahren

Johannes Kaspar

Angaben zur Veröffentlichung / Publication details:

Kaspar, Johannes. 2023. "Möglichkeiten und Grenzen des Einsatzes von Künstlicher Intelligenz im Jugendstrafverfahren." In *Kriminalität und Kriminologie im Zeitalter der Digitalisierung*, edited by Thomas Bliesener, Lena Deyerling, Arne Dreißigacker, Isabell Henningsmeier, Merten Neumann, Jonas Schemmel, Carl Philipp Schröder, and Laura Treskow, 99–112. Mönchengladbach: Forum Verlag Godesberg.

Nutzungsbedingungen / Terms of use:

licgercopyright

Dieses Dokument wird unter folgenden Bedingungen zur Verfügung gestellt: / This document is made available under these conditions:

Deutsches Urheberrecht

Weitere Informationen finden Sie unter: / For more information see:

<https://www.uni-augsburg.de/de/organisation/bibliothek/publizieren-zitieren-archivieren/publiz/>



Möglichkeiten und Grenzen des Einsatzes von Künstlicher Intelligenz im Jugendstrafverfahren

Johannes Kaspar

Gliederung

- | | |
|---|---|
| 1. Einführung und Begriffsbestimmungen | 4. Potenzielle Anwendungsfelder im Jugendstrafverfahren |
| 2. Formen von KI | 5. Denkbare Vorteile |
| 3. Potenzielle Anwendungsfelder im Strafverfahren | 6. Probleme und Risiken |
| | 7. Fazit |

1. Einführung und Begriffsbestimmungen

Der Einsatz digitaler Technologien hat sämtliche Lebensbereiche erfasst und auch nicht vor dem juristischen Bereich Halt gemacht.¹ Zunehmend wird in der juristischen Praxis und der rechtswissenschaftlichen Diskussion über „Legal Tech“ bzw. den Einsatz von „Künstlicher Intelligenz“ im Strafverfahren gesprochen.² Erstaunlicherweise wurde die Frage eines Einsatzes dieser Technologien im Bereich des Strafrechts bislang nur selten in spezifischer Form mit Bezug auf das Jugendstrafverfahren diskutiert. Insofern hat es sich der vorliegende Beitrag zum Ziel gesetzt, potenzielle Anwendungsfelder von „Künstlicher Intelligenz“ im Bereich des Jugendstrafverfahrens anzusprechen und die damit einhergehenden Chancen, aber auch potenziellen Risiken auszuloten.

Der Einsatz von „Legal Tech“ hat einen denkbar breiten Anwendungsbereich und reicht von einfachen Formen der Unterstützung durch Computertechnik (etwa durch die Nutzung von juristischen Datenbanken) bis hin zur Zukunftsvision einer vollständigen Delegation juristisch verbindlicher Entschei-

1 Kment/Borchert (2022); Graichen (2022); Nink (2021).

2 Vgl. Peters (2022), S. 575 ff.; Christoph (2021); zur Strafzumessung s. etwa Rosztalski/Völkening (2019), S. 265; Kaspar/Höffler/Harrendorf, (2020), S. 35; Kohn (2020); krit. Ruppert (2021), S. 90; Greco (2020), S. 29; zur Kriminalprognose s. Butz et al. (2021), S. 241; Eisbach/Heghmanns/Hertel (2022), S. 489.

dungen an Computersysteme.³ Der Begriff der „Künstlichen Intelligenz“ (welche als „Legal Tech-Kerntechnologie“⁴ bezeichnet wird) wird nicht einheitlich verstanden, was eine Diskussion erschwert. Im Kern geht es um den Vergleich von natürlicher und künstlicher Informationsverarbeitung, wie es schon auf der berühmten Dartmouth-Conference im Jahre 1955 zu Grunde gelegt wurde, auf der sich eine Gruppe von Wissenschaftlerinnen und Wissenschaftlern über die Grenzen und Möglichkeiten einer computergestützten Informationsverarbeitung Gedanken machte und in diesem Zusammenhang den Begriff der „Artificial Intelligence“ prägte.⁵ Sucht man nach Definitionsversuchen in Normtexten, so stößt man u. a. auf einen Rechtsakt der EU-Kommission aus dem Jahre 2018, in dem Künstliche Intelligenz folgendermaßen beschrieben wird: „Systeme mit einem ‚intelligenten‘ Verhalten, die ihre Umgebung analysieren und mit einem gewissen Grad an Autonomie handeln, um bestimmte Ziele zu erreichen“.⁶ Die offensichtliche Schwäche solcher Ansätze ist, dass sie tautologisch bzw. zirkulär sind, soweit man den Begriff der Intelligenz mit einem „intelligenten“ Verhalten definiert. Man stößt dann auf das Folgeproblem, dass auch der Begriff der „Intelligenz“ bislang nicht einheitlich und übereinstimmend definiert wird.⁷ Vielleicht hat die EU-Kommission auch aus diesem Grund in ihrem aktuellen Vorschlag einer KI-Verordnung einen anderen Weg gewählt. Sie beschreibt nunmehr Künstliche Intelligenz phänomenologisch anhand bestimmter technologischer Ausprägungen und Erscheinungsformen. So wird etwa auf Konzepte des maschinellen Lernens Bezug genommen, aber auch auf andere Konzepte einschließlich Wissensrepräsentation, induktiver (logischer) Programmierung und auf symbolische Schlussfolgerungs- und Expertensysteme.⁸

2. Formen von KI

Damit sind zwei besonders wichtige KI-Technologien benannt, nämlich einerseits das maschinelle Lernen und andererseits regelbasierte Systeme inklusive sog. Expertensysteme.⁹ Bei den regelbasierten Systemen wird ein System

³ Kaspar/Höffler/Harrendorf (2020), S. 35 ff.

⁴ Herberger (2018), S. 2825.

⁵ Kohn (2020), S. 38.

⁶ EU-Kommission, COM (2018) 237 final, 25.04.2018, 1.

⁷ Kohn (2020), S. 26.

⁸ EU-Kommission, Vorschlag KI-Verordnung, COM (2021) 206 final, Anhang I; dazu Engelhard/Schiemann (2022), S. 444.

⁹ Kohn (2020), S. 60.

von Handlungsanweisungen bzw. Algorithmen geschaffen, anhand derer das Computersystem aufgrund eines bestimmten Inputs ein bestimmtes Ergebnis bzw. einen bestimmten Output produziert. Dabei kann es sich zwar um sehr komplexe Regeln handeln, die aber in ihrer einmal von Menschen geschaffenen Form in immer gleicher und damit auch vergleichsweise nachvollziehbarer und transparenter Weise angewendet werden. Anders ist dies bei der zweiten besonders wichtigen Form Künstlicher Intelligenz, die von manchen auch als eigentliche Form Künstlicher Intelligenz (im engen Sinne) verstanden wird, namentlich beim maschinellen Lernen. Das besondere Merkmal des maschinellen Lernens ist die Datenabhängigkeit. In einer Variante des maschinellen Lernens¹⁰ wird dem System anhand der Eingabe sog. „Trainingsdaten“ nach und nach beigebracht, auf welche Art und Weise ein bestimmtes Ergebnis erzielt werden kann. Typischerweise erfolgt eine Leistungskontrolle und darauf aufbauend eine selbstständige Anpassung der Entscheidungsregeln und eine stetige Verbesserung der Ergebnisse. Ein triviales Beispiel ist das Erkennen von Katzen auf Tierfotos, das in der Weise vom System gelernt werden kann, dass man ihm eine ganze Reihe von Tierfotos zuführt und ihm dann zu jedem dieser Fotos das Feedback gibt, ob es sich bei dem dargestellten Tier um eine Katze handelt oder nicht. Wiederholt man diesen Vorgang mit einer ausreichend großen Zahl von Beispielen (was im Zeitalter von Big Data und riesigen vorhandenen Datenmengen zunehmend einfacher wird), dann wird das System im Laufe der Zeit erkennen, anhand welcher Kriterien und Parameter sich Katzen von anderen Tieren auf Abbildungen unterscheiden lassen. Das Besondere, aber vielleicht auch Problematische bei dieser Form der Künstlichen Intelligenz ist, dass sich die Anpassung und mögliche Verbesserung der Entscheidungsregeln im Laufe der Zeit verselbständigen kann und insofern der Entscheidungsvorgang ab einem gewissen Zeitpunkt möglicherweise nicht mehr transparent und nachvollziehbar ist, und zwar auch nicht für diejenigen Personen, die das System ursprünglich programmiert haben. Dass dieses Problem der Intransparenz, das oft auch mit dem Schlagwort der KI als „Black Box“ qualifiziert wird, auch rechtliche Implikationen hat, liegt auf der Hand. Darauf wird sogleich zurückgekommen.

10 Überblick bei *Kohn* (2020), S. 50 ff.

3. Potenzielle Anwendungsfelder im Strafverfahren

Der Einsatz komplexer digitaler Technologien, insbesondere im Wege der eben beschriebenen Formen von Künstlicher Intelligenz, ist auf vielen Ebenen innerhalb des Strafverfahrens denkbar.

Exemplarisch sei zunächst auf die Aufdeckung von Straftaten hingewiesen, zum Beispiel das Durchforsten von Datenmengen im Internet, um sog. Hasskriminalität erkennen zu können. Konkret diskutiert wird auch der Einsatz von Künstlicher Intelligenz im Bereich der Geldwäsche, wo verdächtige Vorgänge mit Hilfe von Algorithmen identifiziert und gemeldet werden könnten.¹¹ Auch im Bereich des Ermittlungsverfahrens ist ein vielfältiger Einsatz digitaler Technologien denkbar und zum Teil auch schon in der Planung oder auch Umsetzung; beispielsweise die Sichtung von Beweismaterial im Bereich Kinderpornografie.¹² Es liegt auf der Hand, dass es sowohl zur Bewältigung großer Datenmengen als auch zur Schonung der menschlichen Akteure sehr hilfreich wäre, wenn man die treffsichere Identifizierung von kinderpornografischem Material von einem Computersystem erbringen lassen könnte.

Auch im Bereich der Hauptverhandlung sind viele Einsatzfelder denkbar. Diskutiert wird unter anderem über eine Glaubwürdigkeitsbeurteilung bei Aussagen und Geständnissen. Eine einschlägige und auch auf das Jugendstrafverfahren bezogene Untersuchung dieses Aspekts von *Rodenbeck* und *Sommerfeld* kommt zu dem Ergebnis, dass „ein wirklich überzeugendes, ausgereiftes Tool“ bislang noch nicht in Sicht sei.¹³ Die weitere technologische Entwicklung könnte aber dazu führen, dass der Einsatz von solchen Systemen, die von manchen als „Lügendetektor 2.0“¹⁴ bezeichnet werden, technisch möglich wird, sodass man dann endgültig auch über die rechtliche Zulässigkeit neu nachdenken müsste. Sollte sich herausstellen, dass diese Systeme in der Tat besser als menschliche Beobachter Falschaussagen und falsche Geständnisse identifizieren können, spricht prima facie nichts gegen den zumindest unterstützenden Einsatz von solchen Technologien. Schließlich ist auch ein Einsatz von Künstlicher Intelligenz im Bereich der Sanktionsentscheidung denkbar, etwa bei den Prognosen, die in vielen Fällen Teil der Sanktionsentscheidung

11 Siehe den Beitrag von *Höffler/Reisch* in diesem Band.

12 Zu einem entsprechenden Modellprojekt in Nordrhein-Westfalen siehe <https://www.land.nrw/pressemitteilung/kuenstliche-intelligenz-im-kampf-gegen-kinderpornographie> [letzter Aufruf: 04.01.2023].

13 *Rodenbeck/Sommerfeld* (2021), S. 188.

14 Vgl. *Rodenbeck* (2020), S. 479.

sind, aber durchaus auch im Bereich der eigentlichen Strafzumessung, wie es im Bereich des allgemeinen Strafrechts schon zunehmend diskutiert wird.¹⁵

4. Potenzielle Anwendungsfelder im Jugendstrafverfahren

Auch im Bereich des Jugendstrafverfahrens gibt es eine Reihe von denkbaren weiteren (spezifischen) Anwendungsfeldern. So wäre es etwa möglich, Künstliche Intelligenz nicht nur im Bereich der Schuldfähigkeitsfeststellung gem. §§ 20, 21 StGB einzusetzen, sondern auch bei der Verantwortlichkeitsbeurteilung gem. § 3 JGG. Nach einer aktuellen einschlägigen Untersuchung dieser Frage durch *Paar* wäre zumindest die Unterstützung dieser richterlichen Entscheidung durch KI-Systeme rechtlich zulässig.¹⁶ In ähnlicher Weise wäre es auch denkbar, den Heranwachsendenstatus gem. § 105 Abs. 1 Nr. 1 JGG, wieder zumindest in unterstützender Form, von einem KI-System bestimmen zu lassen. Das liegt auch deswegen nicht fern, weil in der Praxis keine freie, individuelle Entscheidung getroffen wird, sondern der Heranwachsendenstatus anhand bestimmter Kriterienlisten und damit in gewisser Weise schematisch ermittelt wird.¹⁷ Insofern erscheint es nicht ausgeschlossen, die Entscheidung in einer algorithmisierten Form vorbereiten zu lassen. Auch im Rahmen der jugendstrafrechtlichen Sanktionsentscheidung wäre es denkbar, die Entscheidung von einem KI-System zumindest im Sinne eines Sanktionsvorschlags vorbereiten zu lassen. Das könnte dann auch die im Jugendstrafrecht deutlich flexiblere Kombination verschiedener Sanktionen erfassen (§ 8 JGG). Und schließlich ist darauf hinzuweisen, dass im Jugendstrafrecht Prognoseentscheidungen eine besondere Rolle spielen. Sie sind nicht nur bei der Aussetzung der Jugendstrafe zur Bewährung (§ 21 JGG) bzw. bei der Strafrestausschließung gem. § 88 JGG notwendig, sondern sind in vielen Fällen auch Teil der eigentlichen Sanktionsentscheidung, etwa im Bereich der Bestimmung von „schädlichen Neigungen“ gem. § 17 Satz 1 JGG. Ohnehin ist es in einem spezialpräventiv bzw. erzieherisch ausgerichteten Strafrecht wie dem Jugendstrafrecht auf den ersten Blick durchaus naheliegend, die empirische Fragestellung einer zukünftigen Delinquenzgefahr von einem System, das in der Lage ist, eine Vielzahl von Daten zu verarbeiten, mit vorbereiten

¹⁵ Siehe die Nachweise in Fußnote 2.

¹⁶ *Paar* (2021), S. 117.

¹⁷ Vgl. zu den Marburger Richtlinien und neueren Kriterienkatalogen *Eisenberg/Kölbel* (2022), § 105 Rn. 20 ff. (mit Kritik an einer zu starken Schematisierung).

oder sogar entscheiden zu lassen. Die besonderen Prognoseprobleme im Bereich des Jugendstrafrechts sind bekannt; unter anderem das Problem individueller dynamischer Verläufe und das Problem einer vergleichsweise schmalen Tatsachenbasis bei Tätern, die noch keine lange kriminelle Karriere hinter sich haben. Hinzu kommt, dass in der Praxis die intuitive Prognose mit all ihren Nachteilen überwiegt. Gerade hier könnten KI-gestützte Instrumente gewisse Vorteile bieten. Und dementsprechend wird auch in einer einschlägigen Untersuchung von *Miron und Kollegen* ausgeführt, dass es plausibel sei, Techniken des maschinellen Lernens gerade bei der Bestimmung der Rückfallgefahr von Jugendlichen einzusetzen.¹⁸

5. Denkbare Vorteile

Was könnten nun insgesamt die Vorteile eines Einsatzes digitaler Technologien, insbesondere in Form von Künstlicher Intelligenz, im Bereich des Jugendstrafverfahrens sein? Zunächst einmal wäre es möglich, dass der Einsatz von Algorithmen, die vermeintlich neutral und unbestechlich sind, Schwächen wie Vorurteile und Diskriminierung seitens der menschlichen Entscheider beseitigen könnte.¹⁹ Durch den Einsatz von Entscheidungsregeln und Handlungsanleitungen, wie man Algorithmen auch übersetzen und verstehen kann, könnte man zu einer höheren Bestimmtheit und Vorhersehbarkeit der staatlichen Sanktionierung im Allgemeinen wie auch im Bereich des Jugendstrafrechts gelangen. Sicher ist jedenfalls, dass der Einsatz von Künstlicher Intelligenz das Potenzial hat, eine Vielzahl von Fällen schneller und effizienter zu entscheiden, was dem Beschleunigungsgebot entsprechen würde, was gerade im Bereich des Jugendstrafrechts besonders betont wird. Und schließlich könnte speziell im eben angesprochenen Bereich der Prognose zukünftiger Delinquenz ein Einsatz von Systemen Künstlicher Intelligenz schlicht zu einer höheren Treffsicherheit führen. Diese Frage ist besonders umstritten und wird gerade in Bezug auf das Risk Assessmenttool COMPAS besonders kontrovers diskutiert.²⁰ Unabhängig von dieser konkreten Ausgestaltung eines Prognoseinstruments zeigen die meisten bisherigen Studien eine höhere Genauigkeit dieser Tools.²¹ Es scheint also so zu sein, dass ein datenbasiertes

18 *Miron et al.* (2021), S. 132.

19 Siehe etwa *Donohue* (2019); *Christoph* (2021), S. 61 f.

20 Siehe einerseits (kritisch) *Dressel/Farid* (2018) und andererseits demgegenüber positiver *Lin et al.* (2020); zusammenfassend *Kohn* (2020), S. 129 ff. sowie *Graichen* (2022), S. 129 ff.

21 Siehe etwa *Kleinberg et al.* (2018), S. 237; zusammenfassend *Butz et al.* (2021), S. 241.

generalisierendes bzw. statistisches Prognosesystem, wenn es entsprechend trainiert und mit vielen bisherigen Entscheidungen angereichert wird, einer menschlichen Prognoseentscheidung überlegen sein kann. Sollte sich dies in Zukunft bewahrheiten und der Einsatz digitaler Technologien hier zu einer echten Erhöhung der Trefferrate führen, müsste man sich umso mehr mit dem zunehmenden Einsatz von solchen Instrumenten auseinandersetzen, auch wenn dem, wie gleich diskutiert wird, eine ganze Reihe möglicher Probleme und Risiken entgegensteht.

6. Probleme und Risiken

Wenn es um den Einsatz von Künstlicher Intelligenz in neuralgischen Bereichen wie dem der Strafjustiz geht, gibt es eine ganze Reihe von allgemeinen Problempunkten, die diskutiert werden, und die hier aus Raumgründen nur sehr kurz genannt, aber nicht vertieft erörtert werden können. Es handelt sich dabei zum einen um das oben bereits angesprochene Problem der fehlenden Transparenz bzw. der Annahme der KI als sog. „black box“. Allerdings hat *Sommerer* zu Recht darauf hingewiesen, dass die Redeweise von der black box insofern irreführend ist, als sich die Unverständlichkeit und auch die Verselbständigung eines Systems nicht von selbst ergeben, sondern letztlich ja im Programm, das von Menschenhand geschaffen wurde, angelegt gewesen sein muss.²² Damit wird zu Recht zum Ausdruck gebracht, dass fehlende Transparenz nicht quasi zwingend in der Natur der Sache liegt, sondern mit einer entsprechenden Programmgestaltung möglicherweise auch verhindert werden kann. Genau in diese Richtung gehen Bemühungen um sog. „explainable AI“, also erklärbare Künstliche Intelligenz, die in letzter Zeit offensichtlich Fortschritte machen.²³

Ein zweiter genereller Kritikpunkt gegenüber der Anwendung von Künstlicher Intelligenz, gerade auch im Bereich der Justiz, ist der Vorwurf einer im System angelegten und möglicherweise schwer erkennbaren Diskriminierung.²⁴ Unter anderem gegenüber dem vorhin schon erwähnten System COMPAS wird der Vorwurf erhoben, dass es die Gefahr von Straftaten bei Angehörigen von ethnischen Minderheiten, insbesondere im Bereich der afroamerikanischen Bevölkerung, systematisch überschätze, während dies bei

²² *Sommerer* (2020), S. 195 ff.

²³ *Wang et al.* (2022); *Kohn* (2020), S. 277 ff.

²⁴ Zum Folgenden siehe *Kohn* (2020), S. 137 ff.

Personen mit weißer Hautfarbe genau umgekehrt sei.²⁵ Auch dieser Kritikpunkt ist nicht unumstritten, es gibt auch Stimmen in der Literatur, die davon ausgehen, dass die Algorithmen, die COMPAS zugrunde liegen, nicht diskriminierend sind.²⁶ Unabhängig von diesem konkreten Anwendungsbeispiel ist es sicher richtig, dass die Gefahr von Diskriminierung immer dann besteht, wenn schon die dem System zugrunde liegenden Daten, die zum Beispiel als Trainingsdaten zugeführt werden und aus dem es seine Entscheidungsregeln extrahiert, mit entsprechenden Diskriminierungen behaftet sind. Wenn Angehörige von ethnischen Minderheiten häufiger von der Polizei kontrolliert werden sollten und man deswegen naturgemäß häufiger bei ihnen bestimmte Straftaten aufdeckt, oder wenn polizeiliche Kontrollen in entsprechenden Wohngebieten häufiger sind und auch dies dazu führt, dass mehr Straftaten ins Hellfeld gelangen, so wäre eine erhöhte Straftatenbelastung bei diesem Personenkreis nicht notwendigerweise ein echter tatsächlicher Befund, sondern ein Artefakt, das von einer diskriminierenden Praxis erst hervorgerufen wird. Wenn nun die auf diese Weise zustande gekommenen Daten dem System zugeführt würden, wäre es durchaus denkbar, dass dies dann auch zu diskriminierenden Ergebnissen, etwa im Sinne der eben erwähnten Risikoüberschätzung beim genannten Personenkreis führt. Es ist also größter Wert darauf zu legen, dass die dem System zugeführten Trainingsdaten möglichst frei von solchen Diskriminierungen sind, was zugleich bedingt, dass der eben erwähnte Aspekt der Transparenz schon bei der Gestaltung des Systems möglichst weitgehend beachtet wird.

Ein weiterer allgemeiner Kritikpunkt ist die Annahme eines sog. „automation bias“, also des Phänomens, dass sich professionelle Entscheider von maschinell zustande gekommenen Entscheidungen bzw. Entscheidungsvorschlägen sehr stark in der Weise leiten lassen, dass sie sich im Zweifel diesem Vorschlag anschließen werden und es gar nicht zu einer echten und vertieften Überprüfung anhand des „gesunden Menschenverstandes“ kommt. Entsprechende empirische Erkenntnisse liegen etwa aus dem Bereich des Luftverkehrs vor, wo dem Autopiloten im Zweifel gefolgt wird, auch wenn die menschliche Erfahrung und Intuition dagegensprechen würde.²⁷ Eine offene Frage ist allerdings, ob ein solcher bias auch bei professionell entscheidenden Akteuren aus dem Bereich der Strafjustiz auftreten würde. Genauso denkbar wäre eine Zurückhaltung bis Ablehnung gegenüber solchen Systemen, die man als Übergriff in die eigene Entscheidungsfreiheit empfinden könnte, die

²⁵ Angwin et al. (2016).

²⁶ Rudin et al. (2020).

²⁷ Siehe dazu mit Nachweisen Sommerer (2020), S. 329 f.

ja gerade im Bereich der Gerichte besonders (auch verfassungsrechtlich) abgesichert ist. Der letzte allgemeine Kritikpunkt, der noch kurz erwähnt werden soll, geht in dieselbe Richtung, nämlich das Phänomen einer – bislang jedenfalls – fehlenden Akzeptanz des Einsatzes von Künstlicher Intelligenz im Bereich der Strafjustiz von Seiten der Bevölkerung.²⁸ Hier dominiert eine regelrechte „algorithm aversion“.²⁹ Das bedeutet, dass man etwa bei der Frage, ob Computertechnologie im Bereich der Vorhersage von Straftaten eingesetzt werden soll, nach verschiedenen Studien hierzulande sehr zurückhaltend bis ablehnend ist. So haben in einer Bertelsmann Studie aus dem Jahr 2018³⁰ nur 2 % der Bevölkerung die Frage bejaht, ob die Prognose von Straftaten allein durch ein Computersystem ausgeführt werden solle. Und selbst die Unterstützung durch ein Computersystem wird nur von 34 % der Befragten befürwortet, während sich 54 % der Befragten hier jeglichen ergänzenden Einsatz von Computertechnologie verbitten. Gerade bei neuralgischen grundrechtsintensiven und in ihrer Tragweite besonders wichtigen Entscheidungen bestehen also offenbar noch gewisse Vorbehalte gegenüber dem Einsatz dieser Technologien, wobei sicher in Rechnung zu stellen ist, dass diese Studie schon einige Jahre alt ist und sich die Bevölkerung vermutlich zunehmend an den Einsatz von Künstlicher Intelligenz in verschiedenen Lebensbereichen gewöhnen wird. Entscheidend für eine Steigerung von Akzeptanz dürfte sowohl der bereits thematisierte Aspekt der Transparenz als auch die Notwendigkeit und Garantie einer menschlichen Letztkontrolle gegenüber möglichen Entscheidungsvorschlägen eines KI-Systems sein.³¹

Ein bislang noch nicht ausreichend thematisiertes Problem ist die Frage, ob im Bereich der Strafjustiz, und insbesondere im Bereich der Sanktionierung und Prognose von Rückfallgefahr, überhaupt die Möglichkeit von maschinellem Lernen besteht und wenn ja, was denn der Bezugspunkt dieses Lernens sein soll. Wenn man sich zunächst einmal der Rückfallgefahr und deren Prognose als empirischer Frage zuwendet, dann erscheint hier ein Lerneffekt durchaus denkbar. Denn die Frage, ob jemand strafrechtlich wieder auffällig wird oder nicht, ist eine Frage, deren Beantwortung richtig oder falsch sein kann. Damit ist es in diesem Bereich grundsätzlich möglich, dem Computersystem anhand von konkreten Fällen eine Rückmeldung zu geben, ob seine Prognose richtig oder falsch war, auch wenn natürlich zahlreiche methodische Probleme zu berücksichtigen sind, nicht zuletzt die Frage, ob man bei einem

28 Siehe zu diesem Aspekt auch *Christoph* (2021), S. 64 ff.

29 Dazu näher *Kohn* (2020), S. 233.

30 *Fischer/Petersen* (2018), S. 15.

31 *Fischer/Petersen* (2018), S. 31.

angeblich nicht rückfällig Gewordenen tatsächlich davon ausgehen kann, dass er gar keine Straftaten mehr begangen hat, oder ob sie vielmehr nur im Dunkelfeld verblieben sind. Bei stärker normativ geprägten Entscheidungen wie beispielsweise der Bestimmung der Schwere der Schuld in § 17 JGG oder auch bei der Strafzumessung im Erwachsenenbereich, wo es um die tat- und schuldangemessene Strafe geht (§ 46 Abs. 1 Satz 1 StGB), ist dieser Trainings- und Lernmechanismus schon deutlich schwerer zu verwirklichen. Denkbar wäre zunächst der Rückgriff auf ein regelbasiertes System, wobei sich dann sofort die Folgefrage stellt, ob sich solche normativen Wertentscheidungen in quasi-mathematische Formeln bzw. Algorithmen transportieren lassen und ob auf diese Weise nicht ein unzulässiger Schematismus entsteht, bei dem die Einzelfallgerechtigkeit auf der Strecke bleibt. Denkbar wäre es, echtes maschinelles Lernen zum Einsatz zu bringen und dabei die bisherige Urteilspraxis heranzuziehen. Man könnte dem Computersystem also eine große Zahl von schon entschiedenen Fällen und deren wichtigste Kriterien und Merkmale zur Verfügung stellen, um auf diese Weise das System lernen zu lassen, wie die Richter in der Vergangenheit geurteilt haben. Ein Problem dabei ist natürlich eine möglicherweise drohende Zementierung der bisherigen Verhältnisse und die Übernahme von Fehlentwicklungen in der Vergangenheit. Eine solche Vorgehensweise empfiehlt sich also nur dann, wenn man mit der bisherigen Urteilspraxis im Großen und Ganzen einverstanden ist und man keine neuen Maßstäbe setzen will.

Daneben stellt sich die Frage einer notwendigen Individualisierung der Sanktions- sowie der Prognoseentscheidung. Hier besteht ein generelles Spannungsfeld zwischen einer generellen und einer individuellen Betrachtung.³² Für eine ausreichende Individualisierung der angesprochenen Entscheidungen spricht zum einen, insbesondere wenn es um die Sanktionsart und -höhe geht, das Schuldprinzip. So betont etwa der Bundesgerichtshof im Bereich des allgemeinen Strafrechts, dass die Strafe stets der individuellen Schuld des jeweiligen Täters angemessen sein müsse und sich eine rein vergleichende Strafzumessung mit Blick auf vergleichbare Konstellationen verbiete.³³ Im Bereich der Prognose wird vom Bundesverfassungsgericht betont, dass ein Anspruch auf eine ausreichend individuelle Prognose bestehe – eine bloß abstrakte, auf statistische Wahrscheinlichkeiten bezogene Prognose genüge jedenfalls nicht.³⁴ Diese Anforderungen werden im Jugendstrafrecht noch weiter

32 Siehe zum Beispiel *Ruppert* (2021), S. 90 ff.

33 Siehe etwa BGHSt 56 (2011), S. 262.

34 *BVerfGE* 109 (2004), S. 190 und 240 ff.

verstärkt durch den in § 2 I JGG verankerten Erziehungsgrundsatz.³⁵ Dieser enthält das Gebot, individuell auf die Besonderheiten des konkreten Falls einzugehen, wie sich anhand von vielen Regelungen des JGGs zeigen lässt. Zu nennen ist hier u. a. § 38 JGG, der den Einsatz der Jugendgerichtshilfe regelt, die gerade deswegen in das Verfahren eingebunden wird, damit die Person des Jugendlichen in besonders vertiefter Weise beleuchtet und in das Verfahren eingeführt werden kann. Eine weitere im Jugendstrafrecht enthaltene Wertung, die in diese Richtung zielt, ist die recht weitreichende und zentrale Position der Jugendrichterinnen und Jugendrichter, die nicht nur im Rahmen des Hauptverfahrens zuständig sind, sondern auch anschließend die Vollstreckung leiten, weil man davon ausgeht, dass die persönliche Kenntnis des individuellen Jugendlichen hier zu sachgerechten und flexiblen Entscheidungen führt. Auch dies spricht klar dagegen, die wesentlichen Entscheidungen im Bereich des Jugendstrafrechts vollständig oder überwiegend in die Hand von Computersystemen zu legen, die naturgemäß von der Person des jugendlichen Straftäters keinen persönlichen Eindruck gewinnen können.

Ein weiteres Grundlagenproblem ist die Frage, ob man nicht ohnehin ein Recht auf einen „menschlichen Entscheider“ hat. *Benedikt Kohn* hat in seiner gründlichen Arbeit über den Einsatz Künstlicher Intelligenz und Strafzumessung aus Art. 22 der DGSVO abgeleitet, dass Künstliche Intelligenz auch in rein unterstützender Form im Jugendstrafrecht nicht eingesetzt werden dürfe.³⁶ Das ist eine vertretbare Position, die man auf die genannte Norm aber insofern nur schwer stützen kann, weil sie im Rahmen der Strafverfolgung nicht unmittelbar anwendbar ist. Ein zwingendes Argument gegen die Anwendung von KI im Jugendstrafrecht lässt sich daraus also nicht ableiten. Vielmehr erlaubt § 54 BDSG unter gewissen Voraussetzungen automatisierte Einzelfallentscheidungen, wenn eine entsprechende gesetzliche Grundlage besteht. Diese Wertung spricht also eher dafür, dass es hier keine sektoralen Komplettausnahmen gibt, sondern dass es dem Gesetzgeber freistünde (unter Beachtung der Vorgaben der kommenden EU-KI-Verordnung³⁷) entsprechende Regelungen zu schaffen. Damit ist aber noch nicht gesagt, dass auch im Bereich der Justiz automatisierte Einzelfallentscheidungen im Sinne einer verbindlichen Letztentscheidung durch das KI-System möglich sind. Richtigerweise bestünden hier viele verfassungsrechtliche Hürden und Probleme, etwa mit Blick auf Art. 92 Abs. 1 GG und Art. 101 GG.³⁸ Ergänzend wird das

³⁵ Ähnlich *Kohn* (2020), S. 325.

³⁶ *Kohn* (2020), S. 323 f.

³⁷ *Engelhard/Schiemann* (2022), S. 444.

³⁸ *Kohn* (2020), S. 180 ff.

Recht auf einen menschlichen Entscheider teilweise aus Art. 1 Abs. 1 GG abgeleitet³⁹ oder auf ein allgemeines rechtstheoretisches Fundament gestützt, das man mit der Frage einer zwingend erforderlichen (menschlichen) „Verantwortung“ für eine Sanktionsentscheidung umschreiben könnte.⁴⁰ Diese Erwägungen werden im Jugendstrafrecht noch weiter untermauert durch den Erziehungsgrundsatz und die oben bereits erwähnte zentrale Stellung der Jugendrichterinnen und Jugendrichter, die jedenfalls eine Ersetzung des Jugendgerichts und seiner Entscheidungen durch ein KI-System als unzulässig erscheinen lassen. Dies spricht aber nicht prinzipiell gegen die Unterstützung der jugendrichterlichen Entscheidung durch KI-Systeme, wie es dem Jugendgericht ja auch sonst unbenommen ist, zu seiner Entscheidung ergänzende Informationen, etwa durch Sachverständige, einzuholen. Entscheidend muss letztlich sein, ob die Einbeziehung dieser Informationsquelle zu einer Verbesserung der Qualität rechtlicher Entscheidungen führt. Und das bleibt vorerst (auch mangels ausreichender Erfahrungen mit der Anwendung entsprechender Systeme) eine offene Frage.

7. Fazit

Als Ertrag der vorstehenden Ausführungen kann zunächst festgehalten werden, dass der Einsatz von KI im Bereich der Strafverfolgung einen sensiblen und grundrechtsrelevanten Bereich betrifft, was sowohl für das allgemeine Strafrecht wie auch (erst recht) für das Jugendstrafrecht gilt. Dementsprechend wird der Einsatz von KI in der oben erwähnten neuen EU-KI-Verordnung in Art. 6 Abs. 2 auch als Hochrisikobereich eingestuft. Darin ist aber zugleich die Wertung enthalten, dass der Einsatz von KI im Bereich der Strafjustiz prinzipiell möglich ist und nicht, wie es in anderen Bereichen der Fall ist, bereits im Verordnungswege komplett ausgeschlossen werden soll. Allerdings darf der Einsatz solcher Technologien auch nach den Wertungen der KI-Verordnung nur unter sorgfältiger Abwägung von Nutzen und Risiken und nur auf gesetzlicher Grundlage erfolgen. Ein vollständiger Ausschluss von KI aus dem allgemeinen Strafverfahren würde dem Potential von entsprechenden Technologien nicht gerecht und wäre auch rechtlich nicht zwingend geboten. Allerdings sprechen die Wertungen des Jugendstrafrechts, insbesondere der Erziehungsgrundsatz dafür, dass KI gerade hier nur sehr maßvoll und mit Vorsicht eingesetzt werden sollte, etwa in Form von KI-basierten (unter-

³⁹ Kernchen (2021), S. 108 und 111.

⁴⁰ Greco (2020), S. 29 und 48 ff.

stützenden) Prognoseinstrumenten. Mit Blick auf Grundrechtsschutz und Vertrauensbildung in der Bevölkerung ist die Forderung nach Transparenz und menschlicher Kontrolle entsprechender Systeme besonders wichtig, die auch in der EU-Verordnung hervorgehoben wird (s. Art. 13 und 14 EU-Verordnung). Als Zukunftsperspektive ist zunächst darauf hinzuweisen, dass trotz zunehmender Forschungsaktivität noch deutlich mehr interdisziplinäre Forschung bezüglich Künstlicher Intelligenz unter Berücksichtigung von technologischen, rechtlichen und ethischen Fragen erfolgen muss. Auch in der juristischen Aus- und Weiterbildung sollte noch deutlich stärker auf die Digitalisierung und die verschiedenen Formen von Künstlicher Intelligenz Bezug genommen werden, wie es sich in jüngerer Zeit auch zunehmend abzeichnet. Und schließlich wird eine noch deutlich intensivere gesamtgesellschaftliche Debatte über Nutzen und Risiken von KI geführt werden müssen, die dann letztlich auch die Weiterentwicklung im Bereich der Kriminalpolitik wesentlich beeinflussen wird.

Literatur

- Angwin, J. L./Larson, J./Mattu, S./Kirchner, L.* (2016): Machine Bias – There’s software used across the country to predict future Criminals. And it’s biased against blacks. ProPublica, in: <https://www.propublica.org/article/machine-bias-risk-assessments-in-criminal-sentencing>.
- Butz, F./Christoph, S./Sommerer, L./Harrendorf, S./Kaspar, J./Höffler, K.* (2021): Automatisierte Risikoprognosen im Kontext von Bewährungsentscheidungen. *BewHi*, 68 (3), S. 241-259.
- Christoph, S.* (2021): Vielversprechender Effektivitätsbooster oder problematischer Subsumtionsautomat? Gedanken zum Einsatz künstlicher Intelligenz „auf der Richterbank“ in Strafverfahren. In: Giannini, A. et al. (Hg.): *Strafrecht im Angesicht der Digitalisierung*, S. 59-78.
- Donohue, M. E.* (2019): A replacement for Justitia’s scales?: Machine Learning’s role in sentencing. *Harvard Journal of Law & Technology*, 32 (2), S. 657-678.
- Dressel, J./Farid, H.* (2018): The accuracy, fairness and limits of predicting recidivism. *Science Advances*, 4 (1).
- Eisbach, S./Heghmanns, M./Hertel, G.* (2022): Künstliche Intelligenz im Strafverfahren am Beispiel von Kriminalprognosen. *ZfJStW*, 1 (7-8), S. 489-496.
- Eisenberg, U./Kölbl, R.* (2022): *Jugendgerichtsgesetz* (23. Auflage). München: C.H. Beck.
- Engelhard, J./Schiemann, A.* (2022): Der EU-Kommissionsvorschlag einer Verordnung für künstliche Intelligenz aus rechts- und kriminalpolitischer Perspektive – Zur unumgänglichen aber nachbesserungsbedürftigen Harmonisierung. *KriPoZ*, 7 (6), S. 444-456.
- Fischer, S./Petersen, T.* (2018): *Was Deutschland über Algorithmen weiß und denkt*. Gütersloh: Bertelsmann-Stiftung.

- Graichen, N.* (2022), Die Automatisierung der Justiz. Baden-Baden: Nomos-Verlag.
- Greco, L.* (2020): Richterliche Macht ohne richterliche Verantwortung: Warum es den Roboter-Richter nicht geben darf. *RW*, 11 (1), S. 29-62.
- Herberger, M.* (2018): „Künstliche Intelligenz“ und Recht. Ein Orientierungsversuch. *NJW*, 71 (39), S. 2825-2829.
- Höffler, K./Reisch, K.* (in diesem Band): Die Auswirkungen von KI auf die Geldwäscheprävention.
- Kaspar, J./Höffler, K./Harrendorf, S.* (2020): Datenbanken, Online-Votings und künstliche Intelligenz. *NK*, 32 (1), S. 35-54.
- Kernchen, T.* (2021): Möglichkeiten und Grenzen des Einsatzes intelligenter Systeme für jugendrichterliche Entscheidungen. *ZJJ*, (2), S. 108-117.
- Kleinberg, J./Lakkaraju, H./Leskovec, J./Ludwig, J./Mullainathan, S.* (2018): Human Decisions and Machine Predictions. *The Quarterly Journal of Economics*, 133 (1), S. 237-293.
- Kment, M./Borchert, S.* (2022): Künstliche Intelligenz und Algorithmen in der Rechtsanwendung. München: C.H.Beck Verlag.
- Kohn, B.* (2020): Künstliche Intelligenz und Strafzumessung. Baden-Baden: Nomos-Verlag.
- Lin, Z. J./Jung, J./Goel, S./Skeem, J.* (2020): The limits of human predictions of recidivism. *Science Advances*, 6 (7).
- Miron, M./Tolan, S./Gómez, E./Castillo, C.* (2021): Evaluating causes of algorithmic bias in juvenile criminal recidivism. *Artificial Intelligence and Law*, 29 (2), S. 111-147.
- Nink, D.* (2021): Justiz und Algorithmen. Berlin: Duncker & Humblot.
- Paar, E.* (2021): Künstliche Intelligenz im Rahmen der Beurteilung der strafrechtlichen Schuldfähigkeit Jugendlicher – Grundrechtliche Überlegungen aus deutscher und österreichischer Perspektive. *ZJJ*, (2), S. 117-125.
- Peters, A.* (2022): Kapitel 12 (Strafrecht). In: Kaulartz, M./Braegelmann, T. (Hg.): *Rechtshandbuch Artificial Intelligence und Machine Learning*. München: C.H.Beck/Vahlen, S. 559-579.
- Rodenbeck, J./Sommerfeld, M.* (2021): Technische Lösungen zur Erkennung falscher Geständnisse in Jugendstrafverfahren – ein Blick auf die Aufzeichnungen von Vernehmungen in Bild und Ton sowie auf die Möglichkeiten des Einsatzes Künstlicher Intelligenz. *ZJJ*, (3), S. 188-193.
- Rostalski, F./Völkening, M.* (2019): Smart Sentencing. Ein neuer Ansatz für Transparenz richterlicher Strafzumessungsentscheidungen. *KriPoZ*, 4 (5), S. 265-273.
- Rudin, C./Wang, C./Coker, B.* (2020): The Age of Secrecy and Unfairness in Recidivism Prediction. *Harvard Data Science Review*, in: <https://hdsr.mitpress.mit.edu/pub/7z10o269/release/7>.
- Ruppert, F.* (2021): Strafzumessung am Scheideweg? Legal Tech und Strafzumessung. *KriPoZ*, 6 (2), S. 90-98.
- Sommerer, L.* (2020): Personenbezogenes Predictive Policing. Baden-Baden: Nomos-Verlag.
- Wang, C./Han, B./Patel, B./Rudin, C.* (2022): In Pursuit of Interpretable, Fair and Accurate Machine Learning for Criminal Recidivism Prediction. *Journal of Quantitative Criminology*. <https://link.springer.com/article/10.1007/s10940-022-09545-w#citeas>.