

# BEWÄHRUNGSHILFE

## Soziales • Strafrecht • Kriminalpolitik

### Inhaltsverzeichnis des Jahrgangs 68 – 2021

#### SCHWERPUNKTTHEMEN

Heft 1	Aktuelle Betreuungskonzepte für Sexualstraftäter	5
Heft 2	Selbstevaluation / Praxisreflexion	117
Heft 3	Digitalisierung	213
Heft 4	Bewährungs- und Gerichtshilfe Baden-Württemberg (BGBW)	309

#### SCHWERPUNKTREFERATE UND EINZELBEITRÄGE

*Katja Adam*

„Vermeiden von Ersatzfreiheitsstrafen“  
Ergebnisse eines Pilotprojekts der BGBW 374

*Julia Audick*

Soziale Diagnostik bei der Bewährungs- und Gerichtshilfe Baden-Württemberg (BGBW) 345

*Louisa Bauer • Sonja Schröder • Safiye Tozdan • Jürgen L. Müller • Peter Fromberger*  
@myTabu – Konzept einer therapeuten-gestützten Online-Intervention für verurteilte  
Personen, die Kindesmissbrauch begangen oder Missbrauchsabbildungen konsumiert  
haben 5

*Susanne Beier • Benjamin Clausen*

Chancen und Herausforderungen der ambulanten deliktpräventiven Therapie aus Sicht  
der Hessischen Fachambulanz 23

*Jennifer Burghardt • Robert Lehmann*

Digitalisierung und Bewährungshilfe – Ein Überblick zum wissenschaftlichen Stand  
und zu Perspektiven der Digitalisierung von Beratungsprozessen 213

*Felix Butz • Stephan Christoph • Lucia Sommerer • Stefan Harrendorf •  
Johannes Kaspar • Katrin Höffler*

Automatisierte Risikoprognosen im Kontext von Bewährungsentscheidungen 241

*René Cuadra*

Wie wir inhaftierte Menschen für eine Kooperation gewinnen können  
Und wie wir aktuell die Resozialisierung verfehlen 97

*Liane Czeremin*

Online gut beraten:  
Erfahrungen und Methoden aus der Online-Beratung gegen Rechtsextremismus 290

## Automatisierte Risikoprognosen im Kontext von Bewährungsentscheidungen

FELIX BUTZ • STEPHAN CHRISTOPH • LUCIA SOMMERER • STEFAN HARRENDORF •  
JOHANNES KASPAR • KATRIN HÖFFLER<sup>1</sup>

*Die immer rasanter werdenden Entwicklungen im Bereich von Legal Tech, insbesondere auch bezogen auf Künstliche Intelligenz (im Folgenden: KI), beschäftigen zunehmend Justiz und Rechtswissenschaft. Gerade Bewährungsentscheidungen scheinen aufgrund der dort anzustellenden, letztlich auf Wahrscheinlichkeitsaussagen beruhenden, Legalprognose für den Einsatz algorithmenbasierter Systeme prädestiniert. Der vorliegende Beitrag beleuchtet Möglichkeiten und Grenzen der – ggf. KI-gestützten – Automatisierung auf diesem Gebiet, wobei neben den rechtlichen Rahmenbedingungen für die Prognoseerstellung und der bestehenden Prognosepraxis vor allem verfassungsrechtliche, psychologische und gesellschaftliche Hürden einer „Algorithmisierung“ der Bewährungsentscheidung eingehend dargestellt werden.*

### I. Einleitung

Die rasante technologische Entwicklung im Bereich der KI und ihrer Vorstufen holt auch die Justiz ein; immer häufiger wird über Chancen, aber auch Risiken des Einsatzes von „Legal Tech“ im juristischen Bereich nachgedacht, wobei das Spektrum denkbarer Anwendungsfelder sehr breit ist (Kaspar/Höffler/Harrendorf 2020). Es reicht von der bereits jetzt vielfach praktizierten schlichten Nutzung von Datenbanken bis hin zur Unterstützung oder sogar Ersetzung juristischer Tätigkeiten durch automatisierte Werkzeuge, mit oder ohne Einsatz von KI (zu letzterem z. B. Aumann 2018).

Dabei wird hier unter einem „automatisierten“ Verfahren ein solches verstanden, welches unmittelbar durch nicht psychowissenschaftlich ausgebildete Entscheider\*innen, z. B. Richter\*innen, computer-gestützt angewendet wird und einen Rückfall-Risikowert für die zu beurteilende Person algorithmenbasiert errechnet. KI-gestützte Verfahren sind dabei solche, die durch Algorithmen menschliches Entscheidungsverhalten imitieren bzw. intelligentes automatisiertes Verhalten ermöglichen (vgl. Russell/Norvig 2021, S. 1 ff.). Oft wird dabei – gerade auch bezüglich der Risikoprognose – auf sog. „machine learning“ gesetzt, d. h. darauf, statistische Modelle durch den Computer automatisiert aus Trainingsdaten heraus abzuleiten, die dann wiederum auf neue Fälle extrapoliert werden (Mehta et al. 2020, S. 5110f.; Wang et al. 2020, S. 2).

<sup>1</sup> Alle Verfasser\*innen haben in gleichem Umfang zu dem Artikel beigetragen. Die Reihenfolge der Nennung wurde mit einem Zufallsgenerator ermittelt.

Auch die Strafjustiz ist von derartigen Entwicklungen betroffen. Angesichts des Problems einer intransparenten und uneinheitlichen Strafzumessungspraxis in Deutschland (siehe nur Kaspar 2018) wird z. B. erörtert, ob der systematische Einsatz von Datenbanken oder sonstiger Formen von „Legal Tech“ bis hin zur automatisierten Strafzumessungsentscheidung durch KI insoweit Verbesserungen mit sich bringen könnte (Nink 2021; Rostalski/Völkening 2020; Ruppert 2021; Kaspar/Höffler/Harrendorf 2020).

KI-Systeme, aber auch automatisierte statistisch-aktuarische Tools, sind dabei vor allem aufgrund ihrer – teilweise empirisch gesicherten (vgl. etwa Kleinberg et al. 2018) – prognostischen Fähigkeiten (Crawford 2021) für das Strafjustizsystem anschlussfähig. Denkbar wäre bspw. ein direkter Einsatz automatisierter Prognoseinstrumente in der Praxis der Bewährungshilfe. Die von den Bewährungshelfer\*innen zu leistenden Aufgaben umfassen unter anderem die Gefahreinschätzung und die Gefahrbearbeitung (zu der Tätigkeit der Bewährungshilfe s. etwa Dölling 2015, S. 76ff.). Erstere könnte aufgrund ihres prognostischen Charakters möglicherweise auch von einem Algorithmus übernommen werden, wie bspw. ein Blick nach Philadelphia (USA) zeigt. Um eine Überlastung der dortigen Bewährungshelfer\*innen abzuwenden, wurde ein Algorithmus entwickelt, der die Rückfallwahrscheinlichkeiten berechnet und an jede freigelassene Person einen Risiko-Score (niedrig, mittel, hoch) vergibt. Dieser Score bestimmt dann die Art und Intensität des individuellen Bewährungsprogramms. Auch die Aufteilung an die Bewährungshelfer\*innen folgt den Risiko-Score-Kategorien – diejenigen, die Perso-

nen mit niedrigem Risiko-Score betreuen, bekommen bis zu 400 Freigelassene zugewiesen; Bewährungshelfer\*innen der hohen Risiko-Score-Kategorie betreuen hingegen „nur“ bis zu 50 Freigelassene (Labi 2012).

Für den vorliegenden Beitrag soll das besondere Augenmerk auf ein weiteres mögliches Anwendungsfeld einer Automatisierung gelegt werden. Erörtert werden soll, inwiefern ein entsprechendes Prognose-Tool bei der gerichtlichen Entscheidung über die Aussetzung von Freiheitsstrafen zur Bewährung Anwendung finden könnte. Denn hier steht mit der von Wahrscheinlichkeitsaussagen abhängigen Legalprognose eine für den Einsatz von Algorithmen, insbesondere auch KI, prädestinierte Fragestellung im Zentrum. Es erscheint zumindest denkbar, dass ein Computer, den man mit einer Menge an entsprechenden Daten „füttert“, aufgrund seiner Rechenleistung zu einer schnelleren und präziseren Beantwortung dieser Frage in der Lage ist als ein Mensch. Sollte man ihm dann diese Entscheidung nicht besser gleich ganz überlassen – oder ihn zumindest eine (unverbindliche) Einschätzung des Rückfallrisikos erstellen lassen, die dann von menschlichen Entscheider\*innen in Betracht gezogen werden kann oder muss? Im vorliegenden Beitrag soll der Frage nachgegangen werden, ob sich der Einsatz KI-gestützter oder sonstiger automatisierter Prognosetools auch in Deutschland aktuell oder künftig empfehlen könnte oder ob dem technische, methodische oder normative Bedenken entgegenstehen.

## II. Rechtliche Grundlagen

### 1. Kriminalprognosen im geltenden Recht

Das deutsche Strafrecht erfordert an vielen Stellen eine Kriminalprognose. Im hier näher behandelten Bereich der Aussetzung einer Freiheitsstrafe zur Bewährung bedarf es Prognosen vor allem gem. § 56 StGB, wenn schon die Verhängung der Freiheitsstrafe zur Bewährung ausgesetzt werden soll, und gem. §§ 57, 57a StGB, wenn es um die Aussetzung des Restes einer zeitigen oder einer lebenslangen Freiheitsstrafe zur Bewährung nach teilweiser Verbüßung der Straftat geht.

Das Gesetz differenziert die Anforderungen an eine Aussetzung zur Bewährung bei § 56 StGB bekanntlich in drei Stufen: Freiheitsstrafe von unter 6 Monaten, von 6 bis 12 Monaten und von über 12 Monaten bis 2 Jahren. In der ersten Stufe ist nur eine günstige Legalprognose erforderlich, in der zweiten Stufe muss noch zusätzlich geprüft werden, ob nicht die Vollstreckung der Freiheitsstrafe zur „Verteidigung der Rechtsordnung“ geboten ist (§ 56 III StGB). Auf Stufe drei muss neben diesen beiden Voraussetzungen geprüft werden, ob „nach der Gesamtwürdigung von Tat und Persönlichkeit des Verurteilten besondere Umstände vorliegen“, die eine Aussetzung rechtfertigen (§ 56 II StGB). Allen drei Stufen gemein ist, dass sie eine günstige Legalprognose verlangen, also die Erwartung, „daß der Verurteilte sich schon die Verurteilung zur Warnung dienen lassen und künftig auch ohne die Einwirkung des Strafvollzugs keine Straftaten mehr begehen wird“ (§ 56 I StGB). Das Gesetz zählt (nicht abschließend) einige Aspekte auf, die bei der Erstellung der Prognose zu

berücksichtigen sind: Persönlichkeit, Vorleben, Umstände der Tat, Verhalten nach der Tat und Lebensverhältnisse. Ausdrücklich erwähnt werden die „Wirkungen (...), die von der Aussetzung für ihn zu erwarten sind“, so dass nicht nur der Status Quo, sondern auch der (zukünftige) Verlauf der Bewährungszeit sowie die Auswirkungen von Bewährungshilfe und denkbaren Weisungen und Auflagen abzuschätzen ist.

Auch im Bereich der §§ 57, 57a StGB stellt das Gesetz eine günstige Legalprognose ins Zentrum: Die Aussetzung wird gewährt, wenn dies „unter Berücksichtigung des Sicherheitsinteresses der Allgemeinheit verantwortet werden kann“ (§ 57 I Nr. 2 StGB; § 57a I Nr. 3 StGB). Auch hier nennt das Gesetz (erneut nicht abschließend) Aspekte, die in die Aussetzungsentscheidung einfließen sollen (vgl. § 57 I 2 StGB).

Für die vorliegende Fragestellung ist daher festzuhalten, dass die Aussetzungsentscheidungen ganz überwiegend nicht allein auf die im Kern empirische Fragestellung der Legalprognose gestützt werden, sondern weitere Voraussetzungen haben, die eine normative Entscheidung erforderlich machen. Besonders deutlich wird dies bei §§ 57, 57a StGB, die mit der „Verantwortungsklausel“ eine Abwägung verlangen, bei der das ermittelte Risiko zukünftiger Straftaten nur ein Teilelement darstellt (Nink 2021, S. 421). Aber auch im Bereich der Freiheitsstrafen ab 6 Monaten kommen weitere Voraussetzungen hinzu. Schon aus diesem Grund würde der Einsatz von KI im Bereich der Prognoseerstellung die richterliche „Gesamtentscheidung“ nicht vollständig vorherbestimmen oder gar ersetzen können (was nicht aus-

schließt, dass mittelfristig auch bei den anderen Parametern der Aussetzungsentcheidung eine KI-Unterstützung diskutiert oder entwickelt werden könnte). Lediglich bei den Freiheitsstrafen unter 6 Monaten ist nach dem Gesetz quasi ein Automatismus vorgesehen, bei dem eine Aussetzung zu gewähren „ist“, wenn eine günstige Legalprognose erstellt wird. Vor allem hier käme eine Determination der Entscheidung des Gerichts durch eine automatisierte Prognose in Betracht. Das wird derzeit nicht ernsthaft gefordert, zumal das Erfordernis einer Interventionsprognose zu weiteren, zumindest aktuell nicht durch KI lösbaren Anforderungen führt (dazu näher unten). Je nachdem, wie die weitere technische Entwicklung verlaufen wird, könnten aber zukünftig Forderungen nach einem umfassenderen Einsatz von KI oder automatisierter statistisch-aktuarischer Tools im Bereich der Aussetzungsentscheidungen laut werden. So oder so stellt sich die Frage, ob und wenn ja, in welcher Form das Gesetz den Einsatz automatisierter Prognosen überhaupt erlaubt.

## 2. „Methodenoffenheit“ der gesetzlichen Regelungen

Das Gesetz enthält nur wenige explizite Anforderungen im Hinblick auf die bei den Aussetzungsentscheidungen zu stellenden Prognosen (s. o.). Letztlich wird eine umfassende Würdigung aller Gesichtspunkte gefordert, die begründete Rückschlüsse auf ein künftiges Legalverhalten der Betroffenen erlauben (Groß/Kett-Straub 2020, § 56 Rn. 16). Auch ist nicht ausdrücklich geregelt, wie genau das Gericht zu seiner Einschätzung der Legalbewährung gelangen soll (Kinzig 2019, § 56

Rn. 19); weder ist eine bestimmte Prognosemethode noch die Einholung eines Sachverständigengutachtens – bis auf spezielle Fälle (s. sogleich) – gesetzlich vorgeschrieben.

Die Methodik der Prognose delinquenten Verhaltens lässt sich grob in drei Stränge unterteilen (vgl. zum Folgenden Rettenberger et al. 2017, S. 3f.): die intuitive, die statistisch-aktuarische sowie die klinische bzw. klinisch-strukturierte Prognose. Die intuitive Prognose, d. h. eine allein aufgrund der eigenen Erfahrung ohne externe Hilfsmittel durchgeführte Prognose, ist sehr stark durch die subjektive Vorerfahrung der prognostizierenden Person geprägt. Zentrale Kritik richtet sich gegen die eingeschränkte Transparenz und die dadurch bedingte eingeschränkte Nachvollziehbarkeit des Vorgehens, weshalb ein solches nicht den Anforderungen genüge, die an Prognosegutachten gestellt werden (Boetticher et al. 2019). Auch die nur knapp den Zufall schlagende Prognoseleistung des intuitiven Vorgehens ist problematisch (Rettenberger et al. 2017, S. 3 m. w. N.).

Statistisch-aktuarische Vorhersagen hingegen sind streng standardisierte Verfahren, die sich an einer bestimmten Anzahl an Risikofaktoren orientieren, die einen empirisch geprüften Zusammenhang mit Rückfälligkeit aufweisen müssen. Die Risikofaktoren sind operationalisiert und werden anhand eines Manuals bewertet; anschließend wird – auf unterschiedliche Art und Weise – aus den Einzelwerten ein Gesamtwert gebildet, der Aufschluss über absolute (bspw. in Form von Normwerten oder Rückfallraten) und relative (in Form von Prozenträngen bzw. Perzentilen) Risikopotentiale einer Person gibt.

Etwas flexibler sind klinische bzw. klinisch-strukturierte Prognoseinstrumente, mit denen die Besonderheiten des Einzelfalls sowie therapeutische und kriminalpräventive Aspekte stärker berücksichtigt werden können. Verfahren dieser Art kennen zwar ähnlich wie statistische Verfahren ebenfalls Risiko- und Schutzfaktoren, eröffnen aber der prognostizierenden Person einen (größeren) Gestaltungs- und Interpretationsspielraum im Hinblick auf das Prognoseergebnis. Zudem ist die Art der Risikokommunikation zentral, die hier nicht numerisch, sondern in Form sog. nominalen Risikokategorien erfolgt („niedrig“, „moderat“, „hoch“). Dieses Prognoseergebnis wird durch eine qualitative und idiographische Bewertung der Risikofaktoren im konkreten Einzelfall erzielt.

Trotz aller Kritik überwiegt in der Praxis die vom Gericht erstellte intuitive Prognose (Kinzig 2019, § 56 Rn. 17; Grube 2010, S. 761). Nur ausnahmsweise (vgl. §§ 454 II, 463 III StPO) wird die Einholung eines Gutachtens zwingend vorgeschrieben. Allerdings bleiben auch hier nicht unerhebliche Spielräume, da den Sachverständigen die Wahl der konkreten Methode und Vorgehensweise nicht vorgeschrieben ist. Die Probleme der intuitiven Methode werden durch die fachliche Begutachtung daher zwar abgemildert, aber nicht behoben.

Es fragt sich nun, ob sich vor diesem Hintergrund die (unterstützende) Nutzung automatisierter statistisch-aktuarischer oder KI-unterstützter Prognoseverfahren durch die Gerichte anböte, da sich so die Vorteile von intuitiver und klinischer Prognose u. U. verbinden ließen: Einerseits kann die Methode möglicherweise schnell und kostengünstig vom Gericht selbst an-

gewendet werden, ohne dass zwingend Sachverständige eingeschaltet werden müssen, andererseits wird die Prognose nicht aufgrund rein subjektiver Wertungen, sondern auf einer empirischen Basis erstellt (Nink 2021, S. 418). Auch zur Verminderung einer regional ungleichen Anwendung der Normen könnte dies führen, soweit damit eine gewisse Vereinheitlichung und Objektivierung der Entscheidungsmaßstäbe erzielt werden könnte (Ebersbach 2020, S. 32). Ob dies allerdings im Bereich der Kriminalität als komplexem, multifaktoriell bedingten Geschehen empirisch-technisch machbar und auch normativ zulässig ist, soll im Folgenden untersucht werden. Hier kann jedenfalls zunächst festgehalten werden, dass die §§ 56, 57 StGB bzgl. der (erfahrungswissenschaftlichen) Beurteilung des Rückfalls „methodenoffen“ ausgestaltet sind und keine ausdrücklichen Hürden gegenüber der Anwendung von statistisch-aktuarischen oder KI-gestützten Verfahren enthalten.

### III. Prognoseforschung

#### 1. Empirische Erkenntnisse zur Treffsicherheit klinisch-strukturierter und statistisch-aktuarischer Instrumente

Heutzutage sind insb. im angloamerikanischen Rechtskreis Risikoprognose-Tools ein weit verbreitetes Instrument im strafjustiziellen System und waren das bereits vor Aufkommen der neueren informationstechnologischen Entwicklungen (Završnik 2018, S. 144). In den USA wurden bereits 2015 um die sechzig verschiedene Instrumente eingesetzt (Barry-Jester et al. 2015). In Kontinentaleuropa ist man mit

deren Anwendung zurückhaltender (Završnik 2018, S. 144). Dennoch werden sie auch hierzulande in klinischer sowie aktuuarischer Ausprägung durchaus breit genutzt (Rettenberger et al. 2017).

Bis heute entzündet sich Streit um die treffsicherste und zugleich in rechtlicher und ethischer Hinsicht angemessenste Prognosemethode (analysierend Schöch 2015, S. 139; vgl. auch BGH StV 2008, 200; NStZ 2009, 499). In der praktischen Anwendung werden statistisch-aktuarische und klinisch-idiographische Prognose von Praktiker\*innen als gleichwertig eingestuft (Rettenberger et al. 2017, S. 11). Geht es allein um die Treffsicherheit der Prognose, ist die Überlegenheit statistisch-aktuarischer Instrumente hingegen in zahlreichen Studien belegt worden (Rettenberger et al. 2017, S. 3 m. w. N.; Ægisdóttir/White/Spengler 2006; Hanson/Morton-Bourgon 2009).

Um den mangelnden Einzelfallbezug abzumildern, wurde vorgeschlagen, gruppenstatistisch gewonnene Erkenntnisse mit klinisch-strukturierten Prognosen zu kombinieren und in einem Prognosemodell zu vereinen (Dahle/Lehmann 2018; Schöch 2015 m. w. N.). Die empirische Validierung solcher kombinierter Verfahren ist indes noch nicht besonders weit vorangeschritten. Zudem legen Untersuchungen den Schluss nahe, dass eine klinisch-individuelle Korrektur aktuuarischer Prognoseinstrumente zu einer Verschlechterung der mit diesen Instrumenten erstellten Prognosen führt (Rettenberger et al. 2017, S. 3). Dieses Phänomen des sog. „Clinical Override“ (Hanson/Morgon-Bourgon 2009; Wormith/Hogg/Guzzo 2012) bei Risikoprognosen ist zwar trotz seiner Verbreitung noch nicht in einem dafür angemessenen

Umfang untersucht (Guay/Parent 2018, S. 86; vgl. aber McCafferty 2017; Cohen/Pendergast/Vanbenschooten 2016; Wormith/Hogg/Guzzo 2015), müsste aber im Kontext von automatisierten Prognoseentscheidungen genau in den Blick genommen werden. Würde die in der Diskussion um künstliche Intelligenz zentrale Forderung nach menschlicher Entscheidungskompetenz mit dem zwangsläufigen Eingriff in automatisierte Prognoseprozesse (vgl. etwa Europäische Kommission 2020; Enquete-Kommission Künstliche Intelligenz, BT-Drs. 19/23700, S. 83) den beschriebenen Effekt haben, stünde hierdurch ein Vorteil von automatisierten Kriminalprognosen – ihre Treffgenauigkeit – wieder zur Disposition.

Neben dem Aufzeigen von Problemen prognostischer Entscheidungen ist die „Clinical Override“-Forschung im Kontext (automatisierter) Risikoprognosen aber auch sinnvoll, weil sie ermöglichen könnte, Kriterien für eine das ursprüngliche Prognoseergebnis verbessernde „Override“-Entscheidung zu identifizieren (Guay/Parent 2017, S. 86, 95). Das wäre einerseits vorteilhaft für bereits bestehende Prognoseinstrumente und andererseits voraussichtlich notwendig, um das beschriebene Dilemma im Rahmen automatisierter Prognoseinstrumente zu lindern.

## 2. Fragen der Treffsicherheit algorithmenbasierter automatisierter Verfahren

Ausgehend von diesen Grundlagen ist nun zu klären, welche Treffsicherheit automatisierte Verfahren zur Risikoprognose aufweisen, seien diese statistisch-aktuarisch oder KI-gestützt.

Dabei existieren für den Bereich der Rückfallprognose sowohl automatisierte Tools auf Basis „normaler“ statistisch-aktuarischer Prognosen als auch KI-basierte Ansätze (Rudin/Wang/Coker 2020, S. 5). Das wohl bekannteste Prognosetool ist dabei die in den USA in einigen Bundesstaaten eingesetzte, von einem privaten Unternehmen entwickelte Anwendung COMPAS (Correctional Offender Management Profiling for Alternative Sanctions; siehe zudem zum neuen, in den USA für nach Bundesrecht Verurteilte entwickelte Rückfallprognosetool PATTERN Cyphert 2020). Anders als häufig angenommen, handelt es sich dabei zwar um ein algorithmenbasiertes, computergestütztes Risikoprognosetool, zu seiner Entwicklung wurde jedoch kein Maschinenlernen eingesetzt, sodass es sich nicht um eine KI-basierte Anwendung handelt, sondern um ein statistisch-aktuarisches Tool (Rudin/Wang/Coker 2020, S. 5; für PATTERN hingegen ist bisher unbekannt, ob machine learning eingesetzt wurde: Cyphert 2020, S. 361f.). Es basiert zentral auf dem Risk-Need-Responsivity-Ansatz von Andrews und Bonta (grundlegend: Andrews/Bonta/Hoge 1990) und dementsprechend auf kriminologischen Annahmen über Zusammenhänge zwischen Rückfallrisiken und bestimmten kriminogenen Bedürfnissen (Jackson/Mendoza 2020). Insgesamt werden im Rahmen von COMPAS 137 Variablen erhoben (Rudin/Wang/Coker 2020), die allerdings nicht alle unmittelbar für die Risikoprognose relevant sind. Zentrale Prognosetools sind die General Recidivism Risk Scale (GRRS) und die Violent Recidivism Risk Scale (VRRS), die jeweils auf 26 Risikoitems basieren (Rudin/Wang/Coker 2020, S. 7f.; Jackson/Mendoza 2020, S. 4).

An COMPAS lassen sich dabei viele Probleme automatisierter Risikoprognosetools deutlich machen. So wird das Programm u. a. dafür kritisiert, dass es (aus kommerziellen Gründen) intransparent sei (Rudin/Wang/Coker 2020; a.A. Jackson/Mendoza 2020) und afro-amerikanische Menschen diskriminiere (Angwin et al. 2016; a.A. Rudin/Wang/Coker 2020). Auf beide Aspekte wird unten näher eingegangen; doch auch die Treffsicherheit wird teilweise kritisch eingeschätzt. So gelangten Dressel/Farid (2018) zu dem Ergebnis, dass das Programm keine besseren Prognosen anstellt als Personen ohne oder mit nur wenig strafrechtlich-kriminologischen Vorkenntnissen. Dieses Ergebnis wird allerdings in neueren Untersuchungen wieder in Zweifel gezogen (Lin et al. 2020), die durchaus eine generelle Überlegenheit von COMPAS gegenüber intuitiven Prognosen belegen. Die abweichenden Ergebnisse bei Dressel/Farid (2018) lassen sich womöglich damit erklären, dass erstens die Prognosen nur auf sieben Variablen basierten, zweitens die menschlichen Proband\*innen – anders als in der forensischen Realität – jeweils ein unmittelbares Feedback über ihre Treffgenauigkeit bekamen, sodass ein Lern- und Selbstkorrekturprozess schon während der Erhebung wirksam wurde, sowie durchgängig bei Dressel/Farid (2018) nur bezogen auf die allgemeine Basisrate der Rückfälligkeit prognostiziert wurde (48% in der untersuchten Gruppe von Straftäter\*innen) (Lin et al. 2020). Die Untersuchung von Lin et al. (2020) ergab, dass die Treffsicherheit von COMPAS diejenige der menschlichen Proband\*innen insbesondere dann deutlich übertraf, wenn die Proband\*innen kein Feedback über die Richtigkeit ihrer Prognose nach jeder Fallbeurteilung erhielten und zudem die Ba-

sistrate niedrig war (im konkreten Fall: Gewaltrückfall statt allgemeiner Rückfall). Eine Übersicht über AUC-Werte (korrekte Einstufung des Rückfallrisikos) aus verschiedenen Evaluationsstudien zu COMPAS und für verschiedene Rückfalldefinitionen findet sich bei Jackson/Mendoza (2020, S. 6f.); danach war die Treffsicherheit moderat und lag zwischen 0,64 und 0,74 (Zufall = 0,5).

Doch auch tatsächlich KI-gestützte, auf machine learning basierende Risikoprognosetools sind bereits im Einsatz bzw. in Entwicklung. So verglichen Ghasemi et al. (2021) auf drei verschiedenen Formen des machine learning basierende Prognosemodelle mit dem statistisch-aktuarischen Tool des LS/CMI (Level of Service/Case Management Inventory; Andrews/Bonta/Wormith o. J.). Hierbei wurden vergleichbare AUC-Werte im Bereich von bis zu 0,75 erreicht, die KI-basierten Modelle erhöhten die Treffsicherheit aber gerade bei den Personen mit mittlerem Rückfallrisiko, bei denen der LS/CMI nicht besser als eine Zufallseinstufung operiert, auf etwa 0,6. Auch eine Untersuchung von Duwe und Kim (2017) erbrachte eine tendenzielle Überlegenheit von Ansätzen, die auf Maschinenlernen basierten, gegenüber klassischen statistischen Methoden zur Rückfallprognose wie der logistischen Regression (anders noch Tollenaar/van der Heijden 2013).

Die Ergebnisse zeigen auf, dass für die automatisierte Risikoprognose, sowohl die statistisch-aktuarische als auch die KI-basierte, erhebliche Potentiale erkennbar sind – auch für einen Einsatz in Deutschland. Die (schon jetzt gute) Treffsicherheit dürfte sich dabei in den nächsten Jahren aufgrund der fortschreitenden techni-

schon Entwicklung nochmals erhöhen. Damit werden aber Fragen nach den rechtlichen Grenzen ihres Einsatzes und ggf. mit ihrem Einsatz verbundenen problematischen individual- und sozialpsychologischen Aspekten umso bedeutsamer. Diese werden im Folgenden erörtert.

#### **IV. Problempunkte**

##### **1. Verfassungsrechtliche Aspekte**

###### *a) Individualisierung der Entscheidung*

Das BVerfG hat in einer Entscheidung zur nachträglichen Sicherungsverwahrung einige Kautelen für Gefährlichkeitsprognosen aufgestellt und insbesondere eine ausreichende Individualisierung der Prognose gefordert (BVerfGE 109, 190, 240ff.). Die Anforderung steht im Zusammenhang mit der Prüfung der Verhältnismäßigkeit des mit der Freiheitsentziehung verbundenen Grundrechtseingriffs und hat daher verfassungsrechtlichen Gehalt (vgl. BVerfGE 109, 190, 240ff.; Boetticher et al. 2019, S. 558). Das BVerfG verlangt eine „umfassende Prüfung der Täterpersönlichkeit und der begangenen Taten“, die auf eine breite Tatsachengrundlage gestützt werden müsse (BVerfGE 109, 190, 241). Eine bloß abstrakte, auf statistische Wahrscheinlichkeiten gestützte Prognoseentscheidung sei unzulässig; es bedürfe vielmehr „unter Ausschöpfung der Prognosemöglichkeiten einer positiven Entscheidung über die Gefährlichkeit des Betroffenen, um die Freiheitsentziehung zu rechtfertigen“ (BVerfGE 109, 190, 242).

Dem hat sich der BGH in mehreren Entscheidungen angeschlossen. Er betont (erneut im Zusammenhang mit der nachträglichen Sicherungsverwahrung) das Er-

fordernis einer umfassenden Gesamtwürdigung der verurteilten Person, ihrer Taten und ergänzend ihrer Entwicklung während des Strafvollzugs; dabei verbiete sich auch nach Auffassung des BGH „eine abstrakte, auf statistische Wahrscheinlichkeiten gestützte Prognoseentscheidung [...]. Auch wenn bestimmte Persönlichkeitsstörungen von vornherein ein hohes Rückfallrisiko beinhalten, entbindet dies [...] nicht von einer individuellen Gefährlichkeitsprognose“. Das sei im vorliegenden Fall nicht gegeben, da lediglich die allgemeine Rückfallgefahr einer Gruppe vergleichbarer Personen (sog. Basisrate) angegeben worden sei (BGHSt 50, 121, 130 f.; ähnlich BGH NStZ 2009, 323; NStZ-RR 2016, 242; bei Pfister NStZ-RR 2011, 193, 198).

Bezüglich statistisch-aktuarischer Prognoseinstrumente (konkret: Static 99 im Rahmen der für die Sicherungsverwahrung erforderlichen Gefährlichkeitsprognose) führt der BGH aus, dass solche Instrumente für die Prognose zwar Anhaltspunkte über die Ausprägung eines strukturellen Grundrisikos liefern könnten. Sie seien indes nicht in der Lage, eine fundierte Einzelbetrachtung zu ersetzen. Hierfür bedürfe es zusätzlich einer differenzierten Einzelfallanalyse durch die Sachverständigen. Jedes Instrument könne nur eines von mehreren Hilfsmitteln sein, mit denen sich der\*die Gutachter\*in die Prognosebeurteilung erarbeitet (vgl. BGH bei Pfister 2010, 165).

Auch in den 2019 von einer interdisziplinär besetzten Gruppe aus Wissenschaftler\*innen und Praktiker\*innen formulierten „Empfehlungen für Prognosegutachten“ wird der Aspekt der „individuellen Fallbetrachtung“ als unverzichtbar

bezeichnet (Boetticher et al. 2019, S. 559). Für die vorliegende Fragestellung besonders relevant ist noch die knappe Bemerkung in einer Fußnote, wonach die Einschränkungen „erst recht“ für den Einsatz „prädiktiver Algorithmen“ gelten (Boetticher et al. 2019, S. 555, Fn. 28; ebenso Nink 2021, S. 420).

Man muss einschränkend berücksichtigen, dass sich die erwähnte Rechtsprechung ausschließlich auf die Verhängung stationärer Maßregeln bezieht. Letztere stellen insofern einen vergleichsweise intensiven Grundrechtseingriff dar, als sie überwiegend keine zeitlich fixierte Obergrenze aufweisen und in der Regel längere Freiheitsentziehungen mit sich bringen, während bei den §§ 56 ff. StGB Freiheitsstrafen von maximal 2 Jahren Dauer im Raum stehen. Hier könnten daher möglicherweise etwas weniger strenge Maßstäbe gelten.

Auch beziehen sich die Anforderungen der Rechtsprechung auf die von Sachverständigen erstellten Prognosen, nicht auf die (intuitive) Prognosetätigkeit der Gerichte. Allerdings erscheint das Individualisierungsgebot dennoch verallgemeinerbar, denn auch im Rahmen der §§ 56 ff. StGB geht es um grundrechtsintensive Entscheidungen, die sorgfältig und mit Blick auf den konkreten Fall begründet werden müssen. Es ist kein Grund ersichtlich, warum die o.g. Anforderungen nicht auch im Rahmen einer intuitiven Prognose erforderlich sein sollten, die ohne Inanspruchnahme externer Sachverständiger erstellt wird. Eine Entscheidung allein auf die Anwendung abstrakter statistischer Prognoseverfahren zu stützen, kommt also auch im Bereich der Aussetzung von Freiheitsstrafen zur Bewährung nicht in

Betracht (ebenso Nink 2021, 421). Doch steht dies nicht im Widerspruch dazu, statistische Prognoseinstrumente als *Hilfsmittel* für die Entscheidungsfindung einzusetzen. So wie es dem Gericht unbenommen ist, bei seiner Entscheidung kriminologische Gutachten oder technische Instrumente heranzuziehen (Kinzig 2019, § 56 Rn. 22 m.w.N.), kann es sich auch statistischer Prognoseverfahren bedienen, solange die von der Rechtsprechung aufgestellte Forderung einer individuellen, auf eine ausreichend breite Grundlage gestützten Prognose erfüllt wird – dies wird wohl auch für einen unterstützenden Einsatz von KI gelten.

Zu fragen wäre allerdings, ob beim Einsatz von KI, die auf der Basis einer sehr großen Menge aussagekräftiger Daten komplexe und ausdifferenzierte Entscheidungsvorschläge abgeben könnte, nicht ab einem gewissen Punkt der Bereich der bloß abstrakten „Gruppenwahrscheinlichkeit“ verlassen und eine ausreichende Individualisierung erzeugt wird, die dann sogar das Tor zu einer vollständig automatisierten Entscheidung öffnen könnte. Hiergegen wendet sich u.a. Sommerer (2020, S. 101 ff.), da es sich – auch bei hochentwickelten, viele Einzelmerkmale berücksichtigenden Algorithmen – stets nur um Korrelationen im Rahmen von Gruppenwahrscheinlichkeiten handele. Der Algorithmus könne nicht über seine Datengrundlage hinausblicken, auch wenn er darin Muster erkenne, die dem Menschen verborgen blieben. Jedenfalls aktuell wird man für KI-basierte Prognoseentscheidungen festhalten können, dass es sich nicht um „echte“ Einzelfallentscheidungen handelt. Dies führt zu dem – vordergründig paradox anmutenden – Ergebnis, dass trotz der Gefahr des

„clinical override“ (s. o.) weder statistisch-aktuarische noch KI-gestützte Prognosen nach derzeitiger Rechtslage ohne zusätzliche Würdigung durch einen Menschen für die Legalprognose übernommen werden dürften.

#### b) *Menschenwürde und Transparenz*

Aus dem Menschenwürdekern des Rechts auf informationelle Selbstbestimmung, Art. 2 Abs. 1 i. V. m. Art. 1 Abs. 1 GG, ergibt sich, dass Personen bei Informationsverarbeitungen nicht zum Objekt einer Maschine gemacht werden dürfen (Botha 2009; Zöller 2002, S. 44). Nach der „Anerkennungstheorie“ konstituiert sich die Würde des Menschen u. a. in sozialer Anerkennung, also „durch positive Bewertung von sozialen Achtungsansprüchen“ (Hofmann 1993, S. 353 ff.). Wenn bei einer Datenverarbeitung die Subjektivität eines Individuums infrage gestellt wird, kann die Menschenwürde berührt sein (Duwell 2014, S. 28); der Mensch droht durch prognostische Datenverarbeitung zu einem „Informationsobjekt“ „degradiert“ zu werden (Schenke et al. 2019, § 18 Rn. 11; Zöller 2002, S. 163); von einem Verbot der Reduzierung von Personen auf „Informationsobjekte“ sprach auch bereits das BVerfG in seinem Volkszählungsurteil (BVerfGE 65, 1, 48). Dem Individuum muss ein Mindestmaß an Selbstbestimmung und Nachvollziehbarkeit der Verarbeitung seiner personenbezogenen Daten erhalten bleiben (Zöller 2002, S. 44). Wann die Schwelle zur Objektivierung im Rahmen staatlicher Datenverarbeitung genau überschritten ist, ist bisher kaum näher bestimmt worden. Nach Ansicht von Sommerer sei hiervon auszugehen, wenn ein systemimmanent intransparenter, d. h. menschlich nicht mehr nachvollziehbarer

Algorithmus, dessen genaue Abläufe selbst Fachleute nicht mehr verstehen, in einem grundrechtssensiblen Bereich wie der Kriminalitätskontrolle eine Prognose erstellt und Beamt\*innen oder Richter\*innen darauf basierend unmittelbar einen Eingriff in Grundrechte vornehmen (vgl. Sommerer 2020, S. 234 ff.).

Transparenz wird nicht nur ganz allgemein im Hinblick auf die Grundlagen einer Kriminalprognose gefordert (vgl. nur Boetticher et al. 2006, S. 539); sie ist auch ein vieldiskutiertes Kernproblem im Zusammenhang mit dem Einsatz von KI (z. B. bei Sommerer 2020; Nink 2021). Dies gilt insbesondere dort, wo die Maschine nicht einfach rein deduktiv auf Basis programmierter Gesetzmäßigkeiten Rechenergebnisse auswirft (hier wären die Maßstäbe der Entscheidung vollständig bekannt), sondern wo sie induktiv aus den eingespeisten Daten im Wege des maschinellen Lernens eigene Regeln formuliert, die als Basis für künftige Entschlüsse dienen (dazu im Zusammenhang mit Prognosebildung Müller-Mall 2020). Insofern wird regelmäßig das Bild vom Algorithmus als undurchsichtiger „Black Box“ gezeichnet (z. B. Herberger 2018, S. 2828; Müller-Mall 2020; Ramge 2020). Diese Undurchsichtigkeit ist zunächst technischer Natur, kann aber durch gezielte Formen der Geheimhaltung, sei es zum Schutz kommerzieller Interessen oder aufgrund angeblicher staatlicher Geheimhaltungsnotwendigkeiten verstärkt werden (Burrell 2016, S. 3 ff.; Kehl et al. 2017, S. 28). Daher spielt sie z. B. auch bei kommerziell vertriebenen automatisierten, rein statistisch-aktuarien Instrumenten wie COMPAS eine Rolle (kritisch zur dortigen Intransparenz z. B. Rudin et al. 2020; beschwichtigend Jackson/Mendoza 2020). Als Reaktion auf

die Undurchsichtigkeit herkömmlicher KI-Lösungen wird bereits an der Entwicklung sog. „explainable AI“ (erklärbarer KI) gearbeitet (dazu Holzinger 2018; Deeks 2019). Dabei gibt es sogar bereits konkrete Ansätze rückfallprognostischer KI-basierter Algorithmen, deren Entscheidungsbasis transparent und nachvollziehbar ist (ausführlich Wang et al. 2020). Ihre Treffsicherheit ist „Black Box“-Algorithmen sowie den automatisierten Risikoprognosetools COMPAS und Arnold Public Safety Assessment nach Wang et al. (2020, S. 28) fast ebenbürtig.

Zudem wird darauf hingewiesen, dass die Undurchschaubarkeit des richterlichen Entscheidungsprozesses nicht nur im Zusammenhang mit künstlicher Intelligenz Bedeutung gewinne, sondern dass auch menschliche Entscheidungsprozesse nie völlig transparent seien (Kaspar/Höffler/Harrendorf 2020, S. 52; Rudin et al. 2020, S. 32; krit. zu einer Gleichsetzung aber Sommerer 2020, S. 226 ff.). Dies gilt mit Blick auf intuitive Prognoseentscheidungen in besonderem Maße. Zwar muss das Gericht seine Bewährungsentscheidung begründen, allerdings bedeutet dies nicht, dass alle relevanten Erwägungen in den gerichtlichen Ausführungen offengelegt werden müssen – oder auch nur den Entscheider\*innen stets bewusst wären. Eine vollständige Transparenz gibt es also auch bei Richter\*innen nicht. Zudem kann der Mensch seine Entscheidungsbegründung zu einem gewissen Grad anpassen und damit einer auf rechtlich fragwürdigen Prämissen beruhenden subjektiven Entscheidung den Anstrich einer juristisch-rationalen Entscheidung verleihen. Das Transparenzproblem ist also kein rein maschinenspezifisches.

Die Frage ist aber doch, wie man auf das Problem reagiert. Vorschläge reichen von einer möglichst frühzeitigen Mitteilung technischer Daten (Kehl et al. 2017, S. 32), insb. auch der Quell- und Protokoll Daten (Sommerer, 2020), bis zur Aufklärung Betroffener über grundlegende Funktionsweisen des Programms (Nink 2021; Sommerer 2020). Freilich sollte man bei der Frage nach der Offenlegung der technischen Prozesse eine wichtige Begleitfrage nicht aus den Augen verlieren: Was geschieht mit dem so gewonnenen Wissen? Angenommen, die Auswertung der Systemdaten ließe tatsächlich erkennen, dass die KI im Wege des induktiven maschinellen Lernens eine rechtlich unzulässige Regel gebildet hat: Wie und durch wen erfolgt erforderlichenfalls eine Korrektur? Und wie wäre bei proprietären Algorithmen einem potentiellen Einfluss Privater auf die Rechtsprechung zu begegnen?

### c) Richterliche Unabhängigkeit

Eine vollständig automatisierte, computergestützte Entscheidung, bei der dem Gericht im Bereich der Aussetzungsentscheidungen keine Letztentscheidungskompetenz mehr verbleibt, ist derzeit noch Zukunftsmusik, soll hier aber dennoch als mögliches Szenario prospektiv diskutiert werden, zumal die „ausschließlich auf einer automatischen Verarbeitung beruhende Entscheidung“ bereits jetzt im Datenschutzrecht Erwähnung findet (vgl. Art. 22 DS-GVO; Art. 11 JI-Richtlinie). Hier könnten sich verfassungsrechtliche Probleme mit Blick auf die Unabhängigkeit der Gerichte gem. Art. 97 I GG ergeben.

Allerdings müsste man differenzieren: Wie oben aufgezeigt, ist eine ausnahmslose Bindung des Gerichts an die von der

KI angegebene Wahrscheinlichkeit nicht gegeben, da das Gericht selbst die Gefahrprognose aufgrund eines normativen Prozesses schuldet (Boetticher et al. 2019), also weiterhin den entscheidenden Einfluss auf die Frage der Gewährung der Aussetzung hat. Selbst die Bewertung der Prognose i.S.d. § 56 I 1 StGB ist kein rein empirischer Vorgang. Vielmehr spielen eben auch in einem zweiten Schritt normative Aspekte, z.B. zur Hinnehmbarkeit zumeist bestehender Restrisiken, eine Rolle (Meier 2021, § 7 Rn. 45). Des Weiteren ist zumindest auf absehbare Zeit fraglich, wie eine KI-gestützte Lösung auch der geforderten Interventionsprognose vollauf gerecht werden und insofern z.B. auch die Vielzahl der über Bewährungsweisungen und -auflagen zur Verfügung stehenden „Treatments“ bewerten und gewichten können soll.

Man müsste also klären, ob dann, wenn lediglich eines von mehreren Entscheidungsmerkmalen der richterlichen Dezi-sion „entzogen“ würde, eine unzulässige Einflussnahme im Sinne von Art. 97 I GG angenommen werden könnte. Damit verbunden ist die Frage, in welchem Umfang es dem demokratisch legitimierten Gesetzgeber freisteht, ursprünglich richterliche Aufgaben ganz oder teilweise im Wege eines formellen Gesetzes anderweitig erledigen zu lassen. Dabei erschiene es noch nicht einmal ausgeschlossen, auch ein KI-System als „Richter\*in“ im Sinne von Art. 92, 97 GG zu definieren, auch wenn das sehr gewöhnungsbedürftig klingt und vertiefter Überlegung bedürfte. Eine solche Erweiterung des Richter\*innenbegriffs böte sich möglicherweise an, wenn der Telos des Art. 97 I GG, also die Garantie einer möglichst von äußeren Einflussnahmen freien Entscheidung, von

einem Computersystem besser erfüllt werden könnte als von einem mit Menschen besetzten Gericht mit all seinen Schwächen und (auch unbewussten) Vorurteilen (generell ablehnend mit dem zentralen Argument, dass nur Menschen Verantwortung übernehmen können (*conditio humana*), und unter Verweis auf Art. 92 und 101 I 2 GG aber Greco 2020). Dazu müsste die KI freilich derartige Schwächen möglichst überwunden haben. Neutralität und Diskriminierungsfreiheit der KI-Entscheidung hängen dabei wesentlich davon ab, dass sich in den zu Grunde liegenden Algorithmen nicht entsprechende Vorurteile manifestiert haben, etwa durch eine entsprechende „Kontaminierung“ der Trainingsdaten (s. dazu sogleich), was bislang nicht zweifelsfrei zu gelingen scheint. Selbst wenn diese verfassungsrechtlichen Fragen geklärt werden könnten, zeigen die anderen, teilweise verschränkten Problempunkte aber, dass damit noch keineswegs der Weg für einen Einsatz automatisierter Risikoprosen frei wäre.

#### d) Diskriminierungsfreiheit

Fragen stellen sich auch bezüglich der Diskriminierungsfreiheit KI-basierter und sonstiger automatisierter Risikoprosensysteme (Art. 3 III GG). Tatsächlich steht zu befürchten, dass die Verwendung KI-gestützter Prognosetools Vorurteile unter bestimmten Umständen, nämlich dann, wenn die Trainingsdaten selbst in dieser Form (und sei es nur unterschwellig) „biased“ sind, noch verfestigt. Sie bedürfen insofern eines externen ethischen Korrektivs (Martini 2019, S. 46ff.). Hinzu kommt die Gefahr, dass automatisierte Risikoprosentools möglicherweise statische Faktoren zu stark gewichten, z. B. Vorstrafen. Dieses Risiko besteht insbesondere

dann, wenn die Risikoprosen eher nach Aktenlage bzw. jedenfalls auf der Basis bereits vorliegender Daten erhoben wird, weil dann dynamische Faktoren, die die Rückfallprognose entscheidend beeinflussen können, aber noch nicht in der Akte stehen, für ein algorithmisches Prognoseinstrument schlicht unerkannt bleiben. Die einem KI-System fehlende Fähigkeit zur Exploration ließe sich nur teilweise durch den Versuch einer größtmöglichen Aktualität der Aktenlage kompensieren, anhand derer ein automatisiertes Prognoseinstrument dann möglicherweise zu einer besseren Berücksichtigung dynamischer Faktoren käme. Ob und wie das gelingen kann, sind offene Fragen, die idealerweise im Kontext der sog. elektronischen Akte untersucht werden sollten.

— Gerade zum oben schon dargestellten Prognosetool COMPAS gibt es eine Debatte darüber, wie stark gewisse Diskriminierungen dem Programm eingeschrieben sind. Wie bereits oben kurz erwähnt, haben Angwin et al. (2016), basierend auf einer externen Analyse der Rückfallprognosen von COMPAS den Vorwurf erhoben, dass das Tool Menschen afroamerikanischer Herkunft systematisch rückfallprognostisch schlechter einstufte als Weiße. Ob diese Einschätzung zutrifft, wird seither auf der Grundlage verschiedener technischer und fairnessbasierter Überlegungen kontrovers diskutiert (ablehnend z. B. Rudin et al. 2020, S. 19f.; einen Überblick bieten Dressel/Farid 2018).

Denkbar wäre es allerdings auch, KI-gestützte Modelle dazu zu nutzen, menschliche Vorurteile und andere Fehler aufzudecken, indem man bspw. die KI teilweise über entsprechende Fälle laufen lässt und prüft, ob sie vorurteilsbezogen

entscheidet oder nicht. Geschieht dies, weiß man, dass das menschliche Vorurteil in der KI „eingeschrieben“ ist und kann diese umprogrammieren (dazu ausführlich Ramge 2020, S. 61 f.). Eine weitere Möglichkeit zur Verringerung eines diskriminierenden Bias ist es, verschiedene, in der Vorhersagegenauigkeit vergleichbare Algorithmen miteinander bezüglich einer etwaigen diskriminierenden Tendenz zu vergleichen und gezielt die am wenigsten diskriminierenden Tools zu identifizieren und weiterzuentwickeln (Skeem/Lowenkamp 2020).

## 2. Sozial- und individualpsychologische Aspekte

### a) *Automation Bias*

Schließlich sind beim Einsatz von KI zur Entscheidungsunterstützung auch soziale und psychologische Phänomene einer Mensch-Maschine-Interaktion zu berücksichtigen. Dies gilt nicht nur in praktischer Hinsicht für das Design eines Systems, sondern auch mit Blick auf rechtliche Rahmenbedingungen. In der Diskussion um KI wird mitunter argumentiert, dass KI-unterstützte Entscheidungen wenig rechtliche Neuerungen brächten und keiner besonderen Regulierung bedürften, da letztendlich doch stets eine Person entscheide. Diesem stünde lediglich eine zusätzliche Informationsquelle bzw. ein neues Werkzeug zur Verfügung. In der englischsprachigen Literatur wird dies auch als „Human in the Loop“-Argument diskutiert (vgl. etwa Jones 2017): zwischen der „Maschine“ und betroffenen Bürger\*innen sei doch weiterhin eine Person, die korrigierend eingreifen könne, eingebunden. So könnten z. B. durch menschlichen Eingriff Diskriminierungen vermieden und Intrans-

parenzen kompensiert werden. Diese pauschalen Annahmen gehen jedoch fehl, da hier die menschliche Leistungsfähigkeit gravierend überschätzt wird. Dies verdeutlicht ein Blick auf das psychologische Phänomen des sog. Automation Bias (Mosier et al. 1998; Skitka et al. 1999, 2000a; Cummings 2006a, S. 25), d. h. der Unfähigkeit von Menschen, sich in der Praxis den Handlungsvorschlägen eines komplexen algorithmischen Systems entziehen zu können.

Studien zu Entscheidungsunterstützungssystemen im medizinischen Bereich sowie bei Pilot\*innen (Manzey 2012, S. 333; Skitka et al. 1999; Mosier et al. 1998, S. 63) haben bereits gezeigt, dass es Personen schwerfällt, sich gegen das Ergebnis algorithmischer Berechnungen zu entscheiden, und zwar selbst dann, wenn das Ergebnis des Algorithmus eigentlich nur einen von mehreren Entscheidungsfaktoren darstellen sollte. Automation Bias führt dazu, dass Menschen es unterlassen, zusätzlich zu einem algorithmischen Ergebnis selbstständig Informationen einzuholen und zu bewerten, und sogar deutlich gegen das Ergebnis des Algorithmus sprechende Anhaltspunkte bewusst ignorieren (Skitka et al. 1999, 2000a, 2000b; Mosier et al. 1998). Der eigenen Expertise wird weniger vertraut als dem Ergebnis des algorithmischen Prozesses, das dem Individuum mit der gesetzten Selbstverständlichkeit des Faktischen entgegentritt. Im Ergebnis führt dies dazu, dass algorithmische Berechnungen, die lediglich als Entscheidungsunterstützung gedacht waren, faktisch die Entscheidung des Menschen determinieren (Cummings 2006b, S. 7; s. a. Ramge 2020, S. 61) und der Mensch die eigene Verantwortung an die algorithmische Vor-

gabe ausgelagert. Eine erhoffte Kontrolle einer algorithmischen Entscheidung durch Individuen läuft so häufig ins Leere (Citron 2007, S. 1277).

Vorkehrungen gegen Automation Bias sind bisher in keinem in Deutschland eingesetzten Entscheidungsunterstützungssystem getroffen. Auch wenn vereinzelt Techniken zur gezielten Sensibilisierung der Nutzer\*innen für das Phänomen des Automation Bias vorgeschlagen wurden (Ahlstrom/Longo 2003, Kap. 3, S. 20), ist fraglich, ob sich das Phänomen überhaupt grundsätzlich beheben lässt (Skitka et al. 1999, S. 1004). Unklar ist, inwieweit bestehende Bemühungen in den USA, wie jene des Wisconsin Supreme Court, Automation Bias tatsächlich zu verhindern vermögen (Freeman 2016, S. 95f.). Dort wurde bestimmt, dass jeder computerbasierte Risikoscore des COMPAS-Algorithmus (s. o.) mit „Warnhinweisen“ u. a. zur mangelnden Transparenz und potenziell diskriminierenden Wirkung des Systems gegenüber Minderheiten zu versehen sei (State v. Loomis, 881 N.W.2 d 749 (WI 2016), Rn. 66).

Deutlich wird jedenfalls, dass die Einbeziehung von Psycholog\*innen neben Computerwissenschaftler\*innen und Jurist\*innen bei der Entwicklung und Begleitung des Einsatzes von Algorithmen in Zukunft eine zentrale Rolle spielen wird. Bei der Erarbeitung von Strategien zur Eindämmung des Automation Bias-Effekts wird man dabei auch das Phänomen des „Clinical Override“ (s. o.) nicht aus den Augen verlieren dürfen, dessen gegenwärtige Implikation an sich eine möglichst geringe Korrektur automatisierter Entscheidungen durch Menschen wäre und der daher die Lösung der Bias-Problema-

tik aufgrund seiner entgegengesetzten Stoßrichtung verkompliziert.

#### b) Akzeptanz in der Bevölkerung (Automation Aversion)

Schließlich ist aus kriminalpolitischer und letztlich auch straftheoretischer Perspektive durchaus interessant, ob die Bevölkerung sich gegenüber einem Einsatz von KI oder generell einer Automatisierung von Entscheidungen im Bereich des Strafrechts offen zeigt oder ob (im Sinne einer „Automation Aversion“) eher Ablehnung herrscht. Nach den vorliegenden Studien ist letzteres der Fall. Dabei gibt es Hinweise darauf, dass das Unbehagen gegenüber maschinellen Entscheidungen gerade dort besonders groß ist, wo sie sich schwerwiegend auf das Leben der Betroffenen auswirken (Bigman/Gray 2018). In einer neueren Studie des Allensbach-Instituts im Auftrag der Bertelsmann Stiftung befürworteten die Befragten einen vom Menschen losgelösten Einsatz von KI bei recht unspektakulären Aufgaben wie der Verwaltung von Lagerräumen; von den insgesamt 16 konkreten Entscheidungssituationen war dagegen die „Beurteilung des Risikos, ob ein Straftäter rückfällig wird“ diejenige, bei der mit Abstand am wenigsten dem Computer getraut wurde: nur 2% stimmten für dessen alleinige Entscheidung, 37% stimmten der Option eines bloßen Computer-Vorschlags zu und 54% sagten, es solle allein ein Mensch entscheiden (Fischer/Petersen 2018, S. 26; ähnliche Ergebnisse bei Bigman/Gray 2018).

Dabei legen die vorhandenen Studien nahe, dass sich die Aversion gegen die Maschine abschwächen lässt, wenn die Nutzenden die Möglichkeit erhielten, auf

das Ergebnis des Algorithmus notfalls noch einzuwirken (Dietvorst et al. 2016) oder wenn die Maschine lediglich beratende Funktion hatte, so dass die „Letztentscheidungskompetenz“ bei einem Menschen verbleiben sollte (Bigman/Gray 2018, S. 30). Zudem könnte das Vertrauen in KI durch mehr Information über ihre Wirkungsweise sowie über eine zunehmende Integrierung in den Alltag und den damit einhergehenden Gewöhnungseffekt möglicherweise steigen (Nink 2021, S. 342).

## V. Fazit

Das Zeitalter der Massendaten und Algorithmen – so viel lässt sich mit Gewissheit sagen – gestaltet die Art und Weise, wie wir Wissen über die Welt produzieren und daran anknüpfend handeln, grundlegend um. Die beschriebenen Problemlagen von automatisierten Risikoprognosen, die lediglich einen kleinen Ausschnitt der derzeitigen rechtswissenschaftlichen KI-Debatte widerspiegeln, verdeutlichen aber, dass für eine geregelte Nutzbarmachung algorithmenbasierter Systeme in Deutschland noch ein weiter Weg zu gehen ist. Sicher ist nur, dass die potentielle Automatisierung von Prognosen im Strafjustizsystem noch in erheblichem Maße der weiteren wissenschaftlichen Durchdringung der verschiedenen Fragestellungen bedarf. Wichtig ist in diesem Zusammenhang auch das Gebot des ergebnisoffenen Forschens: Wenn die weiteren Erkenntnisse darauf hindeuten, dass die aufgeworfenen Komplexitäten nicht in einer Weise aufgelöst werden können, die den Maximen unserer Verfassungsordnung entspricht, wäre auch das ein zu akzeptierendes Ergebnis.

## Literatur

Ægisdóttir, S., White, M., Spengler, P., Maugherman, A., Anderson, L., Cook, R., Nichols, C., Lampropoulos, G., Walker, B., Cohen, G. & Rush, J. (2006): The Meta-Analysis of Clinical Judgment Project: Fifty-Six Years of Accumulated Research on Clinical Versus Statistical Prediction. *The Counseling Psychologist*, 34, 341–382.

Andrews, D. A., Bonta, J. & Wormith, J. S. (o. J.): LS/CMI. Level of Service/Case Management Inventory. Online: <https://mhs.com/info/lscmi/>.

Andrews, D. A., Bonta, J. & Hoge, R. D. (1990): Classification for Effective Rehabilitation: Rediscovering Psychology. *Criminal Justice and Behavior*, 17, 19–52.

Aumann, Y. (2018): Artificial Intelligence vs. Human in the Legal Profession. *The Legal Revolutionary*, 74–76. Online: [https://legal-revolution.com/images/pdf/Artificial-Intelligence\\_vs\\_Human\\_In\\_The\\_Legal\\_Profession.pdf](https://legal-revolution.com/images/pdf/Artificial-Intelligence_vs_Human_In_The_Legal_Profession.pdf).

Angwin, J., 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. Online: <https://www.propublica.org/article/machine-bias-risk-assessments-in-criminal-sentencing>.

Barry-Jester, A. M., Casselman, B., & Goldstein, D. (2015): The new Science of sentencing. The Marshall Project. Online: <https://www.the-marshallproject.org/2015/08/04/the-new-science-of-sentencing>.

Bigman, Y. & Gray, K. (2018): People are Averse to Machines Making Moral Decisions. *Cognition*, 181, 21–34.

Boetticher, A., Kröber, H., Müller-Isberner, R., Böhm, K., Müller-Metz, R. & Wolf, T. (2006): Mindestanforderungen für Prognosegutachten. *Neue Zeitschrift für Strafrecht*, 10, 537–544.

Boetticher, A., Koller, M., Böhm, K., Brettel, H., Dölling, D., Höffler, K., Müller-Metz, R., Pfister, W., Schneider, U., Schöch, H. & Wolf, T. (2019): Empfehlungen für Prognosegutachten. Rechtliche Rahmenbedingungen für Prognosen im Strafverfahren. *Neue Zeitschrift für Strafrecht*, 10, 553–573.

Botha, H. (2009): Human Dignity in Comparative Perspective. *Stellenbosch Law Review*, 20, 171–220.

- Burrell, J. (2016): How the Machine 'Thinks': Understanding Opacity in Machine Learning Algorithms. *Big Data & Society*, 3, 1–12.
- Citron, D. K. (2007): Technological Due Process. *Washington Law Review*, 85, 1249–1313.
- Cummings, M. L. (2006a): Automation and Accountability in Decision Support System Interface Design. *Journal of Technology Studies*, 32, 23–31.
- Cummings, M. L. (2006b): The Social and Ethical Impact of Decision Support Interface Design. In: Karowski, W. (Hrsg.) *International Encyclopedia of Ergonomics and Human Factors*. New York: Taylor & Francis, 1249–1253.
- Cohen, T. H., Pendergast, B. & VanBenschoten, S. W. (2016): Examining Overrides of Risk Classifications for Offenders on Federal Supervision. *Federal Probation*, 80, 12–21.
- Crawford, K. (2021): Atlas of AI. Power, Politics, and the Planetary Costs of Artificial Intelligence. New Haven: Yale University Press.
- Cyphert, A. B. (2020): Reprogramming Recidivism: The First Step Act and Algorithmic Prediction of Risk. *Seton Hall Law Review*, 51, 331–381.
- Dahle, K. & Lehmann, R. (2018): Zum prognostischen Mehrwert einer integrativen nomothetisch-idiografischen kriminalpsychologischen Prognosebeurteilung – Eine empirische Untersuchung an männlichen Gewalt- und Sexualstraftätern. *Forensische Psychiatrie, Psychologie, Kriminologie*, 12, 37–50.
- Deeks, A. (2019): The Judicial Demand for Explainable Artificial Intelligence. *Columbia Law Review*, 119, 1829–1850.
- Dietvorst, B. J., Simmons, J. P., & Massey, C. (2016): Overcoming Algorithm Aversion: People will Use Imperfect Algorithms If They Can (Even Slightly) Modify Them. *Management Science*, 64, 1155–1170.
- Dölling, D. (2015): Qualitätsstandards für die Bewährungshilfe. In: Bannenberg, B., Brettel, H., Freund, G., Meier, B.-D., Remschmidt, H., Safferling, C. (Hrsg.) *Über allem: Menschlichkeit: Festschrift für Dieter Rössner*. Baden-Baden: Nomos, S. 75–84.
- Dressel, J. & Farid, H. (2018): The Accuracy, Fairness, and Limits of Predicting Recidivism. *Science Advances* 4. Online: <https://advances.sciencemag.org/content/4/1/eaao5580>.
- Duwe, G. & Kim, K. (2017): Out with the Old and in with the New? An Empirical Comparison of Supervised Learning Algorithms to Predict Recidivism. *Criminal Justice Policy Review*, 28, 570–600.
- Duwell, M. (2014): Human Dignity: Concept, Discussions, Philosophical Perspectives: Interdisciplinary Perspectives. In: Duwell, M., Braarvig, J., Brownsword, R. & Mieth, D. (Hrsg.) *Cambridge Handbook on Human Dignity*. Cambridge: Cambridge University Press, 23–50.
- Ebersbach, M. (2020): Big Data, Algorithmen und Bewährungsentscheidungen. In: Momsen, C. & Schwarze, M. (Hrsg.), *Strafrecht im Zeitalter von Digitalisierung und Datafizierung*, Sammelband, KriPoZ Junges Publizieren 2020, 26–37. Online: <https://kripoz.de/wp-content/uploads/2020/06/kripoz-jup-sammelband-strafrecht-im-zeitalter-von-digitalisierung-und-datafizierung.pdf>.
- Europäische Kommission (2020): Weißbuch zur Künstlichen Intelligenz – ein europäisches Konzept für Exzellenz und Vertrauen. COM (2020) 65 final. Online: [https://ec.europa.eu/info/sites/default/files/commission-white-paper-artificial-intelligence-feb2020\\_de.pdf](https://ec.europa.eu/info/sites/default/files/commission-white-paper-artificial-intelligence-feb2020_de.pdf).
- Fischer, S. & Petersen, T. (2018): Was Deutschland über Algorithmen weiß und denkt. Ergebnisse einer repräsentativen Bevölkerungsumfrage. Online: [https://www.bertelsmann-stiftung.de/fileadmin/files/BSt/Publikationen/GrauePublikationen/Was\\_die\\_Deutschen\\_ueber\\_Algorithmen\\_denken.pdf](https://www.bertelsmann-stiftung.de/fileadmin/files/BSt/Publikationen/GrauePublikationen/Was_die_Deutschen_ueber_Algorithmen_denken.pdf).
- Freeman, K. (2016): Algorithmic Injustice: How the Wisconsin Supreme Court Failed to Protect Due Process Rights in State V. Loomis. *North Carolina Journal of Law and Technology*, 18, 75–180.
- Ghasemi, M., Anvari, D., Atapour, M., Wormith, J. S., Stockdale, K. C., Spiteri, R. J. (2021): The Application of Machine Learning to a General Risk–Need Assessment Instrument in the Prediction of Criminal Recidivism. *Criminal Justice and Behavior*, 48, 518–538.
- Greco, L. (2020): Richterliche Macht ohne richterliche Verantwortung: Warum es den Roboter-Richter nicht geben darf. *Rechtswissenschaft*, 11, 21–62.
- Grube, A. (2010): Die Strafaussetzung zur Bewährung. *Juristische Ausbildung*, 32, 759–765.
- Guay, J.-P. & Parent, G. (2018): Broken Legs, Clinical Overrides, and Recidivism Risk: An Analysis of Decisions to Adjust Risk Levels with

the LS/CMI. *Criminal Justice and Behavior*, 45, 82–100.

Hanson, R. K. & Morton-Bourgon, K. E. (2009): The Accuracy of Recidivism Risk Assessments for Sexual Offenders: A Meta-Analysis of 118 Prediction Studies. *Psychological Assessment*, 21, 1–21.

Herberger, M. (2018): „Künstliche Intelligenz“ und Recht – Ein Orientierungsversuch. *Neue Juristische Wochenschrift*, 2825–2829.

Hofmann, H. (1993): Die versprochene Menschenwürde. *Archiv des öffentlichen Rechts*, 118, 353–377.

Holzinger, A. (2018): Explainable AI (ex-AI). *Informatik Spektrum*, 41, 138–143.

Jackson, E. & Mendoza, C. (2020): Setting the Record Straight: What the COMPAS Core Risk and Need Assessment Is and Is Not. *Harvard Data Science Review* 2.1. Online: <https://hdsr.mitpress.mit.edu/pub/hzwo7ax4/release/4>.

Jones, M. L. (2017): The Right to a Human in the Loop: Political Constructions of Computer Automation and Personhood. *Social Studies of Science*, 47, 216–239.

Kaspar, J. (2018): Sentencing Guidelines versus freies tatrichterliches Ermessen – Brauchen wir ein neues Strafzumessungsrecht? Gutachten C zum 72. Deutschen Juristentag.

Kaspar, J., Höffler, K. & Harrendorf, S. (2020): Datenbanken, Online-Votings und künstliche Intelligenz – Perspektiven evidenzbasierter Strafzumessung im Zeitalter von „Legal Tech“. *Neue Kriminalpolitik*, 32, 35–56.

Kehl, D., Guo, P. & Kessler, S. (2017): Algorithms in the Criminal Justice System: Assessing the Use of Risk Assessments in Sentencing. Online: <http://nrs.harvard.edu/urn-3:HUL.InstRepos:33746041>.

Kett-Straub, G. & Groß, K. (2020): Kommentierung zu § 56. In: Joecks, W., Miebach, K. (Hrsg.) *Münchener Kommentar zum Strafgesetzbuch. Band 2. §§ 38–79b*. 4. Aufl. München: C. H. Beck.

Kinzig, J. (2019): Kommentierung zu § 56. In: Schönke, A., Schröder, H. (Hrsg.) *Strafgesetzbuch. Kommentar*. 30. Aufl. München: C. H. Beck.

Kleinberg, J., Himabindu, L., Leskovec, J., Ludwig, J. & Mullainathan, S. (2018): Human Decisions and Machine Predictions. *The Quarterly Journal of Economics*, 133, 237–293.

Labi, N. (2012): Misfortune Teller. A Statistics

Professor Says He Can Predict Crime before It Occurs. The Atlantic. Online: <https://www.theatlantic.com/magazine/archive/2012/01/misfortune-teller/308846/>.

Lin, Z., Jung, J., Goel, S. & Skeem, J. (2020): The Limits of Human Predictions of Recidivism. *Science Advances* 6. Online: <https://advances.sciencemag.org/content/6/7/eaaz0652>.

Manzey, D. (2012): Systemgestaltung und Automatisierung. In: Badke-Schaub, P., Hofinger, G. & Lauche, K. (Hrsg.) *Human Factors – Psychologie sicheren Handelns in Risikobranchen*. Berlin/Heidelberg: Springer, S. 307–324.

Martini, M. (2019): Blackbox Algorithmus – Grundfragen einer Regulierung künstlicher Intelligenz. Berlin: Springer.

McCafferty, J. T. (2017): Professional Discretion and the Predictive Validity of a Juvenile Risk Assessment Instrument: Exploring the Overlooked Principle of Effective Correctional Classification. *Youth Violence and Juvenile Justice*, 15, 103–118.

Mehta, H., Shah, S., Patel, N. & Kanani, P. (2020): Classification of Criminal Recidivism Using Machine Learning Techniques. *International Journal of Advanced Science and Technology*, 29, 5110–5122.

Meier, B.-D. (2021): *Kriminologie*. 6. Aufl., München: C. H. Beck.

Mosier, K. L., Skitka, L. J., Heers, S. & Burdick, M. (1998): Automation Bias: Decision Making and Performance in High-Tech Cockpits. *International Journal of Aviation Psychology*, 8, S. 47–63.

Müller-Mall, S. (2020): *Freiheit und Kalkül. Die Politik der Algorithmen*. Stuttgart: Reclam.

Nink, D. (2021): *Justiz und Algorithmen. Über die Schwächen menschlicher Entscheidungsfindung und die Möglichkeiten neuer Technologien in der Rechtsprechung*. Berlin: Duncker & Humblot.

Office of the Assistant Attorney General. (2014): *The Promise and Danger of Data Analytics in Sentencing and Corrections Policy*. Online: <https://www.justice.gov/sites/default/files/criminal/legacy/2014/08/01/2014annual-letter-final-072814.pdf>.

Pfister, W. (2010): Die Beurteilung der Schuldfähigkeit in der Rechtsprechung des Bundesgerichtshofs. *Neue Zeitschrift für Strafrecht – Rechtsprechungs-Report*, 6, 161–166.

- Ramge, T. (2020): *Augmented Intelligence. Wie wir mit Daten und KI besser entscheiden.* Stuttgart: Reclam.
- Rettenberger, M., Eher, R., Desmarais, S. L., Hurducas, C., Arbach-Lucioni, K., Condemarin, C., Dean, K., Doyle, M., Folino, J. O., Godoy-Cervera, V., Grann, M., Ho, R. M. Y., Large, M. M., Pham, T. H., Nielsen, L. H., Rebocho, M. F., Reeves, K. A., de Ruiter, C., Seewald, K., & Singh, J. P. (2017): *Kriminalprognosen in der Praxis: Die Ergebnisse des International Risk Surveys (IRiS) aus Deutschland.* *Diagnostica*, 63, S. 2–14.
- Rostalski, F. & Völkening, M. (2019): *Smart Sentencing. Ein neuer Ansatz für Transparenz richterlicher Strafzumessungsentscheidungen.* *Kriminalpolitische Zeitschrift*, 265–273.
- Ruppert, F. (2021): *Strafzumessung am Scheideweg? Legal Tech und Strafzumessung.* *Kriminalpolitische Zeitschrift*, 90–98.
- Rudin, C., Wang, C. & Coker B. (2020): *The Age of Secrecy and Unfairness in Recidivism Prediction.* *Harvard Data Science Review* 2.1. Online: <https://hdsr.mitpress.mit.edu/pub/7z10o269/release/4>.
- Russell, S. & Norvig, P. (2021): *Artificial Intelligence: A Modern Approach.* 4. Aufl. Hoboken: Pearson.
- Schenke, W.-R., Graulich, K. & Ruthig, J. (2019): *Sicherheitsrecht des Bundes.* München: C. H. Beck.
- Schöch, H. (2015): *Prognosefall.* In: Kaiser, G., Schöch, H., Kinzig, J. (Hrsg.) *Kriminologie, Jugendstrafrecht, Strafvollzug.* München: C. H. Beck, S. 131–150.
- Skeem, J. & Lowenkamp, C. (2020): *Using Algorithms to Address Trade-Offs Inherent in Predicting Recidivism.* *Behavioral Sciences & the Law*, 38, 259–278.
- Skitka, L. J., Mosier, K. L. & Burdick, M. D. (1999): *Does Automation Bias Decision-Making?* *International Journal of Human-Computer Studies*, 51, 991–1006.
- Skitka, L. J., Mosier, K. & Burdick, M. D. (2000a): *Accountability and Automation Bias.* *International Journal of Human-Computer Studies*, 52, 701–717.
- Skitka, L. J., Mosier, K. L., Burdick, M. & Rosenblatt, B. (2000b): *Automation Bias and Errors: Are Crews Better than Individuals?* *International Journal of Aviation Psychology*, 10, 85–97.
- Sommerer, L. (2020): *Personenbezogenes Predictive Policing. Kriminalwissenschaftliche Untersuchung über die Automatisierung der Kriminalprognose.* Baden-Baden: Nomos.
- Tollenaar, N. & van der Heijden, P. G. M. (2013): *Which Method Predicts Recidivism Best? A Comparison of Statistical, Machine Learning and Data Mining Predictive Models.* *Journal of the Royal Statistical Society, Series A*, 176, 565–584.
- Wang, C., Han, B., Patel, B., Mohideen, F. & Rudin, C. (2020): *In Pursuit of Interpretable, Fair and Accurate Machine Learning for Criminal Recidivism Prediction.* Online: <https://arxiv.org/abs/2005.04176>.
- Wormith, J. S., Hogg, S. & Guzzo, L. (2012): *The Predictive Validity of a General Risk/Needs Assessment Inventory on Sexual Offender Recidivism and an Exploration of the Professional Override.* *Criminal Justice and Behavior*, 39, 1511–1538.
- Završnik, A. (2018): *Algorithmic crime control.* In: Završnik, A. (Hrsg.) *Big Data, Crime and Social Control.* New York: Routledge, S. 131–153.
- Zöller, M. A. (2002): *Informationssysteme und Vorfeldmaßnahmen von Polizei, Staatsanwaltschaft und Nachrichtendiensten: zur Vernetzung von Strafverfolgung und Kriminalitätsverhütung im Zeitalter von multimedialer Kommunikation und Persönlichkeitsschutz.* Heidelberg: C. F. Müller.

FELIX BUTZ, wiss. Mit.

Dr. LUCIA SOMMERER LL.M. (Yale), wiss. Mit.

Prof. Dr. KATRIN HÖFFLER

Lehrstuhl für Strafrecht und Kriminologie

Georg-August-Universität Göttingen

lehrstuhl.hoeffler@jura.uni-goettingen.de

für die Autorin Sommerer zudem:

Yale Law School Information Society Project

New Haven, CT, USA

Akad. Rat a.Z. Dr. STEPHAN CHRISTOPH

Professor Dr. JOHANNES KASPAR

Lehrstuhl für Strafrecht, Strafprozessrecht,

Kriminologie und Sanktionenrecht

Universität Augsburg

Prof. Dr. STEFAN HARRENDORF

Lehrstuhl für Kriminologie, Strafrecht, Straf-

prozessrecht und vergleichende Strafrechts-

wissenschaften, Universität Greifswald