

ALLGEMEINE BEITRÄGE

Digitalisierung als Chance für die Strafzumessung?

von Prof. Dr. Johannes Kaspar*

Abstract

Über die Schwächen und Probleme des Strafzumessungsrechts und der Strafzumessungspraxis wird schon seit längerer Zeit diskutiert, vor allem mit Blick auf die wenig präzisen rechtlichen Vorgaben sowie die regional unterschiedlichen Strafzumessungsniveaus. Auch der 72. Deutsche Juristentag 2018 in Leipzig hatte sich intensiv mit der Thematik beschäftigt und mehrheitlich (insgesamt eher moderate) Reformen gefordert, darunter die Einführung einer bundesweiten Strafzumessungsdatenbank. Umgesetzt wurde davon bislang nichts. Der Beitrag geht der Frage nach, ob der Einsatz von Datenbanken und anderer digitaler Technologien einschließlich des Einsatzes von Formen „künstlicher Intelligenz“ zur Lösung der genannten Probleme im Bereich der Strafzumessung beitragen könnte.

The deficiencies and problems of German sentencing law and sentencing practice have been discussed for a long time, with a special regard to quite vague legal ramifications and regional sentencing disparities. Participants of the 72nd German Convention of Legal Practitioners and Scholars (Deutscher Juristentag) in 2018 in Leipzig debated the topic intensely and came up with (all in all quite moderate) demands for legal reform. However, so far none of these demands has been adopted by the legislator. The article addresses the question whether the use of data bases or other digital technologies including the use of „artificial intelligence“ could contribute to solving the before mentioned problems in the area of sentencing.

I. Probleme im Bereich der Strafzumessung

Die Probleme im Bereich der Strafzumessung sind bekannt¹ und sollen daher nur noch einmal kurz in Erinnerung gerufen werden, um dann der Frage nachzugehen, ob die „Digitalisierung“ in ihren unterschiedlichen Erscheinungsformen hier vielleicht Abhilfe schaffen kann.

Wir haben es mit unklaren Maßstäben zu tun, die im Hinblick auf die Bestimmtheit und Gleichheit staatlichen Strafsens zumindest Fragen aufwerfen.² Die gesetzlichen

Strafrahmen sind sehr weit und bieten daher wenig Orientierung. Die gesetzlichen Regelungen zum Strafzumessungsvorgang, insbesondere § 46 StGB als Zentralnorm, sind eher vage gehalten. Die „Schuld“ soll nach Abs. 1 S. 1 die Grundlage der Strafzumessung sein, die „Wirkungen“ der Strafe auf den Täter sollen nach S. 2 (irgendwie) auch berücksichtigt werden. Absatz 2 enthält einen ausdrücklich nicht abschließenden Katalog recht heterogener Strafzumessungsfaktoren, die vom Gericht herangezogen werden können und die ohne Gewichtung oder Hierarchisierung aneinandergereiht werden. Vor allem aber fehlt es an gesetzlichen Regelungen für zwei praktisch besonders wichtige Fragen: Erstens das Problem der „Einstiegsstelle“ in den Strafrahmen, also die Frage, wie der konkrete Fall innerhalb des vom Strafrahmen eröffneten Spektrums in einem ersten Schritt grob eingeordnet werden kann.³ Zweitens fehlt es an Regelungen, wie eine angenommene Tat- und Schuldschwere (nach h.M. unter Berücksichtigung präventiver Belange) in ein konkretes Strafmaß „umgerechnet“ werden kann.

Vieles hängt hier also von den Wertungen des zur Entscheidung berufenen Gerichts ab, und so überrascht es auch nicht, dass man bundesweit deutliche Unterschiede feststellen kann, wenn es um die Bestrafung an sich vergleichbarer Delikte geht. Es ist empirisch belegt, dass persönliche Eigenschaften der Richterinnen und Richter (z.B. Strafzweckpräferenzen oder politische Einstellungen) einen Einfluss auf die Strafhöhe haben.⁴ Möglicherweise gilt dies, wie eine israelische Studie herausgefunden haben will, auch für hungrige Richter, die vor der Mittagspause zu härteren Entscheidungen neigen als danach.⁵ Noch größeren Einfluss haben aber vermutlich lokale und regionale Strafzumessungstraditionen: Das vor Ort übliche Sanktionsniveau wird durch den Rat der älteren und erfahrenen Kolleg:innen an die Neulinge weitergegeben und perpetuiert. Das ist menschlich gut nachvollziehbar und trägt zu einer gewissen Konsistenz und Gleichheit des Strafsens bei, aber eben nur lokal begrenzt. Hinzu kommen weitere Probleme, da der entscheidende „Ankerwert“⁶, der hier gesetzt wird, auf sehr schmaler empirischer Basis,

* Prof. Dr. Johannes Kaspar ist Inhaber des Lehrstuhls für Strafrecht, Strafprozessrecht, Kriminologie und Sanktionenrecht an der Universität Augsburg. Der Beitrag geht zurück auf einen Vortrag, den der Verfasser auf der Tagung des Kriminalpolitischen Kreises am 11.11.2022 in Würzburg gehalten hat. Die Vortragsform wurde weitgehend beibehalten.

¹ Zusammenfassend Kaspar, Gutachten C für den 72. Deutschen Juristentag, 2018; s. auch Kohn, Künstliche Intelligenz und Strafzumessung, 2020, S. 62 ff.

² Zum Gleichheitsproblem und Wegen zu dessen Behebung s. nur Kaspar, in: GS Tröndle, 2019, S. 279.

³ Dies wird von Schöch als Schlüsselproblem der Strafzumessung bezeichnet, s. Schöch, in: Göppinger/Hartmann (Hrsg.), Kriminologische Gegenwartsfragen, 1972, S. 128.

⁴ Nachweise bei Kaspar (Fn. 1), C 18.

⁵ Danziger et al., Proc. Natl Acad. Sci. USA, advance online publication doi:10.1073/pnas.1018033108 (June 2011).

⁶ Allgemein zu Anker-Effekten (etwa durch den Strafantrag der Staatsanwaltschaft) s. English/Mussweiler, Journal of Applied Social Psychology 2003, 1535; s. zu Anker-Effekten auch Kohn (Fn. 1), S. 118 ff.

ohne jede gesetzliche oder sonstige Regulierung und damit völlig intransparent zustande kommt. Und schließlich stellt sich die Frage, warum der Umstand, dass im betroffenen Gerichtsbezirk für vergleichbare Taten „üblicherweise“ vier Jahre Freiheitsstrafe verhängt werden, normative Relevanz haben sollte – lokale Tradition als solche ist kein legitimer Strafzumessungsfaktor. Das Ergebnis sind beträchtliche regionale Unterschiede, die in Studien ermittelt und zum Teil auch in grafischer Form aufbereitet werden – oft mit dem bekannten Nord-Süd-Gefälle, also einer härteren Strafpraxis in den südlichen Bundesländern.⁷

Dass diese Probleme existieren und Reformen nötig sind, war der von vielen geteilte Ausgangspunkt der Diskussion auf dem 72. Deutschen Juristentag in Leipzig 2018. Ich war damals im Vorfeld mit der Abfassung des entsprechenden Gutachtens beauftragt worden.⁸ Viele meiner Vorschläge, etwa die Einführung einer Strafzumessungskommission oder die Einengung und v.a. Absenkung von Strafrahmen, wurden damals von der Mehrheit der Abstimmenden, darunter vor allem Angehörige der Gerichte und der Staatsanwaltschaften, abgelehnt. Komplexere Formen „künstlicher Intelligenz“ und deren Einsatz im Rahmen der Strafzumessung waren in Leipzig noch nicht Gegenstand vertiefter Diskussionen. Was eine Mehrheit fand, war die von mir vorgeschlagene Einführung einer bundesweiten Strafzumessungsdatenbank, die die Suche nach „vergleichbaren“ Urteilen und einer „üblichen Sanktionspraxis“ auf eine transparente und breitere empirische Basis stellen würde (und damit nebenbei bemerkt auch die revisionsrichterliche Kontrolle von Strafzumessungsentscheidungen erleichtern würde). Konkrete politische Pläne zur Umsetzung dieses Vorhabens wurden bislang nicht entwickelt, aber immerhin enthält der aktuelle Koalitionsvertrag (leider ohne Angabe der Hintergründe) auf S. 85 folgenden Passus: „*Gerichtsentcheidungen sollen grundsätzlich in anonymisierter Form in einer Datenbank öffentlich und maschinenlesbar verfügbar sein*“. Das könnte also in die angesprochene Richtung gehen – eine Datenbank als einfache Orientierungshilfe für Gerichte – vielleicht aber auch als Grundlage für komplexere, von Computersystemen eigenständig durchgeführte Auswertungen, immerhin soll das Ganze „maschinenlesbar“ sein. Insofern stellt sich aktuell durchaus die Frage, ob die „Digitalisierung“ der Strafzumessung, einschließlich des Einsatzes von Formen „künstlicher Intelligenz“, der Schlüssel zur Lösung der oben nur grob skizzierten Defizite im Bereich der Strafzumessung sein könnte.⁹

II. Begriffsklärungen

Zunächst sind jedoch Begriffsklärungen nötig. Mit „Digitalisierung“ ist auf einer technischen Ebene der Umstand gemeint, dass Informationen in eine Folge von Einsen und

Nullen überführt werden, sodass sie von einem Computersystem verarbeitet werden können.¹⁰ Teilweise wird mit dieser Bezeichnung auch die „Gesamtheit grundsätzlicher Veränderungen von Situationen und Handlungen in unserem privaten und beruflichen Alltag“ beschrieben, die sich durch die Nutzung digitaler Technologien ergibt.¹¹ Beide Perspektiven sind hilfreich zum Verständnis von Digitalisierung und den Umwälzungen (manche sprechen von „Disruptionen“), die sie schon gebracht hat und noch bringen wird.

Auch das Rechtswesen ist von dieser Entwicklung natürlich nicht ausgenommen. Das Schlagwort des „Legal Tech“, mit dem der Einsatz von Informationstechnologie in juristischen Kontexten umschrieben wird, ist in aller Munde.¹² Man unterscheidet dabei grob zwischen verschiedenen denkbaren Entwicklungsstufen, die die Intensität der Einbindung von digitalen Technologien wieder spiegeln: Mit Legal Tech 1.0 ist die noch ganz einfache Einholung von Informationen etwa durch Benutzung juristischer Datenbanken wie beck-online oder Juris gemeint, die heute schon zum selbstverständlichen juristischen Tagesgeschäft gehört; Legal Tech 2.0 bedeutet, dass nicht nur die Informationssuche im Vorfeld betroffen ist, sondern auch Teile der eigentlichen juristischen Arbeit (z.B. Identifikation juristischer Probleme; Subsumtion und/oder Ermittlung von Rechtsfolgen) vom System in automatisierter Form vorgenommen werden; von Legal Tech 3.0 würde man erst sprechen, wenn das System eigenständig juristische Aufgaben übernimmt und selbst rechtlich verbindliche Ergebnisse produziert.

Sobald man in den Bereich von komplexeren Aufgaben kommt, die vom System vorbereitet oder sogar übernommen werden, wird oft von „künstlicher Intelligenz“ (KI) gesprochen, die als „Legal-Tech-Kerntechnologie“¹³ bezeichnet wird. Der Begriff der KI ist bislang noch wenig konturiert¹⁴ – schon deshalb, weil schon der Begriff der menschlichen Intelligenz vielschichtig ist und uneinheitlich definiert wird.¹⁵ Im Kern geht es darum, dass Computersysteme Informationen erfassen, verarbeiten und auf dieser Grundlage in der Lage sind, Ergebnisse zu produzieren bzw. Entscheidungen zu treffen; auch hier sind die Erscheinungsformen und Entwicklungsstufen vielfältig. Wichtige Anwendungsfelder sind sog. regelbasierte Expertensysteme, bei denen das System anhand der eingegebenen Daten („Input“) auf der Basis vorab festgelegter Entscheidungsregeln bzw. „Algorithmen“ ein Ergebnis („Output“) produziert.

Anders ist dies bei der zweiten wichtigen Fallgruppe, die in der Diskussion oft mit künstlicher Intelligenz gleichgesetzt wird (man könnte hier von KI i.e.S. sprechen), namentlich dem Maschinellen Lernen. Hier entscheidet die Maschine anhand von Regeln, die sie selbst erst aus den

⁷ S. nur *Grundies*, in: Hermann/Pöge (Hrsg.), *Kriminalsoziologie*, 2018, S. 295; zusammenfassend auch *Kohn* (Fn. 1), S. 97 ff.

⁸ S. *Kaspar* (Fn. 1).

⁹ Zu entsprechenden Ansätzen eines „Smart Sentencing“ s. *Rostalski/Völkening*, *KriPoZ* 2019, 265.

¹⁰ *Hilgendorf*, in: *Kasprowicz/Rieger* (Hrsg.), *Handbuch Virtualität*, 2020, S. 406.

¹¹ S. *Staab*, in: *Staab/Stahl*, *Don't worry, be digital*, 2019.

¹² Zum Folgenden s. *Kaspar/Harrendorf/Höffler*, *NK* 2020, 35.

¹³ *Herberger*, *NJW* 2018, 2325.

¹⁴ Vgl. nur *Herberger*, *NJW* 2018, 2325; *Kohn* (Fn. 1), S. 26 ff. (auch zur Entstehungsgeschichte des Begriffs).

¹⁵ S. nur *Kment/Borchert*, *Künstliche Intelligenz und Algorithmen in der Rechtsanwendung*, 2022, S. 6 ff.

eingeebenen Daten extrahiert, den sog. „Trainingsdaten“. Das Besondere ist, dass diese Entscheidungsregeln im Laufe der Zeit vom System angepasst und verfeinert werden, um die Ergebnisqualität weiter zu verbessern, ohne dass dies notwendigerweise vom menschlichen Beobachter bzw. vom ursprünglichen Programmierer nachvollzogen werden – vor allem hier stellt sich dann das bekannte Problem der Intransparenz bzw. der KI als „black box“.

Auf den Bereich der Strafzumessung bezogen wäre es also denkbar, dass sich die verantwortlichen Richter:innen im Rahmen der Strafzumessungsentscheidung entweder über relevante einschlägige Urteile z.B. auf Juris informieren (1.0), sich von einem „Decision-Support-System“ einen datenbasierten Strafzumessungsvorschlag für den konkreten Fall anzeigen lassen (2.0) oder frühzeitig in die Kaffeepause gehen, weil die Strafzumessungsentscheidung auf der Basis der eingegebenen Daten vom System selbst festgelegt wird (3.0) – wobei letzteres Modell, also eine vollständige Ersetzung des menschlichen Entscheiders im Bereich der Strafzumessung offensichtlich mit besonderen technischen wie rechtlichen Problemen behaftet wäre, die eine Verwirklichung in nächster Zeit unrealistisch erscheinen lassen. Daher werde ich mich in der Folge auf die ersten beiden Dimensionen konzentrieren, die jeweils eine Unterstützung der menschlichen Entscheidung beinhalten und nicht das utopische (und für viele auch dystopische¹⁶) Szenario des verbindlich entscheidenden Roboter-Richters.

III. Beispiele

Um dies etwas zu veranschaulichen, möchte ich kurz auf einige Praxisbeispiele eingehen, bei denen digitale Hilfsmittel, wie ich es einmal neutral formulieren will, bei der Strafzumessung bereits jetzt eine prominente Rolle spielen. Es ist vermutlich kein Zufall, dass sich die meisten Anwendungsbeispiele bislang eher im außereuropäischen Bereich finden lassen, v.a. in Asien und den USA.

1. Strafzumessungsdatenbank in Japan

2009 wurde in Japan eine Strafzumessungsdatenbank eingeführt, die allerdings einen nur sehr begrenzten Anwendungsbereich hat, nämlich Schwerdelinquenz, bei der die Todesstrafe im Raum steht und bei der eines der ebenfalls 2009 eingeführten Gerichte mit Beteiligung von Laienrichtern (sog. Saiban'in) zuständig ist.¹⁷ Der Hintergrund war, dass man im Vorfeld der Einführung der Beteiligung von Laien fürchtete, dass diese das zuvor in Japan erstaunlich homogene Strafzumessungsniveau durcheinander bringen könnten. Durch die Strafzumessungsdatenbank sollte dieses Niveau in vergleichbaren Fällen mit Computer-Unterstützung angezeigt werden und dann zur Grundlage der Strafzumessungsdiskussion zwischen Richter:innen und Laien werden. Die Datenbank ist nicht online und nicht für alle Interessierten zugänglich, sondern nur für

autorisierte Personen aus den Bereichen Strafjustiz und Strafverteidigung an Terminals im Gebäude des japanischen Obersten Gerichtshofs. Transparenz und ein öffentlicher Diskurs in Bezug auf das vorherrschende Strafzumessungsniveau waren offensichtlich nicht das Ziel der Reform.

Ich habe mir die Funktionsweise in Japan vor Ort angesehen und erklären lassen – man gibt bestimmte Merkmale des Falles (z.B. Deliktstyp, Schadenshöhe, etc.) in das System ein (was einigermaßen benutzerfreundlich über Dropdown-Menüs erfolgt) und bekommt dann das bisherige Strafniveau in diesen Fällen grafisch aufbereitet angezeigt. Die entsprechende Software wurde von einem privaten Anbieter in Kooperation mit der Verwaltung des *Obersten Gerichtshofs* entwickelt. Eine gesetzliche Grundlage gibt es meines Wissens nicht. Die Hinzuziehung der Datenbank ist für die Gerichte nicht verpflichtend, erst recht sind die dort angezeigten Durchschnittswerte in keiner Weise bindend. Faktisch haben sie aber offenbar ihren Zweck erfüllt: Wie im Vorfeld erhofft, hat die Laienbeteiligung insgesamt zu keinen großen Umwälzungen beim Strafzumessungsniveau geführt, wobei sich allerdings gewisse deliktsspezifische Verschiebungen ergeben haben:¹⁸ Mitleidstötungen sowie Brandstiftungsdelikte wurden durchschnittlich etwas milder beurteilt, sexuelle Gewalt und Tötungsdelikte ohne Mitleidsmotiv etwas härter. Defizite in Bezug auf Transparenz und die fehlende gesetzliche Grundlage sind aber nicht zu bestreiten. Immerhin erfüllt die Datenbank aber offenbar den Zweck, der hierzulande vermutlich oft im Rahmen eines informellen Gesprächs in der Behörden-Kaffeerunde verfolgt wird: Herauszufinden, welches Strafmaß in vergleichbaren Fällen denn in etwa „üblich“ ist. Nur erfolgt dies in Japan zumindest in den Saiban'in-Verfahren auf einer breiten empirischen Basis, die sich auf die Praxis im gesamten Land bezieht. Im Hinblick auf die Gleichheit, Konsistenz und Vorhersehbarkeit staatlichen Strafens ist das ein Vorteil gegenüber der Situation hierzulande.

2. Sentencing-E-System in Südkorea

In Südkorea existiert seit 15 Jahren eine Strafzumessungskommission.¹⁹ Es handelt sich um ein interdisziplinär besetztes Gremium, das dem Zweig der Judikative zugeordnet ist, aber unabhängig agiert. Es hat nach und nach deliktsspezifische Sentencing Guidelines entworfen, die den Strafzumessungsprozess der Gerichte anleiten sollen. Abweichungen sind möglich, allerdings enthalten die Guidelines selbst gewisse Spielräume; auch wurden sie auf der Basis der bisherigen Strafzumessungspraxis entworfen, so dass es nicht überrascht, dass die „Befolgungsrate“ seitens der Gerichte mit ca. 90 % recht hoch ist. Interessant ist, dass die Einführung der Kommission und der Guidelines ausdrücklich auch das Ziel verfolgte, das Vertrauen der Bevölkerung in die Strafjustiz zu stärken. Daher wurde bei der Entwicklung der Guidelines nicht nur das bisherige Strafniveau zugrunde gelegt, sondern ergänzend auch

¹⁶ Vgl. nur *Greco*, RW 2020, 29.

¹⁷ S. näher *Schmidt*, Das japanische Saiban'in System und das deutsche Schöffensystem, 2020.

¹⁸ S. dazu näher *Schmidt* (Fn. 17).

¹⁹ Zum Folgenden s. *Choi*, in: Korean Sentencing Commission (ed.), Reasonable Sentencing: Current Trends and Future Challenges, 2022, S. 59.

die diesbezügliche öffentliche Meinung (im Wege von „public hearings“) erhoben und mitberücksichtigt. Mit Blick auf die „Digitalisierung“ der Strafzumessung ist die Entwicklung in Südkorea vor allem auch deshalb von Interesse, weil mittlerweile den Gerichten ein sogenanntes „Sentencing-E-System“ zur Verfügung steht, eine Software also, die die Anwendung der Guidelines mit Computerunterstützung erleichtern soll. Einen inhaltlich eigenständigen „Strafzumessungsvorschlag“ macht das System also, soweit ich es verstehe, nicht.

3. Decision-Support-System Shanghai

Das scheint in China anders zu sein. Aus Shanghai wird berichtet, dass ein Decision-Support-System entwickelt wurde und auch in der Praxis bereits eingesetzt wird, das sog. Shanghai AI Assistive System on Criminal Cases.²⁰ Es wurde nach Angaben der Verantwortlichen seit 2017 unter Mitarbeit von Justiz-Praktiker:innen und IT-Expert:innen entwickelt. Im Frühjahr 2019 bestand es danach aus 9 Datenbanken mit 45 Millionen digitalisierten Dokumenten. Eine von insgesamt 26 Funktionen des Systems sei das erwähnte KI-basierte Strafzumessungs-Unterstützungs-System, das auf früheren Urteilen als „Trainingsdaten“ beruhe. Über die genaue Funktionsweise des Systems ist bisher leider wenig bekannt; da es jedenfalls in englischer Sprache kaum unabhängige wissenschaftliche Publikationen gibt, ist es momentan schwer, sich ein Bild vom System und seinen Auswirkungen auf die Strafzumessungspraxis zu machen. Interessant ist jedenfalls, dass auch in China betont wird, dass man mit der Entwicklung des Systems nicht etwa die Gerichte ersetzen, sondern in ihrer Entscheidung nur unterstützen wolle.²¹

4. COMPAS (USA)

Zum Schluss möchte ich noch ein Beispiel aus den USA darstellen, das einen Teilbereich der Strafzumessung betrifft, nämlich die Prognose zukünftiger Straftaten. Gerade hier sind „Automatisierung“ und der Einsatz von digitalen Technologien bis hin zum Einsatz von Formen künstlicher Intelligenz weit fortgeschritten.²² Vor allem in den USA werden zunehmend sog. Risk Assessment Tools eingesetzt. Das berühmteste (und zugleich sehr umstrittene) Beispiel ist COMPAS, ein Akronym für „Correctional Offender Management Profiling for Alternative Sanctions“. Das System wurde von einem privaten Unternehmen entwickelt und wird bis heute durch dieses Unternehmen (früher: Northpointe, jetzt Equivant) kommerziell vertrieben. Es basiert auf dem anerkannten und empirisch fundierten „Risk-Need-Responsivity“-Ansatz und dient der Einschätzung der Rückfallgefahr (in Bezug auf allgemeine Delinquenz und Gewaltdelinquenz). Ursprünglich war es mit Blick auf Bewährungsentscheidungen entwickelt worden, wurde dann aber zunehmend auch für die

Erstellung von Prognosen als Teil der Strafzumessungsentscheidung herangezogen. Bei COMPAS geht es nicht um „Machine Learning“; es findet kein fortlaufender „Trainingsprozess“ mit Ergebniskontrolle statt. Vielmehr handelt es sich um ein statisches regelbasiertes Expertensystem, bei dem insgesamt 137 Variablen in das System eingegeben werden, die dann zu einer Einschätzung des Rückfallrisikos (durch Ermittlung eines Risikoscores) führen.

Gegen COMPAS werden einige Kritikpunkte erhoben, die generell gegen den Einsatz von künstlicher Intelligenz im Bereich der Prognose von Straftaten und Strafzumessung eingewandt werden:²³ Es fehle an Transparenz, weil die dem System zugrundeliegenden Algorithmen vom Unternehmen (das sich auf den Schutz von Betriebsgeheimnissen beruft) nicht offen gelegt werden. Das System sei unfair und diskriminiere Minderheiten, bei denen die Quote der „false positives“, also der zu Unrecht als gefährlich qualifizierten Personen, besonders hoch sei.²⁴ Und schließlich wird die überlegene Treffsicherheit des Systems bezweifelt: Die Quote von ca. 65 % richtiger Vorhersagen würden auch juristische Laien erzielen. Trotz dieser Kritik (die nicht unwidersprochen geblieben ist²⁵), hat der *Supreme Court von Wisconsin* in der Entscheidung *Loomis vs. State of Wisconsin* 2016 entschieden, dass der Gebrauch des Systems im Rahmen der Strafzumessungsentscheidung zulässig sei.²⁶ Das Gericht hat allerdings Einschränkungen formuliert: Der Risikoscore dürfe nicht die einzige Basis der Entscheidung sein und dürfe auch nur für Prognoseentscheidungen, aber nicht im Rahmen der Festlegung der Strafhöhe (also der Strafzumessung im engen Sinne) benutzt werden; auch müsse den Gerichten vom Hersteller eine Art „Warnhinweis“ erteilt werden, in dem auf die Schwächen und Limitierungen des Systems hingewiesen wird.²⁷ Ob die Probleme damit gelöst sind, ist natürlich fraglich, wie sich ohnehin die Frage stellt, wie die Gerichte mit solchen technischen Hilfsmitteln hierzulande umgehen würden: Denkbar sind viele Varianten, von der vollständigen Ablehnung bis hin zur unkritischen Übernahme der Ergebnisse im Sinne eines „Automation Bias“. Hier besteht noch großer Forschungsbedarf.

IV. Digitalisierung als Lösung?

Abschließend soll nun der Frage nachgegangen werden, ob die Digitalisierung eine Chance für die Strafzumessung darstellt, ob sie insbesondere eine Lösung der oben skizzierten Probleme beinhaltet oder ob die nicht zu leugnenden Bedenken und auch von Vielen schon artikulierten Einwände überwiegen. Dabei stehen wir natürlich vor dem Problem, dass eine kritische Diskussion des Einsatzes von digitalen Technologien und KI oft eine Art „Schattenboxen“ darstellt, weil wir ja noch gar nicht wissen, wie bestimmte Tools und Systeme in Zukunft aussehen werden.

²⁰ Dazu *Kohn* (Fn. 1), S. 152 ff.

²¹ S. etwa *Jiang*, in: Beck u.a. (Hrsg.), *Digitalisierung, Automatisierung, KI und Recht*, 2020, S. 557 ff.; weitere Nachweise bei *Kohn* (Fn. 1), S. 157.

²² S. nur *Butz u.a.*, *BewHi* 2021, 241.

²³ S. zum Folgenden *Kohn* (Fn. 1), S. 137 ff.

²⁴ *Angwin et al.*, *Machine Bias*, *ProPublica* (May 23, 2016).

²⁵ Zum Vorwurf der Diskriminierung s. nur *Rudin u.a.*, *The Age of Secrecy and Unfairness in Recidivism Prediction*, *Harvard Data Science Review* 2020, 2.1.

²⁶ *State vs. Loomis* 881 N.W.2d 749 (2016); s. dazu *Kohn* (Fn. 1), S. 145 ff.; s. auch *Jiang* (Fn. 21).

²⁷ *Jiang* (Fn. 21), S. 557, 567.

Die von mir genannten Beispiele sind jedenfalls sicher noch nicht das Ende der Entwicklung.

Dennoch wäre es natürlich ein Fehler, die weitere technologische Entwicklung erst abzuwarten. Es ist wichtig, dass wir uns als Jurist:innen aktiv und frühzeitig an der Diskussion beteiligen und das Feld nicht den IT-Fachleuten überlassen – allein dass etwas technisch machbar ist, heißt natürlich auch in diesem Fall nicht, dass es wünschenswert ist. Vielmehr gilt es, den technologischen Entwicklungsprozess einschließlich zumindest denkbarer Zukunftsszenarien von vornherein kritisch zu begleiten.²⁸

1. Mögliche Vorteile

Zunächst also zu den möglichen Chancen und Vorteilen, die die Digitalisierung bieten könnte.

a) Auf der Hand liegt der Vorteil, dass digitale Technologien zu einer schnellen und effizienten Erledigung von Strafverfahren beitragen können²⁹ – natürlich nicht nur im Bereich der Strafzumessung, sondern auch im Vorfeld, etwa bei der Frage der Einstellung von Verfahren gem. §§ 153 ff. StPO. Das würde dem Beschleunigungsgebot Rechnung tragen.

b) Ein weiterer Vorzug könnte darin liegen, dass digitale Technologien bestimmte menschliche Schwächen (z.B. Vorurteile, Wahrnehmungsfehler und „biases“) nicht aufweisen und damit die Strafzumessungspraxis insgesamt fairer und gleichmäßiger werden könnte.

c) Kommen Algorithmen in der Strafzumessung zum Einsatz, könnte dies auch zu mehr Bestimmtheit und Vorhersehbarkeit der Strafzumessung beitragen (was natürlich eine gewisse Transparenz der Systeme voraussetzt, dazu sogleich).

d) Und schließlich wäre die Hoffnung, dass digitale Technologien zu „besseren“ Ergebnissen führen, wobei sich hier gleich eine wichtige Differenzierung auftut: Wenn es um die Treffsicherheit von Prognosen geht, geht es um eine empirische Frage (Rückfälligkeit ja/nein). Eine Entscheidung kann sich hier zumindest prinzipiell im Nachhinein als „richtig“ oder „falsch“ herausstellen. Damit besteht hier auch die Möglichkeit, ein System im Sinne des „Machine Learning“ zu „trainieren“, indem man ihm nach jeder Prognose ein Feedback gibt, ob sich diese so bewahrheitet hat oder nicht. Eine höhere Treffsicherheit bedeutet dann eine „Verbesserung“ der Prognose-Ergebnisse. Die meisten Studien deuten darauf hin, dass dies durch den Einsatz von digital unterstützten Risk-Assessment-Tools möglich ist. Es ist daher zu erwarten, dass solche Systeme auch in Deutschland zunehmend zum Einsatz kommen werden.

Bei normativ geprägten Wert-Entscheidungen wie der „Schuldangemessenheit“ einer Strafe als Kernfrage der Strafzumessung im engen Sinn ist das ganz anders (was bislang selten thematisiert wird). Das Fehlen objektiver Maßstäbe und das Fehlen der Möglichkeit von empirischer Überprüfung bzw. Falsifikation führt hier dazu, dass man eine Strafzumessungsentscheidung als solche kaum als schlicht „richtig“ oder „falsch“ qualifizieren kann. Das hat Konsequenzen: Machine Learning in dem Sinne, dass man dem Computer laufend mitteilt, ob seine Entscheidung „falsch“ oder „richtig“ war, funktioniert hier nicht, solange man keinen Maßstab fixiert hat, anhand dessen diese Bewertung erfolgen soll. Und nur dann lässt sich auch beurteilen, ob der Einsatz von Computersystemen zu einer „Verbesserung“ der Strafzumessung führt.

Als ein solcher (empirisch messbarer) Maßstab, den man auch in Datenform aufbereiten und dem System „füttern“ könnte, kommt insbesondere die bisherige Strafpraxis in Betracht. Das setzt allerdings voraus, dass man sich der Idee einer „komparativen Strafzumessung“³⁰ annähert und die bisherige Strafzumessungspraxis überhaupt als legitimen Orientierungswert akzeptiert. Das erscheint beides gut vertretbar, ist aber voraussetzungsreich, wie sogleich noch etwas vertieft wird. Als weiterer (ergänzender) Richtwert könnten empirisch ermittelte Strafvorstellungen der Bevölkerung³¹ in digital aufbereiteter Form fungieren – aber auch hier sind natürlich noch viele Fragen offen, nicht nur, wer in welcher Form zu diesen Vorstellungen befragt werden sollte, sondern auch (erneut), ob ein solches empirisches Faktum überhaupt normative Relevanz haben sollte.³² Vielleicht ließen sich beide Aspekte zusammenführen im Sinne wechselseitiger „checks and balances“: Jedenfalls dort, wo die Bevölkerung in ihren durchschnittlichen Straferwartungen unterhalb des Niveaus der Justizpraxis bleibt (und solche Fälle gibt es), spricht nichts gegen eine Anpassung „nach unten“³³ – natürlich alles nur im Rahmen von Strafzumessungsempfehlungen, die die Zuständigkeit und auch Verantwortung des Richters für die Entscheidung nicht beseitigen.

2. Einwände und Probleme

Zum Schluss sollen noch in aller Kürze einige Einwände und Probleme angesprochen werden.

a) Individuelle Schuldangemessenheit

Zum Teil werden ganz grundlegende Bedenken gegen den Einsatz insbesondere von Formen der künstlichen Intelligenz, aber auch von einfachen Datenbanksystemen im Bereich der Strafzumessung geltend gemacht. *Ruppert* etwa wendet ein, dass die damit vollzogene Hinwendung zu einer komparativen Strafzumessung nicht mit § 46 Abs. 1

²⁸ So auch bereits *Sommerer*, Personenbezogenes Predictive Policing, 2018.

²⁹ S. nur *Aumann*, Artificial Intelligence vs. Human in the Legal Profession, *The Legal Revolutionary* 2018, S. 74 ff.

³⁰ S. dazu nur *Maurer*, Komparative Strafzumessung, 2005.

³¹ Vgl. zu den methodischen Problemen in diesem Bereich nur *Christoph*, *MschKrim* 2020, 58; sehr kritisch hierzu *Köbel/Singelstein*, *NSStZ* 2020, 333.

³² S. zum Ganzen die Beiträge in *Kaspar/Walter* (Hrsg.), *Strafen im Namen des Volkes?*, 2019.

³³ Näher dazu *Kaspar*, in: *Kaspar/Walter* (Fn. 32), S. 61.

StGB und der dort kodifizierten individuellen Schuldangemessenheit der Strafe kompatibel sei.³⁴ Dafür sprechen auch Entscheidungen des *BGH*, wonach der Tatrichter die schuldangemessene Strafe stets mit Blick auf den konkreten Fall festzulegen hat und sich dabei nicht entscheidend auf die in anderen, womöglich ganz ähnlichen Konstellationen verhängte Strafe stützen darf.³⁵ Auf einer theoretischen (um nicht zu sagen: idealistischen) Ebene sind das nachvollziehbare Bedenken, die aber der Realität der Strafzumessungspraxis nicht gerecht werden. Von dort wird berichtet, dass man – nachvollziehbarerweise – selbst nach Richt- und Ankerwerten sucht, indem man durch eigene Recherchen, inklusive des Gesprächs im Kreis der Kolleginnen und Kollegen, vergleichbare Fälle und das dort übliche Strafniveau ermittelt – also das, was *Dreher* als „geheimes Metermaß“³⁶ der Gerichte bezeichnet hat.

Auch mit der ausdifferenzierten individuellen Schuldbeurteilung ist es in Wahrheit nicht ganz so weit her, wie es die Kritiker offenbar annehmen. Natürlich verantwortet das Gericht die von ihm verhängte Strafe als im konkreten Fall „schuldangemessen“. Die empirische Forschung zeigt aber, dass die Feindifferenzierung an Grenzen stößt: Erstens sind es nur wenige Faktoren, die sehr starken Einfluss auf die Strafhöhe haben, vor allem die Tatschwere, die Vorstrafenbelastung und das Vorliegen eines Geständnisses.³⁷ Zweitens gibt es das bekannte, von *Rolinski* so bezeichnete Phänomen der „Prägnanztendenz“, also die Tatsache, dass Gerichte dazu tendieren, „glatte“ Strafen in bestimmten Abstufungen zu verhängen, z.B. 3 Jahre und 3 Monate Freiheitsstrafe, 3 Jahre und 6 Monate Freiheitsstrafe usw.³⁸ Man geht also mit anderen Worten ohnehin schon (vielleicht auch notgedrungen) in gewisser Weise schematisch vor. Vor diesem Hintergrund spricht m.E. nichts dagegen, dass man Gerichten Informationen über ein empirisch tatsächlich so ermitteltes „übliches“ Strafniveau an die Hand gibt, weil sie sich diese Information ohnehin auf dem Weg zur Findung der schuldangemessenen Strafe besorgen werden, bislang allerdings eben auf sehr intransparente und selektive Art und Weise. Dass damit das „komparative“ Element der Strafzumessung mehr Gewicht bekommt, ist vor diesem Hintergrund unschädlich, zumal der *BGH* selbst in jüngerer Zeit dazu tendiert, Strafzumessungsentscheidungen in Einzelfällen als rechtsfehlerhaft zu kassieren, weil sie zu weit vom in vergleichbaren Fällen üblichen Strafmaß entfernt sind.³⁹ Wenn das der Maßstab ist, kann eine Datenbank bzw. ein Computersystem, das dieses übliche Strafmaß in methodisch belastbarer Weise darstellt, auch für die revisionsrechtliche Kontrolle eigentlich nur von Vorteil sein.

b) Richterliche Verantwortung

Ein weiterer Kritikpunkt betrifft den Punkt der richterlichen Verantwortung. *Greco* warnt in einem 2021 erschienenen Beitrag eindringlich vor dem „Roboter-Richter“,

nachdem er zuvor gezeigt hat, dass die technischen Möglichkeiten KI-gestützter Strafzumessung bereits jetzt schon weit fortgeschritten sind und es weniger um das Problem von deren „Machbarkeit“, sondern mehr um deren Legitimation geht⁴⁰. Nach *Greco* muss die Strafzumessung in der Verantwortung eines Menschen aus Fleisch und Blut bleiben, der Empathie mit dem Verurteilten empfinden kann. Was zunächst nur gegen die vollständige Ersetzung des Richters durch KI spricht, wird von *Greco* dann aber auch gegen Unterstützungssysteme eingewandt: Auch letztere würden, im Sinne des oben erwähnten „Automation Bias“, einen zu starken Einfluss auf die freie (menschliche) richterliche Entscheidung nehmen. Letzteres ist allerdings eine offene Frage. Auch ist die Vorstellung des empathischen, mit dem Schicksal des Verurteilten quasi mitleidenden Richters vielleicht etwas zu einseitig – auch der zynische Richter ist denkbar, der dem vor Gericht unsympathisch und renitent auftretenden Angeklagten (bewusst oder unbewusst) eine etwas härtere Strafe auferlegt – in dieser Konstellation wäre der nüchtern Daten analysierende Computer aus Sicht des Verurteilten die bessere Wahl gewesen.

c) Fairness und Transparenz

Mangelnde Fairness und Diskriminierung von Minderheiten sind Vorwürfe, die generell gegen den Einsatz von KI erhoben werden. Wie oben erwähnt, werden sie auch gegen das System COMPAS vorgebracht. Entscheidend ist, welcher „Input“ dem System zugeführt wird – „garbage in, garbage out“ lautet ein geflügeltes Wort aus dem IT-Bereich. Enthalten die Daten, mit denen das System trainiert wird, diskriminierende Elemente, wird auch der zukünftige Output des Computersystems möglicherweise diskriminierend sein. Beispielhaft und auf die Strafzumessungspraxis bezogen: Wäre es so, dass Richter ausländischen Angeklagten gegenüber systematisch allein aufgrund ihrer Herkunft höhere Strafen verhängen würden, bestünde die Gefahr, dass diese Diskriminierung vom System als Muster und (legitime) Entscheidungsregel erkannt und fortgeführt wird. Entscheidend ist also, dass man auf die Qualität und „Diskriminierungsfreiheit“ der Trainingsdaten achtet. Dies hängt eng mit der Forderung nach Transparenz zusammen, die zunehmend betont wird. Die Funktionsweise des Systems einschließlich der zugrunde liegenden Algorithmen soll möglichst nachvollzogen werden können – man spricht von dem Ziel der „explainable AI“ (Artificial Intelligence), an dem zunehmend entwickelt und geforscht wird.⁴¹ Transparenz und die Möglichkeit der menschlichen Kontrolle sind konsequenterweise auch zwei wichtige Aspekte, die im aktuellen Entwurf einer KI-Verordnung der EU-Kommission in Art. 13 und Art. 14 hervorgehoben werden.

d) Akzeptanz in der Bevölkerung

Schließlich soll noch die Akzeptanz in der Bevölkerung

³⁴ *Ruppert*, KriPoZ 2021, 90.

³⁵ S. etwa *BGHSt* 56, 262.

³⁶ *Dreher*, MDR 1961, 343 (344).

³⁷ Nachweise bei *Kaspar* (Fn. 1), C 88.

³⁸ *Rolinski*, Die Prägnanztendenz im Strafurteil, 1969.

³⁹ Nachweise bei *Kaspar* (Fn. 1), C 101.

⁴⁰ *Greco*, RW 2020, 29.

⁴¹ S. nur *Wang u.a.* in: Rudin, In Pursuit of Interpretable, Fair and Accurate Machine Learning for Criminal Recidivism Prediction, online abrufbar unter: <https://arxiv.org/abs/2005.04176> (zuletzt abgerufen am 6.1.2023).

erwähnt werden. Ein Rechtssystem sollte sich zumindest darum bemühen, mit Maßstäben zu operieren, die von der Mehrheit der Bevölkerung grundsätzlich als gerecht und legitim anerkannt werden. Fehlt es daran, leidet das Vertrauen in die Rechtsordnung und (so legen empirische Studien nahe) auch die eigene Normbefolgungsbereitschaft.⁴² Insofern ist es durchaus von Interesse, ob die Bevölkerung denn mit einer digitalisierten Strafzumessung bis hin zur KI-Unterstützung einverstanden wäre. Was einfache Datenbank-Systeme angeht hätte ich insofern keine Bedenken, auch wenn empirische Studien dazu soweit ersichtlich fehlen. Was von der Bevölkerung (bislang jedenfalls) sehr deutlich abgelehnt wird ist der Einsatz von Computern bei der Prognose von zukünftiger Delinquenz: In einer Studie der Bertelsmann-Stiftung aus dem Jahr 2018 gaben nur 2 % der Befragten an, dass sie diese Aufgabe allein einem Computersystem überantworten würden; nur 34 % meinten, dass der Computer dabei unterstützend eingesetzt werden sollte; 54 % lehnten selbst dies ab und meinten, dass diese Aufgabe allein von einem Menschen ausgeführt werden sollte.⁴³ Was sich hier zeigt ist eine recht stark ausgeprägte „Algorithm Aversion“, die sich vermutlich auch gegenüber der Vorstellung eines KI-gestützten Decision-Support-Systems im Bereich der Strafzumessung ergeben würde. Nun sind diese Zahlen nur eine Momentaufnahme; es wird sich zeigen, ob man mit zunehmendem Wissen über den Einsatz von KI auch offener und aufgeschlossener gegenüber dieser Technologie sein wird. Transparenz und menschliche Kontrolle sind dabei sicher Aspekte, die das Potenzial haben, das Vertrauen in KI-Systeme zu stärken.⁴⁴

V. Fazit

Ein kurzes, vorsichtig optimistisches Fazit:

1. Die Digitalisierung hat das Potenzial, zu einer gleichmäßigeren, besseren und gerechteren Strafzumessung beizutragen. Allerdings hängt das sehr vom jeweiligen Einsatzgebiet und der Ausgestaltung des jeweiligen Systems ab.

2. Das betrifft den Bereich der Prognose, wo sich möglicherweise abzeichnet, dass Risk-Assessment-Tools, die zunehmend auch mit KI-Unterstützung eingesetzt werden, bessere Ergebnisse produzieren als herkömmliche Verfahren. Die weitere Entwicklung bleibt abzuwarten, empirische Forschung ist dringend nötig. Die Akzeptanz der Bevölkerung gegenüber solchen Verfahren ist bislang allerdings nur gering.

3. Im Bereich der Strafzumessung im engeren Sinn wäre die Einführung einer bundesweiten Strafzumessungsdatenbank (etwa nach japanischem Vorbild) ein Schritt in die richtige Richtung, da sie den Gerichten eine sinnvolle Orientierung über das im konkreten Fall übliche Strafmaß in vergleichbaren Fällen liefern könnte. Rechtliche oder straftheoretische Bedenken bestehen hiergegen m.E. nicht.

4. Auf der Grundlage einer solchen Datenbank könnte weitergehend ein Support-System entwickelt werden, das sich nicht auf die Deskription von Strafmaßen in ähnlichen Fällen beschränkt, sondern auf der Basis dieser Daten einen eigenen Strafzumessungsvorschlag erarbeitet. In dieser Hinsicht stellen sich allerdings noch viele Fragen, u.a. im Hinblick auf die Qualität der zugrundeliegenden Daten und die Transparenz des Systems. Ob die Gerichte einer solchen Entwicklung eher ablehnend gegenüberstehen oder dazu tendieren werden, den Vorschlag im Sinne eines „Automation Bias“ kritiklos zu übernehmen, ist eine offene und spannende Frage, die man empirisch erforschen sollte. Gelegenheit dazu wird es vermutlich – trotz einer bislang eher skeptischen Haltung innerhalb der Justiz gegenüber dem Einsatz von KI im Bereich des Strafrechtsrechts – in nicht allzu ferner Zukunft geben.

⁴² Vgl. nur *Robinson*, *The utility of desert*, 2013.

⁴³ Bertelsmann-Stiftung (Hrsg.), *Was Deutschland über Algorithmen weiß und denkt*, 2018, S. 15.

⁴⁴ Bertelsmann-Stiftung (Fn. 43), S. 31.