

Strafzumessung durch Algorithmen?

Von **Hannah Offerdinger**, Hamburg*

I. Einleitung

In der heutigen Welt sind wir ständig mit Algorithmen¹ konfrontiert, die anhand über uns gesammelter Daten Präferenzen, Interessen oder Bewertungen errechnen. Dies reicht von vorgeschlagenen Suchergebnissen bei Google über Verkaufsangebote bei Amazon bis hin zu unserer Kreditwürdigkeit nach der Schufa.² Trotzdem wird dem justiziellen Einsatz von Computerprogrammen sowohl vonseiten der Praxis als auch vonseiten der Wissenschaft mit tiefer Skepsis begegnet. Hier wird vor allem auf Sicherheitsrisiken bei Computereinsätzen, Intransparenz der Rechenprozesse und mögliche Datenschutzverstöße hingewiesen.³ Zudem wurde in der Vergangenheit eine Schematisierung und Mathematisierung der Strafzumessung immer wieder durch die Revisionsgerichte abgelehnt.⁴ Dennoch zeigt sich wiederholt, dass sich Richterinnen und Richter an „Regelfallempfehlungen“ oder „Strafmaßtabellen“ orientieren.⁵ Auch wird in empirischen Studien immer wieder deutlich, dass häufig unterschiedlich hohe Strafen bei Delikten mit gleichem Strafraum und vergleichbaren Tatumsständen verhängt werden.⁶ Würde es also nicht sinnvoll sein, solche „Bewertungen“ von Fällen durch einen bundeseinheitlichen Algorithmus zu berechnen? Und könnte ein Algorithmus hier nicht eine legitime Hilfestellung anbieten? Dieser Beitrag soll Möglichkeiten und Grenzen des Einsatzes von Algorithmen in der Justiz beleuchten. Dabei sollen auch Fragen aufgeworfen werden, die sich bei der Programmierung und dem Einsatz eines Algorithmus im Strafprozess im Zusammenhang mit der Strafzumessung stellen können. Schließlich soll auch hinterfragt werden, ob

die durchaus anzutreffende Skepsis und Kritik in Bezug auf bestimmte Aspekte gerechtfertigt ist.

Ein praktischer Anwendungsfall eines Algorithmus in der Justiz zeigt sich bei einem Blick über den Atlantik. In den USA werden bereits Algorithmen im Strafverfahren eingesetzt. Seit fast zwei Jahrzehnten nutzen Richterinnen und Richter Computerprogramme zur Bewertung der Rückfallwahrscheinlichkeit von Straftäterinnen und Straftätern.⁷ In vielen US-Bundesstaaten werden diese Computerprogramme bei der Entscheidung über die Freilassung einer oder eines Untersuchungsgefangenen auf Kautions- oder die vorzeitige Haftentlassung genutzt. In neun verschiedenen Bundesstaaten werden diese Computerprogramme auch während des Gerichtsverfahrens verwendet.⁸ Die dabei eingesetzten Algorithmen zur Risikobewertung sind so programmiert, dass sie große Datenmengen über die kriminelle Vergangenheit und die Biografie sowie psychologische Informationen einer Straftäterin oder eines Straftäters verarbeiten und daraus einen Risikowert errechnen.⁹ Je nach Einsatzbereich des Computerprogramms wird dieser Wert genutzt, um eine Entscheidung über die bedingte Haftentlassung, Rehabilitierungsmaßnahmen oder sogar die Bestrafung zu treffen.¹⁰

Die Verwendung solcher Computerprogramme stößt jedoch unter anderem deswegen auf Ablehnung, weil deren Funktion und Arbeitsweise oft nicht richtig bekannt sind oder durch Medienberichte nicht richtig dargestellt werden. So erschien auf der Website der Frankfurter Allgemeinen Zeitung erst letztes Jahr ein Artikel, in dem es hieß: „Sechs Jahre Haft für unerlaubtes Benutzen eines fremden Autos und mangelnde Kooperation mit der Polizei: Diese Strafe hatte Eric Loomis einem Algorithmus zu verdanken.“¹¹ Ferner hieß es in der Überschrift „Algorithmen sprechen Recht“.¹² Allerdings gab dieser Artikel weder die wirkliche Arbeitsweise des Algorith-

* Die *Verfasserin* ist wissenschaftliche Mitarbeiterin am Lehrstuhl für Strafrecht und Strafprozessrecht einschließlich ihrer internationalen und historischen Bezüge der Universität Hamburg (Prof. Dr. Dr. Milan Kuhli).

¹ Die Verwendung der Begriffe „Algorithmen“ oder „Computerprogramme“ meint im Folgenden Software, die aus einer Reihe von Eingabewerten einen einzigen Wert berechnet, siehe *Zweig/Krafft*, in: Mohabbat Kar/Resa/Thapa/Basanta E.P./Parycek/Peter (Hrsg.), (Un)Berechenbar? Algorithmen und Automatisierung in Staat und Gesellschaft, 2018, S. 204 (208).

² Vgl. zu weiteren Beispielen *Mohabbat Kar/Parycek*, in: Mohabbat Kar/Resa/Thapa/Basanta E.P./Parycek/Peter (Fn. 1), S. 7 f.

³ Darstellung von Kritik und gute Lösungsvorschläge bei *Martini*, JZ 2017, 1017.

⁴ Vgl. etwa nur BGH NJW 1987, 3014 (3015); BGH StV 2010, 480.

⁵ *Eschelbach*, in: Satzger/Schluckebier/Widmaier (Hrsg.), Kommentar zum Strafgesetzbuch, 4. Aufl. 2018, § 46 Rn. 4.

⁶ Siehe *Kudlich/Koch*, NJW 2018, 2762 (2763), mit Verweis auf *Kaspar*, Sentencing Guidelines vs. freies tatrichterliches Ermessen – Brauchen wir ein neues Strafzumessungsrecht?, Gutachten C zum 72. DJT, 2018, C 16 ff.

⁷ New York State begann die Nutzung des COMPAS-Programms im Jahre 2001, vgl. *Angwin/Larson/Mattu/Kirchner*, Pro Publica, Machine Bias. There's software used across the country to predict future criminals. And it's biased against blacks, 2016, abrufbar unter

<https://www.propublica.org/article/machine-bias-risk-assessments-in-criminal-sentencing> (29.8.2020).

⁸ *Angwin/Larson/Mattu/Kirchner* (Fn. 7).

⁹ Siehe unter II.

¹⁰ *Freeman*, Algorithmic Injustice: How the Wisconsin Supreme Court Failed to Protect Due Process Rights in *State v. Loomis*, North Carolina Journal of Law & Technology 18 (December 2016), 75.

¹¹ Frankfurter Allgemeine Zeitung Rhein-Main v. 11.6.2019, abrufbar unter

<https://www.faz.net/aktuell/rhein-main/algorithmen-werden-in-amerika-bei-gerichtsprozessen-genutzt-16230589.html> (29.8.2020).

¹² Frankfurter Allgemeine Zeitung Rhein-Main v. 11.6.2019 (Fn. 11).

mus wieder noch stimmte die Aussage, der Algorithmus würde Recht sprechen.

Der Schwerpunkt dieses Beitrags liegt in der Diskussion der Frage, ob ausgewählte Strafzumessungsumstände des deutschen Rechts durch einen Algorithmus beurteilt werden könnten (III.). Zuvor soll jedoch ein Blick in die USA und auf das dort eingesetzte Programm geworfen werden (II.). Diese Betrachtung soll vor allem dazu dienen, die Programmierung und Arbeitsweise eines Algorithmus in der Strafjustiz zu skizzieren. Zuletzt sollen die eventuell zu bewältigenden Probleme im Rahmen der Programmierung und Anwendung eines Computerprogramms in Deutschland herausgearbeitet werden (IV.).

II. Die Berechnung der Rückfallwahrscheinlichkeit eines Straftäters – ein Blick in die USA

Bei der Bewertung der Rückfallwahrscheinlichkeit einer Straftäterin oder eines Straftäters kann unter anderem mit klinischen oder statistischen Methoden gearbeitet werden.¹³ Während einem Menschen beide Methoden offenstehen, dürften die derzeit verwendeten Algorithmen prinzipiell nach der statistischen Methode arbeiten.¹⁴ Bei der Erstellung eines Computerprogramms zur Bewertung der Rückfallwahrscheinlichkeit einer Straftäterin oder eines Straftäters wird auf zwei unterschiedlichen Stufen gearbeitet: Auf der ersten Stufe müssen eine Datenbasis geschaffen, Rückfallrisiken definiert und die Rückfallrisiken in ihrer Datenbasis identifiziert werden. Mithilfe dieser Informationen wird sodann ein Algorithmus kreiert, der Angaben zur Wahrscheinlichkeit eines Rückfalls liefern soll. Auf der zweiten Stufe muss dann eine Entscheidung darüber getroffen werden, wie die Ergebnisse des Algorithmus für den tatsächlichen Gebrauch in der Justiz wiedergegeben werden sollen.¹⁵ Hier muss also bestimmt werden, wie die Daten und die Auswertung für den bewerteten Einzelfall für die jeweiligen Anwender dargestellt werden sollen.

Das in den USA eingesetzte Programm nennt sich „Correctional Offender Management Profiling for Alternative Sanctions“ oder kurz: COMPAS. Es wurde von der Firma Equivant (ehem. Northpointe) entwickelt, einem privatwirtschaftlichen Unternehmen, das nach eigenen Angaben „Werkzeuge“ für die Nutzung im Justizsystem anbietet.¹⁶ Auch wenn die Firma nicht alle Informationen die Programmierung betreffend offengelegt hat, so sind doch einige grundlegende Infor-

mationen bekannt. Die Datenbasis oder Normgruppe bei COMPAS bildeten 7.381 Straftäterinnen und Straftäter. Entweder waren sie inhaftiert oder die Strafe bzw. der Strafrest waren zur Bewährung ausgesetzt.¹⁷ Anhand dieser Normgruppe wurden verschiedene Kategorien identifiziert und in insgesamt 43 Skalenwerte eingeteilt.¹⁸ Dabei wurden zum einen Risikofaktoren, zum anderen Hilfsbedarfe herausgearbeitet. Die jeweils möglichen Ergebnisse einer analysierten Person werden von COMPAS im konkreten Fall in 43 Balkendiagrammen dargestellt.¹⁹ Die 43 Skalenwerte werden schließlich in eine Skala mit Werten von 1–10 übersetzt, wobei ein Wert von 1–4 als niedrig, 5–7 als mittel und 8–10 als hoch im Vergleich zu der Normgruppe eingestuft wird.²⁰ Die Werte geben also an, welchem Zehntel der Normgruppe die konkret bewertete Person ähnelt, also einer Personengruppe mit niedrigem, mittlerem oder hohem Rückfallrisiko, und fassen die ermittelten Skalenwerte zusammen. Dabei wurde nicht offengelegt, welches Gewicht den einzelnen Skalenwerten bei der Übersetzung in die Skala von 1–10 beigemessen wird. So ist nicht öffentlich bekannt, ob bestimmten Faktoren ein höheres Gewicht bei der Beurteilung des Rückfallrisikos gegeben wurde als anderen.²¹

Liegt der RichterIn oder dem Richter die Prognoseentscheidung des Computerprogramms vor, muss sie oder er diese interpretieren und entscheiden, ob und wie diese in das konkrete Urteil einbezogen werden soll.²² In ihrem Practitioners Guide weist Equivant darauf hin, dass das für die konkrete Person ermittelte Risiko relativ und immer im Vergleich zu der Normgruppe interpretiert werden muss.²³ Ferner stellt der Hersteller klar, dass die Fähigkeit zur zutreffenden Interpretation der Skalen erlernt werden muss und die verschiedenen Skalen auch in ihren wechselseitigen Beziehungen zueinander gesehen werden müssen.²⁴ Die von dem COMPAS-Programm ausgegebenen Risikowerte für die einzelnen Skalen werden aus der Biografie der oder des Beschuldigten und den Antworten auf 137 Fragen abgeleitet. Zusammengestellt werden die Angaben zum einen aus Polizeiakten und zum anderen aus Fragebögen, die von der oder dem Beschuldigten selbst ausgefüllt wurden.²⁵ Dabei wird ebenfalls getestet, ob die Antworten des Straftäters oder der Straftäterin der Wahrheit entsprechen können oder ob diese suspekt erscheinen.²⁶

¹³ Sohn, Angloamerikanische Untersuchungen zur Rückfalligkeit gewalttätiger Sexualstraftäter: Zwischenresultate einer Sekundäranalyse, Wiesbaden: Kriminologische Zentralstelle, 3. Aufl. 2007, S. 46 f.

¹⁴ John Howard Society of Alberta, Offender Risk Assessment, 2000, S. 1 f., abrufbar unter <https://www.johnhoward.ab.ca/document/offender-risk-assessment-2000/> (29.8.2020).

¹⁵ Eaglin, Constructing Recidivism Risk, Emory Law Journal 67 (2017), 59 (64).

¹⁶ Das Unternehmen selbst bezeichnet die entwickelten Programme als „tools“ vgl. „Über uns“ Website, abrufbar unter <https://www.equivant.com/about-us/> (29.8.2020).

¹⁷ Equivant, Practitioners Guide to Compas Core, 2019, S. 11, abrufbar unter <https://www.equivant.com/practitioners-guide-to-compas-core/> (29.8.2020).

¹⁸ Equivant (Fn. 17), S. 2.

¹⁹ Equivant (Fn. 17), S. 3.

²⁰ Equivant (Fn. 17), S. 8.

²¹ Lischka/Klingel, Wenn Maschinen Menschen bewerten, Internationale Fallbeispiele für Prozesse algorithmischer Entscheidungsfindung – Arbeitspapier –, 2017, S. 9.

²² Lischka/Klingel (Fn. 21), S. 9.

²³ Equivant (Fn. 17), S. 11.

²⁴ Equivant (Fn. 17), S. 4.

²⁵ Lischka/Klingel (Fn. 21), S. 9.

²⁶ Siehe zum Testverfahren Freeman (Fn. 10), S. 79 f.

Nicht grundlos wird der Einsatz von COMPAS und anderer Computerprogramme zur Bestimmung der Rückfallwahrscheinlichkeit in den USA kritisiert.²⁷ Eine Untersuchung von ProPublica, einer unabhängigen US-Nachrichtenorganisation, kam zu der Wertung, dass COMPAS zum einen unzuverlässig bei der Vorhersage zukünftiger Straftatbegehung sei und zum anderen Rassenunterschiede in seinen Ergebnissen aufweise.²⁸ Nach dieser Untersuchung wurden 61 % derjenigen Straftäterinnen und Straftäter, denen ein hohes Rückfallrisiko prognostiziert wurde, innerhalb von zwei Jahren nach der ersten Einstufung festgenommen – ein Wert, den die Autoren mit der Zuverlässigkeit eines Münzwurfs vergleichen.²⁹ Des Weiteren wies der Algorithmus in seiner damaligen Programmierung eine Tendenz auf, Schwarzen Straftäterinnen und Straftätern öfter fälschlicherweise ein hohes Rückfallrisiko zu prognostizieren als weißen. Nach den Autoren der Untersuchung wurden Schwarze Straftäterinnen und Straftäter fast doppelt so oft mit einem hohen Rückfallrisiko bewertet wie weiße.³⁰

Auch wenn Equivant in seiner Antwort auf die Untersuchung deren Ergebnisse heftig bestritt,³¹ so deutet die Untersuchung zumindest an, dass bestimmte Faktoren mittelbar die Einstufung des Rückfallrisikos beeinflussen können. Dies beginnt bereits bei der Bildung der Normgruppe. Hier bestand diese aus einem Personenkreis, der durch ein Gericht wegen einer Straftat verurteilt wurde. Dabei kann nicht ausgeschlossen werden, dass die jeweilige Richterin oder der jeweilige Richter bei ihrer oder seiner Urteilsfindung bestimmten Vorurteilen unterlegen ist. Auch wenn es sich nicht um offensichtliche rassistische oder sexistische Einstellungen handeln muss, so können doch genau solche Faktoren Richterinnen und Richter unterschwellig beeinflussen.

Ein weiteres Problem besteht in der Verzerrung der Falsifikation.³² Diese tritt dann auf, wenn Richterinnen und Richter aufgrund der Risikoprognose des Computerprogramms eher zur Verhängung von Haftstrafen tendieren. Durch einen Gefängnisaufenthalt kann die Gefahr eines Rückfalles erheblich steigen, da Häftlinge in neue kriminelle Kontexte integriert und nach ihrer Entlassung auch häufig rückfällig werden.³³ Das eingesetzte Computerprogramm wird damit schließlich in seiner Einschätzung bestätigt, da die betreffende Person tatsächlich rückfällig geworden ist. Hier haben fälschlicherweise mit hohem Risiko bewertete Straftäterinnen

und Straftäter kaum eine Möglichkeit zu beweisen, dass sie ohne die Verhängung einer Gefängnisstrafe nicht rückfällig geworden wären.³⁴

Dieser Verzerrung kann nur dann entgegengewirkt werden, wenn zunächst eindeutig festgelegt wird, ob das Computerprogramm eher falsch-negative (fälschlicherweise mit niedrigem Risiko bewertete) oder falsch-positive (fälschlicherweise mit hohem Risiko bewertete) Ergebnisse liefern soll. Bereits bei der Programmierung muss also entschieden werden, ob es der Gesellschaft eher zuzumuten ist, dass Verurteilte während ihrer Bewährungszeit erneut Straftaten begehen, oder ob zum Schutze der Allgemeinheit jegliches Rückfallrisiko durch Haftstrafen minimiert werden soll.³⁵

III. Ausgewählte Strafzumessungsumstände und deren Algorithmisierung

Die Strafzumessung ist bekanntlich die Bestimmung der Rechtsfolgen der Tat durch die Richterin oder den Richter.³⁶ § 46 StGB ist die zentrale Norm für die Strafzumessung durch das Gericht.³⁷ Sie benennt zunächst die Schuld als Grundlage der Strafzumessung (§ 46 Abs. 1 S. 1 StGB). Zudem sind die Umstände, die für und gegen die Täterin oder den Täter sprechen, durch das Gericht gegeneinander abzuwägen (§ 46 Abs. 2 S. 1, 2 StGB). Dabei bleibt das Gesetz jedoch unbestimmt, da es weder den Schuldbegriff noch die Art und Weise der Abwägung genauer definiert.³⁸ An dieser Stelle soll weder auf den Sinn und Zweck der Strafe bzw. auf den Begriff der Schuld eingegangen werden, noch kann im Einzelnen auf die durch das Gericht vorzunehmende Abwägung Bezug genommen werden. Hier bleibt nur grundsätzlich festzuhalten, dass die Regelung des § 46 StGB einen weiten Spielraum eröffnet, innerhalb dessen das Tatgericht im jeweiligen Einzelfall eine konkrete Strafe zu bestimmen hat.³⁹ Dieser Spielraum ist im Rahmen der Rechtsmittel auch nur eingeschränkt überprüfbar.⁴⁰

³⁴ Lischka/Klingel (Fn. 21), S. 9.

³⁵ Vgl. auch Spielkamp, *Inspecting Algorithms for Bias*, MIT Technology Review, 2017.

³⁶ Horn/Wolters, in: Wolters (Hrsg.), *Systematischer Kommentar zum Strafgesetzbuch*, Bd. 2, 9. Aufl. 2016, § 46 Rn. 2; v. Heintschel-Heinegg, in: v. Heintschel-Heinegg (Hrsg.), *Beck'scher Online-Kommentar, Strafgesetzbuch*, Stand: 1.2.2020, § 46 Rn. 1.

³⁷ Eschelbach (Fn. 5), § 46 Rn. 1; Miebach/Maier, in: Joecks/Miebach (Hrsg.), *Münchener Kommentar zum Strafgesetzbuch*, Bd. 2, 3. Aufl. 2016, § 46 Rn. 1.

³⁸ Streng, in: Kindhäuser/Neumann/Paeffgen (Hrsg.), *Nomos Kommentar, Strafgesetzbuch*, Bd. 1, 5. Aufl. 2017, § 46 Rn. 19; Eschelbach (Fn. 5), § 46 Rn. 13, 189.

³⁹ Kinzig, in: Schönke/Schröder, *Strafgesetzbuch, Kommentar*, 30. Aufl. 2019, § 46 Rn. 68; Miebach/Maier (Fn. 37), § 46 Rn. 3.

⁴⁰ BGH NSZ-RR 2006, 340 (341) m.w.N.; Eschelbach (Fn. 5), § 46 Rn. 3, 12; Miebach/Maier (Fn. 37), § 46 Rn. 3; Wenske, in: Schneider (Hrsg.), *Münchener Kommentar zur Strafprozessordnung*, Bd. 2, 2016, § 267 Rn. 387.

²⁷ Insbesondere Angwin/Larson/Mattu/Kirchner (Fn. 7); Israni, *The New York Times* v. 26.10.2017, abrufbar unter <https://www.nytimes.com/2017/10/26/opinion/algorithm-compas-sentencing-bias.html> (29.8.2020), aber auch Eaglin, *Emory Law Journal* 67 (2017), 121.

²⁸ Angwin/Larson/Mattu/Kirchner (Fn. 7).

²⁹ Angwin/Larson/Mattu/Kirchner (Fn. 7).

³⁰ Angwin/Larson/Mattu/Kirchner (Fn. 7).

³¹ Stellungnahme abrufbar unter

<https://www.equivalent.com/response-to-propublica-demonstrating-accuracy-equity-and-predictive-parity/> (29.8.2020).

³² Siehe auch Lischka/Klingel (Fn. 21), S. 10.

³³ Lischka/Klingel (Fn. 21), S. 9.

Insbesondere subjektive Komponenten sind unter Umständen wesentlich für die Entscheidungsfindung.⁴¹ So kann etwa die subjektive Wahrnehmung der Richterin oder des Richters von der oder dem Angeklagten einen Einfluss auf die Entscheidungsfindung haben. Auch kann das äußere Erscheinungsbild, die Sprache oder das Verhalten der oder des Angeklagten die Entscheidung prägen.⁴²

Zur Ermittlung des Maßes des konkreten Tatunrechts sowie der Höhe der Tatschuld zählt § 46 Abs. 2 S. 2 StGB beispielhaft Umstände auf, die je nach Lage des Falles zugunsten oder zulasten der oder des Angeklagten ins Gewicht fallen können. Bei den ersten vier handelt es sich um tatbezogene, während bei den letzten das Vorleben und Verhalten der Täterin oder des Täters die Bezugspunkte darstellen.⁴³ Im Folgenden soll anhand dreier beispielhaft gewählter Umstände untersucht werden, ob eine Bewertung dieser Aspekte möglicherweise auch mittels eines Algorithmus vollzogen werden könnte. Zudem wird kurz dargestellt, ob eine Bewertung des jeweiligen Umstandes durch einen Algorithmus überhaupt einen Vorteil erbringen würde. Die ausgewählten Umstände sind zunächst die wirtschaftlichen Verhältnisse der Täterin oder des Täters, die verschuldeten Auswirkungen der Tat sowie die Gesinnung, die aus der Tat spricht.

1. Wirtschaftliche Verhältnisse

Unter den wirtschaftlichen Verhältnissen der Täterin oder des Täters versteht man die finanziellen Verhältnisse der oder des Angeklagten zum Zeitpunkt der Verurteilung.⁴⁴ Dabei sind grundsätzlich Einkommen und Vermögen zu berücksichtigen, es können aber auch Schulden oder persönliche Verpflichtungen (z.B. gegenüber Angehörigen) bei der Ermittlung der wirtschaftlichen Verhältnisse berücksichtigt werden.⁴⁵ Die wirtschaftlichen Verhältnisse der oder des Angeklagten sind oft bei der Bemessung einer Geldstrafe von Bedeutung.⁴⁶ Dies gilt insbesondere bei Vermögens- und Steuerdelikten.⁴⁷ Darüber hinaus können die wirtschaftlichen Verhältnisse auch das Tatmotiv kennzeichnen, so dass auch die wirtschaftliche Lage zur Tatzeit Bedeutung erlangen kann. Besteht eine innere Beziehung zwischen der Tat und den wirtschaftlichen Verhältnissen, etwa weil eine Begehung aus wirtschaftlicher „Not“ erfolgte, können sie im Rahmen der Schuld relevant sein.⁴⁸ Die wirtschaftlichen Verhältnisse werden hier aber ambivalent beurteilt. So soll eine ungünstige wirtschaftliche

Lage strafmildernd ins Gewicht fallen können.⁴⁹ Ob dagegen eine gute wirtschaftliche Lage, etwa bei der Begehung eines Vermögensdelikts, strafscharfend wirken kann, soll aber nach dem Einzelfall beurteilt werden.⁵⁰

Ein Algorithmus kann hier durchaus die Rolle der Richterin oder des Richters übernehmen: eine schlicht mathematische Berechnung von Einkommen abzüglich aller Verpflichtungen. Im Rahmen der Familiengerichtsbarkeit werden bereits ähnliche Programme zur Berechnung des Versorgungsausgleichs eingesetzt.⁵¹ Neben einer reinen Berechnungsaufgabe kann man einen Algorithmus aber auch mit weiteren Informationen „füttern“, wodurch er das Ergebnis der Berechnung in einen Kontext setzen kann. So wäre es etwa möglich abzubilden, ob das Einkommen der oder des Angeklagten über der Armutsgrenze liegt oder wie es sich im Verhältnis des Durchschnittseinkommens darstellt. Zudem könnte spezifischer ermittelt werden, ob das Einkommen am konkreten Wohnort zur adäquaten Lebensführung hinreichend ist. So können die errechneten wirtschaftlichen Verhältnisse konkret mit den durchschnittlichen Lebenshaltungskosten am Lebens- und Wohnort der oder des Angeklagten verglichen werden.

Sofern der Algorithmus aber neben einer schlicht mathematischen Berechnung zusätzlich eine Klassifikation und Bewertung der Daten vornehmen soll, stellt sich zwangsläufig die Frage, wer die dafür zu berücksichtigenden Kriterien festlegen darf. Zudem betont der BGH ausdrücklich eine fallbezogene Beurteilung der materiellen Lage der Täterin oder des Täters.⁵² Eine fallbezogene Bewertung wird durch den Einsatz eines Algorithmus jedoch nicht per se ausgeschlossen. Hier ist hervorzuheben, dass der Algorithmus in keinem Fall die gesamte Beurteilung der Richterin oder des Richters ersetzen darf. Es sollen lediglich einheitliche Maßstäbe und Grundwertungen geschaffen werden. Ein Algorithmus kann unvoreingenommen und auch emotionslos schlicht Zahlen bewerten und diese in einen Kontext setzen. Für die Richterinnen und Richter kann auf diese Weise ein Hilfsmittel geschaffen werden, um die Angaben der oder des Angeklagten besser einordnen und besser bewerten zu können. Auf die Frage, wer die für einen Algorithmus zu berücksichtigenden Kriterien festlegen soll oder darf, soll am Schluss dieses Beitrages zurückgekommen werden.

2. Verschuldete Auswirkungen der Tat

Die verschuldeten Auswirkungen der Tat sind von zentraler Bedeutung für die Praxis der Strafzumessung bei Erfolgsdelikten.⁵³ Dabei können alle Arten von Folgen, wie materielle,

⁴¹ Eschelbach (Fn. 5), § 46 Rn. 4; Miebach/Maier (Fn. 37), § 46 Rn. 4.

⁴² Miebach/Maier (Fn. 37), § 46 Rn. 4.

⁴³ Kühl, in: Lackner/Kühl, Strafgesetzbuch, Kommentar, 29. Aufl. 2018, § 46 Rn. 33, 36.

⁴⁴ Miebach/Maier (Fn. 37), § 46 Rn. 245.

⁴⁵ Streng (Fn. 38), § 46 Rn. 73.

⁴⁶ Kinzig (Fn. 39), § 46 Rn. 37; Streng (Fn. 38), § 46 Rn. 73.

⁴⁷ Kinzig (Fn. 39), § 46 Rn. 38; a.A. Horn/Wolters (Fn. 36), § 46 Rn. 141.

⁴⁸ Kühl (Fn. 43), § 46 Rn. 39; Horn/Wolters (Fn. 36), § 46 Rn. 141; Streng (Fn. 38), § 46 Rn. 73.

⁴⁹ Eschelbach (Fn. 5), § 46 Rn. 159; Kinzig (Fn. 39), § 46 Rn. 38.

⁵⁰ Kinzig (Fn. 39), § 46 Rn. 38; a.A. Horn/Wolters (Fn. 36), § 46 Rn. 141, wonach das Unrecht in allen Fällen gleich zu werten sein soll.

⁵¹ Hierzu krit. Vogel, FPR 2004, 242.

⁵² BGH NStZ 1987, 450.

⁵³ Streng (Fn. 38), § 46 Rn. 57 m.w.N.

körperliche, seelische oder ideelle erheblich sein.⁵⁴ Dabei ist lediglich zu beachten, dass sie der Täterin oder dem Täter nur dann zur Last gelegt werden können, wenn sie zweifellos von ihr oder ihm verschuldet sind.⁵⁵ Ob nur diejenigen Tatfolgen, die ausdrücklich im jeweiligen Tatbestand genannt sind, oder darüber hinaus auch nicht ausdrücklich erfasste mögliche Folgen als verschuldet anzusehen sind, ist umstritten.⁵⁶ Dieser Streit soll hier jedoch weder diskutiert noch entschieden werden. Wenn im Folgenden von Auswirkungen oder Tatfolgen gesprochen wird, so meint dies zumindest die ausdrücklich vom jeweiligen gesetzlichen Tatbestand benannten Folgen.

Auf den ersten Blick erscheint die Annahme befremdlich, man könne die verschuldeten Auswirkungen der Tat durch einen Algorithmus auswerten lassen. Blicken wir jedoch in das Zivilrecht, so sehen wir, dass dort Tabellen zur Bestimmung von Schmerzensgeldsummen genutzt werden.⁵⁷ Dort wird aufgelistet, wie viel Schmerzensgeld in der Vergangenheit bei einer bestimmten Verletzung von einem Gericht zugesprochen wurde. Die Summen in der Tabelle sind also Ausdruck dessen, welche Gravität einer Verletzung und ihren Folgen in der Vergangenheit beigemessen wurde. Es erscheint also nicht fernliegend, das Ausmaß der Tat durch einen Algorithmus bestimmen zu lassen. So könnte ein Algorithmus beispielsweise in einer Weise programmiert werden, in der er Verletzungen als gering, mittel oder schwer einstuft. Auch kann der Algorithmus die gesamten Verletzungen der betroffenen Person im Zusammenhang bewerten, um eine vollständige Einschätzung zum Ausdruck zu bringen. Dabei könnten als Datenbasis sowohl vergangene Gerichtsentscheidungen als auch ärztliche Einstufungen oder Gutachten dienen.

An dieser Stelle ist hervorzuheben, dass es den Vorgaben des Gesetzgebers oder der Idee des Schuldstrafrechts widersprechen würde, wenn die Schadenshöhe die Strafzumessung streng determinieren würde.⁵⁸ Allerdings weist der BGH (insbesondere im Steuerstrafrecht) immer wieder darauf hin, dass ab einer gewissen Schadenssumme eine Geldstrafe nur noch in Ausnahmefällen in Betracht komme.⁵⁹ So bestehen also durchaus gewisse Bewertungsmaßstäbe, die Einfluss auf die Strafzumessung nehmen. Diese Bewertungsmaßstäbe in einen Algorithmus zu übersetzen, ist somit nur eine Frage der Datenbasis.

Auch hier bleibt jedoch anzumerken, dass ein Algorithmus die richterliche Bewertung nicht ersetzen kann und soll.

Zum einen wird dies schon deswegen nicht möglich sein, weil die Auswirkungen immer im Kontext der jeweiligen Tat zu betrachten sind, zum anderen aber auch, weil das Ausmaß der Verletzungen oder die Schadenssumme eben nicht alleiniges Strafzumessungskriterium sein können. Algorithmen können hier aber eine relative Gleichheit schaffen. So verhält sich der Algorithmus indifferent zu der Frage, ob das Opfer in der Gerichtsverhandlung weint oder nicht, wenn es von seiner Schürfwunde am Knie erzählt. Denn eine Schürfwunde bleibt eben nur eine solche und ist auf einer Bewertungsskala geringer einzustufen als ein gebrochenes Bein. Ob und inwieweit emotionale Belastungen für die betroffenen Personen auch durch den Algorithmus einbezogen werden sollen, ist wiederum eine Frage der bei der Programmierung genutzten Datenbasis. Werden hier vergangene Gerichtsentscheidungen oder Gutachten genutzt, so werden zwangsläufig auch emotionale Belastungen in die Programmierung einfließen. Denn sowohl Richterinnen und Richter als auch Gutachterinnen und Gutachter werden bei der Bemessung von Verletzungen in der Regel von der (emotionalen) Schilderung der betroffenen Person beeinflusst sein.

3. Aus der Tat sprechende Gesinnung

Die Gesinnung, die aus der Tat spricht, ist losgelöst von der allgemeinen Gesinnung der Täterin oder des Täters zu betrachten.⁶⁰ Allerdings kann zu ihrer Feststellung auf die Persönlichkeit der Täterin oder des Täters zurückgegriffen werden.⁶¹ Bei der Beurteilung ist eine klare Abgrenzung zu den Beweggründen und Zielen der Täterin oder des Täters kaum möglich.⁶² So liefern die Beweggründe und Ziele in der Regel Kriterien für die Beurteilung der Gesinnung.⁶³ Die psychischen Hintergründe der Tat können sich auch bei der konkreten Ausführung zeigen, diese beeinflussen und insofern auch bei der Suche nach der Gesinnung heranzuziehen sein.⁶⁴ Dabei besteht jedoch auch schnell die Gefahr, dass moralisierende Erwägungen herangezogen werden, welche den Bestand des Urteils gefährden können.⁶⁵

Die Frage, ob ein Algorithmus die Gesinnung einer Täterin oder eines Täters beurteilen oder bewerten könnte, führt zunächst zu der Frage, wie (Tat-)Gesinnung genau zu definieren ist. Im Duden wird sie als geistige und sittliche Grundeinstellung eines Menschen umschrieben.⁶⁶ Die Synonyme umfassen Anschauung, Ansicht, Auffassung, Einstellung, Geisteshaltung, Überzeugung, Vorstellung, Weltbild, Ethos und

⁵⁴ Kinzig (Fn. 39), § 46 Rn. 26; Miebach/Maier (Fn. 37), § 46 Rn. 214; Schäfer/Sander/van Gemmeren, Praxis der Strafzumessung, 6. Aufl. 2017, Rn. 584.

⁵⁵ Kinzig (Fn. 39), § 46 Rn. 26; Miebach/Maier (Fn. 37), § 46 Rn. 213.

⁵⁶ Kinzig (Fn. 39), § 46 Rn. 26ff.; Kühl (Fn. 43), § 46 Rn. 34; Miebach/Maier (Fn. 37), § 46 Rn. 214 f.

⁵⁷ Z.B. Slizyk, Schmerzensgeld 2020, Handbuch und Tabellen, 16. Aufl. 2020.

⁵⁸ Streng (Fn. 38), § 46 Rn. 57.

⁵⁹ BGH NJW 2009, 528 (531 f.); BGH NJW 2012, 5599 f.; BGH NStZ 2012, 634 (636).

⁶⁰ BGH NJW 1979, 1835; Kinzig (Fn. 39), § 46 Rn. 16; Miebach/Maier (Fn. 37), § 46 Rn. 193; Schäfer/Sander/van Gemmeren (Fn. 54), Rn. 615.

⁶¹ v. Heintschel-Heinegg (Fn. 36), § 46 Rn. 33; Schäfer/Sander/van Gemmeren (Fn. 54), Rn. 615.

⁶² Kinzig (Fn. 39), § 46 Rn. 16; Streng (Fn. 38), § 46 Rn. 53.

⁶³ Kinzig (Fn. 39), § 46 Rn. 16.

⁶⁴ Streng (Fn. 38), § 46 Rn. 52.

⁶⁵ Miebach/Maier (Fn. 37), § 46 Rn. 193; Schäfer/Sander/van Gemmeren (Fn. 54), Rn. 614.

⁶⁶ Abrufbar unter

<https://www.duden.de/rechtschreibung/Gesinnung> (29.8.2020).

viele weitere.⁶⁷ In der höchstrichterlichen Rechtsprechung finden sich fast ebenso viele Umschreibungen dessen, was unter der aus der Tat sprechenden Gesinnung zu verstehen ist. So wird auf die sich in der Handlung äußernde innere Einstellung,⁶⁸ das Beruhen der Tat auf einer verwerflichen Gesinnung⁶⁹ oder die besondere innere Einstellung zu der Tat⁷⁰ bezogen. Wenn die Gesinnung alle Werte, Haltungen und die Weltanschauung einer Person umfasst, führt diese Erkenntnis sodann jedoch zu der Folgefrage, ob Werte und Weltanschauungen überhaupt zu bewerten sind. Daran schließt sich sodann die Frage an, wann eine Gesinnung, welche in der Tat Ausdruck gefunden hat, strafscharfend und wann strafmildernd zu berücksichtigen ist.

Nach *Kühl* soll die Gesinnung schwerpunktmäßig nach rechtlichen, unterstützend jedoch auch nach ethischen Gesichtspunkten beurteilt werden.⁷¹ Dies führt jedoch unweigerlich zu der normativen Frage, welche ethischen Gesichtspunkte maßgeblich sein sollen. Diese Frage kann hier nicht beantwortet werden. Und vielleicht sollte sie auch gar nicht beantwortet werden. Dies gilt zumindest dann, wenn man von der Erwartung abrückt, dass sämtliche zum jetzigen Zeitpunkt von Richterinnen und Richtern anzustellenden Erwägungen algorithmisiert werden sollten. Es könnte also durchaus ausreichen, einige Kriterien einer algorithmenbasierten Bewertung zu unterwerfen, andere dabei aber bewusst auszulassen.

IV. Programmierung, Kriterien-Bestimmung und Verwendung

Zuletzt bleiben noch die Fragen, wer einen Algorithmus für die Strafzumessung programmieren könnte oder wer dies tun sollte. Zudem ist zu fragen, wer die Bewertungskriterien des Algorithmus festlegen darf und wie dessen Ergebnisse für Richterinnen und Richter dargestellt und von ihnen verwendet werden können. Dabei gilt es zunächst festzustellen, welche Funktion ein Algorithmus im Strafverfahren einnimmt. Einerseits sprechen Gründe dafür, den Algorithmus als konkretisierendes Instrument bestehender Gesetze zu verstehen. Ist es nämlich die Aufgabe des Algorithmus, die unbestimmten Strafzumessungsumstände des § 46 Abs. 2 S. 2 StGB genau zu definieren, so stellt er gesetzliche Regelungen auf. Andererseits könnte der Algorithmus auch als „Organ der Rechtsprechung“ verstanden werden. Dies ist dann der Fall, wenn es die Aufgabe des Algorithmus ist, bereits zuvor konkretisierte Strafzumessungsumstände im Einzelfall zu subsumieren. Hier würde der Algorithmus allein Aufgaben wahrnehmen, die bisher Richterinnen und Richtern vorbehalten sind.

Versteht man den Algorithmus als „gesetzliches Instrument“, so bedarf die Programmierung und Ingebrauchnahme einer parlamentarischen Legitimation. Die Zuständigkeit und das Verfahren der Kriterien-Bestimmung und deren Überset-

zung in einen Algorithmus müssen denselben Regelungen folgen wie eine Gesetzesänderung desjenigen Gesetzes, welches der Algorithmus konkretisiert. Sieht man den Algorithmus dagegen als „Organ der Rechtsprechung“ so bedürfte es zunächst keiner parlamentarischen Legitimation. Hier führt jedoch das eingangs angeführte Beispiel aus den USA zu der Frage, ob es mit unserem Verständnis von Rechtsprechung vereinbar ist, wenn Programmiererinnen und Programmierer den rechtsprechenden Organen eine „Richtung“ bei der Urteilsfindung vorgeben. Dieses Problem ist jedoch bei weitem nicht unbekannt. So stellt sich diese Frage auch, wenn Richterinnen und Richter zur Beurteilung eines Sachverhalts und somit letztendlich auch zur Urteilsfindung ein Sachverständigengutachten anfordern. In diesem Fall beurteilt auch eine nicht juristisch ausgebildete Person einen Sachverhalt und die Richterin oder der Richter muss sich mangels Sachkunde auf dessen Urteil verlassen.

Es kann nicht von Juristinnen und Juristen erwartet werden, dass sie in der Lage sind, Algorithmen zu programmieren, aber gleichermaßen kann nicht von Programmiererinnen und Programmierern erwartet werden, dass sie juristisch fundierte Einschätzungen abgeben können. Eine Kooperation zwischen diesen beiden Berufsgruppen wird hier der einzig gangbare Weg sein, um zufriedenstellende Lösungen zu entwickeln. Ebenso kann bei der Beurteilung der Kriterien nur eine Kooperation zwischen Praxis und Wissenschaft zu Ergebnissen führen, welche allgemein vertretbar erscheinen. Die in den USA gewählte Methode, sich bei der Programmierung auf vergangene Urteile und Fälle zu beziehen, ist ein möglicher Weg, die Datenbasis eines Algorithmus zu speisen. Allerdings muss auch hier eine Vielzahl an Faktoren berücksichtigt werden. In Deutschland führt dies etwa zu den Fragen, ob die Urteile proportional zur Anzahl der Gerichte in den einzelnen Bundesländern ausgewählt werden und ob die jeweils als solche bezeichneten „Regelfälle“ überhaupt vergleichbar sind. Darüber hinaus berühren die festzulegenden Kriterien aber auch noch viel grundlegendere juristische und gesellschaftliche Problematiken. Denn fragen wir nach der Nutzung algorithmischer Entscheidungssysteme, so müssen wir uns zugleich auch mit Fragen der Gerechtigkeit der Bewertung von Menschen beschäftigen. Zudem müssen wir fragen, ob sich Menschen überhaupt von Algorithmen bewerten lassen wollen und ob es darauf ankommen kann.

Ein wesentliches Merkmal eines zu entwickelnden Algorithmus muss die Offenlegung der von diesem genutzten Kriterien und deren Gewichtung sein. Richterinnen und Richter können sich nur dann sinnvoll auf Ergebnisse einer algorithmischen Berechnung beziehen, wenn sie zumindest im Ansatz erkennen, was überhaupt berechnet wurde. Auch kann es nicht Sinn und Zweck des Algorithmus sein, weitere undurchschaubare Beurteilungsspielräume zu schaffen. Dient ein Algorithmus nämlich dazu, Kriterien zu veranschaulichen und einheitliche Wertungsmaßstäbe zu schaffen, dann kann dies nur erfolgen, wenn die Kriterien und Maßstäbe offen und einsehbar dargelegt werden. Zudem sollte der Einsatz eines Algorithmus nicht dazu führen, dass sich Richterinnen und Richter in der Entscheidungsfindung eingeengt fühlen. So muss immer die Möglichkeit bleiben, bewusst von dem Er-

⁶⁷ Siehe Fn. 66.

⁶⁸ BGH NSTZ 1995, 128 (129).

⁶⁹ BGH NSTZ 2018, 533 (534).

⁷⁰ BGH NSTZ 2019, 657.

⁷¹ *Kühl* (Fn. 43), § 46 Rn. 33.

gebnis der Berechnung abzuweichen, wie es auch im Rahmen von gutachterlichen Einschätzungen möglich ist.

V. Fazit

Die aufgezeigten Erwägungen zeigen, dass im Bereich der Digitalisierung und des damit verbundenen Einsatzes von Computerprogrammen im Strafverfahren große Herausforderungen bestehen. Neben den Risiken kann der Einsatz von Algorithmen jedoch auch Potential bieten. So könnten Abweichungen in „Regelfällen“ oder „Strafmaßtabellen“ vielleicht vermieden oder zumindest verringert werden. Dennoch sollte zum jetzigen Zeitpunkt ein gewisses Maß an richterlicher Autonomie in der Entscheidungsfindung verbleiben. So erscheint es nötig, trotz aller Undurchsichtigkeit des Beurteilungsspielraumes, auf persönliche Erfahrungen und menschliche Empathie der Richterinnen und Richter bei der Urteilsfindung zurückzugreifen. Erfahrungsgemäß werden zwei Fälle, wie ähnlich gelagert sie auch sein mögen, niemals identisch sein. Nach dem jetzigen Stand der Technik kann nicht davon ausgegangen werden, dass es einem Algorithmus gelingen wird, jede noch so kleine Differenz auszugleichen. Ein solcher Ausgleich und die von der Rechtsprechung immer wieder betonte Beurteilung des Einzelfalles muss jedoch weiterhin gewährleistet bleiben.