

Die Nutzung maschineller Lernsysteme für den Erlass verwaltungsrechtlicher Ermessensentscheidungen

Blumröder, Lucas von; Breiter, Andreas

Veröffentlichungsversion / Published Version

Zeitschriftenartikel / journal article

Zur Verfügung gestellt in Kooperation mit / provided in cooperation with:

Verlag Barbara Budrich

Empfohlene Zitierung / Suggested Citation:

Blumröder, L. v., & Breiter, A. (2020). Die Nutzung maschineller Lernsysteme für den Erlass verwaltungsrechtlicher Ermessensentscheidungen. *der moderne staat - dms: Zeitschrift für Public Policy, Recht und Management*, 13(2), 448-463. <https://doi.org/10.3224/dms.v13i2.04>

Nutzungsbedingungen:

Dieser Text wird unter einer CC BY-SA Lizenz (Namensnennung-Weitergabe unter gleichen Bedingungen) zur Verfügung gestellt. Nähere Auskünfte zu den CC-Lizenzen finden Sie hier: <https://creativecommons.org/licenses/by-sa/4.0/deed.de>

Terms of use:

This document is made available under a CC BY-SA Licence (Attribution-ShareAlike). For more information see: <https://creativecommons.org/licenses/by-sa/4.0>

Lucas von Blumröder, Andreas Breiter

Die Nutzung maschineller Lernsysteme für den Erlass verwaltungsrechtlicher Ermessensentscheidungen

Zusammenfassung

Der vorliegende Artikel greift die aktuelle Thematik der „Künstlichen Intelligenz“ in der öffentlichen Verwaltung auf und geht der Frage nach, ob maschinelle Lernsysteme (ML-Systeme) genutzt werden können, um beim Erlass verwaltungsrechtlicher Ermessensentscheidungen zu unterstützen. Ausgehend von einer synergetischen Betrachtung beider Gebiete sowie der theoretischen Herleitung der grundsätzlichen Modellierbarkeit von Ermessensentscheidungen durch ML-Systeme schließt sich die Frage an, in welcher Form eine Unterstützung vorliegend möglich ist. Im Anschluss wird eine Modellierung anhand zweier sozialrechtlicher Ermessensentscheidungen konstruiert und in einem Wizard-of-Oz-ähnlichen Praxisversuch geprüft, wie derartige Ausgaben von Mitarbeiterinnen und Mitarbeitern in der Verwaltung wahrgenommen werden und deren Entscheidungsprozess beeinflussen können. Im Ergebnis hat sich gezeigt, dass eine theoretische Modellierung entsprechender Ermessensentscheidungen in der Praxis mit einer Vielzahl an Herausforderungen verbunden ist. Kann diesen jedoch begegnet werden, können die Anwenderinnen und Anwender von einem intelligenten System profitieren.

Schlagerworte: Digitalisierung, Verwaltungsmodernisierung, Künstliche Intelligenz, Machine Learning, Verwaltungsrecht

Abstract

Using machine learning systems for discretionary decision making in administrative law

The following article examines and discusses if and how machine learning systems can be applied to model successfully discretionary decision-processes in German administrative law. By analyzing both fields and showing their synergetic relation – the assumption can be derived that machine learning systems are suitable to model discretionary decision-making processes in administrative law. Subsequently, the article analyzes how and under which circumstances these systems can be applied to support decision-making. For this, a wizard-of-oz-experiment was developed that appeared to predict the likelihood of a request being lawful and expedient on a legal basis in social law. Test subjects using the system were more likely to make a decision instead of the ones not using it (1), spent more thought into their decision (2) and actively negated the cases that were wrongfully predicted (3). The article shows that using a machine learning system to model a discretionary decision-making process comes with a multitude of challenges that have to be dealt with. If the system bears these challenges, executive officers and customers can benefit from a machine learning system.

Keywords: Digitalization, Administrative Modernization, Artificial Intelligence, Machine Learning, Administrative Law

1 Die „natürliche“ Synergie von Recht und IT

Spätestens seit der Publikation der Eckpunkte für die Strategie „Künstliche Intelligenz“ (KI) der Bundesregierung am 18. Juli 2018 ist „KI“ aus den Fachdebatten der Verwaltungsmodernisierung kaum wegzudenken. Dem Einsatz intelligenter Systeme in der öffentlichen Verwaltung wird dabei das Potenzial zugeschrieben, die bisherige Aufgabenorganisation und -wahrnehmung wesentlich neu aufzustellen und den Wandel der Digitalisierung zu beschleunigen, ja disruptiv in der Gänze zu verändern. Wie genau dieser Wandel aussehen wird, ist allerdings in der Praxis noch nicht recht ersichtlich.

Grundsätzlich ist die öffentliche Verwaltung für die Automatisierung bestehender Prozesse durch informationstechnische Systeme prädestiniert – ihr Handeln beruht immer auf einem formellen Programm: einer Rechtsgrundlage. Diese impliziert zumeist eine konditionale, also eine sich bedingende Struktur, die einer „Wenn-Dann“-Logik entspricht (Demaj, 2018, S. 134). Diese Konditionalität des auf einer Rechtsgrundlage basierenden Verwaltungshandelns kann z. B. die Vollautomatisierung von Verwaltungsverfahren durch den § 35a Verwaltungsverfahrensgesetz (VwVfG) ermöglichen.

Die Nutzung sogenannter „intelligenter“ Systeme ist in konditionalen Verwaltungsverfahren für das Treffen einer Rechtsfolgeentscheidung nicht nötig, weil es solchen Prozessen an der Entscheidungskomponente hinsichtlich einer zu erlernenden Entscheidungspräferenz selbst mangelt. Gleichermäßen basiert nicht alles staatliche Handeln auf einer konditionalen Rechtsgrundlage. Das beruht auf der Tatsache, dass es eine Vielzahl von Sachverhalten gibt, welche nicht zwangsweise unter äquivalente Fallparameter einer konditionalen Rechtsgrundlage zu subsumieren sind und nicht bei allen Fällen, auf welche die gleichen Tatbestandsvoraussetzungen zutreffen, eine einzige Rechtsfolge für die Verfolgung des Zwecks der Ermächtigungsgrundlage förderlich ist. Daher gibt es unter anderem das verwaltungsrechtliche Ermessen, welches nicht konditional ist und dem zuständigen Entscheider einen Ermessensspielraum hinsichtlich der Wahl der Rechtsfolge einräumt (Schmidt-Aßmann, 2018, Rn. 189). An diesem Punkt stellt sich die Frage, ob der Einsatz intelligenter Systeme für eine Modellierung von Ermessensentscheidungen geeignet ist, um auf Grundlage eines geeigneten Datensatzes eine Entscheidungspräferenz zu lernen und auf neue Fälle zu replizieren (Auperle, Langkabel & Ramsauer, 2018, S. 5). Diese Frage fügt sich ein in die von der Bundesregierung in der Strategie aufgeworfene Notwendigkeit, zunächst überhaupt zu erörtern, inwieweit Einsatzmöglichkeiten von Künstlicher Intelligenz in der öffentlichen Verwaltung gegeben sind (Bundesregierung, 2018, S. 9).

2 Die Funktionsweise intelligenter Systeme in der Verwaltungspraxis

Die Terminologie der „Künstlichen Intelligenz“ bezeichnet eine breitere Forschungsdisziplin der Informatik, welche sich mit der Frage beschäftigt, wie „Maschinen“ zu intelligentem Handeln wie bspw. Lernen oder der Problemlösung befähigt werden können (Executive Office of the President, 2016, pp. 6 f.). Hierbei gibt es unterschiedliche Forschungsgebiete: Deduktionssysteme, wissensbasierte Systeme, Musteranalyse und Mustererkennung, Robotik oder intelligente multimodale Mensch-Maschine-Interaktion.

Eine terminologische Differenzierung zum Begriff der Künstlichen Intelligenz ist die „Augmented Intelligence“ (bspw. Jain, Padmanbhan, Pavlou & Santanam, 2018, S. 250). Ziel der Augmented Intelligence ist es, durch die Kombination von Datenwissenschaft, dem maschinellen Lernen und der menschlichen Intelligenz in der Entscheidungsfindung – hier der Rechtmäßigkeit des Verwaltungshandelns – zu profitieren. Wesentliches Unterscheidungskriterium beider Begriffe soll dabei die Art der Anwendung sein: Die Terminologie der Künstlichen Intelligenz impliziert demzufolge ein autonomes, eigenständiges Handeln der Anwendung. Die Augmented Intelligence hingegen setzt auf die Synergie von Mensch und Maschine, wobei die Augmented Intelligence die menschliche Entscheidungsinstanz unterstützt und zu besseren Entscheidungen befähigt (Rouse & Spohrer, 2018, pp. 11 ff.). Die Diskussion um diese begriffliche Differenzierung ist noch vergleichsweise neu (Jain, Padmanbhan, Pavlou & Santanam, 2018) und ähnelt der Differenzierung zwischen schwacher und starker Künstlicher Intelligenz (Kerns, 2017). Für den folgenden Artikel wird daher weiterhin der Begriff der Künstlichen Intelligenz verwendet.

Das maschinelle Lernen bezeichnet eine Methode der Künstlichen Intelligenz, in welcher Systeme befähigt werden, eigenständig aus Erfahrungen zu lernen und Wissen zu generieren, welches auf neue Fälle angewendet werden kann. So definiert Tom M. Mitchell das maschinelle Lernen wie folgt: „A computer program is said to learn from experience E with respect to some task T and some performance measure P, if its performance on T, as measured by P, improves with experience E“ (Mitchell, 1997, p. 2). Wird folgend von einem „intelligenten System“ gesprochen, bezieht sich der Terminus auf ebendieses maschinelle Lernsystem.

Grundsätzlich werden informationstechnische Systeme durch die Nutzung von maschinellem Lernen befähigt, auf der Grundlage vorhandener Datenbestände und Algorithmen Muster und Gesetzmäßigkeiten zu erkennen, um einen bestimmten Output vorherzusagen. Dazu wird ein intelligentes System auf der Grundlage eines Datensatzes trainiert. Im Rahmen des Trainingsprozesses wird das System so verbessert, dass sich die Fehlerquote der Ausgabe, in Mitchells Definition mit P bezeichnet, verringert, bis es optimaler Weise an einem globalen Minimum angelangt – also dem Punkt, an dem das Modell so oft wie möglich richtigliegt (Schutt & O’Neill, 2013, p. 124). Das Modell kann dann auf Grundlage der trainierten Systematik eine bestimmte Ausgabe für neue Fälle, die nicht im Trainingsdatensatz waren, vorhersagen.

In einer hypothetischen verwaltungsrechtlichen Praxisanwendung könnte die individuelle Betrachtungseinheit im Datensatz einen verwaltungsrechtlichen Fall darstellen, in welchem die Frage ist, ob ein Verwaltungsakt recht- und zweckmäßig (1) oder unrecht- bzw. unzulässig (0) ist. Das System lernt dann die Zusammenhänge der einzelnen Parameter, die die Recht- und Zweckmäßigkeit bedingen, im Rahmen eines iterativen „Trainingsprozesses“ und wiederholt diesen, bis das System in der Replikation des Outputs so oft wie möglich richtigliegt.

Intelligente Systeme lernen damit nicht einfach vorhandene Beispiele auswendig, sondern erkennen Muster und Gesetzmäßigkeiten in den vorliegenden Trainingsdaten, die zuvor durch Menschen bewertet worden sind. Diese Funktionsweise eröffnet somit die Möglichkeit, auch in gering strukturierten Entscheidungsverfahren ein Muster in den vorliegenden Fallparametern zu erkennen und auf deren Grundlage das System zu trainieren. Gerade in Verfahren mit einer hohen Fallzahl und einem zumindest minimalen Grad der Strukturierung könnten maschinelle Lernsysteme durch dem Folgen eines

auf dem Datensatz basierenden Modells zu einer strukturierteren Bescheidung und einer damit einhergehenden Rationalisierung der Entscheidungsbegründung führen (Demaj, 2018, S. 135 f.).

Praktische verwaltungsrechtliche Beispiele gibt es in der deutschen Verwaltung noch sehr wenige. In der Schweiz wird ein intelligentes System eingesetzt, das „lernt“, welche Geflüchteten mit welchen Eigenschaften bzw. Parametern in welchen Bezirken des Landes wie erfolgreich in den Arbeitsmarkt integriert wurden (Bansak et.al., 2018a, pp. 325 ff.). Die Ausgangsannahme dabei ist, dass bestimmte Menschen mit bestimmten Charakteristiken, Fähigkeiten, Eigenschaften und auch bestimmter Nationalitäten in manchen Bezirken eine höhere Integrationswahrscheinlichkeit haben als in anderen. Das ML-System lernt also einen Zusammenhang zwischen den individuellen Faktoren und der erfolgreichen Integration in den Arbeitsmarkt, um die Wahrscheinlichkeit der Integration zu erhöhen.

Als relevante Input-Faktoren haben die Forscherinnen und Forscher drei Arten von Faktoren ausgemacht, welche in signifikantem kausalem Zusammenhang mit der Integrationswahrscheinlichkeit stehen sollen: Der geographische Kontext, die persönlichen Charakteristiken sowie die Synergien zwischen beiden; Einzelparameter sind dabei bspw. das Alter, Geschlecht, Herkunftsland, Ankunftszeitpunkt, Ankunftsmonat oder die Religion (Bansak et al., 2018b, p. 12). Das System wurde im Rahmen des überwachten ML – einer Art des maschinellen Lernens, in welcher die Ergebnisse im Trainingsdatensatz bereits gegeben sind – auf Grundlage eines historischen Datensatzes mit Geflüchteten seit dem Jahr 2014 trainiert, wobei die einzelne Betrachtungseinheit x_i der individuelle Einzelfall eines Geflüchteten bzw. einer Geflüchteten bezeichnet. Weitere Beispiele der Anwendung finden sich in der Verwaltung beispielsweise bei erkennungsdienstlichen Maßnahmen nach § 16 AsylG (Frank, 2018) oder dem „Predictive Policing“ (Gerstner, 2017).

Darüber hinaus durchdringt die „Künstliche Intelligenz“ stetig weitere Bereiche des öffentlichen Bereichs, so zum Beispiel im Gesundheitswesen in der Krebsprognostik und -diagnostik (Korou, Exarchos, Exarchos, Karamouzis & Fotiadis, 2015), im Bereich der Risikoprognostik bei der Erstellung einer positiven bzw. negativen Bewährungsprognose (Kerbs, Jones & Jolley, 2009) oder bei Prognosen zur Einschätzung von Misshandlungs- und Vernachlässigungsrisiken (Schwartz, York, Nowakowski-Sims & Ramos-Hernandez, 2017). Diese Anzahl an intelligenten Anwendungen in einzelnen Fachverfahren wächst stetig. Gleichermäßen mangelt es in der Breite noch an einem systematischen Zugang intelligenter Systeme in die bestehenden Strukturen und Prozesse der öffentlichen Verwaltung (Wirtz & Müller, 2019, p. 1096).

3 Rechtsgrundlagen verwaltungsrechtlichen Ermessens im Einzelfall

Aufbauend auf diesen Erkenntnissen ist zu eruieren, ob maschinelle Lernsysteme als Methode der Künstlichen Intelligenz aufgrund ihrer Funktionsweise dazu geeignet sind, verwaltungsrechtliche Ermessensentscheidungen zu modellieren.

Aus dem deutschen Recht kommend kann man Ermessen immer als diejenige Entscheidungssituation definieren, bei welcher die entscheidungsbefugte Person nach formellem Programm unter mehr als einer Handlungsalternative wählen kann, wobei auch

das Unterlassen einer Handlung eine Alternative darstellt (Lautmann, 2011, S. 129). Die Wahl der Entscheidungsalternative ist jedoch niemals frei und unterliegt Anforderungen der Recht- und Zweckmäßigkeit (bspw. § 40 VwVfG). Gleichermäßen impliziert die Wahl der wie auch immer gearteten „besten“ Handlungsalternative in einer bestimmten Situation auch immer eine subjektive Wertung des Entscheidenden, welche auf den individuellen Eigenschaften und Erfahrungen des Individuums beruht. Die letztliche Ermessensentscheidung muss dann transparent darlegen, wie sich die Entscheidung begründet. Für eine recht- und zweckmäßige Entscheidung ist eine Abwägung öffentlicher Belange und der Interessen Einzelner sachgerecht durchzuführen. Diese Abwägung impliziert die Ermittlung aller tatbestandsrelevanten Umstände und somit eine Einzelfallentscheidung. So kann das Gewicht der abzuwägenden Belange je nach Kontextabhängigkeit variieren; auch können die Gegebenheiten des Einzelfalls es erfordern, unzumutbare Maßnahmen zurückzustellen. Für die Zweckmäßigkeit sind unter anderem einerseits „Ausmaß, Schnelligkeit und Sicherheit mit denen die gesetzliche Regelung umgesetzt wird und andererseits der dazu erforderliche Verwaltungsaufwand sowie positive oder negative Nebenfolgen sowie Gerechtigkeitserwägungen relevant“ (Sachs, 2018, Rn. 14). Gelingt es, diese Entscheidungserfordernisse und –gründe für ein bestimmtes Verfahren zu parametrisieren und eine Datenbank aufzubauen, ist das Trainieren eines intelligenten Systems vorstellbar, um einen bestimmten Output auszugeben. Diese Parametrisierung scheint vorliegend möglich: Handelt es sich beispielsweise um eine Rücknahme eines Verwaltungsaktes, bei welcher der Öffentlichkeit ein Schaden entstanden ist, stellt die Schadenssumme einen nominal skalierten Fallparameter einer Einzelfallentscheidung dar. Handelt es sich um die Frage, wer bei der Rücknahme eines Verwaltungsaktes Schuld an dessen ursprünglicher Rechtswidrigkeit hat, lässt sich dieser Tatbestand in eine binäre Skalierung ($y_i \in \{0, 1\}$) transformieren. Schon die Ausführung der „Ermittlung aller tatbestandsrelevanten Umstände“ lässt darauf schließen, dass es in vielen Verfahren eine Vielzahl von Tatbeständen gibt, welche für eine recht- und zweckmäßige Ermessensausübung zu berücksichtigen sind. Und aus der Erfordernis der recht- und zweckmäßigen Ausübung von Ermessensentscheidungen nach § 40 VwVfG und gleichzeitig der Vielzahl von potenziell relevanten Tatbeständen, welche die gesuchte Entscheidung beeinflussen, lässt sich die Hypothese ableiten, dass es in einem verwaltungsrechtlichen Datensatz einer Ermessensentscheidung grundsätzlich einen Grad der Strukturierung gibt – denn eine Zufallsverteilung würde Willkür bedeuten. Auf der anderen Seite gibt es nur einen geringen Grad der Strukturierung, welcher im Wesentlichen abhängig von der Vielzahl der entscheidungsrelevanten Parameter eines individuellen Verfahrens ist. Und da intelligente Systeme auch in Verfahren mit einem geringen Grad der Strukturierung geeignet sind, auf Grundlage dieser geringen Strukturierung zu lernen und ein statistisches Modell zu generieren, welches auf neue Fälle repliziert werden kann, lässt sich die hinreichende Annahme ableiten, dass ML-Systeme grundsätzlich für die Modellierung von Ermessensentscheidungen geeignet sind.

Gleichermäßen ist zunächst nicht anzunehmen, dass ML-Systeme in Ermessensentscheidungen mittelfristig die entscheidungsbefugte Person vollständig ersetzen werden. Unabhängig von der Debatte um die rechtliche Zulässigkeit einer entsprechenden Anwendung – gem. § 35a VwVfG ist eine Vollautomation von Verwaltungsentscheidungen mit Ermessen ausgeschlossen – dürfte sich vor dem Hintergrund der rechtlichen Erfordernisse einer Ermessensentscheidung eine Vollautomation zumindest mittelfristig als nicht möglich erweisen (zur rechtlichen Zulässigkeit intelligenter Systeme in

Ermessensentscheidungen gem. § 35a VwVfG Stegmüller, 2018, Rn. 38 f.; Stelkens et al., 2018, S. 357). Dabei dürfte weniger der Bereich der Anhörung, z. B. nach § 28 VwVfG, schwierig sein – deren maschinelle Verarbeitung über ein entsprechend trainiertes System könnte grundsätzlich in der Lage sein, die Eingaben der Anhörung in die definierten Parameter zu transformieren (Ashley, 2017, pp. 234 ff.). Doch insbesondere vor dem Erfordernis einer transparenten Entscheidungsfindung ist die Entscheidung rechtslinguistisch sicher abzubilden und transparent darzulegen. Aus den individuellen Entscheidungsparametern durch ein maschinelles Lernsystem eine rechtslinguistisch korrekte Einzelfallbegründung zu konstruieren, die sich wiederum aus dem konstruierten statistischen Modell ableitet, dürfte zumindest mittelfristig nicht möglich sein. Verschiedene Ermessenskriterien wie die Zumutbarkeit einer Maßnahme dürften im Rahmen der Einzelfallprüfung hochgradig individuell und über eine parametrische Abbildung hinaus schwierig zu differenzieren sein. Weitere Punkte sind die Repräsentativität des Datenmaterials und die reelle Besorgnis um Einzelfallgerechtigkeit in Fällen, die in die Datenbasis bisher keinen Einklang gefunden haben (Herold, 2018, S. 463). Gerade in Verfahren mit einer hohen Anzahl möglicher Fallkonstellationen kann dies zu einer Verzerrung und letztendlich unrechtmäßigen Bescheidung führen. Für eine Vollautomatisierung wäre es ferner notwendig, die Auslegung unbestimmter Rechtsbegriffe zu parametrisieren – z. B. den Parameter der Schuld – und das System so zu konstruieren, dass es lernt, in welchem Sachverhalt das Verschulden einen bestimmten Wert hat, der wiederum in der Ermessensnorm selbst genutzt werden kann. Eine Vollautomatisierung von Ermessensentscheidungen erscheint daher zunächst nicht möglich – gleichermaßen schließt das einen komplementären Einsatz im Rahmen einer Entscheidungsassistenz nicht grundsätzlich aus.

Eine Vollautomation von Ermessensentscheidungen würde dem Grunde nach implizieren, dass diese Entscheidung sowohl zweck- als auch rechtmäßig sein müsste. Geht man von einem komplementären Einsatz aus, ist bei der Zielsetzung in der Anwendung eines ML-Systems wesentlich zwischen den beiden Begrifflichkeiten zu unterscheiden, denn eine zweckmäßige Maßnahme muss nicht zwangsläufig auch rechtmäßig sein. Je nach Konstruktion eines Modells und der entsprechenden Prädiktion kann diese Unterscheidung relevant sein.

Aus den theoretischen Grundlagen lässt sich die Annahme ableiten, dass maschinelle Lernsysteme grundsätzlich dafür geeignet sind, maschinell lesbare Ermessensentscheidungen zu modellieren bzw. bei der Rechtsanwendung von Ermessensentscheidungen zu unterstützen (also im Sinne einer „augmented intelligence“). Durch das Erlernen von Mustern und Gesetzmäßigkeiten in historischen Daten können intelligente Systeme dazu beitragen, Ermessensentscheidungen rechtsgleicher zu erlassen und individuelle Fehler in der Anwendung zu vermeiden.

4 Die theoretische Modellierung einer verwaltungsrechtlichen Ermessensentscheidung

Aufbauend auf diesen Annahmen stellen sich zwei interdependente Fragen: (1) Welche Ermessensentscheidungen eignen sich konkret für eine Modellierung und wie könnte diese in der Praxis aussehen, und (2) wie würde diese Assistenz im Einzelfall das Entscheidungsverhalten der Entscheiderinnen und Entscheider beeinflussen.

Da in der Verwaltung keine Fachverfahren bekannt sind, in denen ML-Systeme zum verwaltungsrechtlichen Erlass von Ermessensentscheidungen zum Einsatz kommen, sollten diese Erkenntnisse mit Entscheidungsträgerinnen und Entscheidungsträgern der Verwaltung diskutiert werden, um Anhaltspunkte zu praktischen Rechtsgrundlagen mit Ermessensspielraum zu gewinnen, in denen eine Modellierbarkeit möglich sein könnte. Dazu wurde auf Grundlage der theoretischen Annahmen der Forschungsarbeit ein halb-strukturierter Interviewleitfaden um die zentralen Hypothesen gebildet. Die Interviews wurden transkribiert und im Rahmen einer qualitativen Inhaltsanalyse nach Philipp Mayring (Mayring, 2010, S. 601 ff.) durch eine einheitliche Kodierung ausgewertet. Aufgrund des sukzessiven Vorgehens bei der Auswahl der Interviewpartnerinnen und Interviewpartner wurde der Interviewleitfaden aufbauend auf den generierten Erkenntnissen fallbezogen angepasst.

Als erster Interviewpartner konnte der Leiter der Abteilung Digitalisierung einer Landesverwaltung gewonnen werden. Dieser verwies unter anderem auf Ermessensentscheidungen im Bereich der Sozialverwaltung, in denen es auch eine hohe Fallzahl gäbe. Aufbauend auf diesen Erkenntnissen und den sich daraus ergebenden Hinweisen wurden weitere Interviewpartner aus dem Bereich der Sozialverwaltung generiert – einen Direktor eines Sozialgerichts wie auch einen Lehrbeauftragten einer Verwaltungsschule, die letztlich verschiedene Rechtsgrundlagen des Sozialrechts mit Ermessensspielraum nannten, die bezüglich einer Modellierung zu untersuchen wären. Die genannten Entscheidungen betrafen vor allem den § 16 SGB II sowie §§ 44-48 SGB X. Insgesamt konnte in den Interviews festgestellt werden, dass eine Entscheidungsassistenz im Sinne der Augmented Intelligence in der Praxis sehr hilfreich sein könnte, dass die genannten Rechtsgrundlagen unterschiedlich stark strukturiert sind und dass deren Fallparameter sehr unterschiedlich gelagert sind. Die Expertinnen und Experten führten auch aus, dass eine Vollautomation von Ermessensentscheidungen durch eine KI als Entscheidungsinstanz in der Praxis nicht vorstellbar wäre – aus Gründen der Transparenz, der Potenzierung von Fehlern, der anwenderbezogenen Perzeption und der Akzeptanz der Kundinnen und Kunden. Das Ziel der Interviews, die theoretischen Annahmen der Untersuchung zu erörtern und Rechtsgrundlagen mit Ermessen genannt zu bekommen, anhand derer die Annahmen geprüft werden können, konnte erreicht werden.

Aufbauend auf den Erkenntnissen der Interviews wurden beide Rechtsgrundlagen aufbauend auf den theoretischen Annahmen der Untersuchung geprüft. § 16 SGB II bezeichnet die Gewährung von Leistungen zur Eingliederung in Arbeit. Das Ziel des Paragraphens ist die möglichst schnelle und wirksame Eingliederung von Arbeitsuchenden in den Arbeitsmarkt, um die individuelle Hilfsbedürftigkeit zu überwinden (Luik et al., 2017, S. 695). Die formale Ausübung und Überprüfung des Ermessens folgen den o. g. Kriterien. Welche Maßnahme zur Erreichung des Ziels die „zweckmäßigste“ Handlungsalternative darstellt, obliegt dem Ermessen des Sachbearbeiters bzw. der Sachbearbeiterin. Für die Bewilligung einer Maßnahme im Rahmen der Ermessensausübung ist eine dokumentierte Prognoseentscheidung über die erforderlichen Leistungen zur Eingliederung in Arbeit unter Berücksichtigung der Grundsätze von Wirtschaftlichkeit und Sparsamkeit nach § 3 Abs. 1 S. 4 i.V.m. § 14 Abs. 4 SGB II von Nöten. Diese Prognoseentscheidung ist unter Würdigung sämtlicher Umstände des Einzelfalls und der Potenzialanalyse vorzunehmen; insoweit entspricht der Abwägungsvorgang der Einzelfallentscheidung einer rechtmäßigen Ermessensausübung (Bundesagentur für Arbeit, 2018, S. 1).

Die Annahme der Anwendbarkeit maschineller Lernsysteme für das Treffen von Ermessensentscheidungen in diesen Verfahren ist vor dem Hintergrund ihrer Konzeption kritisch zu hinterfragen. Für ein erfolgreiches Training des Systems – also für eine bestmögliche Abbildung der Gesetzmäßigkeiten der vorliegenden Daten – ist dafür relevant, dass die Daten in X , also die Anzahl der individuellen Parameter einer Betrachtungseinheit, auch in kausalem Zusammenhang zum Output stehen. Fraglich ist dabei, welche Parameter diesen „Erfolg“ der aktiven Arbeitsaufnahme kausal bedingen würden. So ist vorstellbar, dass die im Rahmen eines maschinellen Lernmodells prädiktierete Outputvariable bei zwei statistischen Zwillingen X_1 und X_2 unterschiedlich ist, weil nicht im Datensatz vorhandene bzw. nicht zu parametrisierende Entscheidungsdeterminanten den „Erfolg“ wesentlich beeinflusst haben. Denkbar wäre hier bspw. die subjektive Motivationslage oder die psychische bzw. körperliche Verfassung. Dabei steht fest, dass die Motivation einer ALG2-berechtigten Bezugsperson ganz wesentlich zum Eingliederungserfolg beitragen kann (Luik et al., 2017, S. 705 f.), diese jedoch als Parameter für das Trainieren eines ML-Datensatzes mathematisch kaum korrekt skaliert darstellbar ist. Auch bleibt offen, ob die Berücksichtigung statischer Einflussvariablen wie dem Alter, dem Betreuungsbedarf und bestimmter biographischer Merkmale vor dem Hintergrund des Art. 3 GG rechtlich zulässig wäre, vor allem, wenn diese in einem eher ungewissen Zusammenhang zum Maßnahmeerfolg stehen.

Hierzu sei angemerkt, dass auch das oben angeführte Schweizer System anhand von statischen Daten wie dem biografischen Hintergrund lernt und die Motivation des Individuums dort keinen Einfluss auf dessen Integrationswahrscheinlichkeit hat; vielmehr wird diese für die Prädiktion der Integrationswahrscheinlichkeit selbst als gleich angenommen. Diese Annahme dürfte im Verfahren des § 16 SGB II nicht übertragbar sein, weil Hintergrund und Zielsetzung beider Systeme sich unterscheiden.

Anders verhält es sich jedoch bei der Rücknahme rechtswidrig begünstigender Verwaltungsakte nach § 45 SGB X. Die Norm beinhaltet Fälle wie die rückwirkende Aufhebung von am Datum der Ausstellung bzw. Bewilligung rechtswidriger Verwaltungsakte. Die Vorschrift bezweckt einen Ausgleich zwischen dem individuellen Bestandsschutzinteresse und dem öffentlichen Interesse an der Durchsetzung der materiell rechtlich zutreffenden Rechtslage bei rechtswidriger Begünstigung von Leistungsempfängerinnen und -empfängern (Schütze, Bieresborn & Wulffen, 2014, S. 507). Demzufolge kann ein Verwaltungsakt zurückgenommen werden, wenn das öffentliche Interesse der Verwaltung an der Rücknahme eines rechtswidrig begünstigenden Verwaltungsakts ein schutzwürdiges Interesse des Betroffenen an der Aufrechterhaltung überwiegt (ebd.). Dabei ist anzuführen, dass die von der Norm geforderte Interessenabwägung einer Ermessensabwägung ähnlich ist, von dem in § 45 SGB X eingeräumtem Ermessen aber zu unterscheiden ist, da diese Abwägung Teil der Tatbestandsvoraussetzungen ist und damit der vollen verwaltungsgerichtlichen Kontrolle unterliegt (Padé, 2017, S. 542). Dabei überschneiden sich die Abwägungsvorgänge, sodass es in der Ermessensabwägung eine gewisse Deckungsgleichheit zur Vertrauensschutzabwägung nach § 45 SGB X Abs. 2 S. 1 und 2 vorliegt. Aufgrund der Deckungsgleichheit der Interessenabwägung nach § 45 Abs. 2 SGB X mit der Ermessensabwägung einer in dem Sinne typischen verwaltungsrechtlichen Ermessensentscheidung wird für die Modellierung eines Assistenzsystems folgend nicht gesondert differenziert.

Die Frage einer Entscheidung in Verfahren nach § 45 SGB X ist, ob ein rechtswidrig begünstigender Verwaltungsakt mit Wirkung für die Zukunft und/oder mit Wirkung

für die Vergangenheit zurückgenommen werden kann. Dafür sind im Rahmen der Interessenabwägung verschiedene rechtliche Güter maßgeblich, welche für das Trainieren eines intelligenten Systems maßgeblich sind, so u. a.:

- Die Schadenssumme, denn je höher der Schaden ist, desto höher ist das Interesse der Allgemeinheit an der Wiederherstellung des gesetzmäßigen Zustands (Padé, 2017, S. 542),
- Die Dauer der Bestandskraft, denn die Stellung des rechtswidrig Begünstigten wächst mit zunehmendem zeitlichen Abstand vom Zeitpunkt der Leistungsbewilligung; gleichermaßen ist eine kurze temporäre Diskrepanz nicht vertrauensschutzbe gründend und schwächt die Position des Begünstigten (Schütze, Bieresborn & Wulfen, 2014, S. 524).
- Das Verschulden, denn ein leicht zu vermeidender, grober Fehler der Verwaltung bei Erlass des Verwaltungsaktes kann das Vertrauen des Begünstigten in die Bestandskraft der Leistungsbewilligung nachhaltig stärken (ebd.).

Weitere, die Entscheidung beeinflussende Faktoren sind bspw. eine Dauerwirkung, die Zumutbarkeit, der Verwaltungsaufwand, der Verlust weiterer Sozialleistungen, eine etwaig perpetuierende Bestärkung durch (fehlerhaftes) Behördenhandeln, ein etwaiger Verbrauch bzw. eine Disposition der Leistungen oder unzumutbare Nachteile bei der Rückabwicklung (ebd.).

Systematisiert und klassifiziert man historische Daten von Verwaltungsverfahren nach § 45 SGB X entsprechend der genannten kausalen Parameter mit der gesuchten Outputvariable y , also der Rechtmäßigkeit einer Rücknahme, würde so ein mathematisches Modell konstruiert, welches auf neue Fälle übertragen werden könnte; die individuelle Betrachtungseinheit x_i ist dabei eine vollzogene Rücknahme in einem Fall mit differierenden Parametern, welche entweder rechtmäßig oder rechtswidrig ist. Der Output ist nicht die Wahrscheinlichkeit, wie oft in ähnlichen Fällen gleich entschieden wurde, sondern vielmehr die, dass eine Rücknahme vollzogen wurde und diese entweder recht- und zweckmäßig ist oder nicht – also eine binäre Skalierung. Das lässt sich dadurch begründen, dass nicht jede erfolgte Rücknahme recht- und zweckmäßig ist, und sich durch die Nutzung einer Datenbank aus bestandskräftigen Verwaltungsakten eine fehlerhafte Rücknahmep Praxis potenzieren könnte. Ferner geht es vorliegend um eine Entscheidungsassistenz, welche einem Sachbearbeiter bzw. einer Sachbearbeiterin die Information liefern soll, ob im vorliegenden Sachverhalt X_1 bspw. eine Rücknahme recht- und zweckmäßig ist oder nicht. Dafür hilft die Ausgabe von Wahrscheinlichkeiten im Rahmen eines probabilistischen ML-Systems nicht weiter. Außerdem bedingt bereits die Eröffnung verwaltungsrechtlichen Ermessens eine gewisse Entscheidungsfreiheit der Sachbearbeiterinnen und Sachbearbeiter, sodass es bei der Nutzung bestandskräftiger Verwaltungsakte dazu führen dürfte, dass es eine Vielzahl von Einzelfällen mit gleicher Parameterkonstellation gibt, in der unterschiedlich entschieden wurde, was die Exaktheit der Ausgabe negativ beeinflussen würde.

Für die Modellierung wäre es ferner nötig, die Parameter maschinell lesbar abbilden zu können. Bei der Schadenssumme, der Dauer der Bestandskraft oder der Dauerwirkung ist dies nominal möglich. Weitere Parameter wie die Frage des Verschuldens, der wertmäßigen Aufzehrung oder der Vertrauensstärkung durch perpetuierendes Behördenhandeln könnten – sehr vereinfacht – binär i. S. von Ja = 1 oder Nein = 0 transformiert werden. Gleichermäßen würde es für die adäquate Beurteilung der jeweiligen

binären Zuordnungen der Fälle einer weiteren Ausdifferenzierung entsprechender Subsumtion der jeweiligen Verhaltensweisen unter die jeweiligen unbestimmten Rechtsbegriffe bedürfen.

Somit könnten Sachbearbeiterinnen und Sachbearbeiter eine Ausgabe erhalten, ob in ihrem vorliegenden Fall mit den jeweiligen Parametern eine Rücknahme recht- und zweckmäßig ist oder nicht. Dabei wären in der Praxis verschiedene Herausforderungen zu bewältigen, auf die zum Abschluss genauer eingegangen wird. Insofern lässt sich festhalten, dass der § 45 SGB X für eine Modellierung durch ein intelligentes System geeignet wäre. Diese Art der Modellierung impliziert aber gleichermaßen eine Prädiktion der wahrscheinlichen Rechtmäßigkeit im Einzelfall. Daher ist auch die Art der Assistenz als Entscheidungsunterstützung im Abgleich mit historischen Fällen klar definiert.

Nun stellt sich die Frage, ob eine derartige Anwendung bzw. maschinelle Ausgabe zur Recht- und Zweckmäßigkeit eines Einzelfalls tatsächlich die Rechtssicherheit erhöhen oder die Verfahren beschleunigen würde. Komplexe Fälle bleiben vor allem bei hohem Arbeitsdruck länger liegen – kann diese Tatsache durch eine derartige Ausgabe verringert werden? Diese Fragestellung impliziert, dass die Forschungsarbeit auch als „Augmented Intelligence“ eingeordnet werden kann (siehe oben).

In einem sozialwissenschaftlichen Experiment mit zwei Gruppen – einer fünfköpfigen Kontroll- und einer fünfköpfigen Versuchsgruppe – wurden Mitarbeiterinnen und Mitarbeiter einer Sozialverwaltung zu zehn verschiedenen Fällen zum §45 SGB X befragt. Die Probandinnen und Probanden konnten im Rahmen einer Interessenbekundung in einem Sozialamt gewonnen werden. Es handelte sich also vorliegend um Praktikerinnen und Praktiker, die unregelmäßig mit der Rückforderung rechtswidrig gewährter Leistungen konfrontiert sind bzw. waren. Die Anzahl von zehn Probandinnen und Probanden wird in Untersuchungen von Think-Aloud-Ansätzen, insbesondere im Bereich der Human-Computer-Interaction als angemessene Größe für eine entsprechende Auswertung bedacht (bspw. Eccles & Aarsal, 2017, pp. 514 ff.). Die Empirie der quantitativen Auswertung der Entscheidungspräferenzen ist aufgrund der geringen Anzahl an Testsubjekten für eine quantitative Auswertung gleichermaßen bedingt repräsentativ, kann aber die aus der qualitativen Auswertung entstehenden Annahmen bestätigen.

Die Testsubjekte erhielten jeweils zehn Sachverhaltskonstellationen zu § 45 SGB X, verbunden mit der binären Frage, ob sie im jeweiligen Fall den Verwaltungsakt mit Wirkung für die Zukunft bzw. mit Wirkung für die Vergangenheit zurücknehmen würden. Die Darstellung der Sachverhaltskonstellationen lautete bei den Probandinnen und Probanden der Versuchsgruppe wie folgt:

Systemgenerierte Ausgabe: Unter Berücksichtigung der Parameter Schadenssumme, Bestandskraft, Vermögensdisposition, Verschulden, Dauerwirkung und der Zumutbarkeit der Aufhebung ist im vorliegenden Fall eine Rücknahme (nicht) recht- und zweckmäßig. Sind darüberhinausgehende Tatbestände entscheidungsevident, ist meine Ausgabe nicht valide.

Genauigkeit der Ausgabe: 90% (in 90 von 100 Fällen ist die Ausgabe des Systems richtig)“.

Die Probandinnen und Probanden mit der Genauigkeit der systemgenerierten Ausgabe zu konfrontieren, ist für die Reflektion der Ausgabe wichtig. Sind in den Trainingsdaten beispielsweise Unstimmigkeiten bzw. ein „Rauschen“, weil vergleichbare Fälle un-

terschiedlich entschieden wurden, kann sich die Genauigkeit, mit der das System Fälle nachvollziehen kann und zur richtigen Entscheidung finden kann, reduzieren. Wichtig war beim Design der Untersuchung auch, herausfinden zu wollen, wie die Probandinnen und Probanden mit dieser Metrik umgehen. Der zweite Aspekt der systemgenerierten Ausgabe stellt die Auflistung der abschließenden Parameter dar, welche durch das System berücksichtigt wurden. Ermessensentscheidungen sind durch einen geringen Grad der Strukturierung gekennzeichnet, was wiederum die Gefahr birgt, dass Tatbestände im Trainingsdatensatz nicht berücksichtigt wurden, die aber für den zu entscheidenden Einzelfall zu berücksichtigen sind. Daher wurden die berücksichtigten Parameter aufgeführt, um zu eruieren, ob die Probandinnen und Probanden dieses System bzw. den zu entscheidenden Fall auch dahingehend analysieren.

Bei der Beantwortung der Fragebögen waren die Probandinnen und Probanden gehalten, aufgrund der Nutzung der Think-Aloud-Methode laut zu denken (Eccles & Arsal, 2017, pp. 514 ff.). Die Methode eignet sich vor allem, wenn das Ziel der Untersuchung ist, den Gedankengang der Probandinnen und Probanden zu erheben und verstehen zu können (ebd.). Diese Aufnahmen des Lautgedachten wurden anschließend transkribiert und ausgewertet, um verstehen zu können, wie die Testpersonen auf die maschinelle Ausgabe reagiert und diese in ihren Entscheidungsprozess eingebaut haben. Ein Zeitlimit, innerhalb dessen die Entscheidungen zu treffen waren, hätte den Entscheidungsprozess verzerrt und wurde deshalb nicht gegeben.

Die Auswertung beider Gruppen hat – wie erwartet – ergeben, dass eine Rücknahme von rechtswidrig begünstigenden Verwaltungsakten für die Zukunft die absolute Regel ist. Bei der rechtswidrigen zukunftsbezogenen Begünstigung kann in der Theorie offensichtlich noch keine reelle Vermögensdisposition und damit kein unmittelbarer Schaden für den Adressaten entstanden sein, der ein besonders gewichtiges Rechtsgut eröffnen würde. So kann auch eine Rücknahme nicht unmittelbar unzumutbar sein. Und so wurde in der Testgruppe der Verwaltungsakt 25 von 25-mal zurückgenommen; in der Kontrollgruppe 22-mal, wobei 3-mal „weiß nicht“ angegeben wurde. Die Rücknahme für die Vergangenheit fiel dagegen deutlich variabler aus (siehe *Tabelle 1*).

Tabelle 1: Antwortverhalten bei rücknahmebezogenen Fragen

	Versuchsgruppe	Kontrollgruppe
Ja	10	11
Nein	15	9
Weiß nicht	0	5

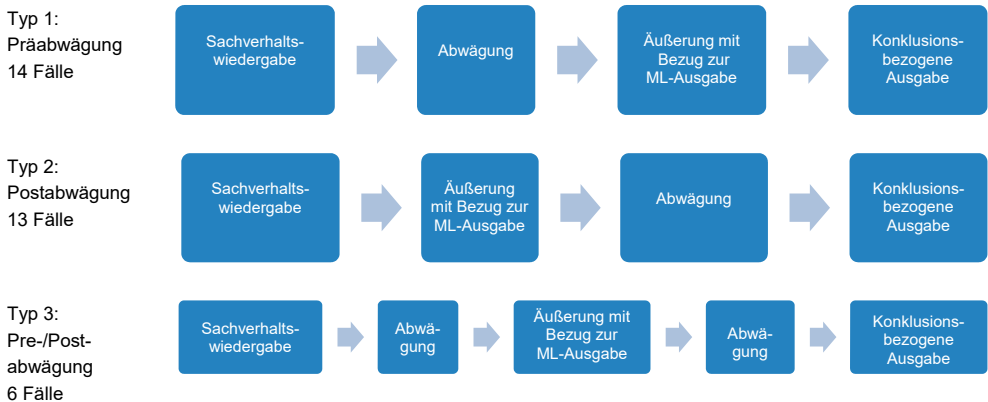
Quelle: Eigene Darstellung.

Dabei waren die Probandinnen und Probanden der Versuchsgruppe eher geneigt, eine Rücknahme für die Vergangenheit nicht zu vollziehen, während die Kontrollgruppe eher geneigt war, „weiß nicht“ als Option zu wählen. Führt man beides zusammen, fällt auf, dass in 16 Prozent der Fälle in der Kontrollgruppe „Weiß nicht“ geantwortet wurde – in der Versuchsgruppe kein Mal. Dabei zeigt sich besonders, dass die Probandinnen und Probanden in der Regel konform mit der systemgenerierten Ausgabe gestimmt haben – allerdings nur in den Fällen, in denen diese auch richtig lag. In drei der zehn Fälle mit differierender Komplexität wurde vom System eine falsche Ausgabe getätigt. Diese wurde aktiv negiert. Insgesamt ließ die Auswertung den Rückschluss zu, dass die

Ausgabe die Entscheidungsbereitschaft gestärkt und gleichermaßen zu einer kritischen Reflexion der individuellen Entscheidungspräferenz beigetragen hat.

Die qualitative Auswertung der Transkripte der Ausführungen unter Nutzung der qualitativen Inhaltsanalyse nach Mayring hat diese Ergebnisse im Wesentlichen verifizieren können – die Auswertungsmethode war aufgrund des Ziels der systematischen Bearbeitung der Kommunikation durch eine Extraktion der für relevant erachteten Daten vorliegend das geeignete Mittel (Mayring, 2010, S. 601 ff.). Im Rahmen der Methode wird ein regelgeleitetes Kategoriensystem gebildet, die einzelnen Textpassagen ihrem Wesensgehalt nach kodiert und dem Kategoriensystem zugeordnet. Anschließend wurden die Ergebnisse ausgewertet (ebd.). So kamen die Probandinnen und Probanden der Kontrollgruppe bei der Beantwortung der Fragen auf 141 auswertungsbezogene Codes, während die Versuchsgruppe auf 172 Codes kam, sich also inhaltlich deutlich intensiver mit den Fällen und der Ausgabe beschäftigt hat. Die intensivere Auseinandersetzung mit der Fragestellung und der Ausgabe kann vorliegend dazu geführt haben, dass sich die Probandinnen und Probanden der Testgruppe in ihrer Entscheidung bzw. Argumentation gestärkt sahen – und entsprechend der quantitativen Auswertung weniger geneigt waren, „weiß nicht“ anzugeben. Das lässt sich auch aus der sequenziellen Abfolge der Entscheidungsprozesse ableiten, welche sich im Wesentlichen wie folgt kategorisieren lassen (siehe *Abbildung 1*):

Abbildung 1: Abfolge der Entscheidungsprozesse bei der Berücksichtigung des ML-Systems



Quelle: Eigene Darstellung.

Gleichermaßen wurden in anderen Fällen keine Äußerungen mit Bezug zur ML-Ausgabe getätigt. Insgesamt hat sich an den Abwägungs- und Argumentationsmustern in der Versuchsgruppe deutlich gezeigt, dass die Probandinnen und Probanden die systemgenerierte Ausgabe vor allem als Reflektionsinstanz für das eigene Entscheidungsverhalten genutzt haben. Entsprach die systemgenerierte Ausgabe beispielsweise ihrer Entscheidungspräferenz, sahen sie sich gestärkt in ihrer Entscheidung und bestätigten diese. Widersprach die systemgenerierte Ausgabe ihrer Entscheidungspräferenz, wurde diese aktiv hinterfragt und neu geprüft. Das entspricht dem Typ 3 der o. g. Klassifikation. Gleichzeitig hat eine bewusst falsche Ausgabe das Entscheidungsverhalten der Probandinnen und Probanden nicht negativ verändert. Vielmehr haben sich diese be-

wusster mit der falschen Ausgabe auseinandergesetzt, diese reflektiert und sind zu der Ansicht gekommen, dass die systemgenerierte Ausgabe nicht richtig ist.

Es ist fraglich, ob diese Annahme generalisierbar ist, da es sich um je drei bewusst falsche Ausgaben bei insgesamt fünf Probandinnen und Probanden handelt, also nur um einen kleinen Datensatz. Zudem bleibt offen, welchen Einfluss die Komplexität der in Rede stehenden Fälle auf das Verhalten der Probandinnen und Probanden hatte. Insofern lässt sich nicht schlussfolgern, ob die Probandinnen und Probanden über ein ausreichendes Vertrauen in die eigenen Fähigkeiten verfügten oder über ein hinreichendes Misstrauen gegenüber der systemgenerierten Ausgabe. Darüber gaben die Transkripte des Think-Aloud-Ansatzes keinen Aufschluss. Nichtsdestotrotz hat die Untersuchung gezeigt, dass die Probandinnen und Probanden in der Breite von der Nutzung des Assistenzsystems profitiert haben. Die tiefere Auseinandersetzung mit der eigenen Entscheidung und deren Reflektion haben zu einer erhöhten individuellen Entscheidungssicherheit beigetragen. Die ermessensrelevanten Tatbestände wurden in der Gesamtschau stärker miteinander abgewogen und haben dazu beigetragen, dass sich die Probandinnen und Probanden in ihrer Entscheidungssicherheit gestärkt sahen. Die kritische Reflexion der systemgenerierten Ausgabe sprach insofern für die Mündigkeit der Probandinnen und Probanden der Testgruppe.

5 Schlussfolgerungen

Zusammenfassend lässt sich festhalten, dass intelligente Systeme grundsätzlich für eine Modellierung verwaltungsrechtlicher Ermessensentscheidungen geeignet sind und in der Praxis durchaus nützlich sein können.

Wesentliches Merkmal der Untersuchung waren allerdings die Schwierigkeiten bei der Suche nach geeigneten Fachverfahren, die sich für eine Modellierung eignen würden. Dafür ist die statistische Verarbeitung der Fallparameter zur maschinellen Lesbarkeit ein wichtiges Merkmal, welches insbesondere in sehr kundennahen Bereichen der Verwaltung oftmals schwierig ist, zumindest wenn es sich um Prognoseentscheidungen handelt.

Dann ist von hoher Relevanz, keine diskriminierenden Merkmale zu berücksichtigen und das Ergebnis nicht von in den Daten liegenden Korrelationen ohne kausalen Bezug zum Output verzerren zu lassen. Auch muss der Datensatz repräsentativ sein und alle entscheidungsrelevanten Parameter beinhalten. Ein weiterer relevanter Punkt ist die Qualität der Daten selbst. Die Nutzung bestandskräftiger Verwaltungsakte dürfte im Output zu einer eher mangelhaften Exaktheit des Systems führen, da auch nicht rechtmäßige Verwaltungsakte bestandskräftig sind und nur existieren, weil sie nicht angegriffen und aufgehoben wurden. Derartige Fälle dürften einer mathematischen Modellkonstruktion anhand der existenten Strukturierung geradezu entgegenwirken.

Ein weiterer wesentlicher Punkt ist die Klarheit über das jeweilige zu prädiktierende Ergebnis: Das Arbeiten mit Wahrscheinlichkeiten hätte keine Aussagekraft. Die Ausgabe, dass in 70% ähnlicher Fälle Handlungsalternative A_1 und in 30% der Fälle A_2 gewählt wurde, führt zu einer Verringerung der Rechtssicherheit in der Antragsbearbeitung und zu mehr Verunsicherung, und ist zumindest in der Frage der Rechtmäßigkeit eines Verwaltungsaktes kein nützliches Ergebnis. Und greift man nur auf Verwaltungsakte zurück, die gerichtlich überprüft und recht- und zweckmäßig bzw. nicht

recht- und zweckmäßig sind, stellt sich die Frage, ob eine ausreichend große Datenbasis zur Verfügung gestellt werden kann, die tatsächlich zu repräsentativen Ergebnissen führen würde. Ferner wurde im o. g. Modell der unbestimmte Rechtsbegriff des Verschuldens binär skaliert. In der Praxis dürfte es bei der Sachverhaltsdeutung häufiger um Graubereiche gehen, denen eine binäre Skalierung keine ausreichende Rechnung trägt. Auch die Auslegung unbestimmter Rechtsbegriffe muss in dem Sinne rechtssicher modelliert und parametrisch wiedergegeben werden können. Abschließend stellt sich die theoretische Frage, wie eine sich ändernde Rechtsprechung Einklang in derartig pfadabhängige Systeme findet.

Gelingt es, diesen Herausforderungen Rechnung zu tragen und ein Verfahren repräsentativ, akkurat und rechtssicher modellieren und trainieren zu können, ist davon auszugehen, dass Mitarbeiterinnen und Mitarbeiter davon in der Rechtsanwendung profitieren können. Gleichmaßen muss ein Mindestmaß an Transparenz geschaffen werden, um zu vermitteln, was solche Systeme können und tun, sowie welche Aussagekraft die Ausgabe hat. Vor allem im Bereich der Human-Computer-Interaction ergeben sich daher eine Vielzahl an weiteren Fragen: Welchen Einfluss hat beispielsweise eine interaktivere Implementation der maschinellen Ausgabe als Entscheidungsassistenz auf das Entscheidungsverhalten der Probandinnen und Probanden – z. B. in Form eines Chatbots. Gerade die Art und Weise der Implementation von Assistenzmodellen kann hier einen wesentlichen Einfluss darauf haben, wie die Entscheidung selbst beeinflusst wird. Auch stellt sich die Frage eines langfristigen Gewöhnungseffekts: Ab einer bestimmten Periode der Zusammenarbeit könnte sich ein grundsätzliches Vertrauen in die systemgenerierten Ausgaben einstellen, wenn diese oft richtigliegen. Was wiederum dazu führen könnte, dass das eigene Entscheidungsverhalten weniger hinterfragt und die vorgeschlagenen Entscheidungen übernommen werden.

Von einer kurzfristigen Umsetzung derartiger Anwendungen sind wir im Verwaltungsrecht noch weit entfernt. Dabei soll der Artikel nicht so verstanden werden, dass er die grundsätzliche Anwendbarkeit von ML-Systemen für den Erlass verwaltungsrechtlicher Ermessensentscheidungen im Verwaltungsrecht negiert – vielmehr sind mit einer ML-basierten Prozessautomatisierung bzw. -unterstützung eine Vielzahl interdisziplinärer Herausforderungen verbunden, denen Rechnung getragen werden muss. Dabei bietet eine Modellierung auch viele Chancen: Schnellere und rechtssichere Bearbeitung von Anträgen, Unterstützung der Mitarbeiterinnen und Mitarbeitern sowie eine Erhöhung der Rechtsgleichheit: Werden neue Fälle entlang der Strukturierung rechtmäßiger „Altfälle“ entschieden, impliziert dies geradezu eine Optimierung von Rechtsgleichheit und eine Minimierung von Zufall und Willkür. Bis dahin ist es aber noch ein weiter Weg.

Literatur

- Ashley, Kevin D. (2017). *Artificial intelligence and legal analytics: new tools for law practice in the digital age*. Cambridge. Cambridge University Press.
- Aupperle, Astrid, Langkabel, Tobias & Ramsauer, Katharina (2018). *Denkimpuls Digitale Ethik: Künstliche Intelligenz – Assistenz oder Konkurrenz in der künftigen Verwaltung?* Initiative D21 (Denkimpuls Digitale Ethik).
- Bansak, Kirk, et al. (2018a). Improving refugee integration through data-driven algorithmic assignment. *Science*, 359 (6373), 325-329. doi: 10.1126/science.aao4408.

- Bansak, Kirk, et al. (2018b). Improving refugee integration through data-driven algorithmic assignment – Supply Materials. *Science*, 359 (6373). doi: 10.1126/science.aao4408.
- Bundesagentur für Arbeit (Hrsg.) (2018). *Zweites Sozialgesetzbuch – Fachliche Weisungen*. § 16 SGB II Leistungen zur Eingliederung.
- Bundesregierung (2018). *Eckpunkte der Bundesregierung für eine Strategie Künstliche Intelligenz*. Verfügbar unter: https://www.bmbf.de/files/Nationale_KI-Strategie.pdf [30. März 2019].
- Demaj, Labinot (2018). Smart Government: Die Verwaltung und den Staat der Zukunft denken. *Informatik-Spektrum*, 41 (2), 123-137. doi: 10.1007/s00287-018-1098-x.
- Eccles, David & Aarsal, Güler (2017). The think aloud method: what is it and how do I use it?. *Qualitative Research in Sport, Exercise and Health*, 9 (4), 514-531. doi:10.1080/2159676X.2017.1331501.
- Executive Office of the President. (2016). *Preparing for the future of artificial intelligence*. National Science and Technology Council. Committee on Technology.
- Frank, Dorothee (2018). *Digitale Unterstützer im Asylverfahren – Wie Sprachbiometrie Asylentscheidungen auf eine noch breitere Grundlage stellt*. Bundesamt für Migration und Flüchtlinge. Verfügbar unter <http://www.bamf.de/DE/Service/Top/Presse/Interviews/20181008-interview-cybersecurity-report/interview-cyber-security-report-node.html> [17. Mai 2019].
- Gerstner, Dominik (2017). *Predictive Policing als Instrument zur Prävention von Wohnungseinbruchdiebstahl*. 1. Auflage 2017. Freiburg im Breisgau: Max-Planck-Institut für ausländisches und internationales Strafrecht (Research in brief).
- Herold, Viktoria (2018). Algorithmisierung von Ermessensentscheidungen durch Machine Learning. In Deutsche Stiftung für Recht und Informatik (Hrsg.), *Tagungsband Herbstakademie 2018* (S. 453-464). Bochum.
- Jain, Hermant, Padmanbhan, Balaji, Pavlou, Paul & Santanam, Raghu (2018). Call for Papers – Special Issue of Information Systems Research – Humans, Algorithms, and Augmented Intelligence: The Future of Work, Organizations, and Society. *Information Systems Research*, 29 (1), 250-251. <https://doi.org/10.1287/isre.2018.0784>.
- Kerbs, John, Jones, Mark & Jolley, Jennifer M. (2009). Discretionary Decision Making by Probation and Parole Officers: The Role of Extralegal Variables as Predictors of Responses to Technical Violations. *Journal of Contemporary Criminal Justice*, 25 (4), 424-441. doi: 10.1177/1043986209344556.
- Kerns, Jeff (2017). *What's the difference between weak and strong AI?*. *MachineDesign*. Verfügbar unter: <https://www.machinedesign.com/markets/robotics/article/21835139/whats-the-difference-between-weak-and-strong-ai> [30. April 2020].
- Kourou, Konstantina, Exarchos, Themis P., Exarchos, Konstantinos P., Karamouzis, M. V. & Fotiadis, D. I. (2015). Machine learning applications in cancer prognosis and prediction. *Computational and Structural Biotechnology Journal*, 13, 8-17. doi: 10.1016/j.csbj.2014.11.005.
- Lautmann, Rüdiger (2011). *Justiz – die stille Gewalt*. Wiesbaden: VS Verlag. doi: 10.1007/978-3-531-93483-9.
- Luik, Steffen et al. (Hrsg.) (2017). *SGB II: Grundsicherung für Arbeitsuchende: Kommentar*. 4. neu bearbeitete Auflage. München: C.H. Beck.
- Mayring, Philipp (2010). Qualitative Inhaltsanalyse. In Günter Mey & Katja Mruck (Hrsg.), *Handbuch Qualitative Forschung in der Psychologie* (S. 601-613). Wiesbaden: VS Verlag. doi: 10.1007/978-3-531-92052-8_42.
- Mitchell, Tom M. (1997). *Machine Learning*. New York: McGraw-Hill (McGraw-Hill series in computer science).
- O'Neill, Cathy & Schutt, Rachel (2013). *Doing Data Science. Straight talk from the frontline*. O'Reilly Media.
- Padé, Christiane (2017). § 45 SGB X. In: Bernd Mutschler & Ingo Palsherm (Hrsg.), *juris Praxis-Kommentar SGB 10*. 2. Auflage 2017. Juris Praxiskommentar. Online-Kommentar.
- Rouse, William & Spohrer, James (2018). Automating versus augmenting intelligence. *Journal of Enterprise Transformation*, 1-21. doi:10.1080/19488289.2018.1424059.

- Sachs, Michael (2018). §40 VwVfG. In Stelkens, Paul et.al. (2018). *Verwaltungsverfahrensgesetz: Kommentar*. 9. Auflage. München: C.H. Beck.
- Schmidt-Aßmann, Eberhard (2018). Art. 19 Abs. 4 Grundgesetz. In Theodor Maunz & Günter Dürig (Hrsg.), *Grundgesetz-Kommentar*. 85. Auflage. München: C.H.Beck-Verlag.
- Schütze, Bernd, Bieresborn, Dirk & Wulffen, Matthias von (Hrsg.) (2014). *SGB X: Sozialverwaltungsverfahren und Sozialdatenschutz. Kommentar*. 8., neubearb. Aufl. München: Beck.
- Schwartz, Ira M., York, Peter, Nowakowski-Sims, Eva & Ramos-Hernandez, Ana (2017). Predictive and prescriptive analytics, machine learning and child welfare risk assessment: The Broward County experience. *Children and Youth Services Review*, 81, 309-320.
doi: 10.1016/j.chilyouth.2017.08.020.
- Stegmüller, Martin (2018). Vollautomatische Verwaltungsakte – eine kritische Sicht auf die neuen § 24 I 3 und § 35 a VwVfG. *Neue Zeitschrift für Verwaltungsrecht*, 6, 353-359.
- Stelkens, Paul et.al. (2018). *Verwaltungsverfahrensgesetz: Kommentar*. 9. Auflage. München: C.H. Beck.
- Wirtz, Bernd & Müller, Wilhelm (2019). An integrated artificial intelligence framework for public management. *Public Management Review*, 21 (7), 1076-1100.
doi: 10.1080/1471903702018.1549268.

Anschriften der Autoren:

Lucas von Blumröder, M.A., Universität Bremen, Institut für Informationsmanagement, E-Mail: lvblumroeder@ifib.de.

Prof. Dr. Andreas Breiter, Universität Bremen, Fachbereich 3: Mathematik und Informatik, Institut für Informationsmanagement, E-Mail: abreiter@uni-bremen.de.