Kanton Zürich
Staatskanzlei

Universität Basel

Einsatz Künstlicher Intelligenz in der Verwaltung: rechtliche und ethische Fragen

Schlussbericht vom 28. Februar 2021 zum Vorprojekt IP6.4

Auftraggeberin

Staatskanzlei Kanton Zürich
Dr. Kathrin Arioli (Auftraggeberin)
Lukas Weibel (Projektleiter)

Autorinnen und Autoren

Prof. Dr. Nadja Braun Binder
Matthias Spielkamp
Catherine Egli
Laurent Freiburghaus
Eliane Kunz
Nina Laukenmann
Dr. Michele Loi
Dr. Anna Mätzener
Liliane Obrecht
Jessica Wulf

Ein Projekt der

Universität Basel
Juristische Fakultät

in Zusammenarbeit mit

ALGORITHM WATCH / CH

Vorwort

Diese Studie ist im Zeitraum August 2020 bis Februar 2021 entstanden – ein für wissenschaftliche Projekte eher kurzer Zeitraum. Erschwerend kam hinzu, dass angesichts der Reisebeschränkungen und Restriktionen hinsichtlich physischer Treffen die ursprünglich als Präsenzveranstaltungen geplanten Workshops und Interviews ausschliesslich online durchgeführt werden konnten. Ohne ein hoch motiviertes Team wäre diese Studie nicht möglich gewesen.

Ein immenses Dankeschön gebührt deshalb den Mitwirkenden aus dem Team der Juristischen Fakultät der Universität Basel, namentlich Catherine Egli, Laurent Freiburghaus, Eliane Kunz, Nina Laukenmann und Liliane Obrecht, sowie den Beteiligten seitens AlgorithmWatch, namentlich Jessica Wulf für die Konzeption, Durchführung und Auswertung der Interviews und Dr. Michele Loi, der als Senior Research Advisor bei AlgorithmWatch Schweiz gemeinsam mit Dr. Anna Mätzener, Leiterin von AlgorithmWatch Schweiz, für die ethischen Erwägungen zuständig war.

Das Interesse des Kantons Zürich an der Thematik manifestiert sich einerseits in der Tatsache, dass diese Studie in Auftrag gegeben und namentlich durch Lukas Weibel, den Projektleiter in der Staatskanzlei des Kantons Zürich, stets kompetent und zielführend begleitet wurde – wofür wir ihm zu grossem Dank verpflichtet sind. Anderseits konnten wir auf die Unterstützung verschiedener Expertinnen und Experten aus dem Kanton und der Stadt Zürich zählen.

Wir danken Michael Bächinger (Sozialversicherungsanstalt, Kanton Zürich), Dr. Dominika Blonski (Datenschutzbeauftragte, Kanton Zürich), Michael Boller (Amt für Raumentwicklung, Kanton Zürich) und Peter Seidler (Steueramt, Kanton Zürich) sowie Rolf Brühlmann (Kompetenzzentrum für Organisation und Information, Stadt Zürich) für ihre Bereitschaft, uns als Interviewpartnerin bzw. Interviewpartner zur Verfügung gestanden zu haben.

Ein besonderer Dank gilt ausserdem Felix Bührer (Compliancebeauftragter, Kanton Zürich), Christian Häberli (AWK Group AG), Stefan Langenauer, Matthias Mazenauer und Dr. Christian Ruiz (alle Statistisches Amt der Direktion der Justiz und des Innern, Kanton Zürich) sowie Sandra Vogel (Juristische Mitarbeiterin Datenschutzbeauftragte, Kanton Zürich), die uns im Rahmen eines Workshops Feedback zu den beiden Ethik-Checklisten gegeben haben.

Wir danken schliesslich den verschiedenen Personen in kantonalen Steuerverwaltungen und einzelnen Unternehmen, die wir im Laufe der Untersuchung telefonisch kontaktiert haben und die uns bereitwillig für Hintergrundgespräche zur Verfügung standen.

Nadja Braun Binder
Matthias Spielkamp

Inhaltsverzeichnis

Kapitel 1

Einleitung 9

A. Projektauftrag und Zielsetzung 9
B. Terminologie 10
C. Eingrenzung und Zielgruppen 11
D. Zeitraum und Methoden 11
I. Instrumente und Arbeitsschritte 12
II. Projektteam 13

Kapitel 2

Potenzial von KI in der öffentlichen Verwaltung 14

A. Auslegeordnung 14
B. Ausgangslage 14
C. Was veranlasst die Verwaltung, KI einzusetzen? 15
I. Entlastung der Arbeitsprozesse 15
II. Allgemeine Effizienzsteigerung 15
III. Verbesserte Servicequalität und Kundenorientierung 16
IV. Überprüfung Verwaltungstätigkeit 16
V. Datenmengen der Verwaltung 17
D. Welche Herausforderungen müssen beachtet werden? 17
I. Technologie 17
1. Ungewollte Effekte: Bias und Diskriminierung 17
2. Fehlende Nachvollziehbarkeit 18
3. Technische Grenzen 18
4. Entscheidung muss beim Menschen liegen 19
II. Öffentliche Verwaltung 19
1. Fehlendes Vertrauen 19
2. Fehlerkultur 19
3. Wenig Datenfluss 20
4. Diversität in der Verwaltung 20
5. Dezentralität der Schweiz 20
6. Fehlende Rechtsgrundlagen 21
III. Mensch 21
1. Bevölkerung 21
2. Mitarbeitende 21
3. Management 22
IV. Möglichkeiten, den Herausforderungen zu begegnen 22
E. Notwendigkeit eines gesellschaftspolitischen Diskurses 23
F. KI in der öffentlichen Verwaltung 23
I. Allgemeine Überlegungen 23
II. KI-Anwendungsbeispiele in Schweizer Verwaltungen 24
1. Steuerverfahren 24
2. Sozialversicherungsverfahren 25
3. Polizeiarbeit 25
a) Predictive Policing 25
i. Ortsbezogenes Predictive Policing in der Schweiz 25
ii. Personenbezogenes Predictive Policing in der Schweiz 25
b) Datenanalyse 26
c) Videoanalyse 26
d) Automatische Fahrzeugerkennung und Verkehrsüberwachung 26
4. Justizvollzug 26
5. Chatbots in unterschiedlichen Einsatzgebieten 27
6. Weitere Einsatzgebiete 27
7. Zwischenfazit 28
III. Internationale Beispiele 28
1. Profiling Arbeitsloser in Dänemark 28
2. Kontrolle von Sozialhilfeempfängerinnen und Sozialhilfeempfängern in Dänemark 29
3. Effizienzsteigerung von Genehmigungsprozessen bei der finnischen Einwanderungsbehörde 29
4. Automatisierte Prozesse bei der finnischen Sozialversicherungsanstalt 30
5. Trelleborg in Schweden 30
6. System für Risikobewertung (SyRi) in den Niederlanden 31
7. Erkennung von Bankkonten, die für illegale Aktivitäten genutzt werden, in Polen 31
8. Das Arbeitsmarkt-Chancen-Assistenzsystem (AMAS) in Österreich 32

Kapitel 3

Rechtliche Rahmenbedingungen für den staatlichen KI-Einsatz im Kanton Zürich 33

A. Zentrale Herausforderungen und Schlussfolgerungen 33
I. Legalitätsprinzip 33
1. Grundlagen 33
2. Legalitätsprinzip und KI 34
a) Erfordernis der Normdichte 34
b) Erfordernis der Normstufe 34
3. Schlussfolgerungen für den Kanton Zürich 35
II. Verfahrensgarantien und Verfahrensgrundsätze 35
1. Grundlagen 35
a) Verfahrensgarantien 35
b) Verfahrensgrundsätze 36
2. Anspruch auf rechtliches Gehör und KI 36
a) Anspruch auf vorgängige Äusserung 36
b) Begründungspflicht 37
3. Untersuchungsgrundsatz und KI 38
4. Schlussfolgerungen für den Kanton Zürich 38
a) Rechtliches Gehör 38
b) Begründungspflicht 38
c) Untersuchungsgrundsatz 39
III. Diskriminierungsverbot 39
1. Grundlagen 39
2. Diskriminierungsverbot und KI 40
a) Mögliche Diskriminierungsquellen in KI-Systemen 40
i. Präexistierender Bias in den Daten 40
ii. Technischer Bias und fehlende Daten 41
iii. Emergenter Bias und statistische Diskriminierungen 41
iv. Diskriminierender Output durch dynamisches Weiterlernen 41
b) Grosses Diskriminierungspotenzial durch KI 41
3. Schlussfolgerungen für den Kanton Zürich 42
IV. Informationelle Selbstbestimmung und Datenschutz 42
1. Grundlagen 42
a) Verfassungsrechtliche Verankerung 42
b) Gesetzliche Konkretisierung des Datenschutzes 43
i. Allgemeine Grundsätze staatlichen Handelns 43
ii. Zweckbindung 44
iii. Erkennbarkeit und Informationspflicht 44

iv. Datenrichtigkeit, Berechtigungsrecht und Löschungsanspruch 44
v. Folgenabschätzung 45
2. Datenschutz und KI 45
a) Automatisierte Einzelentscheidungen 45
b) Profiling 45
3. Schlussfolgerungen für den Kanton Zürich 46
V. Ermessen und unbestimmte Rechtsbegriffe (offene Normen) 46
1. Grundlagen 46
a) Ermessen 47
b) Unbestimmte Rechtsbegriffe 47
2. Offene Normen und KI 47
3. Schlussfolgerungen für den Kanton Zürich 47
VI. Verfügungsbegriff 48
1. Grundlagen 48
2. Verfügungsbegriff und KI 48
3. Schlussfolgerungen für den Kanton Zürich 48
VII. Transparenz 49
1. Grundlagen 49
a) Transparenz als Folge des rechtlichen Gehörs 49
b) Transparenz als Folge datenschutzrechtlicher Anforderungen 49
c) Transparenz als Folge der behördlichen Informationspflicht bzw. des Öffentlichkeitsprinzips 49
2. Transparenz und KI 50
3. Schlussfolgerungen für den Kanton Zürich 50
B. Beispielhafte Einsatzbereiche und Anwendungen 50
I. Steuerverfahren 50
1. Steuerrechtssystem in der Schweiz 50
2. Steuerveranlagungsverfahren 51
3. KI im Veranlagungsverfahren 52
4. Rechtliche Rahmenbedingungen 52
a) Legalitätsprinzip 52
b) Begründungspflicht 53
c) Zusammenspiel Untersuchungsgrundsatz und Mitwirkungspflicht 54
II. Sozialversicherungsverfahren 55
1. Grundlagen des Schweizer Sozialversicherungssystems 55
2. Kompetenzordnung 55
a) Bundesrechtliche Kompetenzen 55
b) Kantonale Kompetenzen 55
3. KI-Anwendungen im Sozialversicherungsbereich 56
4. Rechtliche Rahmenbedingungen 56
a) Legalitätsprinzip 56
b) Diskriminierungsverbot 56
III. Chatbots 57
1. Was ist ein Chatbot? 57
2. Chatbots in der öffentlichen Verwaltung 58
3. Rechtliche Rahmenbedingungen 58
a) Legalitätsprinzip 58
i. Erfordernis des Rechtssatzes 58
ii. Erfordernis der Normstufe 59
iii. Erfordernis der Normdichte 59
b) Datenschutzrechtliche Anforderungen an Chatbots 59
c) Fehlerhafte Auskunft: Vertrauensschutz 60
i. Eignung der Auskunft zur Begründung von Vertrauen 60
ii. Zuständigkeit der auskunftserteilenden Behörde 61
iii. Vorbehaltlosigkeit der Auskunft 61
iv. Unrichtigkeit der Auskunft nicht erkennbar 62
v. Nachteilige Disposition aufgrund der Auskunft 62
vi. Keine Änderung des Sachverhalts oder der Gesetzgebung 62
vii. Interessenabwägung 62
C. Zusammenfassung 62

Kapitel 4

Ethisch vertretbarer Einsatz von KI 65

A. Einführung 65
I. Ethische Richtlinien für den öffentlichen Sektor 65
Supranationale Richtlinien/verschiedene Akteure 65
Nationale Richtlinien 66
II. Zweistufiges Beurteilungsverfahren 66
B. Sieben Grundsätze 68
I. Ethische Grundsätze 68
1. Schadensvermeidung 68
2. Gerechtigkeit und Fairness 68
3. Autonomie 69
4. Benefizienz 70
II. Instrumentelle und aufsichtsrechtliche Grundsätze 70
1. Kontrolle 70
2. Transparenz 72
3. Rechenschaftspflicht 74
C. Checklisten 74
I. Einleitung 74
II. Checkliste 1: Triage-Checkliste für KI-Systeme 75
1. Einleitende Bemerkungen 75
2. Schadensvermeidung 75
3. Gerechtigkeit und Fairness 76
4. Autonomie 76
III. Checkliste 2: Transparenzbericht 76
1. Abschnitt: Bewertungsphase für die Fragen 2.1. bis 2.6.: Bevor Sie Ihr System entwerfen 76
Transparenz hinsichtlich von Werten 76
Transparenz der Rechenschaftspflicht 76
2. Abschnitt: Bewertungsphase für die Fragen 2.7. bis 2.19: Nach dem Testen des Systems 76
Transparenz der Umsetzung und der Steuerung 76
Transparenz hinsichtlich von Leistungen 77
3. Abschnitt: Bewertungsphase für die Frage 2.20: Nach der Implementierung des Systems, wenn das System überwacht wird 77
IV. Beispiel für den Einsatz der Checklisten 1 und 2: Swiss COMPAS 77
1. Vorbemerkungen 77
2. Checklisten 1 und 2 78
3. Transparenzbericht 85
V. Flussdiagramm 90

Kapitel 5

Ausblick 91

Literaturverzeichnis 92
Materialienverzeichnis 99

Zusammenfassung

Der Regierungsrat des Kantons Zürich hat am 25. April 2018 seine Strategie «Digitale Verwaltung 2018–2023» beschlossen. Teil dieser Strategie ist ein Impulsprogramm, in dessen Rahmen auch der Einsatz von Künstlicher Intelligenz (KI) erprobt werden soll (vgl. IP6.4). Zuerst ist ein Vorprojekt durchzuführen, das einen Überblick über rechtliche und ethische Fragestellungen gibt. Mit dem vorliegenden Bericht wird dieses Vorprojekt abgeschlossen. Der Bericht ist in vier Kapitel gegliedert, die in das Thema und die Studie einführen (Kapitel 1), das Potenzial von KI in der öffentlichen Verwaltung untersuchen (Kapitel 2), die rechtlichen Rahmenbedingungen aufzeigen (Kapitel 3) und einen ethisch vertretbaren KI-Einsatz analysieren (Kapitel 4). Den Abschluss bildet ein kurzer Ausblick (Kapitel 5).

Der Bericht liefert in Kapitel 2 eine Auslegeordnung zu Chancen und Risiken beim Einsatz von KI durch die öffentliche Verwaltung und zu Verwaltungsbereichen in der Schweiz und im Ausland, in denen KI eingesetzt wird bzw. ein Einsatz geplant ist. Die Erkenntnisse zu den Vorteilen und Herausforderungen von KI basieren auf Interviews mit Expertinnen und Experten aus der Verwaltung und wurden durch weitere Recherchen ergänzt. Behörden erhoffen sich vom KI-Einsatz verschiedene **Vorteile.** Einfache Arbeitsprozesse zu automatisieren, soll das Personal entlasten, generell die Effizienz steigern und die Qualität der Dienstleistungen verbessern. Automatisierungsbestrebungen bieten zudem Gelegenheit, bestehende Prozesse allgemein zu reflektieren. Schliesslich möchten Behörden durch die Automatisierung stärker von den grossen Datenmengen profitieren, welche bei der Verwaltungstätigkeit naturgemäss anfallen.

Allerdings birgt der Einsatz von KI für die Verwaltung auch **Herausforderungen.** Diese können sich aus der Technologie selbst, aber auch aus den Eigenheiten der öffentlichen Verwaltung sowie aus dem Verhältnis zwischen Mensch und Maschine ergeben. Risiken werden insbesondere in der Diskriminierung gesehen. Daten können fehlerhaft oder unvollständig sein und so alte Diskriminierungsmuster reproduzieren oder gar neue etablieren. Als problematisch wird auch die fehlende Nachvollziehbarkeit betrachtet. Teilweise ist es fast unmöglich, zu verstehen, wie lernende Algorithmen zu ihren Ergebnissen kommen. Schliesslich müssen auch die Grenzen der KI erkannt werden. KI kann grosse Datenmengen analysieren und strukturieren, was allerdings bei komplexen Einzelfällen kaum hilft. Gerade solche sind aber in der öffentlichen Verwaltung häufig. Grosse Zurückhaltung besteht zuletzt hinsichtlich eines Einsatzes von KI für Entscheidungen in sensiblen Bereichen, wo verschiedene Kriterien gegeneinander abgewogen werden müssen. Solche Entscheidungen sollen weiterhin von Menschen getroffen werden. Für die öffentliche Verwaltung stellt auch das fehlende Vertrauen der Bevölkerung ein Problem dar. Viele Menschen teilen ihre Daten offenbar lieber mit Grosskonzernen als mit dem Staat, obwohl für Letzteren sehr viel strengere Regeln gelten. Wie bereits dargelegt, verfügt die öffentliche Verwaltung über grosse Datenmengen aus unterschiedlichen Quellen. Um diese wirklich zu nutzen, müssten sie zwischen verschiedenen Behörden ausgetauscht werden. Solche Datenflüsse sind aber aufgrund von datenschutzrechtlichen Vorgaben oft nicht erlaubt. Eine weitere Herausforderung ergibt sich daraus, dass die Verwaltung unterschiedlichste Aufgaben erfüllt, die jeweils andere technische Lösungen erfordern. Dazu kommt die Dezentralität der Schweiz: Abläufe und Verfahren sind je nach Kanton sehr unterschiedlich, was es erschwert, gemeinsam Lösungen zu finden. In der Bevölkerung ist die Akzeptanz für KI teilweise noch gering, die Befürchtungen dagegen sind gross. Gleichzeitig ist auch bei den Mitarbeitenden der Verwaltung teilweise (noch) keine Akzeptanz vorhanden.

Dargelegt wurden verschiedene Herausforderungen und Gefahren, welchen mit entsprechenden Massnahmen begegnet werden kann und muss. Einige solche Massnahmen werden in den Kapiteln 3 und 4 besprochen. Allerdings können die Risiken eines KI-Einsatzes nie vollständig eliminiert werden. Deshalb braucht es eine konstante Abwägung zwischen Chancen und Risiken. Dieser Prozess kann aber nicht nur verwaltungsintern stattfinden. Vielmehr braucht es einen gesellschaftspolitischen Diskurs. Die Bevölkerung muss die Grundentscheidungen treffen und mittragen.

Die Verwaltung kann KI für ganz **unterschiedliche Aufgaben** einsetzen. Bestimmte Anwendungen bieten nur eine interne Unterstützung und entwickeln keine oder nur eine begrenzte Aussenwirkung. Andere dagegen werden direkt im Rahmen der hoheitlichen Entscheidungstätigkeit eingesetzt. Hier gilt es, zwischen Systemen zu unterscheiden, die Entscheidungen unterstützen, und solchen, die Entscheidungen übernehmen. In der Praxis sind die entscheidungsunterstützenden Systeme verbreiteter. KI eignet sich am besten für den Einsatz in stark strukturierten Prozessen mit grossem Datenvolumen. Potenzial für KI liegt deshalb in erster Linie in der Massenverwaltung, wo Verwaltungsbehörden für eine grosse Anzahl von ähnlich gelagerten Fällen einzelne Verfügungen erlassen müssen. Dies spiegelt sich auch in Ergebnissen der Status-quo-Recherche wider. In verschiedenen Kantonen sind Pläne zum Einsatz von KI im Steuerverfahren weit fortgeschritten. Ein weiteres Gebiet der Massenverwaltung ist das Sozialversicherungsverfahren. Auch hier liegt grundsätzlich einiges an Potenzial für den Einsatz von KI. Allerdings sind die Behörden deutlich zurückhaltender, konkrete Projekte sind kaum vorhanden. Unabhängig von bestimmten Verwaltungsbereichen hat sich gezeigt, dass Chatbots eine weit verbreitete KI-Anwendung darstellen. Diese technischen Dialogsysteme werden vor allem dazu eingesetzt, Anfragen zu beantworten.

Öffentliche Verwaltungen in anderen Ländern setzen teilweise bereits deutlich häufiger KI ein. In Dänemark wird diese zum Profiling von Arbeitslosen und zur Kontrolle von Sozialhilfeempfängerinnen und -empfängern genutzt. In Finnland werden Prozesse im Bereich Migration und Sozialhilfe automatisiert. Die Kommune Trelleborg in Schweden setzt automatische Entscheidungssysteme ein, um Anträge für Sozialleistungen zu bearbeiten. In den Niederlanden war ein datenbasiertes Analyseprogramm im Einsatz, welches Missbrauch im Bereich von Sozialleistungen aufdecken sollte, bevor sein Einsatz gerichtlich untersagt wurde. In Polen hilft ein EDV-System, finanzielle Aktivitäten zu untersuchen und dabei illegale Aktivitäten aufzudecken. In Österreich soll ein Arbeitsmarkt-Chancen-Assistenzsystem Verwendung finden: Auf Basis von statistischen Analysen würden die Chancen von Arbeitssuchenden am Arbeitsmarkt berechnet, um Empfehlungen für den staatlichen Ressourceneinsatz zur Unterstützung bei der Arbeitsmarktintegration zu entwickeln.

Im Rahmen dieser Studie werden verschiedene **rechtliche Herausforderungen** identifiziert, die der Kanton Zürich zu bedenken hat, wenn er KI-Systeme einsetzen möchte.

Zunächst ist aufgrund des **Legalitätsprinzips** sicherzustellen, dass eine sowohl hinsichtlich der Normstufe als auch bezüglich der Normdichte ausreichende Rechtsgrundlage existiert. Weitere Anforderungen ergeben sich aus dem **Anspruch auf vorgängige Äusserung** und Mitwirkung in Verwaltungsverfahren. Mit dem KI-Einsatz im Rahmen von automatisierten Verfahren geht das Risiko einher, dass das Recht auf vorgängige Äusserung eingeschränkt wird. Dabei sind die Limitationen in vollautomatisierten Verfahren tendenziell grös-

ser als in teilautomatisierten. In Abhängigkeit von der konkreten KI-Anwendung ist es deshalb empfehlenswert, zu prüfen, ob ein Anspruch auf vorgängige Äusserung im Rahmen eines KI-Einsatzes im kantonalen Verwaltungsrechtspflegegesetz (VRG) verankert werden soll. Auch aus der **Begründungspflicht** ergeben sich Herausforderungen. Beim KI-Einsatz besteht die Gefahr, dass die entscheidende Behörde der Begründungspflicht nicht vollumfänglich nachkommen kann. Dies gilt insbesondere für Systeme, die statistische Auswertungen vornehmen und bei denen die Entscheidung mithin nach Kriterien erfolgt, die für die Sachbearbeiterin oder den Sachbearbeiter selbst nicht nachvollziehbar sind. Werden solche Systeme für den Erlass von Anordnungen eingesetzt, ist demnach sicherzustellen, dass die Anordnung dennoch rechtskonform begründet wird. Dies könnte etwa durch die Anforderung umgesetzt werden, dass die Logik der Entscheidungsfindung in der Begründung angegeben werden muss. Entsprechende Vorgaben könnten Eingang in das VRG finden, aber auch im Rahmen des kantonalen Gesetzes über die Information und den Datenschutz (IDG) verankert werden.

Der **Untersuchungsgrundsatz** zieht sodann weitere Anforderungen nach sich. Von zentraler Bedeutung ist – auch mit Blick auf die Verhinderung von Diskriminierung und aus datenschutzrechtlicher Sicht –, dass die vom KI-System genutzten Trainingsdaten und weiteren Daten vollständig, korrekt und im zur Eruierung der rechtserheblichen Tatsachen notwendigen Umfang verfügbar sind. Es wird empfohlen, diese Anforderung auf gesetzlicher Stufe, z. B. in § 7 VRG, zu verankern. Um den Amtsermittlungsgrundsatz sicherzustellen, kann es ferner notwendig sein, im Rahmen des spezifischen Fachgesetzes, in dessen Anwendungsbereich ein KI-System eingesetzt werden soll, die notwendige Rechtsgrundlage für den Zugriff auf vorhandene Datensammlungen zu schaffen.

Grosse Herausforderungen stellen sich beim staatlichen KI-Einsatz aufgrund des **Diskriminierungsverbots.** Angesichts der verschiedenen Quellen von Diskriminierung kann deren Verhinderung nicht alleinige Aufgabe der Rechtsetzung sein. Eine der Diskriminierungsquellen bilden unrichtige Daten. Deshalb muss die Verwaltung – nicht nur als Ausfluss des Untersuchungsgrundsatzes und aufgrund datenschutzrechtlicher Vorgaben – sicherstellen, dass die genutzten Trainingsdaten und Sachverhaltsdaten korrekt sind und nur solche verwendet werden, die für das entsprechende Verfahren geeignet sind. Um zu verhindern, dass KI-Anwendungen, die Entscheidungen unterstützen, diskriminieren, sollte sichergestellt werden, dass Sachbearbeiterinnen und Sachbearbeiter über die notwendigen Kenntnisse und Kompetenzen verfügen, um im Einzelfall eine Entscheidung zu treffen, die vom diskriminierenden Vorschlag abweicht. Weitere denkbare Massnahmen, um Diskriminierung beim Einsatz von maschinellem Lernen zu verhindern, sind Kontrollalgorithmen einzusetzen oder Drittorganisationen bzw. staatliche Institutionen damit zu beauftragen, die Systeme regelmässig zu kontrollieren.

Aus dem Recht auf **informationelle Selbstbestimmung** bzw. aus den **datenschutzrechtlichen Grundsätzen** ergeben sich weitere Anforderungen. Dabei kommen im Kontext von KI-Anwendungen dem Grundsatz der Datenrichtigkeit und der Herstellung von Transparenz besondere Bedeutung zu. Eine Verstärkung bzw. Ausweitung dieser beiden Grundsätze könnte mittels Ergänzungen in § 7 IDG bzw. § 12 IDG erreicht werden. Dabei ist zum einen an die Notwendigkeit korrekter, aktueller und vollständiger Daten zu denken, da von unrichtigen Daten im Rahmen von KI-Anwendungen auch ein erhöhtes Diskriminierungsrisiko ausgeht. Die skizzierten Anforderungen beziehen sich auf alle in KI-Systemen genutzten Daten und somit sowohl auf Sachdaten als auch auf Personendaten. Zu prüfen wäre zum anderen, ob § 12 IDG um eine Informationspflicht hinsichtlich einer automatisierten bzw. KI-gestützten Datenbearbeitung zu ergänzen wäre. Das Ziel einer solchen Bestimmung wäre, die betroffene Person darüber zu informieren, dass ihre Daten automatisiert bzw. mithilfe von KI-Anwendungen bearbeitet werden und der daraus resultierende Entscheid Rechtswirkungen für sie entfaltet.

Die Herstellung von **Transparenz** beim staatlichen KI-Einsatz ist nicht nur unter dem Blickwinkel individueller Kontrollmöglichkeiten von Einzelentscheiden, sondern auch in Bezug auf eine allgemeine Kontrolle – etwa durch die Zivilgesellschaft – zu diskutieren. Dabei sind verschiedene Ansatzpunkte vorstellbar, wie die Transparenz zur Ermöglichung von Kontrolle rechtlich konkretisiert werden könnte. Denkbar wäre etwa, ein öffentlich zugängliches Register zu schaffen, aus dem ersichtlich wird, in welchen Bereichen die öffentliche Verwaltung KI-Systeme einsetzt und das u. a. Auskunft über die Art und Herkunft der bearbeiteten Sach- und Personendaten, die Rechtsgrundlage, den Zweck und die Mittel der Bearbeitung, das verantwortliche Organ, die KI-Anwendung und deren Logik sowie diejenigen Akteure, die an der Entwicklung des Systems mitgewirkt haben, gibt. Eine weitere mögliche Herangehensweise zur Herstellung von Transparenz findet sich in dieser Studie in Kapitel 4. Zu diskutieren wäre, wie und wo die Herstellung von Transparenz mittels der in Kapitel 4 vorgeschlagenen Checklisten bzw. der Erstellung eines Transparenzberichts rechtlich zu verankern wäre.

Die ethischen Erwägungen im vierten Kapitel dieser Studie stützen sich auf verschiedene ethische Richtlinien zum Einsatz KI-basierter Systeme. Aus der Fülle der Richtlinien wurden sieben ethische Grundsätze abgeleitet, die der weiteren Analyse zugrunde gelegt werden. Dabei handelt es sich um die folgenden sieben Werte: die Achtung der **menschlichen Autonomie,** die **Schadensvermeidung,** die **Gerechtigkeit** oder **Unparteilichkeit** (Fairness), die **Benefizienz** sowie die drei instrumentellen Grundsätze der **Kontrolle, Transparenz** und **Rechenschaftspflicht,** die technische, organisatorische und aufsichtsrechtliche Anforderungen zusammenfassen, die üblicherweise in praktischen Richtlinien zur KI-Ethik enthalten sind.

Auf Basis der theoretischen Herleitung dieser sieben ethischen Grundsätze wurden im Laufe der Studie sodann zwei Checklisten entwickelt, die als Hilfsmittel dafür zu verstehen sind, Transparenz bei technologischen Automationsprojekten und -anwendungen in der öffentlichen Verwaltung herzustellen.

Die Methode zur Herstellung von Transparenz besteht darin, einen **Transparenzbericht** zu verfassen, der zeigt, dass die wichtigsten ethischen Fragen sowohl erkannt als auch reflektiert wurden und dass eine angemessene Rechenschaftspflicht für den Prozess sichergestellt wurde.

Anhand der ersten Checkliste **(Triage-Checkliste)** beurteilt die Verwaltung, welche ethischen Transparenzfragen während der Projektdurchführung im Detail zu dokumentieren sind, und wählt angemessene Vorgehensweisen, um diejenigen Daten und Bewertungen zu generieren, die es ermöglichen, den Bericht mit informativem Inhalt zu füllen. Die folgenden Fragen helfen bei der Beurteilung:

- Mit wie vielen ethischen Transparenzaspekten muss die Verwaltung sich befassen?
- Wie viele ethische Transparenzverfahren müssen implementiert werden?
- Wie viele Ressourcen müssen für ethische Transparenzverfahren bereitgestellt werden?
- Welche Aspekte der ethischen Transparenz müssen im Bericht detailliert behandelt werden? (Und ist ein solcher Bericht überhaupt notwendig?)

Die zweite Checkliste **(Checkliste Transparenzbericht)** dient als Leitfaden für die Erstellung eines ausführlichen Transparenzberichts. Dieser kann erst am Ende einer Entwicklung und Implementierung eines KI-Systems erstellt werden. Allerdings muss mit der Erstellung des Transparenzberichts bereits während des Projekts begonnen werden: Manche Informationen, die für den Transparenzbericht notwendig sind, können nur in den verschiedenen Phasen der Projektdurchführung und nicht erst nach Projektabschluss generiert werden. Am Ende des Projekts muss der Transparenzbericht klare Informationen über die umgesetzten Prozesse enthalten, der zur Adressierung der in Checkliste 1 (Triage-Checkliste) hervorgehobenen spezifischen ethischen Punkte geeignet ist.

Der vorliegende Bericht kombiniert die Auslegeordnung des Einsatzes von KI-Systemen in der Verwaltung mit der Analyse und Einordnung rechtlicher und ethischer Implikationen. Damit trägt er dazu bei, die Chancen des KI-Einsatzes durch Behörden im Kanton Zürich zu nutzen und zugleich den Risiken effektiv zu begegnen.

Kapitel 1

Einleitung

Nadja Braun Binder
Matthias Spielkamp
Catherine Egli

Einleitend werden Auftrag und Zielsetzung (A.), Eingrenzung und Zielgruppen (C.) sowie Zeitraum und Methoden (D.) der Studie erläutert. Ein besonderes Augenmerkt gilt in diesem ersten Kapitel zudem dem Verständnis des Begriffs «Künstliche Intelligenz» für die Zwecke dieser Studie (B.).

A. Projektauftrag und Zielsetzung

Verschiedene Anwendungen Künstlicher Intelligenz sind heute schon fester Bestandteil des Alltags oder stehen kurz davor, dies zu werden: Streamingplattformen schlagen Nutzerinnen und Nutzern ihre nächsten Lieblingsfilme vor, autonome Busse fahren von A nach B, Computer unterbreiten Ärztinnen und Ärzten Vorschläge für eine Therapie für den Lungentumor vor, Algorithmen berechnen die individuelle Wahrscheinlichkeit einer Langzeitarbeitslosigkeit und überweisen automatisch staatliche Ausbildungsunterstützungen an Studierende.[1] Möglich werden solche Entwicklungen durch die Verfügbarkeit von enorm gestiegenen Datenmengen, die Leistungsentwicklung von Computern sowie die Entwicklungen des maschinellen Lernens.[2] Mit dieser zunehmend vernetzten Automatisierung alltäglicher Prozesse gehen sowohl Chancen als auch Herausforderungen einher. Besonders für staatliche Akteure, die KI-Anwendungen einsetzen, gilt es, Chancen und Herausforderungen möglichst früh zu erkennen. Nur so kann die digitale Transformation rechtskonform, bürgerinnen- und bürgerfreundlich und ohne Risiko des Vertrauensverlusts in staatliche Institutionen gelingen.

Vor diesem Hintergrund hat der Bundesrat 2018 die Strategie «Digitale Schweiz» verabschiedet und in diesem Rahmen eine interdepartementale Arbeitsgruppe ins Leben gerufen, welche die Herausforderungen der Künstlichen Intelligenz darlegen sollte. Die Arbeitsgruppe hat ihre Resultate Ende 2019 in einem Bericht zusammengefasst und dem Bundesrat zur Kenntnis gebracht. Darauf aufbauend hat dieselbe Arbeitsgruppe sodann Leitlinien für den Umgang mit Künstlicher Intelligenz im Bund erarbeitet, die am 25. November 2020 vom Bundesrat verabschiedet wurden.[3]

Künstliche Intelligenz in der kantonalen Verwaltung anzuwenden und zu fördern, liegt in der Kompetenz der Kantone. So hat der Regierungsrat des Kantons Zürich am 25. April 2018 seine Strategie «Digitale Verwaltung 2018–2023» beschlossen. Teil dieser Strategie ist ein Impulsprogramm, welches sieben Strategieziele anhand verschiedener Projekte umsetzt.[4] Die Abteilung Digitale Verwaltung und E-Government der Staatskanzlei des Kantons Zürich ist unter Ziel 6 «Umsetzung des digitalen Arbeitsplatzes für Zusammenarbeit und Geschäftsabwicklung» im Projekt IP6.4 dafür zuständig, den Einsatz von Künstlicher Intelligenz zu erproben, wobei insbesondere auch Fragen der digitalen Ethik zu berücksichtigen sind. Anfänglich sollte die Erprobung des Einsatzes von Künstlicher Intelligenz direkt durch ein Pilotvorhaben erfolgen. Mangels fundierter Grundlagen wurde jedoch entschieden, zuerst ein Vorprojekt durchzuführen, um eine Auslegeordnung zu rechtlichen und ethischen Fragestellungen zu erhalten.

Zu diesem Zweck sollen die Ergebnisse bereits abgeschlossener nationaler und internationaler Studien für die kantonale Verwaltung Zürich konkretisiert werden. Im Fokus stehen zum einen die Empfehlungen der im Auftrag von TA-SWISS erstellten und im April 2020 erschienenen Studie «Wenn Algorithmen für uns entscheiden: Chancen und Risiken der Künstlichen Intelligenz».[5] Diese Empfehlungen sollen so weit wie möglich für die Verwaltung des Kantons Zürich präzisiert werden. Zum anderen sollen auch die verschiedenen Studien der gemeinnützigen Organisation AlgorithmWatch berücksichtigt werden, wobei insbesondere auf den im Oktober 2020 erschienenen und vollständig neu überarbeiteten Automating Society Report 2020 zu verweisen ist, der erstmals auch einen Überblick über die Schweiz enthält.[6] Die zu erarbeitenden anwendungsorientierten Konkretisierungen des bestehenden Wissens sollen dem Kanton Zürich als Grundlage für allfällige weitere Umsetzungsschritte des Projekts IP6.4 dienen. Mit dieser Studie ist mithin aufzuzeigen, wo das Potenzial von KI in der öffentlichen Verwaltung liegt (Kapitel 2) und welche rechtlichen (Kapitel 3) und ethischen (Kapitel 4) Rahmenbedingungen beim KI-Einsatz in der öffentlichen Verwaltung zu berücksichtigen sind. Darüber hinaus werden konkrete rechtliche und ethische Schlussfolgerungen und Vorschläge entwickelt, die als Ausgangspunkt für konkrete KI-Projekte dienen sollen. Die Analyse und/oder die Entwicklung von informationstechnischen Lösungsansätzen waren dagegen explizit nicht Gegenstand der vorliegenden Studie.

1 Vgl. etwa die internationalen Beispiele in Kapitel 2 F. III.
2 TA-SWISS KI, 2020, S. 53.
3 Vgl. für eine Übersicht der Entwicklungen im Rahmen der Strategie «Digitale Schweiz» https://www.sbfi.admin.ch/sbfi/de/home/bfi-politik/bfi-2021-2024/transversale-themen/digitalisierung-bfi/kuenstliche-intelligenz.html.
Alle in diesem Bericht zitierten Internetadressen wurden zuletzt am 19. Februar 2021 überprüft.
4 Vgl. für das Impulsprogramm des Kantons Zürich https://www.zh.ch/de/politik-staat/kanton/kantonale-verwaltung/digitale-verwaltung/strategie-impulsprogramm-digitale-verwaltung.html.
5 TA-SWISS KI, 2020.
6 Automating Society Report 2020, https://automatingsociety.algorithmwatch.org.

B. Terminologie

Eine allgemein gültige und akzeptierte Definition der Künstlichen Intelligenz gibt es nicht.[7] Dies liegt u. a. daran, dass die Erforschung und Entwicklung von KI von mehreren Wissenschaften vorangetrieben wird und ihre Ursprünge über 70 Jahre zurückreichen, wobei sich die jeweiligen Forschungsbereiche inzwischen grundlegend weiterentwickelt haben.[8] Diese Studie hat nicht das Ziel, einen weiteren Beitrag zu dieser Diskussion beizutragen, und wird auch nicht ausführlich die Funktionsweise der technologischen Grundlagen der KI erläutern.[9] Da der Name des Projekts, für welches diese Vorstudie durchgeführt wurde, jedoch ebenfalls auf dem Begriff der KI basiert, wird er auch in dieser Studie verwendet. Folglich ist es notwendig, das Verständnis von KI in diesem Kontext kurz zu erläutern.

Grundsätzlich besteht der Begriff aus den zwei Wörtern «künstlich» und «Intelligenz». «Künstlich» kann dabei als «etwas von Menschenhand Geschaffenes» definiert werden, während die Umschreibungen der «Intelligenz» stark variieren. Stark vereinfacht kann man als «Intelligenz» einen Prozess des Verstehens und Lernens bezeichnen.[10] Die Künstliche Intelligenz umschreibt insofern den langjährigen Traum des Menschen, eine Maschine zu entwickeln, die verstehen und denken kann.[11] Dabei muss jedoch zwischen der allgemeinen, starken und der angewandten, schwachen KI unterschieden werden. Die starke KI ist das langfristige Forschungsziel einiger Forscherinnen und Forscher, Maschinen eine den Menschen vergleichbare «allgemeine» Intelligenz zu attestieren, die ein umfassendes Spektrum von Anwendungen intelligenten Denkens sowie Handelns zeigt, die sie auf jedes beliebige Problem anwenden kann. Diese möglichst umfassende Nachbildung menschlicher Intelligenz soll diejenige des natürlichen Menschen allenfalls gar übertreffen können (Superintelligenzen). Dass dieses Ziel erreicht werden kann, ist höchst zweifelhaft und umstritten, kann jedoch ausser Acht gelassen werden, da alle heutigen und in naher Zukunft zu erwartenden KI-Anwendungen der schwachen KI angehören, die dem Menschen an ein bestimmtes Problem oder eine bestimmte Aufgabe angepasst dienen sollen. Sie sollen ihn folglich nicht ersetzen, sondern unterstützen.[12] Die aktuell prägenden KI-Anwendungen entspringen hauptsächlich der Informatikwissenschaft.[13] Mit ihrer Hilfe können gewisse kognitive Fähigkeiten wie das Wahrnehmen, Denken und Handeln digital abgebildet werden. Am besten lässt sich diese pauschale Funktionsweise anhand des Sense-Think-Act-Modells illustrieren.[14]

In der ersten Phase nehmen KI-Technologien bestimmte Signale aus ihrer Umgebung wahr **(Sense-Phase).** Diesen Schritt können einfachere Informationstechnologien dank perzeptiver Fähigkeiten (sehen, hören, lesen) schon länger erfüllen. Auch eine gewöhnliche Überwachungskamera «sieht» die aufgezeichneten Menschen, ein gewöhnliches Mikrofon «hört» die vortragende Person und auch ein gewöhnliches Suchprogramm kann einzelne Wörter eines Dokumentes lesen und finden.[15]

Dank KI sind nun aber zwei zusätzliche Schritte möglich. Ihre Anwendungen können das Wahrgenommene auch analysieren **(Think-Phase)** sowie im Anschluss darauf basierend bestimmte Aktionen ausführen (Act-Phase). Für die genannten Beispiele bedeutet die Think-Phase folglich, dass die Kamera die Menschen nicht nur sieht, sondern die einzelnen Personen auch «erkennt». Das Mikrofon hört nicht nur die Stimme, sondern «versteht» auch, was gesagt wird. Dokumente können nicht nur gelesen werden, sondern der Inhalt wird auch «verstanden» und kann auf korrespondierende Daten hin durchsucht werden. Hierbei ist zu beachten, dass Begriffe wie denken, lernen, erkennen, verstehen und handeln lediglich analog zum menschlichen Denken, Lernen, Erkennen, Verstehen und Handeln verwendet werden. Denn KI-Systeme verfügen weder über Intentionalität oder einen freien Willen noch über eine Erkenntnisfähigkeit, die mit der menschlichen gleichgesetzt werden kann.

Technisch kann die Think-Phase grundsätzlich mittels zweier unterschiedlicher Ansätze realisiert werden. Entweder verwenden Algorithmen dafür eindeutig determinierte Regeln oder sie erkennen eigenständig Muster in den Daten.[16] Bei der determinierten Vorgehensweise arbeiten Computer eine genau definierte Folge von Schritten ab, um ein bestimmtes Ergebnis strukturiert zu erreichen.[17] Diese Anwendungen werden heute oft gar nicht mehr als KI bezeichnet, sondern der «nicht intelligenten» Automatisierung zugordnet.[18] Diese Abgrenzung verdeutlicht, dass zumindest bestimmte verbreitete KI-Technologien bereits eine Basistechnologie[19] geworden sind.[20] Die zweite Vorgehensweise basiert auf dem eigenständigen Suchen von Mustern in den Daten. Die KI-Anwendung analysiert dazu die Daten, um darin Regelmässigkeiten, Wiederholungen, Ähnlichkeiten oder Bedeutungszusammenhänge zu erkennen. Durch ihre Fähigkeit, aus Daten und Erfahrungen selbstständig dazuzulernen und aus der Beobachtung vieler Fälle eigenständig Regeln abzuleiten sowie anzuwenden (maschinelles Lernen), kann sie Zusammenhänge identifizieren, die dem Menschen zuvor weder

7 Vgl. für eine Übersicht über verschiedene Definitionen, OECD, 2019, S. 22 ff.

8 Vgl. Braun Binder, 2019b, S. 468; Ertel, 2016, S. 6 ff.; Görz/Schmid/Wachsmuth, 2014, S. 4 ff.; Mainzer, S. 7 ff.; Reichwald/Pfisterer, 2016, S. 210; Wischmeyer, 2018, S. 9.

9 Vgl. jedoch weiterführend für kurze Definitionen von einzelnen im Kontext von KI verwendete Begriffe: https://www.wissenschaftsjahr.de/2019/uebergreifende-informationen/glossar/.

10 TA-SWISS KI, 2020, S. 71.

11 Vgl. Görz/Schmid/Wachsmuth, 2014, S. 1; Stiemerling, S. 762.

12 Bitkom/DFKI, 2017, S. 29; TA-SWISS KI, 2020, S. 72; Von Lucke/Etscheid, 2020, S. 248; Wischmeyer, 2018, S. 3 Fn. 5; vgl. auch Ramge, 2018, S. 19.

13 Ertel, 2016, S. 1 ff.

14 TA-SWISS KI, 2020, S. 82, wobei das Modell auf Albus, 1991 basiert.

15 Die verwendeten Beispiele im Folgenden stützen sich teilweise auf die ebenfalls vereinfachte, aber übersichtliche Skizze zur Darstellung der KI, abrufbar unter https://www.heise.de/select/tr/2020/11/20252134312545 92846#&gid=1&pid=1.

16 Vgl. zu dieser Zweiteilung ebenfalls https://www.heise.de/select/tr/2020/11/20252134312545 92846#&gid=1&pid=1.

17 Kirn/Müller-Hengstenberg, 2013, S. 226; Reichwald/Pfisterer, 2016, S. 209; Zweig, 2016, S. 209.

18 Oft ist die determinierte Arbeitsweise der Algorithmen folglich in den Definitionen von KI gar nicht zu finden, ausdrücklich aber als Automatisierung definiert z. B. durch Kirn/Müller-Hengstenberg, 2013, S. 226.

19 Vgl. zum Begriff «Basistechnologie» Bericht Herausforderungen, 2018, S. 22 f. Grundsätzlich wird darunter eine Querschnittstechnologie verstanden, die das Potenzial hat, alle Branchen zu durchdringen, und eine hohe Produktivitätswirkung auf eine Vielzahl verschiedener Wirtschaftsbereiche ausüben kann.

20 TA-SWISS KI, 2020, S. 78.

bekannt noch aufgefallen waren.[21] Durch die lernende Mustererkennung werden die Daten folglich nicht nur auf zuvor definierte Zusammenhänge überprüft, sondern können auch ergebnisoffen ausgewertet werden.[22] Die Mustererkennung umfasst dabei verschiedene Formen der Datenanalyse wie die Text-, Sprach-, Gesichts-, Bild-, Bewegtbild- und Raumbilderkennung.[23]
Unabhängig von den technischen Ansätzen[24], die genutzt werden, um den menschlichen «Denkprozess» nachzubilden (Think-Phase), können Systeme auf der Grundlage seiner Ergebnisse zusätzliche Aktionen (etwa Benachrichtigungen, Empfehlungen, Prognosen, Sprachübersetzung, Chat- und Textkommunikation, Programmierung, Navigation und zeitnahe Entscheidungsfindung)[25] ausüben **(Act-Phase).** Diese stellen entweder ohne jegliches menschliche Zutun einen endgültigen Entscheid dar (Vollautomation) oder bilden eine Entscheidungsgrundlage für eine natürliche Person (Teilautomation)[26]. Die KI-Technologie kann folglich beispielsweise eine gesuchte Person auf zahlreichen Überwachungskamerabildern finden und identifizieren, auf eine gestellte Frage mündlich oder schriftlich interaktiv antworten (Voice- oder Chatbot)[27] oder Bedeutungszusammenhänge zwischen gewissen Daten aufzeigen.
Die Fähigkeit, diese drei Phasen zu bestreiten, wird in der vorliegenden Studie als «intelligent» bezeichnet. Folglich wird von einem sehr breiten Verständnis der Künstlichen Intelligenz ausgegangen. Weitere Eingrenzungen aufgrund der unterschiedlichen technischen Möglichkeiten für die Umsetzung des Denkprozesses sowie des Handlungsprozesses werden hier nicht vorgenommen. Wenn in der Folge von KI und KI-Anwendungen die Rede ist, wird darunter folglich die gesamte Bandbreite von qualifizierten[28] Automatisierungsprozessen verstanden. Spezielle Anforderungen an die «Denkleistung» der Technologie mittels maschinellen Lernens oder ähnlicher Techniken werden nicht gestellt. Dieses umfassende Verständnis dient dem praxisorientierten Fokus dieser Studie, denn so sind die Empfehlungen für eine Vielzahl von KI-Projekten unabhängig von ihrer Funktionsweise anwendbar.

C. Eingrenzung und Zielgruppen

In der vorliegenden Studie sollen die rechtlichen und ethischen Rahmenbedingungen eines KI-Einsatzes in der öffentlichen Verwaltung thematisiert und Vorschläge dazu für den Kanton Zürich entwickelt werden.[29] Die Studie stellt keine Gesamtsicht dar, sondern konzentriert sich auf bestimmte Schwerpunkte. Der Untersuchungsgegenstand wird sachlich folgendermassen eingeschränkt: Zielgruppe der Studie ist die öffentliche Verwaltung des Kantons Zürich. Der Einsatz von KI durch private Akteure oder die Judikative wird nicht berücksichtigt. Weiter soll die Studie gemäss Auftraggeberin einen anwendungsorientierten Fokus für die Verwaltung aufweisen. Auf weiterführende technische Erläuterungen bezüglich der Funktionsweise von KI ist zu verzichten.[30] Die untersuchten Bereiche der Verwaltung wurden von den Autorinnen und Autoren weiter konkretisiert, um eine vertiefte Analyse der ethischen und rechtlichen Herausforderungen zu ermöglichen. Unberücksichtigt blieb daher einerseits der Einsatz von KI in der medizinischen Diagnostik, da dies kein hoheitliches Handeln der Verwaltung im klassischen Sinne darstellt.[31] Anderseits wurde weitgehend auf Ausführungen zum Predictive Policing verzichtet, da diese KI-Technologien bereits mehrfach untersucht worden sind und die juristischen Vorgaben zudem sehr spezifisch und insoweit nicht für weitere Zweige der kantonalen Verwaltung anwendbar sind.[32] Allerdings wird im Rahmen der Auslegeordnung (Status-quo-Recherche) von KI-Anwendungen in der Schweiz aufgrund ihrer weiten Verbreitung auf Predictive Policing eingegangen.[33] Schliesslich wurden noch zwei weitere Eingrenzungen vorgenommen: In zeitlicher Hinsicht werden die heute zu erwartenden Entwicklungen der nächsten fünf bis zehn Jahre (angewandte Künstliche Intelligenz) untersucht. Allfällige zukunftsorientierte «Superintelligenzen» sind nicht Thema dieser Studie. Geografisch konzentriert sich die Studie schliesslich auf die Entwicklungen in der Schweiz und insbesondere im Kanton Zürich, wobei punktuell dort auf die internationale Sach- und Rechtslage Bezug genommen wird, wo dies einem vertieften Erkenntnisgewinn dient.

D. Zeitraum und Methoden

Die Autorinnen und Autoren erarbeiteten die vorliegende Studie zwischen August 2020 und Februar 2021 und setzten zur Beantwortung der Fragestellungen verschiedene Instrumente und Arbeitsschritte ein. Zunächst wurden eine Status-quo- sowie eine Literaturrecherche durchgeführt. Im Anschluss fanden Interviews mit verschiedenen Expertinnen und Experten aus der Verwaltung des Kantons Zürich statt. Dieses praktische Wissen wurde sodann mit dem erarbeiteten theoretischen Wissen zusammengeführt. Die ermittelten Ergebnisse der verschiedenen Etappen wurden in vier bzw. fünf[34] unterschiedlichen Workshops innerhalb des Projektteams diskutiert und der Auftraggeberin präsentiert. Im Einzelnen haben die Instrumente und Arbeitsschritte wie in der Folge dargestellt ausgesehen.

21 Bomhard, 2019, Rn. 14; Etscheid/Von Lucke/Stroh, 2020, S. 9; Vieth/Wagner, 2017, S. 10.
22 Etscheid/Von Lucke/Stroh, 2020, S. 9; Stiemerling, 2015, S. 763.
23 Vgl. Von Lucke, 2019, S. 56.
24 Vgl. insbesondere für eine differenzierte Unterscheidung der Denkweise in deduktives, induktives, abduktives und analoges Denken TA-SWISS KI, 2020, S. 85 ff.
25 Von Lucke, 2019, S. 56.
26 Vgl. Degrandi, 1977, S. 52 f.; Eifert, 2006, S. 121 f.
27 Vgl. zu Chatbots Kapitel 2 F. II. 5. und Kapitel 3 B. III.
28 Vgl. etwa die Voraussetzung einer gewissen Komplexität für die Charakterisierung als automatisierte Einzelentscheidung im Sinne von Art. 21 DSG, Botschaft E-DSG, S. 7057.
29 Vgl. Kapitel 1 A.
30 Vgl. für die technischen Grundlagen z. B. TA-SWISS KI, 2020, S. 67 ff.
31 Vgl. dazu z. B. Vokinger/Mühlematter/Becker/Boss/Reutter/Szucs, 2017.
32 Vgl. zur Smart Criminal Justice z. B. Simmler/Brunner/Schedler, 2020.
33 Vgl. Kapitel 2 F. II. 3. a).
34 Der Workshop vom 22. Januar 2021 wurde in zwei Teilen durchgeführt.

I. Instrumente und Arbeitsschritte

Status-quo-Recherche: Um zu eruieren, in welchen Verwaltungsbereichen ein KI-Einsatz in naher Zukunft wahrscheinlich sein könnte, wurde zuerst ein Überblick über den derzeitigen Stand des Einsatzes von KI in der öffentlichen Verwaltung auf Bundes- sowie kantonaler Ebene in der Schweiz ermittelt. Für diese Recherche wurden öffentlich zugängliche schriftliche Unterlagen konsultiert und verschiedene Gespräche mit Verwaltungsmitarbeitenden geführt. Die so eruierten bestehenden KI-Anwendungen bzw. laufenden Projekte wurden sodann nach dem Gesichtspunkt ihrer Eignung für den KI-Einsatz aus einer theoretischen Perspektive analysiert. Als Ergebnis dieser Recherche wurden einerseits Bereiche, die für den KI-Einsatz in der öffentlichen Verwaltung infrage kommen könnten, und anderseits Verwaltungsstellen bzw. Expertinnen und Experten im Kanton Zürich, die in Interviews befragt werden sollten, identifiziert.

Literaturrecherche: Zeitgleich mit der Status-quo-Recherche wurde die aktuelle einschlägige Literatur ausgewertet, um einen Überblick über die Herausforderungen des KI-Einsatzes in der öffentlichen Verwaltung in der Schweiz zu erhalten. Aufbauend auf den bereits erwähnten KI-Studien[35] wurden sodann bestehende Lösungsansätze für die erkannten Herausforderungen zusammengestellt. Die Literaturrecherche diente ausserdem dem Ziel, die Leitfragen für die Interviews mit den Expertinnen und Experten des Kantons Zürich zu entwickeln.

Interviews mit Expertinnen und Experten: Die Interviews mit Expertinnen und Experten aus der Verwaltung des Kantons Zürich dienten dazu, einen praktischen Einblick in ausgewählte Verwaltungsbereiche zu erhalten, in denen KI eingesetzt wird bzw. in naher Zukunft eingesetzt werden könnte. Eruiert wurde, wo die Verantwortlichen die zentralen Herausforderungen und Chancen für einen KI-Einsatz in der Verwaltung verorten. Die Interviewten haben jeweils einen guten Einblick in die öffentliche Verwaltung allgemein, Berührungspunkte mit neuen technischen Anwendungen und sind in die strategische Ausrichtung ihres Bereichs involviert und darüber informiert. Konkret wurden mit fünf Personen (vier Männer, eine Frau) in Leitungsfunktionen Interviews zwischen 45 und 70 Minuten geführt. Die Interviewten des Kantons Zürich sind Michael Bächinger (Sozialversicherungsanstalt, SVA) und Dr. Dominika Blonski (Datenschutzbeauftragte), Michael Boller (Amt für Raumentwicklung) und Peter Seidler (Steueramt) sowie Rolf Brühlmann (vom Kompetenzzentrum für Organisation und Information) von der Stadt Zürich.

Die Interviews wurden zwischen dem 16. November und dem 7. Dezember 2020 durchgeführt und fanden aufgrund von Anweisungen zur Kontaktbeschränkung, um die Ausbreitung von Covid-19 einzudämmen, über Videokonferenz statt. Die Interviews wurden von zwei Personen geführt und die Audiospur zur besseren Transkription aufgenommen. Bei den Interviews handelte es sich um leitfadenbasierte Experteninterviews, d. h., dass den fünf Expertinnen und Experten dieselben Fragen gestellt wurden mit leichten Abweichungen bei den vertiefenden Fragen und Rückfragen. Der Leitfaden basierte auf der Status-quo- sowie der Literaturrecherche dieses Projekts. Die Fragen drehten sich um folgende Themen mit jeweiligem Bezug auf den Bereich der Interviewten: Planung und Umsetzung von KI, Vorteile von KI-Anwendungen, Fallstricke von KI-Anwendungen und Risiken spezifisch für die öffentliche Verwaltung. Der Fokus lag immer auf den Erfahrungen und der Expertise der jeweils interviewten Person. Die Audiospuren der Interviews wurden von einer der Interviewerinnen transkribiert und mit der qualitativen Inhaltsanalyse[36] ausgewertet. Basierend auf der Literaturrecherche wurden Kategorien entwickelt und die Auswertungen der Interviews in diese gegliedert. Themen, die sich in der Literatur nicht fanden, jedoch in den Interviews präsent waren, wurden hinzugefügt.

Die Kategorien wurden in einem weiteren Schritt wo möglich in Cluster zusammengefasst. Um die Nachvollziehbarkeit und eine Triangulation[37] zu gewährleisten, wurden die Kategorien und Cluster mit verschiedenen Projektmitarbeitenden besprochen und im gesamten Projektteam in zwei Workshops diskutiert.

An dieser Stelle soll betont werden, dass es bei den Ergebnissen der Interviews um die Erfahrungen von fünf Expertinnen und Experten in ihren spezifischen Bereichen geht. Durch die Auswahl der Interviewten wurde eine möglichst grosse Aussagefähigkeit sichergestellt. Die Ergebnisse sind jedoch nicht erschöpfend und nur teilweise auf die gesamte Verwaltung des Kantons Zürich generalisierbar. Nicht berücksichtigt werden konnte die Betroffenensicht der Bürgerinnen und Bürger. Dies hätte einerseits den Rahmen der Untersuchung gesprengt, und anderseits gibt es – wohl aufgrund der erst zögerlichen Einführung von KI in der öffentlichen Verwaltung in der Schweiz – noch keine repräsentativen Betroffenenorganisationen, die als Interviewpartnerin geeignet wären. In Zukunft wären diesbezüglich weiterführende Arbeiten jedoch wünschenswert.

Workshops: Die Meilensteile der Status-quo-Recherche, Literaturrecherche, Interviews mit Expertinnen und Experten sowie die daraus abgeleiteten Vorentwürfe der einzelnen Kapitel sind in vier Workshops zwischen dem 23. Oktober 2020 und dem 12. Februar 2021, teilweise unter Teilnahme der Auftraggeberin, intensiv diskutiert worden. Im Rahmen eines zusätzlichen Workshops mit dem Projektleiter der Staatskanzlei (Lukas Weibel), dem Compliancebeauftragten des Kantons Zürich (Felix Bührer), einer juristischen Mitarbeiterin der kantonalen Datenschutzbeauftragten (Sandra Vogel), dem Leiter des Statistischen Amtes der Direktion der Justiz und des Innern des Kantons Zürich (Stefan Langenauer), dem stv. Amtsleiter des Statistischen Amtes (Matthias Mazenauer), einem Datenwissenschaftler aus dem Statistischen Amt (Dr. Christian Ruiz) sowie einem Senior Consultant des Management- und IT-Beratungsunternehmens AWK-Group AG (Christian Häberli) wurden darüber hinaus die ethischen Empfehlungen und deren Einbindung in bestehende (Projektleitungs-)Prozesse des Kantons Zürich gesondert besprochen.[38]

Schlussbericht: Die zentralen Ergebnisse der beschriebenen Arbeitsschritte wurden schliesslich für den vorliegenden Bericht zusammengefasst und als ethische sowie rechtliche Schlussfolgerungen ausformuliert. Sie richten sich einerseits an die Staatskanzlei des Kantons Zürich und anderseits in allgemeiner Weise an kantonale Verwaltungsstellen, die KI einsetzen oder deren Einsatz planen.

35 Vgl. Kapitel 1 A.

36 Vgl. zur qualitativen Inhaltsanalyse Döring/Bortz, 2016, S. 541 ff.

37 Nachvollziehbarkeit und Triangulation sind wichtige Güte- und Qualitätskriterien qualitativer Forschungsmethoden. Triangulation meint in diesem Falle die Kombination der Perspektive und Interpretation unterschiedlicher Forschender, welche die Ergebnisse absichern und ergänzen. Zu Triangulation vgl. z. B. Flick, 2004.

38 Selbstverständlich obliegt die alleinige Verantwortung für den Inhalt der Ausführungen in Kapitel 4 den dort genannten Autorinnen und Autoren.

II. Projektteam

Prof. Dr. iur. Nadja Braun Binder
ist Professorin für Öffentliches Recht an der Juristischen Fakultät der Universität Basel. Zu ihren Forschungsschwerpunkten zählen u. a. die Digitalisierung in Staat und Verwaltung. Braun Binder ist Auftragnehmerin der vorliegenden Studie; nadja.braunbinder@unibas.ch.

Matthias Spielkamp
ist Mitgründer und Geschäftsführer von AlgorithmWatch und Mitglied der Global Partnership on Artificial Intelligence (GPAI). Spielkamp bzw. AlgorithmWatch ist Unterauftragnehmer der vorliegenden Studie; spielkamp@algorithmwatch.org.

Catherine Egli
hat Rechtswissenschaft an den Universitäten Basel sowie Genf studiert und ist wissenschaftliche Mitarbeiterin an der Universität Basel (Professur Braun Binder).

Laurent Freiburghaus
studiert Rechtswissenschaft an der Universität Basel und ist ebenda als Student in Assistenzfunktion tätig (Professur Braun Binder).

Eliane Kunz
studiert Rechtswissenschaft an der Universität Basel und ist ebenda als Hilfsassistentin tätig (Professur Braun Binder).

Nina Laukenmann
studiert Rechtswissenschaft an der Universität Zürich und ist als Studentin in Assistenzfunktion an der Universität Basel tätig (Professur Braun Binder).

Dr. Michele Loi
ist Senior Research Advisor bei AlgorithmWatch Schweiz, Senior Researcher am Institut für Biomedizinische Ethik und Medizingeschichte und Research Fellow bei der Digital Society Initiative an der Universität Zürich.

Dr. Anna Mätzener
ist Leiterin von AlgorithmWatch Schweiz. Sie ist promovierte Mathematikerin mit den Nebenfächern Philosophie und italienische Sprachwissenschaft und hat an der Universität Zürich studiert; maetzener@algorithmwatch.ch.

Liliane Obrecht
studiert Rechtswissenschaft an der Universität Basel und ist ebenda als Studentin in Assistenzfunktion tätig (Professur Braun Binder).

Jessica Wulf
ist Projektmanagerin bei AlgorithmWatch. Sie studierte Arbeits- und Bildungspsychologie sowie Internationale Entwicklung und beschäftigt sich vor diesem Hintergrund mit dem Themenfeld automatisierte Entscheidungsfindung und Diskriminierung.

Kapitel 2

Potenzial von KI in der öffentlichen Verwaltung

Jessica Wulf
Catherine Egli

A. Auslegeordnung

Der Fokus der vorliegenden Studie richtet sich auf die Erarbeitung ethischer und rechtlicher Grundlagen für einen qualitätsgesicherten Entwicklungs- und Entscheidungsprozess[39] hinsichtlich des staatlichen Einsatzes von KI.[40] Zu diesem Zweck werden im vorliegenden Kapitel zunächst die tatsächlichen Grundlagen in der Verwaltung analysiert. Dabei geht es insbesondere darum, auszuloten, wo ein KI-Einsatz überhaupt möglich und wann er auch sinnvoll ist. Ferner soll an dieser Stelle skizziert werden, wo KI bereits heute in der Verwaltung eingesetzt wird. Dieser Überblick beruht überwiegend auf den Ergebnissen von Interviews mit Expertinnen und Experten. Die Interviews ermöglichen einen Einblick in fünf Bereiche der öffentlichen Verwaltung im Kanton Zürich und in der Stadt Zürich in Bezug auf Erfahrungen mit Automatisierung und dem (geplanten oder hypothetischen) Einsatz von Künstlicher Intelligenz. Die Interviewten verfügen über einen guten Einblick in ihre Bereiche als Teil der öffentlichen Verwaltung, weisen Berührungspunkte mit neuen technischen Anwendungen auf und sind in die strategische Ausrichtung ihres Bereichs involviert bzw. darüber informiert. Die Ergebnisse der Interviews spiegeln folglich die Erfahrungen, Expertise und Einschätzung der interviewten Personen wider, können aber kein umfassendes Bild liefern. Ergänzt wurden die Interviewergebnisse deshalb durch die Resultate der Literatur- sowie der Status-quo-Recherche.

Die Auslegeordnung beginnt mit einer Kontextualisierung des behördlichen KI-Einsatzes (B.). Im Anschluss wird erläutert, welche Beweggründe die Verwaltung für einen KI-Einsatz grundsätzlich hat (C.) und wo sie aber auch damit verbundene Gefahren und Hindernisse sieht (D.). Dabei wird auch die wichtige politische Komponente eines staatlichen KI-Einsatzes unterstrichen (E.). Zum Schluss folgt eine Darstellung von Verwaltungsbereichen, in welchen KI in der Schweiz und im Ausland eingesetzt wird oder demnächst geplant ist (F.).

B. Ausgangslage[41]

«Bei den guten Projekten kommt der Kunde auf uns zu mit einem konkreten Bedürfnis und nicht andersherum: wir haben eine technische Anwendung und suchen Anwendungsfälle.» *Brühlmann*[42]

Steht ein behördlicher KI-Einsatz zur Diskussion, wird dieses Vorhaben in erster Linie als IT-Projekt wahrgenommen. Auch wenn sich aufgrund der neuartigen Technologien neue Herausforderungen stellen, welchen gegebenenfalls mit neuen, geeigneten Massnahmen entgegengetreten werden muss, verfügt die kantonale Verwaltung über bestehende Prozesse und Abläufe, die auch ein KI-Projekt durchlaufen muss. Dasselbe gilt für die Umsetzung bestehender rechtlicher Anforderungen. Die Ausgangslage einer KI-Anwendung ist demnach grundsätzlich dieselbe wie bei anderen informationstechnischen Neuerungen und insbesondere von den verwaltungsrechtlichen Grundsätzen und Leitbildern bestimmt.

Für die Beantwortung der Frage, *ob* eine KI-Anwendung überhaupt infrage kommt, stellt – neben dem Legalitätsprinzip[43] – das Gemeinwohl[44] einen entscheidenden Orientierungspunkt dar.[45] Das Gemeinwohl bildet eine inhaltliche Schranke allen Staatshandelns. Staatliche Organe sind in ihrem Handeln nicht frei, sondern müssen ihre Tätigkeiten stets mit dem Interesse des Gemeinwohls rechtfertigen, damit sie legitim erscheinen. Die Definition des Gemeinwohls ist dabei vordergründig das Ergebnis eines politischen Prozesses und unterliegt folglich einem ständigen geschichtlichen Wandel.[46] In den Rechtsgrundlagen wird diese politische Ethik in Art. 5 Abs. 2 BV[47] mit den Begriffen «öffentliches Interesse» und «Verhältnismässigkeit» umschrieben, womit die vorhandenen Ermessensspielräume der Verwaltung überprüft werden können. Auch wenn heute fast täglich über aktuelle Entwicklungen der KI berichtet wird,[48] stellt deren Einsatz in der öffentlichen Verwaltung mithin keinen Selbstzweck dar, sondern muss im Interesse des Gemeinwohls liegen und mit klar ersichtlichen und realisierbaren Vorteilen einhergehen. Wie jede informationstechnische Neuerung muss demnach auch eine neue KI-Anwendung den allgemeinen verwaltungsrechtlichen Kriterien der Zweckmässigkeit, Wirksamkeit und Wirtschaftlichkeit genügen.[49] Ein staatlicher KI-Einsatz muss daher vordergründig das anvi-

39 Vgl. zu den Begriffen Zweig, 2019a, S. 9.
40 Vgl. Kapitel 3 und 4.
41 Die hier vorgestellten Vorüberlegungen wurden zwar zum Teil in den Interviews gestreift, jedoch basieren die Ausführungen hauptsächlich auf der Literaturrecherche.
42 Alle Zitate aus den Interviews mit Expertinnen und Experten wurden im gesamten Schlussbericht zur besseren Lesbarkeit sprachlich umformuliert; inhaltlich wurden sie nicht geändert. Die sprachlich angepassten Zitate wurden von den Interviewten überprüft und zur Veröffentlichung freigegeben.
43 Vgl. Kapitel 3 A. I.
44 Vgl. zu den Ausführungen des Gemeinwohls im Folgenden Karlen, 2018, S. 39 f. und 53 ff.
45 Die anderen verwaltungsrechtlichen Grundprinzipien wie der Grundsatz der Rechtsgleichheit, das Willkürverbot und der Grundsatz von Treu und Glauben betreffen nach der hier vertretenen Ansicht nicht die Frage, ob KI eingesetzt werden darf, sondern diejenige, wie sie implementiert werden muss. Vgl. umfassend zu allen Grundprinzipien des Verwaltungsrecht Häfelin/Müller/Uhlmann, 2020, Rn. 565 ff.
46 Vgl. zu den politischen Aspekten beim Einsatz von KI Kapitel 2 E.
47 Bundesverfassung der Schweizerischen Eidgenossenschaft vom 18. April 1999, SR 101.
48 Vgl. für eine Übersicht der Medienberichterstattung TA-SWISS KI, 2020, S. 67 ff.
49 Vgl. dazu die fünf Kriterien für die parlamentarische Oberaufsicht der Bundesversammlung über die Geschäftsführung des Bundes-

sierte, im öffentlichen Interesse liegende Ziel tatsächlich erreichen. Massnahmen, die an diesem Ziel vorbeischiessen oder zu schwach sind, um es zu erreichen, sind unzulässig. Jedoch ist das reine Erreichen eines Ziels noch nicht ausreichend. Der KI-Einsatz muss dem erzielten Ergebnis ausserdem angepasst und angemessen sein. Schliesslich müssen die erreichten, angemessenen Wirkungen auch im richtigen Verhältnis zu den eingesetzten Ressourcen stehen. Mit einem Einsatz von staatlichen KI-Anwendungen stossen Behörden folglich nicht in jeder Beziehung in Neuland vor. Bei jeder noch so revolutionären und beeindruckenden neuen Technologie kann deren Implementierung in der kantonalen Verwaltung nur zu Diskussion stehen, wenn dabei den allgemeinen verwaltungsrechtlichen Grundsätzen entsprochen wird.

«Es geht um einen sinnvollen Einsatz. Wir wollen Technologien einsetzen, weil sie sinnvoll sind, nicht, weil es etwas Lustiges ist.» *Blonski*

C. Was veranlasst die Verwaltung, KI einzusetzen?[50]

Behördliche KI-Anwendungen müssen somit einem klar definierten, fachlichen Ziel dienen und dürfen nicht zum Selbstzweck eingeführt werden. Doch wie soll KI öffentlichen Interessen dienen können? Was erhofft sich die kantonale Verwaltung von einem KI-Einsatz für sich selbst und bzw. oder für die Bevölkerung? Die Chancen der Künstlichen Intelligenz lassen sich grundsätzlich nur schwer beurteilen, da diese von ihrem jeweiligen Anwendungskontext und der konkreten KI-Technologie abhängen.[51] Hinzu kommt, dass mit Chancen der KI oft deren konkrete, aktuelle sowie zukünftige Einsatzfelder in der öffentlichen Verwaltung verbunden werden. Diese Anwendungsfelder werden jedoch gesondert am Ende dieses Kapitels dargestellt.[52] Dennoch können an dieser Stelle die allgemeinen Erwartungen der interviewten Expertinnen und Experten an KI grob in fünf Themenfelder eingeordnet werden.

I. Entlastung der Arbeitsprozesse

«Wir haben bei uns steigende Fallzahlen. Ausserdem werden die Gesetze komplexer und der Klärungsbedarf steigt. Wir sind froh, wenn wir einen Teil davon abfedern können, indem wir sehr einfache Routinetätigkeiten automatisieren und diese nicht mehr von Mitarbeitenden bearbeitet werden müssen.» *Bächinger*

Zunächst sollen bestimmte Arbeitsprozesse durch eine zunehmende Automatisierung vor allem entlastet werden. In verschiedenen Verwaltungsbereichen nehmen die zu erfüllenden Aufgaben zu bzw. steigt der Kostenspardruck.[53] Insbesondere für Steuerbehörden erhöht sich die Arbeitslast durch die steigende Anzahl von zu verarbeitenden Steuererklärungen, straflosen Selbstanzeigen sowie Daten aus dem automatischen Informationsaustausch. Aktuelle Steuerrechtsrevisionen erfordern zudem zusätzliche Veranlagungshandlungen, während die Einzelfälle immer komplexer werden.[54]

II. Allgemeine Effizienzsteigerung

«Für mich steht die Effizienz im Vordergrund.» *Brühlmann*

Aber auch dort, wo der Einsatz von KI zurzeit keine Notwendigkeit darstellt, kann ganz allgemein die Effizienz in der Verwaltung gesteigert werden, wenn durch die Automatisierung mehr Arbeit in kürzerer Zeit zu erledigen ist.[55] Grosse Datenmengen, welche früher maschinell nicht verarbeitet werden konnten, können in Zukunft schneller, effizienter und jederzeit bearbeitet werden. In der öffentlichen Verwaltung gibt es viele Tätigkeiten und Aufgaben, die immer auf gleiche Weise ablaufen, etwa die Überprüfung, ob ein Antrag an die richtige Abteilung gestellt wurde oder ob alle notwendigen Unterlagen eingereicht wurden. Diese einfachen und repetitiven Aufgaben eignen sich laut den Interviewten, durch KI-Anwendungen erledigt zu werden, was zur Effizienzsteigerung beiträgt. Die freiwerdenden Ressourcen können dann für kompliziertere Aufgaben oder Fälle eingesetzt werden, wodurch die Qualität der Verwaltungstätigkeit steigt.

«Der eigentliche Clou ist, dass wir mit komplexen Modellen, also Anwendungen Künstlicher Intelligenz, die einfachen Fälle bearbeiten, nicht die komplexen Fälle.» *Brühlmann*

Durch diese Verschiebung in Richtung höherwertiger und stärker wertschöpfender Verwaltungstätigkeit beschleunigen sich die Prozesse, was zu einer Kostensenkung führt. Die Wirtschaftlichkeit von KI-Systemen kann folglich massgebend zur Effizienz der Verwaltung beitragen.

rates, der Bundesverwaltung, der eidgenössischen Gerichte, der Aufsichtsbehörde über die Bundesanwaltschaft, der Bundesanwaltschaft und anderer Träger von Aufgaben des Bundes in Art. 26 Abs. 3 Bundesgesetz über die Bundesversammlung (Parlamentsgesetz, ParlG) vom 13. Dezember 2002, SR 171.10. Die folgenden Ausführungen lehnen sich daher an die diesbezüglichen Erläuterungen von Sägesser, 2014, Art. 26 N. 39 f. ParlG an.

50 Dieses Kapitel reflektiert zusammen mit Kapitel 2 D. die Hauptergebnisse der Interviews mit den Expertinnen und Experten. Die Literaturrecherche dient in diesem Abschnitt lediglich dazu, gewisse angesprochene Aspekte genauer zu erklären und vereinzelt zusätzliche Thematiken zu ergänzen. Wo diese Ergänzungen anhand der Literatur vorgenommen wurden, werden sie entsprechend gekennzeichnet.

51 TA-SWISS KI, 2020, S. 292.

52 Vgl. Kapitel 2 F.

53 Vgl. z. B. Bericht Herausforderungen, 2019, S. 87; Etscheid/Von Lucke/Stroh, 2020, S. 37 ff.; Hammerschmid/Raffer, 2020, S. 16; Hanania/Knobloch, 2020, S. 11.

54 Vgl. anschaulich für die zunehmende Arbeitslast der Steuerbehörden im Kanton Bern, Fischer/Daepp, 2019, S. 327; Fischer, 2020, Rn. 15 ff.

55 Vgl. allgemein zur Effizienzsteigerung der Verwaltungstätigkeit durch KI z. B. Bericht Herausforderungen, 2019, S. 87; Etscheid/Von Lucke/Stroh, 2020, S. 41; Hammerschmid/Raffer, 2020, S. 16; Opiela/Mohabbat Kar/Thapa/Weber, 2018, S. 27 ff.

III. Verbesserte Servicequalität und Kundenorientierung

Durch die Effizienzsteigerung der Verwaltungstätigkeit werden Ressourcen frei, die zur Verbesserung der Servicequalität und Kundenorientierung genutzt werden können.[56] Einerseits erhalten Bürgerinnen und Bürger dank des Beschleunigungseffekts der Automatisierung schneller Rückmeldungen zu ihren Anträgen und Anfragen. Anderseits verfügen Verwaltungsmitarbeitende durch ihre Entlastung von repetitiven Hintergrundtätigkeiten über mehr Zeit für individuelle Anliegen, was insbesondere bei sensiblen Verwaltungstätigkeiten wie beispielsweise Entscheidungen über IV-Renten sehr wertvoll ist. Dadurch kann die Verwaltung auch kundennäher auftreten und sich besser um Einzelfälle kümmern. Ausserdem kann Künstliche Intelligenz direkt mit dem Ziel, die Kundenorientierung zu fördern, eingesetzt werden. Da sich der Alltag der gesamten Gesellschaft mehr und mehr auf digitale Wege verschiebt, wird in den Interviews eine gewisse Erwartungshaltung gegenüber der Verwaltung beschrieben, sich ebenfalls ein digitales, agiles Handeln anzueignen und insbesondere den Kontakt mit ihr so einfach und angenehm wie möglich zu gestalten.

«Wir sind als Kompetenzzentrum für Geoinformation in der Pflicht, technologische Entwicklungen im Auge zu behalten und abschätzen zu können. Das ist notwendig, weil wir für die kantonale Verwaltung zeitgemässe Werkzeuge einsetzen müssen und wollen.»
Boller

Zum einen können dabei intelligente Dialogsysteme, Chatbots[57], helfen, Anfragen und Anträge von Bürgerinnen und Bürgern zeit- und ortsunabhängig automatisch zu beantworten und so die Serviceleistung vor allem in zeitlicher und örtlicher Hinsicht zu verbessern. Zum anderen kann aber auch ein zunehmendes Angebot von E-Government-Dienstleistungen die Kundenzufriedenheit fördern. Den Bürgerinnen und Bürgern könnte dabei sogar Arbeit abgenommen werden, indem in Zukunft z. B. automatisch vorausgefüllte Steuererklärungen versendet würden, welche von den steuerpflichtigen Personen nur noch überprüft werden müssten. Dafür müssten jedoch zunächst die notwendigen gesetzlichen Grundlagen geschaffen werden.

IV. Überprüfung Verwaltungstätigkeit

KI-Anwendungen können nicht nur dazu benutzt werden, vorhandene Verwaltungsprozesse zu automatisieren, sondern die Automatisierung kann auch zum Anlass genommen werden, die Prozesse zu überarbeiten, von unnötigen Komplexitäten zu befreien und dadurch zu vereinfachen.

«Idealerweise automatisieren wir nicht das, was wir heute haben, sondern evaluieren, was die Technologie leisten kann, und schauen, wie wir unsere Prozesse dazu passend vereinfachen und anpassen können.»
Bächinger

Auch wenn dies in den durchgeführten Interviews nicht weiter vertieft wurde, regt eine Automatisierung von Prozessen gezwungenermassen auch eine grundlegende Reflexion über diese an.[58] Um bisherige menschliche Überlegungen zahlreicher Mitarbeitenden der Verwaltungsbehörde in einen maschinellen Algorithmus zu «übersetzen», ist eine vertiefte Abklärung der zugrunde liegenden Entscheidungskriterien notwendig. Welchen Kriterien wird in welchen Szenarien welches Gewicht zugemessen? Wie sehen diese Entscheidungsprozesse heute aus und wie sollen sie zukünftig aussehen? Darüber hinaus wird dank der Entwicklung von KI-Anwendungen (insbesondere, wenn die Anwendung spezifisch für die Verwaltungsbehörde produziert wurde und diese bei der Entwicklung involviert war) die Verwaltung gewissermassen automatisch geöffnet, denn die Entscheidungsprozesse werden nicht nur von Verwaltungsmitarbeitenden mit juristischem Hintergrund überprüft, sondern auch mit Informatikerinnen und Informatikern sowie Ethikerinnen und Ethiker diskutiert, welche für die Produktion der Technologie verantwortlich sind. Dies ermöglicht eine breite Reflexion über Entscheidungskriterien. Dadurch kann die Verwaltung ihre Arbeit nicht nur verbessern, sondern auch zusätzlich legitimieren. Doch auch nach der Implementierung (zumindest in der Pilotphase) einer KI-Anwendung, kann diese durch ihre systematische Arbeitsweise dazu beitragen, etwa diskriminierende Praktiken offenzulegen, welche auch bei der Entwicklung der Anwendung nicht erkannt wurden. Den dank der Automatisierung identifizierten Diskriminierungen kann folglich gezielt entgegengetreten werden.
Durch die Automatisierung der Verwaltungstätigkeit können jedoch nicht nur die konkreten praktischen Prozesse der einzelnen Verwaltungsbehörden überprüft werden, vielmehr kann die digitale Transformation gemäss Stimmen aus der Literatur auch aus juristischer Perspektive eine Fortentwicklung des Verwaltungsrechts bewirken.[59] Bestimmte Grundannahmen oder ungeklärte Fragen können neu reflektiert werden. So könnte eine Einführung einer gesetzlichen Grundlage für automatisierte Verfügungen einen Anlass bieten, auch das Widerrufsrecht von Verfügungen normativ festzuhalten oder anlässlich der Einführung von Smart Cities[60] könnte die umstrittene Frage geklärt werden, ob die Benutzung des Verwaltungsvermögens und öffentlicher Sachen im Gemeingebrauch durch die Exekutive kraft ihrer Sachherrschaft geregelt werden darf oder es hierfür ebenfalls einer gesetzlichen Grundlage bedarf.

56 Vgl. allgemein zur verbesserten Servicequalität und Kundenorientierung z. B. Bericht Herausforderungen, 2019, S. 87; Etscheid/Von Lucke, 2020, S. 253; Etscheid/Von Lucke/Stroh, 2020, S. 27; Hammerschmid/Raffer, 2020, S. 16; Hanania/Knobloch, 2020, S. 11.
57 Vgl. dazu Kapitel 2 F. II. 5. sowie Kapitel 3 B. III.
58 Die Ausführungen dieses Abschnitts beruhen auf schriftlichen und mündlichen Hinweisen in folgenden Quellen: Stellungnahme von Johannes Kopf zur Kritik am AMS-Algorithmus in Der Standard vom 25.09.2019, https://www.derstandard.at/story/2000109032448/ein-kritischer-blick-auf-die-ams-kritiker; Zweig, 2019b, S. 211; Interview mit Prof. Dr. Zweig über Chancen und Risiken der Künstlichen Intelligenz, abrufbar unter https://kattascha.de/prof-dr-zweig-ueber-chancen-und-risiken-kuenstlicher-intelligenz/; Podiumsdiskussion im Anschluss an den Vortrag von Prof. Dr. Wagner-Pinter über den Prozess der Technikfolgenabschätzung beim Einsatz von ADM-Systemen im Rahmen des Projekts Fair and Good ADM, vgl. dazu https://fairandgoodadm.cs.uni-kl.de.
59 Vgl. zum Folgenden Glaser, 2018, S. 190.
60 Vgl. zu den Smart Cities Kapitel 2 F. II. 6.

V. Datenmengen der Verwaltung

Für jeden KI-Einsatz sind zwingend grosse Datenmengen erforderlich. Die öffentliche Verwaltung besitzt grundsätzlich riesige Mengen an Daten und Informationen aus unterschiedlichen Quellen, weshalb der öffentliche Sektor in Kombination mit seinen typischerweise etablierten Abläufen und weitgehend standardisierten sowie strukturierten Prozessen für den Einsatz Künstlicher Intelligenz fast prädestiniert ist.[61] So bieten sich aufgrund der jährlich grossen Anzahl von Daten, die im Steuerverfahren und Sozialversicherungsverfahren erhoben werden, insbesondere diese Verwaltungsbereiche für zukünftige Einsätze von KI an. Die grossen Datenmengen der öffentlichen Verwaltung könnten im Zusammenhang mit einem KI-Einsatz folglich eine grosse Chance für die Verwaltung darstellen.

«Die letzten Jahrzehnte haben wir immer mehr Daten gesammelt. Die händische Auswertung der Daten stösst durch den grossen Umfang an Informationen an ihre Grenzen. Jetzt geht es darum, die gesammelten Daten sinnvoll auszuwerten, Verknüpfungen herzustellen und Muster zu erkennen, vielleicht in einem nächsten Schritt aus gewonnenen Daten Prognosen für die Zukunft zu erstellen.» *Boller*

Durch KI-Anwendungen können folglich mehr Daten schneller als durch Menschen verarbeitet werden. Im Zuge dessen können durch die Anwendung Zusammenhänge und Muster erkannt werden, welche ein Mensch nicht entdecken würde.[62] Die Erkennung von Mustern und Zusammenhängen in grossen Datenmengen wird in den Interviews als eine potenziell gewinnbringende Anwendung von KI in der öffentlichen Verwaltung beschrieben. Im Steuerverfahren können Mustererkennungen so beispielsweise die Vorbearbeitung von bestimmten Formularen (z. B. Steuererklärungen) ermöglichen. In der Raumplanung werden schon heute KI-Anwendungen verwendet, allerdings von externen Anbietern und nicht von der Verwaltungseinheit selbst. So wird KI beispielsweise eingesetzt, um Luftbildaufnahmen zu analysieren und zu strukturieren und darauf basierend in Kartenmaterial umzuwandeln.

«Datenanalysen und das Erkennen von Mustern in unstrukturierten Daten werden ein wichtiger Einsatzbereich werden.» *Boller*

Im Kanton Zürich sehen Expertinnen und Experten somit u. a. in der Mustererkennungsmethode als Unterstützung der bisherigen Verwaltungstätigkeiten das grösste Potenzial von KI in naher Zukunft. In der Literatur gehen die Erwartungen teilweise noch viel weiter. Aufgrund der stark zunehmenden neuen Datenbestände, die durch smarte Objekte[63] wie etwa Smartphones und smarte Strassenlaternen sowie cyberphysische Systeme[64] gesammelt und durch regionale sowie nationale Datenräume vernetzt werden könnten, würde sich das zukünftige intelligent vernetzte Regierungs- und Verwaltungshandeln signifikant von der bisherigen öffentlichen Verwaltung unterscheiden.[65] Inwiefern die Verwaltung und ihre Aufgabenerfüllung dadurch in ferner Zukunft disruptiv verändert werden, wird in dieser Studie nicht untersucht. Sie konzentriert sich auf die absehbaren Entwicklungen in den nächsten fünf bis zehn Jahren.[66]

D. Welche Herausforderungen müssen beachtet werden?[67]

Von KI-Anwendungen werden somit verschiedene Vorteile für die Verwaltung erwartet. Daneben bestehen jedoch auch einige Befürchtungen hinsichtlich eines staatlichen KI-Einsatzes. Die Analyse der Interviews bezogen auf Fallstricke, Hindernisse, Schwierigkeiten und Risiken rund um den Einsatz von Künstlicher Intelligenz haben dabei drei Cluster ergeben. Im ersten Cluster «Technologie» geht es um Schwierigkeiten, die sich aus Eigenschaften der angewendeten Technologien selbst ergeben können. Das Cluster 2 «Öffentliche Verwaltung» fasst Hindernisse zusammen, die sich aus den spezifischen Merkmalen der öffentlichen Verwaltung in der Schweiz ergeben. Das dritte Cluster «Mensch» dreht sich schliesslich um die Beziehung zwischen Mensch und Maschine.

I. Technologie

Zunächst können in der KI als Technologie bestimmte Schwierigkeiten und Risiken liegen. In diesem Cluster geht es folglich um Hindernisse und Risiken, die in den Technologien selbst verortet werden.

1. Ungewollte Effekte: Bias und Diskriminierung

«Ein grosses Risiko, das ich bei dem Einsatz von Künstlicher Intelligenz oder Algorithmen sehe: Sie müssen diskriminierungsfrei sein und dürfen nicht tendenziös sein.» *Boller*

61 Vgl. dazu ebenfalls z. B. Etscheid/Von Lucke/Stroh, 2020, S. 27, 37 und 45; Hanania/Knobloch, 2020, S. 11.
62 Vgl. zur Mustererkennung auch Kapitel 1 B.
63 Unter smarten Objekten sind Gegenstände zu verstehen, welche das Nutzerverhalten der Personen genau erfassen und diese über Protokolle auch Dritten zur Auswertung bereitstellen; Von Lucke, 2019, S. 52.
64 Unter cyberphysischen Systemen im öffentlichen Raum sind Technologien zu verstehen, welche die Daten aus den smarten Objekten nutzen und auswerten, um auf dieser Basis Aktoren, Menschen und Dinge steuern; Von Lucke, 2019, S. 53.
65 Vgl. dazu Von Lucke, 2019, S. 49 ff.
66 Vgl. auch Kapitel 1 C.
67 Dieses Kapitel reflektiert zusammen mit Kapitel 2 C. die Hauptergebnisse der Interviews mit den Expertinnen und Experten. Die Literaturrecherche dient in diesem Abschnitt lediglich dazu, gewisse angesprochene Aspekte genauer zu erklären und vereinzelt zusätzliche Thematiken zu ergänzen. Wo diese Ergänzungen anhand der Literatur vorgenommen wurden, werden sie entsprechend gekennzeichnet.

Die Ergebnisse von KI-Anwendungen sind im Grunde Vorhersagen statistischer Natur. Um Fragestellungen wie etwa jene, ob auf der Überwachungskamera Person X oder Y zu sehen ist, welches Wort bei einer Übersetzung im konkreten Satz das passendste ist oder welche Antwort am besten auf die im Chatbot gestellte Frage passt, zu beantworten, werden die Algorithmen so modelliert, dass die Antwort eine Wahrscheinlichkeit darstellt.[68] (Zufällige) Fehler sind folglich nie auszuschliessen. Problematisch ist es jedoch, wenn die Ergebnisse systematisch fehlerhaft bzw. verzerrt sind. Die Rede ist in diesem Zusammenhang auch von einem Bias.[69] Die Gründe dafür können in den Daten liegen, wenn z. B. ein statistisch gemessener Zusammenhang (Korrelation) besteht, der jedoch keine kausale Ursache für das Ergebnis bildet.[70] Verzerrungen können jedoch auch durch fehlende Daten entstehen, wenn die Stichprobe etwa eine nicht repräsentative Bevölkerungsgruppe umfasst.[71]

> **«Auch unsere Daten, so leid es mir tut, können fehlerhaft sein. Die Daten sind ein Abbild der Wirklichkeit, das fehlerhaft und unvollständig sein kann.»** *Boller*

Eine solche Verzerrung aufgrund von KI kann dazu führen, dass die unerwünschten statistischen Korrelationen und Zusammenhänge Grundlagen für neue Diskriminierungsmuster[72] bilden oder historische Diskriminierungen reproduziert werden.[73] Dabei ist hervorzuheben, dass historische Diskriminierungen nicht in erster Linie technische Fehler, sondern gesellschaftliche Ungleichheiten und Herausforderungen sind, die in technische Systeme eingeschrieben bzw. in diesen reproduziert werden,[74] denn gerade durch die Konsistenz von KI-Entscheidungen werden die diskriminierenden Effekte systematisiert.[75] Verstärkt werden historische Diskriminierungen jedoch auch durch die sogenannte Feedback-Schleife in Algorithmen. Wurden statistisch gesehen mehr Straftaten von Personen mit Migrationshintergrund registriert, wird der Algorithmus das Kriminalitätsrisiko dieser Bevölkerungsgruppe höher einschätzen. Dadurch wird die Polizei diese Bevölkerungsgruppe vermehrt kontrollieren (Ergebnis: mehr Festnahmen) als diejenigen, die weniger aufmerksam beobachtet werden. Dadurch wird die Vorstellung, von den betreffenden Personengruppen ginge ein erhöhtes Kriminalitätsrisiko aus, weiter verfestigt.[76] Die öffentliche Verwaltung muss zwingend sicherstellen, dass die technische Anwendung, die eingeführt werden soll, keine diskriminierenden Resultate liefert und insbesondere keine diskriminierenden Prozesse fortführt oder verstärkt.

2. Fehlende Nachvollziehbarkeit

Die fehlende Nachvollziehbarkeit ist in der Literatur ein häufig beschriebenes Risiko, das besonders bei neuartigen KI-Technologien auftritt, welche auf Methoden des maschinellen Lernens beruhen. Kurz zusammengefasst beruhen diese Anwendungen auf Methoden, in denen eine Grundstruktur zwar vorgegeben ist, die Konnektivität und Gewichtung der Verbindungen sich aber im Verlaufe der verschiedenen Trainingszyklen verändern.[77] Wie die Algorithmen schliesslich zu ihren Ergebnissen kommen, ist dann auch für die Entwicklerinnen und Entwickler nicht mehr nachvollziehbar.[78] Da im Kanton Zürich – soweit ersichtlich – noch keine Anwendungen von maschinellem Lernen in der öffentlichen Verwaltung eingesetzt werden, wurde diese Thematik in den Interviews nicht besprochen.[79]

3. Technische Grenzen

> **«In meiner Tätigkeit ist der Einzelfall immer wichtig. Also es geht um Einzelfälle, die sich voneinander unterscheiden, und die sind schwierig automatisch abzuarbeiten. Einzelfälle können nicht alle gleich abgehandelt werden.»** *Blonski*

Ein grosses Hindernis von staatlichen KI-Einsätzen stellen die aktuellen technischen Grenzen von KI dar. Diese kann zwar grosse Datenmengen analysieren, strukturieren und darauf basierend einfache Fälle entscheiden. Doch gerade in der öffentlichen Verwaltung sind komplexe, individuelle Einzelfälle, welche nach heutigem Stand der Wissenschaft nicht automatisiert anhand eines statistischen Musters behandelt werden können, weit verbreitet. Hier kommt die KI an ihre Grenzen. Diese komplexen Fälle müssen wohl auch in naher Zukunft weiterhin von Menschen bearbeitet werden.

> **«Komplexitäten, die in vielen Fällen vorhanden sind, verunmöglichen aus meiner Sicht zumindest derzeit noch eine treffsichere Beurteilung durch Künstliche Intelligenz.»** *Seidler*

Des Weiteren eignen sich Menschen durch viel Erfahrung eine gewisse Fachexpertise an, welche die KI nicht ersetzen kann. Diese Fachexpertise ist jedoch notwendig, um schwierige Fälle richtig und angemessen interpretieren zu können, weshalb diese Fälle nicht mit KI bearbeitet werden können.

68 Sehr anschaulich das Beispiel unter https://www.lernen-wie-maschinen.ai/ki-pedia/wahrscheinlichkeit-das-rueckgrat-der-kuenstlichen-intelligenz/: Ein Tierbild wird zu 75% als Hund, zu 20% als Katze, zu 3% Delphin und zu 2% als Fisch identifiziert. Anhand der Wahrscheinlichkeitsrechnung identifiziert der Algorithmus das Tierbild dann als Hund.

69 Bericht Herausforderungen, 2019, S. 26 und 32.

70 Die Armlänge von Kindern korreliert mit der Rechenfähigkeit von Kindern. Kinder rechnen nicht besser aufgrund von längeren Armen, sondern ältere Kinder rechnen meist besser als jüngere Kinder und haben aufgrund ihres Alters einfach längere Arme als jüngere Kinder, Bericht Herausforderungen, 2019, S. 32.

71 Bericht Herausforderungen, 2019, S. 32. Vgl. zu den unterschiedlichen Ursachen von diskriminierenden Algorithmen auch Kapitel 3 A. III. 2.

72 So könnten Männer aufgrund ihrer geringeren Lebenserwartung bei Organspenden benachteiligt werden, Opiela/Mohabbat Kar/Thapa/Weber, 2018, S. 38.

73 Bericht Herausforderungen, 2019, S. 32 und 38.

74 Wenn die Kreditwürdigkeit bestimmter Bevölkerungsgruppen historisch schlecht war, dann ist die Fortführung dieser Historie kein technischer Fehler, https://newsroom.haas.berkeley.edu/minority-homebuyers-face-widespread-statistical-lending-discrimination-study-finds/.

75 Opiela/Mohabbat Kar/Thapa/Weber, 2018, S. 38.

76 Bericht Herausforderungen, 2019, S. 38 f.

77 Vgl. zum maschinellen Lernen auch Kapitel 1 B.

78 Vgl. zur fehlenden Nachvollziehbarkeit etwa Martini, 2019, S. 28 ff.; Bericht Herausforderungen, 2019, S. 31; TA-SWISS KI, 2020, S. 56.

79 Es besteht jedoch ein direkter Zusammenhang zu den Transparenzforderungen aufgrund der fehlenden Akzeptanz in der Bevölkerung; vgl. dazu Kapitel 2 D. II. 1.

«Fachexpertise geht über das reine Analysieren der Daten hinaus. Man zieht aus den Daten Informationen, analysiert und kontextualisiert diese mit Erlerntem und eigenen Erfahrungen. KI-Anwendungen stossen aus meiner Sicht (noch) an ihre Grenzen, wenn gerade in der öffentlichen Verwaltung Fachanalysen gefragt sind.» *Boller*

4. Entscheidung muss beim Menschen liegen

Müssen folgenschwere Entscheidungen für Menschen in sensiblen Lebensbereichen getroffen werden, wo ein gewisser Ermessensspielraum besteht, möchte man keine Algorithmen entscheiden lassen. Besteht kein Ermessensspielraum, sondern müssen beispielsweise lediglich die exakten Rentenbeiträge berechnet werden, ist eine Automatisierung möglich und allenfalls auch sinnvoll. Gibt es bei einer einschneidenden Entscheidung jedoch einen Spielraum, um verschiedene Kriterien abzuwägen, möchte man sicherstellen, dass ein Mensch diese Entscheidung vollumfänglich trifft.

«Wir entscheiden über IV-Anträge, also wirklich wichtige und folgenschwere Entscheidungen. Diese Entscheidungen müssen seriös abgeklärt werden und weiterhin durch Menschen beurteilt und entschieden werden. Das wollen wir nicht an Maschinen delegieren.» *Bächinger*

Diese in den Interviews angesprochene Thematik hängt wohl eng mit dem Aspekt zusammen, dass Algorithmen «kein Taktgefühl haben».[80] Damit wird der Umstand beschrieben, dass es schwierig ist, Algorithmen zu «erklären», in welchem Kontext welche Regeln gelten. Bestimmte Regeln haben in verschiedenen Situationen unterschiedliche Bedeutungen. Bei Menschen würde man von der Fähigkeit des «Feingefühls» sprechen. Dieses Feingefühl ist nur schwer mathematisch darzustellen.[81] Aus diesen Gründen möchte man bestimmte Entscheidungen nicht vollständig an Maschinen abgeben.

«Als Unterstützung ja, aber ich würde davon abraten, Künstliche Intelligenz auch entscheiden zu lassen.» *Seidler*

Die Expertinnen und Experten waren sich daher in dieser Angelegenheit einig: KI kann sehr gut Informationen systematisieren und Entscheidungen vorbereiten. Die Entscheidungen müssen jedoch besonders in bestimmten Bereichen den Verwaltungsmitarbeitenden vorbehalten bleiben.

II. Öffentliche Verwaltung

Die öffentliche Verwaltung verfügt als Organisation über spezifische Merkmale. Einige davon beeinflussen und erschweren den Einsatz von Künstlicher Intelligenz: fehlendes Vertrauen, Fehlerkultur, wenig Datenfluss, Diversität in der Verwaltung, Dezentralität und unklarer gesetzlicher Rahmen. Diese sollten bei der Planung und Umsetzung von KI-Anwendungen bedacht und adressiert werden.

1. Fehlendes Vertrauen

«Warum vertraut man Facebook mehr, wenn man seine Daten preisgibt?» *Boller*

Das mangelnde Vertrauen der Bürgerinnen und Bürger gegenüber dem Staat und damit einhergehend die Zurückhaltung, der öffentlichen Verwaltung Daten bekannt zu geben, war in den Interviews sehr präsent. Das fehlende Vertrauen führt nach der Meinung der Interviewten dazu, dass es schwieriger wird, KI-Anwendungen umzusetzen, obwohl diese Vorteile für die Bürgerinnen und Bürger haben würden. Das beschriebene Misstrauen gegenüber dem Staat steht für die Interviewten im Gegensatz dazu, dass der Staat strengeren Gesetzen als private Unternehmen unterliegt.

«Die Bevölkerung ist dem Staat gegenüber ganz grundsätzlich relativ misstrauisch, gerade wenn es um Datenbearbeitung geht. Das finde ich paradox, weil die rechtlichen Rahmenbedingungen und die Vorgaben für den Staat viel strenger sind. Der Staat darf wirklich nur das tun, was im Gesetz steht, und das Gesetz wurde demokratisch legitimiert.» *Blonski*

2. Fehlerkultur

«Die Erfahrung fehlt einfach. Erfahrungen und damit auch Fehler zu machen, das muss auch der Verwaltung zugestanden werden, sonst können keine Rückschlüsse auf den Umgang mit einer neuen Technologie gezogen werden.» *Boller*

KI-Anwendungen sind überwiegend sehr neu und dementsprechend existieren wenig Anwendungsbeispiele, von denen gelernt werden kann – vor allem für den spezifischen Kontext der öffentlichen Verwaltung. Bei neuen Technologien ist eine Pilotphase, in der man ausprobieren darf und aus Fehlern lernen kann, zentral, um sinnvolle neue Technologien einzusetzen. Aufgrund der erhöhten Anforderungen an die Verwaltung wird eine solche Fehlerkultur für die öffentliche Verwaltung vermisst. Auch in der Literatur wird bestätigt, dass das Vertrauen in staatliche Technologien äusserst fragil ist. Werden einmal Fehler in einer eingesetzten staatlichen KI-Anwendung festgestellt, birgt dies die Gefahr, dass Teile der Bevölkerung das Vertrauen in sämtliche staatlichen KI-Anwendungen in diesem Kontext oder in allen Verwaltungsbereichen verlieren. Breit diskutiert wurde das Risiko des Vertrauensverlusts in den staatlichen Technologieeinsatz insbesondere im Zusammenhang mit dem E-Voting. Werden bei einer elektronischen Wahl oder Volksabstimmung einmal Fehler festgestellt, ist das Risiko gross, dass in Zukunft alle (teilweise elektronischen) Wahl- und Abstimmungsergebnisse bezweifelt werden.[82]

«Eine Fehlerkultur und rechtsstaatliches Handeln, das schliesst sich eben aus.» *Boller*

80 Titel des Buches von Zweig, 2019b: «Ein Algorithmus hat kein Taktgefühl».
81 Interview mit Prof. Dr. Zweig über Chancen und Risiken der Künstlichen Intelligenz, abrufbar unter https://kattascha.de/prof-dr-zweig-ueber-chancen-und-risiken-kuenstlicher-intelligenz/.
82 Vgl. Brändli/Braun Binder, 2003, S. 128; vgl. auch zu jüngeren Rückschlägen NZZ vom 29.03.2019, https://www.nzz.ch/schweiz/fehler-im-quellcode-post-setzt-ihr-e-voting-system-befristet-aus-ld.1471126.

Einige Expertinnen und Experten wünschen sich daher auch bei staatlichen KI-Anwendungen einen lösungsoffeneren Umgang mit neuen Technologien, der ein Ausloten der Anwendungen und von deren Grenzen zulässt.

3. Wenig Datenfluss

Die öffentliche Verwaltung erhält und bearbeitet im Zuge ihrer vielfältigen Aufgaben Daten von der gesamten Bevölkerung aus den unterschiedlichsten Lebensbereichen, z. B. Finanzen und Steuern, Meldedaten sowie Sozialleistungen. Gleichzeitig unterliegt die öffentliche Verwaltung strengen Datenschutzgesetzen.[83] Die unterschiedlichen Abteilungen dürfen Daten ohne gesetzliche Grundlage nicht miteinander austauschen. Das bedeutet, wenn zwei Abteilungen von derselben Person dieselben Daten benötigen, müssen diese in den meisten Fällen doppelt erhoben werden. Ein Beispiel, das mehrfach in den Interviews genannt wurde, sind das Finanzamt und das Sozialamt. Der strenge Datenschutz wird von den Expertinnen und Experten zwar als wichtig wahrgenommen, jedoch für die Automatisierung, Digitalisierung und die erfolgreiche Anwendung von KI auch als grosses Hindernis gesehen.

> **«Der Aufwand, den man bezüglich Datenschutz betreibt, ist wirklich enorm.»** *Brühlmann*

Der Datenschutz regelt nicht nur, dass Daten zwischen den Abteilungen grundsätzlich nicht geteilt werden dürfen, sondern auch der Verwendung in der Abteilung selbst. Die Gesetze rund um den Datenschutz werden als sehr streng und hinderlich für den Einsatz von Künstlicher Intelligenz beschrieben. Diese Hürde wurde in mehreren Interviews als das grösste Hindernis genannt, das die öffentliche Verwaltung im Umgang mit KI bewältigen muss.

> **«Wir dürfen die Daten innerhalb der Verwaltung nicht austauschen. Das sehe ich fast als grösstes Risiko bei der Ausbreitung Künstlicher Intelligenz, dass man diese Hürde nicht eliminieren kann. Selbstverständlich datenschutzkonform, selbstverständlich so, dass der Bürger sicher sein kann, dass die Daten für den erwähnten Zweck und nichts anderes verwendet werden. Aber für mich ist die grösste Hürde, dass die öffentliche Verwaltung mit Mauern getrennt ist.»** *Seidler*

Diese hohen datenschutzrechtlichen Hürden sind beim Einsatz neuer Technologien immer am grössten, da man am Anfang die datenschutzrechtlichen Folgen schlechter abschätzen kann.

> **«Das ist juristisch so definiert: Es gibt eine Liste mit den Fällen, bei denen immer eine Vorabkontrolle durchgeführt werden muss. Das wäre bei Künstlicher Intelligenz immer der Fall, weil es sich um eine neue Technologie handelt, die noch nicht häufig eingesetzt wurde.»** *Blonski*

4. Diversität in der Verwaltung

> **«Wenn man die Verwaltung als Gesamtes anschaut, sind wir ein sehr diversifizierter Laden: Wir machen ganz unterschiedliche Sachen und bieten verschiedene Dienstleistungen an.»** *Seidler*

Eine weitere Schwierigkeit beim Einsatz und bei der Entwicklung von staatlicher KI stellt die Diversität der öffentlichen Verwaltung dar. Zum einen unterscheiden sich die Abteilungen, die zur öffentlichen Verwaltung gehören, stark in ihren Inhalten, Prozessen und Formularen. Die öffentliche Verwaltung kümmert sich um Steuerklärungen, Sozialleistungen, Bauvorgaben, Datenschutz und vieles mehr. Das führt dazu, dass technische Lösungen nicht einfach auf andere Abteilungen übertragbar sind:

> **«Lösungen, die wir in der Geoinformation erfolgreich anwenden, lassen sich nicht eins zu eins auf die gesamte Verwaltung anwenden. Die Verwaltung muss analog einem Gemischtwarenladen viele Fachbereiche und verschiedene Aufgabenbereiche und damit verschiedene Aspekte des öffentlichen Lebens abdecken.»** *Boller*

Zum anderen ist die Verwaltung für die gesamte Bevölkerung zuständig und kann sich in dem Sinne ihre «Kundinnen und Kunden» nicht aussuchen. Die öffentliche Verwaltung ist für unterschiedlichste Bevölkerungsgruppen und Menschen mit diversen Einstellungen, Erwartungen, Bildungshintergrund und sozioökonomischem Hintergrund zuständig.

> **«Als kantonales Amt bzw. Verwaltung erbringen wir Leistungen für alle unsere Kunden; die ganze Breite der Öffentlichkeit. Wir haben ein sehr breites Kundensegment von jung bis alt, von gut bis schlecht ausgebildet, verschiedene ethnische Gruppen, verschiedene kulturelle Hintergründe. Die Digitalisierung hat die Herausforderung, diese vielfältigen Ansprüche abdecken zu müssen, um das Thema für alle weiterzubringen.»** *Boller*

Die Diversität der «Kundschaft» spiegelt sich u. a. auch in den Einstellungen zu Künstlicher Intelligenz wider. Manche Bürgerinnen und Bürger fordern eine deutlich stärker automatisierte Verwaltung, während andere möglichst keine Automatisierung wünschen. Für die Verwaltung stellt es folglich eine grosse Herausforderung dar, diesen unterschiedlichen Erwartungen und Anfragen gerecht zu werden.

5. Dezentralität der Schweiz

Die Schweiz ist dezentral organisiert, was zur Folge hat, dass es je nach Kanton und oft sogar auf Ebene der Gemeinden unterschiedliche Verfahren und teilweise unterschiedliche Gesetze gibt, z. B. die Steuergesetzgebung.

> **«Wir haben ein föderales System. In der Regel handelt es sich um Bundesgesetze, die jeder Kanton ein bisschen anders durchsetzt. Zu denselben Gesetzen existieren unterschiedliche Prozesse und andere Arbeitswege. All dem müssen die Systeme gerecht werden. Der Prozess hin zu einer gemeinsamen Softwarelösung kann dann sehr aufwendig sein, weil jeder Fall abgedeckt werden muss.»** *Bächinger*

Für KI-Anwendungen und die Automatisierung sind eine Vereinheitlichung und Standardisierung aber sinnvoll, um sie auf möglichst viele Fälle anwenden zu können. Hier Lösungen zu finden, die alle Bedürfnisse abdecken, ist herausfordernd. Aus den Interviews ergibt sich die Empfehlung, eine Automatisierung von Verfahren möglichst einheitlich in der Schweiz umzusetzen und gemeinsame Lösungen zu finden, damit die Lösungen kompatibel sind und möglichst viele davon profitieren.

83 Vgl. zum mangelnden Datenfluss in der Verwaltung auch z. B. Bericht Herausforderungen, 2019, S. 87; Opiela/Mohabbat Kar/Thapa/Weber, 2018, S. 34 und 37 f.

«Ich denke, man muss das grosse Ganze im Auge behalten. Man sollte nicht punktuelle Veränderungen vornehmen, sodass ein Bereich vorzieht und andere nachziehen, sondern die Politik muss eine Klammerbewegung machen.» *Seidler*

6. Fehlende Rechtsgrundlagen

Aufgrund des Legalitätsprinzips, an welches jede Verwaltungsbehörde gebunden ist, wird die Verwaltung bei der Entwicklung und Anwendung von neuen Technologien zusätzlich stark durch die rechtlichen Grundlagen gebremst. Bei den Expertinnen und Experten bestand das Gefühl, dass die Gesetze oftmals der technischen Entwicklung «hinterherhinken» würden.

«Bei neuen Technologien muss eventuell die Frage der Gesetzgebung gestellt werden: ob eine weitere Regelung erforderlich ist oder ob diese neue Technologie unter die aktuell bestehenden Regelungen fällt.» *Blonski*

Diese Frage stellt sich insbesondere bei Anwendungen der Künstlichen Intelligenz mit selbstlernenden Komponenten, die sich, nachdem sie programmiert wurden, verändern. Deshalb wurde u. a. explizit vorgeschlagen, KI-Technologien mit selbstständigen Lernelementen gesetzlich separat zu regeln.

III. Mensch

Das letzte Cluster bezieht sich, wie eingangs bereits erwähnt, schliesslich noch auf die Beziehung zwischen Menschen und Maschinen, welche ebenfalls zentrale Hindernisse für einen erfolgreichen KI-Einsatz darstellen können. Dabei geht es einerseits um die persönliche Einstellung gegenüber der digitalen Transformation der Verwaltung und anderseits um das Verständnis der neuen Technologien.

1. Bevölkerung

«Aufseiten der Bürger, das merken wir immer wieder, fehlt teilweise die Akzeptanz.» *Seidler*

Damit KI-Anwendungen in der öffentlichen Verwaltung erfolgreich eingesetzt werden können, müssen diese zwingend von der Bevölkerung akzeptiert werden. Da sich der KI-Einsatz zurzeit im Kanton Zürich wie auch in der gesamten Schweiz erst in einer Einführungsphase befindet, ist auch die Bevölkerung wenig mit dieser absehbaren Transformation vertraut und steht ihr dementsprechend teilweise kritisch gegenüber. In den Interviews wurden vier mögliche Ansätze erwähnt, welche diese Akzeptanz in der Bevölkerung erhöhen könnten.

«Die Frage ist, wie sorgfältig wird die KI-Anwendung eingeführt, wie wird sie kontrolliert und sichergestellt, dass das Ergebnis stimmt. Diese Lösungen müssen das tun, was sie sollen, und fehlerfrei arbeiten, damit die Qualität stimmt.» *Bächinger*

Erstens stand zunächst die Qualitätssicherung im Zentrum. Die Verwaltungsbehörden müssen wohl mehr als private Unternehmen sichergehen, fehlerfreie Anwendungen anzubieten. Wenn die Qualität der digitalen Dienstleistungen gesichert ist, wird auch das Vertrauen der Bevölkerung steigen.

«Das Wichtigste ist Transparenz: Die Bevölkerung muss wirklich wissen, wie etwas gemacht wird. Dadurch kann das Vertrauen in die Digitalisierung und die Anwendung neuer Technologien gestärkt werden.» *Blonski*

Zweitens wurde mehrfach die Transparenz vonseiten der Verwaltung gegenüber der Bevölkerung rund um KI-Anwendungen und deren Planung betont. Je mehr die Bevölkerung über die Technologien informiert ist und deren Funktionsweise nachvollziehen kann, desto höher ist auch ihr Vertrauen in sie.

«Das Gerechtigkeitsgefühl spielt hier eine Rolle: dass ich sicher sein kann, dass ich gleich und fair behandelt werde wie die anderen.» *Seidler*

Drittens muss sichergestellt sein, dass die KI-Anwendungen fair sind. Von zentraler Bedeutung ist, dass die Bürgerinnen und Bürger überzeugt sind, gleich wie ihre Mitmenschen behandelt zu werden. Gibt es Anzeichen, dass eine KI-Technologie bestimmte Bevölkerungsgruppen benachteiligt oder einem einzelnen Kriterium unverhältnismässig viel Gewicht zumisst, besteht die Gefahr, dass die Bevölkerung das Vertrauen in die Technologie rasch verliert.
Die Akzeptanz ist schliesslich das Ergebnis eines kontinuierlichen Prozesses, in welchem die Bevölkerung mehr über die neuen Technologien und ihre Chancen sowie Schwierigkeiten lernen muss, damit sie diese besser einordnen und ihre Meinung dazu bilden kann. Die Bevölkerung darf auf keinen Fall das Gefühl haben, die Kontrolle über die Entwicklung zu verlieren, sondern muss bei wichtigen Grundsatzentscheidungen beteiligt sein.[84]

2. Mitarbeitende

«Bei den Mitarbeitenden braucht es Geduld, die Automatisierung muss schrittweise eingeführt werden. Man muss die Mitarbeitenden mitnehmen, viel kommunizieren, Zeit lassen und Übergangsangebote bieten.» *Seidler*

Die Verwaltungsmitarbeitenden sind auch Teil der Schweizer Bevölkerung, somit spiegeln sich im Umkehrschluss die Ängste der Bevölkerung auch in den Verwaltungsmitarbeitenden wider. Fehlende Akzeptanz bei den Mitarbeitenden beeinträchtigt einen erfolgreichen KI-Einsatz folglich ebenfalls. Im Zusammenhang mit einer kontinuierlichen Transformierung muss daher sichergestellt werden, dass ihre Bedenken und Zukunftsängste adressiert werden. Dies bedeutet einerseits, den Mitarbeitenden KI als Chance für ihre Tätigkeit und nicht als Gefahr für ihre Arbeitsplätze zu präsentieren, sowie anderseits, die Mitarbeitenden bei der Veränderung auch zu unterstützen und auszubilden.[85]

«Die Leute in meinem Umfeld sind fasziniert von KI-Anwendungen. Sie finden es aber schwierig, die Folgen und Auswirkungen einzuschätzen.» *Blonski*

84 Vgl. auch Kapitel 2 E.
85 Vgl. dazu ebenfalls Bericht Herausforderungen, 2019, S. 87.

Die Ausbildung von Verwaltungsmitarbeitenden sowie die KI-Kompetenzsicherung innerhalb der Verwaltung sind auf drei verschiedenen Ebenen anzusetzen: Um die Akzeptanz bei den Mitarbeitenden gegenüber KI sicherzustellen und zu verhindern, dass entwickelte Technologien nicht angewendet werden, müssen Verwaltungsmitarbeitende zuerst ganz allgemein im Umgang mit digitalen Technologien und KI ausgebildet werden (Digital Literacy oder Data Literacy).[86] Die Verwaltung könnte dadurch eine Führungsrolle dabei übernehmen, die Bevölkerung durch die Ausbildung ihrer Mitarbeitenden über KI aufzuklären, denn durch ihre Bildung können Berührungsängste und digitale Gräben überwunden werden, was sich dann in einer erhöhten Akzeptanz für den Einsatz von KI auch in der Bevölkerung widerspiegelt.[87] Ferner müssen die Verwaltungsmitarbeitenden ebenfalls zwingend bei der Anwendung und Implementation von konkreten KI-Anwendungen begleitet werden, da fehlende Investitionen in KI-Kompetenzen das Risiko von Fehlbedienungen und Fehlinterpretationen erhöhen.[88] Werden die KI-Anwendungen von den Mitarbeitenden nicht richtig verstanden, steigt auch das Risiko, die Verantwortung an IT-Abteilungen abzugeben. Hinzu kommt, dass dann, wenn kein Verständnis vorhanden ist, wie die KI-Ergebnisse zustande kommen, weniger Selbstvertrauen bei den Verwaltungsmitarbeitenden vorhanden ist, der KI-Anwendung zu widersprechen. Auch wenn dieser Aspekt der «Maschinenhörigkeit» in den Interviews nicht direkt angesprochen wurde, besteht eine bestimmte Tendenz von Menschen, ein KI-Ergebnis auch bei offensichtlich überraschenden Ergebnissen erklären zu wollen und bei einer abweichenden Position von einem KI-System das Gefühl zu haben, diese vertieft begründen zu müssen.[89] Insofern muss bei der Anwendung von KI-Systemen in der Verwaltung für die Mitarbeitenden auch Raum für Anpassungen und eigenständige Entscheidungen existieren. Die Systeme müssen bis zu einem bestimmten Grad auch übergangen werden dürfen, ohne dabei bestraft zu werden.[90] In der Literatur wird schliesslich zusätzlich noch angefügt, dass bei einer fehlenden eigenen KI-Expertise das Risiko einer Abhängigkeit von einigen wenigen nationalen oder internationalen Unternehmen steigt.[91]

3. Management

«Ein wesentlicher Punkt ist die Besetzung der Direktion: Was sind das für Leute? Wie funktionieren sie? Wollen sie den Einsatz von KI?» *Brühlmann*

Ein letztes Hindernis kann gemäss den Expertinnen und Experten schliesslich auch das Management sein. Werden in der Direktion KI-Anwendungen nicht gefördert oder akzeptiert, kann KI unabhängig davon, wie überzeugend und erfolgreich die Technologien an anderen Orten sind oder sein könnten, auch nicht zum Einsatz kommen.

IV. Möglichkeiten, den Herausforderungen zu begegnen[92]

Das Feld der Herausforderungen eines staatlichen KI-Einsatzes ist somit sehr breit. Je nach Blickwinkel können die Schwierigkeiten in Hindernisse und Gefahren unterschieden werden. Herausforderungen, welche einer Verwaltungsbehörde einen KI-Einsatz verunmöglichen, können als Hindernisse für die Verwaltung betrachtet werden. Kritische Auswirkungen von KI-Anwendungen auf die Gesellschaft bzw. auf die betroffene Person können als Gefahren oder Risiken für die Bevölkerung bezeichnet werden. Die Unterscheidung von Hindernissen und Risiken ist zentral für die Entscheidung, wie ihnen begegnet werden kann. Manche Hindernisse wie etwa die Diversität der Verwaltung lassen sich nicht überwinden und müssen bei der Planung und Umsetzung von KI berücksichtigt werden. Andere Hindernisse können hingegen mit gezielten Massnahmen zumindest teilweise beseitigt werden.[93] Die Gefahren eines staatlichen KI-Einsatzes lassen sich jedoch nicht grundsätzlich eliminieren. Sie können mit geeigneten Regulierungsmassnahmen aber zumindest reduziert werden.

Schon bei den Chancen der KI hat sich gezeigt, dass es schwierig ist, diese allgemein aufzuführen. Ähnlich verhält es sich mit den Risiken, die ebenfalls stark von der jeweiligen Technologie und der konkreten Implementation des Systems im jeweiligen sozialen Prozess abhängig sind. Das Risikopotenzial einer einzelnen KI-Anwendung kann dabei anhand von zwei Dimensionen beurteilt werden.[94] Auf der einen Seite werden das Schadenspotenzial von Fehlurteilen einer KI-Anwendung für die betroffenen Individuen sowie der darüber hinausgehende mögliche gesamtgesellschaftliche Schaden betrachtet. Auf der anderen Seite steht die Frage nach der Wahlfreiheit über eine Unterwerfung unter das KI-System bzw. der Menge an Einsichts-, Widerspruchs- und Reevaluierungsmöglichkeiten. Selbst wenn eine Software monopolartig genutzt wird (wie das bei staatlichen KI-Anwendungen regelmässig der Fall sein wird), reduziert sich das Risiko, sofern es einfach ist, Fehlurteile zu entdecken, anzufechten und zu ändern. Diese beiden Dimensionen können die Gefahr einer konkreten KI-Anwendung genauer beschreiben. Je nach Risiko sind unterschiedliche Massnahmen wie etwa Transparenz- und Kontrollpflichten notwendig. Die Grundlagen für die Erarbeitung dieser Regulierungsforderungen werden in den Kapiteln 3 und 4 dargelegt. Die Identifikation der Chancen einer spezifischen KI-Anwendung dienen schliesslich beim Entscheid über deren Einsatz dazu, diese mit den konkreten Risiken und dementsprechend notwendigen Massnahmen abzuwägen.[95]

86 Vgl. für den Begriff Data Literacy auch Bericht Herausforderungen, 2019, S. 8; CAHAI Study, 2020, S. 74.
87 Vgl. ebenfalls Opiela/Mohabbat Kar/Thapa/Weber, 2018, S. 37.
88 Vgl. ebenfalls Opiela/Mohabbat Kar/Thapa/Weber, 2018, S. 37.
89 Vgl. dazu TA-SWISS KI, 2020, S. 276.
90 Opiela/Mohabbat Kar/Thapa/Weber, 2018, S. 32.
91 Vgl. Martini, 2019, S. 28 ff.; Opiela/Mohabbat Kar/Thapa/Weber, 2018, S. 37; TA-SWISS KI, 2020, S. 58.
92 Die hier vorgestellten Überlegungen beruhen nicht direkt auf den in den Interviews getroffenen Aussagen, sondern stellen eine zusammenfassende Reflexion der Autorinnen und Autoren dar.
93 Vgl. dazu exemplarisch die erwähnten Lösungsansätze in den vorangehenden Abschnitten.
94 Vgl. zu diesem risikobasierten Ansatz und der Einteilung in verschiedene Regulierungsklassen, jedoch sowohl für staatliche als auch private KI-Anwendungen ausführlich Zweig, 2019b, S. 234 ff.
95 Vgl. dazu im Folgenden Kapitel 2 E.

E. Notwendigkeit eines gesellschaftspolitischen Diskurses[96]

«Wir als Verwaltung können zur Akzeptanz beitragen, indem wir sichere Verfahren anbieten. Eine grundlegende Skepsis oder Abneigung muss in der Gesellschaft abgeholt werden. Das kann kein Amt allein machen.» *Seidler*

In den vorangegangenen Abschnitten wurden summarisch die mit einem staatlichen KI-Einsatz verbundenen Erwartungen und Herausforderungen dargestellt. Bestimmte Risiken können dabei durch geeignete Regulierungsmassnahmen reduziert werden. Doch auch dann wird stets ein KI-Restrisiko bestehen, welches mit dem Nutzen der KI-Anwendungen abgewogen werden muss, um zu entscheiden, ob die Implementierung der Technologie dennoch im Interesse des Gemeinwohls liegt. Da die Definition des Gemeinwohls ein politisches Ergebnis ist, braucht es vordergründig einen gesellschaftlichen und politischen Diskurs, um die jeweiligen Interessen- und Güterabwägungen vorzunehmen. Sehr anschaulich ist diese Abwägung bei der zunehmenden Kameraüberwachung im öffentlichen Raum.[97] Möchte die Bevölkerung hier die öffentliche Sicherheit oder individuelle Grundrechte priorisieren? Überwiegt das Bedürfnis nach öffentlicher Sicherheit, sollten intelligente Überwachungskameras in der Lage sein, möglichst viele gesuchte Personen zu erkennen, auch wenn dabei vereinzelt Unschuldige fälschlicherweise beobachtet werden. Priorisiert man hingegen individuelle Grundrechte, sollten intelligente Überwachungskameras so programmiert werden, dass möglichst wenig unschuldige Bürgerinnen und Bürger fälschlicherweise in das Visier der Behörden kommen, wobei dafür in Kauf genommen wird, dass vereinzelt verdächtige Personen nicht erkannt werden. Eine solche Grundsatzentscheidung muss von der Bevölkerung als Gesellschaft getroffen werden.

«Die Bevölkerung muss auf den Weg der Digitalisierung mitgenommen und partizipativ beteiligt werden: nicht nur als Konsumentinnen und Konsumenten staatlicher Leistung, sondern diese auch mitgestalten.» *Boller*

Allerdings bestehen beim «gewöhnlichen» Verwaltungshandeln ohne Einsatz von KI ähnliche Zielkonflikte. Mit dem Einsatz von staatlicher KI gehen jedoch Datenverarbeitungen in neuen Grössenordnungen sowie neuartige Risiken der fehlenden Nachvollziehbarkeit und Fehleranfälligkeit aufgrund der Arbeitsweise von KI-Anwendungen mit Wahrscheinlichkeiten einher. Diese Zielkonflikte müssen transparent ausgewiesen und von der Gesellschaft in zentralen Grundentscheidungen mitbestimmt werden. Die Bevölkerung muss hier gemeinsam entscheiden, in welche Richtung sich die Verwaltungstätigkeit entwickeln soll. Nur so kann das Vertrauen der Bürgerinnen und Bürgern in das staatliche algorithmische Handeln gesichert werden.[98] In dieser Hinsicht ist die Schweiz dank ihrer zahlreichen direktdemokratischen Partizipationsmöglichkeiten anderen Ländern gegenüber im Vorteil. Die technisch komplexen Aspekte bergen jedoch auch für die Schweiz eine zusätzliche Herausforderung. Die heutigen politischen Rechtsetzungsprozesse, nach welchen die Verwaltung die Vorlage ausarbeitet und die wirkliche öffentliche Debatte erst beginnt, wenn die Vorlage im Parlament ist oder den Stimmberechtigten zur Abstimmung unterbreitet wird, genügen bei grossen digitalen Projekten nicht mehr.[99] Der Abstimmungskampf ist der falsche, ungeeignete und zu späte Zeitpunkt, um über technische Aspekte zu diskutieren. Der gesetzgeberische Prozess muss daher von Beginn an einen offenen und transparenten Dialog mit Expertinnen und Experten anstreben, welcher auch der Öffentlichkeit zugänglich ist. Dabei muss auch schon früh die technische Umsetzung wie die Architektur von IT-Lösungen mitgedacht werden, denn die wichtigsten Entscheidungspunkte sind eng mit ihr verknüpft. Ist die Technik zum Zeitpunkt der Abstimmung noch unklar oder beginnt erst dann der Dialog zwischen Fachpersonen und der Öffentlichkeit, bleibt nicht mehr ausreichend Zeit, um die Grundsatzentscheidungen noch fundiert zu treffen. Der gesellschaftspolitische Diskurs bezüglich eines staatlichen KI-Einsatzes ist folglich nicht nur zwingend notwendig, sondern auch besonders gut und teilweise neu zu gestalten.

«Es braucht eine gesellschaftliche Diskussion, die dann letztendlich auch in die Politik mündet. Bei Fragen zu neuen Technologien müssen wir also auch beim Kantonsrat ansetzen. Bei jedem neuen Einsatz einer Technologie muss diskutiert werden: Wie möchten wir als Gesellschaft damit umgehen? Wir sind eine demokratische und freiheitliche Gesellschaft und haben uns für Grundrechte entschieden. Es ist also schlussendlich eine politische Entscheidung darüber, wie wir zusammenleben wollen und was wir zulassen oder nicht zulassen wollen – immer unter der Berücksichtigung der jeweiligen Konsequenzen.» *Blonski*

F. KI in der öffentlichen Verwaltung

I. Allgemeine Überlegungen

KI-Anwendungen können unterschiedliche Aufgaben in den verschiedensten Verwaltungsbereichen übernehmen.[100] Die Verwaltung kann KI z. B. zur Kommunikation mit Bürgerinnen und Bürgern[101] oder zur Unterstützung bei administrativen Tätigkeiten im Hintergrund nutzen. So gibt es Anwendungen, die basierend auf der aktuellen Auslastung der Mitarbeitenden automatisiert Aufgaben verteilen, Dokumente übersetzen oder Gesprächsprotokolle anfertigen. Diese möglichen KI-Einsatzfelder betreffen keine rechtlich verbindlichen Anordnungen an Bürgerinnen und Bürger, weshalb sie generell als weniger risi-

96 Die hier gemachten Ausführungen wurden zwar zum Teil in den Interviews angesprochen. Aufgrund der zentralen Bedeutung eines gesellschaftspolitischen Diskurses wurde dies jedoch als Anlass genommen, die vereinzelten Kommentare ausführlicher darzustellen und zu ergänzen.

97 Vgl. dazu Opiela/Mohabbat Kar/Thapa/Weber, 2018, S. 37.

98 Opiela/Mohabbat Kar/Thapa/Weber, 2018, S. 37.

99 NZZ vom 10.02.2021, https://www.nzz.ch/meinung/der-bund-kann-digitalisierung-nicht-er-muss-sie-lernen-ld.1599750.

100 Ausführlich zu den folgenden Anwendungsfelder Etscheid/Von Lucke/Stroh, 2020, S. 22 ff.

101 Vgl. dazu insbesondere die Ausführungen zu den Chatbots in Kapitel 2 F. II. 5.

kobehaftet einzustufen sind.[102] KI kann aber auch im Rahmen der Entscheidungstätigkeit der Verwaltung eingesetzt werden. Dabei ist zwischen Systemen, die Verwaltungsmitarbeitende in der Entscheidung unterstützen (Teilautomation), und solchen, die ihnen die Entscheidung gänzlich abnehmen (Vollautomation), zu unterschieden. In der Praxis ist die Entscheidungsunterstützung die relevantere Form. Diese kann weiter unterteilt werden. Ex ante können KI-Systeme dazu dienen, relevante Informationen als Entscheidungsgrundlage zur Verfügung zu stellen und/oder Entscheidungsvorschläge zu unterbreiten. KI-Systeme können jedoch auch ex post zur Entscheidungsunterstützung eingesetzt werden, indem sie die menschlichen Entscheidungen im Nachhinein überprüfen.
Diese möglichen Anwendungen eröffnen für den Staat im Unterschied zur Privatwirtschaft keine neuen Betätigungsfelder, denn staatliche Aufgaben sind grundsätzlich vorgegeben. KI wird in der öffentlichen Verwaltung folglich zur Erfüllung bestehender Aufgaben eingesetzt (werden).[103] Die Entwicklung von innovativen, neuartigen KI-Anwendungen ist – im Unterschied zur Privatwirtschaft – kein Treiber für den KI-Einsatz in der öffentlichen Verwaltung. Vielmehr geht es darum, bestehende Aufgaben effizienter, rechtssicherer oder exakter erledigen zu können. Dies gelingt am besten bei strukturierten Prozessen mit grossem Datenvolumen. Aufgaben, welche komplexe Einzelfallbetrachtungen erfordern, eignen sich zumindest heute noch nicht für den KI-Einsatz. Das grösste staatliche KI-Potenzial liegt demnach in der Massenverwaltung, wo Verwaltungsbehörden für eine grosse Anzahl von ähnlich gelagerten Fällen einzelne Verfügungen erlassen müssen.[104] Dies spiegelt sich auch in den identifizierten aktuellen (geplanten) KI-Anwendungen auf kantonaler und nationaler Ebene wider. Im Folgenden werden – ohne Anspruch auf Vollständigkeit – diese Bereiche umschrieben (II.) sowie einzelne konkrete Beispiele von KI-Anwendungen aus anderen Ländern skizziert (III.).

II. KI-Anwendungsbeispiele in Schweizer Verwaltungen

Nadja Braun Binder
Catherine Egli
Laurent Freiburghaus
Eliane Kunz
Nina Laukenmann
Liliane Obrecht

1. Steuerverfahren

Durch das Vorliegen von grossen strukturierten Datenmengen sind KI-Anwendungen im Steuerverfahren als klassisches Massenverwaltungsverfahren grundsätzlich sehr naheliegend.[105] So prüfen aktuell verschiedene Kantone einen KI-Einsatz und gehen davon aus, dass er innerhalb der nächsten drei bis zehn Jahre gesamtschweizerisch eine wesentliche Rolle spielen wird.[106] Bereits heute werden in den meisten Kantonen die digital eingereichten Steuererklärungen durch automatische Veranlagungsprogramme zumindest teilweise automatisiert bearbeitet. In Zukunft soll KI die Steuerverwaltungen jedoch noch weiterführend unterstützen. Vordergründig wird zurzeit insbesondere die vollautomatische Veranlagung gefördert. Im Kanton St. Gallen betrug der Anteil der vollautomatisierten Verfahren für die Steuerperiode 2016 beispielsweise bereits 5 Prozent,[107] während in Bern sogar fast 20 Prozent der steuerpflichtigen natürlichen Personen vollständig automatisiert veranlagt werden.[108] Auch die Steuerverwaltung des Kantons Obwalden führte ein KI-Projekt ein, um einen Automatisierungsgrad von 20 Prozent zu erreichen.[109] Der Kanton Bern möchte nun jedoch noch weiter gehen und versucht, den Automatisierungsgrad durch maschinelles Lernen gar auf 40 Prozent zu erhöhen.[110] Neben der Ermöglichung einer vollautomatisierten Veranlagung könnte KI aber auch als Entscheidungsunterstützung eingesetzt werden. Einerseits könnten bei den manuell zu veranlagenden Fällen die natürlichen Fachpersonen mittels konkreter Hinweise auf Fehlerbereiche sowie mit Lösungsvorschlägen unterstützt werden.[111] Anderseits könnte KI für einen automatischen Abgleich zwischen den Steuererklärungen und den eingereichten Dokumenten eingesetzt werden.[112]
Auf Bundesebene könnte KI mittelfristig bei der Eidgenössischen Zollverwaltung (EZV) eingesetzt werden.[113] Das Programm DaziT[114], in dessen Rahmen bis 2026 alle Zollprozesse digitalisiert werden sollen, soll eine automatisierte Erhebung verschiedener Verbrauchsteuern ermöglichen. Entsprechende Rechtsgrundlagen für einen Einsatz von automatisierten Verfügungen betreffend Zollabgaben, Tabaksteuern, Mineralölsteuern sowie Schwerverkehrsabgaben sind im Anhang des

102 Vgl. zum risikobasierten Ansatz auch Kapitel 2 D. IV.
103 TA-SWISS KI, S. 211.
104 Djeffal, 2020, S. 7.
105 Nufer, 2019/2020, S. 264.
106 Die folgenden Informationen basieren u. a. auf telefonischen Auskünften von D. Widmer, Steueramt Aargau, 23.09.2020; Ph. Moos, Steueramt Zug, 01.10.2020; A. Daepp, Steuerverwaltung Bern, 25.09.2020; M. Nufer, Steuerverwaltung Obwalden, 01.10.2020; T. Hunkeler, Serviceverantwortlicher für die Veranlagungssoftware NEST bei der KMS AG, 30.09.2020.
107 Vgl. St. Galler Tagblatt vom 08.04.2018, Digitalisierung: Ostschweizer Steuerverwaltungen setzen auf Roboter, https://www.tagblatt.ch/ostschweiz/digitalisierung-ostschweizer-steuerverwaltungen-setzen-auf-roboter-ld.1006738.
108 Fischer/Daepp, 2019, S. 329.
109 Projektpräsentation «Data Analytics in der Steuerverwaltung», https://www.egovernment-wettbewerb.de/praesentationen/2020/Data_Analytics_in_der_Steuerveranlagung_Obwalden_Automatische_Veranlagung.pdf.
110 Fischer/Daepp, 2019, S. 327.
111 Vgl. so z. B. das Projekt im Kanton Obwalden, Projektpräsentation «Data Analytics in der Steuerverwaltung», https://www.egovernment-wettbewerb.de/praesentationen/2020/Data_Analytics_in_der_Steuerveranlagung_Obwalden_Automatische_Veranlagung.pdf.
112 Nufer, 2019/2020, S. 264; vgl. dazu auch die Ausführungen zu Risikomanagementsystemen in Kapitel 3 B. I.
113 Vgl. für die folgenden Ausführungen Braun Binder, 2020b, S. 261 ff., sowie https://www.ezv.admin.ch/ezv/de/home/themen/projekte/dazit.html.
114 DaziT steht dabei für «Dazi», das rätoromanische Wort für Zoll, und für Transformation.

neuen Datenschutzgesetzes bereits eingeführt worden.[115] Die EZV wird somit grundsätzlich berechtigt, vollautomatische Zollbescheide ohne menschliches Zutun zu erlassen.[116]

2. Sozialversicherungsverfahren

Ein weiteres klassisches Gebiet der Massenverwaltung ist das Sozialversicherungsverfahren. Auch im Sozialrecht bieten sich daher viele routinegeprägte, eher rechen- als entscheidungsintensive Verfahren für eine Automatisierung wie etwa die Gewährung von Prämienverbilligungsansprüchen an.[117] Dennoch finden sich in der Schweiz abgesehen vom Einsatz von Chatbots[118] wenig KI-Technologien.[119] Der Kanton Genf ist soweit ersichtlich der erste Kanton, welcher KI zur Sozialversicherungsbetrugsbekämpfung einsetzen möchte. Im Januar 2019 trat das Gesetz zur Gewährung eines Investitionskredits für die Entwicklung neuer Informations- und Kommunikationssysteme im Bereich der Sozialleistungen in Kraft. Damit sollen u. a. ungerechtfertigte Sozialleistungsbezüge aufgedeckt werden können. KI-Systeme werden als Mittel zur Erreichung dieser Ziele betrachtet. Die Entwicklung von Algorithmen für Warn- und Erinnerungssysteme, zur Durchführung von Konsistenzkontrollen und Kreuzanalysen der Einkommens- und Vermögensschwankungen werden als erforderlich angesehen, um die teilweise ausserordentlich grossen Datenmengen bewältigen zu können.[120]
Das Bundesamt für Statistik hat wiederum 2017 seine Dateninnovationsstrategie veröffentlicht, um aus den zunehmenden Datenmengen statistische Erkenntnisse zu ziehen, welche einen gesellschaftlichen Mehrwert schaffen könnten. Mit dem Pilotprojekt «Machine Learning Soziale Sicherheit» wird zurzeit insbesondere versucht, typische Sozialleistungsbezugsverläufe darzustellen. Darauf aufbauend soll mittelfristig frühzeitig erkannt werden können, ob eine Person sich rasch wieder in den Arbeitsmarkt integrieren wird oder überdurchschnittliche Disqualifikations- und Ausgrenzungstendenzen aufweist. Derzeit ist jedoch noch nicht zu beurteilen, in welchem Rahmen diese Ergebnisse in die bestehenden Prozesse eingebaut werden würden.[121]

3. Polizeiarbeit

Der Einsatz von Künstlicher Intelligenz durch die Polizei hat einerseits durch die breite mediale Berichterstattung und anderseits durch den wissenschaftlichen Diskurs bereits erhebliche Aufmerksamkeit gewonnen und wird in dieser Studie deshalb nicht weiter vertieft.[122] Für eine bessere Übersicht werden hier trotzdem einige Entwicklungen präsentiert.

a) Predictive Policing

Predictive Policing (vorausschauende Polizeiarbeit) bezeichnet die Verwendung von Softwareprogrammen durch Polizeibehörden, welche basierend auf historischen Kriminalitätsdaten mittels Algorithmen Vorhersagen über zukünftige Delikte treffen.[123] Die Erstellung der algorithmenbasierten Prognose kann sich entweder auf den Tatort (ortsbezogenes Predictive Policing) oder die Täterschaft bzw. gefährdete Personen (personenbezogenes Predictive Policing) beziehen.[124]

i. Ortsbezogenes Predictive Policing in der Schweiz

In der Schweiz betreiben die Kantonspolizeien Aargau und Basel-Landschaft sowie die Stadtpolizei Zürich ortsbezogenes Predictive Policing mit der kommerziellen, in Deutschland entwickelten Software PRECOBS. Die Software wird in der Schweiz zurzeit ausschliesslich zur Bekämpfung von Wohnungseinbruchsdiebstählen eingesetzt. Allerdings ist die Anwendung auf weitere Deliktstypen grundsätzlich möglich und zurzeit in Planung.[125] PRECOBS basiert auf der kriminologischen Near-Repeat-Theorie, welche besagt, dass professionelle Einbrecher häufig serienmässig arbeiten und damit örtlich und zeitlich konzentriert zuschlagen.[126] Dieses Phänomen soll genutzt werden, um Vorhersagen über erhöhte Wahrscheinlichkeiten für Wohnungseinbrüche in bestimmten Gebieten zu bestimmten Zeiten zu treffen.[127] Im Juni 2020 hat auch der Kanton Basel-Stadt ein Projekt zur automatisierten Analyse von Lagedaten in Auftrag gegeben, wobei ebenfalls Predictive Policing-Anwendungen geprüft werden sollen.[128]

ii. Personenbezogenes Predictive Policing in der Schweiz

Sechs Kantone (Glarus, Luzern, St. Gallen, Solothurn, Thurgau und Zürich) arbeiten (bzw. arbeiteten im Falle Zürichs) mit dem Analysetool DyRiAS-Intimpartner.[129] Das Tool analysiert das Risikopotenzial einer männlichen Person, ein Gewaltdelikt gegen die aktuelle oder ehemalige Partnerin zu begehen. Das Tool gibt auf der Grundlage eines beantworteten Fragekatalogs eine Risikoeinschätzung ab, ob die Person eine schwere Gewalttat gegen die Partnerin begehen wird. Neuerdings kann das System zusätzlich auch Handlungsoptionen für das Fallmanagement aufzeigen.[130] Während DyRiAS aus einer Stichprobe mit Daten von tatsächlich straffällig gewordenen Gewalttätern zwar 82 Prozent der Täter in den beiden höchsten Risikostufen einordnete,[131] verübten jedoch nur 28 Prozent der von DyRiAS als gefährlich eingestuften Personen tatsächlich

115 Vgl. Art. 38 Abs. 2 revidiertes Zollgesetz (ZG) vom 18. März 2005, SR 631.0, Art. 11 Abs. 4 revidiertes Bundesgesetz über eine leistungsabhängige Schwerverkehrsabgabe (Schwerverkehrsabgabegesetz, SVAG) vom 19. Dezember 1997, SR 641.81, Art. 18 Abs. 4 revidiertes Bundesgesetz über die Tabakbesteuerung (Tabaksteuergesetz, TStG) vom 21. März 1969, SR 641.31, Art. 21 Abs. 2bis revidiertes Bundesgesetz über die Biersteuer (Biersteuergesetz, BStG) vom 6. Oktober 2006, SR 641.411.
116 Vgl. ausführlich zur Automation des Zollveranlagungsverfahrens Braun Binder, 2020b, S. 261 ff.
117 Martini/Nink, 2017, S. 3.
118 Vgl. dazu Kapitel 2 F. II. 5.
119 Vgl. jedoch für internationale Beispiele Kapitel 2 F. III. sowie NZZ vom 07.12.2020, https://www.nzz.ch/sponsored-content/machine-learning-ohne-schlaue-koepfe-gehts-nicht-ld.1581399 für die Bestrebungen der SUVA, die Automatisierung von Prozessen zu fördern.
120 PL 12386 Projet de loi, S. 4 f.
121 Vgl. zum Ganzen https://www.experimental.bfs.admin.ch/expstat/de/home/innovative-datenwissenschaft/uebersicht.html.
122 Vgl. nur etwa die verschiedenen Beiträge in: Simmler, 2021. Siehe ferner die Hinweise in Kapitel 1 C.
123 Knobloch, 2018, S. 9.
124 Knobloch, 2018, S. 17.
125 Leese, 2018, S. 57 f.
126 Gerstner, 2017, S. 19; Leese, 2018, S. 61.
127 Gerstner, 2017, S. 19.
128 Vgl. BZ vom 30.06.2020, https://www.bzbasel.ch/basel/basel-stadt/die-blitzschnelle-basler-polizei-mit-digitaler-analyse-dem-verbrechen-einen-schritt-voraus-ld.1301733.
129 Simmler/Brunner/Schedler, 2020, S. 14 ff.
130 Vgl. http://dyrias.com/de/instrument/dyrias-intimpartner.html.
131 Vgl. Hoffmann/Glaz-Ocik, 2012, S. 53 f.

ein Gewaltdelikt.[132] DyRiAS arbeitet folglich mit einer Risikoüberschätzung.[133] Im Frühjahr 2018 präsentierte der Kanton Zürich Octagon, ein Programm, welches das gleiche Ziel wie DyRiAS verfolgt. Auch die Funktionsweise erscheint ähnlich, sie ist aber eine Eigenentwicklung des für Justizvollzug und Wiedereingliederung zuständigen Amtes und soll das Risiko deutlich realistischer einschätzen. Dabei handelt es sich auch um eine Reaktion auf die Probleme, welche DyRiAS aufgeworfen hat. Mittlerweile sollen auch die drei Kantone Solothurn, Neuenburg und Tessin mit dem Tool arbeiten.[134]

b) Datenanalyse

Des Weiteren ist die Polizei bei Ermittlungen mit immer grösseren Datenmengen konfrontiert, deren Auswertung auf traditionellem Weg innert nützlicher Frist nicht mehr zu bewältigen ist. Dies betrifft insbesondere die Wirtschaftskriminalität, aber auch die organisierte Kriminalität oder Cyberkriminalität.[135] Eine mögliche Lösung ist der Einsatz von KI bei der Datenanalyse. In verschiedenen Kantonen setzt die Polizei daher auf das Programm WATSON von IBM. WATSON kann Eingaben in natürlicher Sprache verstehen und riesige Datenmengen auf korrespondierende Daten durchsuchen.[136] Im Kontext der Polizeiarbeit bedeutet das, dass WATSON den Ermittlerinnen und Ermittlern Namen, Adressen, Kontonummern oder ähnliche Informationen, welche in der durchsuchten Datenmenge besonders häufig vorkommen, anzeigen sowie Zusammenhänge zwischen den einzelnen Treffern darstellen kann. In der Schweiz arbeitet die Kantonspolizei Zürich bereits seit 2016 mit WATSON, die Kantonspolizei Luzern plante einen operativen Einsatz ab 2020.[137] Die Kantonspolizei St. Gallen hat das Tool getestet, setzt es aber noch nicht vollumfänglich ein, weil es nach ihrer Meinung noch nicht genug Nutzen zeigt. Dennoch möchte sie das Programm beobachten und es für einzelne Arbeiten einsetzen, z. B. für die gezielte Durchsuchung von Polizeiberichten auf bestimmte Inhalte.[138]

c) Videoanalyse

Besonders gut funktioniert KI bereits bei der Bilderkennung. Deshalb ist es nicht überraschend, dass sich dies auch die Polizei zunutze machen will, insbesondere, weil im Bereich von Videos die menschliche Auswertung aufgrund wachsender Datenmengen zunehmend schwierig wird. Die Kantonspolizei St. Gallen verwendet verschiedene Programme zur automatischen Durchsuchung und Auswertung von Videomaterial.[139] Die Software soll sowohl Gesichter als auch Objekte erkennen können. Dabei geht es um Bilder von Überwachungskameras und um Aufnahmen von Zeugen. Die Evaluation sollte Ende 2020 abgeschlossen werden.[140] Schon einen Schritt weiter ist die Kantonspolizei Luzern: Sie nutzt eine Software, welche Videoaufnahmen auswertet und dabei Menschen nach Geschlecht und Alter, Objekte und Farben unterscheiden kann.[141]

d) Automatische Fahrzeugerkennung und Verkehrsüberwachung

Automatische Fahrzeugfahndung und Verkehrsüberwachung (AFV) bezeichnet das Erfassen der Kontrollschilder von Fahrzeugen mittels einer Kamera. Dabei werden die Identität der Fahrzeughalterin oder des Fahrzeughalters sowie Zeitpunkt, Standort, Fahrtrichtung und andere Fahrzeuginsassen ermittelt. Diese Daten werden dann automatisch mit anderen Datenbanken abgeglichen.[142] Die AFV kann vielfältigen Zielen dienen und etwa gestohlene Fahrzeuge finden oder Kriminelle verfolgen. Der grösste Anteil der aktiven AFV-Kameras steht unter Kontrolle des Bundes: Das Grenzwachtkorps setzt rund 300 davon zur Bekämpfung von grenzüberschreitender Kriminalität ein.[143]

4. Justizvollzug

In der Urteilsfindung hat KI in der Schweiz zwar (noch) keine Anwendung gefunden, dafür werden im Strafvollzug aber anhand des Programms ROS (Risikoorientierter Sanktionenvollzug) die Möglichkeiten von Vollzugslockerungen mit dem Ziel geprüft, das Rückfallrisiko während und nach dem Vollzug zu senken. ROS überführt Daten zu einer Person aus dem Strafregisterauszug wie Alter, begangene Gewaltdelikte vor dem 18. Lebensjahr, Anzahl der Vorstrafen oder Strafmass in das vollautomatisierte Fall-Screening-Tool (FaST). Dieses realisiert eine Triage in drei Risikokategorien bezüglich der Flucht- und Rückfallgefahr der inhaftierten Person, welche als Basis für die Entscheidung dient, ob eine vertiefte risikoorientierte Einzelfallanalyse notwendig ist.[144] ROS wurde zwischen 2010 und 2013 in einem vom Bundesamt für Justiz unterstützten Modellversuch in den Kantonen Luzern, St. Gallen, Thurgau und Zürich getestet. Inzwischen wurde ROS in der ganzen Deutschschweiz umgesetzt.[145]

132 SRF vom 05.04.2018, «Predictive Policing» – Polizeisoftware verdächtigt zwei von drei Personen falsch, https://www.srf.ch/news/schweiz/predictive-policing-polizei-software-verdaechtigt-zwei-von-drei-personen-falsch.
133 Vgl. Automating Society Report 2020, S. 257 f.
134 SRF vom 05.04.2018, «Predictive Policing» – Polizeisoftware verdächtigt zwei von drei Personen falsch, https://www.srf.ch/news/schweiz/predictive-policing-polizei-software-verdaechtigt-zwei-von-drei-personen-falsch.
135 Luzerner Zeitung vom 26.10.2019, Ein Computer soll für die Luzerner Polizei ermitteln, https://www.luzernerzeitung.ch/zentralschweiz/luzern/kuenstliche-intelligenz-ein-computer-soll-fuer-die-luzerner-polizei-ermitteln-ld.1163004.
136 Söbbing, 2018, S. 64.
137 Vgl. Das Abraxas Magazin für die digitale Schweiz, Auf der Datenspur des Verbrechens, https://magazin.abraxas.ch/fokus/big-data-bei-der-polizei.
138 Vgl. St. Galler Tagblatt vom 28.01.2020, «Wir werten sogar Roboter-Staubsauger aus»: Die St. Galler Kantonspolizei ist bei Ermittlungen mit immer grösseren Datenmengen konfrontiert, https://www.tagblatt.ch/ostschweiz/wir-werten-sogar-roboter-staubsauger-aus-die-stgaller-kantonspolizei-ist-bei-ermittlungen-mit-immer-groesseren-datenmengen-konfrontiert-ld.1189325.
139 Vgl. St. Galler Tagblatt vom 28.01.2020, «Wir werten sogar Roboter-Staubsauger aus»: Die St. Galler Kantonspolizei ist bei Ermittlungen mit immer grösseren Datenmengen konfrontiert, https://www.tagblatt.ch/ostschweiz/wir-werten-sogar-roboter-staubsauger-aus-die-stgaller-kantonspolizei-ist-bei-ermittlungen-mit-immer-groesseren-datenmengen-konfrontiert-ld.1189325.
140 Vgl. inside-it.ch vom 18.06.2020, Facial Recognition soll bald St. Galler Polizei unterstützen, https://www.inside-it.ch/de/post/facial-recognition-soll-bald-st-galler-polizei-unterstuetzen-20200618.
141 Luzerner Zeitung vom 26.10.2019, Ein Computer soll für die Luzerner Polizei ermitteln, https://www.luzernerzeitung.ch/zentralschweiz/luzern/kuenstliche-intelligenz-ein-computer-soll-fuer-die-luzerner-polizei-ermitteln-ld.1163004.
142 NZZ vom 23.10.2019, Automatische Verkehrsüberwachung: Kanton Thurgau muss gesetzliche Grundlage stärken, https://www.nzz.ch/schweiz/automatische-verkehrsueberwachung-kanton-thurgau-muss-gesetzliche-grundlage-staerken-ld.1517103?reduced=true.
143 NZZ vom 28.10.2019, Fahrzeugfahndung darf nicht zur totalen Überwachung führen, https://www.nzz.ch/schweiz/bundesgericht-stellt-automatisch-fahrzeugfahndung-infrage-ld.1517764?reduced=true.
144 Simmler/Brunner/Schedler, 2020, S. 31 f.; Treuthardt/Loewe-Baur/Kröger, 2018, S. 25 ff.
145 Vgl. https://www.rosnet.ch; Simmler/Brunner/Schedler, 2020, S. 14 ff.

5. Chatbots in unterschiedlichen Einsatzgebieten

Unabhängig vom Einsatzgebiet sind Chatbots die in der Schweiz zurzeit wohl am meisten verbreiteten oder getesteten KI-Anwendungen. Gemeinhin sind Chatbots technische Dialogsysteme, mit denen über natürliche Sprache text- oder sprachbasiert kommuniziert werden kann. Sie werden insbesondere dazu eingesetzt, Anfragen zu beantworten oder Aktionen einzuleiten. Als konversationsbasierte Schnittstelle zwischen Mensch und Maschine werden sie in der Verwaltung vor allem zur Information der Bürgerinnen und Bürgern genutzt.[146] Die Kantone St. Gallen, Luzern und Aargau setzen Chatbots im Bereich der Prämienverbilligung bei der Krankenkasse ein.[147] Der Bereich Prämienverbilligung eignet sich besonders für den Einsatz von Chatbots, weil die Verwaltung einerseits mit enorm vielen verhältnismässig simplen, aber dennoch zeit- und personalaufwendigen Anfragen konfrontiert ist und es sich anderseits um einen hochstrukturierten Prozess handelt.[148] In Dialogform liefern Chatbots Informationen, klären einen allfälligen Anspruch ab, leisten Unterstützung bei der Anmeldung und helfen bei der Meldung von Änderungen.[149] Die SVA Aargau konnte ihre telefonischen Anfragen zu Prämienverbilligung beispielsweise um 30 Prozent reduzieren.[150]

Chatbots werden jedoch nicht nur im Bereich der Sozialleistungen eingesetzt, sondern sind fast überall denkbar. Die Stadt St. Gallen verfügt seit zwei Jahren über den Chatbot «Gallus», welcher Informationen zu Veranstaltungen, Parkplätzen, Mobilität, Umzug und Entsorgung liefern kann.[151] Gallus soll stetig mit neuem Wissen gefüttert werden, um so seinen Anwendungsbereich nach und nach auszudehnen. Vorbild ist dabei die Stadt Wien, die über einen Chatbot verfügt, der zu beinahe allen relevanten Themenbereichen informieren kann. Gallus soll in Zukunft jedoch nicht mehr lediglich Informationen zur Verfügung stellen, sondern auch in der Lage sein, Prozesse selbstständig abzuwickeln: Zum Beispiel soll er nicht nur informieren, wie die Frist für die Einreichung der Steuererklärung verlängert werden kann, sondern dies gleich selbst übernehmen.[152]

Auch das Handelsregister- und Konkursamt des Kantons Zug setzt einen digitalen Verwaltungsassistenten ein. Der Chatbot versorgt Kundinnen und Kunden mit Informationen und unterstützt beispielsweise beim Bestellen von Auszügen und Belegen, beim Ausfüllen von Formularen oder bei der Anmeldung eines Unternehmens.[153] Eine Chatbot-Lösung der Stadt Winterthur hilft Kundinnen und Kunden bei der Informationsbeschaffung sowie der Wahl des richtigen Einbürgerungsverfahrens.[154]

Im Kanton Zürich arbeiten zurzeit verschiedene Verwaltungseinheiten an Chatbot-Projekten.[155] Dabei handelt es sich um Auskunftslösungen. Das Amt für Informatik entwickelt ferner eine Grundlagen-Bot-Lösung für den Kanton Zürich. Im Laufe des Jahres 2021 soll diese von interessierten Verwaltungseinheiten bestellt werden können. Die Grundlagenlösung wird dann mit dem notwendigen Wissen für den Einsatz im jeweiligen konkreten Verwaltungszweig ausgerüstet sein und eine Vielzahl von Funktionen übernehmen können: Geplant ist eine Lösung, die eine intelligente Suche im Internetauftritt des Kantons Zürich ermöglicht, Unterstützung beim Ausfüllen von E-Formularen leistet, Auskunft im Sinne von FAQ (frequently asked questions) gibt und die Bestellung von Leistungen erlaubt. Jugendlichen soll es möglich sein, mit diesem Chatbot eine Konversation zur Unterstützung bei der Berufswahl zu führen. Schliesslich soll der Chatbot eine Authentifizierungsfunktion anbieten, wodurch Portalzugriffe und das Zurücksetzen von Passwörtern ermöglicht werden soll. Die Zürcher Grundlagen-Bot-Lösung ist dabei sowohl als Chatbot als auch als Voicebot geplant, wobei der Voicebot auf Auskünfte im Sinne von FAQ beschränkt sein wird.

6. Weitere Einsatzgebiete

Neben den bereits erwähnten Beispielen sind zahlreiche weitere KI-Anwendungen in den unterschiedlichsten Gebieten der öffentlichen Verwaltung möglich. Dazu zählen etwa automatische Übersetzungshilfen für Verwaltungsdokumente[156] oder der Einsatz von Spracherkennungssoftware, um beispielsweise die Büroautomation (Erstellung von Akten und Dokumenten mittels Spracheingabe) zu fördern.[157] Auf Bundesebene soll weiter mittels eines lernenden Algorithmus eine arbeitsmarktorientierte Kantonsverteilung von Asylsuchenden erfolgen. Das Projekt befindet sich jedoch noch in der Pilotphase, weshalb eine Evaluation und eine Entscheidung über eine dauerhafte Weiterführung erst in den nächsten Jahren zu erwarten sind.[158]

Eine weitere Entwicklung im Zusammenhang mit KI ist die Bewegung in Richtung Smart Cities und damit einhergehend die Verknüpfung von Daten in Städten.[159] Der Staat will durch die Anwendung von Informations- und Kommunikationstechnologien im Rahmen der Erfüllung seiner Aufgaben das Leben

146 Vgl. m. w. H. Kapitel 3 B. III.

147 Für den Kanton St. Gallen vgl. St. Galler Tagblatt vom 28.02.2018, Digitalisierung: Der Beamten-Bot, https://www.tagblatt.ch/ostschweiz/digitalisierung-der-beamten-bot-im-kanton-stgallen-gibt-ein-chat-roboter-auf-facebook-auskunft-ueber-die-praemienverbilligung-ld.1006812; für den Kanton Aargau vgl. Aargauer Zeitung vom 21.01.2020, Entlastung der Mitarbeiter, https://www.aargauerzeitung.ch/aargau/kanton-aargau/entlastung-der-mitarbeiter-maxi-unterstuetzt-die-sozialversicherung-bei-anfragen-zu-praemien-136252636; für den Kanton Luzern vgl. https://www.ahvluzern.ch/online-schalter/chatbot-wasi/.

148 Ringeisen/Bertolosi-Lehr/Demaj, 2018, S. 58; Aargauer Zeitung vom 21.01.2020, Entlastung der Mitarbeiter, https://www.aargauerzeitung.ch/aargau/kanton-aargau/entlastung-der-mitarbeiter-maxi-unterstuetzt-die-sozialversicherung-bei-anfragen-zu-praemien-136252636.

149 Vgl. Fn. 147.

150 Vgl. https://www.sva-ag.ch/sites/default/files/media/document/Medienmitteiung%20Chatbot%20Maxi.pdf.

151 Vgl. für den Chatbot in Wien, https://smartcity.wien.gv.at/site/wienbot/.

152 St. Galler Tagblatt vom 15.10.2018, In St. Gallen kann man bald mit einem Roboter chatten, https://www.tagblatt.ch/ostschweiz/stgallen/in-stgallen-kann-man-bald-mit-einem-roboter-chatten-ld.1061149.

153 Vgl. Medienmitteilung Kanton Zug vom 01.07.2019, https://www.zg.ch/behoerden/volkswirtschaftsdirektion/handelsregisteramt/aktuell/medienmitteilung-chatbot?searchterm=chatbot.

154 Vgl. Medienmitteilung Stadt Winterthur vom 21.09.2020, https://stadt.winterthur.ch/gemeinde/verwaltung/stadtkanzlei/kommunikation-stadt-winterthur/medienmitteilungen-stadt-winterthur/chatbot-fuer-den-bereich-einbuergerungen.

155 Der folgende Abschnitt basiert auf zwei Telefongesprächen mit S. Höltschi, Adretis AG, vom 18.11.2020 und vom 16.12.2020.

156 Bericht Herausforderungen, 2019, S. 87.

157 Vgl. zur Spracherkennung in Behörden https://www.cancom.info/2019/01/spracherkennung-in-behoerden-ueber-digitale-entlastung-bis-hin-zum-datenschutz/.

158 Vgl. Medienmitteilung ETH Zürich vom 18.01.2018, https://ethz.ch/de/news-und-veranstaltungen/eth-news/news/2018/01/algorithmus-verbessert-erwerbschancen-von-fluechtlingen.html.

159 Zanella/Bui/Castellani/Vangelista/Zorzi, 2014, S. 22 f.

in Städten besser und nachhaltiger gestalten.[160] Unter der Bezeichnung Smart City Basel laufen im Kanton Basel-Stadt derzeit beispielsweise über 30 Projekte, welche u. a. eine intelligente Fussgängersteuerung im Strassenverkehr, die Sauberkeitsmessung der Strassen mittels eines Kamerasystems, die nachhaltige Quartierentwicklung, elektronische Patientendossiers, den digitalen Erwerb eines Fischerpatents sowie ein intelligentes Energienetz umfassen.[161]
Hinzuweisen ist schliesslich noch auf weitere KI-Projekte wie Arealstatistik Deep Learning (ADELE), wodurch Luftbildinterpretationen zur Identifizierung und Klassifizierung von Veränderungen[162] automatisiert werden sollen, sowie die Verteilung von Kodierungen der wirtschaftlichen Tätigkeiten von Unternehmen (NOGAuto).[163]

7. Zwischenfazit

Zusammenfassend lassen sich mehrere Charakteristika aktueller und künftiger KI-Einsatzbereiche erkennen. Ein wiederkehrendes Merkmal ist die Standardisierung von Prozessabläufen. Sich häufig wiederholende, strukturierte Tätigkeiten ohne Ermessensspielraum lassen sich gut automatisieren. Dies zeigt sich am bereits hohen Automatisierungsgrad in Steuerverfahren und den verschiedenen dort geplanten Anwendungen. Auch der Einsatz von Chatbots bestätigt diese Tendenz, da diese vor allem dort genutzt werden, wo häufig wiederkehrende ähnliche Informationsanfragen auszumachen sind.
Ein weiteres – mit dem Standardisierungspotenzial verbundenes – Merkmal ist die Massenverwaltung.[164] Typische Beispiele der Massenverwaltung sind das steuerrechtliche und das sozialversicherungsrechtliche Verfahren. Auch das Asylwesen kann dazu gezählt werden.[165] Die im Rahmen dieser Studie eruierten (geplanten) Anwendungen in der kantonalen öffentlichen Verwaltung sind denn auch im Steuerbereich und im Sozialversicherungsverfahren (dort insbesondere die Chatbots im Bereich der Prämienverbilligung) zu finden. Im Asylbereich befindet sich auf Bundesebene derzeit ein Projekt zur KI-unterstützten Kantonszuteilung von Asylsuchenden in der Pilotphase. Es ist naheliegend, davon auszugehen, dass ein KI-Einsatz in Bereichen der Massenverwaltung vor allem aus Gründen der Effizienzsteigerung bei gleichzeitiger Wahrung verfahrensrechtlicher Vorgaben infrage kommt.[166]
Generell bieten sich für den KI-Einsatz auch Bereiche an, in denen eine intelligente Durchsuchung verschiedenster Informationen erforderlich ist. Dieses Merkmal findet sich etwa bei der Polizeiarbeit im Bereich der Wirtschaftskriminalitätsbekämpfung oder bei der intelligenten Videoanalyse. Auch die geplante Zürcher Grundlagen-Bot-Lösung zur Durchsuchung des kantonalen Internetauftritts nach Informationen kann hier zugeordnet werden. Schliesslich wird KI dort eingesetzt, wo es darum geht, Prognosen zu treffen. So dienen Anwendungen im Rahmen des Sanktionenvollzugs und im Bereich des Predictive Policing der statistikbasierten Vorhersage von Verhaltensweisen oder Ereignissen.

III. Internationale Beispiele

Matthias Spielkamp

Im internationalen Vergleich verwenden einige Länder bereits deutlich mehr staatliche KI-Anwendungen als die Schweiz. Im Folgenden werden deshalb ausgewählte aktuelle Beispiele behördlicher KI-Anwendungen aus anderen Ländern näher vorgestellt. Die Beschreibungen wurden mit Ausnahme der Fälle aus Österreich und Schweden aus dem vor einigen Monaten publizierten und aktualisierten Bericht von AlgorithmWatch übernommen und ins Deutsche übersetzt.[167]
In diesem Bericht wird anstelle der Terminologie «KI» und «KI-Systeme» häufig der Begriff «automated decision-making» (ADM) verwendet, da mit dem Einsatz von KI-Anwendungen im Grunde Entscheidungen an automatisierte Verfahren delegiert werden. Die Umschreibung als ADM-Systeme hilft dabei, den Fokus weniger auf die technische Natur («Künstliche Intelligenz») der Anwendungen zu legen, sondern die automatisierten Vorgänge als gesamtes und komplexes System zu betrachten. Ein ADM-System besteht dabei im Kern aus einem – von Menschen entwickelten – Entscheidungsmodell, einem Algorithmus, der dieses Modell in einen ausführbaren Code übersetzt, den Daten, die dieser Code als Eingabe verwendet, um daraus zu «lernen» oder sie durch Anwendung des Modells zu analysieren, und einem Mechanismus, der eine Handlung ausführt. Diese Handlung kann beispielsweise darin bestehen, Daten zu visualisieren, um eine Entscheidung zu unterstützen oder eine Verfügung zu erstellen, die eine rechtliche Wirkung entfaltet. Da ADM-Systeme soziotechnische Systeme sind, ist es wichtig, das politische und wirtschaftliche Umfeld, in dem ihre Verwendung stattfindet, ebenfalls zu betrachten. Das beinhaltet die Entscheidung, ein ADM-System für einen bestimmten Zweck anzuwenden, die Art und Weise, wie es entwickelt wird (d. h. von einer öffentlichen Einrichtung oder einem kommerziellen Unternehmen), wer es beschafft und wie es eingesetzt wird.

1. Profiling Arbeitsloser in Dänemark

Als im Mai 2019 in Dänemark ein neues Gesetz verabschiedet wurde, entbrannte eine öffentliche Diskussion über die eingesetzte automatische Entscheidungshilfe. Der Gesetzestext enthält einen Passus, der es der Arbeitsministerin erlaubte, Richtlinien für den Einsatz eines «digitalen Klärungs- und Dialogwerkzeugs zu entwickeln, das von Jobcentern und Arbeitslosenämtern genutzt werden soll». Das neue Werkzeug sollte die Einschätzung des Arbeitspotenzials der neu gemeldeten arbeitslosen Person mit anderen Daten kombinieren und dann mit den Merkmalen von Bürgerinnen und Bürgern abgleichen, die zuvor in der Langzeitarbeitslosigkeit gelandet waren. Die Software sollte dann diejenigen markieren, die ein höheres Risiko für Langzeitarbeitslosigkeit haben.

160 Djeffal, 2017, S. 810; Glaser, 2018, S. 185; Harasgama, 2017, S. 29.
161 Vgl. dazu https://www.smartcity.bs.ch/projekte-smart-city.html.
162 Vgl. https://www.experimental.bfs.admin.ch/expstat/de/home/innovative-datenwissenschaft/adele.html.
163 Vgl. https://www.experimental.bfs.admin.ch/expstat/de/home/innovative-datenwissenschaft/nogauto.html.
164 Als Massenverwaltung (auch: Massenverfahren) werden Verwaltungsbereiche umschrieben, in denen eine Behörde in einem Aufgabenbereich eine grosse Masse an einzelnen Verfügungen zu erlassen hat; vgl. Kiener/Rütsche/Kuhn, 2015, Rn. 879.
165 Kiener/Rütsche/Kuhn, 2015, Rn. 880.
166 Eine der Herausforderungen der Massenverwaltung besteht darin, die vielen Verwaltungsverfahren innert angemessener Frist zu bewältigen und dabei gleichzeitig die rechtsstaatlichen Anforderungen zu gewährleisten; vgl. Kiener/Rütsche/Kuhn, 2015, Rn. 881.
167 Automating Society Report 2020. Die herangezogenen Quellen sind daher dort verzeichnet. Das Beispiel aus Schweden beruht auf einem bislang unveröffentlichten Artikel von Prof. Dr. Anne Kaun, hier abgedruckt mit Einverständnis der Autorin. Der Text zum AMAS-Fall aus Österreich wurde für diesen Report verfasst.

Die öffentliche Diskussion betraf sowohl den Inhalt des verwendeten Modells als auch den Prozess, wie das Gesetz verabschiedet wurde ohne die zuständige öffentliche Datenschutzbehörde in Dänemark, Datatilsynet, zu konsultieren. In der Folge nahm Datatilsynet seine Evaluation des Gesetzes wieder auf und betonte, dass das Ministerium auf den rund 1000 Seiten Hintergrundmaterial zum Gesetzesentwurf nicht auf das automatisierte System hingewiesen habe. Im August 2019 kam Datatilsynet jedoch zu dem Schluss, dass die Pläne für die automatisierte Profilerstellung von Arbeitslosen den Vorschriften entsprechen, schlug aber auch vor, das System neu zu bewerten, um abzuschätzen, ob das «mathematische Modell weiterhin relevant ist». Das Ziel sei, dabei sicherzustellen, dass die Rechte der registrierten Personen respektiert und «unsachliche Diskriminierung(en) vermieden» werden.
Zwei politische Parteien – beide Unterstützer der aktuellen sozialdemokratischen Minderheitsregierung – kündigten an, dass sie sich gegen das Profiling aussprechen und sich für eine erneute Evaluierung einsetzen. Als Grund gaben sie an, dass sich die Bürgerinnen und Bürger unwohl fühlen könnten und weniger Vertrauen in die Weitergabe von Informationen an Beamte haben würden. Obwohl das Tool nur als Entscheidungshilfe gedacht war, zeigte sich eine Oppositionspolitikerin zurückhaltend: «Die Realität in den Arbeitsagenturen ist, dass die Beamtinnen und Beamten viel zu viele Fälle auf dem Tisch haben. Daher fällt es leicht, ein Profiling-Werkzeug einzusetzen, weil es schneller geht als ein gründliches Gespräch», sagte Eva Flyvholm von Enhedslisten der Nachrichtenagentur Altinget. Flyvholm plädierte auch für eine ausreichende Personalausstattung der Jobcenter.
Eine kommerzielle Lösung, ein sogenannter Assistent namens Asta, ist bereits verfügbar und wird von einer Spezialfirma angeboten. In dem zugehörigen Werbematerial beschreibt der Hersteller, dass Asta «mit einem Werkzeugkasten ausgestattet ist, der verschiedene Technologien wie Robotertechnik und Künstliche Intelligenz enthält». Ferner vergleicht das Unternehmen Asta mit der Person, die für eine Fernsehköchin oder einen Fernsehkoch vor einer Sendung die Kartoffeln schält. Asta wird bereits von der Arbeitsagentur in Kopenhagen eingesetzt und wurde im Onlinemagazin von KL, dem Verband dänischer Kommunen, positiv beurteilt.

2. Kontrolle von Sozialhilfeempfängerinnen und Sozialhilfeempfängern in Dänemark

Udbetaling Danmark (UDK) ist eine öffentliche Behörde, die für Zahlungen wie Stipendien, Kindergeld, die Mindestrente, Elternschaft- und Krankenunterstützung für Arbeitslose sowie für viele andere Leistungen im Rahmen des Sozialstaatssystems zuständig ist. Seit 2012 obliegen dieser zentralisierten Behörde Auszahlungen, die zuvor von den lokalen Gemeinden bearbeitet und kontrolliert wurden. Heute ist die UDK nicht nur für die Zahlungen selbst, sondern auch dafür zuständig, deren Rechtmässigkeit zu kontrollieren. Einige Zahlungen erfolgen automatisch, z. B. die Statens Uddannelsesstøtte (die staatliche Ausbildungsunterstützung, SU), die auf klaren Kriterien basiert, z. B. darauf, ob eine Bürgerin als Studentin oder ein Bürger als Student an einer Universität eingeschrieben ist oder nicht.
Jedoch ist Kritik an den Kontrollfunktionen der UDK aufgekommen. Ein im Juli 2019 veröffentlichter Bericht stellte die Verhältnismässigkeit der enthaltenen Überwachung infrage, die sich in einer gesonderten Prüfung von 2,7 Millionen nicht eindeutigen Empfängerinnen und Empfängern von Zahlungen manifestierte. Diese Prüfung resultierte jedoch darin, dass im Jahr 2017 nur 705 und im Jahr 2018 weitere 1094 Fälle von Sozialleistungen aufgrund von Fehlern gestoppt, angepasst oder zurückgezahlt wurden. Im Jahr 2017 übergab die UDK 25 Fälle zur weiteren Untersuchung an die Polizei.
Diese Kritik an der Überwachung grosser Teile der dänischen Bevölkerung ohne tatsächlichen Verdacht ist nicht neu. Das grenzenlose Sammeln persönlicher Daten wird in den Kommentaren zu einem Fall deutlich, der die persönlichen Daten der gesetzlichen Vertretung einer Begünstigten betraf. Die Daten des Vertreters sowie dessen naher Familie wurden offenbar gesammelt. UDK gab an, ihre gesetzlichen Verpflichtungen nicht erfüllen zu können, ohne diese Daten zu sammeln. Neu ist die Beschreibung, wie die Analyse der Kontrollen durchgeführt wird. Die Datenanalyse basiert auf einer Reihe von Kriterien, die ständig weiterentwickelt werden, darunter auch «modulare Kriterien», bei denen eine Datenanalyse zu verschiedenen Fragen durchgeführt wird und alle resultierenden Teilergebnisse dann in die Gesamtschlussfolgerung einbezogen werden.

3. Effizienzsteigerung von Genehmigungsprozessen bei der finnischen Einwanderungsbehörde

Migri (die finnische Einwanderungsbehörde) ist eine Regierungsorganisation, die für Entscheidungen im Zusammenhang mit Einwanderung, finnischer Staatsbürgerschaft sowie dem Asyl- und Flüchtlingsstatus zuständig ist. Im Jahr 2019 hat Migri insgesamt 104 000 Entscheidungen in diesen Angelegenheiten getroffen.
Die oft lange Bearbeitungsdauer von Genehmigungen führt zu Ängsten und Unsicherheit bei den Menschen, die von diesen Entscheidungen betroffen sind. Dies gilt insbesondere für Asylsuchende. Andere Themen, die immer wieder in der öffentlichen Diskussion angesprochen werden, sind die Aufenthaltsgenehmigungen von Mitarbeitenden, die von finnischen Unternehmen aus dem Ausland eingestellt werden, sowie internationale Studierende, die den Beginn ihres Studiums aufgrund von Verzögerungen bei der Genehmigungsbearbeitung verpassen. Die Agentur prüft derzeit Automatisierungsmöglichkeiten, um die Genehmigungsprozesse zu beschleunigen und ihre Abläufe kostengünstiger zu gestalten. Nach Angaben eines Sprechers des beauftragten IT-Beratungsunternehmens zielt das laufende Projekt «Smart Digital Agency» darauf ab, Migris Dienstleistungen zu verbessern, interne Prozesse zu beschleunigen und zu lernen, Inhalte genauer zu entfernen.
Da der Agentur nach eigenen Angaben die Ressourcen fehlen, um eine steigende Anzahl von Entscheidungsprozessen abzuwickeln, soll die menschliche Beteiligung in verschiedenen Stadien des Genehmigungsprozesses reduziert werden. Vor diesem Hintergrund erscheint die Automatisierung als ein attraktiver Weg, die verfügbaren menschlichen Ressourcen auf die Teile des Genehmigungsprozesses zu konzentrieren, in denen sie zwingend notwendig sind.
Das Ziel von Migri ist, die Notwendigkeit des persönlichen Kontakts mit Beamtinnen und Beamten zu reduzieren, sodass bis Ende 2020 etwa 90 Prozent aller Interaktionen über Selbstbedienungskanäle stattfinden. Ein praktisches Beispiel dafür ist der Einsatz von Chatbots, um grundlegende Serviceanfragen automatisch abzuwickeln. Laut Migri können Chatbots die überforderten Mitarbeitenden im Callcenter entlasten. Vor der Einführung der Chatbots im Mai 2018 kam die Mehrheit der Anfragen gar nicht durch: Nur etwa 20 Prozent der eingehenden Anrufe wurden beantwortet. Nach der Chatbot-Automatisierung ist der Anteil beantworteter Anrufe auf rund 80 Prozent gestiegen. Dies war vor allem möglich, da viele der Anrufe Routineprobleme betrafen. Die Chatbot-Automatisierung leitet die Anfragen auch an andere Behörden weiter. Bei der Gründung eines finnischen Unternehmens etwa sind neben Migri auch andere Behörden wie die Steuerverwaltung oder das Patent- und Registrierungsamt involviert.
In ihren öffentlichen Äusserungen hat Migri auf die notwendigen organisatorischen Veränderungen für die Automatisierung einschliesslich der Neuordnung der behördeneigenen Abläufe hin-

gewiesen. Zu den noch dringenderen Änderungen gehört, die Datenschutzgesetze der Einwanderungsverwaltung anzupassen. Ein Vorschlag im finnischen Parlament zielt darauf ab, die Regelungen zur Verarbeitung personenbezogener Daten durch Migri zu erneuern, um den Anforderungen im Zusammenhang mit der Digitalisierung gerecht zu werden und automatisierte Entscheidungen (ADM) zu ermöglichen, wenn bestimmte Voraussetzungen erfüllt sind. Der Vorschlag wurde bereits vom Verfassungsrechtsausschuss des finnischen Parlaments geprüft, der sich in seiner Stellungnahme kritisch zu dem Vorschlag äusserte. Nach Ansicht des Ausschusses sind vor allem Defizite bei guter Regierungsführung (good governance) und der Verantwortung der Beamtinnen und Beamten für die Rechtmässigkeit ihres Handelns zu bemängeln. Darüber hinaus würde der Ausschuss eine allgemeine gesetzliche Regelung beim Thema bevorzugen, die bei Bedarf durch sektorspezifische statt einer behördenspezifischen Gesetzgebung zur Entscheidungsautomatisierung Regelungen ergänzt werden sollte. Migris Pläne für den Einsatz von ADM stehen somit im Kontext von internem und externem Druck auf die Genehmigungsprozesse der Behörde, der durch eine erhöhte Anzahl von Anträgen, die bearbeitet werden müssen, den Wunsch, Ressourcen auf Themen zu konzentrieren, die menschliches Urteilsvermögen erfordern, Forderungen nach einer schnellen Antragsbearbeitung und eine Gesetzgebung, die derzeit den Einsatz von ADM einschränkt, charakterisiert ist. Gleichzeitig scheinen Vorschläge zur Änderung der Gesetzgebung auf verfassungsrechtliche Konflikte zu stossen, welche die Automatisierungspläne der Behörde aufhalten oder zumindest verzögern.

4. Automatisierte Prozesse bei der finnischen Sozialversicherungsanstalt

Die finnische Sozialversicherungsanstalt, bekannt als Kela, ist für die Abrechnung von jährlich etwa 15,5 Milliarden Euro an Leistungen im Rahmen der nationalen Sozialversicherungsprogramme verantwortlich. Kela sieht Künstliche Intelligenz, maschinelles Lernen und Softwarerobotik als integrale Bestandteile seiner zukünftigen EDV-Systeme an. Zu den laufenden Entwicklungen von Kela im Bereich der Automatisierung gehören Chatbots, um mit Bürgerinnen und Bürgern zu kommunizieren, Missverständnisse zu erkennen und Betrug zu verhindern, sowie die Datenanalyse. Die Gesetzgebung im Feld der Sozialleistungen ist komplex und umfasst Hunderte von Einzelvorschriften, die in den vergangenen 30 Jahren erlassen wurden.
Ein besonderes Problem, das im Automating Society Report von 2019 identifiziert wurde, war die Notwendigkeit der Feststellung, wie Leistungsentscheidungen an die Bürgerinnen und Bürger kommuniziert werden, damit die Ergebnisse und Schlussfolgerungen der Automatisierung in eine verständliche Entscheidung übersetzt werden können. Die Frage der Kommunikation von automatischen Entscheidungen hat inzwischen eine Untersuchung von Kelas Automatisierungsprozessen ausgelöst. Bereits 2018 erhielt der finnische Justizkanzler – ein Beamter, der die Einhaltung der Gesetze durch die Behörden überwacht und den Rechtsschutz der Bürgerinnen und Bürger sicherstellt – eine Beschwerde über die Kommunikation von Kelas Entscheidungen zum Arbeitslosengeld. Bei der Untersuchung der Beschwerde wurde der Kanzler auf die Tatsache aufmerksam, dass Kela automatische Entscheidungsverfahren zur Abrechnung von Sozialleistungen eingeführt hatte. Es gab Zehntausende automatisierte Entscheidungen und keine einzige Ansprechperson, die zusätzliche Informationen zu diesen geben konnte. In der Folge wurde im Oktober 2019 eine weitere Untersuchung des Einsatzes von Automatisierung in Kelas Verwaltungsprozessen eingeleitet.
Das Informationsersuchen des Justizkanzlers an Kela gibt einen Überblick über Compliance- und Rechtsschutzfragen, die für ADM-Projekte im öffentlichen Sektor relevant sind. Das Ersuchen konzentrierte sich auf die Anforderungen der guten Regierungsführung, die rechtliche Verantwortlichkeit von Beamtinnen und Beamten für ihre Handlungen und die Auswirkungen dieser Anforderungen und Verantwortlichkeiten auf automatisierte Entscheidungsprozesse. Der Justizkanzler stellte fest, dass Kela zwar öffentlich erklärt, das Recht zu haben, Entscheidungen automatisch zu treffen, aber keine Informationen darüber bereitstellt, welche Entscheidungen es automatisiert trifft, ob sich diese auf eine teilweise oder vollständige Automatisierung beziehen oder als Entscheidungsfindungen zu betrachten sind, die durch Automatisierung unterstützt werden. Aufgrund dieser Überlegungen bat der Justizkanzler um detailliertere Informationen über Kelas ADM-Prozesse einschliesslich der Frage, welche Leistungsentscheidungen die Behörde automatisiert und auf welcher Rechtsgrundlage dies erfolgt. Kela wurde auch um Einschätzungen gebeten, wie eine gute Verwaltung und der Rechtsschutz der betroffenen Bürgerinnen und Bürger sichergestellt werden können. Die Behörde wurde angehalten, Angaben darüber zu machen, wie sie die Entscheidungsautomatisierung ausbauen möchte, z. B., welche Arten von ADM in Zukunft eingesetzt werden könnten. Darüber hinaus wurde Kela aufgefordert, zu ADM Stellung zu nehmen, wenn es um die Rechenschaftspflicht der Beamtinnen und Beamten geht, und insbesondere darzulegen, wie die Verantwortlichkeiten zwischen der Behördenleitung, den ADM-Systementwicklerinnen und -Systementwicklern und den einzelnen Beamtinnen und Beamten verteilt sind. Schliesslich sollte Kela Details darüber zur Verfügung stellen, wie Auskunftsrechte von Bürgerinnen und Bürgern zur Funktionsweise der Entscheidungsalgorithmen sowie der darunterliegenden Logik realisiert werden sollen.
Das Ersuchen des Justizkanzlers an Kela weist auch auf eine mögliche Unterscheidung zwischen teilweiser und vollständiger Automatisierung von Entscheidungen hin. Dies legt nahe, dass die Beantwortung der Frage, ob Beamtinnen und Beamte die Automatisierung als Werkzeug oder Hilfsmittel für die Entscheidungsfindung einsetzen oder Entscheidungen vollständig automatisch getroffen werden, Auswirkungen darauf haben kann, wie die Ausübung einer guten Regierungsführung und die Verantwortlichkeit von Beamtinnen und Beamten beurteilt werden. In einem breiteren Kontext wird ADM manchmal als eine inkrementelle Entwicklung dargestellt – eine Fortsetzung älterer, technologisch einfacherer Mittel der automatischen Informationsverarbeitung. Im Fall von Kela wurden in den vergangenen Jahrzehnten eher traditionelle Mittel wie Batch-Prozesse und ein einfacher Softwarecode eingesetzt, um Teile von behördeninternen Prozessen zu automatisieren. Die Unterscheidung zwischen vollständiger und teilweiser Automatisierung deutet jedoch darauf hin, dass ab einem bestimmten Punkt die Erhöhung des Automatisierungsgrads von Entscheidungen nicht einfach eine Verschiebung von weniger zu mehr Automatisierung darstellt, sondern eine qualitative Verschiebung zu einer anderen Art von Prozess mit sich bringt, die neue Auswirkungen auf die Einhaltung der gesetzlichen Rahmenbedingungen hat.

5. Trelleborg in Schweden

Seit 2017 setzt die schwedische Kommune Trelleborg mit rund 46 000 Einwohnerinnen und Einwohnern automatische Entscheidungssysteme ein, um Anträge für Sozialleistungen zu bearbeiten. Das in Trelleborg genutzte System wurde nicht zuletzt wegen der damit verbundenen Rechtsstreitigkeiten zu einem prominenten Beispiel für die Automatisierung von Verwaltungsprozessen in Schweden.
Das System basiert auf einem recht einfachen Entscheidungsmodell, das bestimmte Variablen mit staatlichen Datenbanken, z. B. der Steuerbehörden oder der staatlichen Krankenversicherung, abgleicht. Alle Erstanträge werden manuell durch Beamtinnen und Beamten beschieden. Folgeanträge, die ein-

mal im Monat gestellt werden müssen, werden jedoch automatisch verarbeitet. Betroffen von diesem automatisierten Entscheidungssystem sind Einwohnerinnen und Einwohner der Gemeinde, die finanzielle Unterstützung (ekonomisk bistånd) einschliesslich Sozialleistungen (försörjningsstöd) beantragen, die entweder ganz oder teilweise die Kosten für Wohnung, Lebensmittel, Kleidung, Telefon und Internetzugang abdecken.
Die Kritik an dem System reichte von dem Vorwurf, Entscheidungen würden algorithmischen Systemen überlassen, die dazu nicht gesetzlich legitimiert sind, bis hin zu Fragen der Transparenz sowie der Zukunft der Arbeit und des Status von Beamtinnen und Beamten im Allgemeinen. Im Rahmen einer breiten Diskussion darüber, wie sich das Sozialsystem durch die Digitalisierung und insbesondere durch ADM verändert, begannen mehrere Akteure, das sogenannte Trelleborg-Modell weiter zu untersuchen. Zunächst schaltete sich der Journalist Fredrik Ramel in die Diskussion über die Transparenz des in Trelleborg verwendeten ADM-Systems ein. Nach mehreren gescheiterten Versuchen, von der Stadtverwaltung Trelleborg Zugang zum Quellcode zu erhalten und die dänische Firma, die das System hergestellt hat, zu erreichen, unternahm Ramel rechtliche Schritte. Im Gegensatz zu Simon Vinge, der zuvor die Gemeinde Trelleborg beim Parlamentarischen Ombudsmann (Justitieombudsmannen) angezeigt und von diesem keine Antwort erhalten hatte, reichte er eine Berufung beim Verwaltungsgericht ein und argumentierte, dass der Quellcode der verwendeten Software unter das schwedische Offenlegungsprinzip (Offentlighetsprincipen) fällt. Dies sieht vor, dass jeder Bürgerin und jedem Bürger der Zugang zu offiziellen Dokumenten ermöglicht werden muss. Das Gericht folgte seiner Berufung und entschied, der Quellcode müsse der Öffentlichkeit zugänglich gemacht werden und falle vollständig unter das Prinzip der Offenlegung.

6. System für Risikobewertung (SyRi) in den Niederlanden

Das Systeem Risico Inventarisatie (System für Risikobewertung, SyRI) ist ein datenbasiertes Analyseprogramm, das vom Ministerium für Arbeit und Soziales der Niederlande eingesetzt wurde. Das System sollte dazu dienen, zu Unrecht oder missbräuchlich bezogene Sozialleistungen aufzudecken. Seine Einführung geht auf Anfragen von verschiedenen «Kooperationspartnern» des Ministeriums, z. B. der Steuer- (Belastingdienst) und Einwanderungsbehörden (IND) oder der Sozialversicherungsanstalt (SVB), zurück. SyRI kombinierte Daten, die von Bürgerinnen und Bürger etwa in Steuererklärungen angegeben werden, mit solchen aus einer Reihe von anderen Quellen. Ein Algorithmus berechnete dann auf Basis eines Risikomodells, dessen Kriterien nicht transparent sind, ob bei Bürgerinnen und Bürgern ein erhöhtes Risiko vermutet wird, dass sie Sozialleistungsbetrug begehen.
Eine Gruppe von Bürgerrechtsinitiativen mit dem Namen Bij Voorbaat Verdacht (Im Voraus verdächtigt) wurde aktiv und reichte im Jahr 2018 eine Klage gegen die Verwendung des Systems ein. Der Fall erhielt gegen Ende 2019 internationale Aufmerksamkeit, als Philip Alston, der UN-Sonderberichterstatter für extreme Armut und Menschenrechte, das Verfahren als die erste ihm bekannte juristische Massnahme bezeichnete, die den Einsatz eines ADM-Systems in einem Wohlfahrtsstaat aus menschenrechtlichen Gründen «grundlegend und umfassend» infrage stellt. Im vergangenen Jahr stellte schliesslich der Gerichtshof in Den Haag fest, dass SyRI gegen Art. 8 der Europäischen Menschenrechtskonvention verstösst, und untersagte den weiteren Einsatz. Dabei betonten die Richterinnen und Richter in ihrem Urteil, dass Regierungen eine «besondere Verantwortung» im Hinblick auf den Schutz der Menschenrechte haben, wenn sie neue Technologien wie automatisierte Profiling-Systeme einsetzen.

7. Erkennung von Bankkonten, die für illegale Aktivitäten genutzt werden, in Polen

STIR (System Teleinformatyczny Izby Rozliczeniowej, EDV-System der Clearingstelle) ist ein Werkzeug, das Informationen von Banken sowie Spar- und Kreditgenossenschaften sammelt. Das Ziel dieses Systems ist, finanzielle Aktivitäten zu untersuchen und mögliche illegale Aktivitäten aufzudecken. Die Software wird von der Krajowa Izba Rozliczeniowa, der staatlichen Rechnungskammer in Polen, betrieben. Sie ist eine Schlüsselinstanz der polnischen Infrastruktur für Zahlungsverkehr, die komplexe Clearingdienste anbietet und Lösungen für den Banken- und Zahlungsverkehrssektor erstellt. Das System für automatische Entscheidungen in diesem System liefert den Steuerbehörden Verdachtsfälle für möglichen Betrug, indem es Bankkonten Risikoindikatoren zuordnet. Wenn z. B. der Verdacht besteht, dass im Zusammenhang mit einem bestimmten Konto eine Straftat begangen wurde, kann die Bank (auf Anfrage der Steuerbehörden) das Konto für 72 Stunden sperren (was verlängert werden kann).
Die gesetzlichen Regelungen zu STIR sind seit Januar 2018 in Kraft, als die Steuerverordnung von 1997 geändert wurde.
Der von STIR verwendete Algorithmus ist weder öffentlich zugänglich noch transparent. Ausserdem besagt das Gesetz zur Einführung von STIR, dass die Offenlegung oder Verwendung von Algorithmen oder Risikoindikatoren ohne Berechtigung eine Straftat darstellt. Eine Person, die Algorithmen offenlegt, kann mit einer Freiheitsstrafe von bis zu fünf Jahren bestraft werden (wenn die Tat unbeabsichtigt war, kann eine Geldstrafe verhängt werden).
Die Risikoindikatoren werden durch die von der Clearingstelle entwickelten Algorithmen bestimmt. Sie berücksichtigen die übliche Praxis des Bankensektors im Bereich der Bekämpfung von Finanz- und Steuerdelikten und sind ein Schlüsselfaktor bei der Entscheidung, ob ein Konto gesperrt wird oder nicht.
Diese Kriterien können (gemäss der Steuerverordnung) folgende sein:

1. Ökonomie: basierend auf einer Bewertung der gesamten wirtschaftlichen Aktivitäten eines Unternehmens, insbesondere unter Berücksichtigung von Transaktionen, die nicht durch die Art des Geschäfts gerechtfertigt sind,
2. Geografie: bestehend aus Transaktionen mit Subjekten aus Ländern, in denen ein hohes Risiko von Steuerbetrug besteht,
3. Subjektspezifität bestimmter Unternehmen: Durchführung von Geschäftsaktivitäten mit hohem Risiko, wenn es um die Möglichkeit der Steuererpressung geht,
4. Verhalten: jedes ungewöhnliche Verhalten der Entität
5. Verbindungen: das Vorhandensein von Verbindungen zu Unternehmen, bei denen ein Risiko zum Steuerbetrug besteht.

Die Verrechnungskammer kann für die Wartung oder Änderung des STIR-Systems externe Dienstleister beauftragen. Ein solcher Auftrag erfolgt auf der Grundlage eines zivilrechtlichen Vertrages, aber die öffentliche Einsicht in den Prozess, in dem ein Unternehmen ausgewählt wird, ist nicht ausreichend. Dies liegt an der eingeschränkten Anwendbarkeit der Regeln für öffentliche Vergabeprozesse, denn die staatliche Clearingstelle ist eine Aktiengesellschaft. Dies könnte zu einer unzureichenden Transparenz und Kontrolle des ADM-Systems führen. In der Antwort auf eine Informationsfreiheitsanfrage der ePaństwo-Stiftung nach mehr Informationen über dieses Unternehmen und den Dienstleistungsvertrag weigerte sich das Finanzministerium, die Informationen offenzulegen. Als Begründung führt es an, dass die staatliche Verrechnungskammer eine private Einrichtung ist und ihre Tätigkeiten nicht gemäss den Informationsfreiheitsgesetzen offengelegt werden können.

Im Dezember 2018 veröffentlichte das Verwaltungsgericht in Warschau ein Präzedenzurteil im Fall STIR. Das Gericht stellte fest, dass im Falle einer Verlängerung der Kontosperre um drei Monate kein Beweisverfahren durchgeführt werden muss. In einem solchen Fall reicht es aus, wenn Amtsträgerinnen und Amtsträger die Geldströme analysieren und davon ausgehen, dass bei dem Bankkonto die Gefahr für Steuerbetrug besteht.

8. Das Arbeitsmarkt-Chancen-Assistenzsystem (AMAS) in Österreich

Das Arbeitsmarkt-Chancen-Assistenzsystem (AMAS) soll die Arbeit des österreichischen Arbeitsmarktservice (AMS)[168] verbessern, indem es auf Basis einer statistischen Analyse historischer Daten die zukünftigen Chancen von Arbeitsuchenden am Arbeitsmarkt – die sogenannte Integrationschance (IC) – berechnet und sie anhand der Ergebnisse in drei verschiedene Gruppen einteilt. Je Gruppe werden dann verschiedene Ressourcen zur Weiterbildung bereitgestellt. Merkmale, die in die Berechnung der Integrationschance Arbeitsuchender einbezogen werden, umfassen neben Alter, Herkunft, Ausbildung und früheren Beschäftigungen auch gesundheitliche Beeinträchtigungen, frühere Kontakte mit dem AMS, das Arbeitsmarktgeschehen am Wohnort der Arbeitsuchenden sowie möglicherweise vorliegende Betreuungspflichten.

Im Oktober 2019 berichtete AlgorithmWatch[169] über Kritik am AMAS. Es zeige sich, dass eines der verwendeten Regressionsmodelle insbesondere Frauen und Menschen mit Behinderungen diskriminiere, indem diese Attribute als negativer Faktor gewertet würden – d. h., eine Frau zu sein, wirke sich negativ auf den IC-Wert aus, den das System berechnet.

Johannes Kopf, Vorstand des AMS, bestritt in einer schriftlichen Reaktion[170] auf die vielerorts vorgebrachte Kritik zwar nicht, dass das System verzerrte Modelle enthält, betonte jedoch, dass diese nicht zu diskriminierenden Entscheidungen führten. Vielmehr würden insbesondere Frauen vom ADM-System profitieren, denn sie seien in mittleren Gruppen über- und in unteren Gruppen sogar unterrepräsentiert, sodass sie häufiger in den Genuss von Weiterbildungsmassnahmen kämen als ohne den Einsatz des Algorithmus.

Im August 2020 untersagte die österreichische Datenschutzbehörde den Einsatz von AMAS,[171] weil das System in mehreren Punkten gegen die EU-Datenschutzgrundverordnung verstosse. Die Behörde stellt hierbei insbesondere den Einsatz von Profiling fest, der nach der aktuellen Gesetzeslage nicht gestattet ist. Im Dezember 2020 wurde bekannt, dass das Österreichische Bundesverwaltungsgericht einer Beschwerde des AMS stattgegeben und den Bescheid der Datenschutzbehörde aufgehoben hatte. Medienberichten zufolge[172] hat das Gericht der Einschätzung der Behörde klar widersprochen und festgestellt, dass das Gesetz dem AMS erlaube, personenbezogene Daten von Arbeitsuchenden zu verwenden, um etwa zu beurteilen, welchen Förderbedarf bestimmte Jobsuchende haben. Untersagt sei nur, dass «wichtige Entscheidungen einer Behörde oder eines öffentlichen Dienstleisters auf einer Entscheidung beruhen, die ausschliesslich von einem Computerprogramm getroffen wird». Das sei aber nicht der Fall, denn der AMS habe in seinen internen Leitlinien festgelegt, dass der Algorithmus nur eine zusätzliche Information für Beraterinnen und Berater bieten soll. Demnach könnten sie weiter selbst entscheiden, was sie mit der Information tun, beispielsweise welche Fördermassnahmen sie den Jobsuchende vorschlagen. Die Datenschutzbehörde hatte angeführt, diese Leitlinien reichten nicht aus, da den Beraterinnen und Beratern nur wenig Zeit bei Beratungsgesprächen zur Verfügung stehe und somit zu befürchten sei, dass sie die Prognosen des Algorithmus übernehmen. Das Bundesverwaltungsgericht wendete ein, dass dies nicht einfach behauptet werden könne, sondern nachgewiesen werden müsse. Die Datenschutzbehörde kann gegen die Entscheidung Rechtsmittel einlegen. AMS-Chef Kopf verkündete daraufhin, nun zum einen abzuwarten, ob die Datenschutzbehörde gegen das Urteil vorgehen wird. Zum anderen müsse aber eingeschätzt werden, ob das Verfahren in der aktuellen Version noch eingesetzt werden kann oder überarbeitet werden muss, denn es sei fraglich, wie die Daten der coronabedingten Arbeitsmarktkrise in den Computeralgorithmus integriert werden könnten, da aufgrund der Coronapandemie im Jahr 2020 sehr viele Menschen unverschuldet arbeitslos wurden.[173] Eine Studie[174] der Österreichischen Akademie der Wissenschaften (ÖAW) in Zusammenarbeit mit Forscherinnen und Forschern der TU Wien und der Universität Michigan aus dem Jahr 2020 kam zu dem Ergebnis, dass das System keine Mechanismen enthält, um Bias vorzubeugen. Nach den Ergebnissen der Studie bewirkt die grobe Einteilung in drei Gruppen eher einen effektiveren Einsatz der Mittel als eine bessere und zielgenauere Zuweisung von Weiterbildungsmassnahmen. Die Autorinnen und Autoren merken an, dass umfassende Einsichts- und Einspruchsrechte für Betroffene, öffentliche Konsultationen sowie die Vermittlung neuer Kompetenzen für AMS-Beraterinnen und -Berater, aber auch für Arbeitsuchende geeignete Mittel sein könnten, Diskriminierung durch algorithmische Systeme zu verhindern und zu einem offenen Umgang mit KI-Systemen beizutragen.

168 Das österreichische Äquivalent der Regionalen Arbeitsvermittlungszentren.
169 Kayser-Bril, 2019.
170 Stellungnahme von Johannes Kopf zur Kritik am AMS-Algorithmus in Der Standard vom 25.09.2019, https://www.derstandard.at/story/2000109032448/ein-kritischer-blick-auf-die-ams-kritiker.
171 Zavadil, 2020.
172 Der Standard vom 21.12.2020, Gericht macht Weg für umstrittenen AMS-Algorithmus frei, https://www.derstandard.at/story/2000122684131/gericht-macht-weg-fuer-umstrittenen-ams-algorithmus-frei?ref=rss.
173 WienerZeitung vom 21.12.2020, AMS-Algorithmus vor Neustart: BVwG kippte Datenschutzbehörde-Bescheid, https://www.wienerzeitung.at/nachrichten/digital/digital-news/2086179-AMS-Algorithmus-vor-Neustart-BVwG-kippte-Datenschutzbehoerde-Bescheid.html.
174 Allhutter/Mager/Cech/Fischer/Grill, 2020.

Kapitel 3

Rechtliche Rahmenbedingungen für den staatlichen KI-Einsatz im Kanton Zürich

Nadja Braun Binder
Catherine Egli
Laurent Freiburghaus
Eliane Kunz
Nina Laukenmann
Liliane Obrecht

Das dritte Kapitel dient der rechtlichen Auseinandersetzung mit KI-Anwendungen in der öffentlichen Verwaltung. Dabei ist zu betonen, dass die rechtlichen Herausforderungen sich einerseits aus dem Einsatz von KI selbst, anderseits aber auch aus dem mithilfe von KI hergestellten Automatisierungsgrad ergeben können. Im Folgenden wird deshalb bei Bedarf zwischen teilautomatisierten Verfahren, in denen KI-Systeme unterstützend eingesetzt werden, und vollautomatisierten Verfahren, in denen KI-Systeme genutzt werden und keine menschliche Intervention mehr stattfindet, unterschieden. Das hier zugrunde liegende Verständnis von vollautomatisierten Verfahren deckt sich mit dem Ausdruck «automatisierte Einzelentscheide» nach dem revidierten Datenschutzgesetz.[175]

In einem ersten Schritt (A.) wird zunächst ein Überblick über die wichtigsten in der Literatur besprochenen allgemeinen rechtlichen Herausforderungen des staatlichen KI-Einsatzes gegeben. Für diese Herausforderungen werden sodann spezifisch mit Blick auf die Rechtslage im Kanton Zürich konkrete Schlussfolgerungen gezogen. In einem zweiten Schritt (B.) werden drei konkrete Einsatzbereiche bzw. Einsatzfelder und deren rechtliche Rahmenbedingungen im Kanton Zürich näher beleuchtet, in denen KI inskünftig eine Rolle spielen könnte: Steuerverfahren, Sozialversicherungsverfahren und Chatbots.[176] Die zentralen Erkenntnisse aus diesem Kapitel werden abschliessend zusammengefasst (C.).

A. Zentrale Herausforderungen und Schlussfolgerungen

Im Unterschied zu privaten Akteuren, die KI einsetzen, gelten für den Staat besondere Voraussetzungen und Rahmenbedingungen. So handelt der Staat in der Regel einseitig, d. h. aus einer übergeordneten Stellung heraus, und ist meist nicht auf das einzelfallbezogene Einverständnis der Adressaten seines Handelns angewiesen. Unter bestimmten Voraussetzungen kommt dem Staat eine Anordnungs- und eine Zwangsbefugnis gegenüber den Privaten zu. Im Gegenzug sind der Staat bzw. Private, die staatliche Aufgaben wahrnehmen, an die Grundrechte und an das Legalitätsprinzip gebunden. Diese schützen Private vor Übergriffen durch den Staat. Nichts anderes gilt im Bereich des staatlichen Einsatzes von KI. Setzen der Staat oder Private, die staatliche Aufgaben wahrnehmen, KI ein, müssen sie die Garantien der Grundrechte und das Legalitätsprinzip beachten.

Im Folgenden werden daher zentrale Herausforderungen, die sich aus dem Legalitätsprinzip (I.) und aus den Grundrechten mit Blick auf einen staatlichen KI-Einsatz ergeben, erörtert. Aus der grundrechtlichen Perspektive werden dabei insbesondere die rechtsstaatlichen Verfahrensgarantien (II.), das Diskriminierungsverbot (III.) und das Recht auf informationelle Selbstbestimmung – unter Berücksichtigung der gesetzlich konkretisierten datenschutzrechtlichen Anforderungen – (IV.) beleuchtet. Weitere Herausforderungen ergeben sich aus offenen Normen, d. h. Rechtssätzen, die Ermessen einräumen oder unbestimmte Rechtsbegriffe enthalten (V.). Als zentrale Voraussetzung des schweizerischen Verwaltungs- und Verwaltungsverfahrensrechts ist ferner zu klären, ob mithilfe von KI erlassene Entscheidungen als Verfügungen gelten (VI.). Abschliessend wird skizziert, welche Transparenzanforderungen für die öffentliche Verwaltung bzw. für den KI-Einsatz existieren (VII.).

I. Legalitätsprinzip

1. Grundlagen

Das Legalitätsprinzip bedeutet, dass das Recht die Grundlage und die Schranke staatlichen Handelns bildet. Daraus folgt einerseits, dass staatliches Handeln, das sich nicht auf eine gesetzliche Grundlage stützen lässt, auch dann unzulässig ist, wenn es nicht im Widerspruch zum geltenden Recht steht. Anderseits sind alle staatlichen Akteure bei ihren Handlungen an das geltende Recht gebunden.

Als allgemeiner Grundsatz rechtsstaatlichen Handelns ist das Legalitätsprinzip auf Bundesebene in Art. 5 Abs. 1 BV[177] verankert. Zum Ausdruck kommt das Legalitätsprinzip auch in Art. 36 BV, wonach für die Einschränkung von Grundrechten eine gesetzliche Grundlage vorausgesetzt wird, sowie in Art. 127 BV, wonach die Ausgestaltung der Steuern in den Grundzügen im Gesetz selbst zu regeln ist. Die meisten neueren Kantonsverfassungen halten das Legalitätsprinzip ebenfalls ausdrücklich

175 Art. 21 des totalrevidierten Datenschutzgesetzes (revDSG), BBl 2020 7639, abrufbar unter https://www.admin.ch/opc/de/federal-gazette/2020/7639.pdf. Vgl. auch Kapitel 3 A. IV. 2. a).
176 Zu den potenziellen Einsatzgebieten von KI in der öffentlichen Verwaltung vgl. die Ausführungen im Kapitel 2 F.
177 Bundesverfassung der Schweizerischen Eidgenossenschaft vom 18. April 1999, SR 101.

fest, so z. B. § 5 Abs. 1 KV BS,[178] Art. 66 Abs. 2 KV BE[179] oder Art. 2 Abs. 1 KV ZH.[180]
Das Legalitätsprinzip erfüllt sowohl rechtsstaatliche als auch demokratische Funktionen.[181] Aus Sicht der Rechtsstaatlichkeit soll das Legalitätsprinzip erstens Rechtssicherheit gewährleisten, zweitens Rechtsgleichheit sicherstellen und drittens die Individuen vor ungerechtfertigten staatlichen Eingriffen schützen. Darüber hinaus garantiert das Legalitätsprinzip durch die Bindung des staatlichen Handelns an demokratisch erlassene Rechtssätze die demokratische Legitimität.[182]
Das Erfordernis des Rechtssatzes besagt, dass der Staat nur nach Massgabe von generell-abstrakten Rechtsnormen handeln darf. Rechtssätze sind Normen, die sich an eine unbestimmte Anzahl von Personen richten und eine unbestimmte Anzahl an Sachverhalten regeln. Umfasst sind Normen aller Normstufen. Keine Rechtssatzqualität kommt dagegen Verwaltungsverordnungen zu.[183]
Das Legalitätsprinzip gilt für das gesamte Staatshandeln und alle Bereiche der Verwaltungstätigkeit.[184] Allerdings ist nicht verlangt, dass staatliches Handeln bis in alle Einzelheiten vorbestimmt wird.[185] Vielmehr ist ein Ausgleich zwischen Bestimmtheit und Flexibilität zu suchen. In vielen Bereichen ist das staatliche Handeln nur durch generelle Vorgaben und Ziele gesteuert.[186] Dabei handelt es sich aber nicht um Ausnahmen vom Grundsatz der Gesetzmässigkeit, sondern um Bereiche, in denen die Anforderungen an Normdichte und Normstufe weniger hoch sind.[187]

2. Legalitätsprinzip und KI

Der Einsatz von KI durch staatliche Organe wirft mit Blick auf das Legalitätsprinzip keine grundsätzlich neuen Fragen auf. Wie bei jedem staatlichen Handeln ist auch für den Einsatz von KI-Systemen eine gesetzliche Grundlage notwendig. Die Anforderungen an die Ausgestaltung bezüglich Normdichte und Normstufe sind nach den allgemeinen Kriterien zu beurteilen.

a) Erfordernis der Normdichte

Rechtssätze müssen hinreichend bestimmt und klar formuliert sein.[188] Nach der Rechtsprechung des Bundesgerichts verlangt dies, dass «der Bürger sein Verhalten danach richten und die Folgen eines bestimmten Verhaltens mit einem den Umständen entsprechenden Grad an Gewissheit erkennen kann».[189] Allerdings ist zu berücksichtigen, dass die generell-abstrakte Struktur sowie die beschränkte Voraussehbarkeit zukünftiger Entwicklungen dem Erfordernis der Normdichte Grenzen setzen.[190]

b) Erfordernis der Normstufe

Das Erfordernis der Normstufe bedeutet, dass inhaltlich wichtige Regelungen in einem Gesetz im formellen Sinne, d. h. in einem Gesetz, welches von einem Parlament – gegebenenfalls unter Mitwirkung der Stimmberechtigten – erlassen wurde, enthalten sind. Damit wird der demokratischen Funktion des Legalitätsprinzips Rechnung getragen: Normen von einer gewissen Bedeutung müssen im demokratisch legitimierten Verfahren der Rechtsetzung durch die Legislative erlassen werden.[191]
Weitere Verfassungsbestimmungen konkretisieren die Anforderungen an die Normstufe. So verlangt Art. 36 Abs. 1 BV für schwerwiegende Eingriffe in Grundrechte eine formell-gesetzliche Grundlage. Art. 127 Abs. 1 BV hält fest, dass die Grundzüge der Besteuerung in einem formellen Gesetz zu regeln sind. Art. 164 Abs. 1 statuiert für Bundesbestimmungen schliesslich einen materiellen Gesetzesvorbehalt, wonach alle wichtigen rechtsetzenden Bestimmungen in Gesetzesform zu erlassen sind.[192] Es folgt eine (nicht abschliessende) Aufzählung von entsprechenden Inhalten. Für die Beurteilung, ob eine Bestimmung als wichtig zu qualifizieren ist, sind die folgenden Kriterien massgebend: die Intensität des Eingriffs, die Zahl der Betroffenen, die finanzielle Bedeutung sowie die Akzeptierbarkeit durch die Betroffenen.[193] Art. 164 BV gilt zwar nur für Bundeserlasse, jedoch statuieren zahlreiche Kantonsverfassungen beinahe identische Bestimmungen, darunter auch Art. 38 Abs. 1 KV ZH. Wenn der Einsatz von KI mit schweren Grundrechtseingriffen einhergeht oder eine Norm aus anderen Gründen als wichtig im Sinne von Art. 164 Abs. 1 BV zu qualifizieren ist, ist demnach eine Grundlage in einem formellen Gesetz notwendig. Wird Art. 38 Abs. 1 KV ZH zugrunde gelegt, dann sind dies z. B. Bestimmungen über die Ausübung der Volksrechte, die Organisation und Aufgaben der Behörden, die Grundzüge der Besteuerung oder auch Art und Umfang der Übertragung öffentlicher Aufgaben an Private.
Im Zusammenhang mit KI ist insbesondere die Konkretisierung des Legalitätsprinzips im Datenschutzrecht relevant, wonach für die Bearbeitung von besonders schützenswerten bzw. besonderen Personendaten durch öffentliche Organe eine formell-gesetzliche Grundlage vorgeschrieben wird (Art. 17 Abs. 2 DSG, Art. 34 Abs. 2 revDSG, § 8 Abs. 2 IDG).[194] Dies entspricht der Notwendigkeit einer formell-gesetzlichen Grundlage bei schweren Grundrechtseinschränkungen (Art. 36 Abs. 1 BV).[195]

178 Verfassung des Kantons Basel-Stadt vom 23. März 2005, SG 111.100.
179 Verfassung des Kantons Bern (KV) vom 6. Juni 1993, BSG 101.1.
180 Verfassung des Kantons Zürich vom 27. Februar 2005, LS 101.
181 Schindler, 2014, N. 8 zu Art. 5 BV.
182 Häfelin/Haller/Uhlmann, 2020, Rn. 329 ff.
183 Schindler, 2014, N. 20 ff. und 32 zu Art. 5 BV.
184 Schindler, 2014, N. 29 zu Art. 5 BV.
185 Vgl. Schindler, 2014, N. 30 f. zu Art. 5 BV.
186 Schindler, 2014, N. 31 zu Art. 5 BV.
187 Schindler, 2014, N. 31 zu Art. 5 BV.
188 Schindler, 2014, N. 33 zu Art. 5 BV.
189 BGE 138 IV 13 E. 4.1.
190 Schindler, 2014, N. 33 zu Art. 5 BV. Vgl. auch Kapitel 3 A. V.
191 Schindler, 2014, N. 36 zu Art. 5 BV.
192 Tschannen, 2014, N. 4 zu Art. 164 BV.
193 Häfelin/Haller/Uhlmann, 2020, Rn. 354.
194 Bundesgesetz über den Datenschutz (DSG) vom 19. Juni 1992, SR 235.1; totalrevidierte Fassung vom 25. September 2020 (revDSG), BBl 2020 7639, abrufbar unter https://www.admin.ch/opc/de/federal-gazette/2020/7639.pdf; Gesetz über die Information und den Datenschutz (IDG) vom 12. Februar 2007, LS 170.4.
195 Vgl. dazu die Ausführungen beim Datenschutz in Kapitel 3 A. IV.

3. Schlussfolgerungen für den Kanton Zürich

Das Legalitätsprinzip stellt für KI-Anwendungen somit keine besondere Herausforderung dar, doch ist es eine der zentralen rechtlichen Rahmenbedingungen für den staatlichen KI-Einsatz. Ohne hinreichende rechtliche Ermächtigungsgrundlage ist ein staatlicher KI-Einsatz nicht zulässig.[196] Das Erfordernis der Gesetzmässigkeit wird sich mithin durch die gesamten weiteren rechtlichen Ausführungen ziehen und liegt insbesondere den Ausführungen zu den aus den Verfahrensgarantien und den datenschutzrechtlichen Anforderungen entwickelten Empfehlungen zugrunde.[197]

II. Verfahrensgarantien und Verfahrensgrundsätze

1. Grundlagen

a) Verfahrensgarantien

Als Verfahrensgarantien werden die in der Bundesverfassung in Art. 29–32 verankerten Mindeststandards bezeichnet, die von Behörden in allen Verfahren zu beachten sind.[198] Im Zusammenhang mit dem Verwaltungshandeln sind dabei insbesondere die in Art. 29 BV geregelten sogenannten «allgemeinen Verfahrensgarantien» von Bedeutung. Zu diesen zählt der Anspruch auf rechtliches Gehör (Art. 29 Abs. 2 BV), der einerseits der Sachverhaltsabklärung dient und anderseits ein persönlichkeitsbezogenes Mitwirkungsrecht darstellt.[199] Aus dem rechtlichen Gehör leiten sich verschiedene Teilgehalte[200] wie etwa der Anspruch auf vorgängige Äusserung und Mitwirkung im Verfahren, das Recht auf Akteneinsicht[201] oder das Recht auf Begründung ab.[202]

Der *Anspruch auf vorgängige Äusserung und Mitwirkung* in Verwaltungsverfahren auf Bundesebene wird in Art. 30 Abs. 1 VwVG konkretisiert.[203] Die Parteien[204] müssen demnach vor dem Erlass einer Verfügung angehört werden bzw. Gelegenheit zur Stellungnahme erhalten.[205] Dies stellt mithin das wichtigste Mittel dar, um der Partei einen Einfluss auf die Ermittlung des rechtserheblichen Sachverhalts zu sichern.[206] Nur in den in Art. 30 Abs. 2 VwVG vorgesehenen Ausnahmesituationen müssen Parteien im Vorfeld nicht angehört werden.[207] Um eine Stellungnahme überhaupt zu ermöglichen, müssen den Betroffenen zumindest die wesentlichen Elemente des voraussichtlichen Verfügungsinhalts bekannt gegeben werden.[208] Die anschliessenden Äusserungen der Betroffenen müssen aufgrund der Korrelation mit der Untersuchungsmaxime von der Behörde auch tatsächlich zur Kenntnis genommen und geprüft werden. Selbst dann, wenn die Behörde den vorgebrachten Argumenten der Betroffenen nicht folgt, haben diese einen Anspruch darauf, dass sich Erstere mit ihren Einwänden in der Entscheidfindung und -begründung sachgerecht auseinandersetzt.[209]

Für Verfahren vor Verwaltungsbehörden im Kanton Zürich wird der Anspruch auf rechtliches Gehör im kantonalen Recht nicht separat geregelt. Der kantonale Gesetzgeber verzichtete bewusst auf eine Verankerung im VRG;[210] auch die kantonale Verfassung enthält keine entsprechende Vorgabe. Das ändert allerdings nichts an der Geltung von Art. 29 Abs. 2 BV vor kantonalen Verwaltungsbehörden.[211]

Das *Recht auf Begründung* leitet sich unmittelbar aus dem verfassungsrechtlichen Anspruch auf rechtliches Gehör ab (Art. 29 Abs. 2 BV)[212] und ist auf Bundesebene in Art. 35 Abs. 1 VwVG und für Verfahren vor Verwaltungsbehörden des Kantons Zürich in Art. 18 Abs. 2 KV ZH bzw. § 10 Abs. 1 VRG verankert.[213] Der Betroffene hat das Recht auf eine sachliche Begründung.[214] Für ihn muss ersichtlich sein, wie ein Entscheid zustande gekommen ist. Ziel sind einerseits die Sicherstellung, dass sich die Behörde mit allfällig vorgebrachten Argumenten tatsächlich auseinandergesetzt hat, und anderseits die Prüfung der Sachlichkeit der behördlichen Entscheidmotive. Ausserdem ermöglicht die Begründung, die Verfügung in voller Kenntnis der Sache anzufechten. Die Begründung einer Entscheidung ist somit besonders wichtig für das Vertrauen der Bürgerinnen und Bürger in den Staat und die Akzeptanz eines Entscheids.[215]

196 Vgl. auch das in den Interviews genannte Hindernis der mangelnden Rechtsgrundlagen, um KI in der Verwaltung einsetzen zu können, in Kapitel 2 D. II. 6.
197 Vgl. dazu Kapitel 3 A. II. und Kapitel 3 A. IV.
198 KIENER/RÜTSCHE/KUHN, 2015, Rn. 45.
199 BGE 135 I 187 E. 2.2.
200 WALDMANN, 2015, N. 44 zu Art. 29 BV; für Ausführungen zu den gehörsrechtlichen Grundgehalten z. B. THURNHERR, 2013, Rn. 402 ff.
201 Das Recht auf Akteneinsicht wird in diesem Bericht nicht weiter thematisiert, da die Gewährleistung der Akteneinsicht in digitalen Verfahren Gegenstand des Projekts IP2.1 DigiLex ist.
202 KIENER/RÜTSCHE/KUHN, 2015, Rn. 229 ff.
203 Das VwVG ist grundsätzlich auf das Verfahren vor Bundesverwaltungsbehörden anwendbar (Art. 1 Abs. 1 VwVG). Da kantonale Behörden dem VwVG nicht unterstehen (BGE 125 V 401 E. 2b; BGE 111 Ib 201 E. 3a), hat der Kanton Zürich das VRG erlassen. Gemäss § 4 VRG gilt dieses nur subsidiär; Bestimmungen in Spezialerlassen sind vorrangig.
204 Parteien sind gemäss Art. 6 VwVG Personen, deren Rechte oder Pflichten die Verfügung berühren soll, sowie andere Personen, Organisationen oder Behörden, denen ein Rechtsmittel gegen die Verfügung zusteht.
205 HÄFELIN/MÜLLER/UHLMANN, 2020, Rn. 1001 f.; KIENER/RÜTSCHE/KUHN, 2015, Rn. 230.
206 KIENER/RÜTSCHE/KUHN, 2015, Rn. 649 ff.
207 Dabei handelt es sich um die folgenden Konstellationen: Zwischenverfügungen, die nicht selbstständig durch Beschwerde anfechtbar sind (Bst. a); Verfügungen, die durch Einsprache anfechtbar sind (Bst. b); Verfügungen, in denen die Behörde den Begehren der Parteien voll entspricht (Bst. c); Vollstreckungsverfügungen (Bst. d); andere Verfügungen in einem erstinstanzlichen Verfahren, wenn Gefahr im Verzuge ist, den Parteien die Beschwerde gegen die Verfügung zusteht und einen keine andere Bestimmung des Bundesrechts einen Anspruch auf vorgängige Anhörung gewährleistet (Bst. e).
208 Vgl. HÄFELIN/MÜLLER/UHLMANN, 2020, Rn. 1010a f.
209 BGE 134 I 83 E. 4.1.; BGE 123 I 31 E. 2c.
210 Verwaltungsrechtspflegegesetz vom 24. Mai 1959 (VRG), LS 175.2. Vgl. dazu GRIFFEL, 2014, N. 2 zu § 8 VRG.
211 Vgl. BGE 121 I 230 E. 2b sowie KIENER/KÄLIN/WYTTENBACH, 2018, § 40 Rn. 8.
212 BGE 138 IV 81 E. 2.2; BGE 133 I 270 E. 3.1; BGE 121 I 54 E. 2c.
213 Auf das Verfahren letzter kantonaler Instanzen, die gestützt auf öffentliches Recht des Bundes nicht endgültig verfügen, findet Art. 35 VwVG Anwendung (Art. 1 Abs. 3 VwVG).
214 Statt vieler WALDMANN, 2015, N. 56 ff. zu Art. 29 BV.
215 Statt vieler HÄFELIN/MÜLLER/UHLMANN, 2020, Rn. 1038.

Die Ausführlichkeit der Begründung beurteilt sich nach der Komplexität der Sachverhaltslage. Bei klarer Sachlage und bestimmten Normen können Hinweise auf die Rechtsgrundlagen genügen, während bei weitem Spielraum der Behörden und einer Vielzahl von in Betracht kommenden Sachverhaltselementen eine ausführliche Begründung erforderlich ist.[216] Auf jeden Fall muss eine potenzielle Anfechtung durch die betroffene Person ermöglicht werden.[217] Die Behörde kann sich dabei auf die entscheidrelevanten Argumente beschränken.[218] Die Begründung muss folglich mindestens jene Überlegungen enthalten, von denen sich die Behörde leiten liess und auf die sie ihren Entscheid stützt, damit dessen Tragweite für die Betroffenen und die Rechtsmittelbehörde erkennbar ist.[219] Die entscheidende Behörde hat mithin den zugrunde gelegten Sachverhalt und die rechtliche Würdigung, d. h. die leitenden Überlegungen, auf denen ihr Entscheid basiert, mitzuteilen.

b) Verfahrensgrundsätze

Verfahrensgrundsätze sind in der Lehre entwickelte Leitlinien, die sich aus den gesetzlich geregelten Verfahrensordnungen ergeben.[220] Dabei handelt es sich in erster Linie um deskriptive Konstrukte, die dazu dienen, die zentralen Merkmale der Verfahrensordnungen aufzuzeigen. Verschiedene dieser Verfahrensgrundsätze sind allerdings positivrechtlich verankert und entsprechend kommt ihnen ein Rechtscharakter zu. Zu den hier interessierenden Verfahrensgrundsätzen zählt der Untersuchungsgrundsatz.[221] Dieser beherrscht das öffentliche Verfahrensrecht. Die Behörde ist demnach von Amtes wegen für die Feststellung des Sachverhalts zuständig (für den Bund vgl. Art. 12 VwVG; für den Kanton Zürich siehe § 7 Abs. 1 VRG).[222] Sie muss die rechtserheblichen Tatsachen ausfindig machen und darf nicht auf die Aussagen der Parteien abstellen,[223] denn die Behörde trägt «die Letztverantwortung für die Entscheidung».[224] Die Parteien sind im öffentlichen Verfahrensrecht allerdings verpflichtet, an der Feststellung des Sachverhalts mitzuwirken (für den Bund vgl. Art. 13 VwVG; für den Kanton Zürich siehe § 7 Abs. 2 VRG). Sie müssen in zumutbarer Weise alle rechtsrelevanten Tatsachen von sich aus offenlegen.[225] Das bedeutet etwa, dass sie in Verfahren, die sie durch ihr Begehren einleiten, den Sachverhalt darlegen müssen. Der Untersuchungsgrundsatz wird von den Mitwirkungspflichten der Parteien aber nicht verdrängt; Letztere ändern nichts daran, dass die Behörde den Beweis führt.[226]

2. Anspruch auf rechtliches Gehör und KI

a) Anspruch auf vorgängige Äusserung

Nicht jeder Einsatz von KI in Verwaltungsverfahren führt zu einer Einschränkung des Anspruchs auf vorgängige Äusserung. In bestimmten Konstellationen ist dies aber möglich. Dazu zählen Verfahren, die statt auf Angaben der Betroffenen auf bereits vorhandene Daten aufseiten der Verwaltung zugreifen. In sogenannten No-Stop-Shop-Konstellationen[227] etwa besteht die Gefahr, dass die von einer Verfügung betroffene Person bzw. die Parteien erst von der Verfügung erfahren, nachdem diese bereits erlassen worden ist.

Auch in weitgehend standardisierten und vollautomatisierten Verfahren, die zwar durch Antragstellung einer Person eingeleitet werden, sodann aber ohne jegliche menschliche Intervention ablaufen, kann der Anspruch auf vorgängige Äusserung beeinträchtigt sein.[228] Zwar wird das Verfahren in diesem Fall durch die Antragstellung und damit durch eine Äusserung eingeleitet, aber spätere Äusserungen sind aufgrund der Vollautomatisierung ausgeschlossen.[229] Wie weit der Anspruch auf vorgängige Äusserung in vollautomatisierten Verfahren eingeschränkt wird, hängt mithin davon ab, welche Angaben bei der Antragstellung möglich sind[230] und ob das System sodann in der Lage ist, die Angaben der betroffenen Person zu berücksichtigen.

In der Literatur wird deshalb dafür plädiert, dass vollautomatisierte Entscheidungen nur zulässig sein sollen, wenn die Entscheidung für die betroffene Person positiv ausfällt (d. h., den Begehren der betroffenen Person wurde entsprochen).[231] Bei Entscheidungen, die zuungunsten der betroffenen Person ausfallen, soll hingegen Vollautomation nicht zulässig sein. Vielmehr müsse das rechtliche Gehör vor einem negativen Entscheid durch die Sachbearbeiterin bzw. den Sachbearbeiter gewährt werden.[232] Eine ähnliche Überlegung scheint auf Bundesebene auch der Regelung in Art. 21 revDSG[233] zugrunde zu liegen. In dieser Bestimmung wird neu für automatisierte Einzelentscheidungen, die für die betroffene Person mit einer Rechtsfolge verbunden sind oder sie erheblich benachteiligen, eine Informationspflicht verankert (Abs. 1). Ausserdem erhält

216 Steinmann, 2014, N. 49 zu Art. 29 BV.
217 BGE 136 I 229 E. 5.2; 133 I 270 E. 3.1;129 I 232 E. 3.2; zum Ganzen statt vieler Waldmann, 2015, N. 57 zu Art. 29 BV.
218 Statt vieler Kiener/Rütsche/Kuhn, 2015, Rn. 244.
219 Waldmann, 2015, N. 57 zu Art. 29 BV.
220 Kiener/Rütsche/Kuhn, 2015, Rn. 78 ff.; Rhinow/Koller/Kiss/Thurnherr/Brühl-Moser, 2014, Rn. 974.; Wiederkehr, 2016, Rn. 183.
221 Vgl. dazu Kiener/Rütsche/Kuhn, 2015, Rn. 92 ff.
222 Vgl. dazu Auer/Binder, Kommentar VwVG, N. 7 zu Art. 12 VwVG bzw. Plüss, 2014, N. 4 zu § 7 VRG; vgl. zur gesamten Thematik statt vieler Wiederkehr, N. 183 ff.
223 Statt vieler Häfelin/Müller/Uhlmann, 2020, Rn. 988.
224 Thurnherr, 2013, Rn. 519.
225 Vgl. etwa Tschannen/Zimmerli/Müller, 2014, § 30 Rn. 24. Die Mitwirkungspflicht ist eng verbunden mit dem Mitwirkungsrecht, was dazu führt, dass die antragstellende Person selbst ein bedeutendes Interesse an der Möglichkeit der Partizipation an der Sachverhaltsabklärung hat, vgl. dazu Meyer, 2019, Rn. 116.
226 Kiener/Rütsche/Kuhn, 2015, Rn. 95.
227 Im No-Stop-Shop initiiert die Behörde antragslos die Dienstleistung und erstellt diese, ohne dass eine (weitere) Aktion der betroffenen Person notwendig wäre. Vgl. dazu etwa Scholta/Mertens/Kowalkiewicz/Becker, 2019, S. 11.
228 Vgl. dazu Braun Binder, 2020a, S. 28 f.; Braun Binder, 2020b, S. 273 f.; Rechsteiner, 2018, Rn. 18 ff.; Weber, 2019, S. 18; Weber, 2020, S. 24.
229 Rechsteiner, 2018, Rn. 19. Weber spricht gar von «technischen Unmöglichkeiten», das rechtliche Gehör im Rahmen automatisierter Verfahren je rechtzeitig gewähren zu können, dazu Weber, 2019, S. 18, und Weber, 2020, S. 24. Eine derart restriktive Einschätzung wird hier nicht geteilt.
230 Im deutschen vollautomatisierten Besteuerungsverfahren soll z. B. ein Freitextfeld im Rahmen der elektronischen Steuererklärung sicherstellen, dass die betroffenen Personen frei formulierte Angaben machen können. Vgl. dazu Braun Binder, 2019b, S. 476, m. w. H. Braun Binder, 2016b, S. 895 f.
231 Rechsteiner, 2018, Rn. 19, 23.
232 Rechsteiner, 2018, Rn. 23.
233 Vgl. auch Kapitel 3 A. IV. 2. a).

die betroffene Person die Möglichkeit, ihren Standpunkt darzulegen, wenn sie dies beantragt, und kann verlangen, dass die Entscheidung von einer natürlichen Person überprüft wird (Abs. 2).[234] Bundesorgane haben vollautomatisierte Verfügungen zu kennzeichnen (Abs. 4). Die Möglichkeit der betroffenen Person, ihren Standpunkt darzulegen, entfällt, wenn nach Art. 30 Abs. 2 VwVG ohnehin kein vorgängiges Recht auf Äusserung besteht (Art. 21 Abs. 4 Satz 2 revDSG).[235]

Während der Mechanismus «Recht auf Äusserung bei negativen Konsequenzen für die betroffene Person» aus datenschutzrechtlicher Sicht durchaus nachvollziehbar ist – das Datenschutzrecht zielt auf den Schutz der Persönlichkeit und der Grundrechte der Person ab, über die Personendaten bearbeitet werden (Art. 1 DSG bzw. Art. 1 revDSG) –, greift der Lösungsansatz im Kontext von Verwaltungsverfahren zu kurz. Der Anspruch auf vorgängige Äusserung steht nicht nur dem Verfügungsadressaten und damit der Person zu, über die personenbezogene Daten verarbeitet werden. Der Kreis der zur vorgängigen Äusserung berechtigten Parteien (vgl. den Wortlaut von Art. 30 Abs. 1 VwVG) umfasst auch Personen, Organisationen oder Behörden, die zwar nicht unmittelbare Verfügungsadressaten sind, deren Rechte aber durch die Verfügung gleichwohl betroffen sein können. Zu denken ist z. B. an Nachbarn in Baugenehmigungsverfahren.

Einleuchtender ist daher die Anknüpfung an im entsprechenden Verfahrensrecht vorgesehene Einsprachemöglichkeiten.[236] Die Einsprache stellt ein Rechtsmittel ohne Devolutiveffekt dar, d. h., die betroffene Person kann sich noch vor der verfügenden Behörde und nicht erst vor der Rechtsmittelinstanz äussern.[237] Diesfalls wird eine allfällige Einschränkung des Rechts auf vorgängige Äusserung durch die Möglichkeit der Einsprache kompensiert.

Wird KI im Rahmen von teilautomatisierten Verwaltungsverfahren zur Unterstützung der Entscheidfindung eingesetzt, kann der Anspruch auf rechtliches Gehör einfacher umgesetzt werden, da die inhaltliche Bewertung durch eine natürliche Person vorgenommen wird,[238] die den Parteien eine Möglichkeit zur Stellungnahme einräumen kann. Für die Wahrung des rechtlichen Gehörs im Kontext der Automatisierung ist somit der Zeitpunkt relevant, in dem die teilautomatisierten Aufgaben übernommen werden. Unproblematisch erscheinen unter dem Gesichtspunkt des Rechts auf vorgängige Äusserung jene Verfahren, in denen die Automationsschritte erst einsetzen, nachdem die Parteien Gelegenheit hatten, sich zu äussern.

b) Begründungspflicht

Im Zusammenhang mit KI-Systemen, insbesondere wenn es um Methoden des maschinellen Lernens geht, wird häufig das sogenannte Black-Box-Problem angesprochen.[239] Damit ist gemeint, dass die KI-basierte Datenverarbeitung für den Menschen nicht nachvollziehbar ist.[240] Die erschwerte bzw. unmögliche Nachvollziehbarkeit stellt sowohl für private als auch für staatliche KI-Anwendungen eine besondere Herausforderung dar.[241] Für den staatlichen Bereich wird davon insbesondere das Recht auf Begründung tangiert. So ist eine Entscheidung, die im Rahmen eines vollautomatisierten Verfahrens unter Nutzung von KI-Anwendungen zustande kommt, unter Umständen nicht mehr nachvollziehbar. Dies kann etwa dann der Fall sein, wenn die Entscheidung auf statistischen Auswertungen basiert und die dafür herangezogenen Kriterien nicht bekannt sind oder die Gewichtung der Kriterien unklar ist. Aus denselben Gründen kann aber auch ein teilautomatisiertes Verfahren problematisch sein, wenn die Entscheidung durch ein KI-System vorbereitet wird. Die Möglichkeit der Begründung der Entscheidung hängt dann insbesondere davon ab, ob die entscheidende Person selbst nachvollziehen kann, wie die KI-basierte Empfehlung zustande kam, und entsprechend ihre eigene Entscheidung (die in einer vollständigen oder teilweisen Abstützung auf die Empfehlung oder in einer kompletten Ablehnung der Empfehlung bestehen kann) begründen kann.

Zu bedenken ist, dass die Nachvollziehbarkeit des KI-Systems und diejenige der Entscheidung (Verfügung) nicht dasselbe bedeuten. Aus verwaltungsverfahrensrechtlicher Sicht ist erforderlich, dass die staatliche Entscheidung nachvollziehbar ist. Zu diesem Zweck ist sie zu begründen. Die Begründung muss dabei den vorstehend skizzierten Anforderungen genügen.[242] Sie muss somit aufzeigen, dass die Behörde sich mit den eventuell vorgebrachten Argumenten tatsächlich befasst hat, sowie die Überprüfung der Sachlichkeit der behördlichen Entscheidmotive und die Anfechtbarkeit der Verfügung ermöglichen. Aufgrund des aus Art. 29 Abs. 2 BV abgeleiteten Rechts auf Begründung muss demnach nicht das KI-System selbst, sondern die darauf gestützte staatliche Entscheidung nachvollziehbar sein.

Inwieweit zu diesem Zweck die eingesetzten Algorithmen selbst nachvollziehbar sein müssen, kann nicht pauschal beantwortet werden. Nach den datenschutzrechtlichen Vorgaben für automatisierte Einzelentscheide soll der betroffenen Person etwa auf deren Verlangen die Logik, auf der die Entscheidung beruht, mitgeteilt werden (Art. 25 Abs. 2 Bst. f revDSG). Gemäss Botschaft des Bundesrates ist damit gemeint, dass die «Grundannahmen der Algorithmus-Logik», auf der die automatisierte Einzelentscheidung beruht, mitgeteilt werden müssen.[243] Dagegen müssen gemäss Bundesrat «die Algorithmen (…), die Grundlage der Entscheidung sind», nicht unbedingt mitgeteilt werden.[244] Das datenschutzrechtliche Auskunftsrecht verfolgt den Zweck, der betroffenen Person diejenigen Informationen zu verschaffen, die sie benötigt, um ihre Rechte nach dem Datenschutzgesetz geltend zu machen (vgl. Art. 25 Abs. 2 revDSG). Die Zielsetzung ist mithin mit jener der verfahrensrechtlichen Begründungspflicht vergleichbar.[245] Je nach Kontext und Anwendungsbereich von KI-Anwendungen könnte zur Begründung KI-gestützter Verfügungen deshalb analog zu den datenschutzrechtlichen Vorgaben gesetzlich die Pflicht vorgesehen werden, die Logik, auf der die Entscheidung beruht, mitzuteilen.[246]

234 Nach der hier vertretenen Ansicht handelt es sich in diesen Fällen allerdings nicht mehr um eine vollautomatisierte Entscheidung.
235 Kritisch zu Art. 21 Abs. 4 revDSG Braun Binder, 2020b, S. 259.
236 Braun Binder, 2019b, S. 476; Rechsteiner, 2018, Rn. 20; Glaser, 2018, S. 188.
237 Kiener/Rütsche/Kuhn, 2015, Rn. 1968.
238 Vgl. Etscheid, 2018, S. 129.
239 Vgl. statt vieler Martini, 2019, S. 28 f.
240 Martini, 2019, S. 28 f.
241 Deshalb verwundert es nicht, dass sich aktuell ein Forschungsgebiet zu «Erklärbarer Künstlicher Intelligenz» (Explainable AI) etabliert, vgl. z. B. Samek/Montavon/Vedaldi/Hansen/Müller, 2019.
242 Vgl. Kapitel 3 A. II. 1. a).
243 BBl 2017 6941 (7067).
244 BBl 2017 6941 (7067).
245 Vgl. Braun Binder, 2020b, S. 260.
246 Art. 25 Abs. 2 revDSG gilt lediglich in vollautomatisierten Verfahren und dann, wenn Personendaten bearbeitet werden. KI-Systeme können aber auch in teilautomatisierten Verfahren eingesetzt werden und nicht personenbezogene Daten nutzen.

In Abhängigkeit von Kontext und Einsatzbereich könnten aufgrund des Rechts auf Begründung andere, unter Umständen weitergehende Angaben notwendig sein. Exemplarisch kann auf einen Entscheid des Saarländischen Verfassungsgerichtshofs verwiesen werden. Das Gericht setzte sich mit der Verfassungsbeschwerde gegen eine Busse wegen fahrlässiger Überschreitung der zulässigen innerörtlichen Höchstgeschwindigkeit auseinander. Die Verteidigung rügte, dass das Messgerät die entsprechenden Rohdaten nicht speicherte und damit dem Beschwerdeführer die Möglichkeit genommen worden sei, die hohen Anforderungen an den Vortrag eines Messfehlers zu erfüllen. Das Gericht ist diesem Einwand unter Bezugnahme auf das Recht auf ein faires Verfahren im Ergebnis gefolgt und hat betont, dass staatliches Handeln für die Betroffenen nicht undurchschaubar sein dürfe und dazu auch die grundsätzliche Nachvollziehbarkeit technischer Prozesse als Grundvoraussetzungen eines rechtsstaatlichen Verfahrens gehöre. Weiter hat es ausgeführt, dass Rechtsstaatlichkeit allgemein die Transparenz und Kontrollierbarkeit jeder staatlichen Machtausübung verlange und dies auch für den Einsatz von Algorithmen gelte.[247]

3. Untersuchungsgrundsatz und KI

Ein KI-Einsatz setzt zwingend das Vorhandensein von Daten voraus. Im Kontext von Verwaltungsverfahren handelt es sich dabei um Informationen zu einem Sachverhalt, der im Rahmen des Verfahrens zu beurteilen ist. Da KI in der öffentlichen Verwaltung vor allem im Rahmen von Automationsvorgängen eingesetzt werden wird,[248] stellt sich demnach die Frage, wie die Sachverhaltsdaten eruiert werden, die dem KI-System als Grundlage dienen. Hierbei spielen aus rechtlicher Sicht der Untersuchungsgrundsatz und die Mitwirkungspflicht der Parteien eine zentrale Rolle. Das automatisierte Verwaltungsverfahren charakterisiert sich durch das automatische Sammeln, Auswerten sowie Verifizieren von Sachverhaltsdaten.[249] Die behördlichen Mitarbeitenden treffen zumindest bei der Vollautomatisierung keinerlei zusätzliche Vorkehrungen bezüglich der Sachverhaltsabklärung,[250] wodurch einzelfallspezifische Informationen unter Umständen im (voll)automatisierten Verfahren nicht berücksichtigt werden können.[251] Werden dabei Daten genutzt, über welche die Verwaltung bereits verfügt, muss die Behörde nach dem Untersuchungsgrundsatz sicherstellen, dass aus den genutzten Datengrundlagen alle rechtserheblichen Tatsachen hervorgehen. Dies setzt voraus, dass zum einen die herangezogenen Daten selbst korrekt und vollständig sind und zum anderen all jene Daten herangezogen werden, die auch tatsächlich notwendig sind.

Sachverhaltsdaten können auch auf anderem Weg eruiert werden, etwa mittels Eingabe durch die antragstellende Person. Dies geschieht beispielsweise bei der Einreichung der Steuererklärung durch die zu veranlagende Person oder beim Antrag für eine Baubewilligung durch die Person, die bauen möchte. Die mit Blick auf eine Automation notwendigerweise standardisierte Erfassung von Sachverhaltsdaten kann allerdings dazu führen, dass die antragstellende Person nicht alle Angaben machen kann, die aus ihrer Sicht für das Verfahren notwendig wären. Wenn ein Formular nur Felder enthält, in die Zahlenwerte eingetragen werden können, ist es nicht möglich, zusätzliche Erläuterungen anzufügen, die unter Umständen notwendig sein könnten, um den Einzelfall korrekt beurteilen zu können. Mithin besteht die Gefahr, dass die antragstellende Person ihrer Mitwirkungspflicht nicht ausreichend nachkommen kann.

4. Schlussfolgerungen für den Kanton Zürich

a) Rechtliches Gehör

Im geltenden VRG ist das rechtliche Gehör nicht spezifisch als Verfahrensgarantie verankert. Lediglich das Recht auf Akteneinsicht (§ 9 VRG), das ohnehin mit Blick auf elektronische Verwaltungsverfahren angepasst werden soll,[252] und das Recht auf Begründung (§ 10 Abs. 1 VRG) werden näher geregelt.

Der Anspruch auf vorgängige Äusserung wurde im VRG dagegen bisher nicht spezifisch verankert. Mit dem Einsatz von KI im Rahmen von automatisierten Verfahren geht allerdings das Risiko einher, dass das Recht auf vorgängige Äusserung eingeschränkt wird. Nicht jeder KI-Einsatz und nicht jede Automatisierung bergen dabei dasselbe Risiko. In Abhängigkeit von den konkreten KI-Anwendungen, die der Kanton Zürich inskünftig plant, empfiehlt es sich allerdings, eine entsprechende Vorgabe, mit der das vorgängige Äusserungsrecht auch beim KI-Einsatz gewährt werden kann, in das VRG aufzunehmen.[253] Dabei wird sich vermutlich die Frage stellen, ob der Anspruch auf rechtliches Gehör im VRG nicht umfassend verankert werden sollte. Zu denken wäre an eine zusätzlichen Bestimmung, in der u. a. das Äusserungsrecht, die Begründungspflicht und die Akteneinsicht geregelt werden könnten.

b) Begründungspflicht

Beim Einsatz von KI-Anwendungen besteht die Gefahr, dass die entscheidende Behörde der Begründungspflicht nicht vollumfänglich nachkommen kann. Dies wäre insbesondere dann denkbar, wenn das KI-System Kriterien zugrunde legt, die für Menschen nicht ersichtlich sind. Für solche Systeme könnte deshalb vorgesehen werden, dass die Logik der Entscheidungsfindung in der Begründung angegeben werden muss.[254] Eine Verfügung, die mithilfe von KI erlassen wurde, sollte diesfalls somit die folgenden Angaben enthalten:[255] den Sachverhalt, die Rechtsnormen, auf die sie sich stützt, die Entscheidlogik des Algorithmus mit Art, Menge, Erhebungszeitraum und Gewichtung der Daten sowie die Anwendung der Entscheidlogik auf den konkreten Einzelfall. Darüber hinaus könnten Angaben über die Vergleichsgruppe, in welche der Algorithmus eine Person einordnet, und die im Einzelfall entscheidenden individuellen Besonderheiten verlangt werden.[256] Sollte die Begründung einer staatlichen Entscheidung anders als mittels Offenlegung des Quellcodes des Algorithmus nicht nachvollzogen werden können, wäre die Einsicht in diesen zu gewährleisten. Allerdings ist die Offenlegung des Quellcodes (allein) grundsätzlich nicht als geeignetes Mittel zur Sicherstellung der Begründung staatlicher Entscheidungen zu betrachten, da dieser für die wenigsten Bürgerinnen und Bürger aussagekräftig ist.

247 Krüger/Vogelgesang/Adam, 2020, S. 52, mit Verweisung auf das Urteil des VerfGH Saarland vom 05.07.2019 – Lv 7/17.
248 Vgl. Kapitel 2 F. II.
249 Glaser, 2018, S. 184.
250 Braun Binder, 2018, S. 112; Etscheid, 2018, S. 129.
251 Braun Binder, 2016c, S. 895; Glaser, 2018, S. 184.
252 Im Kanton Zürich werden aktuell im Rahmen des Projekts IP2.1 DigiLex die notwendigen gesetzlichen Grundlagen für den formellen elektronischen Geschäftsverkehr zwischen Privaten und der öffentlichen Verwaltung geschaffen.
253 Ergänzend könnte in § 12 IDG eine Informationspflicht über automatisierte bzw. KI-gestützte Entscheide vorgesehen werden, die es der betroffenen Person ermöglicht, ihr Recht auf vorgängige Äusserung effektiv wahrzunehmen. Vgl. Kapitel 3 A. IV. 3.
254 Vgl. auch Kapitel 3 A. II. 2. b).
255 Vgl. in diesem Sinne für automatisierte Einzelentscheide Rechsteiner, 2018, Rn. 24.
256 Martini, 2019, S. 190.

Im Kanton Zürich wäre es naheliegend, entsprechende Vorgaben in § 10 VRG zu verankern, da in Abs. 1 dieser Bestimmung die Begründungspflicht festgehalten wird. Denkbar wäre aber auch eine Regelung im IDG,[257] da dieses – anders als das DSG auf Bundesebene – nicht nur die Bearbeitung von Personendaten, sondern allgemein den Umgang von öffentlichen Organen mit Informationen regelt (§ 1 Abs. 1 IDG). Das Gesetz bezweckt, das Handeln der öffentlichen Organe transparent zu gestalten und insbesondere die Kontrolle des staatlichen Handelns zu erleichtern (§ 1 Abs. 2 lit. a IDG).[258]
Die Herausforderungen im Zusammenhang mit der Begründung staatlicher Entscheidungen führen dazu, dass in der Literatur zum Teil vorgeschlagen wird, KI-Anwendungen vorerst nur im Rahmen von Verfahren einzusetzen, die standardisiert und unabhängig von allfälligen Individualitäten durchgeführt werden können.[259] Dies mag auf einer theoretischen Ebene zwar einleuchten, allerdings sind solche komplett standardisierten, gleichförmigen Verfahren in der Praxis wohl eher selten. Auch im Steuerbereich – wo durchaus diskutiert werden kann, ob die Anforderungen an die Begründungspflicht nicht ohnehin niedriger sind und deshalb auch beim KI-Einsatz die Begründung einfacher umzusetzen wäre – ist dies nicht der Fall.[260]
Auch eine Beschränkung auf Verfahren, in denen den Begehren der Verfahrensbeteiligten vollständig entsprochen wird und eine Begründung deshalb gemäss § 10a lit. a VRG hinfällig ist, kann keine Lösung sein. Auch wenn eine Begründung in diesen Fällen nicht vorgesehen ist, muss es der Behörde auf Verlangen der Verfahrensbeteiligten möglich sein, die Verfügung nachträglich zu begründen. Dies ergibt sich zwar nicht unmittelbar aus dem VRG, aber aus dem verfassungsmässigen Recht auf Begründung, wie es Art. 18 Abs. 2 KV ZH ausdrücklich vorsieht.[261]

c) Untersuchungsgrundsatz

Nutzt das KI-System Daten, die bei der kantonalen Verwaltung vorhanden sind, so hat die Behörde sicherzustellen, dass diese Daten vollständig, korrekt und soweit zur Eruierung der rechtserheblichen Tatsachen notwendig auch verfügbar sind. Diese Anforderung könnte auf Gesetzesstufe in § 7 VRG konkretisiert werden. Zur Sicherstellung des Grundsatzes der Amtsermittlung kann es überdies notwendig sein, im Rahmen des spezifischen Fachgesetzes, in dessen Anwendungsbereich ein KI-System eingesetzt werden soll, die notwendige Rechtsgrundlage für den Zugriff auf vorhandene Datensammlungen zu schaffen. Hierbei kann ein Konflikt mit datenschutzrechtlichen Vorgaben entstehen.[262]
Der vollständig automatisierte Erlass einer Anordnung kann ausserdem dazu führen, dass die Ermittlung des Sachverhalts zumindest teilweise auf die antragstellende Person übertragen wird. Dies lässt der Untersuchungsgrundsatz nicht bedingungslos zu.[263] Aus diesem Grund scheint es sinnvoll, gesetzlich festzuhalten, dass der Untersuchungsgrundsatz auch bei vollautomatisierten Verfahren zu berücksichtigen ist. Dies könnte durch eine entsprechende Ergänzung von § 7 VRG erfolgen. Die Umsetzung der Mitwirkungspflicht im Falle automatisierter Verfahren, welche die Erfassung der Sachverhaltsdaten durch die Antragstellerin oder den Antragsteller voraussetzen, ist schliesslich im entsprechenden Fachgesetz vorzusehen, das den Bereich regelt, in dem die KI-Anwendung eingesetzt werden soll.

III. Diskriminierungsverbot

1. Grundlagen

Alle Menschen sind vor dem Gesetz gleich und niemand darf diskriminiert werden. Diese zentralen Forderungen sind in Art. 8 BV verankert. Das Diskriminierungsverbot stellt dabei einen qualifizierten Schutz vor Ungleichbehandlungen dar. Es ist die grundrechtliche Antwort auf historische Erfahrungen der Ausgrenzung, Herabwürdigung und Stigmatisierung von Menschen allein wegen eines sensiblen Persönlichkeitsmerkmals, das so wesentlich ist, dass es den Betreffenden nicht möglich oder nicht zumutbar ist, sich des Merkmals zu entledigen.[264]
Einige dieser sensiblen oder verpönten Merkmale werden in Art. 8 Abs. 2 BV ausdrücklich genannt, wobei die Aufzählung jedoch nicht abschliessend ist:[265] Herkunft, Rasse, Geschlecht, Alter, Sprache, soziale Stellung, Lebensform, religiöse, weltanschauliche oder politische Überzeugung sowie körperliche, geistige oder psychische Behinderung. Nach bundesgerichtlicher Rechtsprechung liegt eine Diskriminierung vor, «wenn eine Person ungleich behandelt wird allein aufgrund ihrer Zugehörigkeit zu einer bestimmten Gruppe, welche historisch oder in der gegenwärtigen sozialen Wirklichkeit tendenziell ausgegrenzt oder als minderwertig angesehen wird. Die Diskriminierung stellt eine qualifizierte Ungleichbehandlung von Personen in vergleichbaren Situationen dar, indem sie eine Benachteiligung von Menschen bewirkt, die als Herabwürdigung oder Ausgrenzung einzustufen ist, weil sie an Unterscheidungsmerkmale anknüpft, die einen wesentlichen und nicht oder nur schwer aufgebbaren Bestandteil der Identität der betroffenen Person ausmachen …».[266] Dabei kann zwischen direkter und indirekter Diskriminierung unterschieden werden:

– Eine direkte (unmittelbare) Diskriminierung liegt vor, wenn ein Rechtsakt ohne genügende Rechtfertigung direkt an ein sensibles Merkmal anknüpft.[267] Es besteht kein formales Verbot der Anknüpfung an diese Merkmale, sondern Anknüpfungen können durch eine qualifizierte Rechtfertigung zulässig sein.[268] Eine solche qualifizierte Rechtfertigung muss drei Voraussetzungen erfüllen. Erstens muss die Sonderbehandlung gesetzlich mit der nötigen Klarheit auf genügender Normstufe verankert sein. Zweitens muss mit der Sonderbehandlung ein Interesse des Gemeinwohls verfolgt werden. Drittens muss dieses Interesse im konkreten Fall die entgegenstehenden Interessen der betroffenen Person an gleicher Behandlung überwiegen. Basiert die Sonder-

257 Gesetz über die Information und den Datenschutz (IDG) vom 12. Februar 2007, LS 170.4.
258 Vgl. auch die Ausführungen zum Datenschutz in Kapitel 3 A. IV. 1. b).
259 Vgl. Weber, 2019, S. 18; Weber, 2020, S. 24.
260 Vgl. zu den rechtlichen Herausforderungen eines KI-Einsatzes in Steuerverfahren Kapitel 3 B. I.
261 Vgl. Plüss, 2014, N. 9 zu § 10a VRG.
262 Vgl. zu den datenschutzrechtlichen Anforderungen Kapitel 3 A. IV.
263 Braun Binder, 2020b, S. 274.
264 Statt vieler Müller/Schefer, 2008, S. 651, 684 ff.; Rhinow/Schefer/Uebersax, 2016, Rn. 1889. Waldmann, 2015, N. 45 und 56 zu Art. 8 BV.
265 Waldmann, 2015, N. 65 zu Art. 8 BV.
266 BGE 139 I 169 E. 7.2.1; 139 I 292 E. 8.2.1; 138 I 305 E. 3.3.; 136 I 297 E. 7.1; 135 I 49 E. 4.1; 132 I 49 E. 8.1; 129 I 232 E. 3.4.1; 126 II 377 E.6a.
267 Waldmann, 2015, N. 62 zu Art. 8 BV.
268 Schweizer, 2014, N. 54 zu Art. 8 BV; Waldmann, 2015, N. 62 zu Art. 8 BV. Müller/Schefer, 2008, S. 693 f., 701 ff.; Schweizer, 2014, N. 59 zu Art. 8 BV.

behandlung auf spezifischen und üblicherweise vorübergehenden Fördermassnahmen zugunsten besonders benachteiligter Gruppen, kann sie sogar geboten sein (affirmative action).[269]

– Eine indirekte (mittelbare) Diskriminierung kann die Folge einer formal neutralen gesetzlichen Regelung sein, die selbst keine offensichtlichen Benachteiligungen von Angehörigen besonders geschützter Gruppen enthält. Die Ungleichbehandlung ergibt sich erst aus den praktischen Auswirkungen der Regelung. Nicht jede mittelbare Ungleichbehandlung ist automatisch auch eine Diskriminierung. Die dadurch herbeigeführten Benachteiligungen müssen gemäss Bundesgericht aufgrund der gesamten Umstände als erheblich einzustufen sein.[270] Die Ungleichbehandlung kann – wie bei der direkten Diskriminierung – durch qualifizierte, nicht diskriminierende Gründe gerechtfertigt sein.[271]

Auf Bundesebene wird der verfassungsrechtliche Diskriminierungsschutz in einigen wenigen spezialgesetzlichen Grundlagen konkretisiert. Dazu gehören das Bundesgesetz über die Gleichstellung von Frau und Mann,[272] das Bundesgesetz über die Beseitigung von Benachteiligungen von Menschen mit Behinderungen[273] sowie die Rassismusstrafnorm in Art. 261bis des Strafgesetzbuches.[274]

In der Verfassung des Kantons Zürich werden das Rechtsgleichheitsgebot und das Diskriminierungsverbot in Art. 11 verankert. Die Diskriminierungstatbestände werden im Vergleich zu Art. 8 Abs. 2 BV noch um «genetische Merkmale» und «sexuelle Orientierung» ergänzt. Auf Gesetzesstufe wurden Bestimmungen erlassen, um Benachteiligungen von Behinderten zu beseitigen, indem einerseits ein bedarfsgerechtes Angebot an Einrichtungen mit Wohn- und Arbeitsplätzen für erwachsene invalide Menschen aus dem Kanton Zürich sichergestellt und anderseits in angemessenem Umfang der individuelle Transport von mobilitätsbehinderten Personen gewährleistet wird.[275] Ferner schuf der Gesetzgeber eine öffentlich-rechtliche Anstalt zum Zweck der Bildung und Förderung von Kindern, Jugendlichen und jungen Erwachsenen mit einer Hör- oder einer schweren Sprachbeeinträchtigung.[276] Zur Förderung der Gleichstellung von Frau und Mann wurden eine Fachstelle für Gleichberechtigungsfragen und eine Kommission für die Gleichstellung von Frau und Mann eingerichtet.[277]

2. Diskriminierungsverbot und KI

Diskriminierungsverbote sind das Ergebnis historischer Erfahrungen. Sie reflektieren somit die Tatsache, dass Menschen in ihren Entscheidungen bewusst oder unbewusst von bestimmten diskriminierenden Haltungen geleitet werden (können). Inwiefern diskriminierende Einstellungen auch in KI-Systeme gelangen und weshalb das Diskriminierungspotenzial durch KI so gross ist, wird im Folgenden näher erläutert.

a) Mögliche Diskriminierungsquellen in KI-Systemen

i. Präexistierender Bias in den Daten

Oftmals werden die in der Gesellschaft etablierten Voreingenommenheiten explizit oder implizit auf die Technologie übertragen. Diese können ihren Ursprung entweder in der Gesellschaft im Allgemeinen haben oder die persönlichen Einstellungen von Einzelpersonen (z. B. Kunden oder Systemdesigner) widerspiegeln, die einen wesentlichen Einfluss auf die Gestaltung des Systems haben.[278] So können die diskriminierenden Ergebnisse eines KI-Systems auf einem präexistierenden Bias in den Trainingsdaten basieren, welche die aktuelle Gesellschaft abbilden.[279] Dies kann am Beispiel eines automatisierten Bewerbungsbewertungssystems von Amazon erläutert werden: Einer KI-Anwendung wurden Bewerbungsunterlagen der letzten zehn Jahre als Trainingsdaten zugeführt, um erfolgreiche Arbeitnehmerinnen und Arbeitnehmer von weniger erfolgreichen zu unterscheiden. Da in der Vergangenheit jedoch hauptsächlich Männer eingestellt wurden, wurden Bewerbungen von Frauen systematisch schlechter bewertet als jene von Männern. Das KI-System lernte folglich den historischen Bias mit.[280] Das Diskriminierungspotenzial wird in der Literatur besonders im Zusammenhang mit Predictive Policing diskutiert.[281] So ist etwa bekannt, dass sich Vorurteile der Polizistinnen und Polizisten in der Auswahl der zu kontrollierenden Orte niederschlagen können. Nutzt man solche Orte als Trainingsdaten, können dadurch Verzerrungen entstehen.[282]

Ein präexistierender Bias kann aber auch von Systementwicklerinnen und -entwicklern ausgehen. Das Design eines jeden Artefakts ist an sich eine Ansammlung von Entscheidungen, angefangen bei den zu berücksichtigenden Eingaben bis hin zu den mit dem System verfolgten Zielen. Ist das Ziel eine grösstmögliche Effizienz oder sollen auch die Wirkungen auf Menschen und die Umgebung berücksichtigt werden? Ist es das Ziel des Systems, so viele potenzielle Betrügerinnen und Betrüger wie möglich zu finden, oder soll die Überwachung von unschuldigen Personen möglichst vermieden werden? Solche Entscheidungen werden auf die eine oder andere Weise von den inhärenten Vorurteilen der Personen, die sie treffen, bestimmt.[283]

269 Müller/Schefer, 2008, S. 693 f., 701 ff.; Schweizer, 2014, N. 59 zu Art. 8 BV.
270 BGE 142 V 316 E. 6.1.2.; vgl. Kiener/Kälin/Wyttenbach, 2018, S. 452.
271 Müller/Schefer, 2008, S. 695 ff.
272 Bundesgesetz über die Gleichstellung von Frau und Mann (Gleichstellungsgesetz, GIG) vom 24. März 1994, SR 151.1.
273 Bundesgesetz über die Beseitigung von Benachteiligungen von Menschen mit Behinderungen (Behindertengleichstellungsgesetz, BehiG) vom 13. Dezember 2002, SR 151.3.
274 Schweizerisches Strafgesetzbuch (StGB) vom 21. Dezember 1937, SR 311.0.
275 Gesetz über Invalideneinrichtungen für erwachsene Personen und den Transport von mobilitätsbehinderten Personen (IEG) vom 1. Oktober 2007, LS 855.2.
276 Gesetz über das Zentrum für Gehör und Sprache vom 11. Februar 2008, LS 412.41.
277 Verordnung über die Fachstelle für Gleichberechtigungsfragen und die Kommission für die Gleichstellung von Frau und Mann vom 30. Juni 1993, LS 172.6.
278 Einführung des Begriffs «preexisting bias» durch Friedmann/Nissenbaum, 1996, S. 333 f.
279 Beck/Grunwald/Jacob/Matzner, 2019, S. 8; Bryson, 2017; Hagendorff, 2019, S. 56; Kessler/Oberlin, 2020, S. 92; Kolleck/Orwat, 2020, S. 32; Weber/Henseler, 2020, S. 31.
280 Reuters Business News vom 11. Oktober 2018, Amazon scraps secret AI recruiting tool that showed bias against women, abrufbar unter https://www.reuters.com/article/us-amazon-com-jobs-automation-insight-idUSKCN1MK08G.
281 Vgl. m. w. H. TA-SWISS KI, S. 220 f.
282 Beck/Grunwald/Jacob/Matzner, 2019, S. 8.
283 Bryson, 2017; CAHAI-Bericht, 2020, Rn. 39; Kolleck/Orwat, 2020, S. 32.

ii. Technischer Bias und fehlende Daten

Im Gegensatz zum präexistierenden Bias entsteht der technische Bias durch technische Vorgaben.[284] Beispielsweise kann bei der Vergabe von Organspenden die Bildschirmgrösse entscheidend sein, da die Ergebnisse auf der ersten Seite einer Suchmaschine mit grösserer Wahrscheinlichkeit als die Resultate auf den folgenden Seiten aufgerufen werden. Wie viele Ergebnisse jeweils auf der ersten Seite angezeigt werden, hängt damit u. a. vom Bildschirm ab.[285] Ein weiteres Beispiel stellt ein automatischer Seifenspender dar, dessen Einstellung und Auswahl der Sensoren nur mit einer weissen Person als Testperson erfolgte. Hände von nicht weissen Personen erkannte der Seifenspender folglich nicht, da die Sensoren nur weisse Haut erfassen konnten.[286] Diese Form der Diskriminierung hängt somit stark mit dem Risiko von fehlenden Daten zusammen, wenn eine Bevölkerungsgruppe in den Trainingsgruppen unterrepräsentiert ist.[287] So hatte eine australische Spracherkennungssoftware, welche die Englischkenntnisse von Arbeitsvisa beantragenden Personen beurteilte, das Englisch einer gebürtigen Irin als ungenügend eingestuft.[288] Das KI-System wurde offenbar nicht mit genügend unterschiedlichen Akzenten trainiert.

iii. Emergenter Bias und statistische Diskriminierungen

Diskriminierungen können auch im Nutzungskontext mit realen Benutzerinnen und Benutzern entstehen. Eine solche Voreingenommenheit entwickelt sich typischerweise einige Zeit nach der Fertigstellung eines Designs als Ergebnis von Änderungen gesellschaftlichen Wissens, Änderungen in der Bevölkerung oder hinsichtlich kultureller Werte.[289] Ein emergenter Bias kann auch durch die falsche Interpretation der Ausgaben eines KI-Systems entstehen; dies ist ein Risiko, das bei statistischen Werten häufiger auftritt.[290] Die Diskriminierung kann insbesondere aus Korrelationen resultieren, die an sogenannte Proxies anknüpfen. Bei Proxies handelt es sich um harmlos erscheinende Merkmale, die mit verpönten Merkmalen jedoch stark korrelieren können. So kann etwa der Wohnort mit dem ethnischen Hintergrund oder der sozialen Stellung zusammenhängen.[291] Es wird der Eindruck einer Kausalität erweckt, obwohl eine solche nicht vorhanden ist (Scheinkausalität). Eine solche statistische Ungleichbehandlung stützt sich mithin häufig auf Ersatzinformationen und stellt eine Sonderform der Diskriminierung dar.[292]

iv. Diskriminierender Output durch dynamisches Weiterlernen

Bestimmte Formen der Diskriminierung entstehen aber auch erst in der Anwendung, insbesondere dann, wenn die KI-Anwendung dynamisch weiterlernen kann.[293] Vor einigen Jahren wurde z. B. ein Chatbot entwickelt, welcher auf Twitter mit Menschen interagieren sollte, indem er lernte, wie diese sich artikulierten. Der Chatbot wurde jedoch absichtlich mit rassistischen und sexistischen Tweets gefüttert, weshalb sich dieser innert Kürze ebenfalls entsprechend äusserte.[294]

b) Grosses Diskriminierungspotenzial durch KI

KI-Anwendungen benötigen (quantifizierbare) Daten, die für das System bearbeitbar sind. Geht es darum, mithilfe von KI-Anwendungen eine personenbezogene Entscheidung zu treffen, ist zu bedenken, dass auf einer solchen Datengrundlage lediglich ein fragmentarisches Bild einer Person entstehen kann. Diese Lückenhaftigkeit in der Beurteilung von Personen anhand von wenigen verfügbaren Informationen stellt zwar kein KI-spezifisches Problem dar. Dennoch wird in der Literatur auf das Risiko der Entstehung eines sogenannten digitalen Positivismus hingewiesen, wonach KI-Systeme so interpretiert werden, als würden sie die Realität abbilden bzw. transparent machen und nicht mehr kritisch hinterfragt werden.[295] Dadurch entsteht die Gefahr, dass KI-basierte Empfehlungen unhinterfragt übernommen werden, auch wenn sie unter Umständen zu einer Diskriminierung führen.

Hinzu kommt, dass KI-Anwendungen häufig auf Korrelationen basieren. Scheinkausalitäten können dabei aufgrund der Regeln des Zufalls in grossen Datenmengen immer entstehen. Dies ist insbesondere dann der Fall, wenn KI-Systeme eigenständig nach Mustern in grossen Datenmengen suchen. KI-Anwendungen können mangels theoretischer und intuitiver Überlegungen nicht zwischen Kausalitäten und Korrelationen unterscheiden und mögliche Probleme erkennen. Somit besteht die Gefahr, dass KI-Systeme Personen aufgrund von Korrelationen klassifizieren, die mit verpönten Merkmalen zusammenhängen, ohne dass diese Merkmale kausal für das gesuchte Verhalten sind.[296] Schliesslich ist zu bedenken, dass KI-Systeme im Gegensatz zu Entscheidungen einzelner Menschen eine viel grössere Breitenwirkung entfalten können, da sie potenziell wesentlich mehr Menschen betreffen.[297] Werden dabei Methoden des maschinellen Lernens angewendet, wobei die Entscheidregeln nur noch mit Schwierigkeiten oder gar nicht mehr nachvollzogen werden können, ist das Risiko gross, dass die Verzerrungen gar nicht erst bemerkt werden.[298] Wenn die betroffenen Personen mangels Nachvollziehbarkeit nicht gegen die KI-Entscheidung vorgehen (können), findet sodann keine negative Rückkopplung zur fehlerhaften Entscheidung statt. Vielmehr bestärkt das Ergebnis ein fehlerhaftes Entscheidungsmuster, wodurch diskriminierende Ungleichbehandlungen gar verfestigt werden können.[299]

284 Einführung des Begriffs «technical bias» durch Friedmann/Nissenbaum, 1996, S. 335 f.
285 Beck/Grunwald/Jacob/Matzner, 2019, S. 9; Friedmann/Nissenbaum, 1996, S. 335 f.
286 Gizmodo vom 17.08.2017, Why Can't This Soap Dispenser Identify Dark Skin?, https://gizmodo.com/why-cant-this-soap-dispenser-identify-dark-skin-1797931773.
287 Bryson, 2017; Zweig, 2019b, S. 214 f.; vgl. auch Goodman/Flaxman, 2016, S. 4.
288 Australien Associated Press vom 09.09.2017, https://www.abc.net.au/news/2017-08-09/voice-recognition-computer-native-english-speaker-visa-limbo/8789076.
289 Einführung Begriff «emergent bias» durch Friedmann/Nissenbaum, 1996, S. 336.
290 Beck/Grunwald/Jacob/Matzner, 2019, S. 9.
291 Weber/Henseler, 2020, S. 31 f.
292 Vgl. ausführlich Kolleck/Orwat, 2020, S. 34 ff., ohne jedoch den Begriff «emergent bias» zu verwenden.
293 Beck/Grunwald/Jacob/Matzner, 2019, S. 9; Zweig, 2019b, S. 2018.
294 The Verge vom 24. März 2016, Twitter taught Microsoft's AI chatbot to be a racist asshole in less than a day, abrufbar unter https://www.theverge.com/2016/3/24/11297050/tay-microsoft-chatbot-racist.
295 Vgl. m. w. H. Kolleck/Orwat, 2020, S. 33.
296 Vgl. Bericht Herausforderungen 2019, S. 32; CAHAI-Bericht, 2020, Rn. 38.
297 Datenethikkommission, 2019, S. 167.
298 Zweig, 2019b, S. 210; vgl. auch Weber/Henseler, 2020, S. 31; Martini/Nink, 2017, S. 9 f.; Bericht Herausforderungen 2019, S. 38.
299 Martini, 2019, S. 57.

3. Schlussfolgerungen für den Kanton Zürich

Die Verhinderung von Diskriminierung kann angesichts der verschiedenen Diskriminierungsquellen im Zusammenhang mit KI-Anwendungen nicht alleinige Aufgabe entsprechender Rechtsetzung sein. Im Gegenteil: Aus rechtlicher Sicht ist das Ziel – Verhinderung der Diskriminierung – klar. Die Umsetzung des Diskriminierungsverbots ist denn auch eher eine Frage der Rechtsanwendung als eine solche der Rechtsetzung. Dabei sind rechtliche, organisatorische, technische und gegebenenfalls weitere Aspekte zu berücksichtigen.

Eine der Diskriminierungsquellen bilden unrichtige Daten. Deshalb muss die Verwaltung – nicht nur als Ausfluss des Untersuchungsgrundsatzes[300] und aufgrund datenschutzrechtlicher Vorgaben[301] sicherstellen, dass die einem KI-System zugrunde liegenden Trainingsdaten sowie die Sachverhaltsdaten korrekt sind und nur Daten genutzt werden, die für das entsprechende Verfahren geeignet sind.[302] Dies kann in einer entsprechenden Vorschrift, z. B. im VRG, verankert werden.

Auch auf Basis korrekter Daten können allerdings diskriminierende (Zwischen-)Resultate entstehen. Werden diese Zwischenresultate im Sinne einer Entscheidungsempfehlung von Menschen überprüft, ist sicherzustellen, dass auch eine Entscheidung getroffen werden kann, die einer diskriminierenden Empfehlung widerspricht. Das Beispiel des österreichischen AMS-Algorithmus veranschaulicht, was damit gemeint ist. Wie bereits ausgeführt, beurteilten die österreichische Datenschutzbehörde und das österreichische Bundesverwaltungsgericht unterschiedlich, ob die Entscheidung über Weiterbildungsangebote ausschliesslich beim Algorithmus liegt oder dieser nur eine zusätzliche Information liefert und die Mitarbeitenden immer noch genügend Spielräume und Kompetenzen haben, nach eigenem Ermessen zu entscheiden.[303] Genau das gleiche KI-System, das potenziell diskriminierende Zwischenresultate liefert, kann folglich je nach Rolle der Verwaltungsmitarbeitenden zu rechtlich diskriminierenden Entscheidungen führen oder nicht. Liefert der Algorithmus nur Vorschläge und verfügt die Sachbearbeiterin bzw. der Sachbearbeiter über die notwendigen Kenntnisse und Kompetenzen, um im Einzelfall eine Entscheidung zu treffen, die vom diskriminierenden Vorschlag abweicht, kann dem Diskriminierungsverbot Genüge getan werden.

Mit Blick auf KI-Systeme, die auf maschinellen Lernmethoden beruhen, werden in der Literatur zudem weitere Massnahmen vorgeschlagen, um der Diskriminierungsgefahr entgegenzutreten. So sollen die Datensätze von Verzerrungen bereinigt werden.[304] Ferner wird empfohlen, Kontrollalgorithmen einzusetzen, um die Gewichtung der Faktoren durch die KI-Anwendungen zu analysieren.[305] Ausserdem können Drittorganisationen oder staatliche Institutionen beauftragt werden, regelmässige Kontrollen durchzuführen.[306]

Schliesslich ist festzuhalten, dass KI-Anwendungen, die auf maschinellem Lernen beruhen und personenbezogene Entscheidungen betreffen, systembedingt «ausgrenzend» vorgehen. Verschiedene Eigenschaften oder Merkmale bezüglich eines zu lernenden Verhaltens (Output, beispielsweise Missbrauchspotenzial) sollen gefunden und voneinander differenziert werden, um die Personen in die so erstellten Gruppen einzuteilen (Klassifizierung). Sofern diese Klassifizierungen auf tatsächlichen Ursachen und nicht auf Korrelationen (Scheinkausalitäten) basieren, ist dies auch wünschenswert und führt nicht zwangsläufig zu einer Diskriminierung.[307] Bei der Implementierung eines KI-Systems ist aber genau darauf zu achten, dass eine Unterscheidung, die an verpönte Merkmale anknüpft, auch tatsächlich den notwendigen qualifizierten Rechtfertigungsgründen genügt.[308]

IV. Informationelle Selbstbestimmung und Datenschutz

1. Grundlagen

a) Verfassungsrechtliche Verankerung

Unter informationeller Selbstbestimmung wird der grundrechtliche Anspruch auf Datenschutz verstanden.[309] Der Schutzbereich ist weiter, als dies der Wortlaut von Art. 13 Abs. 2 BV vermuten lässt: Erfasst ist nicht nur der Missbrauch persönlicher Daten, sondern jede Bearbeitung oder Speicherung durch den Staat oder Private.[310] Jede Person hat das Recht, zu bestimmen, ob und zu welchem Zweck Daten über sie erhoben, bearbeitet und gespeichert werden. Von Art. 13 Abs. 2 BV geschützt werden persönliche und personenbezogene Daten. Als persönliche Daten sind alle Informationen mit einem bestimmbaren Bezug zu einer (natürlichen oder juristischen) Person, insbesondere zu ihren physischen und psychischen Eigenschaften, sozialen und wirtschaftlichen Verhältnissen oder politischen Anschauungen, zu qualifizieren.[311]

Bestimmt ist eine Person nach der bundesgerichtlichen Rechtsprechung, wenn sich aus der Information selbst ergibt, um wen es sich handelt. Bestimmbar ist die Person, wenn aufgrund zusätzlicher Informationen, d. h. aus den Umständen oder dem Kontext, auf sie geschlossen werden kann. Für die Bestimmbarkeit genügt allerdings nicht jede theoretische Möglichkeit. Ist der Aufwand für die Identifizierung unverhältnismässig gross, wodurch nach der allgemeinen Lebenserfahrung nicht damit gerechnet werden muss, dass jemand diesen Aufwand aufbringen wird, fehlt es an der Bestimmbarkeit. Das Bundesgericht hält explizit fest, dass diese Frage im konkreten Fall und insbesondere unter Berücksichtigung der technischen Möglichkeiten entschieden werden muss.[312]

Die von Art. 13 Abs. 2 BV erfassten Tätigkeiten umfassen sämtlichen Umgang mit personenbezogenen Daten, z. B. Erhebung, Sammlung, Speicherung, Bearbeitung und Weitergabe.[313] Direkt aus Art. 13 Abs. 2 BV ergeben sich die folgenden grund-

300 Vgl. Kapitel 3 A. II. 3. und 4. c).
301 Kapitel 3 A. IV. 1. b) iv.
302 Vgl. TA-SWISS KI, 2020, S. 309, Empfehlung V-2.
303 Vgl. Kapitel 2 F. III. 8.
304 Vgl. m. w. H. Hagendorff, 2019, S. 60; Weber/Henseler, 2020, S. 42.
305 Martini, 2019, S. 349 ff.; Martini/Nink, 2017, S. 9 ff.
306 Vgl. Datenethikkommission, 2019, S. 167; Kolleck/Orwat, 2020, S. 61; Zweig, 2019a, S. 9 und 11.
307 Hagendorff, 2019, S. 55; Zweig, 2019b, S. 210.
308 Vgl. zu den Rechtfertigungsgründen Kapitel 3 A. III. 1.
309 Diggelmann, 2015, N. 32 zu Art. 13 BV.
310 Schweizer, 2014, N. 72 zu Art. 13 BV. Auf die grundsätzliche Problematik der Drittwirkung wird vorliegend nicht eingegangen.
311 Diggelmann, 2015, N. 33 zu Art. 13 BV.
312 BGE 138 II 346 E. 6.1.; BGE 136 II 505 E. 3.2.
313 Schweizer, 2014, N. 74 zu Art. 13 BV.

rechtlichen Ansprüche: der Anspruch auf Berichtigung falscher Daten, der Anspruch auf Löschung ungeeigneter oder nicht mehr benötigter Daten sowie der Anspruch auf Auskunft bzw. Einsicht.[314]

b) Gesetzliche Konkretisierung des Datenschutzes

Der grundrechtliche Anspruch auf Datenschutz wird in den jeweils anwendbaren Datenschutzgesetzen konkretisiert. Auf Bundesebene ist dies das Bundesgesetz über den Datenschutz (DSG).[315] Dieses Gesetz gilt gemäss Art. 2 Abs. 1 für die Datenbearbeitung durch Privatpersonen (Bst. a) und Bundesorgane (Bst. b). Zu beachten ist ferner das künftige Datenschutzgesetz (revDSG). Das totalrevidierte DSG ist am 25. September 2020 vom Parlament verabschiedet worden[316] und wird voraussichtlich im Verlauf des Jahres 2022 in Kraft treten. Für die Datenbearbeitung durch kantonale Behörden sind dagegen die jeweiligen kantonalen Datenschutzgesetze einschlägig. Für die öffentlichen Organe im Kanton Zürich wird der Umgang mit Daten durch das Gesetz über die Information und den Datenschutz (IDG) bestimmt. Das IDG regelt gemäss § 1 Abs. 1 den Umgang von öffentlichen Organen mit Informationen. Das Gesetz bezweckt, das Handeln der öffentlichen Organe transparent zu gestalten und insbesondere die Kontrolle des staatlichen Handelns zu erleichtern (§ 1 Abs. 2 lit. a IDG). Das IDG ist grundsätzlich umfassend auf die öffentlichen Organe des Kantons Zürich anwendbar.[317]

In § 3 IDG werden die im Gesetz verwendeten Begriffe definiert. Informationen sind demnach alle Aufzeichnungen, die der Erfüllung einer öffentlichen Aufgabe dienen, unabhängig von ihrer Darstellungsform und ihren Informationsträgern. Ausgenommen sind Aufzeichnungen, die nicht fertiggestellt oder ausschliesslich zum persönlichen Gebrauch bestimmt sind (§ 3 Abs. 2 IDG). Voraussetzung ist, dass sich die Informationen auf die Erfüllung von öffentlichen Aufgaben beziehen und in einer (allerdings unspezifischen) Art und Weise aufgezeichnet sind. Die Information kann u. a. auf einem elektronischen Datenträger vorhanden sein, in maschinenlesbarem Code erscheinen und als digitale, numerische oder alphanummerische Zeichen dargestellt werden.[318] Somit fallen Algorithmen unter den Begriff der Information im Sinne des IDG. Gemäss § 3 Abs. 3 IDG sind Personendaten eine Unterkategorie von Informationen und beziehen sich auf eine bestimmte oder bestimmbare Person. Hier liegt ein entscheidender Unterschied zum DSG, das gemäss Art. 2 Abs. 1 nur auf Personendaten anwendbar ist. Der Geltungsbereich des IDG ist somit weiter als derjenige des DSG.

Das Datenschutzrecht wird von einigen zentralen Grundsätzen beherrscht, welche sowohl im DSG als auch in zahlreichen kantonalen Datenschutzgesetzen verankert sind. Im Folgenden werden diejenigen Grundsätze skizziert, die mit Blick auf den KI-Einsatz von besonderem Interesse sind.

i. Allgemeine Grundsätze staatlichen Handelns

In Art. 4 DSG (Art. 6 revDSG) werden zunächst die allgemeinen Grundsätze staatlichen Handelns für das Datenschutzrecht statuiert: das Gebot der Rechtmässigkeit, das Verhalten nach Treu und Glauben sowie der Grundsatz der Verhältnismässigkeit.

Für die öffentliche Verwaltung bedeutet *Rechtmässigkeit* nichts anderes als *Gesetzmässigkeit.* Das Bearbeiten von Personendaten bedarf mithin einer gesetzlichen Grundlage (Art. 17 DSG, Art. 34 revDSG).[319] Dabei ist festzuhalten, dass das Datenschutzgesetz selbst nicht die Grundlage für die Datenbearbeitung darstellt. Diese findet sich vielmehr in den jeweiligen Sachgesetzen.[320] Für die Beantwortung der Frage, auf welcher Normstufe[321] die entsprechende Rechtsgrundlage vorgesehen sein muss, ist in erster Linie die Unterscheidung zwischen gewöhnlichen und besonders schützenswerten Personendaten (Art. 3 DSG, Art. 5 revDSG) entscheidend. Besonders schützenswerte Personendaten im Sinne des DSG sind Personendaten, welche besonders sensible Bereiche betreffen:[322] nach dem geltenden DSG z. B. die religiösen oder politischen Ansichten, Gesundheit, strafrechtliche Verfolgung oder Sozialhilfemassnahmen (Art. 3 Bst. c Ziff. 1–4). Das revidierte DSG fügt dieser Kategorie noch genetische und biometrische Daten hinzu (Art. 5 Bst. c Ziff. 3 und 4 revDSG). Für die Bearbeitung von gewöhnlichen Personendaten kann eine Grundlage in einer Verordnung genügen, allerdings ist immer auf die Intensität des konkreten Eingriffs abzustellen. Liegt nach den Gesamtumständen ein schwerer Eingriff vor, ist dennoch eine formell-gesetzliche Grundlage erforderlich.[323] Für die Bearbeitung besonders schützenswerter Personendaten ist dagegen gemäss Art. 17 Abs. 2 DSG eine formell-gesetzliche Grundlage notwendig. Art. 34 Abs. 2 revDSG dehnt das Erfordernis der formell-gesetzlichen Grundlage auf Profiling und schwere Eingriffe in die Grundrechte der betroffenen Person aus.[324]

Neben der angemessenen Normstufe muss die gesetzliche Grundlage auch eine hinreichende Bestimmtheit aufweisen.[325] Die erforderliche Normdichte beurteilt sich nach den Anforderungen des Legalitätsprinzips sowie den Auswirkungen der Datenbearbeitung für die Betroffenen. Je sensibler die zu bearbeitenden Daten sind und je grösser die Zahl der Betroffenen ist, desto höher sind die Anforderungen an die Regelungsdichte. Für die Bearbeitung von besonders schützenswerten Personendaten muss die gesetzliche Grundlage mindestens den Zweck der Bearbeitung, die beteiligten Behörden sowie den Umfang der Bearbeitung festlegen.[326]

Der Grundsatz der Gesetzmässigkeit ergibt sich auch aus dem kantonalen Datenschutzrecht. Gemäss § 8 Abs. 1 IDG dürfen öffentliche Organe Personendaten bearbeiten, soweit dies zur Erfüllung ihrer gesetzlich umschriebenen Aufgabe geeignet und erforderlich ist. Das Bearbeiten von besonderen Personendaten bedarf gemäss § 8 Abs. 1 IDG einer formell-gesetzlichen

314 Biaggini, 2017, N. 14 zu Art. 13 BV.
315 Bundesgesetz über den Datenschutz (DSG) vom 19. Juni 1992, SR 235.1.
316 Totalrevidierte Fassung vom 25. September 2020, BBl 2020 7639, abrufbar unter https://www.admin.ch/opc/de/federal-gazette/2020/7639.pdf.
317 Siehe hierzu § 2 IDG; vgl. auch die Ausnahmen in §§ 2a–2c IDG.
318 Rudin, 2012, N. 7 zu § 3 IDG.
319 Baeriswyl, 2015, N. 1 zu Art. 4 DSG.
320 Rudin, 2012, N. 6 zu Art. 2 DSG.
321 Vgl. dazu die Ausführungen zum Legalitätsprinzip Kapitel 3 A. I. 2. a).
322 Folglich wird auf die «Herkunft» der Daten abgestellt, d. h. darauf, welche Lebensbereiche durch die Daten abgebildet werden, vgl. dazu auch Glass, 2018.
323 Mund, 2015, N. 9 zu Art. 17 DSG.
324 Indem das revDSG nun bei dem Erfordernis der Normstufe sich ebenfalls explizit auf die Grundrechtsrelevanz bezieht, wird es der kantonalen Datenschutzgesetzen etwas angeglichen, vgl. dazu Glass, 2018.
325 Vgl. dazu die Ausführungen zum Legalitätsprinzip Kapitel 3 A. I. 2. a) und b).
326 Mund, 2015, N. 6 f. zu Art. 17 DSG.

Grundlage. «Besondere Personendaten» sind nach § 3 Abs. 4 IDG einerseits Informationen, bei denen wegen ihrer Bedeutung, der Art ihrer Bearbeitung oder der Möglichkeit ihrer Verknüpfung mit weiteren Daten die besondere Gefahr einer Persönlichkeitsverletzung besteht (sogenannte sensitive Daten; § 3 Abs. 4 lit. a IDG), und anderseits Informationen, die eine Beurteilung wesentlicher Aspekte der Persönlichkeit natürlicher Personen erlauben (sogenannte Persönlichkeitsprofile; § 3 Abs. 4 lit. b IDG). Sensitive Personendaten werden beispielhaft aufgeführt und betreffen z. B. Informationen über die politischen oder religiösen Ansichten, die Gesundheit, Intimsphäre, Rassenzugehörigkeit oder ethnische Herkunft, Massnahmen der sozialen Hilfe sowie administrative oder strafrechtliche Verfolgungen oder Sanktionen.

Der *Grundsatz von Treu und Glauben* ist in Art. 5 Abs. 3 und Art. 9 BV verfassungsrechtlich verankert. Er bindet die Staatsorgane (und die Privaten) bei sämtlichen Tätigkeiten. Die Festschreibung in Art. 4 Abs. 2 DSG (Art. 6 Abs. 2 revDSG) stellt deshalb nur eine Wiederholung dar. Im IDG findet sich eine solche nicht; der Grundsatz von Treu und Glauben wird allerdings auch in Art. 2 Abs. 3 KV ZH festgehalten und hätte aufgrund seiner Verankerung in der BV ohnehin auch für kantonale Behörden Gültigkeit. Für den Bereich des Datenschutzes lässt sich aus dem Prinzip von Treu und Glauben insbesondere Folgendes ableiten: Erstens sind sämtliche rechtlichen Regelungen zum Datenschutz nach Treu und Glauben zu handhaben. Zweitens können sich Informationspflichten ergeben, die über das gesetzlich festgelegte Minimum hinausgehen. Drittens besteht bei Datenpannen ein Anspruch auf Information. Viertens und letztens kann auch eine grundsätzlich rechtmässige Datenbearbeitung wegen Verstosses gegen Treu und Glauben rechtswidrig sein, z. B. eine heimliche Datenbearbeitung, mit welcher die betroffene Person nicht rechnen muss.[327]

Auch der Grundsatz der *Verhältnismässigkeit* ergibt sich bereits aus der Bundesverfassung (Art. 5 Abs. 2 BV) bzw. aus der kantonalen Verfassung (Art. 2 Abs. 2 KV ZH) und wird in Art. 4 Abs. 2 DSG (Art. 6 Abs. 2 revDSG) lediglich wiederholt. Das IDG nennt den Grundsatz nicht nochmals, hält aber in § 11 fest, dass Datenverarbeitungssysteme so auszugestalten sind, dass möglichst wenig Personendaten anfallen.

Für das Datenschutzrecht gilt, dass Datenbearbeitungen dann verhältnismässig sind, wenn die bearbeiteten Daten zur Erreichung des Zwecks geeignet sind sowie nur diejenigen Daten bearbeitet werden, welche zur Erreichung des Zwecks erforderlich sind.[328] Konkret ergeben sich aus dem Verhältnismässigkeitsprinzip die folgenden zentralen Grundsätze für das Datenschutzrecht:[329] Personendaten dürfen nur so lange aufbewahrt werden, wie dies für den Zweck geeignet und erforderlich ist; wenn möglich sind die Erhebung und die Bearbeitung von Personendaten zu vermeiden (Prinzip der *Datenvermeidung);* und es dürfen nur diejenigen Personendaten erhoben und bearbeitet werden, die zur Erreichung des Zwecks notwendig sind (Prinzip der *Datensparsamkeit).*

ii. Zweckbindung

Art. 4 Abs. 3 DSG (Art. 6 Abs. 3 revDSG) bzw. § 9 IDG verankern den Grundsatz der Zweckbindung. Demgemäss dürfen Datenbearbeitungen nur zu einem spezifischen Zweck erfolgen. Datenbeschaffungen auf Vorrat verstossen sowohl gegen das Prinzip der Verhältnismässigkeit als auch gegen den Grundsatz von Treu und Glauben. Zudem dürfen Daten nur zu dem Zweck bearbeitet werden, zu dem sie ursprünglich erhoben wurden. Jede weitergehende Verwendung zu einem anderen Zweck ist rechtswidrig, sofern nicht auch für diese Verwendung wiederum eine gesetzliche Grundlage besteht.[330]

Zweckänderungen sind demnach zulässig, wenn dafür eine rechtliche Grundlage besteht (für besondere Personendaten in einem formellen Gesetz).[331] Eine Ausnahme vom Verbot der Zweckänderung normiert § 9 Abs. 2 IDG: Eine Behörde darf die von ihr erhobenen Daten zu nicht personenbezogenen Zwecken verwenden, z. B. für statistische Auswertungen.[332] Diese Daten sind so bald wie möglich zu anonymisieren, nicht anonymisierte Daten müssen nach der Auswertung vernichtet werden.[333]

iii. Erkennbarkeit und Informationspflicht

Art. 4 Abs. 4 DSG (Art. 6 Abs. 4 revDSG) statuiert den Grundsatz der Erkennbarkeit, nach dem sowohl die Erhebung und Bearbeitung an sich als auch der Zweck der Datenbearbeitung für die betroffenen Personen erkennbar sein müssen. Diese Bestimmung ist von grundlegender Bedeutung: Die Erkennbarkeit der Datenbearbeitung bildet die Voraussetzung für die Wahrnehmung der durch Art. 13 Abs. 2 BV vermittelten Ansprüche. Erkennbarkeit bedeutet dabei, aus den konkreten Umständen ersichtlich zu sein.[334] Inhaltlich müssen die Beschaffung der Daten, der Zweck der Bearbeitung sowie die allgemeinen Rahmenbedingungen erkennbar sein.[335]

Zudem existieren spezifische Informationspflichten (vgl. Art. 18a DSG, Art. 19 revDSG).[336] Nach Art. 18a Abs. 1 DSG (Art. 19 Abs. 1 revDSG) bzw. § 12 IDG müssen die Verantwortlichen die betroffenen Personen über die Beschaffung und Bearbeitung der Daten informieren. Die Informationspflicht entfällt, wenn für die Datenbearbeitung eine gesetzliche Grundlage besteht, welche die entsprechenden Angaben enthält (Art. 18a Abs. 4 Bst. a DSG, Art. 20 Abs. 1 Bst. b revDSG bzw. § 12 Abs. 2 lit. b IDG). Die Existenz einer gesetzlichen Grundlage bedeutet hingegen nicht, dass der Informationspflicht zwingend im Gesetz nachzukommen ist. Vielmehr kann der datenschutzrechtlichen Informationsverpflichtung auch auf andere Weise, etwa durch behördliche Datenschutzerklärungen, nachgekommen werden.[337]

iv. Datenrichtigkeit, Berechtigungsrecht und Löschungsanspruch

Art. 5 Abs. 1 DSG (Art. 6 Abs. 5 revDSG) verankert das Prinzip der Datenrichtigkeit. Ziel der Bestimmung ist, Persönlichkeitsverletzungen aufgrund mangelhafter Datenqualität zu verhindern.[338] Daten sind so zu erheben, zu verwalten und zu bearbeiten, dass der Zweck erreicht werden kann und die

327 Vgl. Baeriswyl, 2015, N. 17 ff. zu Art. 4 DSG.
328 Baeriswyl, 2015, N. 21 zu Art. 4 DSG.
329 Vgl. Baeriswyl, 2015, N. 23 zu Art. 4 DSG.
330 Vgl. Baeriswyl, 2015, N. 34 ff. zu Art. 4 DSG.
331 Harb, 2012, N. 6 zu § 9 IDG.
332 Harb, 2012, N. 15 zu § 9 IDG.
333 Harb, 2012, N. 1–2 zu § 9 IDG.
334 Vgl. Baeriswyl, 2015, N. 49 zu Art. 4 DSG.
335 Vgl. Baeriswyl, 2015, N. 52 zu Art. 4 DSG.
336 Vgl. Baeriswyl, 2015, N. 47 f. zu Art. 4 DSG.
337 Vgl. dazu Thouvenin/Braun Binder, i. V.
338 Vgl. Baeriswyl/Blonski, 2015, N. 5 zu Art. 5 DSG.

Datenbearbeitung zu korrekten Ergebnissen führt. Dies stellt Anforderungen sowohl an die Qualität der bearbeiteten Daten als auch an die Prozesse der Datenbearbeitung. Art. 5 Abs. 1 DSG (Art. 6 Abs. 5 revDSG) enthält eine Vergewisserungspflicht: Die für die Bearbeitung verantwortlichen Personen und Organe haben aktiv dafür zu sorgen, dass die erhobenen Daten und die Bearbeitungsprozesse den Qualitätsanforderungen genügen.[339] Parallel dazu besteht eine Berichtigungs- und Vernichtungspflicht, die ebenfalls in Art. 5 Abs. 1 DSG (Art. 6 Abs. 5 revDSG) festgehalten ist: Unrichtige oder unvollständige Daten sind zu berichtigen oder zu vernichten.[340] Der Anspruch auf Löschung von ungeeigneten oder nicht mehr benötigten Daten ergibt sich auch aus Art. 13 Abs. 2 BV.[341]
Im IDG ist zwar kein grundsätzliches Prinzip der Datenrichtigkeit verankert; ein solches liesse sich jedoch aus dem Grundsatz der Gesetzmässigkeit und der Verhältnismässigkeit[342] sowie aus dem Untersuchungsgrundsatz[343] herleiten. Indirekt ergibt es sich auch aus § 21 Abs. 1 lit. a IDG, wonach die betroffene Person vom öffentlichen Organ verlangen kann, unrichtige Personendaten zu berichtigen oder zu vernichten.

v. Folgenabschätzung

§ 10 IDG verpflichtet die öffentlichen Organe, vor einer Datenbearbeitung in einer Folgenabschätzung die grundrechtlichen Folgen zu bewerten sowie bei einem besonderen Risiko Rücksprache mit der oder dem Datenschutzbeauftragten zu halten. Art. 22 revDSG sieht neu auf Bundesebene ebenfalls eine Datenschutz-Folgenabschätzung vor. Damit sollen Risiken, welche für die betroffene Person durch den Einsatz bestimmter Datenbearbeitungen entstehen können, erkannt und bewertet werden.[344] Auf dieser Basis können sodann im Bedarfsfall angemessene Massnahmen formuliert werden. Eine Datenschutz-Folgenabschätzung ist immer dann durchzuführen, wenn die vorgesehene Datenbearbeitung voraussichtlich zu einem hohen Risiko für die Persönlichkeit oder die Grundrechte der betroffenen Person führt (Art. 22 Abs. 1 revDSG).

2. Datenschutz und KI

KI-Systeme sind auf grosse Datenmengen angewiesen, um sinnvoll eingesetzt werden zu können. Vielfach werden dabei auch Personendaten bearbeitet. Die Systeme arbeiten immer nur so gut, wie es die Qualität der verarbeiteten Daten bzw. der Trainingsdaten erlaubt.[345] Die erwähnten gesetzlichen Konkretisierungen[346] sind deshalb auch beim Einsatz von KI-Systemen von zentraler Bedeutung.
Darüber hinaus sind auf Bundesebene im Rahmen der Totalrevision des Datenschutzgesetzes Bestimmungen ergänzt worden, denen ein Bezug zu KI zugeschrieben werden kann. Hierzu zählen Vorgaben zu automatisierten Einzelentscheiden (Art. 21 revDSG, Art. 25 Abs. 2 Bst. f revDSG) und zum Profiling (Art. 5 Bst. f und g revDSG, Art. 34 Abs. 2 Bst. b revDSG). Inwieweit der Kanton Zürich diese oder ähnliche Bestimmungen auf kantonaler Ebene übernimmt, ist zum Zeitpunkt der Erstellung dieser Studie noch nicht absehbar.[347]

a) Automatisierte Einzelentscheidungen

Art. 21 Abs. 1 revDSG sieht eine Informationspflicht für automatisierte Einzelentscheidungen vor. Abs. 4 enthält eine Sonderregel für Bundesorgane: Diese müssen automatisiert ergangene Einzelentscheidungen entsprechend kennzeichnen. Zudem kann gemäss Abs. 2 die Überprüfung der Entscheidung durch eine natürliche Person verlangt werden (vorbehaltlich der Ausnahmen gemäss Abs. 3).
In Bezug auf KI enthält Art. 21 Abs. 4 revDSG kaum rechtsgestaltende Vorgaben. So ist der Anwendungsbereich der Norm auf Entscheidungen beschränkt, die vollautomatisiert getroffen werden. Verfahren, in denen KI-Systeme zur Entscheidungsunterstützung eingesetzt werden, sind dagegen nicht erfasst. Damit fällt ein wichtiger Anwendungsfall von KI aus dem Regelungsbereich der Norm.[348] Ferner sind Anwendungen, die nicht auf personenbezogenen Daten basieren, nicht von der Norm erfasst. Im Übrigen statuiert die Bestimmung nichts, was aufgrund des geltenden Verwaltungsverfahrensrechts nicht ohnehin bereits gilt.[349] Schliesslich bildet Art. 21 Abs. 4 revDSG allein keine ausreichende gesetzliche Grundlage für den Erlass vollautomatisierter Verfügungen. Vielmehr müsste dafür eine hinreichend bestimmte, formell-gesetzliche Grundlage geschaffen werden (Art. 5 Abs. 1 BV i. V. m. Art. 164 Abs. 1 Bst. b und g BV).[350]
Erwähnenswert ist ferner Art. 25 Abs. 2 Bst. f revDSG, der ein Auskunftsrecht bezüglich der Logik, auf welcher die automatisierte Einzelentscheidung beruht, statuiert.[351] Gemäss Botschaft des Bundesrates bedeutet dies, dass die Grundannahmen des Algorithmus offengelegt werden müssen.[352] Auch diese Regelung enthält mit Blick auf Verwaltungsverfahren nichts Neues: Die Offenlegung der Algorithmuslogik stellt nichts anderes als die Angabe der Gründe für die Entscheidung dar, was aufgrund von Art. 35 Abs. 1 VwVG ohnehin verlangt ist.[353]

b) Profiling

Art. 5 Bst. f revDSG führt den Begriff Profiling im DSG ein. Darunter ist jede automatisierte Bearbeitung von Personendaten zu verstehen, welche den Zweck hat, bestimmte Aspekte des persönlichen Lebens zu analysieren und eventuell vorherzusagen, z. B. die wirtschaftliche Lage, die Gesundheit, persönliche Vorlieben und Interessen usw. Art. 5 Bst. g revDSG definiert zusätzlich Profiling, das ein hohes Risiko für die Grund- und Persönlichkeitsrechte der betroffenen Person mit sich bringt, weil die Verknüpfung von Daten erlaubt, wesentliche Aspekte der Persönlichkeit zu beurteilen, als «Profiling mit hohem Risiko». Für ein Profiling jeglicher Art ist eine hinreichend bestimmte

339 Vgl. Baeriswyl/Blonski, 2015, N. 13 zu Art. 5 DSG.
340 Vgl. Baeriswyl/Blonski, 2015, N. 17 zu Art. 5 DSG.
341 Biaggini, 2017, N. 14 zu Art. 13 BV.
342 So mit Blick auf die Informationsrichtigkeit gemäss § 7 Abs. 2 lit. b IDG Baeriswyl, 2012, N. 24 zu § 7 IDG.
343 Vgl. Kapitel 3 A. II. 3.
344 Botschaft E-DSG, S. 7059.
345 Braun Binder, 2019b, S. 475.
346 Kapitel 3 A. IV. 1.
347 Vgl. Medienmitteilung vom 2. Juni 2020 «Mehr Datenschutz und Transparenz im Kanton Zürich», https://www.zh.ch/de/news-uebersicht/medienmitteilungen/2020/06/mehr-datenschutz-und-transparenz-im-kanton-zuerich.html, wonach mit Blick auf die laufende Digitalisierung der kantonalen Verwaltung für die nächsten Jahre eine vollständige Überarbeitung des IDG vorgesehen ist.
348 Braun Binder, 2019b, S. 475 f.
349 Braun Binder, 2020b, S. 257 ff.
350 Braun Binder, 2020b, S. 260 f.
351 Vgl. auch die Ausführungen zur Begründungspflicht in Kapitel 3 A. II. 2. b) und 4. b).
352 Botschaft E-DSG, S. 7067.
353 Braun Binder, 2020b, S. 260.

Grundlage in einem formellen Gesetz erforderlich (Art. 34 Abs. 2 Bst. b revDSG). Damit entspricht das revDSG der bundesgerichtlichen Rechtsprechung zur informationellen Selbstbestimmung.[354]

3. Schlussfolgerungen für den Kanton Zürich

Das Datenschutzrecht ist eines der zentralen Rechtsgebiete, welche bei der Einführung und Nutzung von KI-Systemen zu berücksichtigen sind. Dies gilt jedenfalls, soweit personenbezogene Daten bearbeitet werden. Von den erwähnten Grundsätzen kommt dem Grundsatz der Datenrichtigkeit eine zentrale Bedeutung zu, da von unrichtigen Daten im Rahmen von KI-Anwendungen auch ein erhöhtes Diskriminierungspotenzial ausgeht.[355] Neben der Korrektheit einzelner Personendaten spielen bei der Bearbeitung grosser Datenmengen aber auch das Umfeld und der Zweck der Datenbearbeitung eine grosse Rolle.[356] Des Weiteren sind die datenschutzrechtlichen Transparenzvorgaben (im Kanton Zürich insbesondere die Informationspflicht gemäss § 12 IDG) mit Blick auf KI-Anwendungen interessant. Sie zeigen einen möglichen Weg auf, um nicht nur datenschutzrechtliche, sondern auch allgemeine verfahrensrechtliche Anforderungen wie die Begründungspflicht[357] umzusetzen.

Ausserdem hat sich das Datenschutzrecht auf Bundesebene als Regelungsort für automatisierte Einzelentscheide und Profiling ergeben. Während der Regelungsort für Aspekte der staatlichen KI-Nutzung auf Bundesebene nicht ideal ist, da der KI-Einsatz auch Fragen aufwirft, die über den Regelungsbereich des DSG hinausgehen,[358] könnte im Kanton Zürich das IDG neben dem VRG den zentralen Regelungsort für allgemeine KI-Vorgaben darstellen. Der Anwendungsbereich des IDG ist nicht auf die Bearbeitung von Personendaten beschränkt, sondern enthält auch die Vorgaben zur Umsetzung des Öffentlichkeitsprinzips (§ 1 Abs. 2 lit. a IDG).[359] Durch eine Regelung im IDG könnten mithin nicht nur datenschutzrechtliche Anforderungen, sondern auch Transparenzanforderungen[360] abgedeckt werden und – bei einer breiteren Regelung hinsichtlich der Datenrichtigkeit – das Diskriminierungsverbot (teilweise) umgesetzt werden. Insbesondere könnte zu diesem Zweck § 7 IDG dahingehend ergänzt werden, dass die Datenrichtigkeit und Datensicherheit automatisierter Verfahren im Besonderen sicherzustellen ist, indem regelmässige Kontrollen stattfinden. Die von der öffentlichen Verwaltung für die Nutzung von KI-Systemen verwendeten Daten müssen qualitativ hochwertig und korrekt sein. Eine regelmässige Qualitätsüberprüfung ist somit äusserst wichtig, um den Grundsätzen der Datenrichtigkeit und Datensicherheit zu genügen. Die Daten müssen zu jedem Zeitpunkt aktuell und vollständig sein. Dieser Standard muss bereits für die Erhebung von Trainingsdaten gelten, d. h., dass die Qualität der Daten bereits in der Trainingsphase kontrolliert werden muss, denn bereits falsche Trainingsdaten haben zur Folge, dass die KI-Systeme unkorrekte Resultate liefern.[361] Die skizzierten Anforderungen beziehen sich auf alle in KI-Systemen genutzten Daten, d. h. sowohl Sachdaten als auch Personendaten.

Ausserdem würde es sachdienlich erscheinen, in § 12 IDG die Informationspflicht über die Beschaffung der Personendaten um eine Informationspflicht hinsichtlich einer automatisierten Datenbearbeitung bzw. einer KI-gestützten Datenbearbeitung zu ergänzen. Das Ziel einer solchen Bestimmung wäre, die betroffene Person darüber zu informieren, dass ihre Daten automatisiert bzw. KI-gestützt bearbeitet werden und der daraus resultierende Entscheid Rechtswirkungen für sie entfaltet. Im Unterschied zur Vorgabe in Art. 21 revDSG[362] sollte eine solche Informationspflicht allerdings nicht nur in Fällen einer Vollautomatisierung, sondern auch bei einer KI-gestützten Teilautomatisierung greifen. Die Informationspflicht sollte ausserdem nicht erst nach dem ergangenen Entscheid bzw. der erlassenen Verfügung, sondern bereits im Vorfeld greifen. Aus dem Anspruch auf rechtliches Gehör (Art. 29 Abs. 2 BV) ergibt sich das vorgängige Äusserungsrecht,[363] das ein Recht auf Orientierung umfasst. Die betroffene Person muss einerseits darüber informiert werden, dass und worüber ein Verfahren hängig ist. Anderseits hat sie auch einen Anspruch darauf, über den Verfahrensablauf sowie über die Vorgänge, die für den Entscheid wesentlich sind, in Kenntnis gesetzt zu werden.[364] Ist die betroffene Person mit einer automatisierten Bearbeitung nicht einverstanden, kann sie Einwände dagegen erheben.[365] Mit der Ergänzung in § 12 IDG würde mithin die vorgeschlagene Verankerung des Rechts auf vorgängige Äusserung im VRG ergänzt.[366] Die Information würde es der betroffenen Person ermöglichen, dieses Recht effektiv wahrzunehmen. Ob diese Information als individuelle Mitteilung an die betroffene Person oder eher im Sinne einer allgemeinen Erklärung (etwa als Bestandteil einer behördlichen Datenschutzerklärung)[367] zu erfolgen hat, ist in Abhängigkeit des konkreten Einsatzfeldes von KI zu beurteilen.

V. Ermessen und unbestimmte Rechtsbegriffe (offene Normen)

1. Grundlagen

Der Gesetzgeber ist oft nicht in der Lage, Normen so zu formulieren, dass sie alle möglichen Anwendungskonstellationen erfassen und damit die Verwaltungstätigkeit präzise vorprogrammieren. Häufig ist eine sinnvolle und gerechte Rechtsanwendung erst dann möglich, wenn die konkreten Umstände des Einzelfalls bekannt sind. Deshalb sind viele Normen so formuliert, dass sie der rechtsanwendenden Behörde einen Entscheidungsspielraum einräumen. Die Rede ist diesfalls auch von offenen Normen.[368] Diese verankern die Ziele, Eck-

354 Vgl. nur Urteil zur automatischen Fahrzeugfahndung und Verkehrsüberwachung, BGer 6B_908/2018 E. 3.2.
355 Vgl. Kapitel 3 A. III. 2. Vgl. auch Baeriswyl/Blonski, 2015, N. 27 zu Art. 5 DSG zu den besonders hohen Anforderungen an die Richtigkeit von Daten im Bereich von Big-Data-Analysen.
356 Vgl. Baeriswyl/Blonski, 2015, N. 3 zu Art. 5 DSG.
357 Vgl. Kapitel 3 A. II. 2. b) und 4. b).
358 Vgl. Kapitel 3 A. IV. 2. a).
359 Baeriswyl, 2012, N. 3 ff. zu § 1 IDG.
360 Vgl. Kapitel 3 A. VII.
361 Braun Binder, 2019b, S. 473.
362 Vgl. Kapitel 3 A. IV. 2. a).
363 Vgl. dazu Kapitel 3 A. II. 2. a).
364 Häfelin/Müller/Uhlmann, 2020, Rn. 1010a; BGE 144 I 11 E. 5.3.
365 Häfelin/Müller/Uhlmann, 2020, Rn. 1010 und 1195.
366 Kapitel 3 A. II. 4. a).
367 Vgl. zur Idee behördlicher Datenschutzerklärungen Thouvenin/Braun Binder i. V.
368 Häfelin/Müller/Uhlmann, 2020, Rn. 391.

werte oder den Rahmen für die Verwaltungstätigkeit. Sie umschreiben entweder den Tatbestand oder die Rechtsfolge in offener Weise. Denkbar ist auch eine Kombination von offener Umschreibung sowohl auf der Tatbestands- als auch auf der Rechtsfolgeseite.[369] Offene Normen verfolgen entweder das Ziel, Einzelfallgerechtigkeit herzustellen, oder sie wollen sachliche Richtigkeit erreichen.[370] Sowohl auf der Tatbestandsseite als auch auf der Rechtsfolgeseite kann die Offenheit der Norm durch Ermessensspielräume oder die Verwendung unbestimmter Rechtsbegriffe erreicht werden.

a) Ermessen

Ermessen ist der Entscheidungsspielraum, den der Gesetzgeber der rechtsanwendenden Behörde in Form eines offen formulierten Rechtssatzes einräumt. Die Behörde kann dadurch in einem bestimmten Einzelfall unter Beachtung der konkreten Umstände entscheiden.[371] Die Behörde ist dabei in ihrer Entscheidung allerdings nicht völlig frei. Sie ist insbesondere an das Rechtsgleichheitsgebot, das Willkürverbot und das Verhältnismässigkeitsprinzip gebunden (sogenannte pflichtgemässe Ermessensausübung).[372]

Das Begründungserfordernis[373] spielt bei der Sicherstellung des pflichtgemässen Ermessens der Behörde eine wichtige Rolle. Je weiter der Ermessensspielraum der Verwaltungsbehörde ist, desto besser muss sie den Entscheid begründen.[374] Ermessen kann z. B. dahingehend vorliegen, dass die Verwaltungsbehörde über einen Spielraum beim Entscheid verfügt, ob eine Massnahme zu treffen ist oder nicht (sogenanntes Entschliessungsermessen).[375] Verfügt die Behörde über einen Entscheidungsspielraum hinsichtlich der Wahl zwischen verschiedenen Massnahmen oder bezüglich der näheren Ausgestaltung einer Massnahme, spricht man von Auswahlermessen.[376] Sowohl Entschliessungsermessen als auch Auswahlermessen betreffen die Seite der Rechtsfolge. Auf der Tatbestandsseite angesiedelt ist das sogenannte Tatbestandsermessen. Ein solches liegt vor, wenn die Verwaltungsbehörde über einen Entscheidungsspielraum hinsichtlich der Frage verfügt, ob die Voraussetzungen für die Anordnung von Massnahmen erfüllt sind oder nicht.[377]

b) Unbestimmte Rechtsbegriffe

Von einem unbestimmten Rechtsbegriff wird gesprochen, wenn die Rechtsfolgen oder deren Voraussetzungen offen formuliert sind. Auch hier wird der Behörde ein Entscheidungsspielraum eingeräumt.[378] In der jüngeren Literatur wird argumentiert, auf die Figur des unbestimmten Rechtsbegriffs könne verzichtet werden, da es auch bei solchen Normen darum gehe, dass Ermessen eingeräumt werde.[379]

2. Offene Normen und KI

Wie erwähnt liegt den offenen Normen insbesondere das Anliegen der Herstellung von Einzelfallgerechtigkeit zugrunde. In der Literatur wird argumentiert, dass eine solche Einzelfallgerechtigkeit nur auf Basis einer menschlichen Willensbildung erreicht werden könne.[380] Dies spricht dagegen, KI einzusetzen, wo ein Ermessens- oder Beurteilungsspielraum besteht.

Wo einer Behörde Ermessen eingeräumt wird, darf sie nicht von Anfang an oder teilweise, z. B. durch den Erlass einer automatisierten Verfügung, darauf verzichten, Ermessen auszuüben (Ermessensunterschreitung).[381] Dasselbe gilt auch für unbestimmte Rechtsbegriffe: Diese müssen pflichtgemäss ausgelegt werden, ansonsten spricht man von einer Unterschreitung des Beurteilungsspielraums. Die Ermessensunterschreitung stellt eine Rechtsverletzung dar.[382]

Beim Erlass vollautomatisierter Einzelentscheidungen kann der Ermessensspielraum der Behörde gerade nicht ausgeschöpft werden. Regelbasierte Algorithmen sind immun gegen Regelabweichungen, weshalb sie keine einzelfallgerechten Entscheidungen treffen können. Hingegen kann eine natürliche Person, die ein Ermessen auszuüben hat, den Sachverhalt kritisch würdigen.[383] Auf maschinellem Lernen beruhende Algorithmen beurteilen die Sachverhalte grundsätzlich gestützt auf bereits ergangene und vergleichbare Entscheide. Auf neuartige Fälle, die zu bereits getroffenen Entscheiden keine oder nur vermeintliche Ähnlichkeiten aufweisen, kann ein auf maschinellem Lernen beruhender Algorithmus nicht angewendet werden.[384] Das hat zur Folge, dass die Behörde nur in jenen Fällen eine vollautomatisierte Verfügung erlassen kann, in denen ihr kein Beurteilungs- oder Ermessensspielraum zukommt.[385]

3. Schlussfolgerungen für den Kanton Zürich

Von einer vollautomatisierten Bearbeitung sollte abgesehen werden, wenn ein Ermessens- oder Beurteilungsspielraum besteht oder sich die Behörde mit einem unbestimmten Begriff konfrontiert sieht. Vollautomatisierte Anordnungen können grundsätzlich nur dann ergehen, wenn der Behörde kein Spielraum in der Sachverhaltsbeurteilung zukommt. Ausnahmen sind allerdings denkbar.

Eine solche Ausnahme könnte etwa dann bestehen, wenn der Ermessens- bzw. Beurteilungsspielraum der Behörde in einer Verwaltungsverordnung eingeschränkt wird. Dabei handelt es sich um interne (Dienst-)Anweisungen, die im Besonderen der

369 Waldmann/Wiederkehr, 2019, S. 254.
370 Vgl. Häfelin/Müller/Uhlmann, 2020, Rn. 392.
371 Häfelin/Müller/Uhlmann, 2020, Rn. 396.
372 Häfelin/Müller/Uhlmann, 2020, Rn. 409.
373 Vgl. allgemein zum Begründungserfordernis Kapitel 3 A. II. 1. a).
374 Häfelin/Müller/Uhlmann, 2020, Rn. 411.
375 Häfelin/Müller/Uhlmann, 2020, Rn. 398.
376 Häfelin/Müller/Uhlmann, 2020, Rn. 401.
377 Häfelin/Müller/Uhlmann, 2020, Rn. 403; Waldmann/Wiederkehr, 2019, S. 256. Tatbestandsermessen wird teilweise auch als Beurteilungsspielraum bezeichnet und den unbestimmten Rechtsbegriffen zugeordnet. Vgl. z. B. Tschannen/Zimmerli/Müller, 2014, § 26 Rn. 10 und 27 f. Teilweise wird dafür plädiert, auf eine Unterscheidung zwischen Tatbestandsermessen und Rechtsfolgeermessen zu verzichten; vgl. Schindler, 2010, Rn. 242 ff., 294; Rhinow, 1983, S. 87 ff.
378 Häfelin/Müller/Uhlmann, 2020, Rn. 413 und 415.
379 Schindler, 2010, Rn. 267 ff., 271 f.
380 Braun Binder, 2019a, Rn. 20; Glaser, 2018, S. 188.
381 Rechsteiner, 2018, Rn. 29.
382 Statt vieler Häfelin/Müller/Uhlmann, 2020, Rn. 439 ff.
383 M. w. H. Rechsteiner, 2018, Rn. 29.
384 Rechsteiner, 2018, Rn. 29.
385 Braun Binder, 2016a, Rn. 15; Rechsteiner, 2018, Rn. 30.

Verwirklichung einer sachrichtigen, einheitlichen und gleichmässigen Rechtsanwendung dienen.[386] Diesfalls wäre ein vollautomatisiertes Verfahren durchaus wünschenswert, da standardisierte oder typisierte Entscheidungen der Gleichbehandlung sowie der Transparenz dienen.[387] Ist eine Einzelfallprüfung notwendig, weil eine Standardisierung oder Typisierung nicht möglich ist,[388] kann eine solche Ausnahme allerdings nicht greifen.
Bezüglich einer Teilautomatisierung sind unterschiedliche Konstellationen denkbar. Es stellt sich insbesondere die Frage, über wie viel Spielraum die entscheidende Person bezüglich des Ermessens oder der Beurteilung in einem teilweise automatisierten Verfahren noch verfügt. In diesem Zusammenhang muss der Zeitpunkt beachtet werden, in dem die Automatisierung einsetzt. Im Rahmen des Algorithmeneinsatzes im Vorbereitungsprozess stellen sich hinsichtlich der Ermessensausübung keine rechtlichen Probleme, da eine natürliche Person schliesslich die Entscheidung fällen wird. Zu beachten ist allerdings, dass die Person über die entsprechenden Kenntnisse und Kompetenzen verfügen muss, um gegebenenfalls eine vom Entscheidungsvorschlag abweichende Entscheidung zu treffen.

VI. Verfügungsbegriff

1. Grundlagen

Die Verfügung ist ein zentrales Element des Verwaltungsrechts. Sie erlaubt der öffentlichen Verwaltung, generell-abstrakte Normen auf den Einzelfall anzuwenden und diese somit individuell-konkret umzusetzen.[389] Zugleich liefert sie als Anfechtungsobjekt die notwendige Grundlage dafür, dass Private, die von staatlichem Handeln in rechtsverbindlicher Art und Weise betroffen sind, bei Bedarf dagegen vorgehen können.[390]
Eine Verfügung ist gemäss Art. 5 VwVG eine hoheitliche Anordnung individuell-konkreter Natur, die einseitig erfolgt und eine bestimmte verwaltungsrechtliche Rechtsbeziehung verbindlich regelt.[391] Voraussetzung ist folglich u. a., dass die Anordnung hoheitlich von einem Verwaltungsträger getroffen wird.[392] Mit dem Begriff des Verwaltungsträgers ist jede Stelle gemeint, die im Rahmen ihrer öffentlichen Aufgabenerfüllung Anordnungen erlässt, wodurch auch Private, denen öffentliche Aufgaben übertragen wurden, erfasst sind.
Das VRG basiert auf dem Begriff der «Anordnung»,[393] der jedoch nicht weiter definiert wird. Das Gesetz spricht ausser in zwei Fällen[394] nicht von «Verfügung».[395] Die Anordnung umfasst von der Verwaltungsbehörde erlassene Verfügungen, Entscheidungen und Massnahmen.[396] Sofern es sich um eine erstinstanzliche Anordnung handelt, ist grundsätzlich der Verfügungsbegriff gemäss Art. 5 VwVG einschlägig.[397]

2. Verfügungsbegriff und KI

Setzt der Staat zum Erlass von Verfügungen KI ein, müssen betroffene Personen rechtlich dagegen ebenso vorgehen können wie gegen Verfügungen, die ohne Abstützung auf KI ergehen. Notwendig ist deshalb, dass auch KI-gestützt erlassene Verfügungen als solche im Sinne des Verwaltungsverfahrensrechts qualifiziert werden können.
Stellt man auf den bundesrechtlichen Verfügungsbegriff ab (Art. 5 Abs. 1 VwVG), so ist nicht ohne Weiteres klar, dass KI-gestützt erlassene Entscheide Verfügungen darstellen können.[398] Problematisch ist dies insbesondere dort, wo KI im Rahmen von vollautomatisierten Entscheiden eingesetzt wird.[399] Während Art. 5 Abs. 1 VwVG ein behördliches Handeln voraussetzt, ergeht eine vollautomatisierte Verfügung gerade ohne jegliches menschliche Zutun. Zwar wird der Erlass einer vollautomatisierten Verfügung vom Wortlaut des Art. 5 Abs. 1 VwVG nicht ausgeschlossen. Dennoch ist fraglich, ob eine vollautomatisiert erlassene Entscheidung als Verfügung qualifiziert werden kann, da kein unmittelbares behördliches Handeln im Einzelfall gegeben ist.[400] Dieses ist aber gerade eines der Hauptmerkmale des Verfügungsbegriffs. Ob das behördliche Handeln im Einzelfall durch den Grundsatzentscheid zur Einführung eines vollautomatisierten Verfahrensablaufs kompensiert werden kann, ist zumindest fraglich.[401]

3. Schlussfolgerungen für den Kanton Zürich

Sollte der Kanton Zürich KI im Rahmen von vollautomatisierten Entscheidungen einsetzen, empfiehlt es sich, aus Gründen der Rechtsklarheit und -sicherheit den Anordnungsbegriff im VRG mindestens dahingehend zu klären, dass vollautomatisiert erlassene Verfügungen als Anordnungen zu qualifizieren sind.[402] Dabei wäre bewusst eine technologieneutrale Formulierung vorzusehen. Mithin sollte festgehalten werden, dass auf vollautomatisierte Weise ergangene Entscheide, die den übrigen Kriterien des Verfügungsbegriffs (Art. 5 Abs. 1 VwVG) entspre-

386 Griffel, 2017, Rn. 126.
387 Stelkens, 2018, N. 40 ff. zu § 35a VwVfG.
388 Insbesondere bilden die Grundsätze der Verhältnismässigkeit, der Rechtsgleichheit, des Willkürverbots und die Obliegenheit zur Wahrung des öffentlichen Interesses die Grenze des Ermessensspielraums einer Behörde, vgl. dazu statt vieler Häfelin/Müller/Uhlmann, 2020, Rn. 409.
389 Griffel, 2017, Rn. 25 und 28. Vgl. auch Botschaft VwVG, S. 1362.
390 Rhinow/Koller/Kiss/Thurnherr/Brühl-Moser, 2014, Rn. 1054.
391 Griffel, 2017, Rn. 30 f.; BGE 135 II 38 E. 4.3.
392 Bertschi/Plüss, 2014, N. 19 der Vorbemerkungen zu §§ 4–31 VRG; Griffel, 2017, Rn. 30.
393 Zum Beispiel §§ 5a Abs. 1, 8 Abs. 1, 10a VRG.
394 §§ 8 Abs. 1, 81 lit. a VRG.
395 Im Gegensatz hierzu verwenden das VwVG sowie z. B. das Baselstädtische Gesetz über die Verfassungs- und Verwaltungsrechtspflege (VRPG) vom 14. Juni 1928, SG 270.100, den Verfügungsbegriff.
396 Bertschi/Plüss, 2014, N. 15 der Vorbemerkungen zu §§ 4–31 VRG.
397 Bertschi/Plüss, 2014, N. 18 der Vorbemerkungen zu §§ 4–31 VRG. Der Begriff der Anordnung unterscheidet sich insoweit von der weiten Legaldefinition des Art. 5 VwVG, als er nicht voraussetzt, dass die Anordnung gestützt auf öffentliches Recht des Bundes ergeht.
398 Vgl. auch Glaser, 2018, S. 187.
399 Diese Frage stellt sich nicht nur bei KI-gestützten vollautomatisierten Verfügungen, sondern auch bei vollautomatisiert erlassenen Verfügungen, die keine KI nutzen.
400 Glaser, 2018, S. 187; Braun Binder, 2020b, S. 272; ausführliche Bemerkungen zur Qualifikation einer automatisierten Einzelentscheidung als Verfügung nimmt Egli in ihrer Masterarbeit vom 12. März 2020 vor, S. 18 ff.
401 Braun Binder, 2020b, S. 272, mit Verweisung auf Braun Binder, 2016d, S. 963.
402 Vgl. auch Glaser, 2018, S. 187 f.

chen, unabhängig von der dafür eingesetzten Technologie vom Anordnungsbegriff im Sinne des VRG[403] erfasst werden und damit Rechtskraft entfalten, aber auch angefochten werden können.
Aus Gründen der Rechtsklarheit und -sicherheit könnte im gleichen Zug auch festgehalten werden, dass teilautomatisiert erlassene Verfügungen Anordnungen im Sinne des VRG darstellen. Zwar ist diesbezüglich die Subsumtion unter den Verfügungsbegriff weniger problematisch, da teilautomatisierte Entscheidungen meist ohne Weiteres als Entscheidung einer Behörde eingeordnet werden können. Insoweit würde es sich um eine rein deklaratorische Bestimmung handeln. Eine explizite Verankerung würde allerdings dem Umstand Rechnung tragen, dass die mit KI-Implementierungen einhergehenden potenziellen Risiken, insbesondere die Diskriminierungsgefahr, unabhängig von einer Vollautomation bestehen. Auch mit Blick auf die teilautomatisiert erlassene Anordnung ist deshalb die Möglichkeit einer Anfechtung von zentraler Bedeutung.

VII. Transparenz

Transparenz ist eine im Kontext von KI häufig geäusserte Forderung, um die Rechtmässigkeit des KI-Einsatzes sicherzustellen. Die Transparenz bildet auch in dieser Studie einen zentralen Untersuchungsgegenstand.[404] Aus einer öffentlich-rechtlichen Perspektive soll deshalb an dieser Stelle skizziert werden, wo sich aus dem geltenden Recht Transparenzanforderungen mit Blick auf das Verwaltungshandeln im Kanton Zürich ergeben und welche Schlussfolgerungen sich daraus in Bezug auf den KI-Einsatz ziehen lassen.

1. Grundlagen

a) Transparenz als Folge des rechtlichen Gehörs

Die Forderung nach Transparenz findet sich in verschiedenen Rechtsgrundlagen. So bezweckt die aus dem Anspruch auf rechtliches Gehör (Art. 29 Abs. 2 BV) folgende Begründungspflicht eine Form der Transparenz. Für den Betroffenen muss ersichtlich sein, wie ein Entscheid zustande gekommen ist; mithin wird eine Form der Transparenz[405] behördlichen Handelns hergestellt. Aus dem Anspruch auf vorgängige Äusserung kann zudem eine Herstellung von Transparenz insoweit hergeleitet werden, als die betroffene Person Kenntnis darüber haben muss, dass und worüber ein Verfahren hängig ist, um von ihrem Äusserungsrecht überhaupt Gebrauch machen zu können. Ausserdem kann darunter auch ein Anspruch auf Kenntnis über den Verfahrensablauf sowie über die Vorgänge, die für den Entscheid wesentlich sind, fallen.[406]

b) Transparenz als Folge datenschutzrechtlicher Anforderungen

Transparenz ist zudem ein wichtiger Grundsatz im Datenschutzrecht. Auch wenn das revDSG keine ausdrückliche Transparenzpflicht vorsieht, kommt der Herstellung von Transparenz im Datenschutzrecht sowohl auf Bundesebene als auch auf kantonaler Ebene zentrale Bedeutung zu. Nach Art. 6 Abs. 3 revDSG dürfen Personendaten nur zu einem bestimmten und für die betroffene Person erkennbaren Zweck bearbeitet werden; die gesetzliche Regelung der Erkennbarkeit des Zwecks der Datenbearbeitung setzt dabei zwingend voraus, dass auch die Datenbearbeitung als solche erkennbar ist. Nach Art. 6 Abs. 2 revDSG muss die Bearbeitung von Personendaten zudem nach Treu und Glauben erfolgen. Aus diesem Grundsatz wird schon im DSG derjenige der Transparenz abgeleitet;[407] dies gilt auch für das revDSG.[408] Weiterhin zeigen zentrale Bestimmungen des revidierten Datenschutzgesetzes mit aller Deutlichkeit, dass das Gesetz auf dem Grundsatz der Transparenz beruht, so namentlich die Regelung der Informationspflichten (Art. 19 ff. revDSG) und des Auskunftsrechts (Art. 25 ff. revDSG). Auch im kantonalen Datenschutzrecht kommt die Bedeutung der Transparenz zum Ausdruck, insbesondere in der Regelung der datenschutzrechtlichen Informationspflicht (§ 12 IDG).

c) Transparenz als Folge der behördlichen Informationspflicht bzw. des Öffentlichkeitsprinzips

Art. 180 Abs. 2 BV verpflichtet den Bundesrat, die Öffentlichkeit rechtzeitig und umfassend über seine Tätigkeit zu informieren, soweit nicht überwiegende öffentliche oder private Interessen entgegenstehen. Der Bundesrat ist zur Information verpflichtet und kann nur ausnahmsweise davon absehen. Das BGÖ[409] regelt den Zugang zu amtlichen Dokumenten von Verwaltungsstellen des Bundes (Art. 2 f.). Dabei gilt das allgemeine Öffentlichkeitsprinzip mit Geheimhaltungsvorbehalt.
Art. 49 KV ZH verpflichtet die Behörden, von sich aus und auf Anfrage die Öffentlichkeit über ihre Tätigkeit zu informieren, soweit nicht öffentliche oder private Interessen entgegenstehen. Art. 17 KV ZH gewährleistet jeder Person das Recht auf Zugang zu amtlichen Dokumenten, soweit dem nicht überwiegende öffentliche oder private Interessen widersprechen. Damit wurde im Kanton Zürich ebenfalls das Öffentlichkeitsprinzip mit Geheimhaltungsvorbehalt verankert.[410] Konkretisiert wird dieses Recht im IDG. Ziel ist u. a., die freie Meinungsbildung und die Wahrnehmung der demokratischen Rechte zu fördern sowie die Kontrolle staatlichen Handelns zu erleichtern (§ 1 Abs. 2 lit. a IDG). Gemäss § 14 IDG informieren die Behörden von sich aus über ihre Tätigkeit, soweit dies von allgemeinem Interesse ist. Dabei handelt es sich um eine Informationstätigkeit von Amtes wegen; die Behörden sind folglich zur aktiven Information verpflichtet.[411] § 20 IDG sieht sodann vor, dass Behörden auf Anfrage Einsicht in amtliche Akten gewähren, sofern dem nicht im Einzelfall eine rechtliche Bestimmung oder ein überwiegendes öffentliches oder privates Interesse entgegensteht (§ 23 IDG).

403 Eine solche Regelung würde allerdings auch voraussetzen, dass der Anordnungsbegriff im VRG grundsätzlich definiert wird.
404 Vgl. Kapitel 4 bzw. die dort erwähnten Richtlinien.
405 Vgl. Kapitel 3 A. VII. 1. a).
406 Häfelin/Müller/Uhlmann, 2020, Rn. 1010a; BGE 144 I 11 E. 5.3.
407 Vgl. Botschaft DSG 1988, BBl 1988 413, S. 449; Epiney, 2011, Rn. 22.
408 Rosenthal, 2020, Rn. 15.
409 Bundesgesetz über das Öffentlichkeitsprinzip der Verwaltung (BGÖ) vom 17. Dezember 2004, SR 152.3.
410 Vgl. Jaag/Rüssli, 2019, Rn. 1008.
411 Vgl. Jaag/Rüssli, 2019, Rn. 1008a.

2. Transparenz und KI

In Diskussionen rund um den staatlichen Einsatz von KI tritt regelmässig die Forderung nach Herstellung von Transparenz auf. Häufig wird damit das Anliegen verbunden, den staatlichen Einsatz von KI kontrollieren zu können.[412]

Aus rechtlicher Sicht bestehen, wie soeben skizziert, verschiedene Anknüpfungspunkte, auf die sich die Herstellung von Transparenz stützen lässt. Während die unter dem Aspekt der Begründungspflicht[413] und der datenschutzrechtlichen Informationspflicht[414] umschriebenen Ansatzpunkte für die Herstellung von Transparenz der individuellen Kontrolle einzelner Entscheide dienen, kann die Herstellung von Transparenz auch eine allgemeine Kontrolle – etwa durch zivilgesellschaftliche Akteure – von KI-Anwendungen bezwecken. Das grosse Diskriminierungspotenzial von KI-Systemen kann eine solche allgemeine Kontrolle von staatlichen KI-Systemen durchaus als gerechtfertigt erscheinen lassen.

3. Schlussfolgerungen für den Kanton Zürich

Die Herstellung von Transparenz beim staatlichen KI-Einsatz ist nicht nur aus der Perspektive individueller Kontrollmöglichkeiten von Einzelentscheiden, sondern auch mit Blick auf eine allgemeine Kontrolle zu diskutieren. Dabei sind verschiedene Ansatzpunkte vorstellbar, wie die Transparenz zur Ermöglichung von Kontrolle rechtlich konkretisiert werden könnte.

In der TA-SWISS-Studie wird z. B. die Schaffung eines öffentlich zugänglichen Registers vorgeschlagen, aus dem ersichtlich wird, in welchen Bereichen die öffentliche Verwaltung KI-Systeme einsetzt.[415] Ein solches Register liesse sich zugleich auf datenschutzrechtliche Erwägungen stützen[416] und sollte u. a. Auskunft über die Art und Herkunft der bearbeiteten Sach- und Personendaten, die Rechtsgrundlage, den Zweck und die Mittel der Bearbeitung, das verantwortliche Organ, die Beschreibung und Erklärung der verwendeten KI-Anwendung und deren Logik sowie diejenigen Akteure, die an der Entwicklung des Systems mitgewirkt haben, geben.

Für die gesetzliche Verankerung einer solchen Pflicht zur Führung eines Registers bietet sich das IDG an. § 14 Abs. 4 IDG sieht vor, dass das öffentliche Organ ein Verzeichnis seiner Informationsbestände und deren Zweck öffentlich zugänglich macht. Weiter muss das öffentliche Organ jene Informationsbestände kennzeichnen, die Personendaten enthalten. Von der Verzeichnispflicht sind sowohl Informationsbestände mit als auch ohne Personendaten erfasst.[417] Die Bestimmung könnte um einen weiteren Absatz ergänzt werden, in dem die rechtliche Grundlage für ein solches öffentliches Register für staatliche KI-Anwendungen verankert wird.

Eine weitere mögliche Herangehensweise zur Herstellung von Transparenz findet sich in dieser Studie in Kapitel 4. Zu diskutieren wäre, wie und wo die Herstellung von Transparenz mittels der in Kapitel 4 vorgeschlagenen Checklisten bzw. der Erstellung eines Transparenzberichts rechtlich zu verankern wäre.

B. Beispielhafte Einsatzbereiche und Anwendungen

Im vorangegangenen Kapitel wurden allgemeine (verfassungs-)rechtliche Herausforderungen eines staatlichen KI-Einsatzes erläutert. Die konkrete rechtliche Ausgestaltung des KI-Einsatzes hängt jedoch zentral davon ab, in welchem Fachbereich KI implementiert werden soll und welche genaue Anwendung dabei vorgesehen ist. Im Folgenden werden deshalb einerseits die bereichsspezifischen Herausforderungen von zwei Verwaltungsbereichen vorgestellt, in denen KI-Anwendungen im Kanton Zürich als besonders wahrscheinlich betrachtet werden können: das Steuer- und das Sozialversicherungsverfahren (I. und II.). Anderseits werden die konkreten rechtlichen Rahmenbedingungen von Chatbots näher beleuchtet, da diese eine der zurzeit am weitesten verbreiteten KI-Anwendungen in der öffentlichen Verwaltung darstellen (III.).

I. Steuerverfahren

1. Steuerrechtssystem in der Schweiz

Das schweizerische Steuerrechtssystem ist durch eine hohe Steuerautonomie der Kantone geprägt. Soweit die Bundesverfassung dem Bund das Recht zur Erhebung von bestimmten Steuern nicht ausdrücklich zuweist, verfügen die Kantone über das volle Besteuerungsrecht. Die wenigen Besteuerbefugnisse des Bundes umfassen einerseits das Recht, neben den Kantonen direkte Steuern auf dem Einkommen der natürlichen Personen und auf dem Reinertrag von juristischen Personen zu erheben (Art. 128 BV)[418]. Anderseits sind bestimmte indirekte Steuerarten ausschliesslich dem Bund vorbehalten. Dazu zählen die Mehrwertsteuer (Art. 130 BV)[419], die besonderen Ver-

412 Vgl. dazu die Ausführungen in Kapitel 4.
413 Kapitel 3 A. II. 2. b) und 4. b).
414 Kapitel 3 A. IV. 1. b).
415 Vgl. TA-SWISS KI, 2020, S. 294, Empfehlung 3. Vgl. auch Automating Society Report 2020, S. 12
416 Vgl. z. B. Art. 18 Gesetz über den Datenschutz Kanton Appenzell A. Rh. vom 18.06.2001, bGS 146.1; § 22 Gesetz über die Information der Öffentlichkeit, den Datenschutz und das Archivwesen Kanton Aargau vom 24.10.2006, SAR 150.700; § 24 Gesetz über die Information und den Datenschutz Basel-Stadt vom 09.06.2010, SG 153.260.
417 Fey, 2012, N. 33 zu § 14 IDG.
418 Vgl. zur Bundesgesetzgebung das Bundesgesetz über die direkte Bundessteuer (DBG) vom 14. Dezember 1990, SR 642.11, sowie die entsprechenden Verordnungen.
419 Vgl. zur Bundesgesetzgebung das Bundesgesetz über die Mehrwertsteuer (MWSTG) vom 12. Juni 2009, SR 641.20, sowie die entsprechenden Verordnungen.

brauchssteuern (Art. 131 BV)[420], die eidgenössische Stempel- und Verrechnungssteuer (Art. 132 BV)[421] und die Zölle (Art. 133 BV)[422]. Hinzu kommen bestimmte für spezielle Zwecke erhobene Bundeseinnahmen wie die Schwerverkehrsabgabe (Art. 85 BV)[423], die Nationalstrassenabgabe (Art. 86 Abs. 2 BV)[424], die Spielbankenabgabe (Art. 160 Abs. 3 BV)[425] sowie der Wehrpflichtersatz (Art. 59 Abs. 3 BV)[426].

Neben diesen Bundessteuern dürfen die Kantone ihr Steuersystem grundsätzlich frei ausgestalten. Eingeschränkt werden sie dabei lediglich von den Grundsätzen der Besteuerung (Art. 127 BV), den Vorgaben zur Harmonisierung im Bereich der direkten Steuern (Art. 129 BV)[427] und dem Ausschluss der Erhebung gleichartiger Steuern (Art. 134 BV). Diese Freiheit der Kantone hat eine weitgehende Verschiedenheit der kantonalen Steuergesetzgebungen zur Folge.[428] Im Kanton Zürich sind insbesondere die folgenden Rechtsgrundlagen wesentlich: das Steuergesetz[429], die Verordnung zum Steuergesetz[430], die Quellensteuerverordnungen I[431] und II[432], das Erbschafts- und Schenkungssteuergesetz[433], die Verordnung über die Durchführung des Bundesgesetzes über die direkte Bundessteuer[434], die Verordnung über die Rückerstattung der Verrechnungssteuer[435] sowie die Verordnung über die elektronische Einrichtung der Steuererklärung[436]. Die Kantone können schliesslich wiederum ihre Gemeinden ermächtigen, Steuern zu erheben. Dabei kann es sich entweder um eigene kommunale Steuern oder um Zuschläge zu den geschuldeten Kantonssteuern handeln.[437] Die Besteuerung erfolgt in der Schweiz somit auf Bundes-, Kantons- und Gemeindeebene.[438]

2. Steuerveranlagungsverfahren

Die Steuererhebung durch Bund, Kantone und Gemeinden setzt die Feststellung der individuell-konkreten Steuerschuld der steuerpflichtigen Personen voraus. Dieser Vorgang der Rechtsanwendung wird Steuerveranlagungs- oder Einschätzungsverfahren genannt.[439] Für die Veranlagung der direkten Bundessteuer sowie der Staats- und Gemeindesteuern ist in der Regel der Wohnsitzkanton zuständig, die Veranlagung der Verrechnungssteuer sowie der indirekten Steuern des Bundes obliegt dagegen den Steuerbehörden des Bundes.[440]

Die Veranlagungsverfahren für die unterschiedlichen Steuerarten können u. a. aufgrund der Mitwirkung der betroffenen Personen unterschieden werden.[441] Beim ordentlichen Veranlagungsverfahren wirken die steuerpflichtige Person und das amtliche Organ gemeinsam an der Ermittlung und Feststellung der Steuerforderung mit. Die Steuerpflichtigen haben so die Pflicht zur Selbstdeklaration, während die Steuerbehörden die massgebenden Tatsachen überprüfen (Veranlagungsverfahren i. e. S.) und den Steuerbetrag in der Veranlagungsverfügung festlegen, womit das Verfahren abgeschlossen wird.[442] Beide nehmen eigentliche Veranlagungshandlungen vor, weshalb auch von einem gemischten Veranlagungsverfahren die Rede ist. Hauptanwendungsfälle des ordentlichen Veranlagungsverfahrens sind die direkten Steuern auf Einkommen, Vermögen, Gewinn und Kapital.[443]

Im besonderen Veranlagungsverfahren verschiebt sich die Mitwirkungspflicht stärker auf eine Seite. Bei der Selbstveranlagung ist allein die steuerpflichtige Person für die Feststellung und Meldung der Steuerforderung verantwortlich. Im Unterschied zur Selbstdeklarationspflicht im ordentlichen Verfahren, nennt die steuerpflichtige Person nicht nur die Bemessungsgrundlagen, sondern meldet sich aufgrund der Steuerpflicht selbst an, berechnet die Steuern selbstständig und zahlt den Steuerbetrag unaufgefordert ein. Die zuständigen Steuerbehörden nehmen einzig Kontrollen und unter Umständen Berichtigungen vor. Die Veranlagungsverfügung entfällt daher in der Regel, sofern keine Berichtigungen oder Nachforderungen

420 Die besonderen Verbrauchssteuern des Bundes umfassen insbesondere die Tabaksteuer (Art. 131 Abs. 1 Bst. a BV), die Alkoholsteuer (Art. 131 Abs. 1 Bst. b BV), die Biersteuer (Art. 131 Abs. 1 Bst. c BV), die Automobilsteuer (Art. 131 Abs. 1 Bst. d BV), die Mineralölsteuer (Art. 131 Abs. 1 Bst. e BV). Vgl. zur Bundesgesetzgebung das Tabaksteuergesetz, das Bundesgesetz über die gebrannten Wasser (Alkoholgesetz, AlkG) vom 21. Januar 1932, SR 680, das Biersteuergesetz, das Automobilsteuergesetz (AStG) vom 21. Juni 1996, SR 641.51, das Mineralölsteuergesetz (MinöStG) vom 21. Juni 1996, SR 641.61, sowie die jeweils entsprechenden Verordnungen.

421 Vgl. zur Bundesgesetzgebung Bundesgesetz über die Stempelabgaben (StG) vom 27. Juni 1973, SR 641.10, und Bundesgesetz über die Verrechnungssteuer (Verrechnungssteuergesetz, VStG) vom 13. Oktober 1965, SR 642.21, sowie die entsprechenden Verordnungen.

422 Vgl. zur Bundesgesetzgebung Zollgesetz (ZG) vom 18. März 2005, SR 631.0, und Zolltarifgesetz (ZTG) vom 9. Oktober 1986, SR 632.10, sowie die entsprechenden Verordnungen.

423 Vgl. zur Bundesgesetzgebung Bundesgesetz über eine leistungsabhängige Schwerverkehrsabgabe (Schwerverkehrsabgabegesetz, SVAG) vom 19. Dezember 1997, SR 641.81, sowie die entsprechenden Verordnungen.

424 Vgl. zur Bundesgesetzgebung Bundesgesetz über die Abgabe für die Benützung von Nationalstrassen (Nationalstrassenabgabegesetz, NSAG) vom 19. März 2010, SR 741.71, sowie die entsprechenden Verordnungen.

425 Vgl. zur Bundesgesetzgebung Bundesgesetz über Geldspiele (Geldspielgesetz, BGS) vom 29. September 2017), SR 935.51, sowie die entsprechenden Verordnungen.

426 Vgl. zur Bundesgesetzgebung Bundesgesetz über die Wehrpflichtersatzgabe (WPEG) vom 12. Juni 1959), SR 661 sowie die entsprechende Verordnung.

427 Vgl. dazu auch Bundesgesetz über die Harmonisierung der direkten Steuern der Kantone und Gemeinden (Steuerharmonisierungsgesetz, StHG) vom 14. Dezember 1990, SR 642.14.

428 Vgl. z. B. Blumenstein/Locher, 2016, S. 60 ff.

429 Steuergesetz (StG) vom 8. Juni 1997, LS 631.1.

430 Verordnung zum Steuergesetz (StV) vom 1. April 1998, LS 631.11.

431 Verordnung über die Quellensteuer für ausländische Arbeitnehmer (Quellensteuerverordnung I) vom 2. Februar 1994, LS 631.41.

432 Verordnung über die Quellensteuer für natürliche und juristische Personen ohne steuerrechtlichen Wohnsitz oder Aufenthalt in der Schweiz (Quellensteuerverordnung II) vom 2. Februar 1994, LS 631.42.

433 Erbschafts- und Schenkungssteuergesetz (ESchG) vom 28. September 1986, LS 632.1.

434 Verordnung über die Durchführung des Bundesgesetzes über die direkte Bundessteuer vom 4. November 1998, LS 634.1.

435 Verordnung über die Rückerstattung der Verrechnungssteuer vom 17. Dezember 1997, LS 634.2.

436 Verordnung über die elektronische Einreichung der Steuererklärung vom 18. Oktober 2011, LS 631.121. Vgl. auch Verordnung über die elektronische Zustellung von Verfügungen und Rechnungen vom 7. September 2012, LS 631.122.

437 Zum Beispiel Blumenstein/Locher, 2016, S. 62 ff.

438 Vgl. zu den gesamten allgemeinen Ausführungen z. B. Blumenstein/Locher, 2016, S. 53 ff.

439 Blumenstein/Locher, 2016, S. 487; Zweifel/Casanova/Beusch/Hunziker, 2018, § 3 Rn. 1.

440 Reich, 2020, § 26 Rn. 11 f.

441 Blumenstein/Locher, 2016, S. 488.

442 Meier-Mazzucato, 2015, S. 943; Reich, 2020, § 26 Rn. 52 ff.

443 Blumenstein/Locher, 2016, S. 507.

erfolgen. Daher kommt die Selbstveranlagung nur zur Anwendung, wenn es dem Gemeinwesen unmöglich oder nur mit unverhältnismässigem Aufwand möglich wäre, den Sachverhalt zu ermitteln. Dies ist beispielsweise bei der Mehrwertsteuer ohne Einfuhrsteuer, den Stempelabgaben oder der Quellensteuer der Fall.[444] Bei der amtlichen Veranlagung verhält es sich gerade umgekehrt. Die Steuerforderung wird durch die zuständige Behörde grundsätzlich ohne Mitwirkung der steuerpflichtigen Person aufgrund von amtlichen Tatsachen festgestellt, die der Veranlagungsbehörde bereits bekannt sind. Dies ist etwa bei der Hundesteuer, der Motorfahrzeugsteuer oder bei der Handänderungssteuer auf Grundstücken der Fall.[445]
Ebenfalls zum besonderen Verwaltungsverfahren gehören die Zollveranlagungen, denn hier stehen nicht die fiskalischen Aspekte, sondern die Überwachung und Kontrolle des Personen- und Warenverkehrs an erster Stelle, was sich auf das Verfahren auswirkt. Die Mitwirkungspflicht ist jedoch mit dem gemischten Veranlagungsverfahren vergleichbar.[446] So gelten insbesondere die spezialgesetzlichen Bestimmungen des Zollverfahrensrechts, welche durch das Prinzip der Selbstdeklaration geprägt sind.[447] Gemäss diesem Selbstdeklarationsprinzip trägt die anmeldepflichtige Person die volle Verantwortung für eine ordnungsgemässe Zollanmeldung.[448]

3. KI im Veranlagungsverfahren

Seit Beginn der Diskussion um den staatlichen KI-Einsatz war klar, dass dieser besonders in Bereichen infrage kommt, in denen es vorwiegend um gebundene Entscheide wie Berechnungen geht.[449] Das steuerrechtliche Veranlagungsverfahren stellt diesbezüglich einen geradezu idealen Verwaltungsbereich für die Nutzung von Automatisierung und KI dar. Darüber hinaus handelt es sich beim Veranlagungsverfahren um ein klassisches Verfahren aus dem Bereich der Massenverwaltung; der Sachverhalt kann grösstenteils in standardisierter Form ermittelt werden und wird aufgrund des Selbstdeklarationsprinzips weitgehend von der steuerpflichtigen Person selbst erfasst (ausser bei der amtlichen Veranlagung). Zudem liegen aufgrund der jährlichen bzw. regelmässigen Durchführung des Veranlagungsverfahrens historische und vergleichbare Daten vor, die zur Kontrolle herangezogen werden können. Dies sind Eigenschaften, die einen Einsatz von KI und Automatisierung begünstigen.[450] Konkret ist der staatliche KI-Einsatz im Steuerbereich an zwei Stellen vorstellbar: erstens beim Veranlagungsverfahren i. e. S. und zweitens bei der Überprüfung der Veranlagungen (z. B. in Form von Risikomanagementsystemen).
Dank der teil- oder vollautomatisierten Verfahren können die Steuer- und Abgabebeträge ohne menschliche Bearbeitung festgesetzt werden. Insbesondere die Kantone arbeiten zurzeit an unterschiedlichen KI-Projekten, um die Veranlagungsverfahren zu automatisieren, und auch der Bund hat mit dem revidierten Datenschutzgesetz erste Rechtsgrundlagen für eine Automation des Zollveranlagungsverfahrens geschaffen.[451] Mit dieser Automatisierung wird allerdings eine noch stärkere Verlagerung der Sachverhaltsermittlung auf die steuerpflichtige Person einhergehen, als dies das Prinzip der Selbstdeklaration ohnehin vorsieht.[452] Diese Abstriche am Untersuchungsgrundsatz müssen kompensiert werden, wobei dies mit Risikomanagementsystemen, dem zweiten Anwendungsfall von KI-Systemen im Steuerbereich, erfolgen könnte.[453] Risikomanagementsysteme können z. B. so eingesetzt werden, dass sie Plausibilitätsprüfungen durchführen und risikobehaftete Fälle einer umfassenden Prüfung durch Verwaltungsmitarbeitende zuführen. Im Zuge der Modernisierung des deutschen Besteuerungsverfahrens wurden etwa die Rechtsgrundlagen für den Einsatz solcher Risikomanagementsysteme geschaffen. Über die genaue Funktionsweise dieser in Deutschland eingesetzten Risikomanagementsysteme ist jedoch nur sehr wenig bekannt.[454]

4. Rechtliche Rahmenbedingungen

An dieser Stelle sollen bestimmte rechtliche Rahmenbedingungen im Hinblick auf das Steuerverfahren konkretisiert werden: das Legalitätsprinzip, die Begründungspflicht sowie das Zusammenspiel von Untersuchungsgrundsatz und Mitwirkungspflicht.

a) Legalitätsprinzip

Aus dem Legalitätsprinzip folgt, dass die Staats- und Verwaltungstätigkeit an das Recht gebunden ist. Alle wichtigen rechtsetzenden Bestimmungen sind dabei in der Form eines Gesetzes im formellen Sinn zu erlassen; hierzu gehört insbesondere die Einschränkung von verfassungsmässigen Rechten (Art. 38 Abs. 1 lit. b KV ZH).[455]
Für den vollautomatisierten Erlass von Verfügungen bzw. Anordnungen ist eine formell-gesetzliche Grundlage notwendig, da er zu Einschränkungen im Bereich der Verfahrensgrundrechte führt.[456] Sollten im Kanton Zürich Einschätzungsentscheide[457] zukünftig vollautomatisiert erlassen werden können, müsste dies im Steuergesetz ausdrücklich vorgesehen werden.[458] Insbesondere genügen die Rechtsgrundlagen in der Verordnung über die elektronische Zustellung von Verfügungen und Rechnungen[459], wo Einschätzungsentscheide und Veranlagungsverfügungen explizit genannt werden (§ 1 Abs. 1 lit. a), nicht, um vollautomatisierte Einschätzungsentscheide zu erlassen. Zum einen stellt die Verordnung keine formell-gesetzliche Grundlage dar. Zum anderen bezieht sie sich lediglich auf den elektronischen Schriftverkehr zwischen Parteien und Behörden und betrifft für die Verfügungen nur die Frage nach der Zustellungsform.

444 Meier-Mazzucato, 2015, S. 940 f.; Zweifel/Casanova/Beusch/Hunziker, 2018, § 3 Rn. 4.
445 Blumenstein/Locher, 2016, S. 488 f.
446 Blumenstein/Locher, 2016, S. 529 ff.
447 Braun Binder, 2020b, S. 261; Braun Binder, 2020c, S. 34 und 38.
448 Siehe Art. 21, 25 und 26 Zollgesetz (ZG) vom 18. März 2005, SR 631.
449 M. w. H. Braun Binder, 2020c, S. 28.
450 Braun Binder, 2020c, S. 34.
451 Vgl. dazu Kapitel 2 F. II. 1.
452 Vgl. dazu Kapitel 3 B. I. 2.
453 Vgl. zu den Ausführungen zu den Risikomanagementsystemen Braun Binder, 2020c, S. 30.
454 Vgl. dazu ausführlich Braun Binder, 2020d.
455 Vgl. dazu Kapitel 3 A. I.
456 Vgl. dazu Kapitel 3 A. II.
457 Im Kanton Zürich wird die Veranlagungsverfügung betreffend kantonale Steuern wie in vielen anderen Kantonen Einschätzungsentscheid genannt.
458 Für die Notwendigkeit einer formell-gesetzlichen Grundlage für vollautomatisierte Verfügungen auf Bundesebene Braun Binder, 2020b, S. 271 f.; im Ergebnis ebenso Glaser, 2018, S. 188, und Rechsteiner, 2018, Rn. 15.
459 Verordnung über die elektronische Zustellung von Verfügungen und Rechnungen vom 7. September 2012, LS 631.122.

b) Begründungspflicht

Veranlagungsverfügungen sind grundsätzlich wie «normale» Verfügungen zu begründen,[460] denn aus dem Anspruch auf rechtliches Gehör (Art. 29 Abs. 2 BV) folgt auch die grundsätzliche Pflicht der Steuerbehörden, ihre Entscheide zu begründen.[461] Für die Massenverwaltung, zu der Steuerveranlagungen gehören, sind jedoch herabgesetzte Anforderungen an diese Begründungspflicht möglich.[462]

Im Bereich der direkten Bundessteuern müssen gemäss Art. 131 Abs. 2 DBG[463] Abweichungen von der Steuererklärung nur bekannt gegeben, nicht aber begründet werden.[464] Die steuerpflichtige Person muss im Sinne einer Mitteilungspflicht folglich nur auf die quantitativen Abweichungen aufmerksam gemacht werden. Eine qualitative Begründung der Abweichungen ist hingegen nicht vorgesehen.[465] Der Gesetzgeber hat auf eine weitergehende Begründungspflicht für Veranlagungsverfügungen aus verfahrensökonomischen Gründen verzichtet.[466] Nach dem Bundesgericht gilt eine Veranlagungsverfügung somit von Gesetzes wegen als begründet, wenn die Abweichungen von der Steuererklärung bekannt gegeben werden.[467] Die Lehre beurteilt diese herabgesetzte Begründungspflicht von Steuerveranlagungen im Lichte von Art. 29 Abs. 2 BV allerdings teilweise kritisch und fordert zumindest eine minimale Begründungspflicht. Diese sei notwendig, damit die geänderten Steuerfaktoren für die steuerpflichtige Person erkennbar seien und sie sich gegen die Gründe, die zur Abweichung von den gelieferten Angaben geführt haben, wehren könne. Dabei würden jedoch eine standardisierte Begründung unter Angabe der Faktorenbestandteile und deren Betrages sowie eine zusammenfassende Darstellung der Motive für die Abweichung genügen.[468] Um eine solche Begründungspflicht für die direkten Bundessteuern jedoch zu bejahen, müsste Art. 131 Abs. 2 DBG vom Gesetzgeber angepasst werden. Eine Interpretation und Auslegung der aktuellen Bestimmung als Begründungspflicht erscheint aufgrund von Art. 190 BV problematisch. Allerdings ist es für die Steuerbehörden bereits heute möglich, in ihrer Praxis weiter als die gesetzlichen Minimalforderungen zu gehen.[469]

Für die direkten Steuern der Kantone und Gemeinden hält Art. 41 Abs. 3 StHG[470] in verbindlicher Weise fest, dass Veranlagungsverfügungen, die eine Rechtsmittelbelehrung enthalten müssen, der steuerpflichtigen Person schriftlich zu eröffnen sind. «Andere Verfügungen und Entscheide sind ausserdem zu begründen» (Art. 41 Abs. 3 Satz 2 StHG). Aus dieser Formulierung könnte geschlossen werden, dass Veranlagungsverfügungen überhaupt nicht begründet werden müssen. Dies stellt jedoch nur eine missverständliche Formulierung dar. Gemäss Art. 46 Abs. 2 StHG muss die Steuerbehörde der steuerpflichtigen Person Abweichungen von der Steuererklärung spätestens bei der Eröffnung der Veranlagungsverfügung bekannt geben.[471] Zumindest eine Mitteilungspflicht besteht somit auch für die kantonalen und kommunalen direkten Steuern. Wie diese Mitteilungspflicht umgesetzt wird und ob über die Minimalforderungen des DBG und StHG zu einer Begründungspflicht hinausgegangen werden muss, ist in den Kantonen jeweils unterschiedlich geregelt. Einige kantonale Steuergesetze lassen die Abweichung von der Steuererklärung in Anlehnung an das Bundesrecht als Mitteilungspflicht genügen (z. B. § 159 Steuergesetz BS)[472], andere verlangen, dass die Abweichungen von der Steuererklärung im Einzelnen anzugeben und zu begründen sind (z. B. Art. 205 Abs. 2 Steuergesetz Obwalden)[473].

Das Steuergesetz des Kantons Zürich[474] sieht in § 126 Abs. 1 eine allgemeine Begründungspflicht für Veranlagungsverfügungen vor.[475] Eine ausführliche schriftliche Begründung ist aus § 126 Abs. 1 StG jedoch nicht abzuleiten. Die Begründung muss der steuerpflichtigen Person allerdings zumindest erlauben, die Veranlagung sachgerecht anzufechten. Die konkreten Anforderungen an diese Begründung hängen von den Umständen des Einzelfalls ab. Je grösser der Ermessensspielraum für die Behörde ist und je stärker ein Entscheid in die individuellen Rechte eingreift, desto ausführlicher ist dieser zu begründen.[476] § 126 Abs. 1 Satz 2 StG hält für Einschätzungsentscheide fest, dass die Abweichungen von der Steuererklärung bekannt gegeben werden müssen. Ob diese Formulierung bedeutet, dass in Einschätzungsentscheiden *nur* die Abweichungen bekannt gegeben und – wie Art. 131 Abs. 2 DBG dies vorsieht – überhaupt nicht begründet werden müssen, ist jedoch unklar.[477] Die Gesetzesmaterialien des StG sprechen für eine Mitteilungs- und gegen eine Begründungspflicht. § 126 Abs. 1 Satz 2 StG wurde im Jahr 2005 explizit zur Anpassung des kantonalen Steuerrechts an das Harmonisierungsrecht eingeführt. Mit § 126 Abs. 1 Satz 1 StG sei das kantonale Steuerrecht über den durch das StHG und DBG gebotenen Rahmen hinausgegangen, da diese keine Begründungspflicht für Einschätzungsentscheide bzw. Veranlagungsverfügungen vorsehen. Um dies zu korrigieren, hat der Regierungsrat dem Kantonsrat Zürich 2004 vorgeschlagen, Satz 2 zu § 126 Abs. 1 StG hinzuzufügen, sodass auch im Kanton Zürich nur noch eine Mitteilungspflicht besteht.[478] In der Literatur finden sich hingegen auch Stimmen, die dafür plädieren, dass neben der betragsmässigen Abweichung auch kurz deren Gründe mitgeteilt werden müssen, um der Begründungspflicht zu entsprechen. Die Begründung könne sich dabei aber auch bereits aus dem vorherigen Einschät-

460 Vgl. zu den allgemeinen Ausführungen zum Recht auf Begründung Kapitel 3 A. II. 2. b).
461 Reich, 2020, § 26 Rn. 57.
462 Kneubühler/Pedretti, 2019, N. 19 zu Art. 18 VwVG.
463 Bundesgesetz über die direkte Bundessteuer (DBG) vom 14. Dezember 1990, SR 642.11.
464 Botschaft Steuerharmonisierung, S. 133.
465 Locher, 2015, N. 10 zu Art. 131 DBG.
466 Botschaft Steuerharmonisierung, S. 133.
467 BGer 2C_596/2012 vom 19.03.2013, E. 4.1.
468 Vgl. dazu Albertini, 2000, S. 419 f.; Althaus-Houriet, N. 10 zu Art. 131 DBG; Wiederkehr, 2010, S. 495 f.; Zweifel, 2017, N. 9 zu Art. 131 DBG; Zweifel/Hunziker, 2008/2009, S. 663.
469 Locher, 2015, N. 12 zu Art. 131 DBG.
470 Bundesgesetz über die Harmonisierung der direkten Steuern der Kantone und Gemeinden (Steuerharmonisierungsgesetz, StHG) vom 14. Dezember 1990, SR 642.14.
471 Vgl. Richner/Frei/Kaufmann/Meuter, 2013, N. 1 zu § 126 StG.
472 Gesetz über die direkten Steuern vom 12. April 2000, SG 640.100.
473 Steuergesetz vom 30. Oktober 1994, GDB 641.4.
474 Steuergesetz (StG) vom 8. Juni 1997, LS 631.1.
475 Richner/Frei/Kaufmann/Meuter, 2013, N. 1 zu § 126 StG und Rn. 31 zu § 139 StG.
476 Richner/Frei/Kaufmann/Meuter, 2013, N. 31 zu § 139 StG m. w. H.
477 Richner/Frei/Kaufmann/Meuter, 2013, N. 35 zu § 139 StG.
478 Vgl. dazu ABl ZH 2004, S. 826.

zungsverfahren ergeben.[479] Folglich ist im Kanton Zürich umstritten, ob Abweichungen in Einschätzungsentscheiden bzw. Veranlagungsverfügungen gegenüber der Steuererklärung nur mittgeteilt oder auch begründet werden müssen.
Erlässt ein automatisiertes KI-System, welches allenfalls sogar auf schwer oder nicht nachvollziehbaren maschinellen Lernverfahrensmethoden basiert, die Veranlagungsverfügung oder bereitet zumindest wesentliche Teile davon vor, ist fraglich, wie einer Begründungspflicht entsprochen werden kann.[480] Die aktuell im Einsatz stehenden Systeme sind jedenfalls (noch) nicht in der Lage, Veranlagungsentscheide ohne menschliches Zutun zu begründen.[481]
Während dies für die direkten Bundessteuern und bestimmte kantonale sowie kommunale direkten Steuern zurzeit zwar rechtmässig ist, ist fraglich, ob dies im Hinblick auf die zunehmende Automatisierung bzw. den KI-Einsatz so bleiben kann. Spätestens dann, wenn die Verwaltungsbehörden selbst die Entscheidmotive aufgrund weitgehender Automatisierung bzw. des Einsatzes von KI nur noch schwer nachvollziehen können, ist es geboten, bestimmte Eckpunkte der automatisierten Entscheidung kenntlich zu machen.
Bei der Konzeption von KI-Anwendungen für das zürcherische Einschätzungsverfahren ist es somit empfehlenswert, die Begründungsmöglichkeit von Anfang an einzuplanen, um den Anforderungen des rechtlichen Gehörs zu genügen.[482]

c) Zusammenspiel Untersuchungsgrundsatz und Mitwirkungspflicht

Gemäss dem Untersuchungsgrundsatz müssen Verwaltungsbehörden den Sachverhalt von Amtes wegen abklären.[483] Die Entscheidungsgrundlagen müssen somit grundsätzlich vom Entscheidungsträger beschafft werden.[484] Der Untersuchungsgrundsatz gilt auch für die Steuerbehörden und hier für alle Veranlagungsverfahren mit Ausnahme der Selbstveranlagung. Die Steuerbehörden sind folglich verpflichtet und berechtigt, den massgeblichen Sachverhalt abzuklären und der Steuerveranlagung nur solche Tatsachen zugrunde zu legen, von deren Vorhandensein sie sich selbst überzeugt haben. Bei Bedarf müssen sie auch Tatsachen ergänzen.[485] Insbesondere darf sich die Veranlagungsbehörde nicht darauf beschränken, nur diejenigen Punkte eines Steuerfalls zu untersuchen, die aus den Angaben in der Steuererklärung hervorgehen. Tatsachen, die für eine Verminderung der Steuerlast sprechen, müssen auch dann in Betracht gezogen werden, wenn die steuerpflichtige Person diese nicht ausdrücklich hervorgehoben hat.[486] Die Veranlagungsbehörden haben folglich eine Beratungs- und Hinweispflicht.[487]
Da die steuerpflichtige Person den Sachverhalt jedoch regelmässig am besten kennt, gilt im Steuerverfahren gleichzeitig eine umfassende Mitwirkungspflicht, wonach die Parteien massgebend zur Abklärung beitragen, die erforderlichen Angaben machen und die Beweismittel für deren Richtigkeit vorweisen können müssen.[488] Das Gesetz konkretisiert die Mitwirkungspflicht in besonderen Sachdarstellungs- und Beweispflichten.[489] Die Steuerpflichtigen sind im ordentlichen Veranlagungsverfahren etwa zur Einreichung der Steuererklärung verpflichtet, wobei mittels eines Fragenkatalogs die für die Veranlagung relevanten Tatsachen erhoben werden. Durch die Beilagen (Lohnausweise, Wertschriftenverzeichnisse usw.) kommen die Steuerpflichtigen im Übrigen ihrer Beweispflicht nach.[490]
Bei einer Vollautomatisierung des Einschätzungsverfahrens wird die (elektronisch eingereichte) Steuererklärung ohne jegliche menschliche Intervention bearbeitet. Die Einschätzung erfolgt dann einzig aufgrund der von der steuerpflichtigen Person eingegebenen Daten sowie allenfalls bereits bei der Steuerbehörde vorliegender Daten.[491] Folglich findet eine weitgehende Verlagerung der Sachverhaltsermittlung auf die steuerpflichtige Person statt, wodurch Verwaltungsmitarbeitende ihrer Beratungs- und Hinweispflicht nicht mehr nachkommen können. Eine derartige Verlagerung der Sachverhaltsermittlung ist grundsätzlich nicht ohne Weiteres mit dem Untersuchungsgrundsatz vereinbar. Im Steuerveranlagungsverfahren ist diese Schwierigkeit aufgrund der ausgeprägten Mitwirkungspflicht bzw. des Selbstdeklarationsprinzips jedoch weniger problematisch.[492] Zentral ist jedoch die aus dem Untersuchungsgrundsatz zusätzlich abgeleitete Verantwortung des Staates für die gesetzmässige Durchführung des Veranlagungsverfahrens.[493] Der Staat darf diese Verantwortung nicht einfach aus der Hand geben, weshalb bei einer Vollautomation des Einschätzungsverfahrens geeignete Kontrollen, etwa mithilfe von Risikomanagementsystemen, durchgeführt werden müssen.

479 Richner/Frei/Kaufmann/Meuter, 2013, N. 35 zu § 139 StG.
480 Vgl. zu den allgemeinen Ausführungen zur Begründungspflicht auch Kapitel 3 A. II. 2. b).
481 Nufer, 2019/2020, S. 267.
482 Richner/Frei/Kaufmann/Meuter, 2013, N. 39 zu § 139 StG.
483 Vgl. dazu Kapitel 3 A. II. 3.
484 Häfelin/Müller/Uhlmann, 2020, Rn. 988.
485 Vgl. für das ordentliche Verfahren Meier-Mazzucato, 2015, § 13 Rn. 2.
486 Blumenstein/Locher, 2016, S. 516 f.
487 Braun Binder, 2020c, S. 35.
488 Vgl. Meier-Mazzucato, 2015, § 13 Rn. 2; Reich, 2020, § 26 Rn. 23 ff.; siehe etwa Art. 126 Abs. 1 DBG, Art. 42 Abs. 1 StHG oder § 134 f. StG.
489 Meier-Mazzucato, 2015, § 13 Rn. 2.
490 Meier-Mazzucato, 2015, § 13 Rn. 2; Reich, 2020, § 26 Rn. 39 f.
491 Vgl. etwa § 155 Abs. 4 Satz 1 der Abgabenordnung in der Fassung der Bekanntmachung vom 1. Oktober 2002 (BGBl. I S. 3866; 2003 I S. 61), zuletzt geändert durch Art. 1 des Gesetzes zur Einführung einer Pflicht zur Mitteilung grenzüberschreitender Steuergestaltungen vom 21. Dezember 2019 (BGBl. I S. 2875).
492 Braun Binder, 2020c, S. 34 f.
493 Siehe für das Zollveranlagungsverfahren auch Art. 42 Abs. 3 ZG, wonach eine Vereinfachung des Zollveranlagungsverfahrens nur zulässig ist, soweit die Zollsicherheit nicht beeinträchtigt und namentlich der Zollabgabebetrag nicht geschmälert wird.

II. Sozialversicherungsverfahren

1. Grundlagen des Schweizer Sozialversicherungssystems

Das schweizerische System der sozialen Sicherheit umfasst fünf Bereiche:

1. die Alters-, Hinterlassenen- und Invalidenvorsorge, die ihrerseits auf einem Drei-Säulen-Prinzip basiert. Die erste Säule, bestehend aus der Alters- und Hinterlassenenversicherung (AHV) sowie der Invalidenversicherung (IV), gewährleistet die Existenzsicherung, während die zweite Säule mit der beruflichen Vorsorge (Pensionskasse) die Fortsetzung der gewohnten Lebenshaltung sichert. Die dritte Säule stellt die individuelle, private Vorsorge dar, welche sich aus der Säule 3a (steuerlich privilegiertes Sparen) und der Säule 3b (übriges Sparen) zusammensetzt;[494]
2. die soziale Krankenversicherung und die Unfallversicherung schützen vor den Folgen einer Krankheit und eines Unfalls;
3. die Arbeitslosenversicherung;
4. den Erwerbsersatz für Dienstleistende bei Mutterschaft und bei Vaterschaft sowie
5. die Familienzulagen.[495]

Diese Sozialversicherungen zielen darauf ab, die wirtschaftlichen Folgen von Risiken[496] abzudecken, indem sie Leistungen erbringen oder Kosten bei Krankheit und Unfall tragen. Subsidiäre (nachrangige) Instrumente sind Ergänzungsleistungen bei Nichtdeckung des Existenzminimums durch die Sozialversicherungsleistungen und die Sozialhilfe für Notlagen in Bereichen, die von den Sozialversicherungen nicht abgedeckt werden.[497]
Eingeleitet werden kann das verwaltungsrechtliche Verfahren auf Erlass von Sozialleistungen nur mittels Anmeldung (des Anspruchs); in der Regel durch die versicherte Person. Solange keine Anmeldung erfolgt, wird die Versicherung folglich nicht von Amtes wegen tätig (Art. 29 ATSG[498]). Nach ergangener Anmeldung ist die Versicherung jedoch verpflichtet, ein Verfahren zu eröffnen.[499] Das Verfahren wird mittels Verfügung abgeschlossen. Wenn der Entscheidgegenstand von erheblicher Bedeutung ist, muss dies von Amtes wegen geschehen, ansonsten nur auf Verlangen der versicherten Person.[500]

2. Kompetenzordnung

a) Bundesrechtliche Kompetenzen

Die Bundesverfassung überträgt dem Bund für die verschiedenen Sozialversicherungen weitreichende Kompetenzen; die kantonalen Kompetenzen wurden dagegen zunehmend eingeschränkt.[501] Die Rechtsetzungskompetenzen hat der Bund durch den Erlass zahlreicher Gesetze und Verordnungen, welche die jeweiligen Versicherungen umfassend regeln, wahrgenommen.[502] Das Bundesgesetz über den Allgemeinen Teil des Sozialversicherungsrechts (ATSG) dient dabei mit dem Ziel einer Vereinheitlichung und Harmonisierung als allgemeiner Teil für viele Sozialversicherungen. Die Bestimmungen des ATSG sind auf die bundesgesetzlich geregelten Sozialversicherungen anwendbar, soweit die entsprechenden Sachgesetze dies vorsehen (Art. 2 ATSG). Die einzelnen Sozialversicherungsgesetze sind insoweit Spezialgesetze bzw. dem ATSG gleich-, aber nicht übergeordnet.[503] Weitere Rechtsquellen sind die von den Sozialversicherungsträgern selbst erlassenen Reglemente und die durch die Aufsichtsbehörden erlassenen Verwaltungsverordnungen (Wegleitungen, Kreisschreiben, Richtlinien usw.).[504]

b) Kantonale Kompetenzen

Das kantonale Sozialversicherungsrecht fällt nicht unter das ATSG.[505] Die Kantone gewähren die Prämienverbilligung bei den Krankenversicherungsprämien (Art. 65 f. KVG[506]) und sind für die Regelung der Restfinanzierung der Pflegekosten (Art. 25a Abs. 5 KVG) sowie für Teilbereiche der Zulassung zur Leistungserbringung (Art. 39 KVG) zuständig. Im Kanton Zürich werden die Prämienverbilligungen in §§ 3 ff. EG KVG[507] geregelt. Ferner können die Kantone über das ELG[508] hinausgehende Leistungen vorsehen (Art. 2 Abs. 2 ELG) und sind für die Vergütung der Krankheits- und Behinderungskosten im Rahmen des ELG zuständig (Art. 14 ff. ELG). Zudem verfügen sie über Kompetenzen im Bereich der Familienzulagen (Art. 3 Abs. 2 FamZG[509]) und sind für die Errichtung und Beaufsichtigung der kantonalen Familienausgleichskasse (Art. 17 FamZG)

494 Widmer, 2019, S. 9.
495 Vgl. Kieser, 2019, S. 24 ff.
496 Die klassischen sozialen Risiken sind Krankheit (medizinische Behandlung und Erwerbsausfall), Arbeitsunfälle und Berufskrankheiten, Mutterschaft, Familienlasten, Arbeitslosigkeit, Invalidität, Alter, Tod (Hinterlassensein), zudem Nichtberufsunfälle, Beeinträchtigungen der psychischen Integrität und Erwerbsersatz bei Dienstleistenden, Gächter/Burch, 2014, Rn. 1.3 f.
497 Widmer, 2019, S. 3 f.
498 Bundesgesetz über den Allgemeinen Teil des Sozialversicherungsrechts (ATSG) vom 6. Oktober 2000, SR 830.1.
499 Flückiger, 2014, Rn. 4.5 ff.
500 Kieser, 2019, S. 430.
501 Vgl. Art. 12, 41, 59 Abs. 4 und 5, 61 Abs. 4 und 5, 111–117 BV.
502 Vgl. dazu nur z. B. die Spezialgesetze Bundesgesetz über die Alters- und Hinterlassenenversicherung (AHVG) vom 20. Dezember 1946, SR 831.10, Bundesgesetz über die Invalidenversicherung (IVG) vom 19. Juni 1958, SR 831.20, Bundesgesetz über Ergänzungsleistungen zur Alters-, Hinterlassenen- und Invalidenversicherung (ELG), SR 831.30, Bundesgesetz über die berufliche Alters-, Hinterlassenen- und Invalidenvorsorge (BVG) vom 25. Juni 1982, SR 831.40, Bundesgesetz über die Krankenversicherung (KVG) vom 18. März 1994, SR 832.10, Bundesgesetz über die Unfallversicherung (UVG) vom 20. März 1981, SR 832.20, Bundesgesetz über den Erwerbsersatz für Dienstleistende bei Mutterschaft und bei Vaterschaft (Erwerbsersatzgesetz, EOG) vom 25. September 1952, SR 834.1, Bundesgesetz über die Familienzulagen in der Landwirtschaft (FLG) vom 20. Juni 1952, SR 836.1, Bundesgesetz über die obligatorische Arbeitslosenversicherung und die Insolvenzentschädigung (Arbeitslosenversicherungsgesetz, AVIG) vom 25. Juni 1982, SR 837.0.
503 Bürkle, 2020, N. 9 zu Art. 2 ATSG; Gächter/Burch, 2014, Rn. 1.13; Widmer, 2019, S. 428.
504 Gächter/Burch, 2014, Rn. 1.20 f.
505 Vgl. zum Ganzen Bürkle, 2020, N. 8 zu Art. 1 ATSG; Kieser, 2020, N. 7 zu Art. 1 ATSG.
506 Bundesgesetz über die Krankenversicherung (KVG) vom 18. März 1994, SR 832.10.
507 Einführungsgesetz zum Krankenversicherungsgesetz (EG KVG) vom 29. April 2019, LS 832.01.
508 Bundesgesetz über Ergänzungsleistungen zur Alters-, Hinterlassenen- und Invalidenversicherung (ELG) vom 6. Oktober 2006, SR 831.30.
509 Bundesgesetz über die Familienzulagen und Finanzhilfen an Familienorganisationen (Familienzulagengesetz, FamZG) vom 24. März 2006, SR 836.2.

sowie für die in Höhe und Art das FLG[510] übersteigenden Familienzulagen in der Landwirtschaft und die entsprechende Sonderbeitragserhebung zuständig. Die das FLG übersteigenden Familienzulagen in der Landwirtschaft sind im Kanton Zürich in § 171a LG[511] geregelt. Kantonale Gesetzgebung ist auch in Bereichen möglich, die vom Bund nicht geregelt werden, etwa für besondere Zulagen bei finanziellen Schwierigkeiten (z. B. kantonalrechtliche Beihilfen).[512] Die Beihilfen und Zuschüsse sind im Kanton Zürich in §§ 13 und 20 ZLG[513] geregelt. Die Sozialhilfe unterliegt ebenfalls der kantonalen Regelung, wobei der Vollzug in den Zuständigkeitsbereich der politischen Gemeinden fällt. Im Kanton Zürich wird dies im Sozialhilfegesetz geregelt.[514] Schliesslich fällt auch die Organisation der Träger der unterschiedlichen Versicherungen in die Kompetenz der Kantone. Einige Kantone gründeten dazu Kompetenzzentren, welche verschiedene Versicherungen umfassen. So richtete der Kanton Zürich die Sozialversicherungsanstalt des Kantons Zürich ein, welche in die Rechtsform einer öffentlichen Anstalt gekleidet ist.[515]

3. KI-Anwendungen im Sozialversicherungsbereich

Auch wenn KI-Anwendungen bei den Schweizer Sozialversicherungen aktuell noch nicht vorhanden oder erst in einer sehr frühen Entwicklungsphase sind, zeigen die internationalen Fallstudien eindrücklich eine Entwicklungstendenz und damit auch das Potenzial von KI-Anwendungen im Sozialbereich.[516] So werden KI-Systeme in anderen Ländern eingesetzt, um Anträge für Sozialleistungen automatisch zu bearbeiten, den wahrscheinlichen Verlauf des Sozialhilfeleistungsbezugs vorauszusagen, die Rechtmässigkeit von Sozialhilfeleistungen zu überprüfen und zu Unrecht oder missbräuchlich bezogene Sozialleistungen aufzudecken.[517] Mithin werden KI-Projekte hauptsächlich für drei unterschiedliche Einsatzbereiche entwickelt: erstens für die automatische Bearbeitung von Anträgen für Sozialleistungen, zweitens für eine fortgeschrittene Datenanalyse, um zukünftige Verläufe zu prognostizieren und darauf basierend Massnahmen zu ergreifen, sowie drittens zur Bekämpfung von Sozialleistungsbetrug. Es handelt sich mithin um die Vornahme (automatisches Bearbeiten), Unterstützung (Datenanalyse) und Kontrolle (Betrugsbekämpfung) der Verwaltungstätigkeit im Sozialversicherungsbereich.

4. Rechtliche Rahmenbedingungen

Aus rechtlicher Hinsicht sind hinsichtlich des (potenziellen) KI-Einsatzes im Sozialversicherungsbereich vor allem zwei Aspekte zu erwähnen.[518] Einerseits stellt sich vor dem Hintergrund der zum Teil weitreichenden Datensammlungen die Frage nach ausreichenden gesetzlichen Grundlagen. Anderseits ist aufgrund der persönlichen Daten die erhöhte Gefahr von Diskriminierungen zu betonen.

a) Legalitätsprinzip

Jede Verwaltungstätigkeit darf nur aufgrund und nach Massgabe von generell-abstrakten, genügend bestimmten Rechtsnormen ausgeübt werden. Wichtige Rechtsnormen müssen dabei in einem formellen Gesetz enthalten sein. Sollten KI-Anwendungen zur Bearbeitung oder Kontrolle von Sozialleistungsansprüchen eingesetzt werden, müssen auch diese in einer gesetzlichen Grundlage verankert werden, die den Anforderungen an die Normstufe und die Normdichte entspricht.[519]

Bei Personendaten, die von Sozialversicherungen bearbeitet werden, handelt es sich in der Regel um sensitive Personendaten im Sinne von § 3 Abs. 4 lit. a IDG bzw. besonders schützenswerte Personendaten nach Art. 3 Bst. c DSG (Art. 5 Bst. c revDSG). Zudem würden allfällige KI-basierte Betrugsbekämpfungsanwendungen, welche das jeweilige Missbrauchsrisiko von Sozialleistungsempfängerinnen und -empfängern berechnen sollen, als Persönlichkeitsprofile im Sinne von § 3 Abs. 4 lit. b IDG bzw. Art. 3 Bst. d DSG (bzw. als Profiling im Sinne von Art. 5 Bst. f revDSG, unter Umständen auch als Profiling mit hohem Risiko im Sinne von Art. 5 Bst. g revDSG) qualifiziert werden.

Sowohl Art. 17 Abs. 2 DSG (Art. 34 Abs. 2 revDSG) als auch § 8 Abs. 2 IDG verlangen für die Bearbeitung dieser Kategorie von Personendaten eine Grundlage in einem Gesetz im formellen Sinne.[520] Folglich ist für KI-Anwendungen von Sozialversicherungen regelmässig eine Grundlage in einem solchen Gesetz notwendig.

Aus dem Legalitätsprinzip lassen sich neben der Normstufe auch Anforderungen an die Normdichte ableiten. Sollen etwa vollautomatisierte Verfügungen bzw. Anordnungen möglich sein, müssen die Bereiche, in denen ein automatisierter Verfügungserlass vorgesehen ist, genau bestimmt werden. Zudem müssen die benötigten Datenkategorien festgelegt werden, damit der Kreis der Betroffenen transparent ausgewiesen werden kann.[521] Eine reine Blankettermächtigung zur Bearbeitung besonders schützenswerter Personendaten ist nicht ausreichend.[522] Schliesslich ist auch die Art und Weise der Datenbearbeitung zu beschreiben, insbesondere sind die Art der Beschaffung oder der Bekanntgabe sowie gegebenenfalls die Bekanntgabe an Dritte und technische Abrufverfahren anzugeben.

b) Diskriminierungsverbot

Wie bereits in den Ausführungen zu den allgemeinen Herausforderungen des staatlichen KI-Einsatzes erwähnt wurde, stellt das Diskriminierungspotenzial von KI-Anwendungen eines der grössten Probleme dar. Gerade im Bereich der Sozialversicherungen kann sich diese Gefahr einer diskriminierenden Behandlung akzentuieren, da Sozialversicherungen regelmässig besonders schützenswerte bzw. sensitive Personendaten bear-

510 Bundesgesetz über die Familienzulagen in der Landwirtschaft (FLG) vom 20. Juni 1952, SR 836.1.
511 Landwirtschaftsgesetz (LG) vom 2. September 1979, SR 910.1.
512 Bürkle, 2020, N. 10 zu Art. 1 ATSG.
513 Zusatzleistungsgesetz (ZLG) vom 7. Februar 1971, LS 831.3.
514 Sozialhilfegesetz (SHG) vom 14. Juni 1981, LS 851.1.
515 § 1 Einführungsgesetz zu den Bundesgesetzen über die Alters- und Hinterlassenenversicherung und die Invalidenversicherung (EG AHVG/IVG) vom 20. Februar 1994, LS 831.1.
516 Vgl. Kapitel 2 F. III., aber auch Kapitel 2 F. II. 2 und 7.
517 Vgl. Kapitel 2 F. III.
518 Selbstverständlich sind weitere rechtliche Aspekte wie z. B. die in Kapitel 3 A. II. thematisierten Verfahrensgarantien und Verfahrensgrundsätze ebenfalls zu berücksichtigen.
519 Vgl. zu den allgemeinen Ausführungen zum Legalitätsprinzip Kapitel 3 A. I.
520 Vgl. dazu Kapitel 3 A. IV.
521 Baeriswyl, 2012, N. 18 zu § 8 IDG.
522 Vgl. dazu Rudin, 2017, S. 61.

beiten.[523] Bei materiellen Entscheidungen über Sozialleistungen werden häufig Gesundheitsdaten über eine Person bearbeitet, da medizinische Gutachten in fast allen Sozialversicherungen das zentrale Mittel zur Sachverhaltsabklärung sind.[524] Zudem sind zahlreiche nicht medizinische persönliche Daten wie etwa der Wohnort, Familienverhältnisse und Ausbildung in Gutachten enthalten. Falls nun KI-Anwendungen zur Vornahme oder Unterstützung dieser Anordnungen eingesetzt werden sollen, müssen zunächst präexistierende Diskriminierungsmuster in den Trainingsdaten und der Systemarchitektur ausgeschlossen werden, damit diese nicht reproduziert werden. Weiter müssen eine genügende Repräsentation aller Bevölkerungsgruppen sowie Kontrollen sichergestellt werden, um das Anknüpfen an allfällige Proxies so früh wie möglich zu entdecken.[525]
Werden KI-Anwendungen eingesetzt, um Missbrauchsrisiken einzuschätzen, muss ein besonderes Augenmerk auf allfällige Feedback-Loops geworfen werden. Käme das KI-System z. B. zu der Feststellung, dass bei Personen mit Migrationshintergrund generell ein grösseres Risiko für Sozialversicherungsbetrug besteht, und würden diese Personen aufgrund der KI-Anwendung dann vermehrt kontrolliert, würden bei dieser Bevölkerungsgruppe aufgrund der erhöhten Kontrolle auch mehr tatsächliche Missbräuche als bei den weniger kontrollierten Sozialleistungsempfängerinnen und -empfängern ohne Migrationshintergrund festgestellt werden. Das KI-System würde daher in seiner diskriminierenden Annahme bestätigt werden.[526] Auch wenn es sich hier nur um ein entscheidungsvorbereitendes Warnsystem handeln würde, das einen Menschen auf ein potenzielles Betrugs- bzw. Missbrauchsgeschehen hinweist, ist zu bedenken, dass (vermeintlich objektive) maschinelle Hinweise das Risiko bergen, vom Menschen ohne vertiefte Reflexion befolgt zu werden (Stichwort «Maschinenhörigkeit»). Dies nicht zuletzt auch, weil bei Unterlassen einer durch das KI-System angeregten Überprüfung oder bei der Wahl einer anderen als der vorgeschlagenen Handlungsweise eine Begründung bzw. Rechtfertigung erforderlich sein kann. Bei KI-Anwendungen für Sozialversicherungen sollte folglich sichergestellt werden, dass die Verwaltungsmitarbeitenden einen genügend grossen Entscheidungsspielraum haben, um ihrer Fachexpertise und Erfahrung Rechnung tragen zu können.[527]
Aus dem Gesagten ist ersichtlich, dass gerade im Sozialversicherungsbereich ein erhebliches Diskriminierungsrisiko zu berücksichtigen ist. Hinzu kommt, dass fast die gesamte Bevölkerung in irgendeiner Weise zu einem bestimmten Zeitpunkt Sozialleistungen bezieht. Eine diskriminierende KI-Anwendung in diesem Bereich würde folglich zahlreiche Menschen betreffen.

III. Chatbots

1. Was ist ein Chatbot?

Ein Chatbot ist ein Onlinedialogsystem, das in natürlicher Sprache und in Echtzeit kommunizieren kann.[528] Die Kommunikation ist interaktiv und gleicht derjenigen mit einer natürlichen Person.[529] Chatbots können ohne menschlichen Input Anfragen beantworten und Aktionen einleiten.[530] Sie stellen damit eine konversationsbasierte Schnittstelle zwischen Mensch und Maschine dar.[531] Voicebots kommunizieren nicht in geschriebener, sondern in gesprochener Sprache. Dafür benötigen sie eine zusätzliche Ebene, auf der gesprochene Eingaben in geschriebene Textbausteine und geschriebene Antworten des Chatbots in automatisiert gesprochene Antworten umgewandelt werden.[532]
Technisch gesehen bestehen Chatbots aus vier Komponenten:[533]

- **Instant-Messaging-Plattform:** eine digitale Umgebung, welche die Interaktion mit Benutzerinnen und Benutzern erlaubt, z. B. eine Gemeindewebseite.
- **Protokolle:** verknüpfte Inhalte, die der Chatbot kennt und auf die er zugreifen kann. Im Kontext der Verwaltung sind dies zum einen alle Informationen über relevante Verfahren, Ansprüche und Leistungen, aber zum anderen auch das gesammelte Erfahrungswissen des entsprechenden Verwaltungszweigs.[534]
- **Module:** technische Fertigkeiten, die der Chatbot beherrscht, um seine Dienste auszuführen (z. B. Angaben über den Nutzer speichern, Datenbanken lesen oder Zahlungen tätigen). Im Bereich der Verwaltung könnte das z. B. das Schreiben von Briefen, das Ausfüllen von Formularen oder das Verlängern von Fristen umfassen.[535]
- **Analytics:** Die Arbeit des Chatbots wird ständig ausgewertet, damit dieser verbessert werden kann.

Konzeptuell funktionieren Chatbots folgendermassen: Zunächst analysiert der Chatbot die Eingabe und erfasst das eigentliche Anliegen. Dieses Anliegen bildet die Grundlage für eine Recherche in einer verknüpften Datenbank oder für eine Suchanfrage im Internet. Findet der Chatbot eine passende Antwort, bereitet er sie für die Ausgabe auf und kommuniziert sie dem Gegenüber oder leitet eine entsprechende Aktion ein.[536]
Bezüglich ihrer Leistungsfähigkeit oder Funktionsweise können sich Chatbots stark unterscheiden. Einfache Chatbots erkennen lediglich bestimmte vordefinierte Schlagworte (Keywords) oder Fragestellungen und zeigen eine mit dem jeweiligen Schlagwort bzw. mit der jeweiligen Fragestellung verknüpfte vorprogrammierte Antwort an. Komplexere Chatbots dagegen beruhen auf maschinellen Lernverfahren, d. h., sie lernen im Einsatz dazu, eliminieren Schwachstellen und erweitern fortlaufend das eigene Erfahrungswissen. Dies ermöglicht dem Chatbot, Anfragen in unterschiedlichen Kontexten zu verstehen und differenziert darauf zu reagieren.[537]
Die Einsatzmöglichkeiten für Chatbots sind vielfältig und grundsätzlich unbegrenzt. Werden die einzelnen Anwendungsmög-

523 Vgl. zur Diskriminierungsproblematik auch Kapitel 3 A. III.
524 Aliotta, 2014, Rn. 6.1.
525 Vgl. dazu Kapitel 3 A. III.
526 Vgl. dazu ebenfalls Kapitel 3 A. III.
527 Vgl. Kapitel 3 A. III. 3.
528 BigData-Insider vom 28.02.2018, Was ist ein Chatbot?, https://www.bigdata-insider.de/was-ist-ein-chatbot-a-690591/.
529 Demaj/Sägesser, 2017, S. 1.
530 BigData-Insider vom 28.02.2018, Was ist ein Chatbot?, https://www.bigdata-insider.de/was-ist-ein-chatbot-a-690591/.
531 Demaj/Sägesser, 2017, S. 1.
532 Telefongespräch mit S. Höltschi, Adretis AG, vom 16.12.2020.
533 Vgl. zum Ganzen Demaj/Sägesser, 2017, S. 3.
534 Ringeisen/Bertolosi-Lehr/Demaj, 2018, S. 55.
535 Ringeisen/Bertolosi-Lehr/Demaj, 2018, S. 55.
536 BigData-Insider vom 28.02.2018, Was ist ein Chatbot?, https://www.bigdata-insider.de/was-ist-ein-chatbot-a-690591/.
537 BigData-Insider vom 28.02.2018, Was ist ein Chatbot?, https://www.bigdata-insider.de/was-ist-ein-chatbot-a-690591/.

lichkeiten abstrahiert, ergeben sich fünf Bereiche, in denen der Einsatz von Chatbots besonders sinnvoll erscheint:[538]

- **Kundensupport:** Der Chatbot dient als Orientierungshilfe bei häufig gestellten Fragen.
- **Verkaufsassistent:** Der Chatbot erfasst die Bedürfnisse der Kunden und bietet passende Lösungen an.
- **Formular-Helfer:** Chatbots können das Ausfüllen von Formularen erleichtern oder die in der Konversation erfragten Informationen direkt in ein verknüpftes System einspeisen.
- **Lern-Tutor:** Chatbots können Wissen in Interaktionsform vermitteln und so im Bereich des E-Learnings eingesetzt werden.
- **Newsfeed:** Chatbots können Informationen aufgrund thematischer Vorgaben aufarbeiten und in Konversationsform vermitteln.

2. Chatbots in der öffentlichen Verwaltung

Für den Bereich der öffentlichen Verwaltung stellen Chatbots eine konversationsbasierte Schnittstelle zwischen Privaten, Maschine und öffentlicher Verwaltung dar. Idealerweise sind sie die erste Anlaufstelle für alle denkbaren Anliegen der Bevölkerung, z. B., um eine Leistung zu beantragen, einen Anspruch abzuklären oder etwas zu melden. Der Chatbot liefert entweder direkt die benötigten Informationen, tritt mit den Systemen der Verwaltung in Kontakt (z. B. durchsucht er die Datenbank des Handelsregisters) oder verweist an die zuständige Verwaltungseinheit, wenn er das Problem nicht selbstständig bewältigen kann. Damit reduzieren sich Aufwand und Komplexität einer Interaktion mit der Verwaltung für Private auf ein Minimum.[539]
Der Chatbot kann insbesondere folgende Rollen einnehmen:[540]

- **Chatbot als Verfahrensführer:** Der Chatbot wird quasi als Fachexperte für einen bestimmten Verwaltungszweig eingesetzt. Beispielsweise unterstützt er Private dabei, Bewilligungen zu beantragen oder Formulare auszufüllen. Er erfasst das Bedürfnis, trifft Abklärungen und leitet notwendige Handlungen ein. Diese Aufgabe kann er in allen Bereichen erfüllen, die inhaltlich und prozessual stark strukturiert bzw. normiert sind.
- **Chatbot als Auskunftsgeber:** Der Chatbot ist eine Auskunftsstelle, welche Informationsbedürfnisse bezüglich beliebiger Aspekte der Verwaltungstätigkeit befriedigt, indem er z. B. Auskünfte über Zuständigkeiten, Öffnungszeiten oder Registerinhalte erteilt.
- **Chatbot als Aktivierer:** Der Chatbot versucht mittels persönlicher Kommunikation, ein bestimmtes erwünschtes Verhalten, z. B. die Einhaltung von Fristen, auszulösen.

3. Rechtliche Rahmenbedingungen

In diesem Abschnitt wird der oben skizzierte Einsatz von Chatbots durch die öffentliche Verwaltung rechtlich beurteilt. Ziel ist, einen Überblick über die sich stellenden Rechtsfragen zu bieten. Dabei stehen die Fragen im Zentrum, die sich aus den Einsatzformen ergeben, welche im Kanton Zürich bereits realisiert wurden bzw. mit deren Einführung in naher Zukunft zu rechnen ist. Konkret sind dies die Einsatzmöglichkeiten von Chatbots zu Auskunftszwecken sowie zur Unterstützung beim Ausfüllen von E-Formularen. Weitere Verwendungsalternativen werden soweit möglich und sinnvoll berücksichtigt.

a) Legalitätsprinzip

Es stellt sich die Frage, ob für den Einsatz von Chatbots durch die öffentliche Verwaltung eine gesetzliche Grundlage notwendig ist und wie diese gegebenenfalls ausgestaltet sein muss. Die Anforderungen an die gesetzliche Grundlage ergeben sich aus dem Legalitätsprinzip und dem grundrechtlichen Anspruch auf informationelle Selbstbestimmung[541], welcher im anwendbaren Datenschutzgesetz (in casu das IDG) konkretisiert wird.

i. Erfordernis des Rechtssatzes

Das Gesetzmässigkeitsprinzip gilt für die gesamte Verwaltungstätigkeit einschliesslich der Erteilung von Auskünften oder standardisierten Antworten auf häufig gestellte Fragen.[542] Entsprechend muss auch der Einsatz von Chatbots grundsätzlich auf einer hinreichenden Rechtsgrundlage beruhen. Allerdings betrifft diese Voraussetzung nicht alle Arten von Chatbots.
Da die gesetzliche Regelung der Informationstätigkeit nicht ganz einfach ist, wird in der Lehre grundsätzlich anerkannt, dass die Auskunftserteilung einer Behörde im Einzelfall keine ausdrückliche gesetzliche Grundlage voraussetzt.[543] Solange eine Chatbot ausschliesslich zur Auskunftserteilung im Einzelfall eingesetzt wird, können hierfür keine weitergehenden Anforderungen an die Rechtsgrundlage als an die Erteilung von Auskünften durch Verwaltungsmitarbeitende gestellt werden.
Im Kanton Zürich findet sich für die Informationstätigkeit ausserdem eine Grundlage im Gesetz über die Information und den Datenschutz (IDG). Gemäss § 1 Abs. 2 lit. a IDG bezweckt das Gesetz, das Handeln der öffentlichen Organe transparent zu gestalten und damit die freie demokratische Meinungsbildung sowie die Kontrolle des staatlichen Handelns zu ermöglichen. Konkret verlangt § 14 Abs. 1 IDG, dass öffentliche Organe von sich aus über ihre Tätigkeiten von öffentlichem Interesse, insbesondere über Aufbau, Zuständigkeiten und Ansprechpersonen, informieren. Gestützt auf diese rechtlichen Grundlagen erliess der Regierungsrat des Kantons Zürich die Verordnung über die Information und den Datenschutz (IDV). Als Informationskanäle werden darin explizit die amtlichen Publikationsorgane, die Medien und das Internet genannt, wobei das für die Informationstätigkeit zuständige Organ weitere Informationsmittel bestimmen kann (§ 4 Abs. 1 und 2 IDV). Setzt der Kanton Zürich Chatbots als Informationsmittel ein, kann dies mithin auch unter § 4 Abs. 2 IDV subsumiert werden.
Werden Chatbots zur Unterstützung beim Ausfüllen von Formularen genutzt, kann dies als Erteilung einer Auskunft eingestuft werden, solange diese nur Informationen wiedergeben, die dabei helfen, das Formular richtig auszufüllen (etwa erläuternde Hinweise, welche Informationen in welches Feld einzufügen sind). Das oben Ausgeführte gilt entsprechend. Die Notwendigkeit einer gesetzlichen Grundlage für den Einsatz elektronischer Formulare bzw. eine daran anschliessende elektronische Abwicklung des Verwaltungsverfahrens[544] wird an dieser Stelle nicht weiter thematisiert.

538 Demaj/Sägesser, 2017, S. 10.
539 Ringeisen/Bertolosi-Lehr/Demaj, 2018, S. 54 ff.
540 Ringeisen/Bertolosi-Lehr/Demaj, 2018, S. 57.
541 Art. 13 Abs. 2 BV.
542 Häfelin/Müller/Uhlmann, 2020, Rn. 383.
543 Häfelin/Müller/Uhlmann, 2020, Rn. 383.
544 Gemäss bundesgerichtlicher Rechtsprechung ist für den elektronischen Verkehr im Rahmen von Gerichts- und Verwaltungsverfahren eine spezifische gesetzliche Regelung notwendig; BGE 142 V 152 E. 2.4; BGE 145 V 90 E. 6.2.1. Vgl. dazu Glaser/Ehrat, 2019, S. 3; Glaser, 2018, S. 185. Im Kanton Zürich werden aktuell im Rahmen des Projekts IP2.1 DigiLex die notwendigen gesetzlichen Grundlagen für den formellen elektronischen Geschäftsverkehr zwischen Privaten und der öffentlichen Verwaltung geschaffen.

Führt der Chatbot neben der Auskunfterteilung weitere Schritte aus, so ist zu prüfen, ob dies von der gesetzlichen Grundlage abgedeckt ist, welche die Aufgaben der Behörde umschreibt, die den Chatbot einsetzt. Eine (formell-)gesetzliche Grundlage ist jedenfalls dann notwendig, wenn im Rahmen von Chatbots besondere Personendaten bearbeitet werden (§ 8 Abs. 2 IDG). Dies wäre etwa dann der Fall, wenn der Chatbot im Sinne des Once-only-Prinzips besondere Personendaten, die bereits einmal erhoben worden sind, automatisch in das Formular einträgt. Eine entsprechende Rechtsgrundlage dürfte auch dann angezeigt sein, wenn durch die Nutzung von Chatbots die im Formular eingegebenen Daten neu strukturiert und darauf basierend neue Verknüpfungen und automatisierte Verarbeitungen zu verschiedenen Zwecken ermöglicht werden.[545]

ii. Erfordernis der Normstufe

Für den Einsatz von Chatbots ist dann eine formell-gesetzliche Grundlage notwendig, wenn eine der Alternativen von Art. 38 Abs. 1 KV ZH einschlägig ist. Darüber hinaus haben auch datenschutzrechtliche Überlegungen Einfluss auf die erforderliche Normstufe.

Chatbots arbeiten mit KI, um eine semantische Analyse der Eingaben von Benutzerinnen und Benutzern vorzunehmen. Dieser Vorgang fällt potenziell in den Anwendungsbereich des IDG. Fraglich ist zunächst, ob das Auslesen der Eingaben durch eine KI als Bearbeiten von Daten zu qualifizieren ist. Dies ist zu bejahen: Erfasst ist jeder Umgang mit Daten ungeachtet des Verfahrens der Datenbearbeitung.[546] Entscheidend ist deshalb die Frage, ob die Eingaben, welche Private bei der Benutzung des Chatbots vornehmen, Personendaten darstellen. Dies ist je nach Konzeption und Komplexität des infrage stehenden Chatbots unterschiedlich. Es gibt Chatbots, deren Verwendung keine Personendaten voraussetzt. Wenn der Chatbot nur allgemeine Informationen wiedergibt und dafür keine persönlichen Angaben erfragt, handelt es sich in der Regel nicht um die Bearbeitung von Personendaten. Die Eingabe «Wann ist das Steueramt geöffnet?» erlaubt keinen Rückschluss auf die anfragende Person, sofern der Chatbot keine Daten über die beteiligten Endgeräte der Privaten erfasst (z. B. IP-Adresse), welche es erlauben, die Anfrage einer konkreten Person zuzuordnen. Der Anwendungsbereich des Datenschutzrechts wäre dann nicht betroffen. Andere Chatbots erfragen die zur Beantwortung einer individuell-konkreten Frage benötigten Angaben zur Person. Diese Daten stellen Personendaten im Sinne von § 3 Abs. 3 IDG dar. Gespräche mit Entwicklern haben zudem ergeben, dass viele Chatbots bald über eine Identifikationsfunktion verfügen werden.[547] Auch der Grundlagen-Bot, den die Adretis AG im Auftrag des Amtes für Informatik plant, soll eine Identifikationsfunktion beinhalten.[548] Durch die Möglichkeiten, Benutzer und Benutzerinnen eindeutig zu identifizieren, werden neue Bereiche erschlossen, in welchen Chatbots eingesetzt werden können. Sofern eine eindeutige Identifikation stattfindet, liegt eine Bearbeitung von Personendaten vor.

Gemäss § 8 Abs. 2 IDG ist für die Bearbeitung nicht besonderes schützenswerter Personendaten die sachliche Zuständigkeit der Behörde grundsätzlich ausreichend. Diese Aufgabenumschreibung braucht nicht in einem Gesetz im formellen Sinne enthalten zu sein.[549] Das ist nicht in allen Fällen unproblematisch, denn auch die Bearbeitung von gewöhnlichen Personendaten kann zu schweren Grundrechtseingriffen führen. Dann wird nach Art. 38 Abs. 1 lit. b KV ZH sowie Art. 36 Abs. 1 BV eine Grundlage in einem formellen Gesetz verlangt. Liegt eine Bearbeitung besonderer Personendaten gemäss § 3 Abs. 4 IDG vor, ist dagegen stets eine spezifische gesetzliche Grundlage in einem formellen Gesetz erforderlich.[550]

iii. Erfordernis der Normdichte

Bestimmte Teile der Verwaltungstätigkeit lassen sich naturgemäss nur schwer normieren. Dazu gehören auch die alltäglichen, informellen Informations- und Unterstützungstätigkeiten der Verwaltungsbehörden. § 4 Abs. 2 IDV erlaubt den Behörden, selbstständig weitere Informationskanäle zu schaffen. Die Bestimmung ist sehr offen formuliert. Behörden sollen selbst entscheiden können, welche Informationskanäle nützlich und notwendig sind. Gleiches gilt für die Unterstützung beim Ausfüllen von Formularen. Auch hier stützt sich die Tätigkeit der Behörden auf die allgemeine Umschreibung ihrer Aufgaben und Zuständigkeiten. Dies ist aus Sicht des Legalitätsprinzips nicht zu beanstanden, da Flexibilität in einem bestimmten Umfang wünschenswert und notwendig ist.

Strenger sind die Anforderungen an die Normdichte hingegen, wenn Rechte oder Pflichten von Privaten betroffen sind. Im Bereich der Chatbots ist dies vor allem das Recht auf Datenschutz. Werden durch den Chatbot gewöhnliche Personendaten bearbeitet, so genügt nach § 8 Abs. 1 IDG die sachliche Zuständigkeit. Dann ist aber immerhin zu verlangen, dass diese Zuständigkeit im Gesetz klar umschrieben und die Datenbearbeitung für die betroffenen Privaten erkennbar ist.[551] Werden dagegen besondere Personendaten gemäss § 3 Abs. 4 IDG bearbeitet, verlangt § 8 Abs. 2 IDG eine hinreichend bestimmte Regelung in einem formellen Gesetz. Für die hinreichende Bestimmung muss die Regelung die verantwortliche Behörde und den Zweck der Datenbearbeitung bezeichnen sowie die Datenkategorie, die bearbeitet wird, und die eingesetzten Mittel nennen.[552]

b) Datenschutzrechtliche Anforderungen an Chatbots

Mit Blick auf den Einsatz von Chatbots ergeben sich aus datenschutzrechtlicher Sicht[553] verschiedene Anforderungen.

Daten dürfen ausschliesslich zu dem Zweck verwendet werden, zu dem sie ursprünglich erhoben worden sind. Die Informationen, die ein Chatbot zur Beurteilung eines möglichen Anspruchs auf Prämienverbilligung erhebt, dürfen ausschliesslich dafür eingesetzt werden; jede weitere Verwendung durch die Verwaltung wäre grundsätzlich rechtswidrig bzw. müsste in einer gesetzlichen Grundlage vorgesehen sein. Umgekehrt gilt, dass ein Chatbot nicht ohne entsprechende Rechtsgrundlage Daten, die zu anderen Zwecken gesammelt worden sind, abrufen und nutzen darf.[554] Dies ergibt sich aus dem in § 9 Abs. 1 IDG verankerten Grundsatz der Zweckbindung.

Ferner sind Datenverarbeitungssysteme, darunter auch Chatbots, so zu gestalten, dass möglichst wenige zur Erfüllung der

545 Vgl. nur etwa Guckelberger, 2019, Rn. 402.
546 Rudin, 2012, N. 32 zu § 3 IDG. Vgl. auch Schweizer, 2014, N. 74 zu Art. 13 BV.
547 Telefongespräch mit S. Höltschi, Adretis AG, vom 18.11.2020; Telefongespräch mit L. Demaj, byerley AG, 29.10.2020.
548 Telefongespräch mit S. Höltschi, Adretis AG, vom 16.12.2020.
549 Baeriswyl, 2012, N. 5 zu § 8 IDG.
550 Vgl. dazu auch die Ausführungen in Kapitel 3 A. I.
551 Baeriswyl, 2012, N. 4 zu § 8 IDG.
552 Baeriswyl, 2012, N. 14–21 zu § 8 IDG.
553 Vgl. zu datenschutzrechtlichen Anforderungen allgemein Kapitel 3 A. IV.
554 Harb, 2012, N. 1 f. zu § 9 IDG.

Aufgabe nicht erforderliche Personendaten anfallen. Dies bezieht sich insbesondere auf die sogenannten Randdaten, z. B. die IP-Adresse.[555] Die Datenvermeidung und Datensparsamkeit sind bereits im Rahmen der Programmierung sicherzustellen (privacy by design).[556] Das kann auch die Umgestaltung von (kommerziell hergestellten) Programmen erfordern, da diese in der Regel so konzipiert sind, dass sie möglichst viele Daten sammeln.[557] Diese Anforderung ergibt sich aus § 11 Abs. 1 IDG. Sollten dennoch solche Daten anfallen, sind sie so rasch wie möglich zu löschen, zu anonymisieren oder zu pseudonymisieren (§ 11 Abs. 2 IDG).
Zuletzt ergibt sich aus § 12 Abs. 1 IDG, dass die Beschaffung bzw. Bearbeitung von Daten durch Chatbots für Private erkennbar sein muss. Handelt es sich um besondere Personendaten oder wendet die kantonale Verwaltung Bundesrecht an, ist die betroffene Person explizit zu informieren. Die Informationspflicht umfasst die Mitteilung des verantwortlichen Organs, des Zwecks der Datenbearbeitung und der Empfänger einer allfälligen Bekanntgabe sowie die Nennung der Folgen einer möglichen Weigerung, die verlangten Daten anzugeben.[558]
Bei Voicebots ist ferner zu bedenken, dass aus der Stimme einer Person zahlreiche Informationen abgeleitet werden können (Alter, Geschlecht, Stimmung usw.). Bei einer Anfrage könnten somit Informationen erfasst werden, welche mit der Anfrage selbst überhaupt nichts zu tun haben. Solche Informationen darf die Verwaltung nicht bearbeiten: Dies ergibt sich bereits aus dem Verfassungsgrundsatz der Verhältnismässigkeit in Art. 5 Abs. 2 BV, der verlangt, dass nur diejenigen Daten erhoben werden, die für die Erfüllung der Aufgabe erforderlich sind. Für den Kanton Zürich ist zudem § 11 Abs. 1 IDG einschlägig. Sicherzustellen ist, dass die Chatbots solche Informationen nicht erfassen, im besten Fall bereits durch die Konzeption des Bots (privacy by design).

c) Fehlerhafte Auskunft: Vertrauensschutz

Aktuell übernehmen Chatbots vor allem die Funktion einer Auskunftsstelle.[559] Daher drängt sich die Frage auf, wie eine fehlerhafte Auskunft eines Chatbots rechtlich einzuordnen ist. So wäre etwa denkbar, dass der Chatbot einer Person eine tatsächlich mögliche Antwort gibt, welche im konkreten Fall aufgrund des Sachverhalts aber nicht korrekt ist. In diesem Fall stellt sich die Frage, inwiefern die Benutzerinnen und Benutzer Anspruch darauf haben, in ihrem Vertrauen auf die Antwort des Chatbots geschützt zu werden. Der Grundsatz des Vertrauensschutzes ergibt sich aus Art. 9 BV und besagt, dass Private unter bestimmten Umständen in ihrem durch behördliches Verhalten geweckten Vertrauen geschützt werden.[560] Der Schutz gegen fehlerhafte behördliche Auskünfte stellt dabei einen Spezialfall des Vertrauensschutzes dar.[561] Im Folgenden werden die Voraussetzungen des Vertrauensschutzes bei fehlerhaften behördlichen Auskünften zusammengefasst, und es wird danach gefragt, ob und wie die Voraussetzungen auf die Situation einer (fehlerhaften) Information durch einen Chatbot übertragen werden können.

i. Eignung der Auskunft zur Begründung von Vertrauen

Zunächst reicht nicht jede behördliche Auskunft automatisch als Vertrauensbasis. Die Auskunft muss inhaltlich genügend bestimmt sein.[562] Herrschende Lehre und Rechtsprechung vertreten dabei die Auffassung, nur eine individualisierte, auf einen konkreten Sachverhalt bezogene Aussage sei inhaltlich genügend bestimmt.[563] Ansonsten könnten allgemein zugängliche Informationen seitens der Verwaltung geltendes Recht de facto ausser Kraft setzen.[564] In der Literatur finden sich jedoch ebenfalls Stimmen, nach welchen auch bestimmte allgemeine Auskünfte ein berechtigtes Vertrauen begründen können sollen.[565] Umstritten ist in diesem Zusammenhang, wie allgemeine Informationen auf Webseiten von Behörden zu beurteilen sind. Die Informationstätigkeit der Verwaltung hat sich durch die technologische Entwicklung zunehmend verschoben: Im Zentrum steht nicht mehr der individuelle persönliche oder telefonische Kontakt, sondern die Veröffentlichung generell-abstrakter Informationen, insbesondere im Internet. Dabei wird erwartet, dass Private sich im Internet informieren, bzw. es stehen zunehmend weniger Kapazitäten für die individuelle Anfragenbeantwortung zur Verfügung. Für die Qualifikation als Vertrauensgrundlage ist ausschlaggebend, zu welchem Zweck Behörden Informationen im Internet veröffentlichen. Grundsätzlich ist davon auszugehen, dass Private über die Rechtslage und die Behördenpraxis informiert werden sollen, damit sie ihr Verhalten entsprechend danach ausrichten können. Dies soll die Verwaltungstätigkeit vereinfachen und die Effizienz steigern. Aus diesem Blickwinkel ist offensichtlich, dass Private den im Internet veröffentlichten Informationen ein bestimmtes Vertrauen entgegenbringen dürfen müssen. Wäre dies nicht der Fall, könnte das Verhalten nie danach ausgerichtet bzw. müsste jedes Mal eine individuelle Zusicherung eingeholt werden, dass die Informationen auf der Webseite verlässlich sind, womit der Zweck der Zurverfügungstellung von Informationen auf Webseiten unterlaufen würde.[566] Auch aus der Perspektive der Rechtsgleichheit wirkt es stossend, dass Private, die andauernd die Verwaltung kontaktieren, besser gestellt wären als Private, die sich selbstständig anhand der zu diesem Zweck zur Verfügung gestellten Informationen kundig machen.[567] Nach dem Gesagten kann *unseres Erachtens* nicht daran festgehalten werden, dass Internetauftritte von Behörden wegen des fehlenden Einzelfallbezugs keine taugliche Vertrauensgrundlage darstellen. Vielmehr müssen Private darauf vertrauen können, dass die Informationen richtig sind. Entsprechend sind sie in diesem Vertrauen mindestens dann zu schützen, wenn die Informationen bewusst zur rechtlichen Orientierung von Privaten gedacht sind.[568]
Nun stellt sich die Frage, unter welchen Umständen eine Information durch einen Chatbot eine Vertrauensgrundlage dar-

555 Baeriswyl, 2012, N. 1 zu § 11 IDG.
556 Baeriswyl, 2012, N. 9 zu § 11 IDG.
557 Baeriswyl, 2012, N. 1 zu § 11 IDG.
558 Harb, 2012, N. 5 f. zu § 12 IDG.
559 Vgl. Kapitel 2 F. II. 5.
560 Häfelin/Müller/Uhlmann, 2020, Rn. 624.
561 Häfelin/Müller/Uhlmann, 2020, Rn. 667.
562 Häfelin/Müller/Uhlmann, 2020, Rn. 668.
563 Tschannen/Zimmerli/Müller, 2014, § 22 Rn. 15; Rohner, 2014, N. 49 zu Art. 9 BV; Tschentscher, N. 16 zu Art. 9 BV; BGE 131 II 627, 637; BGE 125 I 267, 274 f.
564 Vgl. nur BGE 125 I 267 E. 4c; Weber-Dürler, 201, 294 f.
565 Häfelin/Müller/Uhlmann, 2020, Rn. 669; I. d. S. auch Uhlmann/Stojanovic, 2017, S. 736.
566 Vgl. Uhlmann/Stojanovic, 2017, S. 736.
567 Vgl. Uhlmann/Stojanovic, 2017, S. 736.
568 I. d. S. auch Häfelin/Müller/Uhlmann, 2020, Rn. 670.

stellen kann. Einerseits gibt es Chatbots, die lediglich als eine Art Suchmaschine dienen, aber keine Liste von Ergebnissen, sondern die richtige Antwort direkt in Konversationsform ausgeben. Dies dürfte z. B. der Fall sein, wenn der Chatbot nach Informationen zu Parkgebühren, Abfallentsorgung oder Ähnlichem gefragt wird. Sofern der Chatbot lediglich eine allgemeine Information in Konversationsform aufbereitet, liegt eine Analogie zum Internetauftritt der Behörde nahe. Nach der hier vertretenen Auffassung stellen nur solche Informationen, welche explizit zur rechtlichen Orientierung von Privaten gedacht sind, taugliche Vertrauensgrundlagen dar.
Anderseits gibt es Chatbots, die allgemeine Regeln auf konkrete Sachverhalte anwenden und zu diesem Zweck auch persönliche Informationen von Benutzerinnen und Benutzern erfragen. Zu denken ist dabei etwa an die Bots im Bereich der Prämienverbilligung. Sobald eine Individualisierung erfolgt, liegt damit eine auf einen konkreten Sachverhalt bezogene Aussage vor. Solche Aussagen stellen auch nach der herrschenden Lehre und Rechtsprechung eine Vertrauensgrundlage im Sinne des Vertrauensschutzes dar.[569] Im Rahmen der Beweisbarkeit wäre es schliesslich wünschenswert, wenn die betroffene Person die Möglichkeit hätte, die geführte Chatbot-Konversation schriftlich zu erhalten. Dies wäre folglich eine Dokumentationspflicht der Behörde, wodurch die betroffene Person das Recht auf Einsicht hätte.

ii. Zuständigkeit der auskunftserteilenden Behörde

Die Behörde muss zur Erteilung der Auskunft zuständig sein, wobei die Zuständigkeit zur Entscheidung in einem Sachbereich diejenige zur Auskunftserteilung mit umfasst. Es reicht jedoch aus, wenn Private nach Treu und Glauben annehmen dürfen, dass die Behörde zur Auskunft berechtigt ist. Der Schutz fällt erst dahin, wenn die Unzuständigkeit offensichtlich ist. Bei dieser Beurteilung sind sowohl objektive als auch subjektive Kriterien, insbesondere die Situation der auskunftsersuchenden Person, zu berücksichtigen. Eine Person, welche über Sonderwissen verfügt, muss sich dieses anrechnen lassen: Von ihr wird eher erwartet, die fehlende Zuständigkeit zu erkennen.[570]
Diese Voraussetzung ist im Kontext von Auskünften durch Chatbots in der Regel unproblematisch: Wenn der Chatbot über das entsprechende Wissen verfügt, um eine Antwort geben zu können, und damit programmiert wurde, um Fragen in einem bestimmten Themenbereich beantworten zu können, wird die Zuständigkeit in der Regel vorliegen. Auch wenn sie im Einzelfall fehlen sollte, dürfen die Privaten in guten Treuen davon ausgehen, dass der Chatbot für alles zuständig ist, was er beantwortet.

iii. Vorbehaltlosigkeit der Auskunft

Eine Auskunft kann nur dann ein schutzwürdiges Vertrauen begründen, wenn sie vorbehaltlos erteilt wird.[571] Viele Chatbots und Webseiten arbeiten nun aber mit Disclaimern. Diese besagen im Wesentlichen, dass alle Informationen ohne Gewähr gegeben werden. In anderen Worten soll man sich nicht auf die Informationen verlassen dürfen. Derartige Disclaimer sind besonders in der Privatwirtschaft weit verbreitet. Fraglich ist allerdings, ob staatliche Akteure solche Disclaimer gleichermassen einsetzen und damit im Resultat den Vertrauensschutz abschwächen dürfen.[572]
Grundsätzlich sind Vorbehalte bei Auskünften zulässig. Mitarbeitende der Verwaltung können jederzeit anmerken, dass die erteilte Auskunft lediglich eine spontane Einschätzung darstellt und keine Verbindlichkeit beansprucht. Die Privatperson kann sich dann nicht auf den Vertrauensschutz berufen.[573] Im Unterschied zu einer spontanen, meist mündlichen und auf einen Einzelfall bezogenen Einschätzung werden Informationen auf einer Webseite jedoch vorgängig ausführlich vorbereitet und überprüft. Sie dienen keiner spontanen Einschätzung. Im Gegenteil sollen sie Privaten ermöglichen, sich im Internet so einfach wie möglich über die Rechtslage zu informieren, um ihr Verhalten entsprechend auszurichten. Dafür sind provisorische Angaben nicht geeignet, was Behörden auch bewusst ist. Daher erscheint es eher zwecklos, die Bevölkerung mittels Internetauftritten aufklären, die dazu notwendigen Angaben aber nicht garantieren zu wollen. Behörden haben bezüglich der öffentlich zur Verfügung gestellten Informationen folglich eine höhere Sorgfaltspflicht als private Akteure.[574] Dies entspricht auch dem verwaltungsrechtlichen Grundsatz von Treu und Glauben (Art. 9 BV), welcher die Behörde zu vertrauenswürdigem Handeln verpflichtet. Zusätzlich spricht für die Verbindlichkeit von Informationen auf behördlichen Webseiten, dass das Internet zunehmend zum primären Informationskanal der Behörden wird,[575] denn je wichtiger ein Informationskanal ist, desto höhere Anforderungen sind an die darauf aufgeführten Informationen zu fordern. Folglich ist es unzulässig, dass Behörden durch umfassende Disclaimer ihrer Kommunikation im Internet (generell) jegliche Verbindlichkeit nehmen.[576]
Fraglich ist, wie diese Überlegungen auf die Informationstätigkeit von Chatbots zu übertragen sind. Dabei ist zu berücksichtigen, dass hinter dem vermehrten Einsatz von Chatbots durch die Verwaltung letztlich die Vision von digitalen Mitarbeitenden der Verwaltung steht, welche eine erste Anlaufstelle für alle denkbaren Anliegen sein sollen. Die Informationsmöglichkeiten für Private werden sich somit zunehmend auf Anwendungen wie Chatbots konzentrieren, da die Verwaltung andere Kommunikationskanäle mit dem Aufbau von Chatbots abbaut. Zwar kann die Verwaltung weiterhin alternative Kommunikationskanäle betreiben, doch werden diese zunehmend mehrheitlich für Ausnahmen und besonders anspruchsvolle Anfragen eingesetzt werden. Somit nehmen Bots eine wesentliche Rolle in der Interaktion von Bürgerinnen und Bürgern mit der Verwaltung ein, sodass die Verwendung von generellen Disclaimern stets kritisch zu prüfen ist. Je mehr Chatbots als massgebender Informationskanal für Private erscheinen und je weniger andere Informationsmöglichkeiten bestehen, desto weniger kann Privaten zugemutet werden, erhaltene Informationen nur als provisorisch zu betrachten und eigenständig auf die rechtliche Korrektheit zu überprüfen. Wird das Vertrauen in die Auskünfte von Chatbots schliesslich durch Disclaimer zu stark ausgeschlossen, werden diese von der Bevölkerung nicht als verlässliche Option wahrgenommen werden und ihr Nutzen für die Verwaltung wird begrenzt bleiben.
Für die konkrete Beurteilung der Zulässigkeit von Disclaimern bei Chatbots ist aber die Ausgestaltung des Chatbots zentral. Bei Chatbots, welche lediglich allgemeine Information aus einer Datenbank in Konversationsform ausgeben können, liegt eine Analogie zu Webseiten von Behörden nahe (vgl. die Überlegun-

569 Tschannen/Zimmerli/Müller, 2014, § 22 Rn. 15; Rohner, 2014, N. 49 zu Art. 9 BV; Tschentscher, N. 16 zu Art. 9 BV.
570 Häfelin/Müller/Uhlmann, 2020, Rn. 676 f.
571 Häfelin/Müller/Uhlmann, 2020, Rn. 682.
572 Zu dieser Frage auch Uhlmann/Stojanovic, 2017, S. 736 f.
573 Häfelin/Müller/Uhlmann, 2020, Rn. 682.
574 I. d. S. auch Häfelin/Müller/Uhlmann, 2020, Rn. 670; Uhlmann/Stojanovic, 2017, S. 737.
575 Uhlmann/Stojanovic, 2017, S. 737.
576 Überzeugend Uhlmann/Stojanovic, 2017, S. 737.

gen im vorherigen Absatz). Nach dem oben Gesagtem bedeutet dies, dass Private in ihrem Vertrauen auf allgemeine Auskünfte geschützt werden sollen. Dieser Schutz darf nicht durch umfassende Disclaimer ausgeschlossen werden. Hingegen sollten Disclaimer bei komplexen Chatbots bzw. Verfahren (eher) zulässig sein. Bei rechtlich komplexen Anfragen muss der Verwaltung zugestanden werden, darauf hinzuweisen, dass die Beantwortung der vorliegenden Anfrage komplex ist. Weiterhin muss ihr sowohl zum eigenen Schutz als auch zum Schutz der Privaten vor vorschnellem Handeln (dessen Folgen allenfalls nur in langwierigen Rechtsstreiten beseitigt oder gemildert werden können) erlaubt bleiben, die erteilten Informationen für nicht verbindlich zu erklären. Alles andere ist mit dem Bedürfnis der Privaten nach verlässlichen Informationen und nach Rechtssicherheit nicht zu vereinbaren.

iv. Unrichtigkeit der Auskunft nicht erkennbar

Der Vertrauensschutz greift nur bei gutem Glauben. Nicht geschützt wird, wer die Unrichtigkeit kannte oder hätte kennen müssen, wobei allerdings kein strenger Massstab angelegt wird. Wiederum ist die Situation der auskunftsersuchenden Person in die Beurteilung einzubeziehen.[577]
Bei gutem Glauben stellt sich die Rechtslage bei der Information durch Chatbots nicht wesentlich anders als bei der Information durch Verwaltungsmitarbeitende dar. Eine gesunde Skepsis ist angebracht, grundsätzlich sollte den Äusserungen von Behörden aber vertraut werden dürfen, auch wenn diese vielleicht überraschend wirken.
Bei Einführung von Chatbots darf von Privaten zu Beginn vorläufig eine erhöhte Wachsamkeit verlangt werden, weil diese doch sehr neue und folglich auch eher noch fehleranfällige Tools darstellen. Dabei ist auf die persönliche Situation der Privatperson abzustellen, d. h., es ist danach zu fragen, ob ihr konkret zugemutet werden konnte, die (mögliche) Fehlerhaftigkeit der Information zu erkennen, insbesondere, ob sie konkret in der Lage war, die Möglichkeiten und Vertrauenswürdigkeit von Chatbots adäquat einzuschätzen. Gegebenenfalls wäre es angebracht, neue Chatbots in den ersten Monaten des Betriebs entsprechend zu kennzeichnen.

v. Nachteilige Disposition aufgrund der Auskunft

Die betroffene Person muss im Vertrauen auf die Richtigkeit der Auskunft etwas getan oder unterlassen haben, was sich für sie nicht ohne Schaden wieder rückgängig machen lässt.[578] Diese Voraussetzung hat mit der Art der Information grundsätzlich nichts zu tun, weshalb sich weitere Ausführungen dazu erübrigen.

vi. Keine Änderung des Sachverhalts oder der Gesetzgebung

Behörden sind an die erteilte Auskunft nicht mehr gebunden, wenn sich der Sachverhalt oder die Rechtslage nachträglich geändert haben.[579] Daran wird sich auch durch die Informationstätigkeit durch Chatbots nichts ändern, denn es spielt schliesslich keine Rolle, über welchen Kanal die Information erteilt wird.
Ferner kann die Auskunft nur Verbindlichkeit in Bezug auf den Sachverhalt haben, wie er der Behörde mitgeteilt wird.[580] Hier stellt sich bezüglich Chatbots die Frage, wie mit fehlerhaften Eingaben bzw. Informationen seitens der Privaten umgegangen werden soll. Nach der hier vertretenen Ansicht darf sich die Situation der Privaten durch den Einsatz von Chatbots nicht verschlechtern, d. h., mindestens diejenigen Fehler, welche in der Kommunikation mit Verwaltungsmitarbeitenden hätten vermieden werden können, dürfen den Privaten nicht angelastet werden. Mit anderen Worten: Wo die falsche Erfassung des Sachverhalts nicht nur durch eine falsche Schilderung seitens der Privatperson, sondern wesentlich auch durch die Konzeption des Bots verursacht wird (und sei dies auch bloss die Tatsache, dass ein Bot eine Information falsch erfasst, die eine natürliche Person korrekt einzuordnen gewusst hätte), darf dies den Privaten nicht vorgehalten werden.

vii. Interessenabwägung

Selbst wenn die Voraussetzungen des Vertrauensschutzes vorliegen, ist im Einzelfall abzuwägen, ob das Interesse an der richtigen Rechtsanwendung, d. h. an der Rechtssicherheit, überwiegt.[581] An diesem Grundsatz ändert sich mit dem Einsatz von Chatbots nichts.

C. Zusammenfassung

Im Rahmen dieser Studie wurden verschiedene allgemeine rechtliche Herausforderungen identifiziert, die der Kanton Zürich zu bedenken hat, wenn er KI-Systeme einsetzen möchte.
Zunächst ist aufgrund des **Legalitätsprinzips** sicherzustellen, dass eine sowohl hinsichtlich der Normstufe als auch hinsichtlich der Normdichte ausreichende Rechtsgrundlage existiert. Dabei ist insbesondere zu berücksichtigen, dass nach den datenschutzrechtlichen Vorgaben für die Bearbeitung von besonderen Personendaten eine formell-gesetzliche Grundlage notwendig ist (§ 8 Abs. 2 IDG). Gerade im *Sozialversicherungsbereich* ist davon auszugehen, dass KI-Anwendungen vorwiegend bei der Bearbeitung besonderer Personendaten eingesetzt werden und diesfalls eine formell-gesetzliche Grundlage erforderlich ist. Im *Steuerbereich* muss im Falle einer Einschränkung von Verfahrensgrundrechten durch KI-Anwendungen ebenfalls eine formell-gesetzliche Grundlage gegeben sein (Art. 38 Abs. 1 lit. b KV ZH). Der Einsatz von Chatbots zu Informationszwecken dürfte hingegen von bestehenden Rechtsgrundlagen abgedeckt sein. Etwas anderes gilt insbesondere dann, wenn Chatbots anderweitige Aufgaben übernehmen und dabei (besondere) Personendaten bearbeiten. Zu denken ist etwa an Unterstützungsleistungen beim Ausfüllen von Formularen.
Weitere Anforderungen ergeben sich aus dem **Anspruch auf vorgängige Äusserung** und Mitwirkung in Verwaltungsverfahren. Mit dem KI-Einsatz im Rahmen von automatisierten Verfahren geht das Risiko einher, dass das Recht auf vorgängige Äusserung eingeschränkt wird. Dabei sind die Einschränkungen in vollautomatisierten Verfahren tendenziell grösser als in teilautomatisierten. In Abhängigkeit von der konkreten KI-Anwendung ist es deshalb empfehlenswert, zu prüfen, ob ein Anspruch auf vorgängige Äusserung im Rahmen eines KI-Einsatzes im VRG verankert werden soll. Dabei wird sich vermutlich die Frage stellen, ob der Anspruch auf rechtliches Gehör im VRG nicht umfassend geregelt werden sollte. Dazu würde auch die **Begründungspflicht** zählen. Beim KI-Einsatz besteht die Gefahr, dass die entscheidende Behörde der Begründungs-

577 Häfelin/Müller/Uhlmann, 2020, Rn. 684.
578 Häfelin/Müller/Uhlmann, 2020, Rn. 688.
579 Häfelin/Müller/Uhlmann, 2020, Rn. 695.
580 Häfelin/Müller/Uhlmann, 2020, Rn. 695.
581 Häfelin/Müller/Uhlmann, 2020, Rn. 699.

pflicht nicht vollumfänglich nachkommen kann. Dies gilt insbesondere für Systeme, die statistische Auswertungen vornehmen und bei denen die Entscheidung mithin nach Kriterien erfolgt, die für die Sachbearbeiterin oder den Sachbearbeiter selbst nicht nachvollziehbar sind. Werden solche Systeme für den Erlass von Anordnungen eingesetzt, ist demnach sicherzustellen, dass die Anordnung dennoch rechtskonform begründet wird. Dies könnte etwa durch die Anforderung umgesetzt werden, dass die Logik der Entscheidungsfindung in der Begründung angegeben werden muss. Entsprechende Vorgaben könnten Eingang in das VRG finden, aber auch im Rahmen des IDG verankert werden.

Eine allgemeine Begründungspflicht gilt im Kanton Zürich gemäss § 126 Abs. 1 StG auch für Entscheide nach dem *Steuergesetz.* Für Veranlagungsverfügungen («Einschätzungsentscheide») ist zwar gemäss § 126 Abs. 2 StG lediglich eine Mitteilungspflicht hinsichtlich von Abweichungen gegenüber der Steuererklärung erforderlich. Auch wenn demnach keine detaillierte Begründung notwendig ist, muss die Behörde dennoch in der Lage sein, auch KI-gestützte Einschätzungsentscheide sachlich nachzuvollziehen. Spätestens im Falle eines Rekurses muss sie die Gründe für die Entscheidung angeben können. Deshalb empfiehlt es sich, bei der Konzeption von KI-Anwendungen für das Einschätzungsverfahren die Begründungsmöglichkeit von Anfang an mit einzuplanen, um den Anforderungen des rechtlichen Gehörs zu genügen.

Der **Untersuchungsgrundsatz** zieht sodann weitere Anforderungen nach sich. Von zentraler Bedeutung ist – auch mit Blick auf die Verhinderung von Diskriminierung und aus datenschutzrechtlicher Sicht –, dass die vom KI-System genutzten Trainingsdaten und weiteren Daten vollständig, korrekt und soweit zur Eruierung der rechtserheblichen Tatsachen notwendig verfügbar sind. Empfohlen wird, diese Anforderung auf gesetzlicher Stufe zu verankern. Zu denken wäre an eine Konkretisierung in § 7 VRG. Je nach Konzeption der Überarbeitung des IDG wäre eine entsprechende Vorgabe auch dort denkbar. Zur Sicherstellung des Amtsermittlungsgrundsatzes kann es ferner notwendig sein, im Rahmen des spezifischen Fachgesetzes, in dessen Anwendungsbereich ein KI-System eingesetzt werden soll, die notwendige Rechtsgrundlage für den Zugriff auf vorhandene Datensammlungen zu schaffen. Dabei sind die datenschutzrechtlichen Vorgaben ebenfalls zu berücksichtigen.

Im Rahmen von vollautomatisierten Verfahren – wie dies etwa im Falle eines vollautomatisierten *steuerrechtlichen Einschätzungsverfahrens* denkbar wäre – verlagert sich die Sachverhaltsermittlung weitgehend auf die steuerpflichtige Person. Aufgrund der im Steuerverfahren ausgeprägten Mitwirkungspflicht und des Selbstdeklarationsprinzips ist dies nicht grundsätzlich problematisch. Dennoch hat die Behörde durch geeignete Massnahmen sicherzustellen, dass die Einschätzung rechts- und gesetzeskonform erfolgen kann. Nach dem Untersuchungsgrundsatz kommt ihr die Verantwortung zu, dem Einschätzungsverfahren korrekte Sachverhaltsdaten zugrunde zu legen. Dies müsste sie bei einer vollautomatisierten Einschätzung verstärkt berücksichtigen und durch geeignete Kontrollmassnahmen – z. B. durch den Einsatz von Risikomanagementsystemen – kompensieren.

Grosse Herausforderungen stellen sich beim staatlichen KI-Einsatz aufgrund des **Diskriminierungsverbots.** Angesichts der verschiedenen Quellen von Diskriminierung kann deren Verhinderung nicht alleinige Aufgabe der Rechtsetzung sein. Die Umsetzung des Diskriminierungsverbots ist vielmehr eine Frage der Rechtsanwendung, wobei neben rechtlichen insbesondere organisatorische und technische Aspekte zu berücksichtigen sind. Eine der Diskriminierungsquellen bilden unrichtige Daten. Deshalb muss die Verwaltung – nicht nur als Ausfluss des Untersuchungsgrundsatzes und aufgrund datenschutzrechtlicher Vorgaben – sicherstellen, dass die genutzten Trainingsdaten und Sachverhaltsdaten korrekt sind und nur Daten genutzt werden, die für das entsprechende Verfahren geeignet sind. Wie bereits erwähnt, wäre die Verankerung einer entsprechenden Vorgabe sowohl im VRG als auch im IDG denkbar. Eine weitere Massnahme zur Verhinderung von Diskriminierung bei KI-Anwendungen, die zur Entscheidungsunterstützung eingesetzt werden, ist die Sicherstellung, dass die Sachbearbeiterin bzw. der Sachbearbeiter über die notwendigen Kenntnisse und Kompetenzen verfügt, um im Einzelfall eine vom diskriminierenden Vorschlag abweichende Entscheidung zu treffen. Zu den weiteren Massnahmen, die zur Verhinderung von Diskriminierung beim Einsatz von maschinellem Lernen vorgeschlagen werden, zählen die Nutzung von Kontrollalgorithmen oder die Beauftragung einer Drittorganisation bzw. staatlichen Institution mit der Durchführung regelmässiger Kontrollen. Auch die verschiedenen Optionen zur Herstellung von Transparenz dienen letztlich der Kontrolle von KI-Systemen und damit der Verhinderung von diskriminierenden Entscheiden.

Von zentraler Bedeutung ist das Diskriminierungsverbot bei einem potenziellen KI-Einsatz im Bereich von *Sozialversicherungsverfahren.* Dies liegt insbesondere daran, dass Entscheidungen über Sozialleistungen häufig auf besonders sensitiven Informationen zu einer Person wie etwa Gesundheitsdaten oder den Familienverhältnissen basieren. Zum einen müssen bei KI-Anwendungen im Sozialversicherungsbereich präexistierende Diskriminierungsmuster in den Trainingsdaten und der Systemarchitektur ausgeschlossen werden, damit diese nicht reproduziert werden. Zum anderen müssen eine genügende Repräsentation aller Bevölkerungsgruppen sowie Kontrollen sichergestellt werden, um das Anknüpfen an allfällige Proxies so früh wie möglich zu entdecken. Werden KI-Anwendungen eingesetzt, um Missbrauchsrisiken einzuschätzen, muss ein besonderes Augenmerk auf eventuelle Feedback-Loops geworfen werden.

Aus dem Recht auf **informationelle Selbstbestimmung** bzw. den **datenschutzrechtlichen Grundsätzen** ergeben sich weitere Anforderungen, da KI-Systeme auf grosse Datenmengen angewiesen sind, um sinnvoll eingesetzt werden zu können, und dabei häufig Personendaten bearbeitet werden. Hier kommen im Kontext von KI-Anwendungen dem Grundsatz der Datenrichtigkeit und der Herstellung von Transparenz besondere Bedeutung zu. Eine Verstärkung bzw. Ausweitung dieser beiden Grundsätze könnte mittels Ergänzungen in § 7 IDG bzw. § 12 IDG erreicht werden. Dabei ist an die Notwendigkeit korrekter, aktueller und vollständiger Daten zu denken, da von unrichtigen Daten im Rahmen von KI-Anwendungen auch ein erhöhtes Diskriminierungspotenzial ausgeht. Die skizzierten Anforderungen beziehen sich auf alle in KI-Systemen genutzten Daten und damit sowohl auf Sachdaten als auch auf Personendaten. Zu prüfen wäre ferner, ob § 12 IDG um eine Informationspflicht hinsichtlich einer automatisierten bzw. KI-gestützten Datenbearbeitung ergänzt werden soll. Das Ziel einer solchen Bestimmung wäre, die betroffene Person darüber zu informieren, dass ihre Daten automatisiert bzw. mithilfe von KI-Anwendungen bearbeitet werden und der daraus resultierende Entscheid Rechtswirkungen für sie entfaltet. Damit würde die vorgeschlagene Verankerung des Rechts auf vorgängige Äusserung im VRG ergänzt. Die Information würde es der betroffenen Person ermöglichen, ihr Recht auf vorgängige Äusserung wahrzunehmen.

Wo ein **Ermessens- oder Beurteilungsspielraum** besteht, sollte von einer vollautomatisierten Bearbeitung grundsätzlich abgesehen werden. Ausnahmen können gerechtfertigt sein, wenn der Ermessens- bzw. Beurteilungsspielraum der Behörde in einer Verwaltungsverordnung eingeschränkt wird. Eine weitere Herausforderung ergibt sich in vollautomatisierten Verfahren hinsichtlich des Anordnungsbegriffs. Aus Gründen der Rechts-

klarheit und -sicherheit empfiehlt es sich, den **Anordnungsbegriff** im VRG mindestens dahingehend zu klären, dass vollautomatisiert erlassene Entscheide ebenfalls als Anordnungen zu qualifizieren sind und damit Rechtskraft entfalten, aber auch angefochten werden können.

Die Herstellung von **Transparenz** beim staatlichen KI-Einsatz ist nicht nur unter dem Blickwinkel individueller Kontrollmöglichkeiten von Einzelentscheiden, sondern auch in Bezug auf eine allgemeine Kontrolle – etwa durch die Zivilgesellschaft – zu diskutieren. Dabei sind verschiedene Ansatzpunkte vorstellbar, wie die Transparenz zur Ermöglichung von Kontrolle rechtlich konkretisiert werden könnte. Denkbar wäre etwa die Schaffung eines öffentlich zugänglichen Registers, aus dem ersichtlich wird, in welchen Bereichen die öffentliche Verwaltung KI-Systeme einsetzt, und das u. a. über die Art und Herkunft der bearbeiteten Sach- und Personendaten, die Rechtsgrundlage, den Zweck und die Mittel der Bearbeitung, das verantwortliche Organ, die KI-Anwendung und deren Logik sowie diejenigen Akteure, die an der Entwicklung des Systems mitgewirkt haben, Auskunft gibt. Eine weitere mögliche Herangehensweise zur Herstellung von Transparenz findet sich in dieser Studie in Kapitel 4. Zu diskutieren wäre, wie und wo die Herstellung von Transparenz mittels der in Kapitel 4 vorgeschlagenen Checklisten bzw. der Erstellung eines Transparenzberichts rechtlich zu verankern wäre.

Abschliessend ist darauf hinzuweisen, dass beim Einsatz von Chatbots der **Vertrauensschutz** nach Art. 9 BV zu bedenken ist. In den Fällen, in denen ein Chatbot eine individualisierte, auf einen konkreten Sachverhalt bezogene Information ausgibt, liegt grundsätzlich eine taugliche Vertrauensgrundlage vor. Doch auch die Wiedergabe von allgemeinen, sozusagen «generell-abstrakten» Informationen durch Chatbots sollte eine taugliche Vertrauensgrundlage darstellen. Dies entspricht einerseits dem Bedürfnis der Privaten nach verlässlichen Informationen über die Rechtslage. Anderseits ist es auch im Interesse der Verwaltungsbehörden: Wenn den Informationen, welche Chatbots zur Verfügung stellen, nicht vertraut werden kann, dann werden diese von der Bevölkerung nicht als sinnvolle Alternative zu bestehenden Kommunikationskanälen akzeptiert werden, und ihr Nutzen für die Verwaltung wird beschränkt bleiben.

Kapitel 4

Ethisch vertretbarer Einsatz von KI

Michele Loi
Matthias Spielkamp
Anna Mätzener

A. Einführung

I. Ethische Richtlinien für den öffentlichen Sektor

In den vergangenen Jahren gab es einen regelrechten Wettlauf um die Entwicklung von ethischen Richtlinien zum Einsatz KI-basierter Systeme. Weltweit veröffentlichten Unternehmen und Unternehmensverbände, Organisationen der Zivilgesellschaft, Interessen- und Berufsverbände, Regierungen, Behörden und überstaatliche Institutionen Handlungsempfehlungen zum Umgang mit künstlicher Intelligenz. Allein im AI Ethics Guidelines Global Inventory[582], das AlgorithmWatch zusammengestellt hat, sind mehr als 160 solcher Dokumente gesammelt.

Die intensive Debatte um eine «KI-Ethik» hat unterschiedliche Reaktionen ausgelöst. Zum einen wurde das Interesse des Privatsektors misstrauisch beäugt, da Kritiker befürchten, dass dadurch entweder ein soziales Problem in ein technisches verwandelt wird oder Unternehmen versuchen, durch Selbstregulierung strengere Gesetze zu vermeiden. Ausserdem wird bisweilen kritisiert, dass Ethikrichtlinien anders als Gesetze nicht demokratisch legitimiert sind. Zum anderen stellen Wissenschaftlerinnen und Wissenschaftler die Frage, ob Einschätzungen dazu, was ein angemessener Einsatz von KI ist und welche Prinzipien die Entwicklung von KI bestimmen werden, konvergieren und sich somit eine Art gemeinsames Verständnis oder gemeinsame Erwartungen an den richtigen Umgang mit KI-basierten Systemen herausbilden.[583]

Im Rahmen dieser Untersuchung treten jene Kritikpunkte, die sich auf den Einsatz von KI-Ethikrichtlinien im Privatsektor beziehen, in den Hintergrund, da der Fokus der Untersuchung auf dem KI-Einsatz in der öffentlichen Verwaltung liegt. Da es bereits eine relevante Anzahl an Handlungsanleitungen gibt, die explizit für die öffentliche Verwaltung als Adressatin verfasst sind, konzentriert sich die vorliegende Studie auf diese Dokumente. Ethikrichtlinien, die sich an alle Entwicklerinnen und Entwickler bzw. alle Anwenderinnen und Anwender richten, werden mithin aus der Untersuchung ausgeklammert.

Dasselbe gilt für sektorspezifische Empfehlungen und Regelungen – unabhängig davon, ob sie sich an Private oder an die öffentliche Verwaltung richten. Gesetze, die Anwendungen von algorithmischen Verfahren regulieren, gibt es bereits seit Langem vom US-amerikanischen Code of Federal Regulations, Section 255.4 – Display of information[584] bis hin zur Richtlinie 2014/65/EU des europäischen Parlaments und des Rates vom 15. Mai 2014 über Märkte für Finanzinstrumente sowie zur Änderung der Richtlinien 2002/92/EG und 2011/61/EU[585]. Alle diese gesetzlichen Regulierungen hier zu betrachten, wäre einerseits aus Ressourcengründen nicht möglich. Anderseits wäre dies auch nicht sinnvoll, weil es sich um sektorspezifische Regelungen handelt, die wenig zur Beantwortung der Frage beitragen können, welche generellen Aspekte die (kantonale) öffentliche Verwaltung beim Einsatz von KI-basierten bzw. automatisierten Entscheidungssystemen beachten sollte. Deshalb werden die genannten Richtlinien und Regelungen hier grundsätzlich ausgeklammert.

Im Einzelnen werden die folgenden Richtlinien einbezogen:

Supranationale Richtlinien/verschiedene Akteure

1. Article 29 Data Protection Working Party
   Guidelines on Automated individual decision-making and Profiling for the purposes of Regulation 2016/679[586]
2. Europarat
   Discrimination, artificial intelligence, and algorithmic decision-making (insbesondere S. 29: Public sector bodies)[587]
3. Europarat
   Recommendation CM/Rec(2020)1 of the Committee of Ministers to member States on the human rights impacts of algorithmic systems (insbesondere Anhang A, Nr. 11 und 12[588])
4. Europarat
   European ethical Charter on the use of Artificial Intelligence in judicial systems and their environment[589]
5. Dataethics
   Data Ethics in Public Procurement[590]
6. World Economic Forum
   Unlocking Public Sector AI: AI Procurement in a Box[591]
7. Cities Coalition for Digital Rights
   Transparency, accountability, and non-discrimination of data, content and algorithms[592]
8. AI Now Institute, City of Amsterdam, City of Helsinki, Mozilla Foundation, Nesta
   Using procurement instruments to ensure trustworthy AI[593]

582 https://inventory.algorithmwatch.org/about.
583 Jobin/Ienca/Vayena, 2019.
584 «Each [airline reservation] system shall provide to any person upon request the current criteria used in editing and ordering flights for the integrated displays and the weight given to each criterion and the specifications used by the system's programmers in constructing the algorithm.»
585 Hier wird «hochfrequente algorithmische Handelstechnik» als algorithmische Handelstechnik definiert, die u. a. «gekennzeichnet ist durch […] b) die Entscheidung des Systems über die Einleitung, das Erzeugen, das Weiterleiten oder die Ausführung eines Auftrags ohne menschliche Intervention».
586 https://ec.europa.eu/newsroom/article29/document.cfm?action=display&doc_id=49826.
587 https://rm.coe.int/discrimination-artificial-intelligence-and-algorithmic-decision-making/1680925d73.
588 https://rm.coe.int/09000016809e1154.
589 https://rm.coe.int/ethical-charter-en-for-publication-4-december-2018/16808f699c.
590 https://dataethics.eu/publicprocurement/.
591 http://www3.weforum.org/docs/WEF_AI_Procurement_in_a_Box_Project_Overview_2020.pdf.
592 https://citiesfordigitalrights.org/#declaration.
593 https://assets.mofoprod.net/network/documents/Using_procurement_instruments_to_ensure_trustworthy_AI.pdf.

9. ePaństwo Foundation
   alGOVrithms: The Usage of Automated Decision Making – Policy Recommendations For Decision Makers[594]

Nationale Richtlinien

10. Australien: Commonwealth Ombudsman
    Automated decision-making better practice guide[595]
11. Kanada: Government of Canada
    Directive on Automated Decision-Making[596]
12. Deutschland: Kompetenzzentrum Öffentliche IT
    KI im Behördeneinsatz: Erfahrungen und Empfehlungen[597]
13. Neuseeland: Government of New Zealand
    Algorithm charter for Aotearoa New Zealand[598]
14. Schweiz: Bundesrat
    Leitlinien «Künstliche Intelligenz» für den Bund[599]
15. Grossbritannien: Government Digital Service and Office for Artificial Intelligence UK
    A guide to using artificial intelligence in the public sector[600]
16. Grossbritannien: National Health Service
    Code of conduct for data-driven health and care technology[601]
17. Vereinigte Staaten von Amerika: New York City
    Automated Decision Systems Task Force Report[602]

II. Zweistufiges Beurteilungsverfahren

Die erwähnten Richtlinien schaffen zwar Eckpunkte und geben Hinweise für einen ethisch vertretbaren Einsatz von KI, sind für die konkrete Umsetzung von KI-Vorhaben in der öffentlichen Verwaltung aber nicht ohne weitere Implementierungsschritte umsetzbar. In diesem Kapitel wird deshalb ein zweistufiges Beurteilungsverfahren entwickelt, das dazu genutzt werden kann, ethische Auswirkungen eines KI-Systems zu erkennen und darauf aufbauend Transparenz über das System herzustellen.

Im Folgenden wird zunächst dargelegt, warum in diesem Bericht zum einen bestimmte *ethische Grundsätze* und zum anderen bestimmte *instrumentelle Grundsätze* als Grundlage einer Beurteilung von KI-Systemen herangezogen werden. Anschliessend wird in den Abschnitten B. I. und B. II. zu jedem der sieben identifizierten Grundsätze detailliert dargestellt, welche Fragen aus diesen Grundsätzen folgen und bei der Beurteilung entsprechend zu beantworten sind. Die Fragen selbst werden in zwei verschiedenen Checklisten in Abschnitt C. dargestellt. Die *Triage-Checkliste für KI-Systeme* (Checkliste 1) hilft bei der Feststellung, welche ethischen Transparenzfragen vor und während der Durchführung eines KI-Projekts im Detail zu dokumentieren sind. Die *Checkliste Transparenzbericht* (Checkliste 2) dient als Leitfaden für die Erstellung eines ausführlichen Transparenzberichts. In den folgenden Ausführungen beziehen sich die in Klammern gesetzten Hinweise jeweils auf die Fragen aus den beiden Checklisten in den Abschnitten C. II. und C. III. Wie das skizzierte Beurteilungsverfahren angewendet werden kann, wird sodann am fiktiven Beispiel eines Swiss-COMPAS-Risikobewertungssystems für Straftäterinnen und Straftäter veranschaulicht (Abschnitt C. IV.). Ein Flussdiagramm gibt schliesslich einen Überblick über das gesamte Vorgehen (Abschnitt C. V.).

Als Grundlage des hier entwickelten Beurteilungsverfahrens dienen die «Ethik-Leitlinien für eine vertrauenswürdige KI» der hochrangigen Expertengruppe für Künstliche Intelligenz, die von der Europäischen Kommission eingesetzt wurde.[603] Dieses Dokument stellt allerdings lediglich eine Vereinfachung derjenigen Grundwerte dar, die in anderen Richtlinien erarbeitet wurden, und muss daher ergänzt werden.

Zu diesem Zweck werden weitere Zusammenfassungen von Richtlinien herangezogen. Die umfassendste Analyse bisher veröffentlichter KI-Richtlinien[604] enthält eine Liste von elf verschiedenen Grundsätzen, die in den analysierten Richtlinien als gemeinsamer Nenner enthalten sind. Von diesen elf Grundsätzen überschneiden sich einige **(Nicht-Missbräuchlichkeit oder Schadensverhütung, Gerechtigkeit und Unparteilichkeit [Fairness]** sowie **Freiheit und Autonomie)** mit den ethischen Grundsätzen der Richtlinien der Expertengruppe. Zwei der elf ethischen Grundsätze sind so nicht in den EU-Leitlinien enthalten: Benefizienz und Achtung der Würde. In den elf Grundsätzen sind zudem instrumentelle, technische oder verfahrenstechnische Anforderungen **(Transparenz** und **Verantwortung/Rechenschaftspflicht)** enthalten, die in den EU-Richtlinien als «Schlüsselanforderungen für vertrauenswürdige KI» bezeichnet werden. Ein weiterer Grundsatz, die **Erklärbarkeit,** ist in den EU-Leitlinien ebenfalls enthalten, wird aber – was auch plausibel ist – als Grundlage für andere Grundsätze wie die Implementierungsanforderungen betrachtet.[605]

Der Grundsatz der **Benefizienz** ist nicht in den EU-Leitlinien enthalten, aber sie ist nicht nur in der erwähnten Zusammenfassung von Jobin et al.[606] aufgeführt, sondern auch im am weitesten verbreiteten Grundgerüst ethischer Grundsätze, jenem der biomedizinischen Ethik, enthalten.[607]

Als weitere Grundsätze werden als Makrokategorien die **Kontrolle,** die **Transparenz** und die **Rechenschaftspflicht** betrachtet. Loi et al. verstehen darunter Massnahmen in den KI-Ethikrichtlinien, die *instrumentelle und verfahrenstechnische Anforderungen* umfassen.[608] Im Folgenden werden daher Kontrolle, Transparenz und Rechenschaftspflicht als *«instrumentelle Grundsätze»* bezeichnet.[609]

594 https://epf.org.pl/en/wp-content/uploads/sites/3/2019/05/alGOVrithms-Recommendations-EN.pdf.
595 https://consult.industry.gov.au/strategic-policy/artificial-intelligence-ethics-framework/supporting_documents/ArtificialIntelligenceethicsframeworkdiscussionpaper.pdf.
596 https://www.tbs-sct.gc.ca/pol/doc-eng.aspx?id=32592.
597 https://oeffentliche-it.de/documents/10181/14412/KI+im+Beh%C3%B6rdeneinsatz+-+Erfahrungen+und+Empfehlungen.
598 https://data.govt.nz/use-data/data-ethics/government-algorithm-transparency-and-accountability/algorithm-charter.
599 https://www.sbfi.admin.ch/dam/sbfi/de/dokumente/2020/11/leitlinie_ki.pdf.download.pdf/Leitlinien%20K%C3%BCnstliche%20Intelligenz%20-%20DE.pdf.
600 https://www.gov.uk/government/collections/a-guide-to-using-artificial-intelligence-in-the-public-sector.
601 https://www.gov.uk/government/publications/code-of-conduct-for-data-driven-health-and-care-technology/initial-code-of-conduct-for-data-driven-health-and-care-technology.
602 https://www1.nyc.gov/assets/adstaskforce/downloads/pdf/ADS-Report-11192019.pdf.
603 Ethics Guidelines for Trustworthy AI, 2019.
604 Jobin/Ienca/Vayena, 2019.
605 Floridi/Cowls, 2019.
606 Jobin/Ienca/Vayena, 2019.
607 Beauchamp/Childress, 2008.
608 Loi/Heitz/Christen, 2020.
609 Loi, 2020.

Die Analyse von 18 weiteren Dokumenten zum Einsatz von KI im öffentlichen Sektor (siehe A. I.) ergibt ebenfalls ethische und instrumentelle Grundsätze, die mit diesem Gerüst vereinbar sind. Im Folgenden konzentriert sich die Analyse daher auf einen Rahmen von sieben Werten:

- drei der vier ethischen Grundsätze, die in den EU-Leitlinien enthalten sind, d. h. die Achtung der **menschlichen Autonomie,** die **Schadensvermeidung** sowie die **Gerechtigkeit** oder **Unparteilichkeit** (Fairness);
- die **Benefizienz** als weithin anerkannter ethischer Grundsatz sowie
- die drei instrumentellen Grundsätze der **Kontrolle, Transparenz** und **Rechenschaftspflicht,** die technische, organisatorische und aufsichtsrechtliche Anforderungen zusammenfassen, die üblicherweise in praktischen Richtlinien zur KI-Ethik enthalten sind.

Im Gegensatz zu den Richtlinien der EU-Expertengruppe wird die *Erklärbarkeit* nicht als eigenständiger Grundsatz, sondern als Bestandteil anderer instrumenteller Grundsätze betrachtet.[610]

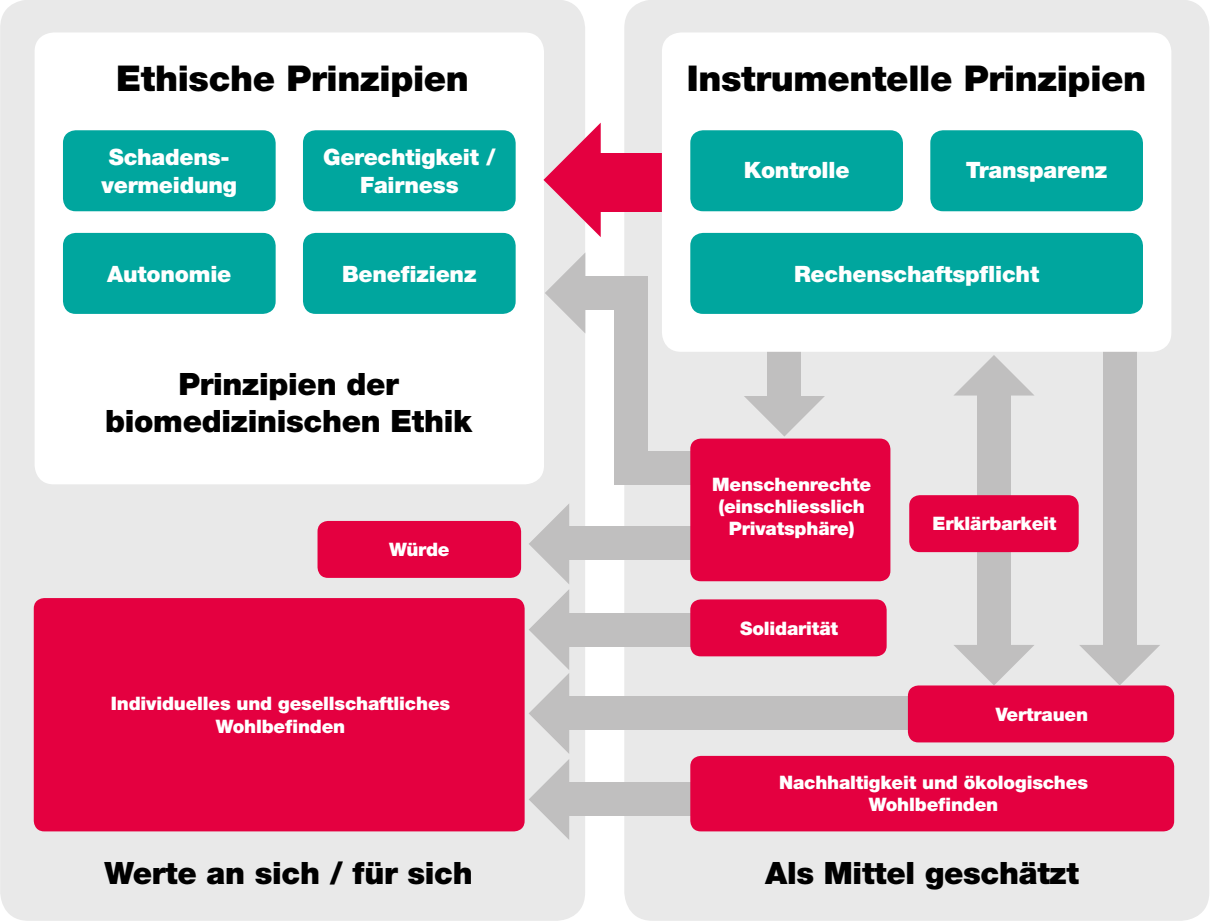


Abb. 1 (eigene Darstellung): Die wichtigsten Grundsätze und Werte in den ethischen Richtlinien zur KI. Grün eingefärbt sind die in dieser Untersuchung berücksichtigten Grundsätze, in Rot sind Werte und Grundsätze aus anderen Richtlinien. Der Pfeil bedeutet «ist erforderlich für».

Im Folgenden wird der ethische Rahmen, der die Grundlage der in dieser Untersuchung entwickelten praktischen Empfehlungen und Checklisten bildet, näher ausgeführt. Für jeden der sieben berücksichtigten Werte wird in Fussnoten auf analoge Konzepte in bestehenden Richtlinien für Anwendungen von KI-Systemen im öffentlichen Sektor verwiesen. Darüber hinaus wird auf die Fragen in den Checklisten 1 und 2 verwiesen, die direkt aus dieser Analyse abgeleitet werden.

Für die Schweiz sind die Leitlinien des Schweizer Bundesrates zu KI von besonderem Interesse. Die darin enthaltene Idee, «den Menschen in den Mittelpunkt zu stellen», findet sich in fünf ethischen Grundsätzen wieder, die hier ebenfalls berücksichtigt werden: die Benefizienz als Förderung des Wohlergehens, die Gerechtigkeit als Achtung der Grundrechte und Vermeidung von Diskriminierung, die Autonomie als Selbstbestimmung, die Schadensvermeidung als Schutz der Privatsphäre und personenbezogener Daten sowie die Würde.[611] Somit kann festgehalten werden, dass die folgenden Erwägungen mit den KI-Leitlinien des Bundesrates grundsätzlich kompatibel sind.

610 Vgl. auch Jobin/Ienca/Vayena, 2019 und Loi/Heitz/Christen, 2020.
611 Bundesrat Leitlinien KI 2020, Leitlinie.

B. Sieben Grundsätze

I. Ethische Grundsätze

1. Schadensvermeidung

Dies ist das Prinzip «Niemandem einen Schaden zufügen». Zivile KI-Systeme dürfen nicht so konzipiert sein, dass sie Menschen schaden oder täuschen, und sie sollten so implementiert werden, dass negative Ergebnisse minimiert werden.[612] Schaden zu vermeiden, bedeutet in erster Linie in der einfachsten Form, Schmerzen und Unbehagen für Menschen zu verhindern. Im weiteren Sinne umfasst Schaden die Verletzung der Privatsphäre[613] (Frage 1.1) und der Rechte (Fragen 1.4 und 1.5) einschliesslich der Menschenrechte.[614] Die Vermeidung von Schäden ist damit verbunden, Sicherheit[615] und Nachhaltigkeit zu fördern sowie allgemeine, technische und institutionelle Schutzmassnahmen aufzubauen.[616] Diese werden oft als Vertrauenselement oder vertrauenswürdige Technologie bezeichnet.[617] Zur Vermeidung von Schäden gehören auch die Verhinderung und Steuerung von Risiken. Häufig erfordert das, zu gewährleisten, dass die Systeme zuverlässig[618] und berechenbar[619] sind. Schadensvermeidung umfasst ebenfalls, eine breite ökologische und soziale Nachhaltigkeit sicherzustellen (Frage 1.10)[620], denn nicht nachhaltige Praktiken führen letztendlich zu menschlichem Schaden und zu einer Verletzung eigener Interessen. Darüber hinaus kann hier die Nachhaltigkeit eines *soziotechnischen Systems* berücksichtigt werden, das auf einer zuverlässigen Technologie basiert.[621] Dies umfasst auch die Gewährleistung der Cybersicherheit (Frage 1.2) einschliesslich des Schutzes der Vertraulichkeit, Integrität und Verfügbarkeit von Informationen.[622]

2. Gerechtigkeit und Fairness

Das ethische Ziel von Gerechtigkeit und Fairness umfasst die Wahrung von sechs Dimensionen ethischer Werte. Die erste (in Bezug darauf, wie oft sie in den KI-Ethikrichtlinien erwähnt wird) ist der Schutz vor einer *ungerechten Diskriminierung* und *nicht zu rechtfertigenden Voreingenommenheit.*[623] Diese Dimension der Fairness gilt für verschiedene Elemente der Datenverarbeitung: Wenn das KI-System soziale oder demografische Daten verarbeitet, sollte es so ausgelegt werden, dass ein Mindestmass an Schadensvermeidung in Bezug auf Diskriminierung erreicht wird. Um dies zu tun,

- sollten nur faire und gerechte Datensätze verwendet werden (Datengerechtigkeit),
- sollten angemessene Funktionen, Prozesse und analytische Strukturen in die Modellarchitektur aufgenommen werden (Design-Fairness),
- sollte verhindert werden, dass das System diskriminierende Auswirkungen hat (Ergebnisgerechtigkeit) sowie
- das System unvoreingenommen implementiert werden (Implementierungsgerechtigkeit).[624]

Wahrscheinlich ist es nicht möglich, jeden Bias zu vermeiden. Voreingenommenheit kann auf unzureichende Daten beim Trainieren des Modells zurückzuführen sein,[625] aber es kann sich schwierig oder unmöglich gestalten, repräsentative Daten zum Trainieren von Modellen zu finden – auch aufgrund von Datenschutzbestimmungen und (paradoxerweise) Antidiskriminierungsvorgaben, die es erschweren, beispielsweise Daten über sexuelle Orientierung, Geschlechter, ethnische Gruppen oder religiösen Status zu sammeln und zu verarbeiten. Aber selbst dann, wenn die Daten völlig ausreichen, um die Gesellschaft in allen ihren Facetten darzustellen, können Entscheidungen, die auf statistischen Verallgemeinerungen beruhen, als ungerechtfertigt voreingenommen angesehen werden – etwa, wenn sie für Personen mit Merkmalen, die mit denen aus privilegierteren Umständen übereinstimmen, günstiger sind. Das ist *insbesondere* dann der Fall, wenn Daten vorhanden sind, die für verschiedene gesellschaftliche Gruppen repräsentativ sind, da dies die Wahrscheinlichkeit erhöht, dass einige Daten (oder eine Kombination davon) als Stellvertreter für Alter, Geschlecht, Religion usw. fungieren. Auch wenn solche Daten, die *explizit* die Kategorien betreffen, die durch Antidiskriminierungsgesetze geschützt sind, nicht in dem Mix enthalten sind, wird jedes effiziente Verfahren maschinellen Lernens, das Algorithmen erzeugt, normalerweise lernen, Stellvertreter zu erkennen. Aus diesem Grund sind alle algorithmischen Schlussfolgerungen, die auf Techniken des statistischen Lernens aus Daten über Menschen basieren, in Bezug auf Bias und indirekte Diskriminierung potenziell moralisch problematisch. Daher muss diesen Algorithmen in der Checkliste (Frage 1.14) besondere Aufmerksamkeit gewidmet werden.

Darüber hinaus ist die Dimension der Ergebnis- und Anwendungsgerechtigkeit besonders wichtig, wenn Algorithmen den Wettbewerb in Politik und Wirtschaft beeinflussen (Fragen 1.12 und 1.13). Wettbewerbsprozesse sind von enormer Bedeutung, da sich die Gesellschaft auf einen fairen Wettbewerb in Markt und Politik stützt, um zu entscheiden, wie soziale Ressourcen und Chancen in der Gesellschaft auf eine Weise verteilt werden müssen, die insgesamt als verfahrensgerecht betrachtet werden kann.[626]

612 Dawson et al., 2020, Principle 2.
613 Dawson et al., 2020, Principle 4.
614 Council of Europe, CM/Rec (2020)1, Principle 3.2; Government of Canada, 2019, Appendix B (risk evaluation framework).
615 Council of Europe, CM/Rec (2020)1; Bundesrat Leitlinien KI 2020, Leitlinie 5; Engelmann/Puntschuh, 2020, 5. Sicherheit.
616 Council of Europe, CM/Rec (2020)1; Engelmann/Puntschuh, 2020, 5. Sicherheit.
617 Leslie, 2019; Ethics Guidelines for Trustworthy AI, 2019.
618 Government of New Zealand, 2020.
619 Vgl. Government of Canada, 2019, Guidelines and laws that require a risk assessment may include: «the rights of individuals or communities, the health or well-being of individuals or communities, the economic interests of individuals, entities, or communities, the ongoing sustainability of an ecosystem».
620 Council of Europe, CM/Rec (2020)1, Principle 6.3.
621 Government UK, 2019.
622 Council of Europe, 2020, Ethical Charter AI, Principle of quality and security.; Dawson et al., 2020, Principle 4; Government of Canada, 2019, 6.3.7; Engelmann/Puntschuh, 2020, 6. Datenhaltung und -qualität.
623 Council of Europe, 2020, Ethical Charter AI, Principle of non-discrimination; Dawson et al., 2020, Principle 5; Government of New Zealand, 2020, Fairness and Justice; Leslie, 2019; New York City, 2019.
624 Government UK, 2019.
625 Council of Europe, CM/Rec (2020)1, Testing on personal data.
626 Rawls, 1999.

Zweitens gehen die Fairnessanforderungen über die Ethik hinaus und beziehen die *Legalität* mit ein, um sicherzustellen, dass Algorithmen nicht gegen bestehende Gesetze einschliesslich der gesetzlichen Rechte verstossen.[627]
Drittens ist die *Achtung aller Rechte,* ob sie im positiven Recht anerkannt sind oder nicht, einschliesslich der Menschenrechte und Persönlichkeitsrechte beinhaltet.[628]
Viertens ist der Wert von *Gleichheit*[629], *Inklusion* und *Solidarität* erfasst (obwohl dies ein umstrittenerer Wertekanon sein kann und sehr kontextabhängig ist). Das Erfordernis der Inklusion, das Elemente der öffentlichen Beteiligung umfasst, scheint in Leitlinien für den öffentlichen Sektor stärker vertreten zu werden[630] als in Leitlinien für die Wissenschaft oder für private Unternehmen, die zumeist früher entwickelt worden sind.
Fünftens sind *Entschädigungen* und *Rechtsmittel* eingeschlossen (Fragen 1.7, 1.8, 1.9), wenn eine Rechtsverletzung nachgewiesen werden kann.[631]
Sechstens geht es um die Frage der *prozessualen Ordnungsmässigkeit,* die in den untersuchten Leitlinien nicht erwähnt wurde, aber in der Literatur berücksichtigt wird. Einige KI-Systeme aktualisieren ihre Modelle kontinuierlich basierend auf neuen Daten, um das Ziel, für das sie programmiert wurden, besser zu erreichen. Ein Nebeneffekt eines kontinuierlich aktualisierten Modells besteht darin, dass es unterschiedliche Ergebnisse für dieselben Eingaben erzeugen kann, wenn dieselben Eingaben vor oder nach einer Modellaktualisierung verarbeitet werden. Dies bedeutet, dass zwei Personen mit denselben Merkmalen (Eingaben) möglicherweise unterschiedliche Entscheidungen (Ausgaben) erhalten, die davon abhängen, wann der Algorithmus ihre personenbezogenen Daten verarbeitet (vor oder nach einer Modellaktualisierung).[632] Dies kann das Recht der Personen auf rechtsgleiche Behandlung verletzen (Frage 1.15).

3. Autonomie

Die Förderung der Autonomie bedeutet, dass Einzelne Entscheidungen über ihr Leben treffen können, die ihre eigenen sind und die ihnen nicht von anderen auferlegt oder von ihnen manipuliert werden. Eine Entscheidung auf der Grundlage unzureichender Informationen oder einer Täuschung gilt nicht als autonom. Das Ziel der menschlichen Autonomie hängt hauptsächlich mit dem prozessualen Erfordernis der Transparenz zusammen, das im Abschnitt II. der Leitlinie beschrieben wird. Transparenz bedeutet, ausreichende Informationen bereitzustellen und die Täuschung von Personen zu vermeiden, die mit den Algorithmen interagieren, was autonome Entscheidungen ermöglicht.
Die bekannteste (und vielleicht am wenigsten geschätzte) Implementierung von Entscheidungsautonomie im digitalen Alltag betrifft personenbezogene Daten. Individuen müssen darüber informiert werden, was mit ihren persönlichen Daten geschehen kann, damit sie ihre Zustimmung dazu geben können, dass ihre Daten auf die eine oder andere Weise verwendet werden.[633] Dies gilt insbesondere im Zusammenhang mit experimentellen Technologien.[634]
Ein weiterer Aspekt der Autonomie ist die «Fähigkeit, unfaire, voreingenommene oder diskriminierende Systeme infrage zu stellen und zu ändern»[635], über die Bürgerinnen und Bürger nur verfügen, wenn sie «verständliche und genaue Informationen über die technologischen, algorithmischen und Künstlichen Intelligenzsysteme erhalten, die sich auf ihr Leben auswirken».[636] Die Anfechtung kann *sowohl* als wertvoll an sich als Element menschlicher Autonomie *als auch* als instrumentell wertvoll als eine Form der Kontrolle des Algorithmus und ein Weg zur Förderung der Rechenschaftspflicht angesehen werden (Letzteres wird in Abschnitt II. dieser Leitlinien betrachtet).
Ein weiterer Aspekt der Autonomie ist die Möglichkeit, die zu verwendenden digitalen Dienste auszuwählen oder deren Einsatz ganz zu vermeiden,[637] insbesondere dann, wenn es sich um experimentelle Dienste handelt[638] (Frage 1.6).
Autonomie hat auch eine kollektive Dimension: Sie ist die Fähigkeit der Bürger, gemeinsam Entscheidungen über ihr kollektives Schicksal als Gemeinschaft zu treffen. Diese kollektive Dimension der Autonomie ist in KI-Richtlinien nicht weit verbreitet,[639] scheint jedoch für öffentliche digitale Infrastrukturen wie die von Smart Cities sehr wichtig zu sein.[640]
Das ethische Erfordernis, die Grundrechte zu respektieren,[641] fördert implizit die Autonomie, da viele dieser Rechte (die typischerweise die Menschenrechte respektieren und in Demokratien verfassungsrechtlich geschützt sind) die Autonomie des Menschen schützen. Beispielsweise gewähren negative Rechte wie die Meinungs- oder Religionsfreiheit der individuellen Autonomie im politischen Raum und bei individueller und kollektiver Meinungsäusserung Schutz. Positive Rechte wie das Recht auf Gesundheitsversorgung und Bildung schützen die Autonomie, indem sie sicherstellen, dass der Einzelne über die Mittel verfügt, die er für ein unabhängiges Leben benötigt. Daher wird die Autonomie durch eine sozial nachhaltige KI gefördert.[642]
Schliesslich kann sich Autonomie in der KI-Ethik auf die Idee beziehen, dass das KI-System «unter Benutzerkontrolle» steht.[643] Das bedeutet, Algorithmen sollten verwendet werden, um die menschliche Entscheidungsfindung zu unterstützen, und nicht dazu dienen, sie vollständig zu ersetzen. Dies wird am plausibelsten als ein eingeschränktes Prinzip verstanden. Philosophisch gesehen kann man argumentieren, Automatisierung könne die Autonomie eher *erweitern* als reduzieren (und sie hat es auch getan), insofern die Automatisierung von Routineaufgaben dazu beigetragen hat, mehr Zeit und Ressourcen für den Menschen freizugeben, damit er sich mit intellektuell her-

627 Dawson et al., 2020, Principle 3; Government of New Zealand, 2020, Fairness and Justice.
628 Government of New Zealand, 2020, Fairness and Justice.
629 Government of New Zealand, 2020, Fairness and Justice.
630 Council of Europe, CM/Rec (2020)1, Barriers, Advancement of public benefits; Dataethical Thinkdotank, 2021, Universal design; Cities for Digital Rights, 2020, Participatory democracy, diversity and inclusion; Bundesrat Leitlinien KI 2020, Leitlinie 7.
631 Council of Europe, CM/Rec (2020)1, Effective remedies; Government of Canada, 2019, Principle 6.4.
632 Kroll et al., 2016/2017; Loi/Ferrario/Viganò, 2020.
633 Cities for Digital Rights, 2020, Privacy, data protection and security; WEF, 2020.
634 Council of Europe, CM/Rec (2020)1, computational experimentation.
635 Cities for Digital Rights, 2020, Transparency, accountability, and non-discrimination of data, content and algorithms.
636 Cities for Digital Rights, 2020, Transparency, accountability, and non-discrimination of data, content and algorithms.
637 Cities for Digital Rights, 2020, Open and ethical digital service standards.
638 Council of Europe, CM/Rec (2020)1, Computational experimentation.
639 Es wurde nur von Jobin et al., 2019, erwähnt.
640 Vgl. Cities for Digital Rights, 2020: «Everyone should have […] the ability collectively to engage with the city through open, participatory and transparent digital processes», Participatory Democracy, diversity and inclusion.
641 Council of Europe, CM/Rec (2020)1, Principle of respect for fundamental rights.
642 Council of Europe, CM/Rec (2020)1, Human-centric and sustainable innovation; Dataethical Thinkdotank, 2021, Human primacy.
643 Council of Europe, CM/Rec (2020)1, Principle under user control.

ausfordernden, kreativen oder emotional lohnenden Aufgaben befassen kann.[644] Das Problem der Autonomie ergibt sich aus KI-Systemen, die kognitiv anspruchsvollere Arten menschlicher Aktivitäten automatisieren sollen, wobei Menschen Befehle *von* den Maschinen entgegennehmen, anstatt ihnen Befehle zu erteilen.[645] Auf dem Spiel steht Autonomie als ethischer Wert daher möglicherweise bei all jenen Automatisierungsprojekten, bei denen KI-Systeme das menschliche Urteilsvermögen ersetzen sollen (Frage 1.16), und bei den Formen der Automatisierung, bei denen unklar ist, ob Benutzerinnen und Benutzer die KI-Entscheidungen ausreichend verstehen, damit diese ihre Autonomie unterstützen, statt sie durch KI-Systeme zu ersetzen (Fragen 1.17 und 1.18). Darüber hinaus kann die Öffentlichkeit die Kontrolle und damit die Autonomie über ihre Prozesse und Entscheidungen verlieren, wenn sie sich auf eine Infrastruktur stützen, die sich vollständig im Besitz von Dritten befindet und von ihnen abgeschottet wird (Frage 1.19). Dies ist ein aufkommendes Problem, das in früheren Überprüfungen nicht erkennbar war,[646] aber in Bezug auf die KI-Systeme des öffentlichen Sektors ziemlich wichtig erscheint.[647]

4. Benefizienz

Benefizienz ist wohl dasjenige Grundprinzip der Ethik, das in den KI-Richtlinien am wenigsten verbreitet ist. Ein plausibler Grund für die unzulängliche Beachtung der Benefizienz ist, dass die meisten Akteurinnen und Akteure, die sich mit KI-Systemen befassen, davon ausgehen, KI-Systeme könnten irgendwelche Vorteile bringen. Effizienz wird häufig als Grund für die Verwendung von KI genannt: Der gleiche Dienst kann für die gleiche Anzahl von Personen bereitgestellt werden, während weniger Ressourcen verwendet oder vorhandene Dienste können verbessert werden (z. B., indem sie genauere Ergebnisse liefern oder mit zusätzlichen Funktionen ausgestattet werden), während sie günstig und damit für die meisten zugänglich bleiben. Und doch ist die Möglichkeit, mithilfe von KI-Systemen *Gutes zu tun,* ethisch grundlegend, da eine Leitlinie für die ethische Verwendung von KI-Systemen zu stark auf Schadensverhütung und zu wenig auf die Schaffung von Nutzen ausgerichtet sein kann. Eine solche Leitlinie wird sich in den meisten Kontexten eher gegen die Einführung von KI-Systemen aussprechen, da Innovation an sich Risiken birgt. Wenn man den potenziellen Nutzen von Innovation vergisst, gibt es keinen Grund, *irgendein Risiko* einzugehen. Eine extreme *Risikovermeidung* ist jedoch nicht immer sinnvoll. Vielmehr sollte das mit Innovationen verbundene Risiko *gemanagt* werden. Beispielsweise haben KI-Systeme wie erwähnt das Potenzial, die Autonomie des Menschen zu verbessern, wenn sie dazu verwendet werden, Prozesse so zu automatisieren, dass menschliche Ressourcen freigesetzt werden, die besser an anderer Stelle eingesetzt werden. Glücklicherweise erwähnen einige Richtlinien, die sich an den öffentlichen Sektor richten, zumindest implizit die Benefizienz, indem sie auf den Nutzen von Innovationen hinweisen.[648] Zum Beispiel: «Erzeugt Nettonutzen. Das KI-System muss Vorteile für Menschen generieren, die mehr wiegen als die Kosten.»[649] Die Benefizienz wird implizit auch in diesen Leitlinien angesprochen, in denen die Förderung des menschlichen Wohlbefindens als übergeordnetes Ziel einer solchen Innovation angegeben ist.[650] Die Richtlinien für den öffentlichen Sektor betonen den Gedanken, dass der Nutzen der KI-Systeme ein *öffentlicher* sein sollte.[651]

Die Benefizienz wird in der Checkliste nicht ausdrücklich aufgeführt. Das Konzept eines Nettonutzens wird jedoch implizit im Transparenzbericht (Checkliste 2) erwähnt, insbesondere in den Fragen 2.1, 2.16 und 2.17, in denen erläutert werden muss, warum die Einführung von KI-Systemen *nützlich* ist und welche Nachweise dafür erbracht werden können. Darüber hinaus führt die Frage nach dem Schaden für das Gemeinwohl (Frage 1.10) (gemäss dem vorgeschlagenen Checklistenalgorithmus) zu einer transparenten Stakeholderanalyse (Frage 2.8) und einer transparenten Darstellung möglicher Kritik, die externe Stakeholder (Frage 2.20) gegen die Kosten-Nutzen-Analyse der öffentlichen Verwaltung vorbringen.

II. Instrumentelle und aufsichtsrechtliche Grundsätze

1. Kontrolle

Das prozessuale Erfordernis der Kontrolle ergibt sich aus der Analyse von 20 Richtlinien, die sich eher auf Handlungstypen als auf Handlungsziele konzentrierten.[652] In den ursprünglich analysierten Leitlinien erschien dieses Erfordernis nicht als eigenständige Anforderung, sondern eher als ein Element – eine gemeinsame Reihe von Handlungen –, das jeweils gleich häufig unter den Rubriken *Transparenz* und *Rechenschaftspflicht* aufgeführt ist. In der Tat umfasst die Kontrolle gemeinsame Aktivitäten, die sowohl für die Transparenz als auch für die Rechenschaftspflicht erforderlich sind: Man kann in Bezug auf Prozesse oder Ergebnisse, die man nicht kennt, nicht transparent sein und Verantwortung im positiven, vorausschauenden Sinne des Begriffs (nicht im Sinne von rückwirkender Schuld oder Haftung) umfasst, Prozesse so zu kontrollieren, dass sie zu den beabsichtigten Ergebnissen führen. Die Kontrolle umfasst alle Aktivitäten, die erforderlich sind, um allen zielbezogenen Aktivitäten *Robustheit* zu verleihen: vom Erreichen des von den Benutzerinnen und Benutzern vorgesehenen Ziels des KI-Systems (welches immer das sein mag) bis hin zur Gewährleistung, dass die anderen ethischen Ziele (Schadensvermeidung, Fairness und Autonomie) ebenfalls gefördert werden. Kontrolle erscheint moralisch neutral, weil ihr ethischer Wert rein *instrumentell* und *ungewiss* ist. Eine Reihe von Praktiken, mit denen Terroristen ein KI-System besser kontrollieren können, um Drohnenangriffe zu koordinieren, ist nützlich, aber nicht ethisch wertvoll, da das Ziel, das Terroristen verfolgen, als solches schlecht ist. Wenn das Ziel des oder der Handelnden jedoch ethisch positiv ist, ist das Fehlen von Kontrolle ethisch schlecht und nicht nur neutral, da gute Absichten ohne Kontrolle möglicherweise nicht das Gute erreichen, auf das sie abzielen, oder sogar unbeabsichtigt schädlich sein können.

644 Danaher, 2016a, Danaher, 2016b; Loi, 2015.
645 Danaher, 2016a, Danaher, 2016b; Loi, 2015.
646 Jobin/Ienca/Vayena, 2019.
647 Cities for Digital Rights, 2020, Participatory Democracy, diversity and inclusion and Open and ethical digital service standards; Council of Europe, CM/Rec (2020)1, Infrastructure.
648 Engelmann/Puntschuh, 2020, 2. Interne Veränderung und 3. Innovationsmarker.
649 Dawson et al., 2020.
650 Leslie, 2019; Government of New Zealand, 2020, Well-Being.
651 Council of Europe, CM/Rec (2020)1, Advancement of public benefit, Rights-promoting technology.
652 Loi/Heitz/Christen, 2020.

Da es sich bei Kontrolle um ein so vielseitiges Mittel handelt, ist es nicht überraschend, dass sie die häufigste Verfahrensanforderung ist, die in den Ethikrichtlinien für KI-Systeme enthalten ist.
Erstens umfasst die Kontrolle die Dokumentation von Prozessen und Ergebnissen sowie die *Aufzeichnung,*[653] *Prüfung*[654] und *Überwachung,*[655] welche die Daten über das liefern, was zu dokumentieren ist.[656] Die Dokumentation unterscheidet sich von der *Transparenz,* da sie mit *interner Kommunikation* einhergehen kann, z. B. zwischen Mitarbeiterinnen und Mitarbeitern innerhalb eines Data-Science-Teams oder eines gesamten Unternehmens,[657] ohne jedoch Benutzerinnen und Benutzern oder der breiten Öffentlichkeit ausreichende Transparenz zu bieten.
Zweitens beinhaltet die Kontrolle das *Messen, Erfassen, Bewerten*[658] und *Definieren von Standards*[659] *und Richtlinien* für alle KI-bezogenen Prozesse und Ergebnisse. Sie umfasst daher, zu *untersuchen,* was erforderlich ist, um *aussagekräftige* Standards und Massnahmen zu erzeugen, d. h. solche, die zu einem authentischen *Verständnis* und *Wissen* über die zu bewertenden Prozesse und Ergebnisse führen, die sowohl Produkte als auch Aspekte der Kontrolle über KI sind. KI-Systeme zu verstehen und zu kennen, erfordert, die Funktionsweise der Systeme zu *erklären.*[660] Daher sind die Ziele der *erklärbaren*[661] KI und der Wert der Erklärbarkeit in den Richtlinien der EU-Expertengruppe eine Facette (möglicherweise die wichtigste Facette) des Erfordernisses der prozessualen Kontrolle. KI-Systeme zu bewerten und einzuschätzen, erfordert nicht nur, Entwicklungsentscheidungen zu *rechtfertigen,*[662] sondern dies auch bei Fehlern, Vorurteilen und Güterabwägungen mit anderen moralischen Zielen zu tun, wenn sie unvermeidbar sind.
Drittens umfasst die Kontrolle die sozialen Aktivitäten, die erforderlich sind, um sicherzustellen, dass die Untersuchung der Prozesse und Ergebnisse angemessen vollständig ist und relevante Perspektiven nicht ausgeschlossen werden. Dies beinhaltet Aktivitäten wie *Schulung*[663] und Verbesserung des *internen Fachwissens,*[664] *Überprüfung durch Expertinnen und Experten*[665] und sogar *Vielfalt in der Belegschaft*[666] sowie *Transparenz als öffentliche Debatte,*[667] wenn sie dazu dienen soll, das Verständnis einer Organisation für die sozialen Auswirkungen von KI-Systemen zu verbessern.
Viertens umfasst die Kontrolle Massnahmen zur Risikominderung, beispielsweise, Backups und Notfallpläne zu erstellen,[668] Prozesse einzugrenzen und zu trennen, zu blockieren und zu unterbrechen,[669] die Möglichkeit für menschliches Eingreifen zu schaffen,[670] Risiken vorherzusagen und zu verhindern, schädliche oder riskante Praktiken zu verbieten,[671] Prozesse anzufechten[672] und Fehler zu korrigieren.[673] Die Bedeutung, die der Risikobewertung und dem Risikomanagement[674] in den hier analysierten Leitlinien beigemessen wird, kann kaum überschätzt werden.
Fünftens und mit besonderer Bedeutung für den Einsatz von KI-Systemen im *öffentlichen* Sektor beinhaltet die Kontrolle, wer über *Schlüsselinfrastrukturen*[675] Bescheid weiss, wem sie gehören und wer sie effektiv kontrolliert – z. B. die Datenbestände und Algorithmen für maschinelles Lernen, die wesentliche Voraussetzung dafür sind, aus Daten zu lernen und die eingesetzte KI weiterzuentwickeln, zu gestalten und zu kontrollieren.
Zu betonen ist, dass in der hier skizzierten Richtlinie kein Instrument zur Risikobewertung zur Verfügung gestellt wird, mit dem das Risiko *quantifiziert* werden soll. Die vorliegend untersuchten Werkzeuge, die von den Verwaltungen Kanadas[676] und Neuseelands[677] verwendet oder in den Richtlinien des Weltwirtschaftsforums[678] empfohlen werden, weisen allesamt Merkmale auf, die als problematisch erachtet werden können – insbesondere, wenn sie eine Rechtsgrundlage schaffen sollen, auf deren Basis bei Verstössen sanktioniert werden soll. Das neuseeländische Risiko-Tool stützt sich vollständig auf subjektive Risikobewertungen und fordert Benutzerinnen und Benutzer auf, anzugeben, ob ein Risiko «gelegentlich» besteht, «unwahrscheinlich» oder «wahrscheinlich» ist und ob die Auswirkungen «gering», «mässig» oder «hoch» sind. Jedoch werden keine konkreten, objektiven Kriterien angegeben, auf die sich diese Einschätzungen stützen können. Bezeichnungen wie «nicht ernst», «mässig», «weit verbreitet» oder «ernst» sind sehr kontextbezogen und vage. In einem Kontext, in dem die Einstufung einer Anwendung als «kann schwerwiegende Konsequen-

653 Dataethical Thinkdotank, 2021, Traceability.
654 Council of Europe, CM/Rec (2020)1, Testing.
655 Council of Europe, CM/Rec (2020)1, Interaction of systems.
656 Bundesrat Leitlinien KI 2020. Leitlinie 3; Engelmann/Puntschuh, 2020, 7. Wirkungsmonitoring.
657 Engelmann/Puntschuh, 2020, 8. Nachvollziehbarkeit.
658 AI Now Institute et al., Key Elements Of A Public Agency Algorithmic Impact Assessment, #1; Reisman et al., 2018; Council of Europe, CM/Rec (2020)1, Ongoing review, Evaluation of datasets and system externalities, Testing on personal data; WEF, 2020, Data Quality; New York City, 2019, Impact determination.
659 Council of Europe, CM/Rec (2020)1, Standards; Engelmann/Puntschuh, 2020, 1. Zielorientierung.
660 Mittelstadt/Russell/Wachter, 2019; New York City, 2019, Explanation; Engelmann/Puntschuh, 2020, 8. Nachvollziehbarkeit.
661 Dataethical Thinkdotank, 2021, Explainability; Floridi/Cowls, 2019; Government of New Zealand, 2020, Transparency.
662 Leslie, 2019, Transparency; Council of Europe, CM/Rec (2020)1, Testing; Loi/Ferrario/Viganò, 2020.
663 Council of Europe, CM/Rec (2020)1, Personnel management; Engelmann/Puntschuh, 2020, 9. Akzeptanz.
664 AI Now Institute et al, Executive Summary; Council of Europe, CM/Rec (2020)1, Independent research and Rights-promoting technology.
665 AI Now Institute et al., 2018, Key Elements Of A Public Agency Algorithmic Impact Assessment, #2; Government of Canada, 2019, Appendix C; Council of Europe, CM/Rec (2020)1, Consultation and adequate oversight and Expertise and oversight.
666 Council of Europe, CM/Rec (2020)1, Principle of Equality and Security and Personnel management.
667 Council of Europe, CM/Rec (2020)1, Public debate.
668 Government of Canada, 2019, Appendix C.
669 Council of Europe, CM/Rec (2020)1, Follow up, Consultation and adequate oversight.
670 Government of New Zealand, 2020; Council of Europe, 2020, Ethical Charter AI, Principle under user control.
671 Council of Europe, CM/Rec (2020)1, Consultation and adequate oversight and Follow up.
672 Council of Europe, CM/Rec (2020)1, Barriers and Effective remedies.
673 Council of Europe, CM/Rec (2020)1, Consultation and adequate oversight and Effective remedies.
674 WEF, 2020, Key variables to consider in a risk assessment; Council of Europe, CM/Rec (2020)1, Human Rights Impact Assessment; Government of New Zealand, 2020, Assessing likelihood and impact, Human oversight and accountability, Reliability, Security and Privacy; Government of Canada, 2019, Algorithmic Impact Assessment.
675 Council of Europe, CM/Rec (2020)1, Infrastructure and Interaction of systems.
676 Government of Canada, 2019, Appendix B.
677 Government of New Zealand, 2020, Assessing likelihood and impact.
678 WEF, 2020.

zen haben» bedeutet, dass kostspielige Dokumentations- und Verwaltungsanforderungen folgen, ist zu erwarten, dass die Unbestimmtheit der Sprache Benutzerinnen und Benutzer dazu veranlasst, die Beschreibung des Risikos herunterzuspielen (wenn sie die Konsequenzen tragen müssen) oder einheitlich die höchste Risikostufe auszuwählen (wenn die Personen, die für das Risiko verantwortlich sind, nicht die höheren Kosten für dessen Management tragen müssen). Darüber hinaus berücksichtigt das Auswirkungskriterium sowohl die Ausbreitung (Anzahl der Betroffenen) als auch die Schwere (wie schwerwiegend der Schaden ist) und lässt ungeklärt, wie diese Dimensionen zum Risiko beitragen.

Aus ähnlichen Gründen sind die Folgenabschätzungsniveaus der kanadischen «Richtlinie über automatisierte Entscheidungsfindung»[679] problematisch, da sie keine Kriterien zur Unterscheidung zwischen «geringer», «mässiger», «hoher» und «sehr hoher» Auswirkung bietet; ausserdem kann das Wort «oft» im Ausdruck «wird oft dazu führen» zu einer Vielzahl widersprüchlicher Interpretationen führen.

In den Richtlinien des Weltwirtschaftsforums heisst es: «Es ist wichtig, Faktoren wie die Anzahl der Betroffenen zu berücksichtigen.» Unklar ist jedoch, ob dies mit den anti-utilitaristischen Grundsätzen vereinbar ist, die die Gesetze vieler EU-Länder beeinflussen. Diese Art von Risikoindikator könnte darauf hindeuten, dass ein geringes Mass an Kontrolle über einen riskanten Algorithmus, der einem Individuum erheblichen Schaden zufügen kann, ethisch zulässig ist, solange nur die Würde einiger weniger Individuen von ihnen negativ beeinflusst wird (während die Mehrheit von ihrer Einführung profitiert). Dies steht auch im Spannungsfeld mit einer der Empfehlungen des Europarates, die bezogen auf die «Verantwortung der Akteure des Privatsektors in Bezug auf Menschenrechte und Grundfreiheiten im Kontext algorithmischer Systeme» eine klare *anti-utilitaristische* ethische Sichtweise annimmt, indem sie klar sagt («1.2. Umfang der Massnahmen»), dass die Verantwortung für die Achtung der Menschenrechte «unabhängig von ihrer Grösse» gilt (auch wenn Umfang und Komplexität der Mittel unterschiedlich sein können).[680]

Bei der Entwicklung der Leitlinien für den Kanton Zürich wurde ebenfalls davon ausgegangen, dass Kontrolle einen hohen Stellenwert besitzt. Diese wird vor allem durch *Dokumentation* umgesetzt. Einen Transparenzbericht zu schreiben, soll nicht nur oder nicht einmal hauptsächlich die externe Kontrolle und Überprüfung ermöglichen, sondern auch und vor allem die Verwaltung verpflichten, klar strukturierte und zielgerichtete Dokumentations-, Mess- und Bewertungsaufgaben zu übernehmen. Fragen 2.1 und 2.2 erfordern die Dokumentation der Ziele des Systems, Fragen 2.13 und 2.15 verlangen die Dokumentation der *Verantwortlichkeiten* und der *menschlichen Kontrollstruktur,* die zur Steuerung des Systems zur Verfügung stehen. Fragen 2.7 und 2.8 und Fragen 2.10, 2.12, 2.16, 2.17, 2.18 und 2.19 erfordern die Dokumentation von *Definitionen, Standards, Tests, Messungen* und *Bewertungen* der Leistung, des Datenschutzes, der Fairness und der beteiligten Stakeholder. Fragen 2.9, 2.10, 2.14, 2.15 und 2.20 setzen die Dokumentation von *Risikomanagementstrukturen* einschliesslich *Überwachung, Feedback, Fehlerkorrektur* und *Cybersicherheit* sowie ihrer Ergebnisse voraus.

Die Checkliste erfordert daher nicht, dass die Verwaltung den *Grad* des Risikos *quantifiziert,* bevor sie sich damit befasst. Sie funktioniert jedoch trotzdem als eine Art Risikobewertungsinstrument, da sie bestimmte Dokumentationsaufgaben unter der Bedingung notwendig macht, dass entsprechende Risikosignale von der Person, welche die Checkliste beantwortet, erkannt werden. Je höher die Anzahl der Risikosignale ist, desto länger und strukturierter sind die Dokumentations- und Transparenzanforderungen und der Aufwand für die Verwaltung bei der Kontrolle der KI-Systeme. Sie ist eher als Checkliste für die *Risikoreaktion* als für die *Risikobewertung* gedacht.

2. Transparenz

Unter Transparenz werden hier ausschliesslich die Vorlage und die Übermittlung von Informationen an Parteien ausserhalb der Institution verstanden, die eine KI-Lösung entwerfen oder implementieren, einschliesslich der Wirtschaftsprüfer, externen Expertinnen und Experten, Journalistinnen und Journalisten, Politikerinnen und Politiker, Verantwortlichen in anderen Bereichen der Verwaltung und der breiten Öffentlichkeit. Die Kommunikation zwischen den Mitgliedern des Data-Science-Teams, zwischen einem Data-Science-Team und der/dem CEO des Datenanalyseunternehmens, dem sie Bericht erstatten, oder zwischen einem privaten Unternehmen und der für die Beschaffung zuständigen Verwaltungseinheit wird hier als intern verstanden, da davon ausgegangen werden kann, dass alle diese Parteien in Bezug auf die Verwendung und Anwendung von KI-Systemen das gleiche Ziel verfolgen.

Es gibt mindestens vier Haupttheorien, warum Transparenz aus ethischer Sicht instrumentell wertvoll ist.[681] Erstens wird die Ansicht vertreten, dass «Sonnenlicht das beste Desinfektionsmittel ist», um Louis Brandeis zu zitieren, das bedeutet die Auffassung, dass Transparenz die Rechenschaftspflicht fördert, was wiederum verhindert, dass zumindest das schlimmste unethische Verhalten auftritt. Zweitens besteht die Meinung, dass Transparenz zur Qualität der Technologie beiträgt, da sie das Crowdsourcing von Expertenmeinungen und das Feedback der betroffenen Bürgerinnen und Bürger ermöglicht, was zu einer besseren Prüfung der Technologie führt und sie vertrauenswürdiger macht. Drittens gibt es die Ansicht, dass Transparenz Endbenutzerinnen und Endbenutzern einer Technologie oder Personen, die davon betroffen sein könnten, ermöglicht, eine fundierte Entscheidung darüber zu treffen, ob sie verwendet werden soll. Viertens wird vertreten, dass Transparenz eine öffentliche Debatte ermöglicht, die für die demokratische Legitimität technologischer Lösungen erforderlich ist, was besondere Relevanz hat, wenn die Implementierung von Technologie nicht wertneutral ist. Alle diese vier Rollen der Transparenz spiegeln sich in den untersuchten Richtlinien wider und es wird allgemein angegeben, dass sie zum Vertrauen in die Technologie beitragen.

Die Plausibilität, Stärke und Reichweite jeder dieser Theorien sind jedoch umstritten.[682] Die Rechenschaftstheorie ist möglicherweise nicht in allen Kontexten, sondern nur in solchen gültig, die aus dem einen oder anderen Grund die öffentliche Aufmerksamkeit auf sich ziehen. Die Crowdsourcingtheorie unterstützt möglicherweise nur schwächere Formen der Transparenz (die Technologie wird der Kontrolle einer begrenzten und ausgewählten Gruppe von Expertinnen und Experten unter-

679 Government of Canada, 2019.
680 Council of Europe, CM/Rec (2020)1, Appendix B.
681 DE LAAT, 2017; FELZMANN ET AL. 2019; LOI/FERRARIO/VIGANÒ, 2020; ZARSKY, 2013.
682 FELZMANN ET AL. 2019; ZARSKY, 2013.

worfen). Die Autonomietheorie kann die begrenzten kognitiven Ressourcen des Individuums, das sich für eine Technologie entscheiden muss, nicht angemessen berücksichtigen und ist nicht anwendbar, wenn Individuen nicht die Wahl haben, von der Technologie betroffen zu sein. Die Theorie der öffentlichen Debatte kann ihren Zweck verfehlen, wenn die breite Öffentlichkeit entweder nicht interessiert oder nicht qualifiziert genug ist, um diese Art von Diskussion zu führen. Dennoch ist Transparenz das am häufigsten zitierte Prinzip in KI-Ethikrichtlinien: Nur sehr wenige Texte erwähnen sie nicht.[683] Unter den Richtlinien, die speziell für diesen Bericht überprüft wurden, gab es keine einzige, in der sie nicht genannt wurde.

Rolle der Transparenz	Ziele	Ethischer Wert
Transparenz als Desinfektionsmittel	Rechenschaftspflicht, Vermeidung von unethischem Verhalten	Schadensvermeidung
Transparenz für Crowdsourcing	Sammeln von Experten- und Laienmeinungen, Verbessern der Technologie	Nutzen
Transparenz für fundierte Auswahl	Informierte individuelle Auswahl ermöglichen	Autonomie (individuell)
Transparenz für eine informierte öffentliche Debatte	Informierte demokratische Deliberation ermöglichen	Autonomie (kollektiv), Demokratie

Tab. 1 (eigene Darstellung): Transparenz

Transparenz über KI-Systeme – oder besser gesagt das soziotechnische System, von dem KI nur ein Teil ist – herzustellen, wird in Bezug auf verschiedene Elemente erwartet: die Existenz automatisierter Entscheidungssysteme,[684] einschliesslich ihres Zwecks, ihrer Reichweite und ihrer tatsächlichen Verwendung (Fragen 2.1 und 2.2),[685] die Definitionen von Kernkonzepten und Kernmassnahmen (z. B. der automatisierten Entscheidung oder der KI,[686] der Fairness[687]) (Fragen 2.8, 2.10, 2.12), die diesbezügliche ethische oder Folgenabschätzung,[688] ihre Rechtfertigung[689] (Fragen 2.18 und 2.19), die zugrunde liegenden Datentypen und Verarbeitungsmethoden[690] sowie deren Gesamtqualität, die als Genauigkeit,[691] Effektivität, Effizienz[692] oder Fähigkeit zur Unterstützung der Verwaltung[693] charakterisiert wird (Fragen 2.16 und 2.17). Im Vergleich zu früheren Richtlinien, die in vorherigen Analysen untersucht wurden,[694] scheint es weniger wichtig zu sein, individuelle Erklärungen der Ursachen oder Gründe zu liefern, warum eine bestimmte Entscheidung von einer KI getroffen wurde[695] – etwas, das im gerade entstehenden Feld der [e]X[plainable] KI im Vordergrund steht.

Während anerkannt wird, dass ein gleiches Mass an Transparenz nicht immer für alle Systeme angemessen ist,[696] sollte dieses jedoch so gross sein, wie es nach einer Güterabwägung mit anderen Zielen möglich ist.[697] Das Zielpublikum der Kommunikation kann variieren und die beteiligten oder betroffenen Personen,[698] die breite Öffentlichkeit[699] oder unabhängige Expertinnen und Experten umfassen.[700] Das Format der Kommunikation kann sich auch abhängig vom Kontext unterscheiden, obwohl dies selten spezifiziert wird, z. B. von allgemeinverständlichen Texten auf einer Webseite (z. B. die Erklärung, dass eine automatisierte Entscheidung getroffen wird) bis zu einem vollständig dokumentierten Bericht.[701] Mitunter wird sogar vorgeschlagen, Veröffentlichungen durch Whistleblower zu schützen und durch staatliche Gesetze und Strukturen in Unternehmen zu unterstützen.[702]

Transparenz kann als Eckpfeiler des hier skizzierten Ansatzes angesehen werden, bei dem Beamtinnen und Beamte einen Transparenzbericht erstellen müssen, der das wichtigste Ergebnis aller ethischen Aktivitäten darstellt. Nach der hier ver-

683 Jobin/Ienca/Vayena, 2019.
684 AI Now Institute et al.; Cities for Digital Rights, 2020; Council of Europe, CM/Rec (2020)1, Identifiability of algorithmic decision-making; Dataethical Thinkdotank, 2021, Fair communication.
685 AI Now Institute et al.; Engelmann/Puntschuh, 2020, 1. Zielorientierung; Engelmann/Puntschuh, 2020, 8. Nachvollziehbarkeit.
686 AI Now Institute et al.
687 Leslie, 2019.
688 AI Now Institute et al.; Council of Europe, CM/Rec (2020)1, Expertise and oversight; Government of Canada, 2019, Appendix C – Notice.
689 AI Now Institute et al.; Government UK, Transparency; Bundesrat Leitlinien KI, Leitlinie 3.
690 Council of Europe, 2020, Ethical Charter AI; Government of Canada, 2019, Appendix C – Notice.
691 Dataethical Thinkdotank, 2021, Fair communication.
692 Government of Canada, 2019, Reporting: 6.5.1.
693 Government of Canada, 2019, Appendix C – Notice.
694 Loi, 2020; Loi/Heitz/Christen, 2020.
695 Government UK, 2019, Transparency; Dataethical Thinkdotank, 2021, Transparency; WEF 2020, Human in the loop; Government of Canada, 2019, 6.2.3.
696 Council of Europe, CM/Rec (2020)1, Levels of transparency; Government of New Zealand, 2020, Transparency.
697 Council of Europe, CM/Rec (2020)1, Levels of transparency.
698 Dawson et al., 2020; Dataethical Thinkdotank, 2021, Transparency; Government UK, 2019, Ongoing review; Council of Europe, CM/Rec (2020)1, Expertise and oversight; Bundesrat Leitlinien KI 2020, Leitlinie 3.
699 Council of Europe, CM/Rec (2020)1, Public debate; New York City, 2019, Available information.
700 Council of Europe, CM/Rec (2020)1, Expertise and oversight; AI Now Institute et al.; Bundesrat Leitlinien KI 2020, Leitlinie 3.
701 Government of Canada, 2019, 6.2, and Appendix C – Notice.
702 Council of Europe, CM/Rec (2020)1, Advancement of public benefit.

tretenen Ansicht steht das gewünschte Mass an Transparenz in einem relativen Verhältnis zum Kontext und zur Art des zu beurteilenden KI-Systems. Anstatt eine Risikobewertung als Grundlage anzubieten, liefert der hier vorgeschlagene Ansatz Hinweise, mit welchen Aspekten von Transparenz die Verwaltung sich befassen muss bzw. welche Aspekte und Themen sie transparent machen muss.

3. Rechenschaftspflicht

Die Rechenschaftspflicht umfasst Massnahmen, Entscheidungen, Rahmenbedingungen und Organisationsstrukturen, die die Verteilung und Identifizierung von Verantwortlichkeiten erleichtern sollen. Nur menschliche Personen können zur Rechenschaft gezogen werden, während dies bei einem KI-System nicht möglich ist. Rechenschaftspflichtige Akteure können Verantwortung für ihre Handlungen übernehmen und mit Sanktionen belegt werden. Die Förderung der Rechenschaftspflicht ist daher gleichbedeutend mit derjenigen der Fähigkeit, festzustellen, wer wofür verantwortlich ist und wer für unethische oder illegale Ergebnisse – nicht nur rechtlich, sondern auch organisatorisch oder durch Imageschäden – sanktioniert werden sollte. Die im Rahmen dieser Studie untersuchten Richtlinien sehen vor, dass Rechenschaftspflichten wie folgt gefördert werden:

1. Indem Verantwortung zugewiesen wird – es sollte möglich sein, zu identifizieren, wer für die Gewährleistung ethischer Ergebnisse und Verhaltensweisen verantwortlich ist (vorausschauende Verantwortung)[703] und wer Sanktionen unterliegt, wenn dies nicht der Fall ist.[704]
2. Durch angemessene Strukturen und eine angemessene Organisation der Prozesse der Datenwissenschaft hinter dem KI-System und den Automatisierungsprozessen.[705] Organisationen sollten «eine kontinuierliche Verantwortungskette für alle Rollen einrichten, die am Entwurfs- und Lebenszyklus des Projekts beteiligt sind».[706] Hier ist zu beachten, dass eine kontinuierliche Verantwortungskette für alle Rollen erfordert, Prozesse und Ergebnisse klar zu dokumentieren, zu überwachen und zu kontrollieren.[707] Mit anderen Worten: Angemessene Strukturen für die Rechenschaftspflicht sind Kontrollstrukturen wie oben definiert (Abschnitt C. II. 1.). Der Kürze halber werden nicht alle Kontrollelemente wiederholt, die bereits in Abschnitt C. II. 1. beschrieben wurden.
3. Indem ermöglicht wird, von scheinbar unpersönlichen Systemen getroffene Entscheidungen anzufechten oder abzulehnen[708] (in einigen Fällen kann das Rechtsstaatsprinzip herangezogen werden)[709] oder das KI-System grundsätzlich aufgrund seiner schädlichen oder diskriminierenden Auswirkungen anzufechten (mit Konzepten, die öffentlichen Beteiligungsverfahren entsprechen).[710]
4. Indem verlangt wird, dass Institutionen, die ein KI-System einsetzen, dafür verantwortlich sind, das Feedback der von ihm betroffenen Personen zu sammeln und die erforderlichen Abhilfemassnahmen umzusetzen.[711]
5. Indem verlangt wird, dass Schäden kompensiert werden, die aufgrund unethischen Verhaltens entstanden sind.

Die hier entwickelten praktischen Richtlinien greifen den Grundsatz der Rechenschaftspflicht auf, indem sie die Rechenschaftspflicht in ein Objekt der Transparenz verwandeln. Die Fragen 2.3 bis 2.6 erfordern, dass die oben genannten individuellen Verantwortlichkeiten im Transparenzbericht angegeben werden. Die Fragen 2.15 und 2.10 befassen sich mit der Existenz von Strukturen, mit denen Entscheidungen des KI-Systems infrage gestellt werden können.

C. Checklisten

I. Einleitung

Bei den folgenden zwei Checklisten handelt es sich um Hilfsmittel zur Herstellung von Transparenz bei technologischen Automationsprojekten und -anwendungen in der öffentlichen Verwaltung.

Die Methode zur Herstellung von Transparenz besteht darin, einen Transparenzbericht zu verfassen, der zeigt, dass die wichtigsten ethischen Fragen sowohl erkannt als auch unter menschliche Kontrolle gebracht wurden und eine angemessene Rechenschaftspflicht für den Prozess sichergestellt wurde. Bei der Prüfung der ethischen Anforderungen für den Einsatz von KI-Systemen sollten zwei verschiedene Einschätzungen die praktischen Aktivitäten des Kantons Zürich leiten. Zu diesem Zweck werden zwei Checklisten zur Verfügung gestellt.

Die Checklisten greifen teilweise Aspekte auf, die bereits durch andere Instrumente abgedeckt sind. Ausserdem können sich die Checklisten mit bestehenden Prozessen in der Verwaltung des Kantons Zürich überschneiden (zu denken ist etwa an Regeln und Abläufe zur Sicherstellung des Datenschutzes oder der Cyber-Security). Dies ist bei der Implementierung der Checklisten zu berücksichtigen, würde aber den Rahmen des vorliegenden Vorprojekts sprengen. Dasselbe gilt für die Koordination mit dem Bund.

Im Rahmen der Entwicklung dieser Checklisten wurde der Vorschlag geäussert, diese im Rahmen von HERMES (Initialisierungsphase) zu integrieren. Allerdings sind die Checklisten nicht nur bei der Durchführung eines Projekts, sondern auch bei späteren Änderungen zu berücksichtigen sind. Sie sind somit ab einem bestimmten Zeitpunkt nicht mehr für die Projektleitung, sondern für die Stammverwaltung massgebend.

Anhand der **ersten Checkliste** (Triage-Checkliste) beurteilt die Verwaltung, welche ethischen Transparenzfragen während der Projektdurchführung im Detail zu dokumentieren sind, und wählt angemessene Vorgehensweisen für die Generierung derjenigen Daten und Bewertungen, die notwendig sind, um den Bericht mit informativem Inhalt zu füllen. Die folgenden Fragen helfen bei der Beurteilung:

– Mit wie vielen ethischen Transparenzaspekten muss die Verwaltung sich befassen?

703 Dawson et al., 2020, Accountability; Bundesrat Leitlinien KI 2020, Leitlinie 4; Engelmann/Puntschuh, 2020, 4. Projektmanagement.
704 Government of New Zealand, 2020, Human oversight and accountability.
705 Engelmann/Puntschuh, 2020, 4. Projektmanagement.
706 Government UK, 2019, Accountability.
707 Government UK, 2019, Accountability.
708 AI Now Institute et al.; Council of Europe, CM/Rec (2020)1, Contestability; Dawson et al., 2020, Contestability.
709 AI Now Institute et al.
710 Cities for Digital Rights, 2020.
711 AI Now Institute et al., Participatory Democracy, diversity and inclusion; Council of Europe, CM/Rec (2020)1, Consultation and adequate oversight; Dawson et al., 2020, Recourse; New York City, 2019, impact address.

– Wie viele ethische Transparenzverfahren müssen implementiert werden?
– Wie viele Ressourcen müssen für ethische Transparenzverfahren bereitgestellt werden?
– Welche Aspekte der ethischen Transparenz müssen im Bericht detailliert behandelt werden? (Und ist ein solcher Bericht überhaupt notwendig?)

Die **zweite Checkliste** (Checkliste Transparenzbericht) dient als Leitfaden für die Erstellung eines ausführlichen Transparenzberichts (im Folgenden: Transparenzbericht).
Der Transparenzbericht kann erst am Ende einer Entwicklung und Implementierung eines KI-Systems (im Folgenden: Projekt) erstellt werden (einschliesslich der Interaktion des soziotechnischen Systems mit der betroffenen Öffentlichkeit in Fällen, in denen die Beurteilung ethischer Fragen eine solche Überwachung erfordert). Allerdings muss mit der Erstellung des Transparenzberichts bereits während des Projekts begonnen werden: Manche Informationen, die für den Transparenzbericht notwendig sind, können nur in den verschiedenen Phasen der Projektdurchführung und nicht erst nach Projektabschluss generiert werden. Die Checkliste Transparenzbericht enthält deshalb auch Hinweise darauf, in welcher Phase des Prozesses spezifische, für die Transparenz notwendige Informationen generiert werden müssen. Am Ende des Projekts muss der Transparenzbericht klare Informationen über die umgesetzten Prozesse enthalten, die zur Adressierung der in Checkliste 1 (Triage-Checkliste) hervorgehobenen spezifischen ethischen Punkte geeignet sind.
Falls sich die Prozesse nach der Einführung des Systems ändern, muss die Verwaltung überprüfen, ob die ursprüngliche Einschätzung noch gültig ist oder sich zusätzliche ethische Transparenzprobleme ergeben haben.
Diese ethischen Transparenzprozesse sollten von der Verwaltung eingeleitet werden und alle potenziellen Transparenzaspekte berücksichtigen, die in der Checkliste angegeben sind. Wenn die Verwaltung nicht in der Lage ist, ein angemessenes Mass an Transparenz über diese ethischen Fragen zu schaffen, oder wenn die Transparenz die Unzulänglichkeit des soziotechnischen Systems in einer Weise aufzeigt, die mit dem Ruf der Behörde, die das Projekt verantwortet, unvereinbar ist, sollte dies dazu führen, das Projektziel zu überdenken und/oder mehr Ressourcen in die Suche nach einer praktikablen Lösung zu investieren.
Unter Transparenz wird im vorliegenden Kontext die Kommunikation an verschiedene Zielgruppen verstanden. Im Falle einiger Anwendungen oder für einige Informationskategorien wird die breite Öffentlichkeit (jede/jeder) die Adressatin sein. Dies kann z. B. durch die Veröffentlichung eines Berichts mit allen Informationen auf einer Webseite umgesetzt werden. Bei anderen Anwendungen oder für einige andere Informationskategorien werden eine vorgesetzte Stelle oder eine verantwortliche Person innerhalb derselben Abteilung oder einer anderen Abteilung, technische Expertinnen und Experten oder Vertreterinnen und Vertreter von Interessengruppen, denen die Dokumentation vertraulich mitgeteilt wird, die Zielgruppe sein.

II. Checkliste 1: Triage-Checkliste für KI-Systeme

1. Einleitende Bemerkungen

Die Fragen in dieser Checkliste sollten so früh wie möglich, d. h. bereits in der Planungsphase, beantwortet werden, da sie dabei helfen, neben den primären Projektzielen zusätzliche Spezifikationen für das Projekt zu berücksichtigen. Die Checkliste 1 hilft bei der Ermittlung der wichtigsten ethischen Transparenzaspekte, die dokumentiert werden müssen. Sie sollte nicht als erschöpfende Liste aller Risiken für alle Zusammenhänge angesehen werden (es können andere Risiken für Personen, Sachen oder die Gesellschaft insgesamt bestehen). Es handelt sich um eine methodische Herangehensweise, jedoch besteht keine Garantie.
Die erste Checkliste ist bewusst so formuliert, dass sie nicht nur für spezifische KI-Technologien gilt. Vielmehr soll mit ihrer Hilfe ermittelt werden, welche Arten von KI-Systemen einen (erhöhten) Transparenzbedarf aufweisen. Es ist einfacher, anhand der Checkliste ein adäquates Mass an Transparenz zu erreichen, wenn
– das Projekt ethische Aspekte während der frühen Planung und bei der Implementierung des Produkts berücksichtigt,
– die Spezifikationen zur Befassung mit solchen ethischen Aspekten von Anfang an einbezogen werden (Ethics-by-Design-Ansatz),
– die Informationen zu den zur Behandlung der ethischen Aspekte ergriffenen Massnahmen in jeder Phase des Projekts ordnungsgemäss dokumentiert werden,
– die Transparenz und die Rechenschaftspflicht der von Menschen durchgeführten Verfahren in allen Phasen des Projekts bestehen.

Die Checkliste impliziert zwei Ebenen der Transparenz: Transparenz mit geringem Detaillierungsgrad und Transparenz mit hohem Detaillierungsgrad:
– Niedriger Detaillierungsgrad der Transparenz: Dokumentieren und speichern Sie nur die Antworten auf die Triage-Checkliste einschliesslich der Begründungen für Ihre Antworten.
– Hoher Detaillierungsgrad der Transparenz: Legen Sie einen detaillierten Transparenzbericht ab (Checkliste 2). Die Antworten auf die Fragen der Triage-Checkliste bestimmen, ob ein Transparenzbericht ausgefüllt werden muss und welche Abschnitte des Transparenzberichts ausgefüllt werden müssen.

Bewertungsstufe für die Triage-Checkliste: ganz am Anfang eines Projekts.

2. Schadensvermeidung

1.1. Befasst sich die Entscheidung mit speziellen Kategorien personenbezogener Daten im Sinne des kantonalen Rechts?
1.2. Haben böswillige Parteien besonders starke Motive, das System zu hacken? Können sie, auch durch Erpressung, einen einfachen und substanziellen finanziellen Gewinn erzielen oder kann ein gehacktes System verwendet werden, um politische Ziele zu erreichen (einschliesslich der Äusserung politischer Opposition gegen das System)?
1.3. Wird das soziotechnische System verwendet, um Entscheidungen über Personen zu treffen, zu empfehlen oder zu beeinflussen, und zwar in einer Weise, die Auswirkungen darauf hat, welche Entscheidung getroffen wird?
[Zum Beispiel ist eine automatische Rechtschreibprüfung Teil eines soziotechnischen Systems. Wenn es von Menschen verwendet wird, die Entscheidungen über Individuen treffen, kann es als «verwendet, um Entscheidungen über Individuen zu treffen» beschrieben werden. Aber es beeinflusst nicht (in irgendeiner erkennbaren und wissenschaftlich plausiblen Weise), welche Entscheidung getroffen wird.]
1.4. Wird das System verwendet, um eine Entscheidung über eine gesetzliche Pflicht oder ein Recht einer Person zu treffen?
[Zur Bedeutung von «für eine Entscheidung verwendet» in diesem Zusammenhang siehe Anleitungstext zu Frage 1.3.]
1.5. Macht es das System mehr oder weniger wahrscheinlich, dass bestimmte Personen den Wesensgehalt eines Rechts geniessen? Oder macht es das System mehr oder weniger wahrscheinlich, dass bestimmte Personen sanktioniert werden?

Oder beeinflusst das System die Wahrscheinlichkeit, dass ein Einzelfall die Aufmerksamkeit der Verwaltung auf sich zieht oder von dieser ignoriert wird?
1.6. Können Einzelpersonen die Entscheidung vermeiden oder verlangen, dass die Entscheidung über ein anderes Verfahren getroffen wird, bei dem nicht dasselbe technische System verwendet wird?
1.7. Kann die Person, über die mithilfe des Tools eine Entscheidung getroffen wurde, beweisen, dass diese falsch ist, ohne vor Gericht zu gehen?
1.8. Ist der Schaden einer falschen Entscheidung vollständig reversibel?
1.9. Ist es möglich, den Einzelnen oder die Familie vollständig und angemessen zu entschädigen, wenn festgestellt wird, dass die Entscheidung falsch war und irreversibel ist?
1.10. Betrifft die Entscheidung einen der folgenden Bereiche des öffentlichen Lebens oder Ressourcen des öffentlichen Sektors:
– die Rechtspflege,
– den Zugang zu Bildungschancen,
– den Zugang zu demokratischen Prozessen,
– den Zugang zur Gesundheitsversorgung,
– Massnahmen im Bereich der öffentlichen Gesundheit,
– die Umwelt?
1.11. Findet durch die Anschaffung bzw. den Einsatz des KI-Systems in einem der folgenden Bereiche eine Änderung statt bei:
– öffentlicher Computer-Infrastruktur,
– öffentlichen Datenbeständen oder
– immateriellen Vermögenswerten (z. B. Kompetenzen) im öffentlichen Sektor?

3. Gerechtigkeit und Fairness

1.12. Besteht das Risiko, dass das System eine politische Entscheidung (z. B. Wahl oder Volksabstimmung) beeinflusst?
1.13. Beeinflusst das technische System die Verteilung öffentlicher Mittel an wirtschaftliche Akteure in der Gesellschaft?
1.14. Beruht das technische System auf einem statistischen Modell des menschlichen Verhaltens oder der persönlichen Merkmale?
1.15. Ist das System so konzipiert, dass es adaptiv ist, damit nicht alle neuen Fälle wie andere behandelt werden, denen es in der Vergangenheit begegnet ist, weil es seine Parameter ändert, z. B. um effizienter zu werden?

4. Autonomie

1.16. Ist es das Ziel des technischen Systems, ein vollständig deterministisches Regelsystem zu automatisieren, das nur ein Minimum an Kreativität und menschlichem Urteilsvermögen durch die derzeitigen menschlichen Anwenderinnen/Anwender erfordert und keine Risiko- oder Wahrscheinlichkeitsabschätzungen beinhaltet?
1.17. Beruht das technische System auf Parametern, Merkmalen, Faktoren oder Entscheidungskriterien, die nicht den von den meisten Fachleuten auf diesem Gebiet normalerweise berücksichtigten Aspekten entsprechen?
1.18. Beurteilt das technische System (durch Vorhersagen oder Empfehlungen), dass den zuständigen Mitarbeitenden der öffentlichen Verwaltung die Kompetenz (im Gegensatz zur Befugnis) zur Kritik und zum Verwerfen einer Entscheidung fehlt?
1.19. Greift das technische System auf die Infrastruktur eines Drittanbieters zurück, über die die öffentliche Einrichtung keine uneingeschränkte Kontrolle und/oder bei der sie keinen Zugriff auf z. B. Datensätze oder die Rechenleistung hat?

III. Checkliste 2: Transparenzbericht

Das Ziel der zweiten Checkliste ist, Transparenz herzustellen und damit die Vertrauenswürdigkeit des Prozesses zu fördern. Das Ziel wird mittels Erstellung eines Berichts erreicht.
Benutzungsanweisung: Wenn die Beantwortung der Checkliste 1 ergibt, dass ein Transparenzbericht zu verfassen ist, wird mithilfe des Flussdiagramms bestimmt, welche Fragen aus Checkliste 2 der Transparenzbericht beantworten muss.

1. Abschnitt: Bewertungsphase für die Fragen 2.1. bis 2.6.: Bevor Sie Ihr System entwerfen

Transparenz hinsichtlich von Werten

2.1. Für welches Problem soll das System eine Lösung liefern?
2.2. Welches sind die weiteren Anforderungen des Systems? Berücksichtigen Sie zumindest:
2.2.1. Privatsphäre
[Bitte gehen Sie in diesem Teil des Berichts spezifisch auf die in Checkliste 1 – Frage 1.1. genannten Aspekte ein.]
2.2.2. Cybersicherheit
[Bitte gehen Sie in diesem Teil des Berichts spezifisch auf die in Checkliste 1 – Frage 1.2. genannten Aspekte ein.]
2.2.3. Fairness
[Bitte gehen Sie in diesem Teil des Berichts spezifisch auf die in Checkliste 1 – Fragen 1.12. bis 1.15. genannten Aspekte ein.]
2.2.4. Erklärbarkeit
[Bitte gehen Sie in diesem Teil des Berichts spezifisch auf die in Checkliste 1 – Fragen 1.17. und 1.18. genannten Aspekte ein.]

Transparenz der Rechenschaftspflicht

[Es wird empfohlen, diesen Abschnitt in jedem Fall auszufüllen. Wenn Frage 1.16. mit «Ja» beantwortet wurde, muss aufgezeigt werden, dass die neuen Zuständigkeiten mindestens gleichwertige Möglichkeiten zur Anfechtung von Entscheidungen bieten wie das zuvor bestehende System.]
2.3. Wer ist für die Konstruktion des Systems verantwortlich (Ebene Projektorganisation)?
2.4. Wer ist für den Einsatz des Systems und dessen Resultate verantwortlich (Ebene Stammorganisation)?
2.5. Wer ist verantwortlich für die Verwaltung der Antworten und Rückmeldungen der Endbenutzerinnen und Endbenutzer, d. h. der Personen, die das System benutzen oder von ihm unterstützt werden?
2.6. Wer ist dafür verantwortlich, auf Zweifel oder Herausforderungen des Einzelnen, der von der Nutzung des Systems betroffen ist, zu antworten?

2. Abschnitt: Bewertungsphase für die Fragen 2.7. bis 2.19: Nach dem Testen des Systems

Transparenz der Umsetzung und der Steuerung

2.7. Mit welchen Methoden wurde die Leistung des Systems getestet und gemessen?
[Bitte geben Sie an, wie Sie die Leistung in Bezug auf das in Checkliste 2 – Frage 2.1. angegebene Hauptziel messen.]
2.8. Welche Methoden wurden verwendet?
2.8.1. Welche Methoden wurden verwendet, um die von den Systemvorhersagen/-empfehlungen/-entscheidungen unmittelbar betroffenen Stakeholder zu identifizieren? Und was sind die voraussichtlichen Auswirkungen auf diese Personen?
2.8.2. Welche Methoden wurden verwendet, um die von der digitalen Transformation in der öffentlichen Verwaltung betroffenen Personen zu identifizieren (z. B. Personal der öffentlichen Verwaltung)? Und was sind die voraussichtlichen Auswirkungen auf diese Personen?

2.9. Welche Protokolle sind vorhanden, um Systemfehler und Fehlfunktionen zu behandeln?
2.10. Welche Methoden wurden zur Definition und zum Schutz der Privatsphäre verwendet?
[Bitte gehen Sie in diesem Teil des Berichts spezifisch auf die in Checkliste 1 – Frage 1.1. genannten Aspekte ein.]
2.11. Welche Massnahmen zum Schutz der Cybersicherheit wurden getroffen?
[Bitte gehen Sie in diesem Teil des Berichts spezifisch auf die in Checkliste 1 – Frage 1.2. genannten Aspekte ein.]
2.12. Welche Methoden wurden verwendet, um die Voreingenommenheit und die Fairness des Systems zu definieren und zu messen?
[Bitte gehen Sie in diesem Teil des Berichts spezifisch auf die in Checkliste 1 – Fragen 1.12. bis 1.14. genannten Aspekte ein, die in diesem Zusammenhang relevant sind.]
2.13. Wie werden den Systemendbenutzerinnen und -endbenutzern sowie den Personen, die vom Einsatz des Systems unmittelbar betroffen sind, individuelle Vorhersagen/Empfehlungen/
Entscheidungen des Systems erklärt?
[Bitte gehen Sie in diesem Teil des Berichts spezifisch auf die in Checkliste 1 – Fragen 1.17. und 1.18. genannten Aspekte.]
2.14. Wird die Systembereitstellung nach der Testphase kontinuierlich überwacht?
a) Zu jedem Zeitpunkt?
b) In einem bestimmten Zeitpunkt?
c) Mit welchen Massnahmen?
2.15. Gibt es Möglichkeiten für Personen, die von einer Entscheidung betroffen sind, den Output des automatisierten Systems zu erfahren und die vom System beeinflussten Vorhersagen/Empfehlungen/Entscheidungen anzufechten?
[Falls zutreffend, beschreiben Sie, wie diese Kanäle mit den organisatorischen Rollen zusammenhängen, die in Checkliste 2 – Fragen 2.5. und 2.6. erwähnt werden.]

Transparenz hinsichtlich von Leistungen

Auf der Grundlage der bisherigen Testläufe:
2.16. Wie verhält sich das System in Bezug auf die ausgewählten relevanten Metriken?
[Bitte beachten Sie alle Ziele und Anforderungen, die in Checkliste 2 – Fragen 2.1. und 2.2. angegeben sind.]
2.17. Wie ist das System im Vergleich zu dem zuvor vorhandenen, falls zutreffend, oder mit etablierten Benchmarks, falls vorhanden?
[Bitte beachten Sie auch die Ziele und Anforderungen, die in Checkliste 2 – Fragen 2.1. und 2.2. angegeben sind.]
2.18. Welches sind die verbleibenden Sicherheits- und Datenschutzrisiken und warum sind sie angemessen?
[Bitte gehen Sie in diesem Teil des Berichts auf alle Anforderungen ein, die in Checkliste 2 – Fragen 2.1. und 2.2. sowie 2.10. und 2.11. genannt werden.]
2.19. Bitte beschreiben Sie relevante ungelöste Voreingenommenheit oder mögliche Ursachen für Ungerechtigkeiten im System und erklären Sie, warum sie nicht gelöst werden können (beispielsweise, indem Sie Kompromisse mit anderen Systemzielen einschliesslich widersprüchlicher Fairnessziele erläutern).
[Bitte gehen Sie in diesem Teil des Berichts auf alle Anforderungen ein, die in Checkliste 2 – Frage 2.2.3. genannt werden, und erläutern Sie, wie die in Checkliste 1 – Fragen 1.12. bis 1.15. identifizierten Aspekte adressiert wurden.]

3. Abschnitt: Bewertungsphase für die Frage 2.20: Nach der Implementierung des Systems, wenn das System überwacht wird

Die Transparenz über die angebotenen Lösungen zu den ethischen Fragen erfordert manchmal eine kontinuierliche Überwachung des Projekts auch über die Testphase hinaus. Kontinuierliche Überwachung bedeutet, dass der Transparenzbericht aktualisiert werden muss.
2.20. Wurden während der Überwachung Vorhersagen/Empfehlungen/Entscheidungen des Systems jemals
a) von der Systemendbenutzerin oder dem Systemendbenutzer oder
b) von Personen, die Entscheidungen unterliegen, hinterfragt?

IV. Beispiel für den Einsatz der Checklisten 1 und 2: Swiss COMPAS

Im Folgenden wird anhand eines imaginären Tools – Swiss-COMPAS-System zur Prognose von Rückfällen (ähnlich dem US-amerikanischen System COMPAS)[712] – die Anwendung der Triage für KI-Systeme (Checkliste 1) und der Vorgaben für den Transparenzbericht (Checkliste 2) veranschaulicht. Dieses imaginäre System wäre eine Anwendung, mit dem anhand von Daten über eine inhaftierte Person die Wahrscheinlichkeit bestimmt wird, dass diese nach ihrer Entlassung aus dem Gefängnis erneut straffällig wird.

1. Vorbemerkungen

Dieser hypothetische Bericht wird nur als Illustration dafür zur Verfügung gestellt, wie ein auf Tatsachen basierender Bericht aussehen würde. Der vorliegende Fall soll kein Modell für «ethische Best Practices» darstellen: Transparenz wird erreicht, wenn die Situation in Bezug auf die Automatisierung einschliesslich der Aspekte, die möglicherweise nicht auf ethisch angemessene Weise behandelt worden sind, ehrlich dargestellt wird. Anzunehmen ist, dass die meisten Praxisfälle unter dem Gesichtspunkt der ethischen Angemessenheit unvollkommen sein werden: Transparenz soll schrittweise, sukzessive Verbesserungen ermöglichen.
Dieser hypothetische Transparenzbericht wird als imaginärer «erster Entwurf» des Kantons Zürich verstanden, der durch den Austausch mit Stakeholdern, die mehr Informationen verlangen, als der aktuelle Bericht bereitstellt, verbessert wird. Transparenz soll den Einsatz von KI-Technologien nicht abschliessend rechtfertigen, sondern diese schrittweise verbesserungsfähig machen.
Das Flussdiagramm unterstützt Nutzerinnen und Nutzer dabei, den Transparenzbericht in einer anderen Reihenfolge als derjenigen der verschiedenen Abschnitte im Abschlussbericht auszufüllen.
Die Struktur des Transparenzberichts ist immer dieselbe, aber die Triage-Fragen und die Flussdiagrammbefehle bestimmen, welche Abschnitte in einem bestimmten Fall ausgefüllt werden müssen und welche nicht. Dies kann von Fall zu Fall unterschiedlich sein und hängt davon ab, wie die Triage-Fragen beantwortet werden. Das Flussdiagramm kann angeben, dass bestimmte Fragen

712 https://en.wikipedia.org/wiki/COMPAS_(software).

mehr als einmal beantwortet werden müssen (z. B. als Antwort auf verschiedene Triage-Fragen). Solche Wiederholungen können ignoriert werden. In diesem Beispiel wird keine Wiederholung erwähnt, die durch die Anforderungen des Flussdiagramms erzeugt wird.

Der Abschnitt «Checklisten 1 und 2» stellt dar, wie eine Person oder Gruppe, die verantwortlich dafür ist, einen Transparenzbericht zu erstellen, die Fragen auf der Grundlage der Checklisten eantworten könnte. Der Abschnitt «Transparenzbericht» stellt den Bericht in einer Fassung dar, die veröffentlicht werden könnte.

2. Checklisten 1 und 2

Frage aus der Triage-Checkliste (Checkliste 1)
Antwort auf die Frage
Konsequenz für den Transparenzbericht (ausgehend von den Fragen in Checkliste 2)

Frage

1.1. Befasst sich die Entscheidung mit speziellen Kategorien personenbezogener Daten im Sinne des kantonalen Rechts?

Antwort

Das (hypothetische) Schweizer COMPAS-Bewertungstool erfordert, das Geschlecht der bewerteten Person zu erheben. Diese Angabe ist durch das Diskriminierungsverbot geschützt bzw. in bestimmten Fällen untersagt. Es werden zudem vertrauliche Informationen gesammelt, z. B., ob die untersuchte Person eine Trennung ihrer Eltern durchlebt hat, oder Strafregistereinträge von Freunden und Verwandten.

Konsequenz

2.2.1. Welche Anforderungen werden an das System in Bezug auf die Privatsphäre gestellt?
[Legen Sie hier die Anforderungen an das System dar. Welche Massnahmen bezüglich des Datenschutzes sollten ergriffen werden? Nehmen Sie beispielsweise Kontakt mit der/dem Datenschutzbeauftragten Ihrer Organisation auf, um diesen Teil des Berichts zu verfassen.]

2.8.1. Welche Methoden wurden verwendet, um die von den Systemvorhersagen/-empfehlungen/-entscheidungen unmittelbar betroffenen Stakeholder zu identifizieren? Und was sind die voraussichtlichen Auswirkungen auf diese Personen?

Beispiel

«Wir haben ein Brainstorming-Meeting mit Staatsanwältinnen und Staatsanwälten, Richterinnen Richtern des Kantons sowie Anwältinnen und Anwälten der Strafjustiz durchgeführt. Bei diesem Treffen wurden die Stakeholder identifiziert, die direkt von den Vorhersagen betroffen sind, nämlich die Angeklagten, ihre Verteidiger, ihre Familien, potenzielle künftige Opfer, wenn die Angeklagten erneut straffällig werden, und Gemeinschaften, in denen Menschen, die möglicherweise erneut straffällig werden, leben.

Unsere Analyse der Stakeholderinteressen sieht wie folgt aus:

A) Angeklagte. Das System zu verwenden, liegt im Interesse der Angeklagten, bei denen es unwahrscheinlich ist, dass sie erneut straffällig werden (oder die statistisch nicht von denjenigen zu unterscheiden sind, bei denen es unwahrscheinlich ist, dass sie erneut straffällig werden). Es ist insbesondere im Interesse derjenigen, die ihr Recht auf ein günstiges Bewährungsurteil am schlechtesten ausüben können, da sie sich nicht die besten Anwältinnen und Anwälte leisten können. Es ist nicht im Interesse von Personen, die sich mithilfe guter Anwältinnen und Anwälte bessere Chancen verschaffen können, auf Bewährung entlassen zu werden.

B) Anwältinnen und Anwälte der Angeklagten. Das System ist nicht in ihrem Interesse, da es ein Bestandteil richterlicher Entscheidung sein wird, den die Anwältinnen und Anwälte nicht anfechten können.

C) Familien. Die Familien der Angeklagten werden dann von der höheren Wahrscheinlichkeit profitieren, dass der Angeklagte auf Bewährung freigelassen wird, wenn der Einsatz von Swiss COMPAS im Vergleich zum Status quo zu einem höheren Anteil an gewährten Bewährungsstrafen führt (es sei denn, Angeklagte sind wegen eines Verbrechens gegen ihre Familien angeklagt). Dies hängt eng mit der Verhältnismässigkeit der Entscheidungen zur Gewährung der Bewährung zusammen.

D) Potenzielle Opfer. Sie werden dann von Swiss COMPAS profitieren, wenn dadurch ein geringerer Anteil der wieder straffällig werdenden Angeklagten freigelassen wird. Dieses Interesse wird nicht unbedingt gefördert, wenn Swiss COMPAS zu einem geringeren Anteil an gewährten Bewährungsstrafen führt. Wenn das Tool einerseits weniger genau ist als menschliche Richterinnen und Richter, kann ein geringerer Anteil der gewährten Bewährungsstrafen dazu führen, dass die Quote der Straftäterinnen und Straftäter, deren Haftstrafe zur Bewährung ausgesetzt wurde, die aber erneut straffällig werden, steigt. Wenn das Tool anderseits jedoch genauer ist als menschliche Richterinnen und Richter, kann ein höherer Anteil der gewährten Bewährungsstrafen damit einhergehen, dass weniger Verbrechen von Straftäterinnen und Straftätern begangen werden, deren Haftstrafe zur Bewährung ausgesetzt wurde.

E) Das Interesse der *Gemeinschaften*, die davon profitieren könnten, kann als Kombination folgender Interessen angesehen werden:
- der Familienmitglieder der Angeklagten, wie oben in Abschnitt C angegeben,
- der Personen, deren Interessen mit denen der Familienmitglieder in Einklang stehen,
- der Interessen der potenziellen Opfer von Straftäterinnen und Straftätern, deren Haftstrafe zur Bewährung ausgesetzt wurde, wie oben in Abschnitt D angegeben,
- der Personen, deren Interessen mit denen der potenziellen Opfer von Straftäterinnen und Straftätern in Einklang stehen, deren Haftstrafe zur Bewährung ausgesetzt wurde (z. B. der Kinder des Opfers).

Ein Tool, das in der Lage ist, mehr Menschen auf Bewährung freizulassen, während gleichzeitig die Häufigkeit der erneuten Straffälligkeit von Straftäterinnen und Straftätern, deren Haftstrafe zur Bewährung ausgesetzt wurde, abnimmt, sollte von den Gemeinschaften der Angeklagten begrüsst werden.»

2.10. Welche Methoden wurden zur Definition und zum Schutz der Privatsphäre verwendet?
[Hier legen Sie dar, welche Methoden tatsächlich in Bezug auf die Privatsphäre zum Schutz der personenbezogenen Daten eingesetzt worden sind.]

2.18. Welches sind die verbleibenden Sicherheits- und Datenschutzrisiken und warum sind sie angemessen?
[Hier erklären Sie, warum das Cybersicherheitsrisiko, das sich aus den Ziffern 2.9. und 2.11. ergibt, angesichts dessen, was auf dem Spiel steht, und der Wahrscheinlichkeit einer Attacke als angemessen beurteilt wird.]

Frage

1.2. Haben böswillige Parteien besonders starke Motive, das System zu hacken? Können sie, auch durch Erpressung, einen einfachen und substanziellen finanziellen Gewinn erzielen oder kann ein gehacktes System verwendet werden, um politische Ziele zu erreichen (einschliesslich der Äusserung politischen Widerspruchs gegen das System)?

Antwort

Da die gesammelten Daten sensibel sind, muss die Cybersicherheit hoch sein, um Attacken durch motivierte Eindringlinge zu verhindern.

Konsequenz

2.2.2. Welche Anforderungen werden an das System in Bezug auf die Cybersicherheit gestellt?
[Legen Sie hier die Anforderungen an das System dar. Fordern Sie von Cybersicherheitsexpertinnen und -experten eine technische Expertise an.]

2.9. Welche Verfahren sind vorhanden, um Systemfehlern und Fehlfunktionen zu begegnen?
[Hier fügen Sie einen Abschnitt zur Cybersicherheit ein. Sie erläutern beispielsweise, wie Sie mit Fehlern von Mitarbeitenden umgehen, die die Integrität, Verfügbarkeit oder Vertraulichkeit der gesammelten Informationen gefährden. Dieser Abschnitt wird am besten von Cybersicherheitsexperten entworfen.]

2.11. Welche Massnahmen zum Schutz der Cybersicherheit wurden getroffen?
[In diesem Abschnitt erläutern Sie die im System integrierten Cybersicherheitsmassnahmen. Dieser Abschnitt wird am besten von Cybersicherheitsexperten entworfen.]

2.18. Welches sind die verbleibenden Sicherheits- und Datenschutzrisiken und warum sind sie angemessen?
[Hier erklären Sie, warum das Cybersicherheitsrisiko, das sich aus den Ziffern 2.9. und 2.11. ergibt, angesichts dessen, was auf dem Spiel steht, und der Wahrscheinlichkeit einer Attacke als angemessen beurteilt wird.]

Frage

1.3. Wird das soziotechnische System verwendet, um Entscheidungen über Personen zu treffen, zu empfehlen oder zu beeinflussen, und zwar in einer Weise, die beeinflusst, welche Entscheidung getroffen wird?

Antwort

Ja, das System liefert Ergebnisse, die von Richterinnen und Richtern verwendet werden, um zu entscheiden, ob dem Angeklagten die Aussetzung der Haft zur Bewährung gewährt wird.

Konsequenz

2.2.3. Welche Anforderungen werden an das System in Bezug auf die Fairness gestellt?
[Dieser Aspekt ist zu komplex, um hier darauf einzugehen – die Beurteilung erfordert eine gemeinsame Analyse von zumindest Expertinnen und Experten für Statistik und Kriminologie, die in der Lage sind, eine begründete Einschätzung dazu zu geben, was sie für ein faires und unvoreingenommenes Urteil halten; im Idealfall sollten Strafverteidigerinnen oder Strafverteidiger und/oder Expertinnen oder Experten für soziale Gerechtigkeit beauftragt werden oder die Ergebnisse überprüfen.]

Beispiel
Wir definieren Fairness wie folgt: Die Vorhersage, ob jemand erneut straffällig wird, ist im Durchschnitt für Männer und Frauen gleichermassen genau.
Die Rechtfertigung dafür, Fairness so zu bestimmen, ist: ...

2.2.4. Welche Anforderungen werden an das System in Bezug auf die Erklärbarkeit gestellt?

Beispielhafte Erwägungen
Da dies eine weitreichende Entscheidung ist, die letztendlich von Menschen (Richterinnen und Richtern) getroffen wird, erscheint es wichtig, dass sie ein mentales Modell der Faktoren und ihrer Gewichtung bilden können, die bei der Erzeugung des Scores berücksichtigt wurden. Sofern das Schweizer COMPAS-Tool einen geheimen Algorithmus verwendet, kann diese Anforderung möglicherweise nicht erfüllt werden. Wenn die Formel jedoch öffentlich bekannt wäre, würden um Bewährung ersuchende Personen unaufrichtig antworten, um ihre Risikobewertung zu verbessern. (Der US-COMPAS-Algorithmus ist geheim.)
Ein realer Bericht würde im Idealfall eine gründlichere Analyse und Diskussion dieses Widerspruchs sowie potenziell realisierbare Lösungen beinhalten.

2.7. Mit welchen Methoden wurde die Leistung des Systems getestet und gemessen?
[Bitte geben Sie an, wie Sie die Leistung in Bezug auf das in Checkliste 2 – Frage 2.1. angegebene Hauptziel messen.]

Beispiel
Wir haben die Leistungsfähigkeit des Algorithmus getestet, indem wir seine Vorhersagen auf der Grundlage der historischen Daten überprüft haben, die von der Kantonspolizei zu Festnahmen nach Bewährung im Kanton erhoben wurden. Wir haben die Gesamtgenauigkeit gemessen, die Falsch-Negativ-Rate und die Falsch-Positiv-Rate. Der technische Anhang ist [hier] verfügbar.

2.8.1. Welche Methoden wurden verwendet, um die von den Systemvorhersagen/-empfehlungen/-entscheidungen unmittelbar betroffenen Stakeholder zu identifizieren? Und was sind die voraussichtlichen Auswirkungen auf diese Personen?
Antwort: siehe oben (Punkt 2.8.1. bei Frage 1.1.)

2.9. Welche Verfahren sind vorhanden, um Systemfehlern und Fehlfunktionen zu begegnen?

Beispiel
«Fehlerhafte Vorhersagen bezüglich Personen, die aus dem Gefängnis entlassen wurden und als risikoarm eingestuft wurden, aber erneut straffällig geworden sind, werden durch ein spezifisches Verfahren erfasst, das den Erfolg des zu einer Bewährungsstrafe verurteilten Straftäters überwacht. Alle Daten werden sicher in X [Angabe der Datenbanken oder des Registers] gespeichert und sind für Y [Rollen der öffentlichen Verwaltung] zugänglich. Die Massnahme zur Behebung der fehlerhaften Vorhersagen lautet: Das Strafrecht sieht bereits rechtliche Konsequenzen für einen Rückfall während des Bewährungszeitraums vor. Der Anteil der Angeklagten, denen eine Bewährung gewährt oder abgelehnt wurde, wird erfasst und X, Y [Rollen in der Organisation] zugänglich gemacht. Die Möglichkeit falscher Prognosen in dem Sinne, dass ihre Risikobewertung auf Datenkorrekturfehlern basiert, wurde berücksichtigt und der folgende Plan wurde erstellt, um diese Art von Fehler zu minimieren [Verfahren zur Fehlerverminderung einfügen, falls vorhanden].»

Frage **1.4.** Wird das System verwendet, um eine Entscheidung über eine gesetzliche Pflicht oder ein Recht einer Person zu treffen?

Antwort Ja, das System wird verwendet, um zu entscheiden, ob jemand das Gefängnis verlassen darf.

Frage **1.6.** Können Einzelpersonen die Entscheidung durch Swiss COMPAS vermeiden oder verlangen, dass die Entscheidung mithilfe eines anderen Verfahrens getroffen wird, bei dem nicht dasselbe technische System verwendet wird?

Antwort Zwei möglich Szenarien:
Ja, Angeklagte können es vermeiden, von Swiss COMPAS beurteilt zu werden, indem sie dies über ihre Anwältin oder ihren Anwalt verlangen, und dies wird immer akzeptiert. Infolgedessen wird die Rückfallgefahr von einer Richterin oder einem Richter ohne die Hilfe des Tools bewertet.
Nein, alle Angeklagten müssen die Bewertung akzeptieren.

Konsequenz Fall A) weiter zu Frage 1.10.
Fall B) weiter zu Frage 1.7.

Frage **1.7.** Kann die Person, über die mithilfe des Tools eine Entscheidung getroffen wurde, beweisen, dass eine falsche Entscheidung getroffen wurde, ohne vor Gericht zu gehen?

Antwort Nein, es kann nicht gezeigt werden, dass die Entscheidung, die Aussetzung zur Bewährung zu verweigern, auf einer falschen Prognose beruht. Der Nachweis, dass jemand, der auf Bewährung freigelassen wurde, tatsächlich erneut straffällig geworden ist, zeigt nicht, dass die Entscheidung falsch war, da die Entscheidung ausdrücklich auf einer unsicheren Risikobewertung beruhte, die (manchmal) damit vereinbar ist, dass eine Angeklagte oder ein Angeklagter mit geringem Rückfallrisiko erneut gegen das Gesetz verstösst.

Konsequenz

2.5. Wer ist verantwortlich dafür, Rückmeldungen der Endbenutzerinnen und Endbenutzer zu bearbeiten, d. h. der Personen, die das System benutzen oder von ihm unterstützt werden?

Beispiel
Richterinnen und Richter können ein Informationsgespräch mit Expertinnen und Experten der Firma Swiss COMPAS anfordern, die in Laienform erklären können, was das Tool berücksichtigt und warum oder welche Art von Bias im System vorliegt.

2.6. Wer ist dafür verantwortlich, auf Zweifel oder Anfechtungen durch Einzelne zu antworten, die von der Nutzung des Systems betroffen sind?
Beispiel: Niemand.

2.14. Wird der Einsatz des Systems nach der Testphase kontinuierlich überwacht?
a) Zu jedem Zeitpunkt?
b) In einem bestimmten Zeitraum?
c) Mit welchen Massnahmen?

Beispiel
Nein. Es ist kein Überwachungssystem vorhanden, mit dem die Leistung des Systems über die Dauer seines Einsatzes verfolgt wird.

2.15. Können Personen, die von einer Entscheidung betroffen sind, den Output des automatisierten Systems erfahren und die vom System beeinflussten Vorhersagen/Empfehlungen/Entscheidungen anfechten?

Beispiel
Nein, es ist kein solches System entwickelt und implementiert worden.

2.17. Wie schneidet das System im Vergleich zu dem zuvor vorhandenen ab, falls es eines gibt, oder mit etablierten Benchmarks, falls vorhanden?

Beispiel
Das zuvor bestehende System war eine Bewährungsentscheidung menschlicher Richterinnen und Richter unter Berücksichtigung von X, Y, Z. Bislang gibt es keine verlässlichen Studien zur Gesamtgenauigkeit von Bewährungsentscheidungen menschlicher Schweizer Richterinnen und Richter (Anteil der auf Bewährung freigelassenen Personen, die erneut straffällig werden). Es gibt einige Studien in den USA, die zeigen, dass US-Prognosetools x% genauer sind als menschliche Richterinnen und Richter, aber es ist unklar, wie sich dies auf den Schweizer Kontext übertragen lässt, da die Leistung der Schweizer Richterinnen und Richter nicht bekannt ist. Um die Genauigkeit von Swiss COMPAS zu messen, gehen wir davon aus, dass für jeden Angeklagten mit einem Swiss-COMPAS-Risikowert von mehr als 0,3 unabhängig vom Geschlecht eine Empfehlung zur Verweigerung der Bewährung angenommen wird (d. h., dass, wenn dieser Schwellenwert verwendet wird, von 1000 Personen, die auf Bewährung freigelassen wurden, ungefähr 300 wieder straffällig geworden sind). Mit diesem Schwellenwert hat das Swiss COMPAS eine Genauigkeit von x%, was niedriger ist als die Genauigkeit des US-amerikanischen COMPAS-Tools mit y%. Die Falsch-Positiv-Rate (Anteil der Personen, bei denen in Testdaten ein erneuter Gesetzesverstoss vorhergesagt wurde, die aber nicht erneut straffällig geworden sind) ist bei Swiss COMPAS jedoch viel niedriger als beim US-COMPAS, während die Falsch-Negativ-Rate etwas höher liegt. Daher ist Swiss COMPAS besser als das US-amerikanische Gegenstück in der Lage, Personen zu identifizieren, die nicht erneut straffällig werden und eine Bewährung verdienen, aber es ist schlechter darin, Personen zu identifizieren, die erneut straffällig werden und deshalb im Gefängnis bleiben sollten.

2.20. Wurden während der Überwachung Vorhersagen/Empfehlungen/Entscheidungen des Systems jemals
a) von der Systemendbenutzerin oder dem Systemendbenutzer oder
b) von Personen, die Entscheidungen unterliegen, hinterfragt?

Beispiel
Das System ist in der Schweiz noch nicht in Betrieb. Es ist möglich, dass ähnliche Systeme in anderen Ländern, in denen sie eingesetzt werden, angefochten wurden, aber unsere Abteilung verfügt nicht über diese Informationen.

Frage

1.10. Betrifft die Entscheidung einen der folgenden Bereiche des öffentlichen Lebens oder Ressourcen des öffentlichen Sektors:
– die Rechtspflege,
– den Zugang zu Bildungschancen,
– den Zugang zu demokratischen Prozessen
– (usw.)?

Antwort

Ja, es wird in der Justizverwaltung (Rechtspflege) verwendet.

Konsequenz **2.8.1.** Welche Methoden wurden verwendet, um die von den Systemvorhersagen/-empfehlungen/-entscheidungen unmittelbar betroffenen Stakeholder zu identifizieren? Und was sind die voraussichtlichen Auswirkungen auf diese Personen?
Antwort: siehe oben (Punkt 2.8.1. bei Frage 1.1.)

2.20. Wurden während der Überwachung Vorhersagen/Empfehlungen/Entscheidungen des Systems jemals
a) von der Systemendbenutzerin oder dem Systemendbenutzer oder
b) von Personen, die Entscheidungen unterliegen, hinterfragt?
Antwort: siehe oben (Punkt 2.20. bei Frage 1.7.)

Frage **1.11** Findet durch die Anschaffung bzw. den Einsatz des KI-Systems in einem der folgenden Bereiche eine Änderung statt bei
– öffentlicher Computer-Infrastruktur,
– öffentlichen Datenbeständen oder
– immateriellen Vermögenswerten (z. B. Kompetenzen) im öffentlichen Sektor?

Antwort Ja.

Konsequenz **2.2.3.** Welche Anforderungen werden an das System in Bezug auf die Fairness gestellt?
Antwort: siehe oben (Punkt 2.2.3. bei Frage 1.3.)

2.8.2. Welche Methoden wurden verwendet, um die von der digitalen Transformation in der öffentlichen Verwaltung betroffenen Personen zu identifizieren (z. B. Personal der öffentlichen Verwaltung)? Und was sind die voraussichtlichen Auswirkungen auf diese Personen?

Beispiel
Die Abteilungsleiterinnen und Abteilungsleiter X, Y, Z des zuständigen Gerichts des Kantons Zürich haben sich getroffen und folgende Stakeholder und Konsequenzen identifiziert:

A) Richterinnen und Richter. Die Verwendung des Algorithmus von Swiss COMPAS wird von den meisten Haftprüfungsrichtern aus zwei Gründen sehr begrüsst. Sie haben das Gefühl, dass sie angesichts der hohen Anzahl von Fällen, bei denen eine Haftprüfung notwendig ist, nicht genügend Zeit haben, um die Profile der Angeklagten bei der Entscheidung über eine Aussetzung zur Bewährung zu überprüfen, und Druck besteht, mehr Zeit für Folgeentscheidungen sowie die Aburteilung eines Verbrechens aufzuwenden; zum Teil aufgrund der kurzen Zeit, die sie dieser Aufgabe widmen sollen, beklagen sie sich über die Subjektivität und die schlechte Genauigkeit ihrer Einschätzungen und hoffen, dass die Genauigkeit ihrer Prognosen sowohl verbessert als auch objektiver oder zumindest uniform gestaltet werden kann. Basierend auf einer internen Umfrage ist es nur einer Minderheit solcher Richterinnen und Richter wichtig, dass die Logik hinter der Risikobewertung für sie zugänglich gemacht wird, solange es starke Beweise dafür gibt, dass das Tool korrekt arbeitet.

B) Kantonale Pflichtanwältinnen und Pflichtanwälte. Die kantonalen Pflichtanwältinnen/-verteidigerinnen und Pflichtanwälte/-verteidiger begrüssen diesen Schritt nicht. Sie beschweren sich über die Undurchsichtigkeit des Tools, die es ihnen unmöglich macht, die Rechte der Menschen zu verteidigen, denen sie helfen. Sie planen den Gang zum obersten Gericht, wenn der Algorithmus hinter dem Score nicht zugänglich gemacht wird.

Frage **1.12** Besteht das Risiko, dass das System eine politische Entscheidung (z. B. Wahl oder Volksabstimmung) beeinflusst?

Antwort Nein.

Frage **1.13.** Beeinflusst das technische System die Verteilung öffentlicher Mittel an wirtschaftliche Akteure in der Gesellschaft?

Antwort Nein.

Frage **1.14.** Beruht das technische System auf einem statistischen Modell des menschlichen Verhaltens oder der persönlichen Merkmale?

Antwort Ja.

Konsequenz **2.2.3.** Welche Anforderungen werden an das System in Bezug auf die Fairness gestellt?
[Hier werden die Anforderungen an das System in Sachen Fairness erläutert. Dieser Punkt ist zu komplex, um hier zusammengefasst zu werden – idealerweise sollte dies nach einer Konsultation mit Expertinnen und Experten und betroffenen Interessenvertreterinnen und -vertretern oder Strafrechtswissenschaftlerinnen und -wissenschaftlern definiert werden.]
Antwort: siehe oben (Punkt 2.2.3. bei Frage 1.3.)

2.8.1. Welche Methoden wurden verwendet, um die von den Systemvorhersagen/-empfehlungen/-entscheidungen unmittelbar betroffenen Stakeholder zu identifizieren? Und was sind die voraussichtlichen Auswirkungen auf diese Personen?
Antwort: siehe oben (Punkt 2.8.1. bei Frage 1.1.)

2.12. Welche Methoden wurden verwendet, um die Voreingenommenheit und die Fairness des Systems zu definieren und zu messen?

Beispiel
«Wir haben die Falscherkennungsrate, die Falschauslassungsrate, die Falsch-Positiv-Rate und die Falsch-Negativ-Rate insgesamt und nach Geschlecht aufgeschlüsselt gemessen, unter der Annahme, dass die Richterinnen und Richter allen Angeklagten mit einer Risikobewertung von mehr als 0,3 die Bewährung verweigern.

Die Quote falscher Auslassungen und falscher Entdeckungen ist jedoch für beide Geschlechter gleich. Risikobewertungen haben unabhängig vom Geschlecht, für das sie verwendet werden, den gleichen Prognosewert (sie bieten einen gleich starken Beleg dafür, dass eine Straftäterin oder ein Straftäter, dessen Haftstrafe zur Bewährung ausgesetzt wird, wieder straffällig wird).
Das System ist nicht adaptiv ausgelegt, da ein adaptives System für Personen mit denselben Merkmalen kein homogenes Ergebnis ergeben würde, d. h., Personen mit denselben Merkmalen erhalten möglicherweise unterschiedliche Empfehlungen, und das in einer Weise, die von den Richterinnen und Richtern nur schwer zu kontrollieren wäre.»

2.13. Wie werden den Systemendbenutzerinnen und -endbenutzern sowie den Personen, die vom Einsatz des Systems unmittelbar betroffen sind, individuelle Vorhersagen/Empfehlungen/Entscheidungen des Systems erklärt?

Beispiel
Die Richterinnen und Richter wurden über die Merkmale informiert, die vom Vorhersagetool berücksichtigt werden. Die Formel und die spezifische Gewichtung dieser Merkmale sind jedoch ein Geschäftsgeheimnis und wurden ihnen aus diesem Grund nicht offenbart.

2.19. Bitte beschreiben Sie relevante Probleme mit Bias, die nicht gelöst werden konnten, oder mögliche Ursachen für Ungerechtigkeiten im System und erklären Sie, warum sie nicht gelöst werden können (beispielsweise, indem Sie Kompromisse mit anderen Systemzielen einschliesslich widersprüchlicher Fairnessziele erläutern).

Beispiel
Es ist unmöglich, alle oben genannten Masse für alle Gruppen auszugleichen, da die durchschnittliche Rückfallwahrscheinlichkeit bei Männern und Frauen unterschiedlich ist. Für diesen Schwellenwert unterscheiden sich die Falsch-Positiv-Rate und die Falsch-Negativ-Rate für die Geschlechter. Weibliche Gefangene haben eine höhere Falsch-Negativ-Rate, aber eine niedrigere Falsch-Positiv-Rate, d. h., es ist im Vergleich zu Männern weniger wahrscheinlich, dass sie zu Unrecht inhaftiert werden, und es ist wahrscheinlicher, dass sie zu Unrecht freigelassen werden.
Männliche Gefangene haben eine höhere Falsch-Positiv-Rate und eine niedrigere Falsch-Negativ-Rate, d. h., es ist wahrscheinlicher, dass sie fälschlicherweise im Gefängnis festgehalten werden, und weniger wahrscheinlich, dass sie fälschlicherweise inhaftiert werden.
Angesichts der in 2.2.3 angegebenen Fairnessziele und der Bedeutung des Ausgleichs der Rate falscher Entdeckungen und falscher Auslassungen (sowie der Übermittlung kalibrierter Bewertungen an die Richterinnen und Richter), die sich aus dieser Analyse der Ziele des Systems ergibt, gehen wir davon aus, dass dies aus Sicht der Fairness die beste Lösung ist.

2.20. Wurden während der Überwachung Vorhersagen/Empfehlungen/Entscheidungen des Systems jemals
a) von der Systemendbenutzerin oder dem Systemendbenutzer oder
b) von Personen, die Entscheidungen unterliegen, hinterfragt?
Antwort: siehe oben (Punkt 2.20. bei Frage 1.7.)

Frage **1.15.** Ist das System so konzipiert, dass es adaptiv ist, sodass nicht alle neuen Fälle wie andere behandelt werden, denen es in der Vergangenheit begegnet ist, weil es seine Parameter ändert, z. B., um effizienter zu werden?

Antwort Nein.

Frage **1.16.** Ist es das Ziel des technischen Systems, ein vollständig deterministisches Regelsystem zu automatisieren, das nur ein Minimum an Kreativität und menschlichem Urteilsvermögen durch die derzeitigen menschlichen Anwenderinnen/Anwender erfordert und keine Risiko- oder Wahrscheinlichkeitsabschätzungen beinhaltet?

Antwort Nein.

Frage

1.17. Beruht das technische System auf Parametern, Merkmalen, Faktoren oder Entscheidungskriterien, die nicht dem entsprechen, was von den meisten Fachleuten auf diesem Gebiet normalerweise berücksichtigt wird?

Antwort

Ja, da das System auf Merkmalen und Faktoren beruht, die von menschlichen Richterinnen und Richtern normalerweise nicht (oder zumindest nicht systematisch) berücksichtigt werden.

Konsequenz

2.2.4. Welche Anforderungen werden an das System in Bezug auf die Erklärbarkeit gestellt?
Antwort: siehe oben (Punkt 2.2.4. bei Frage 1.3.)

2.8.1. Welche Methoden wurden verwendet, um die von den Systemvorhersagen/-empfehlungen/-entscheidungen unmittelbar betroffenen Stakeholder zu identifizieren? Und was sind die voraussichtlichen Auswirkungen auf diese Personen?
Antwort: siehe oben (Punkt 2.8.1. bei Frage 1.1.)

2.13. Wie werden den Systemendbenutzerinnen und -endbenutzern sowie den Personen, die vom Einsatz des Systems unmittelbar betroffen sind, individuelle Vorhersagen/Empfehlungen/Entscheidungen des Systems erklärt?
Antwort: siehe oben (Punkt 2.13. bei Frage 1.14.)

2.14. Wird die Systembereitstellung nach der Testphase kontinuierlich überwacht?
a) Zu jedem Zeitpunkt?
b) In einem bestimmten Zeitrahmen?
c) Mit welchen Massnahmen?
Antwort: siehe oben (Punkt 2.14. bei Frage 1.7.)

2.20. Wurden während der Überwachung Vorhersagen/Empfehlungen/Entscheidungen des Systems jemals
a) von der Systemendbenutzerin oder dem Systemendbenutzer oder
b) von Personen, die Entscheidungen unterliegen, hinterfragt?
Antwort: siehe oben (Punkt 2.20. bei Frage 1.7.)

Frage

1.19. Greift das technische System auf die Infrastruktur eines Drittanbieters zurück, über die die öffentliche Einrichtung keine uneingeschränkte Kontrolle hat und/oder bei der sie keinen Zugriff auf z. B. Datensätze oder die Rechenleistung hat?

Antwort

Ja, der Algorithmus zur Erstellung des Risiko-Scores ist ein Geschäftsgeheimnis der Firma Swiss COMPAS.

Konsequenz

2.8.2. Welche Methoden wurden verwendet, um die von der digitalen Transformation in der öffentlichen Verwaltung betroffenen Personen zu identifizieren (z. B. Personal der öffentlichen Verwaltung)? Und was sind die voraussichtlichen Auswirkungen auf diese Personen?
Antwort: siehe oben (Punkt 2.8.2. bei Frage 1.11.)

Frage

Überprüfung
Resultiert aus der Beantwortung der Fragen in Checkliste 1, dass Sie einen Transparenzbericht schreiben sollen?

Antwort

Ja.

Konsequenz

2.1. Für welches Problem soll das System eine Lösung liefern?

Beispiel
Derzeit gibt es Zweifel an der Qualität von Bewährungsentscheidungen durch Richterinnen und Richter, die auch damit zusammenhängen, dass es ein Missverhältnis gibt zwischen der geringen Zeit, die den Richterinnen und Richtern zur Verfügung steht, um diese Entscheidungen zu treffen, und der Anzahl der Fälle, in denen sie entscheiden müssen. Swiss COMPAS soll Richterinnen und Richtern eine Risikoanzeige in Verbindung mit einer optimierten Empfehlung (Bewährung zu gewähren oder zu verweigern) geben, die sie allerdings auch ignorieren dürfen. Das Ziel des Systems ist, bei der Prognose von Rückfällen sowohl genauer als auch homogener zu sein als die derzeitige Einschätzung durch menschliche Richterinnen und Richter.

2.3. Wer ist für die Konstruktion des Systems verantwortlich?

Beispiel
Die Firma «Southpointe» mit Hauptsitz in Lugano, CH, in der Person von Gianni Contabene, CEO.

2.4. Wer ist für die Implementierung des Systems verantwortlich?
[Geben Sie die Rollen an, z. B. im kantonalen Strafgericht, die für die Haftprüfungsverfahren zuständig sind.]

2.5. Wer ist verantwortlich dafür, Rückmeldungen der Endbenutzerinnen und Endbenutzer zu bearbeiten, d. h. der Personen, die das System benutzen oder von ihm unterstützt werden?
Antwort: siehe oben (Punkt 2.5 bei Frage 1.7.)

2.6. Wer ist dafür verantwortlich, auf Zweifel oder Anfechtungen durch Einzelne zu antworten, die von der Nutzung des Systems betroffen sind?
Antwort: siehe oben (Punkt 2.6 bei Frage 1.7)

2.14. Wird die Systembereitstellung nach der Testphase kontinuierlich überwacht?
a) Zu jedem Zeitpunkt?
b) In einem bestimmten Zeitrahmen?
c) Mit welchen Massnahmen?
Antwort: siehe oben (Punkt 2.14. bei Frage 1.7.)

2.15. Können Personen, die von einer Entscheidung betroffen sind, den Output des automatisierten Systems erfahren und die vom System beeinflussten Vorhersagen/Empfehlungen/Entscheidungen anfechten?
Antwort: siehe oben (Punkt 2.15. bei Frage 1.7.)

2.16. Wie verhält sich das System in Bezug auf die ausgewählten relevanten Metriken?
[Bitte beachten Sie alle Ziele und Anforderungen, die in Checkliste 2 – Fragen 2.1. und 2.2. angegeben sind.]
Das erste Ziel ist, dass die Vorhersage so genau wie möglich ist: Es soll die Anzahl der Personen minimiert werden, denen die Bewährung verweigert wird, obwohl sie nicht wieder straffällig würden, und derjenigen, die auf Bewährung entlassen werden, aber wieder eine Straftat begehen würden. Das System hat eine Gesamtgenauigkeit von x %. Der Vergleich mit relevanten nationalen und internationalen Benchmarks wird weiter oben diskutiert, siehe Abschnitt 2.17.

2.17. Wie schneidet das System im Vergleich zu dem zuvor vorhandenen ab, falls es eines gibt, oder mit etablierten Benchmarks, falls vorhanden?
Antwort: siehe oben (Punkt 2.17. bei Frage 1.7.)

2.19. Bitte beschreiben Sie relevante Probleme mit Bias, die nicht gelöst werden konnten, oder mögliche Ursachen für Ungerechtigkeiten im System und erklären Sie, warum sie nicht gelöst werden können (beispielsweise, indem Sie Kompromisse mit anderen Systemzielen einschliesslich widersprüchlicher Fairnessziele erläutern).
Antwort: siehe oben (Punkt 2.19. bei Frage 1.14.)

2.20. Wurden während der Überwachung Vorhersagen/Empfehlungen/Entscheidungen des Systems jemals hinterfragt?

Frage	**Letzte Nachfrage** Gibt es zusätzliche ethische Fragen?
Konsequenz	**Beispiel** Zusätzliche ethische Fragen sind uns nicht bekannt.

3. Transparenzbericht

Es folgt der Transparenzbericht, der sich aus der Beantwortung der Checkliste 1 für diesen hypothetischen Fall (Swiss COMPAS) ergibt. Er weist Antworten auf alle Fragen auf, doch das ist üblicherweise nicht zu erwarten. Der Transparenzbericht soll Antworten enthalten, die durch die Fragen in Checkliste 1 notwendig werden. Dass hier alle Fragen der Checkliste beantwortet werden, dient lediglich der Veranschaulichung.

a) Was soll das System leisten und welchen Anforderungen an den Schutz von Grundwerten soll es gerecht werden?

2.1. Für welches Problem soll das System eine Lösung liefern?

Beispiel
Derzeit gibt es Zweifel an der Qualität von Bewährungsentscheidungen durch Richterinnen und Richter, die auch damit zusammenhängen, dass es ein Missverhältnis zwischen der geringen Zeit, die den Richterinnen und Richtern zur Verfügung steht, um diese Entscheidungen zu treffen, und der Anzahl der Fälle, in denen sie entscheiden müssen, gibt. Swiss COMPAS soll Richterinnen und Richtern eine Risikoanzeige in Verbindung mit einer optimierten Empfehlung (Bewährung zu gewähren oder zu verweigern) geben, die sie allerdings auch ignorieren dürfen. Das Ziel des Systems ist, bei der Prognose von Rückfällen sowohl genauer als auch homogener zu sein als die derzeitige Einschätzung durch menschliche Richterinnen und Richter.

2.2. Weitere Anforderungen des Systems?

2.2.1. Welche Anforderungen werden an das System in Bezug auf die Privatsphäre gestellt?
[Legen Sie hier die Anforderungen an das System dar. Welche Massnahmen bezüglich des Datenschutzes sollten ergriffen werden? Nehmen Sie beispielsweise Kontakt mit der/dem Datenschutzbeauftragten Ihrer Organisation auf, um diesen Teil des Berichts zu verfassen.]

2.2.2. Welche Anforderungen werden an das System in Bezug auf die Cybersicherheit gestellt?
[Legen Sie hier die Anforderungen an das System dar. Fordern Sie von Cybersicherheitsexperten eine technische Expertise an.]

2.2.3. Welche Anforderungen werden an das System in Bezug auf die Fairness gestellt?
[Dieser Aspekt ist zu komplex, um hier darauf einzugehen – die Beurteilung erfordert eine gemeinsame Analyse von zumindest Expertinnen und Experten für Statistik und Kriminologie, die in der Lage sind, eine begründete Einschätzung dazu zu geben, was sie für ein faires und unvoreingenommenes Urteil halten; im Idealfall sollten Strafverteidigerinnen und Strafverteidiger und/oder Expertinnen und Experten für soziale Gerechtigkeit beauftragt werden oder die Ergebnisse überprüfen.]

Beispiel
Wir definieren Fairness wie folgt: Die Vorhersage, ob jemand erneut straffällig wird, ist im Durchschnitt für Männer und Frauen gleichermassen genau.
Die Rechtfertigung dafür, Fairness so zu bestimmen, ist: ...

2.2.4. Welche Anforderungen werden an das System in Bezug auf die Erklärbarkeit gestellt?

Beispielhafte Erwägungen
Da dies eine weitreichende Entscheidung ist, die letztendlich von Menschen (Richterinnen und Richtern) getroffen wird, erscheint es wichtig, dass sie ein mentales Modell der Faktoren und ihrer Gewichtung bilden können, die bei der Erzeugung des Scores berücksichtigt wurden. Sofern das Schweizer COMPAS-Tool einen geheimen Algorithmus verwendet, kann diese Anforderung möglicherweise nicht erfüllt werden. Wenn die Formel jedoch öffentlich bekannt wäre, würden um Bewährung ersuchende Personen unaufrichtig antworten, um ihre Risikobewertung zu verbessern. (Der US-COMPAS-Algorithmus ist geheim.)
Ein realer Bericht würde im Idealfall eine gründlichere Analyse und Diskussion dieses Widerspruchs sowie potenziell realisierbare Lösungen beinhalten.

b) Wer ist rechenschaftspflichtig?

2.3. Wer ist für die Konstruktion des Systems verantwortlich?

Beispiel
Die Firma «Southpointe» mit Hauptsitz in Lugano, CH, in der Person von Gianni Contabene, CEO.

2.4. Wer ist für die Implementierung des Systems verantwortlich?
[Geben Sie die Rollen an, z. B. im kantonalen Strafgericht, die für die Haftprüfungsverfahren zuständig sind.]

2.5. Wer ist verantwortlich dafür, Rückmeldungen der Endbenutzerinnen und Endbenutzer zu bearbeiten, d. h. der Personen, die das System benutzen oder von ihm unterstützt werden?

Beispiel
Richterinnen und Richter können ein Informationsgespräch mit Expertinnen und Experten der Firma Swiss COMPAS anfordern, die in Laienform erklären können, was das Tool berücksichtigt und warum oder welche Art von Bias im System vorliegt.

2.6. Wer ist dafür verantwortlich, auf Zweifel oder Anfechtungen durch Einzelne zu antworten, die von der Nutzung des Systems betroffen sind?

Beispiel
Niemand.

c) Transparenzinformationen über die Umsetzung und Steuerung des Systems

2.7. Mit welchen Methoden wurde die Leistung des Systems getestet und gemessen?
[Bitte geben Sie an, wie Sie die Leistung in Bezug auf das in Checkliste 2 – Frage 2.1. angegebene Hauptziel messen.]

2.8. Welche Methoden wurden verwendet?

2.8.1. Welche Methoden wurden verwendet, um die von den Systemvorhersagen/-empfehlungen/-entscheidungen unmittelbar betroffenen Stakeholder zu identifizieren? Und was sind die voraussichtlichen Auswirkungen auf diese Personen?

Beispiel
«Wir haben ein Brainstorming-Meeting mit mit Staatsanwältinnen und Staatsanwälten, Richterinnen Richtern des Kantons sowie Anwältinnen und Anwälten der Strafjustiz durchgeführt. Bei diesem Treffen wurden die Stakeholder identifiziert, die direkt von den Vorhersagen betroffen sind, nämlich die Angeklagten, ihre Verteidiger, ihre Familien, potenzielle künftige Opfer, wenn die Angeklagten erneut straffällig werden, und Gemeinschaften, in denen Menschen, die möglicherweise erneut straffällig werden, leben.

Unsere Analyse der Stakeholderinteressen sieht wie folgt aus:

A) Angeklagte. Das System zu verwenden liegt im Interesse der Angeklagten, bei denen es unwahrscheinlich ist, dass sie erneut straffällig werden (oder die statistisch nicht von denen zu unterscheiden sind, bei denen es unwahrscheinlich ist, dass sie erneut straffällig werden). Es ist insbesondere im Interesse derjenigen, die ihr Recht auf ein günstiges Bewährungsurteil am schlechtesten ausüben können, da sie sich nicht die besten Anwältinnen und Anwälte leisten können. Es ist nicht im Interesse von Personen, die sich mithilfe guter Anwältinnen und Anwälte bessere Chancen verschaffen können, auf Bewährung entlassen zu werden.

B) Anwältinnen und Anwälte der Angeklagten. Das System ist nicht in ihrem Interesse, da es ein Bestandteil richterlicher Entscheidung sein wird, den die Anwältinnen und Anwälte nicht anfechten können.

C) Familien. Die Familien der Angeklagten werden dann von der höheren Wahrscheinlichkeit profitieren, dass die oder der Angeklagte auf Bewährung freigelassen wird, wenn der Einsatz von Swiss COMPAS im Vergleich zum Status quo zu einem höheren Anteil an gewährten Bewährungsstrafen führt (es sei denn, Angeklagte sind wegen eines Verbrechens gegen ihre Familien angeklagt). Dies hängt eng mit der Verhältnismässigkeit der Entscheidungen zur Gewährung der Bewährung zusammen.

D) Potenzielle Opfer. Sie werden dann von Swiss COMPAS profitieren, wenn dadurch ein geringerer Anteil der wieder straffällig werdenden Angeklagten freigelassen wird. Dieses Interesse wird nicht unbedingt gefördert, wenn Swiss COMPAS zu einem geringeren Anteil an gewährten Bewährungsstrafen führt. Wenn das Tool einerseits weniger genau ist als menschliche Richterinnen und Richter, kann ein geringerer Anteil der gewährten Bewährungsstrafen dazu führen, dass die Quote der Straftäterinnen und Straftäter, deren Haftstrafe zur Bewährung ausgesetzt wurde, die aber erneut straffällig werden, steigt. Wenn das Tool anderseits jedoch genauer ist als menschliche Richterinnen und Richter, kann ein höherer Anteil der gewährten Bewährungsstrafen damit einhergehen, dass weniger Verbrechen von Straftäterinnen und Straftätern begangen werden, deren Haftstrafe zur Bewährung ausgesetzt wurde.

E) Das Interesse der *Gemeinschaften,* die davon profitieren könnten, kann als Kombination folgender Interessen angesehen werden:
- der Familienmitglieder der Angeklagten, wie oben in Abschnitt C angegeben,
- der Personen, deren Interessen mit denen der Familienmitglieder in Einklang stehen,
- der Interessen der potenziellen Opfer von Straftäterinnen und Straftätern, deren Haftstrafe zur Bewährung ausgesetzt wurde, wie oben in Abschnitt D angegeben,
- der Personen, deren Interessen mit denen der potenziellen Opfer von Straftäterinnen und Straftätern in Einklang stehen, deren Haftstrafe zur Bewährung ausgesetzt wurde (z. B. der Kinder des Opfers).

Ein Tool, das in der Lage ist, mehr Menschen auf Bewährung freizulassen, während gleichzeitig die Häufigkeit der erneuten Straffälligkeit von Straftäterinnen und Straftätern, deren Haftstrafe zur Bewährung ausgesetzt wurde, abnimmt, sollte von den Gemeinschaften der Angeklagten begrüsst werden.»

2.8.2. Welche Methoden wurden verwendet, um die von der digitalen Transformation in der öffentlichen Verwaltung betroffenen Personen zu identifizieren (z. B. Personal der öffentlichen Verwaltung)? Und was sind die voraussichtlichen Auswirkungen auf diese Personen?

Beispiel

Die Abteilungsleiterinnen und Abteilungsleiter X, Y, Z des zuständigen Gerichts des Kantons Zürich haben sich getroffen und folgende Stakeholder und Konsequenzen identifiziert:

A) Richterinnen und Richter. Die Verwendung des Algorithmus von Swiss COMPAS wird von den meisten Haftprüfungsrichtern aus zwei Gründen sehr begrüsst. Sie haben das Gefühl, dass sie angesichts der hohen Anzahl von Fällen, bei denen eine Haftprüfung notwendig ist, nicht genügend Zeit haben, um die Profile der Angeklagten bei der Entscheidung über eine Aussetzung zur Bewährung zu überprüfen, und Druck besteht, mehr Zeit für Folgeentscheidungen sowie die Aburteilung eines Verbrechens aufzuwenden; zum Teil aufgrund der kurzen Zeit, die sie dieser Aufgabe widmen sollen, beklagen sie sich über die Subjektivität und die schlechte Genauigkeit ihrer Einschätzungen und hoffen, dass die Genauigkeit ihrer Prognosen sowohl verbessert als auch objektiver oder zumindest uniform gestaltet werden kann. Basierend auf einer internen Umfrage ist es nur einer Minderheit solcher Richterinnen und Richter wichtig, dass die Logik hinter der Risikobewertung für sie zugänglich gemacht wird, solange es starke Beweise dafür gibt, dass das Tool korrekt arbeitet.

B) Kantonale Pflichtanwältinnen und Pflichtanwälte. Die kantonalen Pflichtanwältinnen/-verteidigerinnen und Pflichtanwälte/-verteidiger begrüssen diesen Schritt nicht. Sie beschweren sich über die Undurchsichtigkeit des Tools, die es ihnen unmöglich macht, die Rechte der Menschen zu verteidigen, denen sie helfen. Sie planen den Gang zum obersten Gericht, wenn der Algorithmus hinter dem Score nicht zugänglich gemacht wird.

2.9. Welche Verfahren sind vorhanden, um Systemfehler und Fehlfunktionen zu behandeln?
[Hier fügen Sie einen Abschnitt zur Cybersicherheit ein. Sie erläutern beispielsweise, wie Sie mit Fehlern von Mitarbeitenden umgehen, die die Integrität, Verfügbarkeit oder Vertraulichkeit der gesammelten Informationen gefährden. Dieser Abschnitt wird am besten von Cybersicherheitsexperten entworfen.]

2.10. Welche Methoden wurden zur Definition und zum Schutz der Privatsphäre verwendet?
[Bitte gehen Sie in diesem Teil des Berichts spezifisch auf die in Checkliste 1 – Frage 1.1. genannten Aspekte ein.]

2.11. Welche Massnahmen zum Schutz der Cybersicherheit wurden getroffen?
[In diesem Abschnitt erläutern Sie die im System integrierten Cybersicherheitsmassnahmen. Dieser Abschnitt wird am besten von Cybersicherheitsexperten entworfen.]

2.12. Welche Methoden wurden verwendet, um die Voreingenommenheit und die Fairness des Systems zu definieren und zu messen?

Beispiel
Wir haben die Falscherkennungsrate, die Falschauslassungsrate, die Falsch-Positiv-Rate und die Falsch-Negativ-Rate insgesamt und nach Geschlecht aufgeschlüsselt gemessen, unter der Annahme, dass die Richterinnen und Richter allen Angeklagten mit einer Risikobewertung von mehr als 0,3 die Bewährung verweigern.
Das System ist nicht adaptiv ausgelegt, da ein adaptives System für Personen mit denselben Merkmalen kein homogenes Ergebnis ergeben würde, d. h., Personen mit denselben Merkmalen erhalten möglicherweise unterschiedliche Empfehlungen, und das in einer Weise, die von den Richterinnen und Richtern nur schwer zu kontrollieren wäre.

2.13. Wie werden den Systemendbenutzerinnen und -endbenutzern sowie den Personen, die vom Einsatz des Systems unmittelbar betroffen sind, individuelle Vorhersagen/Empfehlungen/Entscheidungen des Systems erklärt?

Beispiel
Die Richterinnen und Richter wurden über die Merkmale informiert, die vom Vorhersagetool berücksichtigt werden. Die Formel und die spezifische Gewichtung dieser Merkmale sind jedoch ein Geschäftsgeheimnis und wurden ihnen aus diesem Grund nicht offenbart.

2.14. Wird die Systembereitstellung nach der Testphase kontinuierlich überwacht?
a) Zu jedem Zeitpunkt?
b) In einem bestimmten Zeitrahmen?
c) Mit welchen Massnahmen?

Beispiel
Nein. Es ist kein Überwachungssystem vorhanden, mit dem die Leistung des Systems über die Dauer seines Einsatzes verfolgt wird.

2.15. Können Personen, die von einer Entscheidung betroffen sind, den Output des automatisierten Systems erfahren und die vom System beeinflussten Vorhersagen/Empfehlungen/Entscheidungen anfechten?

Beispiel
Nein, es ist kein solches System entwickelt und implementiert worden.

d) Transparenzinformationen über die Leistungen des Systems

Auf der Grundlage der bisherigen Testläufe:

2.16. Wie verhält sich das System in Bezug auf die ausgewählten relevanten Metriken?
[Bitte beachten Sie alle Ziele und Anforderungen, die in Checkliste 2 – Fragen 2.1. und 2.2. angegeben sind.]

Beispiel
Das erste Ziel ist, dass die Vorhersage so genau wie möglich ist: Es soll die Anzahl der Personen minimiert werden, denen die Bewährung verweigert wird, obwohl sie nicht wieder straffällig würden, und derjenigen, die auf Bewährung entlassen werden, aber wieder eine Straftat begehen würden.
Das System hat eine Gesamtgenauigkeit von x %. Der Vergleich mit relevanten nationalen und internationalen Benchmarks wird weiter oben diskutiert, siehe Abschnitt 2.17.

2.17. Wie schneidet das System im Vergleich zu dem zuvor vorhandenen ab, falls es eines gibt, oder mit etablierten Benchmarks, falls vorhanden?

Beispiel
Das zuvor bestehende System war eine Bewährungsentscheidung menschlicher Richterinnen und Richter unter Berücksichtigung von X, Y, Z. Bislang gibt es keine verlässlichen Studien zur Gesamtgenauigkeit von Bewährungsentscheidungen menschlicher Schweizer Richterinnen und Richter (Anteil der auf Bewährung freigelassenen Personen, die erneut straffällig werden). Es gibt einige Studien in den USA, die zeigen, dass US-Prognosetools x % genauer sind als menschliche Richterinnen und Richter, aber es ist unklar, wie sich dies auf den Schweizer Kontext übertragen lässt, da die Leistung

der Schweizer Richterinnen und Richter nicht bekannt ist. Um die Genauigkeit von Swiss COMPAS zu messen, gehen wir davon aus, dass für alle Angeklagten mit einem Swiss-COMPAS-Risikowert von mehr als 0,3 unabhängig vom Geschlecht eine Empfehlung zur Verweigerung der Bewährung angenommen wird (d. h., dass, wenn dieser Schwellenwert verwendet wird, von 1000 Personen, die auf Bewährung freigelassen wurden, ungefähr 300 wieder straffällig geworden sind). Mit diesem Schwellenwert hat das Swiss COMPAS eine Genauigkeit von x %, was niedriger ist als die Genauigkeit des US-amerikanischen COMPAS-Tools mit y %. Die Falsch-Positiv-Rate (Anteil der Personen, bei denen in Testdaten ein erneuter Gesetzesverstoss vorhergesagt wurde, die aber nicht erneut straffällig geworden sind) ist bei Swiss COMPAS jedoch viel niedriger als beim US-COMPAS, während die Falsch-Negativ-Rate etwas höher liegt. Daher ist Swiss COMPAS besser als das US-amerikanische Gegenstück in der Lage, Personen zu identifizieren, die nicht erneut straffällig werden und eine Bewährung verdienen, aber es ist schlechter darin, Personen zu identifizieren, die erneut straffällig werden und deshalb im Gefängnis bleiben sollten.

2.18. Welches sind die verbleibenden Sicherheits- und Datenschutzrisiken und warum sind sie angemessen?
[Hier erklären Sie, warum das Cybersicherheitsrisiko, das sich aus den Ziffern 2.9. und 2.11. ergibt, angesichts dessen, was auf dem Spiel steht, und der Wahrscheinlichkeit einer Attacke als angemessen beurteilt wird.]

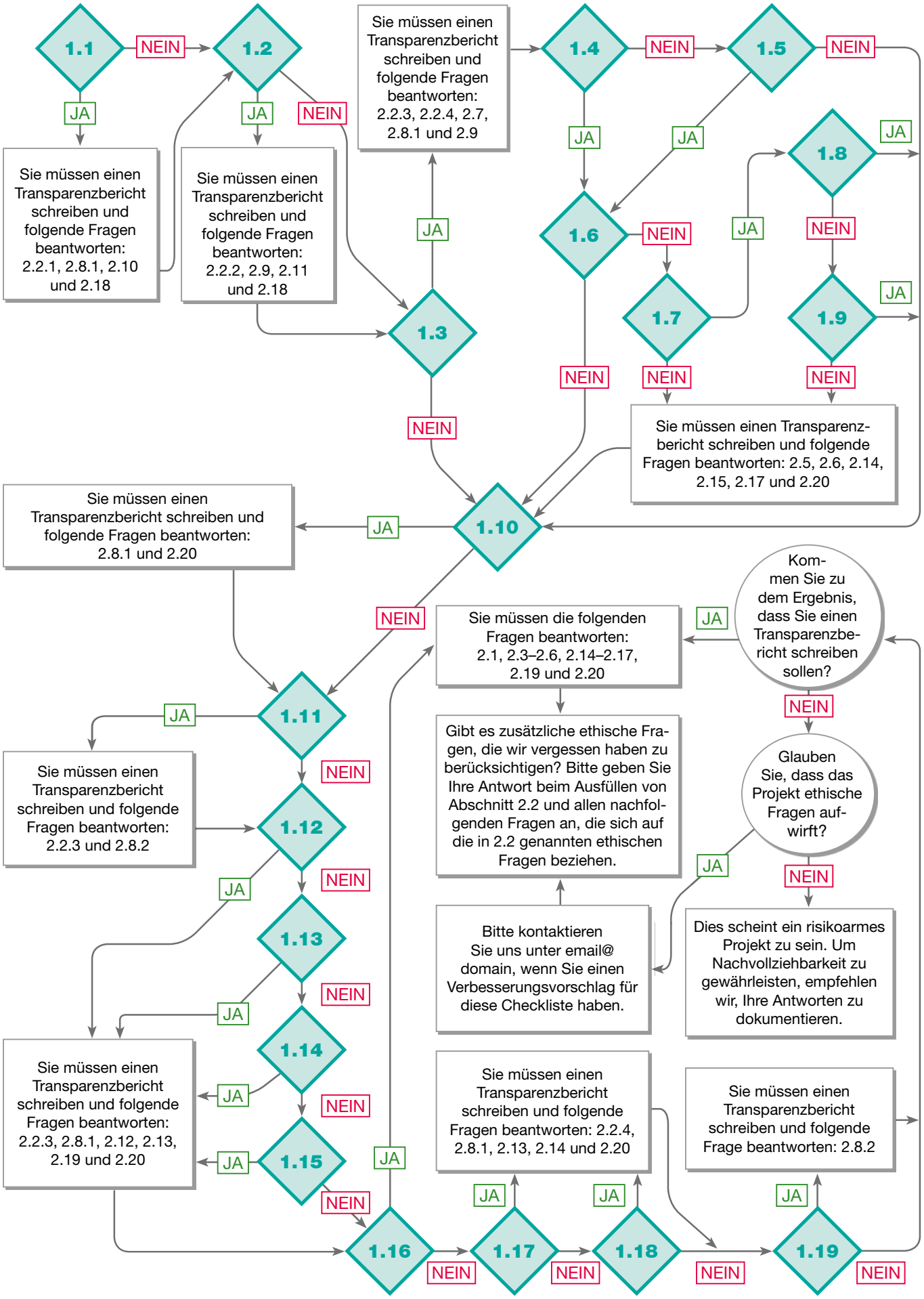
2.19. Bitte beschreiben Sie relevante Probleme mit Bias, die nicht gelöst werden konnten, oder mögliche Ursachen für Ungerechtigkeiten im System und erklären Sie, warum sie nicht gelöst werden können (beispielsweise, indem Sie Kompromisse mit anderen Systemzielen einschliesslich widersprüchlicher Fairnessziele erläutern).

Beispiel
Es ist unmöglich, alle oben genannten Masse für alle Gruppen auszugleichen, da die durchschnittliche Rückfallwahrscheinlichkeit bei Männern und Frauen unterschiedlich ist. Für diesen Schwellenwert unterscheiden sich die Falsch-Positiv-Rate und die Falsch-Negativ-Rate für die Geschlechter. Weibliche Gefangene haben eine höhere Falsch-Negativ-Rate, aber eine niedrigere Falsch-Positiv-Rate, d. h. es ist im Vergleich zu Männern weniger wahrscheinlich, dass sie zu Unrecht inhaftiert werden, und es ist wahrscheinlicher, dass sie zu Unrecht freigelassen werden.
Männliche Gefangene haben eine höhere Falsch-Positiv-Rate und eine niedrigere Falsch-Negativ-Rate, d. h., es ist wahrscheinlicher, dass sie fälschlicherweise im Gefängnis festgehalten werden, und weniger wahrscheinlich, dass sie fälschlicherweise inhaftiert werden.
Angesichts der in 2.2.3 angegebenen Fairnessziele und der Bedeutung des Ausgleichs der Rate falscher Entdeckungen und falscher Auslassungen (sowie der Übermittlung kalibrierter Bewertungen an die Richterinnen und Richter), die sich aus dieser Analyse der Ziele des Systems ergibt, gehen wir davon aus, dass dies aus Sicht der Fairness die beste Lösung ist.

2.20. Wurden während der Überwachung Vorhersagen/Empfehlungen/ Entscheidungen des Systems jemals hinterfragt?

Beispiel
Das System ist in der Schweiz noch nicht in Betrieb. Es ist möglich, dass ähnliche Systeme in anderen Ländern, in denen sie eingesetzt werden, angefochten wurden, aber unsere Abteilung verfügt nicht über diese Informationen.

V. Flussdiagramm



Kapitel 5

Ausblick

Nadja Braun Binder
Matthias Spielkamp
Catherine Egli

Künstliche Intelligenz wird die öffentliche Verwaltung verändern. Festzustellen, in welcher Form und in welchem Ausmass diese Transformation stattfinden wird oder soll, ist schwierig. Grund dafür ist u. a. die schwierige Eingrenzung der Thematik. Als interdisziplinäres Forschungsfeld ist KI ein Oberbegriff für ein breites Themenfeld von verschiedensten Anwendungen, die aus sich ständig weiterentwickelnden technischen Fortschritten und Zielen resultieren. Auch wenn bereits jetzt zahlreiche KI-Anwendungen den Alltag beeinflussen, sind in der öffentlichen Schweizer Verwaltung bisher nur sehr wenige KI-Systeme zu finden, die zudem noch in frühen Entwicklungsphasen stecken. Folglich wurden auch im Kanton Zürich bis heute nur wenige Erfahrungen mit behördlichem KI-Einsatz gemacht, weshalb die Erwartungen an KI und die Herausforderungen, die damit einhergehen, derzeit noch sehr allgemeiner Natur sind.

Sowohl im Ausland als auch im Privatwirtschaftssektor sind KI-Anwendungen verbreiteter und haben daher bereits wichtige ethische Debatten im demokratischen Diskurs im Allgemeinen und in der Wissenschaft im Speziellen angestossen. In der Schweiz und insbesondere im Kanton Zürich beginnt erst langsam ein gesellschaftlicher Diskurs zum Umgang mit KI. Im Rahmen dieser Studie war es folglich mangels spezialisierter gemeinnütziger Organisationen nicht möglich, die Betroffenensicht umfassend zu berücksichtigen – eine Sichtweise, die für die Entwicklung eines erfolgreichen und wünschenswerten KI-Einsatzes jedoch unabdingbar ist. Die vorliegende Untersuchung kann somit trotz interdisziplinärer Fragestellung und Zusammensetzung des Projektkonsortiums keine Gesamtansicht abbilden. Die Ausführungen beschränken sich auf eine Auswahl an zentralen rechtlichen und ethischen Fragestellungen, die unter Berücksichtigung des kurzen Zeitfensters bearbeitet werden konnten. Selbstverständlich spielen weitere Perspektiven wie etwa diejenige der Informationstechnik eine elementare Rolle, die für eine umfassende Beurteilung des staatlichen KI-Einsatzes berücksichtigt werden müssen.

Aus rechtlicher Perspektive kommen mit dem Einsatz von KI auf jeden Fall neue Herausforderungen auf den Rechtsstaat zu. Der Gesetz- und Verordnungsgeber wird sich bei der Schaffung des rechtlichen Rahmens etwa damit auseinandersetzen müssen, wie rechtsstaatlichen Begründungserfordernissen entsprochen, das Äusserungsrecht der Betroffenen gewährleistet sowie der Untersuchungsgrundsatz umgesetzt werden kann. Weiter kommt sowohl datenschutzrechtlichen Überlegungen als auch offenen Normen und Ermessensspielräumen neue Bedeutung zu. Eine zentrale Herausforderung, die alle Disziplinen intensiv beschäftigt und für die noch keine umfassende Lösung gefunden wurde, ist schliesslich die Sicherstellung eines diskriminierungsfreien Staatshandelns. Der Einsatz von KI kann diskriminierende Haltungen nicht nur perpetuieren und gesellschaftliche Veränderungen verhindern, sondern auch zur Entwicklung neuer Diskriminierungsmuster beitragen. Diesem Risiko ist besondere Aufmerksamkeit zu schenken.

Letztendlich variieren Risiken und Chancen von KI jedoch je nach Verwaltungsbereich und konkreter Anwendung stark. Deren Identifikation, Ausschöpfung und Begegnung mit gezielten Lösungsvorschlägen für die Rechtsetzung und -anwendung müssen bei tatsächlichen und konkreten Vorhaben vertieft werden. Die Ausarbeitung dieses rechtsstaatlichen Rahmens kann sich dabei namentlich auf ethische Grundsätze stützen. Ethische Leitlinien zu Schadensvermeidung, Gerechtigkeit, Fairness, Autonomie, Benefizienz, Kontrolle, Transparenz und Rechenschaftspflicht liefern hilfreiche Bausteine dazu, in einer Demokratie erwünschte bzw. akzeptierte Wertehaltungen freizulegen und zu entwickeln. Durch eine entsprechende Gesetzgebung und damit die Herstellung demokratischer Legitimation werden der öffentlichen Verwaltung die notwendigen Rahmenbedingungen vorgegeben, innerhalb deren sie KI einsetzen kann. Ethische Leitlinien können die Verwaltung zudem in der Rechtsanwendung begleiten.

Mit dem vorliegenden Bericht möchten wir einen Beitrag zum notwendigen gesellschaftspolitischen Diskurs leisten. Die Rechtsordnung bezüglich des Umgangs mit staatlichem KI-Einsatz konkret auszugestalten, muss schliesslich aber im Rahmen der etablierten demokratischen Prozesse erfolgen.

Literaturverzeichnis

Die nachstehenden Werke werden, wenn nichts anderes angegeben ist, mit Nachnamen der Autorin oder des Autors, Jahreszahl sowie mit Seitenzahlen oder Randnummern zitiert.

Albertini Michele (2000), Der verfassungsmässige Anspruch auf rechtliches Gehör im Verwaltungsverfahren des modernen Staates, Eine Untersuchung über Sinn und Gehalt der Garantie unter besonderer Berücksichtigung der bundesgerichtlichen Rechtsprechung, Diss. Bern

Albus James (1991), Outline for a theory of intelligence, in: Vol. 21, No. 3, IEEE Transactions on Systems, Man, and Cybernetics

Aliotta Massimo (2014), § 6 Medizinische Begutachtung, in: Steiger-Sackmann Sabine / Mosimann Hans-Jakob (Hrsg.), Recht der Sozialen Sicherheit, Sozialversicherungen, Opferhilfe, Sozialhilfe – Beraten und Prozessieren, Basel, S. 245 ff.

Allhutter Doris / Mager Astrid / Cech Florian / Fischer Fabian / Grill Gabriel (2020), Der AMS-Algorithmus. Eine Soziotechnische Analyse des Arbeitsmarktchancen-Assistenz-Systems (AMAS), Institut für Technikfolgen-Abschätzung der Österreichischen Akademie der Wissenschaften, abrufbar unter http://epub.oeaw.ac.at/0xc1aa5576_0x003bfdf3.pdf

Althaus-Houriet Isabelle (2017), Kommentierung der Art. 122–131 DBG, Noël Yves / Aubry Girardin Florence (Hrsg.), Commentaire Romand de la loi sur l'Impôt fédéral direct, 2. Aufl., Basel

Auer Christoph / Müller Markus / Schindler Benjamin (Hrsg.) (2018), VwVG, Kommentar zum Bundesgesetz über das Verwaltungsverfahren, 2. Aufl., Zürich (zit. Bearbeiter/in, 2018, N. ... zu Art. ... VwVG)

Baeriswyl Bruno / Pärli Kurt (Hrsg.) (2015), Stämpfli Handkommentar zum Datenschutzgesetz, Bern (zit. Bearbeiter/in, 2015, N. ... zu § ... DSG)

Baeriswyl Bruno / Rudin Beat (Hrsg.) (2012), Praxiskommentar zum Informations- und Datenschutzgesetz des Kantons Zürich (IDG), Zürich/Basel/Genf (zit. Bearbeiter/in, 2012, N. ... zu § ... IDG)

Beauchamp Tom L. / Childress James F. (2008), Principles of Biomedical Ethics, 6. Aufl., New York

Beck Susanne / Grunwald Armin / Jacon Kai / Matzner Tobias (2019), Künstliche Intelligenz und Diskriminierung, Herausforderungen und Lösungsansätze, Whitepaper aus der Plattform Lernende Systeme, München, abrufbar unter https://www.plattform-lernende-systeme.de/publikationen-details/kuenstliche-intelligenz-und-diskriminierung-herausforderungen-und-loesungsansaetze.html

Biaggini Giovanni (2017), Kommentar zur Bundesverfassung der Schweizerischen Eidgenossenschaft, 2. Aufl., Zürich (zit. Biaggini, 2017, N. … zu Art. … BV)

Blumenstein Ernst / Locher Peter (2016), System des schweizerischen Steuerrechts, 7. Aufl. von Professur Dr. Peter Locher, Zürich/Basel/Genf

Bomhard David (2019), Automatisierung und Entkollektivierung betrieblicher Arbeitsorganisation, Herausforderungen einer digitalen Arbeitsweilt, Diss. München 2018

Braun Binder Nadja / Brändli Daniel (2003), Vote électronique – Abstimmen und Wählen per Mausklick, in: LeGes 2003, S. 125 ff.

Braun Binder Nadja (2016a), Auf dem Weg zum vollautomatisierten Besteuerungsverfahren in Deutschland, in: Jusletter IT vom 25. Mai 2016

Braun Binder Nadja (2016b), Ausschliesslich automationsgestützt erlassene Steuerbescheide und Bekanntgabe durch Bereitstellung zum Datenabruf, in: DStZ 2016, S. 526 ff.

Braun Binder Nadja (2016c), Vollständig automatisierter Erlass eines Verwaltungsaktes und Bekanntgabe über Behördenportale, in: DÖV 21/2016, S. 891 ff.

Braun Binder Nadja (2016d), Vollautomatisierte Verwaltungsverfahren im allgemeinen Veraltungsverfahrensrecht?, in: NvWZ 2016, S. 960 ff.

Braun Binder Nadja (2018), Algorithmic Regulation – Der Einsatz algorithmischer Verfahren im staatlichen Steuerungskontext, in: Hill Hermann / Wieland Joachim (Hrsg.), Zukunft der Parlamente – Speyer Konvent in Berlin, S. 107 ff.

Braun Binder Nadja (2019a), Vollautomatisiert erlassene Verwaltungsakte und elektronische Aktenführung, in: Seckelmann Margrit (Hrsg.), Digitalisierte Verwaltung – Vernetztes E-Government, Berlin 2019, S. 311 ff.

Braun Binder Nadja (2019b), Künstliche Intelligenz und automatisierte Entscheidungen in der öffentlichen Verwaltung, in: SJZ 15/2019, S. 467 ff.

Braun Binder Nadja (2020a), Automatisierte Entscheidungen: Perspektive Datenschutzrecht und öffentliche Verwaltung, in: SZW 1/2020, S. 27 ff.

Braun Binder Nadja (2020b), Als Verfügung gelten Anordnungen der Maschinen im Einzelfall … – Dystopie oder künftiger Verwaltungsalltag?, in: ZSR 139/2020, S. 253 ff.

Braun Binder Nadja (2020c), Der Untersuchungsgrundsatz als Herausforderung vollautomatisierter Verfahren, in: zsis) 2/2020, A5, S. 26 ff.

Braun Binder Nadja (2020d), AI and Taxation: Risk Management in Fully Automated Taxation Procedures, in: Rademacher Timo / Wischmeyer Thomas (Hrsg.), Regulating Artificial Intelligence, Wiesbaden, S. 295 ff.

Bryson Joanna (2017), Three very different sources of bias in AI, and how to fix them, abrufbar unter https://joanna-bryson.blogspot.com/2017/07/three-very-different-sources-of-bias-in.html

Bürkle Martin (2020), Kommentierungen zu Art. 1 und 2, in: Frésard-Fellay Ghislaine / Klett Barbara / Leuzinger Susanne (Hrsg.), Basler Kommentar zum Allgemeinen Teil des Sozialversicherungsrechts (zit. Bürkle, 2020, N. … zu Art. … ATSG)

Danaher John (2016a), The Threat of Algocracy: Reality, Resistance and Accommodation, in: Philosophy & Technology 03/2016, S. 245 ff., abrufbar unter https://doi.org/10.1007/s13347-015-0211-1

Danaher John (2016b), Will Life Be Worth Living in a World Without Work? Technological Unemployment and the Meaning of Life, in: Science and Engineering Ethics 01/2016, abrufbar unter https://doi.org/10.1007/s11948-016-9770-5

Dawson et al. (2020), Artificial Intelligence: Australia's Ethics Framework, abrufbar unter https://consult.industry.gov.au/strategic-policy/artificial-intelligence-ethics-framework/supporting_documents/ArtificialIntelligenceethicsframeworkdiscussionpaper.pdf

de Laat Paul B. (2017), Algorithmic Decision-Making Based on Machine Learning from Big Data: Can Transparency Restore Accountability?, in: Philosophy & Technology, abrufbar unter https://doi.org/10.1007/s13347-017-0293-z

Degrandi Benno (1977), Die automatisierte Verwaltungsverfügungen, Diss. Zürich

Demaj Labinot / Sägesser Patrick (2017), Chatbots für Organisationen: Anatomie, Potentiale und Anwendungsmöglichkeiten zur direkten Kommunikation mit Anspruchsgruppen, Diskussionspapier, abrufbar unter https://byerley.ch/assets/pdf/bylerley_2017_Chatbots_Diskussionspapier.pdf

Djeffal Christian (2017), Das Internet der Dinge und die öffentliche Verwaltung – Auf dem Weg zum automatisierten Smart Government?, in: DVBl 2017, S. 808 ff.

Djeffal Christian (2020), Künstliche Intelligenz, in: Klenk Tanja / Nullmeier Frank/Wewer Goettrik (Hrsg.): Handbuch Digitalisierung in Staat und Verwaltung, Wiesbaden

Döring Nicola / Bortz Jürgen (2016), Forschungsmethoden und Evaluation in den Sozial- und Humanwissenschaften, 5. Aufl., Berlin

Egli Catherine (2020), Automatisierte Einzelentscheidungen: Regelungsbedarf im VwVG?, Die Einführung des Begriffs «automatisierte Einzelentscheidungen» und dessen Auswirkungen auf das Verwaltungsverfahrensrecht, Masterarbeit vom 12.03.2020 an der Universität Basel

Ehrenzeller Bernhard / Schindler Benjamin / Schweizer Rainer J. / Vallender Klaus A. (Hrsg.) (2014), Bundesverfassung St. Galler Kommentar, 3. Aufl., Zürich (zit. Bearbeiter/in, 2014, N. ... zu Art. ... BV)

Eifert Martin (2006), Electronic Government: Das Recht der elektronischen Verwaltung, Baden-Baden

Engelmann Jan / Puntschuh Michael (2020), KI im Behördeneinsatz: Erfahrungen und Empfehlungen, Kompetenzzentrum Öffentliche IT (Hrsg), abrufbar unter https://oeffentliche-it.de/documents/10181/14412/KI+im+Beh%C3%B6rdeneinsatz+-+Erfahrungen+und+Empfehlungen

Epiney Astrid (2011), in: Belser Eva Maria / Epiney Astrid / Waldmann Bernhard (Hrsg.), § 9 (Allgemeine Grundsätze)

Ertel Wolfgang (2016), Grundkurs Künstliche Intelligenz, Eine praxisorientierte Einführung, 4. Aufl., Wiesbaden

Etscheid Jan (2018), Automatisierungspotenziale in der Verwaltung, in: Mohabbat Kar Resa / Thapa Basanta / Paryecek Peter (Hrsg.), (Un)Berechenbar? Algorithmen und Automatisierung in Staat und Gesellschaft, Berlin, S. 126 ff.

Etscheid Jan / von Lucke Jörn / Stroh Felix (2020), Künstliche Intelligenz in der öffentlichen Verwaltung, Stuttgart

Felzmann Heike et al. (2019), Transparency You Can Trust: Transparency Requirements for Artificial Intelligence between Legal Norms and Contextual Concerns, in: Big Data & Society 6, no. 1, abrufbar unter https://doi.org/10.1177/2053951719860542

Fischer Claudio (2020), Die digitale Steuerverwaltung, in: zsis) 2/2020, A6, S. 39 ff.

Fischer Claudio / Daepp Annemarie (2019), Digitalisierung am Beispiel der Steuerverwaltung des Kantons Bern, in: EF 4/19, S. 327 ff.

Flick Uwe (2004), Triangulation – Eine Einführung, Wiesbaden

Floridi Luciano / Cowls Josh (2019), A Unified Framework of Five Principles for AI in Society, abrufbar unter https://doi.org/10.1162/99608f92.8cd550d1

Flückiger Thomas (2014), § 4 Verwaltungsverfahren, in: Steiger-Sackmann Sabine / Mosimann Hans-Jakob (Hrsg.), Recht der Sozialen Sicherheit, Sozialversicherungen, Opferhilfe, Sozialhilfe – Beraten und Prozessieren, Basel, S. 97 ff.

Friedmann Batya / Nissenbaum Helen (1996), Bias in Computer Systems, in: ACM Transactions of Information Systems, 03/1996, S. 330 ff.

Gächter Thomas / Burch Stephanie (2014), § 1 Nationale und internationale Rechtsquellen, in: Steiger-Sackmann Sabine / Mosimann Hans-Jakob (Hrsg.): Recht der Sozialen Sicherheit, Sozialversicherungen, Opferhilfe, Sozialhilfe – Beraten und Prozessieren, Basel, S. 3 ff.

Gerstner Dominik (2017), Predictive Policing als Instrument zur Prävention von Wohnungseinbruchdiebstahl: Evaluationsergebnisse zum Baden-Württembergischen Pilotprojekt P4, abrufbar unter http://hdl.handle.net/11858/00-001M-0000-002E-384A-F

Glaser Andreas (2018), Einflüsse der Digitalisierung auf das schweizerische Verwaltungsrecht, in: SJZ 114/2018, S. 181 ff.

Glaser Andreas / Ehrat Marco (2019), E-Government-Gesetzgebung durch die Kantone – Integration in die Verfahrenskodifikation oder Auslagerung in Spezialerlasse?, in: LeGes 2019, S. 1 ff.

Glass Philip (2018), Gedanken zur Revision des DSG, abrufbar unter https://www.datalaw.ch/gedanken-zur-revision-des-dsg/

Goodman Bryce / Flaxman Seth (2016), European Union regulations on algorithmic decision-making and a «right to explanation», Oxford, abrufbar unter https://arxiv.org/pdf/1606.08813.pdf

Görz Günther / Schmid Ute / Wachsmuth Ipke (2014), Einleitung, in: Görz Günther / Schneeberger Josef / Schmid Ute (Hrsg.), Handbuch der Künstlichen Intelligenz, 5. Aufl., München, S. 1 ff.

Griffel Alain (2017), Allgemeines Verwaltungsrecht im Spiegel der Rechtsprechung, Zürich/Basel/Genf

Griffel Alain (Hrsg.) (2014), Kommentar zum Verwaltungsrechtspflegegesetz des Kantons Zürich (VRG), 3. Aufl., Zürich/Basel/Genf (zit. Bearbeiter/in, 2014, N. ... zu § ... VRG)

Guckelberger Annette (2019), Öffentliche Verwaltung im Zeitalter der Digitalisierung, Baden-Baden

Häfelin Ulrich / Müller Georg / Uhlmann Felix (2020), Allgemeines Verwaltungsrecht, 8. Aufl., Zürich / St. Gallen

Hagendorff Thilo (2019), Maschinelles Lernen und Diskriminierungen: Probleme und Lösungsansätze, in: ÖZS 01/2019, S. 53 ff.

Hammerschmid Gerhard / Raffer Christian (2020), Künstliche Intelligenz im öffentlichen Sektor: Potenziale nutzen, Risiken bedenken, abrufbar unter https://publicgovernance.de/media/KI_Oeffentliche_Verwaltung.pdf

Hanania Pierre-Adrien / Knobloch Tobias (2020), Künstliche Intelligenz im öffentlichen Sektor – Teil 1, abrufbar unter https://www.capgemini.com/de-de/wp-content/uploads/sites/5/2020/10/PublicGoesAI-PoV-Part1-23122020.pdf

Harasgama Rehana C. (2017), Erfahren – Wissen – Vergessen, Zur zeitlichen Dimension des staatlichen Informationsanspruches, Zürich / St. Gallen

Hoffmann Jens / Glaz-Ocik Justine (2012), DyRiAS-Intimpartner: Konstruktion eines online gestützten Analyse-Instrument zur Risikoeinschätzung von tödlicher Gewalt gegen aktuelle oder frühere Intimpartnerinnen, in: Polizei und Wissenschaft, 2/2012, S. 45 ff.

Jaag Tobias / Rüssli Markus (2019), Staats- und Verwaltungsrecht des Kantons Zürich, 5. Aufl.

Jobin Anna / Ienca Marcello / Vayena Effy (2019), The Global Landscape of AI Ethics Guidelines, in: Nature Machine Intelligence 1, no 9, S. 389 ff., abrufbar unter https://www.nature.com/articles/s42256-019-0088-2

Karlen Peter (2018), Schweizerisches Verwaltungsrecht – Gesamtdarstellung unter Einbezug des europäischen Kontextes, Zürich

Kayser-Bril Nicolas (2019), Austria's employment agency rolls out discriminatory algorithm, sees no problem, abrufbar unter https://algorithmwatch.org/en/story/austrias-employment-agency-ams-rolls-out-discriminatory-algorithm/

Kessler Rainer / Oberlin Jutta Sonja (2020), Künstliche Intelligenz: Quo Vadis?, in: Compliance Berater 2020, S. 89 ff.

Kiener Regina / Kälin Walter / Wyttenbach Judith (2018), Grundrechte, 3. Aufl., Bern

Kiener Regina / Rütsche Bernhard / Kuhn Mathias (2015), Öffentliches Verfahrensrecht, 2. Aufl., Zürich / St. Gallen

Kieser Ueli (2019), Leistungen der Sozialversicherung, 3. Aufl., Zürich

Kieser, Ueli (2020), Kommentar zum Bundesgesetz über den Allgemeinen Teil des Sozialversicherungsrechts ATSG, 4. Aufl., Zürich (zit. Kieser, 2020, N. … zu Art. … ATSG)

Kirn Stefan / Müller-Hengstenberg Claus (2013), Intelligente (Software-)Agenten: Von der Automatisierung zur Autonomie) Verselbständigung technischer Systeme, in: MMR 2013, S. 225 ff.

Knobloch Tobias (2018), Vor die Lage kommen: Predictive Policing in Deutschland – Chancen und Gefahren datenanalytischer Prognosetechnik und Empfehlungen für den Einsatz in der Polizeiarbeit, abrufbar unter https://www.bertelsmann-stiftung.de/fileadmin/files/BSt/Publikationen/GrauePublikationen/predictive.policing.pdf

Kolleck Alma / Orwat Carsten (2020), Mögliche Diskriminierung durch algorithmische Entscheidungssysteme und maschinelles Lernen – ein Überblick, Büro für Technikfolgen-Abschätzung beim Deutschen Bundestag, abrufbar unter https://www.tab-beim-bundestag.de/de/pdf/publikationen/berichte/TAB-Hintergrundpapier-hp024.pdf

Kroll Joshua A. et al. (2016/2017), Accountable Algorithms, in: University of Pennsylvania Law Review 165, S. 633 ff., abrufbar unter https://scholarship.law.upenn.edu/penn_law_review/vol165/iss3/3/

Krüger Jochen / Vogelgesang Stephanie / Adam Lena-Marie (2020), Verantwortungsbewusste Digitalisierung, gerichtliche Entscheidungen und der Gedanke des fairen Verfahrens, in: Jusletter IT vom 28. Februar 2020

Leese Matthias (2018), Predictive Policing in der Schweiz: Chancen, Herausforderungen, Risiken, in: Bulletin zur schweizerischen Sicherheitspolitik 2018, S. 57 ff.

Leslie David (2019), Understanding Artificial Intelligence Ethics and Safety, A guide for the responsible design and implementation of AI systems in the public sector, The Alan Turing Institute, abrufbar unter https://doi.org/10.5281/zenodo.3240529

Locher Peter (2015), Kommentar zum Bundesgesetz über die direkte Bundessteuer, III. Teil – Art. 102–222 DBG, Basel (zit. Locher, 2015, N. … zu Art. … DBG)

Loi Michele (2015), Technological Unemployment and Human Disenhancement, in: Ethics and Information Technology, S. 1 ff., abrufbar unter https://doi.org/10.1007/s10676-015-9375-8

Loi Michele (2020), People Analytics Must Benefit the People, An Ethical Analysis of Data-Driven Algorithmic Systems in Human Resources Management, abrufbar unter https://algorithmwatch.org/wp-content/uploads/2020/03/AlgorithmWatch_AutoHR_Study_Ethics_Loi_2020.pdf

Loi Michele / Ferrario Andrea / Viganò Eleonora (2020), Transparency as Design Publicity: Explaining and Justifying Inscrutable Algorithms, in: Ethics and Information Technology, abrufbar unter https://doi.org/10.1007/s10676-020-09564-w

Loi Michele / Heitz Christoph / Christen Markus (2020), A Comparative Assessment and Synthesis of Twenty Ethics Codes on AI and Big Data, in: 2020 7th Swiss Conference on Data Science (SDS), S. 41 ff., abrufbar unter https://ieeexplore.ieee.org/document/9145014

Mainzer Klaus (2019), Künstliche Intelligenz – Wann übernehmen die Maschinen?, 2. Aufl., Berlin

Martini Mario (2019), Blackbox Algorithmus – Grundfragen einer Regulierung Künstlicher Intelligenz, Berlin

Martini Mario / Nink David (2017), Wenn Maschinen entscheiden … – vollautomatisierte Verwaltungsverfahren und der Persönlichkeitsschutz, in: NVwZ – Extra 36/2017 Nr. 10, S. 1 ff.

Meier-Mazzucato Giorgio (2015), Steuern Schweiz, Grundriss zu den eidgenössischen und kantonalen Steuern mit Beispielen und Darstellungen, Bern

Meyer Christian (2019), Die Mitwirkungsmaxime im Verwaltungsverfahren des Bundes – Ein Beitrag zur Sachverhaltsfeststellung als arbeitsteiligem Prozess, Diss. Luzern

Mittelstadt Brent / Russell Chris / Wachter Sandra (2019), Explaining Explanations in AI, in: Proceedings of the Conference on Fairness, Accountability, and Transparency, S. 279 ff., abrufbar unter https://doi.org/10.1145/3287560.3287574

Müller Jörg Paul / Schefer Markus (2008), Grundrechte in der Schweiz, Im Rahmen der Bundesverfassung, der EMRK und der UNO-Pakte, 4. Aufl., Bern

Nufer Marianne (2019/2020), Künstliche Intelligenz in der Steuerveranlagung, in: ASA 88 (2019/2020), S. 259 ff.

Opiela Nicole / Mohabbat Kar Resa / Thapa Basanta / Weber Mike (2018), Exekutive KI 2030, Vier Zukunftsszenarien für Künstliche Intelligenz in der öffentlichen Verwaltung, Kompetenzzentrum Öffentliche IT (Hrsg.), abrufbar unter https://www.oeffentliche-it.de/documents/10181/14412/Exekutive+KI+2030+-+Vier+Zukunftsszenarien+für+Künstliche+Intelligenz+in+der+öffentlichen+Verwaltung

Ramge Thomas (2018), Mensch und Maschine, Wie Künstliche Intelligenz und Roboter unser Leben verändern, 2. Aufl., Ditzingen

Rawls John (1999), A Theory of Justice, 2. Aufl., Cambridge

Rechsteiner David (2018), Der Algorithmus verfügt, Verfassungs- und verwaltungsrechtliche Aspekte automatisierter Einzelentscheidungen, in: Jusletter vom 26. November 2018

Reich Markus (2020), Steuerrecht, 3. Aufl., Zürich

Reichwald Julian / Pfisterer Dennis (2016), Autonomie und Intelligenz im Internet der Dinge, Möglichkeiten und Grenzen autonomer Handlungen, in: CR 3/2016, S. 208 ff.

Reisman Dillon et al. (2018), Algorithmic Impact Assessments: A Practical Framework for Public Agency Accountability, AI Now Institute, abrufbar unter https://ainowinstitute.org/aiareport2018.pdf

Rhinow René (1983), Vom Ermessen im Verwaltungsrecht: eine Einladung zum Nach- und Umdenken Teil 2, in: recht 1983, S. 87 ff.

Rhinow René / Koller Heinrich / Kiss Christina / Thurnherr Daniela / Brühl-Moser Denise (2014), Öffentliches Prozessrecht, Grundlagen und Bundesrechtspflege, 3. Aufl., Basel

Rhinow René / Schefer Markus / Uebersax Peter (2016), Schweizerisches Verfassungsrecht, 3. Aufl., Basel

Richner Felix / Frei Walter / Kaufmann Stefan / Meuter Hans Ulrich (2013), Kommentar zum Zürcher Steuergesetz, 3. Aufl., Zürich (zit. Richner/Frei/Kaufmann/Meuter, N. … zu Art. … StG)

Ringeisen Peter / Bertolosi-Lehr Andrea / Demaj Labinot (2018), Automatisierung und Digitalisierung in der öffentlichen Verwaltung: digitale Verwaltungsassistenten als neue Schnittstelle zwischen Bevölkerung und Gemeinwesen, in: YSAS 91, 2018, S. 51 ff., abrufbar unter https://doi.org/10.5334/ssas.123

Rosenthal David (2020): Das neue Datenschutzgesetz, in: Jusletter vom 16. November 2020

Rudin Beat (2017), Anpassungsbedarf in den Kantonen, in: digma 2017, S. 58 ff.

Sachs Michael / Schmitz Heribert (Hrsg.) (2018), Kommentar zum Verwaltungsverfahrensgesetz: VwVfG, 9. Aufl., München (zit. Bearbeiter/in, 2018, N. ... zu § ... VwVfG)

Sägesser Thomas (2014), Kommentierungen zu Art. 26 in: Graf Martin / Theler Cornelia / von Wyss Moritz (Hrsg.), Parlamentsrecht und Parlamentspraxis der Schweizerischen Bundesversammlung, Kommentar zum Parlamentsgesetz (ParlG) vom 13. Dezember 2002, Basel (zit. Sägesser, 2014, N. … zu Art. 26 ParlG)

Samek Wojciech / Montavon Grégoire / Vedaldi Andrea / Hansen Lars Kai / Müller Klaus-Robert (Hrsg.) (2019), Explainable AI: Interpreting, Explaining and Visualizing Deep Learning, Cham

Schindler Benjamin (2010), Verwaltungsermessen, Zürich / St. Gallen

Scholta Hendrik / Mertens William / Kowalkiewicz Marek / Becker Jörg (2019), From one-stop shop to no-stop shop: An e-government stage model, in: Government Information Quarterly 36, S. 11 ff.

Simmler Monika / Brunner Simone / Schedler Kuno (2020), Smart Criminal Justice – Eine empirische Studie zum Einsatz von Algorithmen in der Schweizer Polizeiarbeit und Strafrechtspflege, St. Gallen, abrufbar unter https://www.alexandria.unisg.ch/261666/1/Simmler%20et%20al._Smart%20Criminal%20Justice_Forschungsbericht%20vom%2010.12.2020.pdf

Simmler Monika (Hrsg.) (2021), Smart Criminal Justice. Der Einsatz von Algorithmen in der Polizeiarbeit und in der Strafrechtspflege, Basel

Söbbing Thomas (2018), Künstliche Intelligenz im HR-Recruiting-Prozess: Rechtliche Rahmenbedingungen und Möglichkeiten, in: InTer 2018, S. 64 ff.

Stiemerling Oliver (2015), «Künstliche Intelligenz» – Automatisierung geistiger Arbeit, Big Data und das Internet der Dinge, Eine technische Perspektive, in: CR 12/2015, S. 762 ff.

Thouvenin Florent / Braun Binder Nadja (i. V.), Datenschutzerklärungen für den Bund

Thurnherr Daniela (2013), Verfahrensgrundrechte und Verwaltungshandeln, Habil. Basel

Treuthardt Daniel / Loewe-Baur Mirjam / Kröger Melanie (2018), Der Risikoorientierte Sanktionenvollzug (ROS) – aktuelle Entwicklungen, in: SKZ 2/2018, S. 24 ff.

Tschannen Pierre / Zimmerli Ulrich / Müller Markus (2014), Allgemeines Verwaltungsrecht, 4. Aufl., Bern

Uhlmann Felix / Stojanovic Jasna (2017), Vertrauen im Finanzmarktrecht aus öffentlich-rechtlicher Sicht, in: SZW 2017

Vieth Kilian / Wagner Ben (2017), Teilhabe, ausgerechnet, Wie algorithmische Prozesse Teilhabechancen beeinflussen können, Gütersloh

Vokinger Kerstin Noëlle / Mühlematter Urs Jakob / Becker Anton / Boss Andreas / Reutter Mark A. / Szucs Thomas D. (2017), Artificial Intelligence und Machine Learning in der Medizin, in: Jusletter vom 28. August 2017

Von Lucke Jörn (2019), Disruptive Modernisierung von Staat und Verwaltung durch den gezielten Einsatz von smarten Objekten, cyberphysischen Systemen und künstlicher Intelligenz, Digitalisierung von Staat und Verwaltung – Gemeinsame Fachtagung Verwaltungsinformatik (FTVI) und Fachtagung Rechtsinformatik (FTRI) 2019, S. 49 ff.

Von Lucke Jörn / Etscheid Jan (2020), Wie Ansätze Künstlicher Intelligenz die öffentliche Verwaltung und die Justiz verändern könnten, in: Jusletter vom 21. Dezember 2020

Waldmann Bernhard / Belser Eva Maria / Epiney Astrid (Hrsg.) (2015), Basler Kommentar, Bundesverfassung (zit. Bearbeiter/in, 2015, N. ... zu Art. ... BV)

Waldmann Bernhard / Wiederkehr René (2019), Allgemeines Verwaltungsrecht, Zürich/Basel/Genf

Weber Rolf H. (2019), Digitalisierung und der Kampf ums Recht, in: APARIUZ 2019, S. 1 ff.

Weber Rolf H. (2020), Automatisierte Entscheidungen: Perspektive Grundrechte, in: SZW 2020, S. 18 ff.

Weber Rolf H. / Henseler Simon (2020), Regulierung von Algorithmen in der EU und in der Schweiz, in: EuZ 2020, S. 28 ff.

Weber-Dürler Beatrice (2001), Neuere Entwicklung des Vertrauensschutzes, in: ZBl 103/2001

Widmer Dieter (2019), Die Sozialversicherung in der Schweiz, 12. Aufl., Zürich

Wiederkehr René (2010), Die Begründungspflicht nach Art. 29 Abs. 2 BV und die Heilung bei Verletzung, in: ZBl 111/2010, S. 481 ff.

Wiederkehr René (2016), Öffentliches Verfahrensrecht, Bern

Wischmeyer Thomas (2018): Regulierung intelligenter Systeme, in: Di Fabio Udo / Eifert Martin / Huber Peter M. (Hrsg.), AöR 143/2018 Nr. 1, S. 1 ff.

Zanella Andrea / Bui Nicola / Castellani Angelo / Vangelista Lorenzo / Zorzi Michele (2014), Internet of Things for Smart Cities, in: IEEE, Vol. 1 No. 1, 2014, abrufbar unter https://ieeexplore.ieee.org/stamp/stamp.jsp?tp=&arnumber=6740844

Zarsky Tal Z. (2013), Transparent Predictions, in: University of Illinois Law Review, S. 1503 ff., abrufbar unter https://illinoislawreview.org/wp-content/ilr-content/articles/2013/4/Zarsky.pdf

Zavadil Andreas (2020), Datenschutzrechtliche Zulässigkeit des «AMS-Algorithmus», abrufbar unter https://www.dsb.gv.at/dam/jcr:dcc945ef-8666-4859-9362-060dedb12f1c/Newsletter-4-2020.pdf

Zweifel Martin (2017), Kommentierung der Art. 109–119, 122–135 DBG, Zweifel Martin / Beusch Michael, Bundesgesetz über die direkte Bundessteuer (DBG), Kommentar zum Schweizerischen Steuerrecht, 3. Aufl., Basel 2017 (zit. Zweifel, 2017, N. … zu Art. … DBG)

Zweifel Martin / Casanova Hugo / Beusch Michael / Hunziker Silvia (2018), Schweizerisches Steuerverfahrensrecht – Direkte Steuern, 2. Aufl., Zürich/Basel/Genf

Zweifel Martin / Hunziker Silvia (2008/2009), Steuerverfahrensrecht, Beweislast, Drittvergleich, «dealing at arm's length», Art. 29 Abs. 2 BV, Art. 58 DBG, Beweis und Beweislast im Steuerverfahren bei der Prüfung von Leistung und Gegenleistung unter dem Gesichtswinkel des Drittvergleichs («dealing at arm's length»), in: ASA 77 (2008/09), S. 657 ff.

Zweig Katharina (2016), 1. Arbeitspapier, Was ist ein Algorithmus?. Berlin, abrufbar unter https://algorithmwatch.org/publication/arbeitspapier-was-ist-ein-algorithmus/

Zweig Katharina (2019a), Algorithmische Entscheidungen: Transparenz und Kontrolle, in: Analyse und Argumente, Digitale Gesellschaft, Nr. 338

Zweig Katharina (2019b), Ein Algorithmus hat kein Taktgefühl, Wo Künstliche Intelligenz sich irrt, warum uns das betrifft und was wir dagegen tun können, München

Materialienverzeichnis

AI Now Institute et al., Using Procurement Instruments to Ensure Trustworthy, abrufbar unter https://assets.mofoprod.net/network/documents/Using_procurement_instruments_to_ensure_trustworthy_AI.pdf (zit. AI Now Institute) (zit. AI Now Institute et al.)

AlgorithmWatch (2020), Automating Society Report 2020, A report by AlgorithmWatch in cooperation with Bertelsmann Stiftung, supported by the Open Society Foundations, Berlin, abrufbar unter https://automatingsociety.algorithmwatch.org (zit. Automating Society Report 2020)

Amtsblatt Zürich (2004), Antrag des Regierungsrates vom 21. Juli 2004 zur Änderung des Steuergesetzes, Amtsblatt Kanton Zürich, S. 810 ff. (zit. ABl ZH 2004)

Automated Decision Systems Task Force (2019), New York City Automated Decision Systems Task Force Report, abrufbar unter https://www1.nyc.gov/assets/adstaskforce/downloads/pdf/ADS-Report-11192019.pdf (zit. New York City, 2019)

Bitkom/DFKI (2017), Deutsches Forschungszentrum für Künstliche Intelligenz (Hrsg.)

Entscheidungsunterstützung mit Künstlicher Intelligenz, abrufbar unter https://www.bitkom.org/sites/default/files/file/import/171012-KI-Gipfelpapier-online.pdf (zit. Bitkom/DFKI, 2017)

Botschaft vom 15. September 2017 zum Bundesgesetz über die Totalrevision des Bundesgesetzes über den Datenschutz und die Änderung weiterer Erlasse zum Datenschutz, BBl 2017 6941 ff. (zit. Botschaft E-DSG)

Botschaft vom 19. Februar 2003 zur Änderung des Bundesgesetzes über den Datenschutz (DSG) und zum Bundesbeschluss betreffend den Beitritt der Schweiz zum Zusatzprotokoll vom 8. November 2021 zum Übereinkommen zum Schutz des Menschen bei der automatischen Verarbeitung personenbezogener Daten bezüglich Aufsichtsbehörden und grenzüberschreitende Datenübermittlung, BBl 2003 2101 ff. (zit. Botschaft DSG und Zusatzprotokoll)

Botschaft vom 25. Mai 1983 zu den Bundesgesetzen über die Harmonisierung der direkten Steuern der Kantone und Gemeinden sowie über die direkte Bundessteuer, BBl 1983 III 1 ff. (zit. Botschaft Steuerharmonisierung)

Botschaft vom 24. September 1965 des Bundesrates an die Bundesversammlung über das Verwaltungsverfahren, BBl 1965 1348 ff. (zit. Botschaft VwVG)

CAHAI (Ad hoc Committee on Artificial Intelligence) (2020), Towards Regulation of AI Systems, Global perspectives on the development of a legal framework on Arificial Intelligence (AI) systems based on the Council of Europe's standards on human rights, democracy and the rule of law, abrufbar unter https://rm.coe.int/prems-107320-gbr-2018-compli-cahai-couv-texte-a4-bat-web/1680a0c17a (zit. CAHAI 2020)

Cities for Digital Rights (2020), Declaration of Cities Coalition for Digital Rights, abrufbar unter https://citiesfordigitalrights.org/declaration (zit. Cities for Digital Rights, 2020)

Council of Europe (2020), European Ethical Charter on the Use of Artificial Intelligence in Judicial Systems and Their Environment, Strasbourg 2020, abrufbar unter https://rm.coe.int/ethical-charter-en-for-publication-4-december-2018/16808f699c (zit. Council of Europe, 2020, Ethical Charter AI)

Council of Europe (2020), Recommendation CM/Rec(2020)1 of the Committee of Ministers to Member States on the Human Rights Impacts of Algorithmic Systems, Strasbourg 2020, abrufbar unter https://rm.coe.int/09000016809e1154 (zit. Council of Europe, CM/Rec (2020)1)

Dataethical Thinkdotank (2021), White Paper: Data Ethics in Public Procurement, abrufbar unter https://dataethics.eu/publicprocurement/ (zit. Dataethical Thinkdotank, 2021)

Datenethikkommission (2019), Gutachten der Datenethikkommission der Bundesregierung, Berlin, abrufbar unter https://www.bmi.bund.de/SharedDocs/downloads/DE/publikationen/themen/it-digitalpolitik/gutachten-datenethikkommission.pdf?__blob=publicationFile&v=6

Botschaft vom 25. Mai 1983 zu den Bundesgesetzen über die Harmonisierung der direkten Steuern der Kantone und Gemeinden sowie über die direkte Bundessteuer, BBl 1983 III 1 ff. (zit. Botschaft Steuerharmonisierung)

Government Digital Service and Office for Artificial Intelligence UK (2019), A Guide to Using Artificial Intelligence in the Public Sector/Understanding Artificial Intelligence Ethics and Safety, abrufbar unter https://www.gov.uk/guidance/understanding-artificial-intelligence-ethics-and-safety (Government UK, 2019)

Government of Canada and Treasury Board Secretariat (2019), Directive on Automated Decision Making, abrufbar unter https://assets.mofoprod.net/network/documents/Using_procurement_instruments_to_ensure_trustworthy_AI.pdf (zit. Government of Canada, 2019)

Government of New Zealand (2020), Algorithm Charter for Aotearoa New Zealand, abrufbar unter https://data.govt.nz/use-data/data-ethics/government-algorithm-transparency-and-accountability/algorithm-charter (zit. Government of New Zealand, 2020)

Independent High-Level Expert Group On Artificial Intelligence Set Up By The European Commission (2019), Ethics Guidelines for Trustworthy AI (European Commission - Digital Single Market), abrufbar unter https://ec.europa.eu/digital-single-market/en/news/ethics-guidelines-trustworthy-ai (zit. Ethics Guidelines for Trustworthy AI)

OECD (2019), Artificial Intelligence in Society, OECD Publishing, Paris, abrufbar unter https://www.oecd-ilibrary.org/docserver/eedfee77-en.pdf?expires=1610099411&id=id&accname=ocid195445&checksum=5D406E9B-3F37E2B2CA6328E248A2F9AE (zit. OECD, 2019)

SBFI (2019), Bericht der interdepartementalen Arbeitsgruppe «Künstliche Intelligenz» an den Bundesrat über die Herausforderungen der Künstlichen Intelligenz (zit. Bericht Herausforderungen 2019)

Schweizerische Eidgenossenschaft (2020), der Bundesrat, Leitlinien «Künstliche Intelligenz» für den Bund, Orientierungsrahmen für den Umgang mit künstlicher Intelligenz in der Bundesverwaltung, vom 25. November 2020, abrufbar unter https://www.admin.ch/gov/de/start/dokumentation/medienmitteilungen.msg-id-81319.html (zit. Bundesrat Leitlinien KI 2020)

Secrétariat du Grand Conseil, Projet présenté par le Conseil d'Etat Date de dépôt: 29 août 2018 , PL 12386 Projet de loi, abrufbar unter https://ge.ch/grandconseil/data/texte/PL12386.pdf (besucht am: 1. Januar 2021) (zit. PL 12386 Projet de loi)

TA-SWISS KI (2020), Wenn Algorithmen für uns Entscheiden: Chancen und Risiken der Künstlichen Intelligenz, TA-SWISS Publikationsreihe (Hrsg.): TA 72/2020, Zürich (zit. TA-SWISS KI, 2020)

World Economic Forum (2020), AI Government Procurement Guidelines, abrufbar unter http://www3.weforum.org/docs/WEF_AI_Procurement_in_a_Box_AI_Government_Procurement_Guidelines_2020.pdf (zit. WEF, 2020)