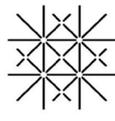




Universität
Zürich ^{UZH}



Universität
Basel

Juristische
Fakultät

WHITE PAPER

Transparenz durch Begründung von Verfügungen

Nadja Braun Binder
Liliane Obrecht

Juni 2024

CENTER FOR
INFORMATION
TECHNOLOGY
SOCIETY AND
LAW — ITSL

e-PIAF

electronic Public Institutions and
Administrations Research Forum

Dieses White Paper wurde im Projekt «**Nachvollziehbare Algorithmen: ein Rechtsrahmen für den Einsatz von Künstlicher Intelligenz**» entwickelt, das vom Center for Information Technology, Society, and Law (ITSL) der Universität Zürich und von der Forschungsstelle electronic Public Institutions and Administrations Research Forum (e-PIAF) der Universität Basel durchgeführt und von der Stiftung Mercator finanziell unterstützt wird. Dieses White Paper ist Teil einer Reihe von White Papers, die sich mit den zentralen Herausforderungen befassen, die mit dem Einsatz von Künstlicher Intelligenz (KI) in Unternehmen und in der Verwaltung verbunden sind.

Die White Papers und weitere Informationen zum Projekt finden sich auf www.itsl.uzh.ch und www.ius.unibas.ch/e-piaf.

Folgende White Papers sind bislang erschienen:

- **Manipulation**
- **Diskriminierung**
- **Datenschutz**
- **Transparenz durch Begründung von Verfügungen**
- **Transparenz durch öffentliches Verzeichnis**

Das Kernprojektteam bestand aus folgenden Personen:

Prof. Dr. Florent Thouvenin, Professor für Informations- und Kommunikationsrecht an der Universität Zürich, Vorsitzender des Lenkungsausschusses des ITSL

Prof. Dr. Nadja Braun Binder, MBA, Professorin für öffentliches Recht an der Universität Basel

Dr. Stephanie Volz, wissenschaftliche Geschäftsführerin des ITSL und Lehrbeauftragte an der Universität Zürich

Liliane Obrecht, MLaw, Wissenschaftliche Mitarbeiterin und Doktorandin an der Juristischen Fakultät der Universität Basel

Verfügungen sind in der Regel zu begründen. Werden beim Erlass von Verfügungen KI-Systeme oder algorithmische Entscheidungssysteme eingesetzt, kann dies für die Begründung eine Herausforderung sein, da je nach technischer Ausgestaltung des Systems dessen Ergebnisse nicht nachvollziehbar sind. Entsprechend kann der Einsatz solcher Systeme dazu führen, dass Verfügungen nicht oder nur mit zusätzlichen Vorkehrungen begründet werden können. Unter Berücksichtigung der Funktionen der staatlichen Begründungspflicht skizziert das vorliegende White Paper konkrete Herausforderungen und zeigt mögliche Lösungsansätze auf.

Problemstellung

Von verschiedenen Seiten, namentlich der Politik, von zivilgesellschaftlichen Organisationen, aber auch der Wissenschaft, wird für den Einsatz von Künstlicher Intelligenz (KI) bzw. von algorithmischen Entscheidungssystemen in der öffentlichen Verwaltung Transparenz gefordert. Dazu zählt die Forderung nach einer allgemeinen, gesellschaftlich breit angelegten Transparenz darüber, welche Behörde zu welchen Zwecken solche Systeme einsetzt. Dies kann (insbesondere) durch ein entsprechendes öffentliches Verzeichnis erfüllt werden (siehe White Paper «Transparenz durch öffentliches Verzeichnis»).

Transparenz wird ferner mit Blick auf konkrete Einzelfälle gefordert, die mithilfe eines algorithmischen bzw. KI-Systems entschieden werden. Der Begriff algorithmische Systeme zielt nicht auf eine bestimmte Technologie oder Methode ab, sondern erfasst alle Anwendungen, die auf (Computer-)Algorithmen basieren. Er ist insofern für die Anliegen des vorliegenden White Papers besser geeignet als der Begriff KI-System, als er auch Technologien umfasst, die die gleichen Auswirkungen auf Einzelne und die Gesellschaft haben wie KI-Systeme.

Die von einer **algorithmenbasierten Entscheidung** betroffene Person soll diese Entscheidung nachvollziehen können. Diese Forderung findet ihre Basis für Verwaltungsverfahren u. a. in den allgemeinen verfassungsrechtlichen Verfahrensgarantien gemäss Art. 29 Bundesverfassung (BV, SR 101). Dazu zählt insbesondere

die **Begründungspflicht**, der im vorliegenden Kontext eine zentrale Bedeutung zukommt.

Mit «Entscheidungen» sind im Kontext von Verwaltungsverfahren «Verfügungen» gemeint. Unter **algorithmenbasierten Verfügungen** werden Verfügungen verstanden, die mithilfe von algorithmischen Systemen vorbereitet und/oder erlassen werden. Im Folgenden geht es demnach um die Bedeutung und Umsetzbarkeit der Begründungspflicht bei algorithmenbasierten Verfügungen.

Begründungspflicht und Begründungsdichte

Als Teilaspekt des rechtlichen Gehörs (Art. 29 Abs. 2 BV) ist die Begründungspflicht staatlicher Entscheidungen Bestandteil der allgemeinen verfassungsrechtlichen Verfahrensgarantien. Das rechtliche Gehör wird im anwendbaren Verfahrensrecht weiter konkretisiert. Für das erstinstanzliche Verwaltungsverfahren auf Bundesebene sind die Bestimmungen des Verwaltungsverfahrensgesetzes des Bundes (VwVG, SR 172.021) massgebend. Dort wird zunächst bestätigt, was bereits aus der Verfassung hervorgeht: Verfügungen sind grundsätzlich zu begründen (Art 35 Abs. 1 VwVG). Dies gilt allerdings nicht, wenn die Behörde den Begehren der Partei(en) vollständig entspricht und die Partei keine Begründung verlangt (Art. 35 Abs. 3 VwVG). Durch die Begründung soll die betroffene Person nachvollziehen können, weshalb sich die Behörde im konkreten Fall so und nicht anders entschieden hat.

Die von der Entscheidung betroffene Person hat Anspruch auf eine **inhaltlich und umfangmässig angemessene Begründung**. Obwohl die Angemessenheit nur anhand des konkreten Einzelfalls beurteilt werden kann, haben Lehre und Rechtsprechung Kriterien entwickelt, die zur Bestimmung der Begründungsdichte herangezogen werden können:

- Intensität des Eingriffs: Je stärker die Entscheidung Rechte und Pflichten der betroffenen Person berührt, desto detaillierter muss die Begründung sein.
- Komplexität der Sach- und Rechtslage: Je komplexer die Sach- und Rechtslage, desto schwieriger ist es für die betroffene Person, die Subsumtion (Unterordnung) des Sachverhalts unter die anwendbaren Normen und damit die behördliche Entscheidung zu verstehen. Eine gesteigerte Begründungsdichte kann dies kompensieren.
- Entscheidungsspielraum der Behörde: Obliegt der entscheidenden Behörde kein Entscheidungsspielraum, genügt regelmässig ein Verweis auf die angewendeten Rechtsnormen. Je grösser allerdings der Entscheidungsspielraum ist, desto detaillierter ist die Entscheidung zu begründen.
- Abweichung von einer behördlichen Praxis: Weicht die Behörde in einem Fall von einer konstanten Praxis der Gesetzesanwendung ab, bedarf es einer detaillierten Begründung.
- Verfügung der Massenverwaltung: In Verwaltungsbereichen, in denen eine Vielzahl an gleichgelagerten Verfügungen ergeht, sind im Sinne der Prozessökonomie geringere Anforderungen an die Begründungsdichte zulässig.

Herausforderungen bei algorithmenbasierten Verfügungen

Die Herausforderungen, die sich bei der Begründung einer algorithmenbasierten Verfügung stellen, können je nach zugrundeliegendem Modell unterschiedlich sein. Dabei kann zwischen regelbasierten und datenbasierten Algorithmen unterschieden werden.

Regelbasierte Algorithmen verarbeiten eine Eingabe nach durch natürliche Personen festgelegten Regeln, um zu einem Ergebnis zu kommen. Sie sind deterministisch, verarbeiten daher eine bestimmte Ein-

gabe immer nach denselben Abläufen. Die Algorithmen sind dabei auch determiniert, eine bestimmte Eingabe erzeugt immer dieselbe Ausgabe. Dies führt dazu, dass die Reihenfolge der einzelnen Schritte zu jedem Zeitpunkt bestimmt ist, was grundsätzlich bewirkt, dass das Zustandekommen des Ergebnisses nachvollziehbar ist. Dies bedeutet allerdings nicht, dass die betroffene Person die Entscheidung ohne Weiteres nachvollziehen kann. Dies trifft insbesondere dann zu, wenn die Grundlage des algorithmischen Entscheidungssystems komplexe Algorithmen bilden, wie etwa vielfach verzweigte Entscheidungsbäume. Die Begründung kann entsprechend herausfordernd sein.

Hingegen verarbeiten **datenbasierte machine learning (ML)-Algorithmen** eine Eingabe nicht anhand vorgegebener Regeln, sondern analysieren i. d. R. enorme Datensätze (Trainingsdaten) auf Korrelationen, um selbst eine verallgemeinerbare Regel aufzustellen. Diese kann sodann auf andere (unbekannte) Eingaben angewendet werden. Korrelation bedeutet, dass zwei Phänomene statistisch miteinander zusammenhängen, was allerdings nichts über die Ursache-Wirkungs-Beziehung (Kausalität) aussagt. Die eingesetzten Algorithmen führen keine im Vorhinein vom Menschen definierten Regeln aus, sondern finden selbständig anhand der zur Verfügung stehenden Daten einen Lösungsweg. Solche Algorithmen sind also wahrscheinlichkeitsbasiert. Daneben können auch Zufälle den Prozess prägen. Dies führt dazu, dass die gleiche Eingabe nicht immer dasselbe Resultat erzeugt.

Das Zustandekommen der Ergebnisse von datenbasierten ML-Algorithmen ist somit oft – je nach Komplexität der angewendeten Technologie – nicht nachvollziehbar. Dies betrifft etwa künstliche neuronale Netze (KNN), eine Klasse von ML-Verfahren, die aktuell besonders viel Aufmerksamkeit auf sich zieht und auch in der öffentlichen Verwaltung als vielfältig einsetzbar gehandelt wird. Auch in solchen Fällen stellt die Herstellung der Nachvollziehbarkeit der algorithmenbasierten Entscheidung im Rahmen ihrer Begründung eine besondere Herausforderung dar.

Auswirkungen auf die Funktionen der Begründung

Die Begründung von Verfügungen dient verschiedenen Funktionen. Eben diese Funktionen sind beim Einsatz von Algorithmen schwieriger zu erfüllen.

Rechtsmittelfunktion

Die Begründung muss es der betroffenen Person erlauben, die Verfügung sachgerecht anzufechten. Insbesondere muss in verständlicher Weise aus der Begründung hervorgehen, welche **tatsächlichen und rechtlichen Erwägungen** zur Entscheidung geführt haben (Kausalität zwischen der rechtlichen Handlung und der erlassenen Verfügung).

Datenbasierte ML-Algorithmen: Es ist möglich, dass sich die Ergebnisse von datenbasierten ML-Algorithmen gerade nicht auf kausale Beziehungen, sondern auf Korrelationen zwischen den einzelnen Daten stützen. Es ist diesfalls kaum darzulegen, welche Sachverhaltsdaten und Rechtsgrundlagen ausschlaggebend für das Ergebnis des Systems im Einzelfall waren.

Regelbasierte Algorithmen: Insbesondere bei komplexen regelbasierten Algorithmen reicht es nicht aus, die einzelnen Rechenschritte aufzulisten. Es bedarf zur Gewährleistung der Rechtsmittelfunktion der Begründung einer Herausarbeitung und verständlichen Formulierung der für die Entscheidung ausschlaggebenden tatsächlichen und rechtlichen Erwägungen.

Akzeptanzfunktion

Verfügungen werden hoheitlich – d. h. einseitig – durch die zuständige Behörde erlassen. Die betroffene Person soll die Erwägungen der Behörde nachvollziehen können und wissen, weshalb sich die Behörde so und nicht anders entschieden hat. Dadurch zeigt die Behörde, dass das **Anliegen der Person gewürdigt** wurde, was die **innere Zustimmung der betroffenen Person** herbeiführen soll. Dies leistet einen massgeblichen Beitrag dazu, dass die Entscheidung nicht an die nächste Instanz weitergezogen wird, und dient somit dem Rechtsfrieden sowie der Prozessökonomie.

Datenbasierte und regelbasierte Algorithmen: Unabhängig vom eingesetzten Modell ist bei algorithmenbasierten Verfügungen sicherzustellen, dass die betroffene Person die Entscheidung versteht und nachvollziehen kann, dass bzw. wie ihre Anliegen dabei gewürdigt wurden. Die Begründung muss für einen Laien bzw. eine Laiin verständlich abgefasst sein.

Rechtssicherheitsfunktion

Die Begründung muss ersichtlich machen, welche Tatsachen bzw. Sachverhaltselemente berücksichtigt wurden und unter welche Rechtsnormen die Behörde diese subsumiert. Dies macht **staatliches Handeln vorhersehbar** und fördert eine **einheitliche Praxis**.

Datenbasierte ML-Algorithmen: Da der Output von ML-Algorithmen nicht auf kausalen Beziehungen, sondern auf Korrelationen und ggf. Zufällen beruht, besteht diesbezüglich eine grosse Diskrepanz. Wenn nicht klar ist, dass gleiche Inputs immer zum gleichen Output führen, kann nicht von einer einheitlichen Praxis und damit auch nicht von Vorhersehbarkeit ausgegangen werden.

Selbstkontrollfunktion

Behörden haben sich in Erfüllung ihrer Aufgaben an die geltende Rechtsordnung zu halten und einzelfallgerecht zu entscheiden. Die Entscheidungsfindung kann durch subjektive (z. B. emotionale Einflüsse), aber auch objektive Faktoren (z. B. Wetterlage) beeinflusst werden. Das Verfassen der Begründung kann zur Selbstreflexion der die Verfügung ausstellenden Person führen, indem sie überprüft, ob sie sich von unsachlichen Motiven hat leiten lassen.

Datenbasierte und regelbasierte Algorithmen:

Auch wenn menschliche Bearbeitungsschritte bei algorithmenbasierten Verfügungen ganz oder teilweise wegfallen, können sich gerade aufgrund der eingesetzten Daten oder Algorithmen Fehler ergeben. Über die Begründung soll auch in diesen Fällen – jedenfalls dann, wenn die algorithmenbasierte Verfügung im Rahmen eines teilautomatisierten Verfahrens ergeht – eine Selbstkontrolle vor Erlass der Verfügung ermöglicht werden. Dies bedingt, dass die entscheidungsrelevanten Faktoren des Systems erkennbar sind und durch einen Menschen nachvollzogen werden können. Die Begründung kann in diesem Fall auch als Mittel zum Zweck verstanden werden, indem z. B. allfällige Diskriminierungsprobleme sichtbar gemacht werden. Im vollautomatisierten Verfahren kann die Selbstkontrollfunktion hingegen gar nicht erfüllt werden.

Fremdkontrollfunktion

Die Begründung erlaubt einer allfälligen Rechtsmittelinstanz, die Erwägungen der verfügenden Behörde nachzuvollziehen, um sich ein sachgerechtes Bild von der Angelegenheit zu machen und prüfen zu können, ob die Vorinstanz rechtskonform entschieden hat. Dazu

benötigt sie Angaben über das Verhältnis zwischen den angewendeten Rechtsnormen und Tatsachen sowie der erlassenen Verfügung.

Datenbasierte und regelbasierte Algorithmen: Die insbesondere bei der Rechtsmittel- und Akzeptanzfunktion angeführten Herausforderungen stellen sich auch bezüglich der Fremdkontrollfunktion.

Lösungsansätze

Die Begründungspflicht verlangt nach einer Erklärung, welche Sachverhaltsinformationen und Rechtsgrundlagen im konkreten Einzelfall berücksichtigt wurden und für das Ergebnis ausschlaggebend waren. Die folgenden Lösungsansätze können dabei helfen, die Begründungspflicht umzusetzen:

Kennzeichnungspflicht

Damit die betroffene Person Kenntnis darüber hat, dass ihr Anliegen durch ein algorithmenbasiertes Entscheidungssystem bearbeitet wurde, bedarf es einer entsprechenden Kennzeichnung der Entscheidung, etwa im Sinne der Kennzeichnungspflicht für automatisierte Einzelentscheide von Bundesorganen in Art. 21 Abs. 4 Datenschutzgesetz (DSG, SR 235.1). Dies stellt zwar einen Baustein zur Erfüllung der Begründungspflicht dar, reicht allerdings für sich allein nicht aus.

Verweis auf weiterführende Informationen

Neben der Kennzeichnung können auch weitere Informationen zum Entscheidungssystem selbst geliefert werden. Dazu zählt etwa ein Verweis auf die Ergebnisse einer vorgängig durchgeführten Folgenabschätzung, die in einem öffentlich zugänglichen Verzeichnis über den Einsatz des algorithmischen Systems zu finden sind.

Wahl des Modells/der Technik

Es existieren ML-Modelle, die aufgrund ihres Designs (intrinsische Eigenschaft) nachvollziehbare Ergebnisse liefern können. So können Ergebnisse basierend auf (einfachen) Entscheidungsbäumen eher nachvollziehbar sein als jene basierend auf künstlichen neuronalen Netzen, da erkenntlich gemacht werden kann, welchen «Weg» das System eingeschlagen hat und welche Datenpunkte für die Entscheidung ausschlaggebend waren. Dementsprechend könnten für behördliche Entscheidungen nur jene ML-Techniken/Modelle erlaubt werden, die der Nachvollziehbarkeit zugänglich sind.

Menschliche Verantwortung

Von verschiedenen Seiten wird gefordert, dass im automatisierten Entscheidungsverfahren immer ein **Mensch in den Entscheidungsprozess involviert** sein muss: sei es bei der Entscheidungsunterstützung, bei der das System einem Sachbearbeiter bzw. einer Sachbearbeiterin Vorschläge zur Entscheidungsfindung unterbreitet, sei es bei der vollautomatisierten Entscheidungsfindung, bei der ein Mensch die durch das System gefällte Entscheidung – inklusive der Begründung – im Nachhinein überprüft. In beiden Konstellationen ist darauf zu achten, dass die entscheidende Person sich nicht durch Verzerrungen im System zu Fehlentscheidungen verleiten lässt bzw. Fehlentscheidungen billigt. In diesem Zusammenhang wird auch von «**Automation Bias**» gesprochen. Dabei misst der Mensch der Computerantwort mehr Bedeutung bei, als ihr eigentlich zukommen sollte. Um dies zu vermeiden, muss einerseits die Datenkompetenz der Behördenmitarbeitenden gefördert werden. Die sachbearbeitende Person muss grundsätzlich in der Lage sein, den Prozess, der zu einer Entscheidung führt, nachzuvollziehen und eine entsprechende Begründung zu formulieren. Andererseits sollen **retrospektive externe Analysen** durchgeführt werden. Denn ob sich jemand von unsachlichen Motiven oder Datenverzerrungen hat leiten lassen und dadurch (indirekte) Diskriminierungen entstanden sind, zeigt sich oft erst in der Gesamtschau der Entscheidungspraxis und nicht in einer einzelnen Verfügung. Damit erfährt die Selbstkontrollfunktion der Begründungspflicht in automatisierten Entscheidungsverfahren eine gewisse Kompensation.

Reduktion der Begründungsdichte?

In Verwaltungsbereichen, in denen eine Vielzahl an Entscheidungen bezüglich gleichartiger Fälle ergeht, stellen Lehre und Rechtsprechung reduzierte Anforderungen an die Begründungsdichte. Dies ist auf die weitgehende Standardisierung des Verfahrens zurückzuführen. Zentral ist dabei, dass gleiche Inputs stets zu gleichem Output führen. Die Kausalität des Entscheidungsprozesses ist gewährleistet. Für regelbasierte Algorithmen könnte diese Praxis allenfalls analog angewendet werden, da sie ebenfalls standardisierte Abläufe ausführen. Allerdings müsste auf Verlangen der von der Entscheidung betroffenen Person trotzdem eine Begründung geliefert werden können. Die Behördenmitarbeitenden müssten also dennoch in der Lage sein, eine Begründung zu verfassen. Demgegenüber beruhen datenbasierte ML-Algorithmen regelmässig auf Korrelationen. Im Lichte der verfassungsrechtlichen Anforderungen

und der bestehenden Praxis im Bereich der Massenverwaltung scheint es deshalb nicht gerechtfertigt, beim Einsatz datenbasierter ML-Algorithmen die Begründungsanforderungen zu reduzieren. Gegebenenfalls ist jedoch ein **Paradigmenwechsel** denkbar. An die Stelle der Nachvollziehbarkeit und Begründung der Entscheidung im Einzelfall könnte eine **retrospektive**

Systemkontrolle und Analyse der **Entscheidungspraxis** treten (z. B. durch einfache Zufallsstichproben, «random sampling»). Die Akzeptanzfunktion kann dadurch allerdings nur kompensiert werden, wenn die Bevölkerung den automatisierten Verfahren vertraut. Auch ist ein solches Kontrollsystem ohnehin nur in nicht sensiblen Bereichen denkbar.

Impressum

© 2024

Herausgeberin:
Center for Information Technology,
Society, and Law (ITSL)
Universität Zürich
Rämistrasse 74|38
8001 Zürich