



Digital
Autonomy Hub
Technik souverän nutzen

POLICY BRIEF #3

Juli 2021

Tools zur Folgenabschätzung

Automatisierte
Entscheidungsfindungssysteme
im öffentlichen Sektor

INHALT

TOOLS ZUR FOLGENABSCHÄTZUNG 4

AUSWIRKUNGEN FÜR DIE
ÖFFENTLICHE VERWALTUNG 6

POLITIKEMPFEHLUNGEN 7

Vom maschinell erstellten Steuerbescheid über Chatbots bis hin zur „vorausschauenden Polizeiarbeit“ (predictive policing) – Systeme zum automatisierten Entscheiden (im Folgenden „ADM-Systeme“, automated decision-making systems) sind in der öffentlichen Verwaltung angekommen.¹ In unserem Policy Brief legen wir dar, warum ihr Einsatz in der Verwaltung besondere Herausforderungen mit sich bringt und wie ihnen begegnet werden kann. Dazu stellen wir ein Instrument vor, mit dem die Folgen von ADM-Systemen eingeschätzt werden können und zeigen, wie es verwendet wird. Zudem machen wir deutlich, welchen Nutzen es sowohl für die Behörden als auch für die Betroffenen – also uns alle – haben kann.

Behörden erhoffen sich vom Einsatz von ADM-Systemen, einfache Arbeitsprozesse zu automatisieren, Personal zu entlasten, generell die Effizienz zu steigern und die Qualität der Dienstleistungen zu verbessern. Automatisierung bietet zudem die Gelegenheit, bestehende Prozesse allgemein zu reflektieren. Nicht zuletzt möchten Behörden durch die Automatisierung stärker von den großen Datenmengen profitieren, die bei der Verwaltungstätigkeit bereits jetzt oder potenziell anfallen.

Die öffentliche Verwaltung hat eine gesellschaftliche Sonderstellung, da sie meistens die einzige Anbieterin bestimmter Dienstleistungen ist. Betroffene sind ihr daher alternativlos ausgeliefert und haben oft auch

keine Möglichkeit, sich ihr zu entziehen. Gleichzeitig haben Behörden nicht nur Zugang zu besonders sensiblen Personendaten – ihre Entscheidungen haben oft auch gravierende Auswirkungen auf unser Leben. Der Einsatz von ADM-Systemen im öffentlichen Sektor bringt also besondere Risiken mit sich. Zugleich sollten Einzelne und die Gesellschaft der öffentlichen Verwaltung mit guten Gründen vertrauen können und damit auch der Automatisierung von Verwaltungsabläufen. Angesichts dieses einzigartigen Kontexts müssen die Auswirkungen von ADM-Systemen systematisch bewertet sowie Transparenz und Rechenschaftspflicht gegenüber Betroffenen gewährleistet werden. Ebenso ist eine informierte, demokratische Auseinandersetzung mit ihnen notwendig.

Eine besondere Herausforderung bei der Bewertung von ADM-Systemen liegt allerdings oft darin, dass wenig über sie bekannt ist: Wie sie funktionieren, wo und von wem sie eingesetzt werden – darüber ist oft kaum etwas zu erfahren.² Dementsprechend ist es wichtig, Transparenz zu schaffen und einzelne Systeme in ihrem Kontext zu betrachten, um Risiken zu identifizieren und gegebenenfalls die Systeme anzupassen.

Vor diesem Hintergrund hat AlgorithmWatch ein konkretes und praktikables **Instrument zur Folgenabschätzung** (Impact Assessment Tool) entwickelt, das für die Bewertung spezifischer ADM-Systeme und

1 Chiusi, Fabio; Fischer, Sarah; Kayser-Bril, Nicolas; Spielkamp, Matthias (Hg.) (2020): Automating Society Report 2020. Unter: <https://automatingsociety.algorithmwatch.org>

2 Chiusi, Fabio; Fischer, Sarah; Kayser-Bril, Nicolas; Spielkamp, Matthias (Hg.) (2020): Automating Society Report 2020, S. 119. Unter: <https://automatingsociety.algorithmwatch.org>

Ethische Prinzipien

Schadens-
vermeidung

Gerechtigkeit /
Fairness

Autonomie

Benefizienz

Instrumentelle Prinzipien

Kontrolle

Transparenz

Rechenschaftspflicht

ihrer potenziellen Risiken eingesetzt werden kann. Das Tool basiert auf einer Studie zum Einsatz von Künstlicher Intelligenz (KI) in der öffentlichen Verwaltung, die AlgorithmWatch und AlgorithmWatch Schweiz in Zusammenarbeit mit der Universität Basel im Auftrag des Kantons Zürich durchgeführt haben.³

Grundannahme ist, dass die ethischen Prinzipien der Schadensvermeidung, der Autonomie, der Gerechtigkeit und Fairness sowie der Benefizienz⁴ gewahrt werden müssen, wenn ADM-Systeme eingesetzt werden. Um das zu erreichen, müssen Transparenz, Kontrolle und Rechenschaftspflicht gewährleistet sein (s. Grafik auf S.3). Das Instrument zur Folgenabschätzung, das im Folgenden näher erläutert wird, soll den Verantwortlichen dabei helfen, diese Ziele zu erreichen, indem sie – geführt von zwei Checklisten – Anforderungen, potenzielle Risiken und mögliche Antworten bereits im Entwicklungsprozess reflektieren können.

TOOLS ZUR FOLGENABSCHÄTZUNG

Es existieren bereits Tools zur Folgenabschätzung. Oft bewerten sie allerdings nur eine Momentaufnahme der ADM-Systeme, sind in sich widersprüchlich oder es bleibt unklar, wie sie praktisch einzusetzen sind. Das hier entwickelte Instrument ermöglicht eine einzelfallbezogene und fortwährende Risikobewertung und legt dabei einen Fokus auf die Verständnis- und Nutzerfreundlichkeit. In einem ersten Schritt wird geprüft, ob und wenn ja welche Risiken von einem

³ Braun Binder, Nadja et al. (2021): Einsatz Künstlicher Intelligenz in der Verwaltung: ethische und rechtliche Fragen. Unter: <https://algorithmwatch.org/de/bwrict-ki-in-der-verwaltung-2021/>

⁴ Benefizienz bezeichnet das ethische Prinzip, so zu handeln, dass das Wohl der Betroffenen gefördert wird.

System ausgehen – die sogenannte „Triage“ von ADM-Systemen. Werden Risiken erkannt, wird im zweiten Schritt Transparenz über diese Risiken hergestellt und darüber, was die Verwaltung unternommen hat, um sie zu minimieren. Je mehr Risikosignale erkannt werden, desto umfassender und anspruchsvoller muss ein folgender Transparenzbericht gestaltet werden.

Checklisten

Die erste Checkliste (**Triage-Checkliste**) deckt die bereits erwähnten vier ethischen Prinzipien Schadensvermeidung (1), Gerechtigkeit und Fairness (2), Autonomie (3) und Benefizienz (4) ab. Anhand der Checkliste beurteilt eine Verwaltung folglich,

1. ob mithilfe eines ADM-Systems das Risiko besteht, dass Menschen geschädigt oder getäuscht werden und ob es so umgesetzt wird, dass negative Folgen minimiert werden. Das umfasst auch, dass die Systeme möglichst sicher sein müssen – also etwa gegen einen Ausfall geschützt sind oder keine privaten Daten von Betroffenen preisgeben – und ökologisch und sozial nachhaltig sind. (Schadensvermeidung)
2. ob ein ADM-System Schutz vor Diskriminierung und Verzerrung („bias“) bietet sowie geltende Rechtsnormen, individuelle Rechte und individuelle moralische Rechte achtet (auch Menschenrechte). Es soll ein hohes Maß an Gleichheit, Inklusion und Solidarität sicherstellen. Betroffene müssen einen Zugang zu Rechtsmitteln und Entschädigungen bei Rechtsverletzungen haben. Nicht zuletzt muss gerade bei KI-basierten ADM-Systemen eine prozedurale Ordnungsmäßigkeit sichergestellt sein: Sollten Modelle geändert werden – etwa auf der Grundlage neuer Daten –, müssen sie erneut überprüft werden. (Gerechtigkeit und Fairness)
3. ob ein ADM-System es ermöglicht, dass Individuen ausreichende Informationen erhalten und nicht getäuscht werden, wenn sie mit Algorithmen interagieren und ob sie die Möglichkeit haben, selbst

zu entscheiden, welche digitalen Dienste sie nutzen oder vollständig vermeiden möchten. Letztlich geht es auch darum, dass KI-basierte ADM-Systeme „unter der Kontrolle der Anwender:innen“ stehen. Das heißt, dass Algorithmen eingesetzt werden sollten, um die menschliche Entscheidungsfindung zu unterstützen und nicht, um sie gänzlich zu ersetzen. (Autonomie)

4. ob der Einsatz eines ADM-Systems mit einem konkreten Nutzen verbunden ist. (Benefizienz)

Die Antworten auf die Fragen der Triage-Checkliste bestimmen zunächst, ob und in welchem Umfang weitere Transparenzfragen im Detail zu beantworten sind. Das Impact Assessment Tool hat daher zwei Ebenen:

- Wenn nur zu wenigen Aspekten Transparenz hergestellt werden muss, kann es ausreichend sein, nur die Antworten auf die Triage-Checkliste und deren Begründungen zu dokumentieren.
- Wenn über viele Aspekte Transparenz hergestellt werden muss, ist es notwendig, einen detaillierten Transparenzbericht (anhand von Checkliste 2) zu erstellen. Durch den Prozess führt ein Flussdiagramm.

Der **schriftliche Transparenzbericht** veranschaulicht dann, dass die wichtigsten ethischen Fragen nicht nur erkannt wurden, sondern auch, wie sie unter menschliche Kontrolle gebracht wurden und dass eine angemessene Rechenschaftspflicht für den Prozess sichergestellt ist. In der Praxis bedeutet das: Je mehr Risikosignale in den Antworten auf die Checklisten identifiziert werden, desto umfassender gestaltet sich der Transparenzbericht.

Wenn die Verwaltung nicht in der Lage ist, ein angemessenes Maß an Transparenz über diese ethischen

Fragen zu schaffen oder wenn der Transparenzbericht aufzeigt, dass das ADM-System mit den Prinzipien der Behörde, die das Projekt verantwortet, unvereinbar ist, sollte dies dazu führen, dass das Projektziel überdacht wird oder dass mehr Ressourcen in die Suche nach einer praktikablen Lösung investiert werden.



Ein Beispiel

Anhand des fiktiven ADM-Systems „Swiss COMPAS-System“ lässt sich die Anwendung des Impact Assessment Tools veranschaulichen – sowohl, wie die Triage für KI-Systeme angewendet wird (Checkliste 1), als auch, wie der Transparenzbericht umgesetzt wird (Checkliste 2).

Das Swiss-COMPAS-System soll Prognosen darüber erstellen, welche Rückfallwahrscheinlichkeit verurteilte Straftäter:innen nach ihrer Haftentlassung haben – ähnlich wie das US-amerikanische System COMPAS⁵. Die Triage-Checkliste stellt im Rahmen des Impact Assessments zum Beispiel Fragen zu der Art der erhobenen und genutzten Daten, wie etwa Daten zum Geschlecht einer Person, die durch Antidiskriminierungsgesetze geschützt sind. Im

5 Siehe unter: [https://en.wikipedia.org/wiki/COMPAS_\(software\)](https://en.wikipedia.org/wiki/COMPAS_(software))

Transparenzbericht muss die Verwaltung, die Swiss COMPAS einsetzen möchte, also Angaben dazu machen, welche Datenschutzmaßnahmen ergriffen werden oder wie gewährleistet werden kann, dass sensible Daten der Betroffenen geschützt bleiben und das System dennoch genutzt werden kann. Ebenso müsste die Verwaltung in ihrem Transparenzbericht Angaben dazu machen, inwiefern relevante Stakeholder wie beispielsweise Straftäter:innen, Anwält:innen, Familien von Straftäter:innen sowie potenzielle und vergangene Opfer durch den Einsatz des COMPAS-Systems betroffen sein können. Der Transparenzbericht würde dann die Antworten dokumentieren, die auf Grundlage der Checkliste geliefert wurden.

Adressat:innen des Transparenzberichts

Die Checklisten selbst geben keine Hinweise darauf, an wen ein solcher Transparenzbericht adressiert wird. Grund dafür ist, dass Adressat:innen all diejenigen sind, die aufgrund ihrer gesellschaftlichen Rolle berechtigt – und nicht nur verpflichtet – sind, auf diese Informationen zuzugreifen.⁶ Die Bestimmung dieser Rechtsansprüche ist eine äußerst komplexe Aufgabe, bei der unterschiedliche moralische, organisatorische und rechtliche Aspekte abgewogen werden müssen. Transparenz ist also nicht immer gleichbedeutend mit vollständiger Offenlegung gegenüber der Öffentlichkeit. In bestimmten Fällen können berechtigte Interessen dagegensprechen, der Öffentlichkeit vollen Zugang zu Transparenzberichten zu gewähren. In diesen Fällen ist jedoch anzugeben, welchem Gremium – etwa welcher Aufsichtsinstanz – der Transparenzbericht offengelegt wurde.

⁶ Loi, Michele; Spielkamp, Matthias (2021): Towards Accountability in the Use of Artificial Intelligence for Public Administrations. Unter: <https://algorithmwatch.org/en/accountability-in-the-use-of-ai-for-public-administrations/>



AUSWIRKUNGEN FÜR DIE ÖFFENTLICHE VERWALTUNG

Das Ziel der Folgenabschätzung ist kein quantifizierbarer Risikograd – etwa eine Punktzahl –, der zum Vergleich verschiedener ADM-Systeme dient. Vielmehr soll sie ein Werkzeug dafür sein, aufgrund einer Einzelfallbeurteilung auf Risikosignale reagieren zu können. Sie kann für alle ADM-Systeme im öffentlichen Sektor verwendet werden. Je weniger Risikosignale festgestellt werden, umso geringer sind die Dokumentationspflichten im Transparenzbericht und somit auch der Aufwand. Im Umkehrschluss wird es anspruchsvoller für die öffentliche Verwaltung, ein ADM-System zu nutzen, wenn es sich als risikoreich erweist. Die Verwaltung selbst erhält mit dem Instrument zur Folgenabschätzung ein Werkzeug, mit dem sie bereits im Entwicklungsprozess Risiken erkennen und minimieren kann. Das kann sowohl die Qualität erhöhen, als auch dazu beitragen, Probleme früh zu erkennen und damit spätere Konflikte – mit Politik, Betroffenen, Aufsichtsbehörden oder auch aufgrund kritischer Medienberichte – zu vermeiden.

Mit dem Impact Assessment Tool wird nicht nur ein Beitrag zur Debatte über den ethischen Einsatz von ADM-Systemen geleistet, sondern gleichzeitig auch eine praktikable Lösung angeboten, um diesen ethischen Einsatz sicherzustellen. Letztlich trägt dies dazu bei, dass ADM-Systeme zum Nutzen von Individuen und Gesellschaft – und nicht zu ihrem Schaden – eingesetzt werden.

POLITIK- EMPFEHLUNGEN

Aus der Analyse ergeben sich folgende Empfehlungen an die Gesetzgebenden und die öffentliche Verwaltung:⁷

- Wenn ADM-Systeme im öffentlichen Sektor eingesetzt werden, sollten Behörden dazu verpflichtet werden, die potenziellen Risiken ihres Einsatzes systematisch zu bewerten und transparent zu machen. Diese Risiken können nicht pauschal, sondern nur durch eine Einzelfallanalyse ermittelt werden. Daher sollten die Behörden verpflichtet sein, **vor** und **während** der Einführung eines ADM-Systems eine Folgenabschätzung durchzuführen.
- Das hier entwickelte zweistufige **Verfahren zur Folgenabschätzung** bietet ein praktikables und sofort einsetzbares Werkzeug, um Transparenz über potenzielle Risiken zu schaffen, basierend auf sieben grundlegenden ethischen Prinzipien. Es ermöglicht eine Triage von ADM-Systemen und zeigt an, ob ein bestimmtes System zusätzliche Transparenzanforderungen erfüllen muss. Ist dies der Fall, sollten die Behörden sicherstellen, dass ein umfassender Transparenzbericht erstellt wird, der es ermöglicht, das System und seinen Einsatz über den gesamten Lebenszyklus hinweg zu bewerten. Transparenz allein garantiert noch nicht die Konformität mit ethischen Anforderungen, ist aber eine notwendige Bedingung, um diese Konformität zu erreichen.
- Es sollte ein **öffentliches Register** für alle ADM-Systeme geben, die im öffentlichen Sektor eingesetzt werden. Neben Informationen zum Zweck, zum zugrundeliegenden Modell und zu den an Entwicklung und Anwendung beteiligten Akteur:innen, sollte dieses Register auch die **Ergebnisse der Folgenabschätzung** enthalten – also gegebenenfalls den Transparenzbericht. In Fällen, in denen es legitime Gründe dafür gibt, den Zugriff auf den Transparenzbericht zu beschränken, sollte genannt werden, welche Gremien den Transparenzbericht einsehen dürfen.

⁷ Loi, Michele in Zusammenarbeit mit Mätzener, Anna; Müller, Angela; Spielkamp, Matthias (2021): Automated Decision-Making Systems in the Public Sector – An Impact Assessment Tool for Public Authorities, S. 2. Unter: <https://algorithmwatch.org/en/adms-impact-assessment-public-sector-algorithmwatch/>



Digital Autonomy Hub

Technik souverän nutzen

Der *Digital Autonomy Hub – Technik souverän nutzen* ist ein Kompetenzzentrum, das ein interdisziplinäres Netzwerk von 43 Instituten und Organisationen koordiniert. Der Hub macht sichtbar, woran die Partner forschen und welche Ideen sie entwickeln, um die individuelle digitale Souveränität zu stärken. Ziel dieses Wissenstransfers ist es, allen Menschen einen reflektierten und selbstbestimmten Umgang mit ihren Daten, Geräten und Anwendungen zu ermöglichen. Das Kompetenzzentrum bereitet aktuelle Forschungsergebnisse für Zivilgesellschaft, Politik, Wissenschaft und Wirtschaft auf und berät die verschiedenen Akteure zu ethischen, rechtlichen und sozialen Aspekten der Datennutzung.

Der *Digital Autonomy Hub* wird vom Bundesministerium für Bildung und Forschung im Rahmen des Forschungsprogramms „Technik zum Menschen bringen“ gefördert und von AlgorithmWatch und Gesellschaft für Informatik e.V. (GI) umgesetzt.

Mehr Informationen unter: www.digitalautonomy.net

Anwendungen zur Folgenabschätzung: Automatisierte Entscheidungsfindungssysteme im öffentlichen Sektor

Policy Brief #3
des Digital Autonomy Hubs
Juli 2021

Autor-innen:

Sarah Michot, Junior Advocacy & Campaign
Managerin, AlgorithmWatch

Matthias Spielkamp, Mitgründer &
Geschäftsführer, AlgorithmWatch

Korrektorat:

Marlene Kienberger

Layout:

Beate Auring

Veröffentlicht von

AW AlgorithmWatch gGmbH

Linienstr. 13

10178 Berlin

Gesellschaft für Informatik e.V. (GI)

Spreepalais am Dom

Anna-Louisa-Karsch-Straße 2

10178 Berlin

Kontakt:

info@digitalautonomy.net

Der Digital Autonomy Hub
wird gefördert vom



Bundesministerium
für Bildung
und Forschung

im Rahmen des Forschungsprogramms
„Technik zum Menschen bringen“



Diese Veröffentlichung ist unter einer Creative Commons Namensnennung
4.0 International Lizenz lizenziert

<https://creativecommons.org/licenses/by/4.0/legalcode.de>